

Contemporary  
Concepts  
in Physics

Volume 3

**A.M. Polyakov**

# Gauge Fields and Strings



# **Gauge Fields and Strings**

# **Contemporary Concepts in Physics**

A series edited by  
**Herman Feshbach**  
Massachusetts Institute  
of Technology

Associate Editors  
**N. Bloembergen**  
Harvard University

**Leo Kadanoff**  
University of Chicago

**Mal Ruderman**  
Columbia University

**S. B. Treiman**  
Princeton University

Founding Editor  
**Henry Primakoff**  
(1914–1983)

**Volume 1**  
**Particle Physics  
and Introduction  
to Field Theory**  
T.D. Lee

**Volume 2**  
**Particle Physics**  
**The Quest for the  
Substance of Substance**  
L.B. Okun

**Volume 3**  
**Gauge Fields  
and Strings**  
A.M. Polyakov

Additional volumes in preparation

# Gauge Fields and Strings

by

**A.M. Polyakov**

L.D. Landau Institute  
for Theoretical Physics  
USSR Academy of Sciences,  
Moscow

 **Routledge**  
Taylor & Francis Group  
LONDON AND NEW YORK

*Open Access funded by SCOAP3*

*Copyright 2021 A.M. Polyakov*

*This eBook was converted to open access in 2021 through the sponsorship of SCOAP3 licensed under the terms of the creative commons Attribution-NonCommercial 4.0 International License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits use, sharing, adaptation distribution and reproduction in any medium or format, as long as you give appropriate credit to the author(s) and the source, provide a link to the creative commons license and indicate if changes were made, this license does not permit the Contribution to be used commercially.*

First published 1987 by Harwood Academic Publishers GmbH

This edition published 2021 by Routledge

2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN  
711 Third Avenue, New York, NY 10017

*Routledge is an imprint of the Taylor & Francis Group, an informa business*

The Open Access version of this book, available at [www.taylorfrancis.com](http://www.taylorfrancis.com), has been made available under a Creative Commons Attribution-Non Commercial 4.0 International.

#### **Library of Congress Cataloging-in-Publication Data**

**Polyakov, A. M.**

**Gauge fields and strings.**

**(Contemporary concepts in physics; v. 3)**

**Includes index.**

**1. Gauge fields (Physics) 2. String models.**

**1. Title. II. Series.**

**QC793.3.F5P66 1987**

**530.1'43**

**87-151**

**ISBN 3-7186-0393-4**

**ISBN 3-7186-0392-6 (soft)**

DOI: 10.1201/9780203755082

# CONTENTS

PREFACE TO THE SERIES	viii
PREFACE	ix
CHAPTER 1 STATISTICAL MECHANICS AND QUANTUM FIELD THEORY	1
1.1 Quantum Particles	1
1.2 Global and Local Symmetries. Preliminary Description	4
1.3 Discrete Global Symmetries	5
1.4 Continuum Abelian Global Symmetries	11
1.5 Non-Abelian Global Symmetries	13
1.6 Discrete Gauge Symmetries	14
1.7 $O(2)$ Gauge Systems	15
1.8 Non-Abelian Gauge Theories	17
CHAPTER 2 ASYMPTOTIC FREEDOM AND THE RENORMALIZATION GROUP	19
2.1 Principal Chiral Fields	19
2.2 The $n$ -Fields	27
2.3 Non-Abelian Gauge Fields for $\mathcal{D} = 4$	29
CHAPTER 3 THE STRONG COUPLING EXPANSION	33
3.1 Ising Model	34
3.2 Continuous Global Symmetry	37
3.3 Gauge Symmetries	40
CHAPTER 4 INSTANTONS IN ABELIAN SYSTEMS	49
4.1 Instantons in Quantum Mechanics and the Ising model	49
4.2 Instantons in the Global $O(2)$ Model	54
4.3 Compact QED ( $O(2)$ Gauge Model)	62

CHAPTER 5	QUARK CONFINEMENT, SUPERFLUIDITY, ELASTICITY. CRITERIA AND ANALOGIES	73
CHAPTER 6	TOPOLOGY OF GAUGE FIELDS AND RELATED PROBLEMS	85
6.1	Instantons for $\mathcal{D} = 2, N = 3n$ -Fields	85
6.2	Instantons in Non-Abelian Gauge Theories	92
6.3	Qualitative Effects of Instantons	99
CHAPTER 7	ANALOGIES BETWEEN GAUGE AND CHIRAL FIELDS. LOOP DYNAMICS	111
7.1	Non-Abelian Phase Factor	111
7.2	Quantum Theory of Loops	119
CHAPTER 8	THE LARGE $N$ EXPANSION	125
8.1	$O(N)$ $\sigma$ -Model	125
8.2	The Principal Chiral Field for $SU(N)$	134
8.3	The $CP^{N-1}$ - $n$ -model	139
8.4	Non-Abelian Gauge Theory	144
CHAPTER 9	QUANTUM STRINGS AND RANDOM SURFACES	151
9.1	Mathematical Preliminaries: Summation of Random Paths	151
9.2	Measures in the Space of Metrics and Diffeomorphisms	157
9.3	Closed Paths	164
9.4	General Theory of Random Hypersurfaces	169
9.5	Two-Dimensional Surfaces. Geometrical Introduction	176
9.6	Computation of Functional Integrals	185
9.7	Scattering Amplitudes	192
9.8	Scattering Amplitudes and the Operator Product Expansion	195
9.9	The Energy-Momentum Tensor in Conformal Quantum Field Theory	203
9.10	Physical States of String Theory in the Critical Dimension	212
9.11	Fermi Particles	222

9.12	Fermionic Strings	228
9.13	Vertex Operators	240
<b>CHAPTER 10</b>	<b>ATTEMPT AT A SYNTHESIS</b>	<b>253</b>
10.1	Long Wave Oscillations of Strings in Critical Dimensions	253
10.2	Possible Applications of Critical Strings	266
10.3	The Three Dimensional Ising Model	273
10.3.1	The Dirac Equation in the Two Dimensional Ising Model	275
10.3.2	The Three Dimensional Case. The Loop Equation	278
10.4	Extrinsic Geometry of Strings	283
	<b>SUBJECT INDEX</b>	<b>289</b>

## **Preface to the Series**

The series of volumes, *Concepts in Contemporary Physics*, is addressed to the professional physicist and to the serious graduate student of physics. The subjects to be covered will include those at the forefront of current research. It is anticipated that the various volumes in the series will be rigorous and complete in their treatment, supplying the intellectual tools necessary for the appreciation of the present status of the areas under consideration and providing the framework upon which future developments may be based.

## Preface

For many years I have been keeping notes on different topics in physics—a kind of scientific diary. They contain occasional new results and mostly derivations of known things, done in a way that seemed nice to me. The notes were very helpful when I needed to recall some subject. It is surely best to consult with one's own self.

This book has arisen from these notes, or better to say, from the part of them devoted to field theory. I decided to publish it because it seems that there are some people who may find it useful.

In many cases I discuss things that have never been completely understood. I do this in the hope that the approach I suggest, although imperfect, will stimulate deeper penetration into the subject.

I do not give many references in this book (except for very recent results). The reason is that although to study the history of physics and to distribute credits is an interesting enterprise, I am not yet prepared for it.

The reader can find extra information and references in many review papers, e.g. J. Kogut and K.G. Wilson, *Physics Reports*, **12**, 75–199 (1974); J. Kogut, *Reviews of Modern Physics*, **55**, 775–836 (1983); A.A. Migdal, *Physics Reports*, **102**, 199–290 (1983); Patashinsky, Pokrovsky, “Fluctuation Theory of Phase Transitions” Pergamon Press, Oxford (1979) and “Superstrings” (J. Schwarz ed.), World Scientific Pub (1985).

Also, below I list (in arbitrary order) some of my favorite papers that had a profound influence on this book. The choice is, by definition, subjective and incomplete:

A. M. Polyakov

1. A. Patashinsky and V. Pokrovsky, *Zhetph* **46**, 994 (1964).
2. V. Gribov and A. Migdal, *Zhetph* **55**, 1498 (1968).
3. V. Vaks and A. Larkin, *Zhetph* **49**, 975 (1975).
4. V. Berezinsky, *Zhetph* **61**, 1144 (1971).
5. K. Wilson, *Phys. Rev. D* **10**, 2445 (1974).
6. M. Gell-Mann and F. Low, *Phys. Rev.* **111**, 582 (1954).

7. L. Landau, A. Abrikosov, and I. Khalatnikov, *Dan.* **95**, 497 (1954).
8. L. Faddeev and V. Popov, *Phys. Lett. B* **25**, 29 (1967).
9. T. Skyrme, *Proc. Roy. Soc. London Section A* **260**, 127 (1961).
10. J. Schwinger, *Phys. Rev.* **94**, 1362 (1954).
11. M. Atiyah, V. Pathody, and I. Singer, *Math. Proc. Camb. Phys. Soc.* **77**, 43 (1975).
12. R. Jackiw and K. Rebbi, *Phys. Rev. D* **14**, 517 (1971).
13. J. Kogut and L. Susskind, *Phys. Rev. D* **11**, 395 (1975).
14. G. 'tHooft, *Phys. Rev. Lett.* **37**, 8 (1976).
15. G. 'tHooft, *Nucl. Phys. B* **72**, 461 (1974).
16. L. Brink, P. DiVecchia, and P. Howe, *Phys. Lett. B* **65**, 471 (1976).
17. S. Deser and B. Zumino, *Phys. Lett. B* **65**, 369 (1976).
18. C. Marshall and P. Raymond, *Nucl. Phys. B* **85**, 375 (1975).
19. M. Green and J. Schwartz, *Nucl. Phys. B* **255**, 93 (1985).
20. A. Migdal, *Nucl. Phys. B* **180**, 71 (1981).
21. D. Gross and F. Wilczek, *Phys. Rev. Lett.* **30**, 1345 (1973).
22. H. Politzer, *Phys. Rev. Lett.* **30**, 1346 (1973).
23. A. Zamolodchikov, *Zhetph Lett.* **17**, 28 (1986).
24. K. Wilson, *Phys. Rev.* **25**, 2003 (1969).
25. D. Amati and M. Testa, *Phys. Lett. B* **48**, 227 (1974).

---

## CHAPTER 1

# Statistical Mechanics and Quantum Field Theory

---

---

DOI: 10.1201/9780203755082-1

### 1.1 Quantum Particles

We have no better way of describing elementary particles than quantum field theory. A quantum field in general is an assembly of an infinite number of interacting harmonic oscillators. Excitations of such oscillators are associated with particles. The special importance of the harmonic oscillator follows from the fact that its excitation spectrum is additive, i.e. if  $E_1$  and  $E_2$  are energy levels above the ground state then  $E_1 + E_2$  will be an energy level as well. It is precisely this property that we expect to be true for a system of elementary particles. Therefore we attempt to identify the Hamiltonian of the particles with the Hamiltonian of coupled oscillators (there is a familiar example from solid state physics: the excitations of a crystal lattice can be interpreted as particles—phonons). All this has the flavour of the XIX century, when people tried to construct mechanical models for all phenomena. I see nothing wrong with it because any nontrivial idea is in a certain sense correct. The garbage of the past often becomes the treasure of the present (and *vice versa*). For this reason we shall boldly investigate all possible analogies together with our main problem.

A very important analogy, which will be extensively used below, is the one between the quantum mechanics of a  $\mathcal{D}$ -dimensional system and the classical statistical mechanics of a  $\mathcal{D} + 1$ -dimensional system. Let us demonstrate it in the simplest case of the  $\mathcal{D} = 1$  quantum mechanics of one particle. According to the Feynman principle, the transition amplitude  $F$  from the point  $x$  to the point  $x'$  is given by the sum over all possible trajectories connecting points  $x$  and  $x'$ , each trajectory entering with the weight  $\exp((i/\hbar)S[x(t)])$  where  $S[x(t)]$  is the

classical action. Therefore:

$$F(x, x', T) = \int_{\substack{x(0)=x \\ x(T)=x'}} \mathcal{D}x(t) \exp \left\{ \frac{i}{\hbar} \int_0^T \left[ \frac{m\dot{x}^2}{2} - v(x(t)) \right] dt \right\} \quad (1.1)$$

Here  $F$  is the amplitude,  $T$  is the time allowed for the transition,  $v(x)$  is an external potential, and the functional integral is defined in the following way. Split the interval  $[0, T]$  into  $N$  small pieces  $[0, t_1]$ ,  $[t_1, t_2], \dots, [t_{N-1}, T]$ . Consider instead of (1.1) the expression:

$$F = \int \prod_{j=1}^{N-1} dx_j \left( \frac{m}{2\pi i \hbar (t_j - t_{j-1})} \right)^{1/2} \left( \frac{m}{2\pi i \hbar (T - t_{N-1})} \right)^{1/2} \\ \times \exp \frac{i}{\hbar} \left\{ \sum_{j=1}^N \frac{m(x_j - x_{j-1})^2}{2(t_j - t_{j-1})} - \sum_{j=1}^N (t_j - t_{j-1}) v(x_{j-1}) \right\} \quad (1.2)$$

(here  $x_0 = x$ ,  $t_0 = 0$ ,  $x_N = x'$ ,  $t_N = T$ ).

Now, it is possible to show that as the mesh  $t_{j+1} - t_j \sim T/N \rightarrow 0$  the expression (1.2) has a finite limit that is precisely the transition amplitude. While I do not intend to prove it (and refer instead to the book by Feynman and Hibbs), I shall explain briefly the origin of the formulae (1.1) and (1.2). It is actually quite simple. According to standard quantum mechanics, the transition amplitude is given by†:

$$F(x, x', T) = \langle x' | e^{-(i/\hbar)HT} | x \rangle \quad (1.3)$$

where  $H$  is the Hamiltonian. We can rewrite (1.3) in the following way:

$$F(x, x', T) = \langle x' | e^{-(i/\hbar)H(T-t_{N-1})} e^{-(i/\hbar)H(t_{N-1}-t_{N-2})} \dots e^{-(i/\hbar)Ht_1} | x \rangle \\ = \int \langle x' | e^{-(i/\hbar)H(T-t_{N-1})} | x_{N-1} \rangle \langle x_{N-1} | e^{-iH(t_{N-1}-t_{N-2})/\hbar} | x_{N-2} \rangle \\ \times \dots \times \langle x_1 | e^{-(i/\hbar)Ht_1} | x \rangle dx_{N-1} \dots dx_1 \quad (1.4)$$

It is easy to check that as all time intervals  $t_{j+1} - t_j \rightarrow 0$  we obtain:

$$\langle x_{j+1} | e^{-(i/\hbar)H(t_{j+1}-t_j)} | x_j \rangle \xrightarrow{t_{j+1}-t_j \rightarrow 0} \\ (2\pi i \frac{\hbar}{m} (t_{j+1} - t_j))^{-1/2} \exp \frac{i}{\hbar} \left\{ \frac{m}{2} \frac{(x_{j+1} - x_j)^2}{t_{j+1} - t_j} - v(x_j)(t_{j+1} - t_j) \right\} \quad (1.5)$$

---

† We put  $\langle x' | x \rangle = \delta(x' - x)$

After substitution of (1.5) into (1.4) we obtain (1.2). Notice also that without the potential  $v$  the formula (1.5) is exact for any value  $t_{j+1} - t_j$  and describes the propagation of a free particle.

In order to establish the analogy with classical statistical mechanics one has to consider the propagation for imaginary time  $T$ . Namely, let us look at

$$Z(x, x', T) = \langle x' | e^{-(HT/\hbar)} | x \rangle = F(x, x', -iT) \quad (1.6)$$

We can repeat the splitting procedure again with the only difference that the  $t_j$  in (1.4) will acquire an extra factor  $-i$ . In this way we obtain:

$$Z(x, x', T) = \int_{\substack{x(0)=x \\ x(T)=x'}} \mathcal{D}x(t) \exp \left\{ -\frac{1}{\hbar} \int_0^T \left( \frac{m}{2} \dot{x}^2 + v(x(t)) \right) dt \right\} \quad (1.7)$$

which is to be understood in the same way as (1.1). The mnemonic rule for passing from (1.1) to (1.7) is very simple: consider the expression:

$$i \int_0^{iT} \left( \frac{m}{2} \dot{x}^2 - v(x(t)) \right) dt \quad (1.8)$$

and introduce  $t = -i\tau$ . We obtain:

$$(1.8) = - \int_0^T \left\{ \frac{m}{2} \left( \frac{dx}{d\tau} \right)^2 + v(x(\tau)) \right\} d\tau \quad (1.9)$$

The derivation (1.9) shows also that we have even more freedom in computing the functional integral. Namely, we can chose the splitting points  $\{t_j\}$  to lie on an arbitrary contour in the complex plane, and therefore time not only can be imaginary but also can go along some complex path (provided that the convergence condition for (1.5),  $\text{Im } \Delta t < 0$ , is satisfied). For some problems this freedom is very useful. At the moment, however, we are interested in a different aspect of all this. Namely, that formula (1.9) has an important physical interpretation. Let us consider an elastic string of length  $T$  and tension  $m$  with the ends fixed at  $x$  and  $x'$ . Suppose that this string stays in an external potential  $v(x)$ . The potential energy of such a string will be given by:

$$\mathcal{E}_{\text{pot}}[x(t)] = \int_0^T \left\{ \frac{m}{2} \left( \frac{dx}{d\tau} \right)^2 + v(x(\tau)) \right\} d\tau \quad (1.10)$$

Notice, again, that now  $\tau$  is *not* a time but the length of the elastic string. According to the Boltzmann principle, the classical partition function of the string is proportional to:

$$Z \sim \int \mathcal{D}x(t) e^{-\beta \mathcal{E}_{\text{pot}}[x(t)]} \quad (1.11)$$

( $\beta$  being the inverse temperature), (we have omitted the contribution from the kinetic energy, since in classical statistical mechanics it factors out and does not depend on  $x$  and  $x'$ ). Comparison of (1.11) and (1.9) shows the first analogy between classical statistical mechanics and quantum mechanics: *The transition amplitude for a quantum particle for the time ( $-iT$ ) is equal to the classical partition function for a string of length  $T$  computed at the value of  $\beta = 1/\hbar$ .*

The second analogy follows from the fact that the *quantum* partition function for the particle is given by  $Z_{\text{qu}} = \text{Tr } e^{-\beta H}$  and hence:

$$Z_{\text{qu}} = \int dx F(x, x, -i\beta\hbar) \quad (1.12)$$

Therefore our second rule is that *in the quantum case the inverse temperature acts as imaginary time.*

Our derivation of these analogies was technical. I feel that there are deep reasons for them, connected with the properties of space-time. Although no real explanation exists, I shall give some comments on this below, when discussing gravity. At the moment our aims are more modest—we are going to exploit these analogies in concrete problems. It is quite clear that, although we have derived everything for one particle, both of our analogies are true for an arbitrary number of degrees of freedom.

## 1.2 . Global and Local Symmetries. Preliminary Description

Elementary particles existing in nature resemble very much excitations of some complicated medium (ether). We do not know the detailed structure of the ether but we have learned a lot about effective lagrangians for its low energy excitations. It is as if we knew nothing about the molecular structure of some liquid but did know the Navier-Stokes equation and could thus predict many exciting things. Clearly, there are lots of different possibilities at the molecular level

leading to the same low energy picture. For theoretical purposes we can take any model we like if it has desirable low energy properties.

In this section we shall discuss the most fundamental symmetry properties of particle physics in context of such specially chosen models. Perhaps the most important discovery of modern particle physics is the gauge principle. According to it, all interactions in nature arise from the claim that the Lagrangian has to be invariant under local symmetry transformations, i.e. symmetry rotations that may be different at different space-time points. It is remarkable that this claim predicts the low energy structure of the Lagrangian.

The first (and most complicated) example of this phenomenon was general relativity, in which, due to the presence of the gravitational field, it is possible to perform Lorentz rotations, different at each point. The second (and easiest) example was quantum electrodynamics, in which the gauge group is abelian (the arbitrariness of the phase of the electron wave function). And lastly, we have the Yang–Mills fields, which are supposed to mediate strong and weak interactions. The study of the dynamics of gauge fields is the most important problem of modern physics.

Using the analogies described in the preceding section, we shall first examine certain classical systems, and then formulate results in the language of particle theory.

### 1.3 Discrete Global Symmetries

Let us begin with the case of global (nongauge) symmetries. The simplest example is the well-known Ising model. Its partition function is given by:

$$\begin{aligned}
 Z &= \sum_{\{\sigma_x\}} e^{-\beta \mathcal{H}[\sigma_x]} \\
 \mathcal{H}[\sigma_x] &= - \sum_{\langle x, \delta \rangle} \sigma_x \sigma_{x+\delta}
 \end{aligned} \tag{1.13}$$

Here  $x$  denotes a site of a cubic lattice,  $\delta$  is a unit vector connecting this site with one of its nearest neighbours and the variable  $\sigma_x$  is  $\pm 1$ . It is clear that the system is invariant under the  $Z_2$  group:  $\sigma_x \rightarrow -\sigma_x$ . If the dimensionality of the  $x$  space is more than 1, the system (1.13) has two different phases. In the high temperature (small  $\beta$ ) phase the  $Z_2$  symmetry is unbroken and we do not have long range order. By that I

mean that if one considers a large but finite system and fixes the value of  $\sigma_x$  at the boundary  $B$  by the condition:

$$\sigma_x|_{x \in B} = 1 \quad (1.14)$$

we have the average value of  $\langle \sigma_x \rangle$  inside the system vanishing as the size of the system goes to infinity. To prove this, let us compute the correlation function in the small  $\beta$  limit. We have:

$$\begin{aligned} \langle \sigma_0 \sigma_R \rangle &= Z^{-1} \left( \sum_{\sigma_x} e^{\beta \sum \sigma_x \sigma_{x+\delta}} \sigma_0 \sigma_R \right) \\ &\approx \beta^{|\mathbf{R}|} Z^{-1} \left( \sum_{\sigma_x} \sigma_0 (\sigma_0 \sigma_\delta) \cdots (\sigma_{R-\delta} \sigma_R) \sigma_R \right) \\ &= \beta^{|\mathbf{R}|} \end{aligned} \quad (1.15)$$

(In (1.15) we have expanded the exponent in  $\beta$  and left the lowest nonvanishing order obtained by the string of  $\beta(\sigma_x, \sigma_{x+\delta})$  along the shortest path connecting the points 0 and  $\mathbf{R}$ ). We conclude that, since the correlation length is small, being of the order  $(\log(1/\beta))^{-1}$ , the influence of the boundary condition inside the system must also be small. So, one expects that for small  $\beta$ :

$$\langle \sigma_x \rangle \sim e^{-L \log(1/\beta)} \xrightarrow{L \rightarrow \infty} 0 \quad (1.16)$$

( $L$  being the size of the system).

Now let us look at the case of large  $\beta$  (low temperature phase). The maximal contribution to (1.13) in this case will be given by the configuration with all  $\sigma_x = 1$ . The probability for a spin to flip is of the order of  $e^{-4\mathcal{D}\beta}$ , so one expects:

$$\langle \sigma_x \rangle = 1 - O(e^{-4\mathcal{D}\beta}) \quad (1.17)$$

Here  $\mathcal{D}$  is the dimensionality of space and  $2\mathcal{D}$  is equal to the number of nearest neighbours. However, (1.17) is not completely true. For  $\mathcal{D} = 1$  the entropy effects spoil the order completely for all  $\beta$ . In order to see how this happens, let us examine a one dimensional Ising chain. In the ground state all the spins point up. The general configuration can be described by marking the links that connect opposite spins. If there are  $n$  such links, then the energy factor of the system is just  $e^{-2\beta n}$  but the number of such configurations is  $2(N!/n!(N-n)!)$ . ( $N$  is the total number of links). As a result:

$$Z = \sum_n 2 \frac{N!}{n!(N-n)!} e^{-2\beta n} \quad (1.18)$$

We see that the average value of  $n$  is of the order of  $Ne^{-2\beta}$ , giving for the correlation length  $r_c \sim N/n \sim e^{2\beta}$ . Therefore, for any value of  $\beta$  the influence of the boundary conditions is negligible and the spontaneous magnetization  $\langle \sigma \rangle$  is zero. In the two dimensional case a simple and important argument due to Peierls shows that the long range order persists for large  $\beta$ . The essence of the argument is the following. Let us consider a “drop” of reversed spins plunged into the sea of “up” spins. If the boundary of the drop has length  $L$ , then the energy factor for the configuration is given by  $e^{-2\beta L}$ . At the same time the number of loops of length  $L$  that can be drawn on the lattice behaves as  $C^L$  (where  $C$  is some constant): this combinatorial result will be discussed below in great detail. Therefore, if  $\beta > \frac{1}{2} \log C$ , creation of these dissident drops is strongly suppressed, and we have long range order in our system. For  $\beta < \frac{1}{2} \log C$  we have proliferation of drops which spoil the long range order. For  $\mathcal{D} > 2$  the argument is similar. So, the conclusion is that in the case of  $Z_2$  symmetry we have a phase transition for  $\mathcal{D} \geq 2$  that separates the phases with spontaneously broken symmetry (ferromagnetic phase) and with restored symmetry (paramagnetic). We could consider in a similar way more complicated discrete groups, like  $Z_p$ ; the phase structure would be more rich in these cases, and we postpone their discussion until later.

Now, we would like to explain in more detail the relationship between the qualitative behaviour of the Ising model just discussed and quantum field theory. The statement to be proved is that, if we consider the continuum limit of the quantum field theory with the Lagrangian:

$$\begin{aligned} \mathcal{L} &= \frac{1}{2}(\hat{c}_\mu \varphi)^2 - v(\varphi) \\ v(\varphi) &= v(-\varphi) \end{aligned} \tag{1.19}$$

it will be described by the continuum limit of an Ising model, more or less independent of the detailed form of  $v(\varphi)$ . The general reason for this is that the neighbourhood of the second order phase transition, where the lattice system appears continuous because of the large correlation length, possesses remarkable universality properties. Usually a change in the lattice interaction changes the transition temperature but not the correlation functions expressed in terms of the correlation length  $r_c$ . This universality will be explained below by means of operator algebra. For the moment we shall content ourselves with a less sophisticated derivation. Let us first obtain a diagrammatic representation for the

Ising model correlation functions. We have the set of identities:

$$\begin{aligned}
Z &= \sum_{\{\sigma_x\}} \exp\left(\beta \sum_{x,x'} \mathcal{K}_{x,x'} \sigma_x \sigma_{x'}\right) \\
&= \exp\left(\beta \sum_{x,x'} \mathcal{K}_{x,x'} \frac{\partial^2}{\partial \varphi_x \partial \varphi_{x'}}\right) \sum_{\{\sigma_x\}} \exp\left(\sum_x \varphi_x \sigma_x\right) \Big|_{\varphi=0} \\
&= \exp\left(\beta \sum_{x,x'} \mathcal{K}_{x,x'} \frac{\partial^2}{\partial \varphi_x \partial \varphi_{x'}}\right) \exp\left(\sum_x \log(2 \cosh \varphi_x)\right) \Big|_{\varphi=0} \quad (1.20) \\
&\quad \left(\mathcal{K}_{x,x'} = \sum_{\delta} \delta_{x,x'+\delta}\right)
\end{aligned}$$

We have thus obtained the standard functional representation for the set of Feynman diagrams with the bare propagator  $\beta \mathcal{K}_{x,x'}$  and vertices generated by the potential  $\log(2 \cosh \varphi)$ . If we define Dyson's self-energy part  $\Sigma$  as the sum of diagrams that do not contain parts connected by just one line (one line irreducible) we have the Dyson equation (in momentum space):

$$\begin{aligned}
G(\mathbf{p}) &= \frac{1}{(\beta \mathcal{K}(\mathbf{p}))^{-1} - \Sigma(\mathbf{p})} = \frac{\beta \mathcal{K}(\mathbf{p})}{1 - \beta \mathcal{K}(\mathbf{p}) \Sigma(\mathbf{p})} \quad (1.21) \\
\Sigma(\mathbf{p}) &= \text{---} \bullet \text{---} + \text{---} \bigcirc \text{---} + \text{---} \bigcirc \text{---} + \dots
\end{aligned}$$

( $G$  is the exact propagator for the  $\varphi$ -field). For a generic value of  $\beta$ , (1.21) has singularities in  $\mathbf{p}$  for  $|\mathbf{p}| \sim 1$ , which means that the correlation length is of the order of the lattice spacing. In that case there is no rotational symmetry and no universality in the system. However, there should exist a phase transition temperature  $\beta_c$  defined by the relation:

$$1 = \beta_c \mathcal{K}(0) \Sigma(0) \quad (1.22)$$

At this point we have a singularity at  $\mathbf{p} = 0$  and power-like behaviour of the correlation functions. Expanding  $\mathcal{K}(\mathbf{p})$  in  $\mathbf{p}$  for  $|\beta - \beta_c| \ll \beta_c$  we have:

$$G(\mathbf{p}, \tau) = \frac{Z}{\mathbf{p}^2 + \tau + \Sigma(\mathbf{p}, \tau) - \Sigma(0, 0)} \quad (1.23)$$

(here  $Z \sim 1$  is a constant that will be absorbed below into a redefinition of the  $\varphi$ -field, and  $\tau \sim |(\beta - \beta_c)/\beta_c|$ ). Equation (1.23) permits us to estimate quickly the situation at the critical point. Let us take the bare Green function  $G_0(\mathbf{p}) = 1/(\mathbf{p}^2 + \tau)$  and estimate the first diagram for  $\Sigma$ :

$$\Sigma^{(1)} = \text{---} \bigcirc \text{---} \sim \int \frac{d^D \mathbf{p}}{\mathbf{p}^2 + \tau} = A + B\tau^{\varphi/2-1} \quad (1.24)$$

We see that if  $\mathcal{D} > 4$  then the contribution of  $\Sigma^{(1)}$  is negligible in comparison with  $p^2 + \tau$ ; for  $\mathcal{D} = 4$  the correction is of relative order  $\log(1/\tau)$  and for  $\mathcal{D} < 4$  we have a very large powerlike correction. Similar estimates of higher diagrams show that for  $\mathcal{D} > \mathcal{D}_{\text{cr}} = 4$  they are irrelevant. Moreover, we see that the most singular terms arise from the  $\varphi^4$ -type of interaction, and all higher powers in  $\varphi$  may be presumed to be irrelevant. This is indeed the case, as will be shown later. The detailed form of the interaction  $\mathcal{K}_{x,x'}$  is also irrelevant—we have seen that it was sufficient to keep only the  $p^2$ -term in the expansion of  $\mathcal{K}(p)$ . All these arguments are not proofs, but they give correct guidance in complex situations and are therefore worth mentioning.

We have arrived at the following statement. Let us consider a  $\mathcal{D}$ -dimensional Ising model with short ranged interaction:

$$\mathcal{E} = - \sum_{x,x'} \mathcal{K}_{x,x'} \sigma_x \sigma_{x'}$$

Take the temperature  $\beta$  close to the critical one:  $|(\beta - \beta_c)/\beta_c| \ll 1$ . Then all the correlation functions are the same as those for a field theory with Lagrangian:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \varphi)^2 + \frac{1}{2}m_0^2 \varphi^2 + \lambda_0 \varphi^4 \quad (1.25)$$

also defined in the  $\mathcal{D}$ -dimensional Euclidean space. An important point about (1.25) is that  $m_0^2$  must be chosen in such a way that the physical mass  $m_{\text{phys}}^2 = \tau \ll \Lambda^2$  (where  $\Lambda$  is a momentum cut-off). This last condition means that we are in the critical region for (1.25). The critical point itself corresponds to the value  $m_0^2 = m_{0,\text{cr}}^2$  at which  $m_{\text{phys}}^2 = 0$ . In order to obtain the continuum limit (or, in other words, to renormalize the theory) we must take the limit  $m_0^2 \rightarrow m_{0,\text{cr}}^2$  and  $\Lambda^2 \rightarrow \infty$  in such a way that  $m_{\text{phys}}^2$  remains fixed. If this is possible, we get a rotationally invariant theory that does not depend on the way it has been defined in the cut-off region.

We also see that the theory (1.25) has two phases for  $\mathcal{D} > 1$ , one with  $\langle \varphi \rangle = 0$  and another, the broken symmetry phase, with  $\langle \varphi \rangle \neq 0$ . In the case  $\mathcal{D} = 1$  we have just the example we analysed in the beginning, which is the quantum mechanics of a single particle in a potential. We know that the ground state wave function must be even, and therefore  $\langle \varphi \rangle = 0$ . This conclusion is identical to the one we found when analysing the  $\mathcal{D} = 1$  Ising model. We shall clarify this coincidence in the chapter about instantons. How does it happen that the theory with discrete variables  $\sigma_x = \pm 1$  appears to be equivalent to the one with a continuum field  $\varphi$ ? This can be understood as follows. Let us change

the lattice spacing by fixing the value of the so-called block-spins (e.g., we fix the sum of the spins occupying the vertices of each hypercube of our lattice) and then summing over configurations with those fixed values. As a result, we obtain an effective energy that depends on the block-spins  $S$ , which are no longer restricted by  $S^2 = 1$ . Repeating this transformation many times we shall eventually come to an effective action depending on a continuum field  $\varphi$ .

There is a tendency in particle physics to consider Lagrangians like (1.25) as fundamental. It seems to me that it is more appropriate to imagine some kind of  $\sigma$ -like variables at very small distances, because they carry a quintessence of the symmetry properties. This difference, however, is not noticeable at large distances, and a theory of small distances (of order the Planck length) still does not exist. The last thing about the Ising model that we need to discuss in this preliminary section is its Hamiltonian form. In order to derive it, let us split the  $\mathcal{D}$ -dimensional coordinate  $x$  into a  $\mathcal{D} - 1$  dimensional  $y$  and one "time" dimension  $t$ :  $x = (y, t)$ . Let us chose the coupling in the time direction to be much stronger than in the space ones (universality should permit us to play this trick without changing the critical properties). We have:

$$Z = \sum_{\{\sigma_{y,t}\}} \exp(\beta_1 \sum_{y,\delta} \sigma_{y,t} \sigma_{y+\delta,t}) \exp(\beta_0 \sum_y \sigma_{y,t} \sigma_{y,t+1}) \quad (1.26)$$

This sum can be presented in a convenient form if we introduce the so-called transfer-matrix  $T$  which is defined by:

$$\langle \{\tilde{\sigma}_y\} | T | \{\sigma_y\} \rangle = \exp(\beta_0 \sum_y \tilde{\sigma}_y \sigma_y) \exp(\beta_1 \sum_{y,\delta} \sigma_y \sigma_{y+\delta}) \quad (1.27)$$

and has the order  $2^N \times 2^N$  where  $N$  is the number of  $y$ -points. From this definition it immediately follows that:

$$Z = \text{Tr } T^L \quad (1.28)$$

where  $L$  is the lattice length in the  $t$ -direction. Therefore it is enough to diagonalize the  $T$ -matrix in order to solve the system. For our purpose (1.27) can be further simplified. Let us use the identity:

$$\begin{aligned} e^{\beta_0 \sigma_y \tilde{\sigma}_y} &= \cosh \beta_0 + (\tilde{\sigma}_y \sigma_y) \sinh \beta_0 = \frac{1}{2} e^{\beta_0} (1 + \tilde{\sigma}_y \sigma_y) \\ &+ \frac{1}{2} e^{-\beta_0} (1 - \tilde{\sigma}_y \sigma_y) = \langle \{\tilde{\sigma}_y\} | e^{\beta_0} + e^{-\beta_0} \tau_y^z | \{\sigma_y\} \rangle \end{aligned} \quad (1.29)$$

(Here we have introduced the Pauli matrix  $\tau_y^z$ , with the states  $|\{\sigma_y\}\rangle$  satisfying  $\tau_y^z |\{\sigma_y\}\rangle = \sigma_y |\{\sigma_y\}\rangle$ ). If we take  $\beta_1 \ll 1$ , we have:

$$\exp\left(\beta_1 \sum_{y,\delta} \sigma_y \sigma_{y+\delta}\right) \approx \langle \{\sigma_y\} | 1 + \beta_1 \sum_{y,\delta} \tau_y^z \tau_{y+\delta}^z | \{\sigma_y\} \rangle \quad (1.30)$$

From these two relations we deduce

$$\begin{aligned} T &= e^{N\beta_0(1 + \hat{H})} \\ \hat{H} &= e^{-2\beta_0} \sum_y \tau_y^x + \beta_1 \sum_{y,\delta} \tau_{y+\delta}^z \end{aligned} \quad (1.31)$$

The critical point of the original model corresponds to the value of the parameters at which  $\hat{H}$  has a vanishing gap in the spectrum†).

#### 1.4 Continuum Abelian Global Symmetries

Next in order of increasing complexity we shall consider now a system with global  $O(2)$  symmetry. Instead of  $\sigma$ -variables with  $\sigma = \pm 1$  we have to introduce at each site a two dimensional unit vector  $\mathbf{n} = (\cos \alpha, \sin \alpha)$  and to consider the energy:

$$\mathcal{E} = - \sum_{\mathbf{x}, \mathbf{x}'} \mathcal{K}_{\mathbf{x}, \mathbf{x}'} \mathbf{n}_{\mathbf{x}} \cdot \mathbf{n}_{\mathbf{x}'} = - \sum_{\mathbf{x}, \mathbf{x}'} \mathcal{K}_{\mathbf{x}, \mathbf{x}'} \cos(\alpha_{\mathbf{x}} - \alpha_{\mathbf{x}'}) \quad (-\pi < \alpha_{\mathbf{x}} < \pi) \quad (1.32)$$

The partition function is defined by:

$$Z = \int_{-\pi}^{\pi} \prod_{\mathbf{x}} d\alpha_{\mathbf{x}} e^{-\beta \mathcal{E}[\alpha_{\mathbf{x}}]} \quad (1.33)$$

We can repeat the trick we used in Ising case in order to transform (1.33) to the theory of a continuous complex field  $\phi = \phi_1 + i\phi_2$ . For this purpose we write:

$$\begin{aligned} Z &= \exp\left(\beta \sum_{\mathbf{x}, \mathbf{x}'} \mathcal{K}_{\mathbf{x}, \mathbf{x}'} \frac{\partial^2}{\partial \phi_{\mathbf{x}}^a \partial \phi_{\mathbf{x}'}^a}\right) \Bigg|_{\phi_{\mathbf{x}}^a=0} \prod_{\mathbf{x}} \int_{-\pi}^{\pi} d\alpha_{\mathbf{x}} \exp(\phi_{\mathbf{x}}^1 \cos \alpha_{\mathbf{x}} + \phi_{\mathbf{x}}^2 \sin \alpha_{\mathbf{x}}) \\ &= \exp\left(\beta \sum_{\mathbf{x}, \mathbf{x}'} \mathcal{K}_{\mathbf{x}, \mathbf{x}'} 2 \left[ \frac{\partial^2}{\partial \phi_{\mathbf{x}}^* \partial \phi_{\mathbf{x}'}} + \frac{\partial^2}{\partial \phi_{\mathbf{x}} \partial \phi_{\mathbf{x}'}} \right]\right) \\ &\quad \times \exp\left(\sum_{\mathbf{x}} \log 2\pi I_0(i\sqrt{\phi_{\mathbf{x}}^* \phi_{\mathbf{x}}})\right) \Bigg|_{\phi_{\mathbf{x}}=0} \end{aligned} \quad (1.34)$$

---

† This condition implies that  $e^{-2\beta_0} \sim \beta_1 \ll 1$  which allows us to take for  $\hat{H}$  the approximation linear in  $\tau_y^z$ .

( $I_0$  here is a Bessel function). Repeating the arguments of the previous section we find that the theory (1.33) lies in the same universality class as the theory described by the Lagrangian:

$$\mathcal{L} = \partial_\mu \phi^* \partial_\mu \phi + m_0^2 \phi^* \phi + \frac{1}{2} \lambda_0 (\phi^* \phi)^2 \quad (1.35)$$

Again, we have no phase transition for  $\mathcal{D} = 1$ , and we have two different phases for  $\mathcal{D} \geq 2$ . The crucial difference from the Ising case is the existence of gapless excitations for all  $\beta > \beta_c$ , dictated by Goldstone's theorem. The physical origin of these excitations is very simple. Suppose that we have a broken symmetry,  $\langle \phi \rangle \neq 0$  (we shall see that this is true for  $\mathcal{D} \geq 3$ ). Then the states with different orientation:  $\langle \tilde{\phi} \rangle = e^{i\alpha} \langle \phi \rangle$  must have the same energy. If we form a state with slowly varying  $\alpha(x)$ , its energy will go to the vacuum value as the wave vector tends to zero. Hence, there should be no gap in the spectrum. To see this more formally let us introduce the conserved current  $\mathcal{J}_\mu$ :

$$\mathcal{J}_\mu = \frac{1}{i} \left[ \phi^* \frac{\partial}{\partial x_\mu} \phi - \left( \frac{\partial}{\partial x_\mu} \phi^* \right) \phi \right] \quad (1.36)$$

for which we have the Ward identity:

$$\frac{\partial}{\partial x_\mu} \langle \mathcal{J}_\mu(x) \phi(y) \rangle = i \delta(x-y) \langle \phi(y) \rangle \quad (1.37)$$

If we pass to the momentum representation

$$q_\mu \langle \mathcal{J}_\mu(q) \phi(-q) \rangle = \langle \phi(0) \rangle \quad (1.38)$$

we conclude by taking  $q \rightarrow 0$  that  $\langle \mathcal{J}_\mu(q) \phi(-q) \rangle$  must be singular in this limit, having a singularity:

$$\langle \mathcal{J}_\mu(q) \phi(-q) \rangle_{q \rightarrow 0} = \langle \phi(0) \rangle q_\mu / q^2 + \dots \quad (1.39)$$

For  $\mathcal{D} = 2$  the situation is more tricky. It is quite clear in this case that the propagator of the  $\mathbf{n}$ -field of the model (1.33) cannot have a Goldstone pole. Indeed, since

$$1 = \langle \mathbf{n}^2(0) \rangle = \int \frac{d^2 q}{(2\pi)^2} \langle \mathbf{n}(q) \mathbf{n}(-q) \rangle \quad (1.40)$$

such a pole in the right hand side would lead to an infrared contradiction. The answer is that the pole is softened and replaced by some powerlike singularity. There is also no naive order parameter and no true symmetry breaking in this case:  $\langle \mathbf{n} \rangle = 0$ . Nevertheless a phase transition at some  $\beta_c$  does take place and the observable properties of the phases are quite different from each other. We shall discuss them in

detail in later sections. We observe here a feature, characteristic for all continuous symmetries. Namely, we have *two* critical dimensions—the upper one, ( $\mathcal{D} = 4$ ) at which fluctuations at the phase transition point become irrelevant; and the lower one, ( $\mathcal{D} = 2$ ) at which Goldstone bosons start to interact strongly.

### 1.5 Non-Abelian Global Symmetries

There are several Non-Abelian generalizations of the preceding models. The most straightforward possibility is to consider again the expression (1.32) for the energy but to take the unit vectors  $\mathbf{n}_x$  to be  $N$ -dimensional. The symmetry group in this case will be  $O(N)$ . The major qualitative difference from the Non-Abelian case reveals itself for  $\mathcal{D} = 2$ . Due to the strong interaction of the Goldstone bosons, they acquire an energy gap for all values of  $\beta$  and the Non-Abelian system does not have a phase transition at all. At our preliminary level this qualitative difference can be explained as follows. In the continuum limit the Lagrangian (1.32) for  $N = 3$  has the form:

$$\mathcal{L} \sim (\partial_\mu \mathbf{n})^2 = (\partial_\mu \theta)^2 + \sin^2 \theta (\partial_\mu \varphi)^2 \quad (1.41)$$

(here  $\theta, \varphi$ -are polar and azimuthal angles). We conclude from (1.41) that the scattering amplitude  $F$  for the Goldstone bosons behaves like  $F \sim k^2$  where  $K$  is a characteristic momentum. The first radiative correction to this amplitude is given by:

$$F^{(1)} = \text{fish diagram} \sim F^2 \frac{1}{k^4} k^{\mathcal{D}}$$

$$\frac{F^{(1)}}{F} \sim k^{\mathcal{D}-2} \text{ for } \mathcal{D} \neq 2 \quad (1.42)$$

$$\frac{F^{(1)}}{F} \sim \log \frac{1}{k} \text{ for } \mathcal{D} = 2$$

(The last estimate is a consequence of the logarithmic divergence of dimensionless integrals.)

This result shows that the interaction is infrared-strong for  $\mathcal{D} = 2$  (in contrast with the Abelian case where  $F \sim k^4$ ).

It is a matter of more complicated analysis to see the consequences of this fact. We shall devote a special chapter to it.

For  $\mathcal{D} > 2$  the system has a phase transition and spontaneous symmetry breaking.

Another important Non-Abelian case is described by having group elements attached to each link. Let us consider matrices  $g \in G$  where  $G$  is some compact Lie group. The energy is given by:

$$\mathcal{E} = - \sum_{x, x'} \mathcal{N}_{x, x'} \text{Tr}(g_x^{-1} g_{x'}) \quad (1.43)$$

The partition function is

$$Z = \int \prod_x d\mu(g_x) e^{-\beta \mathcal{E}} \quad (1.44)$$

(where  $d\mu(g)$  is the Haar measure on the group).

The energy  $\mathcal{E}$  is invariant under  $G \otimes G$  transformations, described by the formula:

$$g_x \rightarrow u g_x v; \quad u, v \in G \quad (1.45)$$

The qualitative features of this theory are the same as for the  $O(N)$  model.

There are also many other Non-Abelian models in which fields belong not to the group itself but to some coset space  $G/H$ . They have some interesting features which we touch upon later.

## 1.6 Discrete Gauge Symmetries

Let us start from the discrete gauge group. The basic variables are quantities  $\sigma = \pm 1$ . But in this case they are attached to the links and not to the sites of the lattice. If we denote a link by the pair  $(x, \alpha)$ , where  $x$  is its beginning and  $\alpha$  its direction, the expression for the energy has the form:

$$\mathcal{E} = - \sum_{x, \alpha, \beta} \sigma_{x, \alpha} \sigma_{x+\alpha, \beta} \sigma_{x+\beta, \alpha} \sigma_{x, \beta} \quad (1.46)$$

(Here  $\alpha$  is a unit vector in the direction  $\alpha$ ). The rule, according to which (1.46) is constructed, is quite transparent. We have “1-forms” or “vector potentials”  $\sigma_{x, \alpha}$  associated with links. Then we take four links, forming a plaquette, and take a corresponding product around the given plaquette. As a result we obtain a “2-form” or “field strength”—the quantity, associated with each plaquette. The most remarkable property of this construction is that these 2-forms are gauge invariant. Indeed, if we change

$$\sigma_{x, \alpha} \rightarrow \eta_x \sigma_{x, \alpha} \eta_{x+\alpha} \quad (1.47)$$

(with  $\eta_x = \pm 1$ ) the field strength is unchanged:

$$f_{x, \alpha\beta} = \sigma_{x, \alpha} \sigma_{x+\alpha, \beta} \sigma_{x+\beta, \alpha} \sigma_{x, \beta} \rightarrow f_{x, \alpha\beta}$$

Therefore, the action has the symmetry  $(Z_2)^N$  where  $N$  is the number of lattice sites. We conclude that while in the global  $Z_2$  case the energy  $\mathcal{E}$  had precisely two degenerate minima (all  $\sigma_x = +1$  or  $-1$ ), in the gauge case the degeneracy is enormous, namely any configuration which is a "pure gauge":

$$\sigma_{x,\alpha} = \eta_x \eta_{x+\alpha} \quad (1.48)$$

with arbitrary  $\{\eta_x\}$  is the ground state.

These peculiar properties imply, first of all that there could be no order parameter in such systems,  $\langle \sigma_{x,\alpha} \rangle = 0$  and moreover, only gauge invariant quantities are nonzero. This follows from the fact, that by fixing the values of  $\sigma_{x,\alpha}$  at the boundary of our system we do not spoil gauge invariance inside it. All this does not have any deep influence on the phase structure of gauge systems. Different phases are easily distinguished by the different behaviour of gauge invariant correlation functions. The situation is reminiscent of what we had in the  $O(2)$  global model, where the second order phase transition took place without explicit violation of symmetry.

The physical properties of the  $Z_2$ -gauge system are the following: for  $\mathcal{D} = 2$  the model is trivially solved, being equivalent to the decoupled set of  $\mathcal{D} = 1$  Ising models (which follows from the fact that by the transformation (1.47) we can easily set  $\sigma_{x,1} = 1$ ), and therefore has no phase transitions. For  $\mathcal{D} = 3$  we shall show that it is equivalent to the  $\mathcal{D} = 3$  Ising model (by Kramers-Wannier duality.) This model is of great interest since it describes most of the  $\mathcal{D} = 3$  phase transitions in Nature. We shall devote a special chapter to its study. Now, for  $\mathcal{D} \geq 4$  numerical studies of  $Z_2$ -models show that there is a first order transition in this case. That means that the correlation length never becomes infinite and the theory does not have a continuum limit.

## 1.7 $O(2)$ Gauge Systems

In this case the system is constructed of unit vectors (which we write in complex form) attached to the links,  $\phi_{x,\alpha} = e^{iA_{x,\alpha}}$  ( $-\pi < A_{x,\alpha} \leq \pi$ ). The expression for the energy is:

$$\begin{aligned} \mathcal{E} &= - \sum_{x,\alpha,\beta} \frac{1}{2} (\phi_{x,\alpha} \phi_{x+\alpha,\beta} \phi_{x+\beta,\alpha}^* \phi_{x,\beta}^* + \text{c.c.}) \\ &= \sum_{x,\alpha,\beta} \cos(A_{x,\alpha} + A_{x+\alpha,\beta} - A_{x+\beta,\alpha} - A_{x,\beta}) \end{aligned} \quad (1.49)$$

The invariance properties of (1.49) are given by:

$$\phi_{x,\alpha} \rightarrow e^{i\varphi_x} \phi_{x,\alpha} e^{-i\varphi_{x+\alpha}} \quad (1.50)$$

or

$$A_{x,\alpha} \rightarrow A_{x,\alpha} + \varphi_x - \varphi_{x+\alpha}$$

(with arbitrary  $\{\varphi_x\}$ ). The formal continuum limit of (1.49) would give a familiar expression:

$$\begin{aligned} \mathcal{E} &= \text{const.} + \frac{1}{2} \int d\mathbf{x} F_{\alpha\beta}^2 \\ F_{\alpha\beta} &= \partial_\alpha A_\beta - \partial_\beta A_\alpha \end{aligned} \quad (1.51)$$

Should we conclude from this that the continuum limit of our model is just the free Maxwell field? To understand what goes on, let us compare the situation with the one described by (1.32). In the latter case the energy in the formal continuum limit is:

$$\mathcal{E} = \text{const.} + \frac{1}{2} (\partial_\mu \alpha)^2 \quad (1.52)$$

It describes a massless scalar field, which is nothing but the Goldstone field associated with symmetry breaking. In this Abelian case the perturbative interactions (described by the omitted terms  $\sim (\partial_\mu \alpha)^4$ ) are irrelevant as we saw. But, in the chapter devoted to instantons, we shall show that due to nonperturbative effects, associated with vortices, there exists a phase transition after which the field  $\alpha$  acquires a mass. So, we conclude that for  $\beta > \beta_c$  the system is indeed described by the massless free Goldstone field, that in the critical region  $|\beta - \beta_c| \ll \beta_c$  we have some complicated interacting continuum theory with both massless and massive particles, and that for  $\beta < \beta_c$  the massless particles disappear. All these effects are nonperturbative. Notice also an interesting phenomenon for  $\mathcal{D} = 2$ : we do not strictly speaking have spontaneous symmetry breaking, and the  $\phi$ -field from (1.35) has the property:

$$\langle \phi \rangle = 0 \quad (1.53)$$

Nevertheless we do have massless Goldstone modes, described by  $\alpha(x)$ , which disappear at the phase transition point. One of the possible ways to understand this is to introduce the decomposition

$$\phi(x) = \rho(x) e^{i\alpha(x)} \quad (1.54)$$

It is possible to show that

$$\langle \rho(x) \rangle_{V \rightarrow \infty} \begin{cases} 0, & \beta < \beta < \beta_c \\ \text{const.}, & \beta > \beta_c \end{cases} \quad (1.55)$$

(where  $V$  is the volume of our system), all of which is quite analogous to what we have in the case (1.49) and (1.51). If  $\mathcal{D} = 4$ , then for large  $\beta$  we have the theory of free photons. We see that these photons are in some sense Goldstone fields associated with gauge invariance, although this gauge invariance is never strictly broken. As we decrease  $\beta$ , at some  $\beta_c$  we shall have a phase transition due to instanton effects. For  $\beta < \beta_c$  the theory will contain only massive excitations. For  $\mathcal{D} = 3$  the situation is even more interesting. It will be shown that in this case nonperturbative effects extinguish the photons for all  $\beta$  and we have only one, massive phase. Therefore, the formal continuum limit (1.51) for  $\mathcal{D} = 3$  has nothing to do with reality. This is just one of many examples in which owing to quantum corrections the effective Lagrangian differs drastically from the classical one.

It remains to say here that for  $\mathcal{D} = 2$  the gauge model is trivial and for  $\mathcal{D} > 4$  it presumably has a first order phase transition.

## 1.8 Non-Abelian Gauge Theories

In this case we associate with each link a matrix of some compact Lie group  $G$ :  $B_{x,\alpha} \in G$ . The energy is given by:

$$\mathcal{E} = - \sum_{x,\alpha,\beta} \frac{1}{2} [\text{Tr}(B_{x,\alpha} B_{x+\alpha,\beta} B_{x+\beta,\alpha}^{-1} B_{x,\beta}^{-1}) + \text{c.c.}] \quad (1.56)$$

The invariance property of (1.56) is:

$$B_{x,\alpha} \rightarrow h_x^{-1} B_{x,\alpha} h_{x+\alpha} \quad (1.57)$$

In order to find the naive continuum limit one takes  $B_{x,\alpha}$  to be close to the identity element:

$$B_{x,\alpha} \simeq I + A_{x,\alpha} \quad (1.58)$$

with  $A_{x,\alpha}$  small and slow varying. That gives:

$$\mathcal{E} = \text{const} - \frac{1}{2} \int \text{Tr}(F_{\alpha\beta}^2) dx \quad (1.59)$$

$$F_{\alpha\beta} = \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta] \quad (1.60)$$

which is known as the Yang-Mills action. Just as in the case of global symmetries for  $\mathcal{D} = 2$ , in the gauge case at  $\mathcal{D} = 4$  the perturbative interaction is important. This is demonstrated by the estimates:

$$\begin{aligned} F &\sim \times \sim 1 & F^{(1)} &\sim \sim k^{\mathcal{D}-4} \\ F^{(1)}F &\sim \log(1/k) & \text{for } \mathcal{D} &= 4 \end{aligned} \quad (1.61)$$

(Here  $F$  is an amplitude for the scattering of  $A_\alpha$ -quanta and  $k$  is a characteristic wavelength.) Therefore, even for large  $\beta$  the true infrared behaviour has nothing to do with the naive continuum limit. In fact we shall give arguments that the theory has a mass gap and a rather peculiar spectrum for large  $\beta$ . Investigation of this limit is in fact a central problem of the theory of strong interactions. This is because the theory (1.56) reaches its continuum limit at  $\beta \rightarrow \infty$  only, and there are numerous pieces of evidence that this continuum limit for  $G = SU(3)$  describes the world of strong interactions. Most of our efforts in the next chapters will be devoted to this problem.

Here it remains to say that for  $\mathcal{D} = 3$  the theory has similar properties, namely no phase transition for any  $\beta$  and a massive phase at  $\beta \rightarrow \infty$ . For  $\mathcal{D} > 4$  we have first order transitions.

We shall pass now to a more systematic study of some of the properties listed above.

---

## CHAPTER 2

# Asymptotic Freedom and the Renormalization Group

---

DOI: 10.1201/9780203755082-2

### 2.1 Principal Chiral Fields†

In this section we shall study the large  $\beta$  limit for the  $\mathcal{D} = 2$  principal chiral field described by the Lagrangian:

$$\mathcal{L} = \frac{1}{2e_0^2} \text{Tr}(\hat{\partial}_\mu g^{-1} \hat{\partial}_\mu g) \quad (2.1)$$
$$(1/e_0^2 = \beta \gg 1 \quad \text{and} \quad g \in G)$$

As we discussed in Chapter 1, the infrared interaction of the massless particles described by (2.1) is logarithmically strong. It is our aim now to reveal the structure of this logarithmic interaction.

Let us study the effective Lagrangian which arises from (2.1) in the loop approximation. In order to find it we write the quantum field  $g(x)$  in the form:

$$g(x) = h(x) \cdot g_{\text{cl}}(x) \quad (2.2)$$

where  $g_{\text{cl}}(x)$  is some classical solution for the Lagrangian (2.1), namely

$$\hat{c}_\mu R_\mu^{\text{cl}} = 0 \quad (2.3)$$
$$R_\mu = \hat{c}_\mu g \cdot g^{-1}$$

Our programme is to integrate over the field  $h(x)$  so as to obtain an effective action depending on  $g_{\text{cl}}(x)$ . This approach is more or less standard in field theory but it requires some clarification. At the first sight, due to the invariance properties of the integration measure,  $\mathcal{D}h(x) = \mathcal{D}(h(x)g_{\text{cl}}(x))$ , the result of such integration would seem not to be dependent on  $g_{\text{cl}}(x)$  at all. Also it is not quite clear, what kind of

---

† A principal chiral field is one which defines the principal bundle over the base space.

classical solutions must be included. Both of these questions can be answered if we consider a finite size system and fix the value of  $g(x)$  at the boundary  $\Gamma$ :

$$g(x)|_{\Gamma} = g(\xi) \quad \xi \in \Gamma \quad (2.4)$$

We are considering the functional

$$\Psi[g(\xi)] = \int_{g(x)|_{\Gamma} = g(\xi)} \mathcal{D}g(x) e^{-(1/\hbar)S[g(x)]} \quad (2.5)$$

(where  $S$  is the action). This functional is just the Euclidean analogue of the Shrödinger wave function. The classical limit  $\hbar \rightarrow 0$  would correspond to taking the minimum of the action  $S$  with the Dirichlet boundary conditions (2.4). Integration over  $h(x)$  in the decomposition (2.2) corresponds to the inclusion of quantum fluctuations. It is clear from the above discussion that we should fix the boundary condition for  $h$ :

$$h(x)|_{\Gamma} = I \quad (2.6)$$

To sum up, we shall compute the integral over  $h$  and obtain an effective action depending on  $g_{cl}(x)$ . It must be understood, however, that  $g_{cl}(x)$  is not an independent variable: in fact all classical solutions are parametrized by their boundary values  $g(\xi)$  ( $\xi$  is  $\mathcal{D} - 1$ -dimensional). So, we are effectively computing the  $\Psi$ -functional (2.5), which as we already said is an analogue of the Shrödinger wave function and, on the other hand is analogous to the on-shell amplitudes of the Minkowskian theory (they also depend on  $\mathcal{D} - 1$ -dimensional fields).

All this information about boundary conditions and  $\Psi$ -functionals can be kept subconsciously so far as we are interested in infinite systems. Actually, most often we need not bother to express  $g_{cl}(x)$  through  $g(\xi)$ . Substituting (2.2) into (2.1) we get:

$$\begin{aligned} L_{\mu} &= g^{-1} \partial_{\mu} g = L_{\mu}^{cl} + g_{cl}^{-1} (h^{-1} \partial_{\mu} h) g_{cl} \\ \mathcal{L} &= -\frac{1}{2e_0^2} \text{Tr}(L_{\mu}^2) \\ &= -\frac{1}{2e_0^2} \text{Tr}(L_{\mu}^{cl})^2 + \frac{1}{2e_0^2} \text{Tr}(\partial_{\mu} h^{-1} \partial_{\mu} h) \\ &\quad + \frac{1}{e_0^2} \text{Tr}(R_{\mu}^{cl} (h^{-1} \partial_{\mu} h)), \quad R_{\mu}^{cl} = (\partial_{\mu} g_{cl}) g_{cl}^{-1} \end{aligned} \quad (2.7)$$



the effective action will be finite in the given order. The result of the above computation can be interpreted in two different ways. The most naive interpretation has already been given: namely we compute the effective action and see that the ultraviolet divergence in it can be absorbed into a redefinition of the coupling constant. Another interpretation is the following. Let us suppose that we have a theory with the cut-off  $\Lambda$  (which is of the order of the inverse lattice spacing). If we integrate out the fields  $h(x)$  with the wave vectors  $\tilde{\Lambda} < p < \Lambda$  we shall obtain the effective action which in the low energy limit has again the form (2.1) but with a renormalized value of  $e_0^2$ ,  $e_0^2(\tilde{\Lambda})$ :

$$e_0^2(\tilde{\Lambda}) = e_0^2(\Lambda) + \frac{e_0^4}{4\pi} C_v(G) \log \frac{\Lambda}{\tilde{\Lambda}} = \tilde{e}_0^2 \quad (2.13)$$

This formula follows directly from (2.11), provided that we restrict the  $p$ -integration by the condition  $\tilde{\Lambda} < |p| < \Lambda$ . As a result we conclude that the physical theory, formulated with the cut-off  $\Lambda$  and the bare charge  $e_0^2$  must be equivalent (for small momenta  $|p| \ll \Lambda$ ) to the one with the cut-off  $\tilde{\Lambda}$  and the specially chosen new bare charge  $\tilde{e}_0^2$ . This statement is called renormalizability and the transformations from  $e_0^2$  to  $\tilde{e}_0^2$  and from  $\Lambda$  to  $\tilde{\Lambda}$  the renormalizability group. The formula (2.13), as is seen from its derivation, is correct provided that:

$$e_0^2 \log \frac{\Lambda}{\tilde{\Lambda}} \ll 1 \quad e_0^2 \ll 1 \quad (2.14)$$

Nothing prevents us from repeating the procedure and passing from  $\tilde{\Lambda}$  to  $\tilde{\tilde{\Lambda}} < \tilde{\Lambda}$  etc.

The most important consequence of renormalizability is that it controls the momentum dependence of different physical quantities. Let us consider as an example the behaviour of the effective charge  $e^2(p)$  for the fluctuations with momentum  $p$ . This quantity can be defined in different ways. One of the possibilities is to consider the four point function with all momenta equal to  $p$ . There are many other options and we shall comment on this ambiguity later.

Since our theory does not contain any dimensional parameter except for  $\Lambda$ , we must have:

$$e^2(p) = e^2(\log(\Lambda/p), e_0^2) \quad (2.15)$$

Let us express  $e^2(p)$  in terms of  $e^2(\mu)$ , where  $\mu$  is some fixed value of the momentum. Inverting (2.15) we have:

$$e_0^2 = e_0^2(\log(\Lambda/\mu), e^2(\mu)) \quad (2.16)$$

Substitution of (2.16) into (2.15) gives:

$$e^2(p) = e^2(\log(p/\mu), e^2(\mu), \log(\Lambda/\mu)) \quad (2.17)$$

The first thing renormalizability tells us is that (2.17) actually does not depend on its last argument, since we can change  $\Lambda$  and compensate for it by changing  $e_0^2$  without changing  $e^2(p)$ . Hence:

$$e^2(p) = e^2(\log(p/\mu), e^2(\mu)) \quad (2.18)$$

We see that  $e^2(p)$ , being expressed in terms of  $e^2(\mu)$ , does not contain any divergences and does not depend on the structure of the theory at distances of the order of lattice spacing. But this is not the end of the story. Actually we have a further constraint on (2.18) which follows from the fact that the point  $\mu$  was quite arbitrary. Hence, just as it was with  $\Lambda$ , it must be possible to compensate for a shift in  $\mu$  by changing  $e^2(\mu)$ . We have, therefore, a functional constraint on  $e^2(p)$  which is quite easy to solve. Namely, it is clear that:

$$e^2(p) = f(\log(p/\mu) + g(e^2(\mu))) \quad (2.19)$$

with

$$f(g(x)) = x \quad (2.20)$$

The structure (2.19) makes the above property explicit: a shift in  $\log(p/\mu)$  is obviously compensated by a change of  $e^2(\mu)$ . The formula (2.20) follows from the fact that  $e^2(p)|_{p=\mu} = e^2(\mu)$ . The relation (2.19) presents a very strong constraint on the structure of momentum dependence. It is not fulfilled in a fixed order of perturbation theory and permits us to obtain nonperturbative expressions.

It follows from (2.19) that in a theory without dimensional parameters a so-called dimensional transmutation takes place:

$$e^2(p) = f(\log(p/\lambda)) \quad \lambda = \mu e^{-g(e^2(\mu))} \quad (2.21)$$

All quantities depend on a universal correlation length  $\lambda^{-1}$  which should be kept fixed as the lattice spacing  $\Lambda^{-1}$  goes to zero. No other arbitrary parameters enter into the theory. (The last is not generally true: there are theories with several effective charges, like massless scalar QED, in which physical quantities depend on the ratios of these charges). In order to see how (2.19) improves perturbation theory, let us write it in differential form:

$$\frac{de^2(p)}{d \log(p/\mu)} = \beta(e^2(p)) \quad (2.22)$$

$$\beta(x) = f'(g(x))$$

This is the Gell-Mann-Low equation. Iterations of its solution in  $\log(p/\mu)$  give:

$$e^2(p) \simeq e^2(\mu) + \beta(e^2(\mu)) \log(p/\mu) + O(\log^2(p/\mu)) \quad (2.23)$$

At the same time, from (2.13) we deduce:

$$e^2(p) \simeq e^2(\mu) - \frac{C_v(G)}{4\pi} e^4(\mu) \log(p/\mu) \quad (2.24)$$

Comparison gives:

$$\beta(e^2) = -\frac{C_v(G)}{4\pi} e^4 + O(e^6)$$

Solving the differential equation (2.22) we get:

$$e^2(p) = \frac{e^2(\mu)}{1 + \frac{C_v(G)}{8\pi} e^2(\mu) \log\left(\frac{p^2}{\mu^2}\right)} \quad (2.25)$$

It is easy to check that while (2.24) does not have the form (2.19), (2.25) does (with  $f(x) = 4\pi/C_v(G) \cdot x$ ). Another useful expression for  $e^2(p)$  is:

$$e^2(p) = \frac{e_0^2}{1 - \frac{C_v(G)}{8\pi} e_0^2 \log \frac{\Lambda^2}{p^2}} \quad (2.26)$$

( $\Lambda$  is the inverse lattice spacing,  $p \ll \Lambda$ ).

What is the range of applicability of (2.25) and (2.26)? It is defined by the fact that we have neglected all higher powers in  $e^2(p)$  in the expression for the  $\beta$ -function. Therefore the condition is:

$$e^2(p) \ll 1 \quad (2.27)$$

The real meaning of this improvement to perturbation theory, invented by Gell-Mann and Low, is that it replaces the expansion in the bare charge  $e_0^2$  which may not be small by the expansion in  $e^2(p)$ , which in many important cases is small.

For example if we rewrite (2.25) as:

$$e^2(p) = \frac{8\pi}{C_v(G)} \frac{1}{\log(p^2/\lambda^2)} \quad (2.28)$$

we conclude that this is a true asymptotic expansion for  $e^2(p)$  when  $p \gg \lambda$ . As we shall see, in this region all correlation functions can be computed for the reason that the interaction is small. This ultraviolet smallness is called asymptotic freedom. For  $p \lesssim \lambda$  perturbation theory

is untrue and qualitatively different methods are needed. They will be discussed in later chapters.

Let us proceed to the computation of the correlation functions. Take as an example:

$$\mathcal{D}(x - y) = \langle \text{Tr}(g^{-1}(x)g(y)) \rangle \quad (2.29)$$

The tactic is again the following. Let us integrate over rapid fluctuations with  $\tilde{\Lambda} < |p| < \Lambda$  and use the renormalization group argument. If we write

$$g(x) = e^{\phi(x)} g_0(x) \simeq (1 + \phi(x) + \frac{1}{2} \phi^2(x)) g_0(x) \quad (2.30)$$

and integrate over the  $\phi$  with the wavelengths in our interval we get:†

$$\mathcal{D}(p) = \mathcal{D}_0(p) \left( 1 - \frac{C_2}{\pi} e_0^2 \log \frac{\Lambda}{\tilde{\Lambda}} \right) \quad (2.31)$$

From (2.31) we deduce by the repetition of the arguments leading to the Gell-Mann-Low equation, that if we set

$$\mathcal{D}(p) = d(p)/p^2 \quad (2.32)$$

then:

$$\frac{\partial \log d(p)}{\partial \log(p/\mu)} = \gamma(e^2(p)) \quad (2.33)$$

where

$$\gamma(e^2) = \frac{C_2}{\pi} e^2 + O(e^4) \quad (2.34)$$

Integrating (2.33) we get:

$$\begin{aligned} \mathcal{D}(p) &= \frac{1}{p^2} \left( 1 + \frac{C_v e^2(\mu)}{8\pi} \log(p^2/\mu^2) \right)^{4C_2/C_v} \\ &\sim \frac{1}{p^2} \left( \log \left( \frac{p^2}{\lambda^2} \right) \right)^{4C_2/C_v} \quad \text{for } p \gg \lambda \end{aligned} \quad (2.35)$$

which is the desired answer. Of course, the range of applicability is again  $p \gg \lambda$ .

As we said before the definition of  $e^2(p)$  is ambiguous. This ambiguity leads to an ambiguity of the  $\beta$ -function and of the  $\gamma$ -function. However

---

† Here we put  $t^A t^A = -C_2 I$ . For the fundamental representation of  $G = SU(N)$ :  $C_2 = (N^2 - 1)/2N$ ;  $C_v(SU(N)) = N$ .

the next result for any correlation function is unique—the  $\beta$ -, and  $\gamma$ -ambiguities compensate each other. Moreover, the first two coefficients in the  $\beta$ -function are more or less universal. This can be checked by a change in the definition of  $e^2(p)$ . Suppose we chose:

$$\tilde{e}^2(p) = e^2(p) + C_1 e^4(p) + C_2 e^6(p) + \dots \quad (2.36)$$

and

$$\frac{de^2}{d \log(p/\mu)} = \beta_1 e^4 + \beta_2 e^6 + \dots \quad (2.37)$$

On substituting (2.36) into (2.37) we get:

$$\begin{aligned} \frac{d\tilde{e}^2}{d \log(p/\mu)} &= \frac{de^2}{d \log(p/\mu)} (1 + 2C_1 e^2 + \dots) \\ &= (1 + 2C_1 e^2 + \dots)(\beta_1 e^4 + \beta_2 e^6 + \dots) = \beta_1 e^4 + (\beta_2 + 2C_1 \beta_1) \\ &\quad \times e^6 + \dots = \beta_1 (\tilde{e}^2 - C_1 \tilde{e}^4)^2 + (\beta_2 + 2C_1 \beta_1) \tilde{e}^6 + \dots \\ &= \beta_1 \tilde{e}^4 + \beta_2 \tilde{e}^6 + \dots \end{aligned} \quad (2.38)$$

The coefficient  $\beta_3$  will be changed, but up to two loop order the  $\beta$ -function is indeed universal. In the case when several leading coefficients of the  $\beta$ -function vanish, this remark relates to the first nonvanishing one.†

So, our conclusion is that the Gell-Man-Low renormalization group is a useful tool in the region where the effective charge is small. In order to use it one has to compute the coefficients before the leading powers of  $\log(p/\mu)$  in physical quantities. This is usually easy in the lowest order, but becomes increasingly tedious in the higher orders, since one has to separate a nonleading contribution  $\sim \log(p/\mu)$  from the leading ones  $\sim (\log(p/\mu))^n$ . In the asymptotically free theories the renormalization group gives order by order a small distance expansion which goes in inverse powers of  $\log(p/\mu)$ . For example the two loop order solution of equation (2.22) gives:

$$e^2(p) = -\frac{1}{\beta_1 \log(p/\lambda)} \left\{ 1 + \frac{\beta_2 \log \log(p/\lambda)}{\beta_1^2 \log(p/\lambda)} + O\left(\frac{1}{\log^2(p/\lambda)}\right) \right\} \quad p \gg \lambda \quad (2.39)$$

Another useful thing to remember is that the inverse correlation length  $\lambda$  is expressed in terms of the bare parameters  $\Lambda$ ,  $e_0$  as

$$\lambda = \text{const} \cdot \Lambda (e_0^2)^{\beta_2/\beta_1^2} e_0^{1/\beta_1} \quad (2.40)$$

---

† In the general case we have freedom to redefine coupling constants, leading to Riemannian geometry in the space of coupling 'constants'.

which is the consequence of the relations:

$$\frac{\partial \lambda}{\partial \Lambda} + \frac{\partial \lambda}{\partial e_0^2} \frac{\partial e_0^2}{\partial \Lambda} = 0 \quad (2.41)$$

$$\lambda = \Lambda \phi(e_0^2)$$

(the first is the renormalization condition and the second is just a naive dimensional statement). The equation (2.40) indicates a consistent limiting procedure necessary to reach the continuum limit of the theory. For the asymptotically free case ( $\beta_1 < 0$ ) it is described by the conditions:

$$e_0 \rightarrow 0 \quad \Lambda \rightarrow \infty \quad \lambda \rightarrow \text{const.} \quad (2.42)$$

## 2.2 The $n$ -Fields

In this section we shall consider another important example of an asymptotically-free system with a global symmetry. It is the theory of the field of an  $N$ -dimensional unit vector  $\mathbf{n}$ ,  $\mathbf{n}^2 = 1$ . The action is given by:

$$S = \frac{1}{2e_0^2} \int d^2x (\partial_\mu \mathbf{n})^2 \quad (2.43)$$

In order to exploit the renormalization group, we set:

$$\mathbf{n}(x) = (1 - (\varphi)^2)^{1/2} \mathbf{n}_0(x) + \sum_{a=1}^{N-1} \varphi_a \mathbf{e}_a(x) \quad (2.44)$$

(where  $\mathbf{n}_0(x)$  is a slowly varying vector and the  $\{\mathbf{e}_a\}$  are orthogonal to it and to each other;  $\varphi_a$  represents "fast" fluctuations with  $\tilde{\Lambda} < |p| < \Lambda$ ). The vectors  $(\mathbf{n}_0, \mathbf{e}_a)$  obey the relations

$$\begin{aligned} \partial_\mu \mathbf{n}_0 &= \sum_a B_\mu^a \mathbf{e}_a \\ \partial_\mu \mathbf{e}_a &= \sum_b A_\mu^{ab} \mathbf{e}_b - B_\mu^a \mathbf{n}_0 \end{aligned} \quad (2.45)$$

(Here  $B_\mu^a$  and  $A_\mu^{ab}$  are some potentials characterizing the  $\mathbf{n}_0$ -field; (2.45) is the consequence of  $\mathbf{n}_0 \cdot \mathbf{e}_a = 0$  and  $\mathbf{e}_a \cdot \mathbf{e}_b = \delta_{ab}$ ). Substituting (2.44) into (2.43) gives:

$$\begin{aligned} S &= \frac{1}{2e_0^2} \int \{ (\partial_\mu (1 - \varphi^2)^{1/2} - B_\mu^a \varphi^a)^2 \\ &\quad + (\partial_\mu \varphi^a - A_\mu^{ab} \varphi^b + B_\mu^a (1 - \varphi^2)^{1/2})^2 \} d^2x \end{aligned} \quad (2.46)$$

Selecting quadratic terms in  $\phi$  we get:

$$S^{(0)} = \frac{1}{2e_0^2} \int \{(\partial_\mu \phi^a - A_\mu^{ab} \phi^b)^2 + B_\mu^a B_\mu^b (\phi^a \phi^b - \phi^2 \delta^{ab})\} d^2x \\ + \frac{1}{2e_0^2} \int (B_\mu^a)^2 d^2x \quad (2.47)$$

The logarithmic contribution to the renormalization of  $e_0^2$  comes from the second term in (2.47). Integrating over  $\phi$  with momenta

$$\bar{\Lambda} < |p| < \Lambda$$

gives:

$$\langle \phi^a \phi^b - \phi^2 \delta^{ab} \rangle = \delta^{ab} [1 - (N-1)] \int_{\bar{\Lambda}}^{\Lambda} \frac{d^2p}{(2\pi)^2} \frac{e_0^2}{p^2} \\ = \delta^{ab} (2-N) \frac{e_0^2}{2\pi} \log \frac{\Lambda}{\bar{\Lambda}} \quad (2.48) \\ \frac{1}{\bar{e}_0^2} = \frac{1}{e_0^2} + \frac{N-2}{2\pi} \log \frac{\bar{\Lambda}}{\Lambda}$$

Recalling the arguments of the preceding section we obtain:

$$\beta(e^2) = -\frac{N-2}{2\pi} e^4 \quad (2.49) \\ e^2(p) = e^2(\mu) \left( 1 + \frac{N-2}{4\pi} e^2(\mu) \log(p^2/\mu^2) \right)$$

The correlation function

$$\mathcal{D}(x-y) = \langle \mathbf{n}(x) \mathbf{n}(y) \rangle \quad (2.50)$$

and the function  $\gamma(e^2)$  are also easily computed to leading order. We have:

$$\langle \mathbf{n}(x) \mathbf{n}(y) \rangle_\Lambda \approx \langle \mathbf{n}_0(x) \mathbf{n}_0(y) \rangle_\Lambda (1 - \langle \phi^2 \rangle_{\Lambda, \Lambda}) \\ = \langle \mathbf{n}_0(x) \mathbf{n}_0(y) \rangle \left( 1 - \frac{N-1}{2\pi} e^2(\Lambda) \log \frac{\Lambda}{\bar{\Lambda}} \right) \quad (2.51)$$

and

$$\gamma(e^2) = \frac{N-1}{2\pi} e^2 \quad (2.52) \\ \mathcal{D}(p) = \frac{1}{p^2} \left( 1 + \frac{N-2}{4\pi} e^2(\mu) \log \frac{p^2}{\mu^2} \right)^{(N-1)/(N-2)}$$

Notice that  $A_\mu^{ab}$  did not participate in our derivation because it enters in a gauge invariant way. As a consequence it appears only in combinations containing  $(F_{\mu\nu}^{ab})^2$  (where  $F_{\mu\nu}^{ab}$  is the Yang-Mills field strength) which have dimension 4 and cannot contain logarithmic divergences. This gauge invariance is just the reflection of the arbitrariness of the fields  $\{e_a\}$  (remember that only  $n$  is physical).

This kind of derivation can be easily generalized to fields belonging to an arbitrary coset space,  $G/H$ . (Above we have worked out the cases  $H = I$ , and the case of the  $n$ -field:  $n \in S^N = SO(N + 1)/SO(N)$ ,  $S^N$  being a unit sphere).

### 2.3 Non-Abelian Gauge Fields for $\mathcal{D} = 4$

In this case the action has the form:

$$S = -\frac{1}{2e_0^2} \int \text{Tr } F_{\mu\nu}^2 d^4x \quad (2.53)$$

We shall follow the same procedure as before. Let us set

$$A_\mu = \bar{A}_\mu + a_\mu \quad (2.54)$$

and expand (2.53) up to terms quadratic in  $a_\mu$ ;

$$\begin{aligned} \bar{S} &= -\frac{1}{2e_0^2} \int \text{Tr } \bar{F}_{\mu\nu}^2 d^4x \\ &\quad -\frac{1}{2e_0^2} \text{Tr} \int d^4x \{(\nabla_\mu a_\nu - \nabla_\nu a_\mu)^2 + 2\bar{F}_{\mu\nu} \cdot [a_\mu, a_\nu]\} \\ &= \bar{S} - \frac{1}{e_0^2} \text{Tr} \int d^4x \{(\nabla_\mu a_\nu)^2 + 2\bar{F}_{\mu\nu} \cdot [a_\mu, a_\nu] - (\nabla_\lambda a_\lambda)^2\} \end{aligned} \quad (2.55)$$

(Here  $\nabla_\mu a_\nu = \partial_\mu a_\nu + [A_\mu, a_\nu]$ .)

We have to integrate over all possible  $a_\mu$ -fields with momenta  $\bar{\Lambda} < |p| < \Lambda$ . At this point we have to fix the gauge of  $a_\mu$ , because due to the gauge invariance of the action the quadratic form (2.55) has zero eigenvalues. Namely if we take

$$a_\mu^{(0)} = \nabla_\mu w = \partial_\mu w + [\bar{A}_\mu, w] \quad (2.56)$$

we find that  $S^{(1)} = 0$  for an arbitrary function  $w$ . This simply means that we have to integrate in directions in our functional space which are



The expression for  $\Pi_{\mu\nu}^{ab}$  in (2.60) is exactly the polarization operator for massless scalar QED multiplied by an isotopic factor:

$$\begin{aligned}\Pi_{\mu\nu}^{ab}(q) &= 2f^{acd}f^{bcd}\Pi_{\mu\nu}^{(\text{scalar})}(q) \\ &= 2C_v(G)\delta^{ab}\Pi_{\mu\nu}^{(\text{scalar})}(q)\end{aligned}\quad (2.61)$$

The factor 2 in (2.61) arises from the two physical polarizations of  $a_\mu$ . Computation of  $\Pi_{\mu\nu}^{(\text{scalar})}$  is easy:

$$\begin{aligned}\Pi_{\mu\nu}^{(\text{sc})} &= -\frac{1}{4}\int\frac{d^4p}{(2\pi)^4}\frac{(2p+q)_\mu(2p+q)_\nu}{p^2(p+q)^2}+\frac{1}{2}\delta_{\mu\nu}\int\frac{d^4p}{(2\pi)^4p^2} \\ &= \pi(q^2)(q^2\delta_{\mu\nu}-q_\mu q_\nu)\end{aligned}\quad (2.62)$$

(This equality follows from the conservation of current  $q^\mu\Pi_{\mu\nu}^{(\text{sc})}=0$ . It ensures cancellation of the quadratic divergence in  $\Pi_{\mu\nu}^{(\text{sc})}$ ). Taking the trace in  $\mu\nu$  of (2.62) and keeping only the terms proportional to  $q^2$  we have:

$$\begin{aligned}3q^2\pi(q^2) &= \frac{1}{2}\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^4}\left(4p^2-\frac{q^2+4pq+4p^2}{2\left(1+\frac{2pq}{p^2}+\frac{q^2}{p^2}\right)}\right) \\ &\simeq\frac{1}{4}\int\frac{d^4p}{(2\pi)^4}\frac{1}{p^4}\left\{-q^2-\frac{8(pq)^2}{p^2}+4p^2\frac{q^2}{p^2}\right\} \\ &= \frac{q^2}{64\pi^2}\log\frac{\Lambda^2}{\bar{\lambda}^2}\end{aligned}\quad (2.63)$$

Therefore, to within logarithmic accuracy we have the following diamagnetic part of the coupling renormalization:

$$\left(\frac{1}{\bar{e}^2}\right)_{\text{diam.}} = \frac{1}{e^2} + \frac{1}{48\pi}C_v(G)\log\frac{\Lambda^2}{\bar{\lambda}^2}(\bar{\lambda} < \Lambda)\quad (2.64)$$

We see that this part of the effect decreases the coupling as we go to larger wave lengths. The physical interpretation of this in Minkowski space is that a charge introduced into the vacuum gets screened by virtual pairs of particles and antiparticles, just as happens in a dielectric medium. In Euclidean space the explanation of the sign of (2.64) is also simple. The second order effect in general should decrease the effective action (or free energy in the language of statistical mechanics), as is seen from:

$$\begin{aligned}S_{\text{eff}} &= -\log\int e^{-(S(\varphi)+V(\varphi))}\mathcal{D}\varphi \\ S_{\text{eff}}^{(2)} &= -\frac{1}{2}\langle V^2(\varphi)\rangle = -\frac{1}{2}\frac{\int V^2(\varphi)e^{-S(\varphi)}d\varphi}{\int e^{-S(\varphi)}d\varphi}\end{aligned}\quad (2.65)$$

(In this formula  $\varphi$  is any field and  $V(\varphi)$  is any perturbation with  $\langle V(e) \rangle = 0$ ). So the normal sign in (2.64) would be negative. However, vector fields enter the action with an extra factor  $i$  (in Euclidean space the extra term in the amplitude for a scalar particle is  $\exp\{i\oint_C \mathcal{A} dL\}$  where  $C$  is the particle's trajectory). This effect in second order gives an  $i^2 = -1$  factor, which changes the sign in (2.65).

From this consideration it is obvious that the spin term in (2.55) will give a negative sign contribution to the charge renormalization. This paramagnetic effect is readily evaluated:

$$\begin{aligned}
 S_{\text{eff}}^{(\text{paramagnetic})} &= -\frac{1}{2e_0^4} \int \bar{F}_{\mu\nu}^a(q) \bar{F}_{\lambda\rho}^b(-q) \frac{d^4q}{(2\lambda)^4} \\
 &\quad \times f^{acd} f^{bef} \langle a_\mu^c a_\nu^d(-q), a_\lambda^e a_\rho^f(q) \rangle \\
 &\simeq \int d^4x (\bar{F}_{\mu\nu}^a)^2 \frac{C_v(G)}{2e_0^4} \left( \bigcirc_p^{p+q} \right) \Big|_{q=0} \\
 &= -\frac{C_v(G)}{2e_0^4} \int (\bar{F}_{\mu\nu}^a)^2 d^4x \int \frac{d^4p}{(2\lambda)^4} \frac{2e_0^4}{p^2(p+q)^2} \Big|_{q=0} \\
 &= -\frac{C_v(G)}{16\pi^2} \log \frac{\Lambda^2}{\bar{\Lambda}^2} \int (\bar{F}_{\mu\nu}^a)^2 d^4x \tag{2.66}
 \end{aligned}$$

Putting together these two effects we obtain:

$$\frac{1}{\bar{e}^2} = \frac{1}{e^2} - \frac{11}{3} C_v(G) \frac{1}{16\pi^2} \log \frac{\Lambda^2}{\bar{\Lambda}^2} \tag{2.67}$$

Now we can repeat all the previous arguments concerning the renormalization group and we find that the effective coupling in the nonabelian gauge theories is asymptotically free:

$$e^2(p) = \frac{48\pi^2}{11C_v(G)} \frac{1}{\log(p^2/\lambda^2)}, \quad p \gg \lambda \tag{2.68}$$

This is the famous result of Gross, Wilzek and Politzer.

It is not difficult to evaluate different gauge invariant Green functions at small distances. As before, they differ from the free ones by factors of the type  $(\log(p^2/\lambda^2))^\gamma$  where  $\gamma$  depends on the type of the Green function. However, the methods described above are absolutely inadequate for the investigation of the infrared region  $p^2 \lesssim \lambda^2$ , which is physically the most interesting. To accomplish this task we proceed to a quite different approach.

---

## CHAPTER 3

# The Strong Coupling Expansion

---

---

DOI: 10.1201/9780203755082-3

In the previous chapter we have seen how simple perturbative methods permit us to examine the short distance behaviour in asymptotically free theories. The reason for such success was that the effective coupling for high momentum fluctuations is logarithmically small. In the region  $p \sim \lambda$  however, this coupling becomes of the order of unity (if we take the formula (2.68), naively it becomes infinite.) In order to examine the infrared structure of the theory we have to develop some nonperturbative methods. In this chapter we shall describe the simplest (though in many respects imperfect) method—the strong coupling expansion. Unfortunately this phrase means expansion not in the physical coupling (which may really be large) but in the bare one,  $e_0^2$ . At first sight the enterprise may seem completely meaningless because as was explained above, the continuum limit of our lattice models is achieved when

$$e_0^2 \sim \frac{1}{\log \Lambda} \xrightarrow{\Lambda \rightarrow \infty} 0 \quad (3.1)$$

The reason the large  $e_0^2$  expansion is interesting is that there are grounds to believe that in most asymptotically free systems there are no phase transitions in  $e_0^2$ . If true, this implies that the qualitative character of the spectrum and correlation functions is unchanged as we go from small to large  $e_0^2$ . For instance, the masses of elementary excitations must have no singularities in  $e_0^2$  and hence their  $1/e_0^2$  expansions can be continued to rather small values of  $e_0^2$ . This numerical aspect will not be discussed here. Instead we shall be concerned with the qualitative picture which arises from the strong coupling limit. Though our main interest lies in nonabelian gauge systems, we start from the easier cases because it is always satisfactory to realize a special theory as belonging to some larger variety.

### 3.1 Ising Model

In this section we shall reveal the strong coupling (small  $\beta$ ) expansion in the Ising model. By using the relation

$$e^{\beta\sigma\sigma'} = \cosh \beta(1 + \sigma\sigma' \tanh \beta) \quad (3.2)$$

we get:

$$Z = (\cosh \beta)^N \sum_{\{\sigma\}} \prod_{\langle x, \delta \rangle} (1 + \zeta \sigma_x \sigma_{x+\delta})$$

$$\zeta = \tanh \beta \quad (3.3)$$

Expanding the product (3.3) we see that only those terms contribute in which in each site we have  $\sigma$  raised to an even power. On each link we can have the term 1 or  $\zeta \sigma_x \sigma_{x+\delta}$ . If we draw a solid line on this link in the second case we shall obtain the following diagrammatic expansion for  $Z$ . We have to draw all closed paths on a lattice, such that at each site only an even number of lines meets, and each link can be covered once or not covered at all. If we look at the correlation function  $\langle \sigma_0 \sigma_{\mathbf{R}} \rangle$ , the rules will be the same except that at the points 0 and  $\mathbf{R}$  we must have an odd number of lines meeting. The contribution of a given graph is just  $\zeta^L$  where  $L$  is the total length of the solid line.

The interpretation of these rules is the following. Since the correlations are exponentially small ( $\sim \zeta^{|\mathbf{R}|}$  for  $\beta \ll 1$ ) we have a massive excitation with the gap  $\sim \log(1/\beta)$ . As we increase  $\beta$  (decrease the coupling) the mass gap decreases. At some point  $\beta_c$ , the number of paths of length  $L$  which is of the order of  $\exp(\text{const. } L)$  becomes larger than the damping factor  $\zeta^L$ . At this point the paths will get condensed —there will be a finite density of lines in the system. In terms of the correlation function, this phase transition will mean that  $\langle \sigma_0 \sigma_{\mathbf{R}} \rangle \xrightarrow{R \rightarrow \infty} \text{const.}$  This expansion in terms of lines is typical for all systems with global symmetries. The lines themselves are nothing but the world lines of the elementary excitations: to see this more explicitly let us use the hamiltonian formulation of the Ising model, described by (1.31):

$$\hat{H} = u \sum_{\mathbf{y}} \tau_{\mathbf{y}}^1 + v \sum_{\mathbf{y}, \delta} \tau_{\mathbf{y}}^3 \tau_{\mathbf{y}+\delta}^3 (u = e^{-\beta_0}, v = \beta_1) \quad (3.4)$$

The strong coupling limit corresponds to  $u \gg v$ . In some sense the first term in (3.4) is a kinetic energy, while the second is a potential one (because the first term describes the change of  $\tau_{\mathbf{y}}^3$  in time).

In the leading approximation the ground state is described by the wave function:

$$\begin{aligned} \tau_y^1 |\phi\rangle &= -|\phi\rangle \\ |\phi\rangle &= \prod_y \begin{pmatrix} 1 \\ -1 \end{pmatrix}_y \quad \begin{pmatrix} a \\ b \end{pmatrix}_y \stackrel{\text{def}}{=} \frac{a|\uparrow\rangle_y + b|\downarrow\rangle_y}{(a^2 + b^2)^{1/2}} \end{aligned} \quad (3.5)$$

(The Hilbert space in our problem is formed by the direct product of the two dimensional spaces at each site  $y$ . The operators  $\tau_y$  act as Pauli matrices on the space labelled by  $y$  and as a unit operator on all other spaces.) The ground state energy corresponding to (3.5) is given by:

$$E_0 = -Nu \quad (3.6)$$

The first excited state in this leading approximation is obtained as:

$$|\phi_1\rangle = \prod_{y \neq y_0} \begin{pmatrix} 1 \\ -1 \end{pmatrix}_y \times \begin{pmatrix} 1 \\ 1 \end{pmatrix}_{y_0} \stackrel{\text{def}}{=} |y_0\rangle \quad (3.7)$$

These states (labelled by  $y_0$ ) have energy:

$$E_1 - E_0 = 2u \quad (3.8)$$

We see that in the leading approximation we have a nondegenerate ground state and a highly degenerate first excited level, separated by the gap  $2u$ . We shall see now that in the next approximation this degeneracy is removed.

If we denote the second term in (3.4) by  $V$  we have, first of all to consider matrix elements  $\langle y' | V | y \rangle$ . Due to the obvious relations

$$\begin{aligned} \tau_y^3 |y\rangle &= |0\rangle \\ \tau_{y_1}^3 |y_2\rangle &= |y_1, y_2\rangle \quad (y_1 \neq y_2) \\ \tau_y^3 |0\rangle &= |y\rangle \end{aligned} \quad (3.9)$$

we find that:

$$\begin{aligned} \langle y' | V | y \rangle &= v \sum_{\delta} \delta_{y', y \pm \delta} \\ \langle y' | H_0 | y \rangle &= 2u \delta_{y, y'} \end{aligned} \quad (3.10)$$

We see that while  $H_0$  described individual and independent spins, the term  $V$  describes hopping of their excitations from one site to another.

Diagonalizing  $H_0 + V$ , we find that in the first order of degenerate perturbation theory we have the Eigenstates:

$$\begin{aligned} |\mathbf{p}\rangle &= \sum_y e^{i\mathbf{p}\cdot\mathbf{y}} |y\rangle \\ \mathcal{E}(\mathbf{p}) &= E_1(\mathbf{p}) - E_0 = 2u + 2v \sum_{\delta} \cos(\mathbf{p}\delta) \end{aligned} \quad (3.11)$$

As expected, we obtain point-like elementary excitations with a finite gap, which are characterized by the quasimomentum  $\mathbf{p}$ . This conclusion will be true in all further orders of perturbation theory, which gives an expansion in  $v/u$ . As we have already said, in the Ising model this expansion will diverge at a certain critical value of  $v$ . This value corresponds to the phase transition point at which the gap

$$E_1(0) - E_0 = 0 \quad (3.12)$$

Near this point it is expected (and will be explicitly shown for  $\mathcal{D} = 2$ ) that:

$$m = E_1(0) - E_0 \ll 1$$

and

$$\mathcal{E}(\mathbf{p}) = (m^2 + \mathbf{p}^2)^{1/2} \quad (3.13)$$

for  $|\mathbf{p}| \ll 1$ .

After the phase transition point a condensate of these particles is formed and the strong coupling expansion is impossible. In this phase one has to start from the opposite limit, treating the first term in (3.4) as a perturbation.

In the zero approximation we have a strictly ordered vacuum

$$|\phi\rangle = \prod_{\substack{\mathbf{y} = 2n_i\delta_i \\ (n_i \in \mathbb{Z})}} \left( \binom{1}{0} \prod_{\delta} \binom{0}{1}_{\mathbf{y}+\delta} \right) \quad (3.14)$$

It should be clear from the above that the lines which appear in the expansion of the Euclidean version of our model are just the world lines of the particles we treated in the Hamiltonian version. It is an interesting exercise to establish the correspondence between Schrödinger perturbation theory for the Hamiltonian and diagrams for the Euclidean approach.

### 3.2 Continuous Global Symmetry

Let us start from the abelian case. In order to obtain the strong coupling limit it is helpful to use the expansion (the analogue of (3.2)):

$$e^{\beta \cos(\varphi - \varphi')} = \sum_{n=0}^{\infty} \lambda_n(\beta) e^{in(\varphi - \varphi')} \quad (3.15)$$

( $\lambda_n(\beta) = I_n(-i\beta)$ ,  $I_n$  being a Bessel function). Substituting (3.15) into the partition function we get:

$$\begin{aligned} Z &= \lambda_0^N \sum_{\{n_{x,\delta}\}} \exp \left\{ \sum_{x,\delta} \log(\lambda_{n_{x,\delta}}(\beta)/\lambda_0(\beta)) \right\} \\ &\times \int_{-\pi}^{\pi} \prod_x \frac{d\varphi_x}{2\pi} \exp \left\{ i \sum_{x,\delta} n_{x,\delta} (\varphi_x - \varphi_{x+\delta}) \right\} \\ &= \lambda_0^N \sum_{\{n_{x,\delta}: \sum_{\delta} (n_{x,\delta} - n_{x-\delta,\delta}) = 0\}} \prod_{x,\delta} \frac{\lambda_{n_{x,\delta}}(\beta)}{\lambda_0(\beta)} \end{aligned} \quad (3.16)$$

As a result, we obtain the following graphical rules for computing (3.16). We have again a graph on the lattice, such that each line is characterized by a nonzero integer  $n$ . This  $n$  is conserved (owing to the condition  $\sum_{\delta} (n_{x,\delta} - n_{x-\delta,\delta}) = 0$ ). That means for example, that if three lines meet at some site then  $n_1 + n_2 + n_3 = 0$ . To each bit of the path we associate the factor  $\lambda_{n_{x,\delta}}(\beta)/\lambda_0(\beta) < 1$ , and take a product of all such factors. Notice that in the Ising case the rules were the same except that the permitted values of  $n_{x,\delta}$  were 0, 1 and the conservation was true mod 2. If we wish to compute a correlation function

$$\langle e^{im(\varphi(0) - \varphi(\mathbf{R}))} \rangle = \mathcal{D}_m(\mathbf{R}) \quad (3.17)$$

( $m$  being some integer) we have the same rules except that now:

$$\sum_{\delta} (n_{x,\delta} - n_{x-\delta,\delta}) = m(\delta_{x,\mathbf{R}} - \delta_{x,0}) \quad (3.18)$$

which means that we have “sources” of  $n$ -flux placed at the points  $\mathbf{O}$  and  $\mathbf{R}$ . The result for the correlation functions is again that they decay exponentially for small  $\beta$ . Hence in this phase we expect to have only massive excitations. It is worthwhile to study the same theory in its

Hamiltonian version. In order to obtain it we introduce a large anisotropy in the "time" direction:

$$\begin{aligned}
 -S = & \beta_0 \sum_{\mathbf{y}} \cos(\varphi_{\mathbf{y},t} - \varphi_{\mathbf{y},t+1}) \\
 & + \beta_1 \sum_{\mathbf{y},\delta} \cos(\varphi_{\mathbf{y},t} - \varphi_{\mathbf{y}+\delta,t})
 \end{aligned} \quad (3.19)$$

(Here  $\mathbf{y}$  belongs to the  $\mathcal{D} - 1$ -dimensional lattice). In the limit of large  $\beta_0$  we can replace (3.19) by:

$$-S = \text{const} - \int dt \left\{ \beta_0 \sum_{\mathbf{y}} \frac{1}{2} \left( \frac{d\varphi_{\mathbf{y}}}{dt} \right)^2 - \beta_1 \sum_{\mathbf{y},\delta} \cos(\varphi_{\mathbf{y},t} - \varphi_{\mathbf{y}+\delta,t}) \right\} \quad (3.20)$$

Comparing (3.20) with (1.10) we see that it is just the action for imaginary time of the system of coupled rotators sitting at the lattice sites and described by the angles  $\{\varphi_{\mathbf{y}}\}$ . In the Minkowskian time their lagrangian is:

$$\mathcal{L} = \sum_{\mathbf{y}} \frac{\beta_0}{2} \left( \frac{d\varphi_{\mathbf{y}}}{dt} \right)^2 + \beta_1 \sum_{\mathbf{y},\delta} \cos(\varphi_{\mathbf{y},t} - \varphi_{\mathbf{y}+\delta,t}) \quad (3.21)$$

Passing to the Hamiltonian by the standard procedure we obtain:

$$\begin{aligned}
 H = & \frac{1}{2\beta_0} \sum_{\mathbf{y}} l_{\mathbf{y}}^2 - \beta_1 \sum_{\mathbf{y},\delta} \cos(\varphi_{\mathbf{y}} - \varphi_{\mathbf{y}+\delta}) \\
 l_{\mathbf{y}} = & \frac{1}{i} \frac{\partial}{\partial \varphi_{\mathbf{y}}}
 \end{aligned} \quad (3.22)$$

(Notice the analogy with (3.4)). Now the strong coupling limit  $\beta_1 \beta_0 \ll 1$  corresponds to neglect of the potential energy in (3.22):

$$\begin{aligned}
 H_0 = & \frac{1}{2\beta_0} \sum_{\mathbf{y}} l_{\mathbf{y}}^2 \\
 \Psi = & \exp\left(i \sum_{\mathbf{y}} n_{\mathbf{y}} \varphi_{\mathbf{y}}\right), E = \frac{1}{2\beta_0} \sum_{\mathbf{y}} n_{\mathbf{y}}^2
 \end{aligned} \quad (3.23)$$

Here  $\{n_{\mathbf{y}}\}$  is an arbitrary set of integers; the ground state corresponds to all  $n_{\mathbf{y}} = 0$ . The first excited state is obtained by taking some one of the  $n_{\mathbf{y}_0} = \pm 1$ , and has a mass gap. The reason for this is simple—in our approximation all the rotators are decoupled. The elementary excitation we have described is just the excitation of a single rotator placed at the point  $\mathbf{y}_0$ . As before taking into account the potential energy causes hopping of the excitation from  $\mathbf{y}$  to  $\mathbf{y} + \delta$ :

$$\langle \mathbf{y}' | V | \mathbf{y} \rangle = - \frac{\beta_1}{2} \sum_{\delta} \delta_{\mathbf{y}', \mathbf{y} + \delta} \quad (3.24)$$

Therefore, as in the Ising case, we have an elementary excitation with energy depending on quasimomentum. The gap is nonzero in any finite order in  $\beta_1\beta_0$ . However, again we expect a phase transition to occur. Let us clarify the nature of the large  $\beta_1\beta_0$  phase. While for small  $\beta_1\beta_0$  we had an almost decoupled set of rotators, for large  $\beta_1\beta_0$  all these rotators are tightly bound. If we consider them as something like a rigid body, then the excitation spectrum will be

$$E = L^2/2I \quad (3.25)$$

where  $L = \sum_{\mathbf{y}} l_{\mathbf{y}}$  is the total angular momentum and  $I$  is the moment of inertia. In the tightly bound phase we should have  $I \sim N$  where  $N$  is the number of rotators. If the rotators are weakly coupled, then  $I \sim 1$  (each one rotates separately). The phase transition in  $\beta_0\beta_1$  corresponds to the change between these two regimes. In the small  $\beta_1\beta_0$  phase we have a gap in the spectrum and in the large  $\beta_1\beta_0$  phase the gap is absent (as  $N \rightarrow \infty$ ). In the next chapter we shall explore this phase transition in more detail.

The Non-Abelian case in the strong coupling region is not much different from the above. We have (in the case of the  $\mathbf{n}$ -field) the following Hamiltonian:

$$H = \frac{1}{2\beta_0} \sum_{\mathbf{y}} l_{\mathbf{y}}^2 - \beta_1 \sum_{\mathbf{y}, \mathbf{s}} (\mathbf{n}_{\mathbf{y}} \cdot \mathbf{n}_{\mathbf{y}+\mathbf{s}}) \quad (3.26)$$

In this case  $l_{\mathbf{y}}$  is the standard operator of angular momentum with eigenvalues of  $l_{\mathbf{y}}^2: l(l+1)$  (with degeneracies  $2l+1$ ). All conclusions are the same as above except that the elementary excitation is vector-like (it has  $l=1$ ) and that for  $\mathcal{D}=2$ , as will be explained later, there is no phase transition. The last fact is of great importance. It means that even as we take  $\beta_1\beta_0 \rightarrow \infty$  (or  $e_0^2 \rightarrow 0$ ) we have a gap in the energy spectrum. This gap can be interpreted as arising through the strong interaction of Goldstone's bosons. On the basis of the strong coupling expansion we expect that the Lagrangian

$$\mathcal{L} = \frac{1}{2e_0^2} (\partial_{\mu} \mathbf{n})^2 \quad (3.27)$$

describes massive particles with isotopic spin 1. Dimensional transmutation, described in Chapter 2, predicts that all scattering amplitudes depend on  $p_j/m$  (where  $p_j$  are the momenta of the particles) and do not contain any free parameters (like the coupling constant). All these expectations turn out to come true as follows from the exact solution.

The last model with global symmetry is the principal chiral field. Its Lagrangian is:

$$\mathcal{L} = \sum_{\mathbf{y}} \frac{\beta_0}{2} \text{Tr}(\dot{g}_{\mathbf{y}}^{-1} \dot{g}_{\mathbf{y}}) + \frac{\beta_1}{2} \sum_{\mathbf{y}, \delta} (\text{Tr}(g_{\mathbf{y}}^{-1} g_{\mathbf{y}+\delta}) + \text{c.c.}) \quad (3.28)$$

We see that it describes a set of symmetric tops (in the case  $g \in SU(2)$ ). The Hamiltonian has the form:

$$H = \sum_{\mathbf{y}} \frac{1}{2\beta_0} \hat{l}_{\mathbf{y}}^2 - \frac{\beta_1}{2} \sum_{\mathbf{y}, \delta} \{\text{Tr}(g_{\mathbf{y}}^{-1} g_{\mathbf{y}+\delta}) + \text{c.c.}\} \quad (3.29)$$

Here  $\hat{l}_{\mathbf{y}}$  is an operator of left rotations which, in the case of the  $SU(2)$  group, can be expressed in terms of derivatives in Euler angles (see any book on quantum mechanics). Eigenvalues of  $\hat{l}^2$  are again  $l(l+1)$  but the degeneracy is  $(2l+1)^2$  due to the fact that the symmetry group of (3.29) is  $SU(2) \otimes SU(2)$  (body and frame rotations in quantum mechanics of the top). The quantum numbers of elementary excitations are such that they transform by fundamental representations of both groups. In the case of  $SU(2)$  it is a vectorlike excitation of  $SO(4) \simeq SU(2) \otimes SU(2)$ . Again, this conclusion will be confirmed by exact results.

### 3.3 Gauge Symmetries

We have seen that in the strong coupling region all systems with global symmetries look roughly the same. In all cases we had massive pointlike excitations which propagate through the lattice. In the Non-Abelian cases with  $\mathcal{D} = 2$  this picture remained valid even for small coupling.

In the case of a gauge system the strong coupling region is again rather insensitive to the type of symmetry. However, gauge invariance introduces qualitatively new features to this region which will be discussed now.

Let us look first at the small  $\beta$  expansion for the partition function in the  $Z_2$  case. Since each term in the energy (1.46) is associated with a plaquette, the result of the expansion can be presented as a collection of plaquettes, such that at each link an even number of plaquettes meet. Therefore we have something like closed surfaces instead of closed paths for a nongauge system. The contribution of a given surface to the partition function is given by  $(\tanh \beta)^A$  where  $A$  is the number of plaquettes or, in other words, the area of the surface. For more

complicated gauge groups, like  $O(2)$  or  $SU(N)$  the small  $\beta$  expansion is similar with the only difference that the plaquettes carry conserved quantum numbers just as it was in the case of paths, describing systems with global symmetries.

We see that there exists a very general rule that the strong coupling expansion for a gauge system is obtained from the corresponding global system by replacing paths by surfaces. The physical meaning of this rule becomes transparent in the Hamiltonian language. Let us pass to the hamiltonian in the  $O(2)$  case (other cases are similar). As before, to do this we introduce anisotropy in the “time” direction into the formula (1.49). We obtain:

$$\begin{aligned} \mathcal{L} = & \frac{\beta_0}{2} \sum_{y,\alpha} (\dot{A}_{y,\alpha} + \phi_y - \phi_{y+\alpha})^2 \\ & + \beta_1 \sum_{y,\alpha,\beta} (\cos(A_{y,\alpha} + A_{y+\alpha,\beta} - A_{y+\beta,\alpha} - A_{y,\beta}) - 1) \end{aligned} \quad (3.30)$$

Here  $\alpha = 1, \dots, \mathcal{D} - 1$ ;  $y$  belongs to a  $\mathcal{D} - 1$  dimensional lattice, and we denote by  $\phi_y$  the time component of the vector potential  $A_{y,0}$ . The first term in (3.30) was obtained by expansion of the corresponding cosine in (1.49) and making time continuous. In order to pass to the hamiltonian, let us introduce canonical momenta:

$$E_{y,\alpha} = \frac{\partial \mathcal{L}}{\partial \dot{A}_{y,\alpha}} = \beta_0 (\dot{A}_{y,\alpha} + \phi_y - \phi_{y+\alpha}) \quad (3.31)$$

The Hamiltonian is given by:

$$\begin{aligned} H = & \sum_{y,\alpha} E_{y,\alpha} \dot{A}_{y,\alpha} - \mathcal{L} \\ = & \frac{1}{2\beta_0} \sum_{y,\alpha} E_{y,\alpha}^2 + \beta_1 \sum_{y,\alpha,\beta} \{1 - \cos(A_{y,\alpha} + A_{y,\alpha,\beta} \\ & - A_{y+\beta,\alpha} - A_{y,\beta})\} - \sum_{y,\alpha} \phi_y (E_{y,\alpha} - E_{y-\alpha,\alpha}) \end{aligned} \quad (3.32)$$

Since we have no time derivatives for the field  $\phi_y$  we must just minimize  $H$  with respect to this field, which gives the condition:

$$\Gamma_y = \sum_{\alpha} (E_{y,\alpha} - E_{y-\alpha,\alpha}) = 0 \quad (3.33)$$

With this condition, the Hamiltonian is just:

$$H = \frac{1}{2\beta_0} \sum_{y,\alpha} E_{y,\alpha}^2 + \beta_1 \sum_{y,\alpha,\beta} (1 - \cos F_{y,\alpha,\beta}) \quad (3.34)$$

Notice, that the operators  $\Gamma_y$  are generators for the gauge transformations of  $A_{y,\alpha}$ . Indeed:

$$\delta A_{y,\alpha} = \left[ \sum_z i\omega_z \Gamma_z, A_{y,\alpha} \right] = \omega_y - \omega_{y+\alpha} \quad (3.35)$$

Therefore:

$$[\Gamma_y, H] = 0 \quad (3.36)$$

In quantum theory we have to substitute

$$E_{y,\alpha} = \frac{1}{i} \frac{\partial}{\partial A_{y,\alpha}}$$

and to solve the equations:

$$\begin{aligned} H\psi &= \mathcal{E}\psi \\ \Gamma_y\psi &= 0 \end{aligned} \quad (3.37)$$

The strong coupling limit corresponds to neglecting the second, potential term in (3.34). The general solution to (3.37) in this approximation is:

$$\begin{aligned} \psi &= \exp\left(i \sum_{y,\alpha} n_{y,\alpha} A_{y,\alpha}\right) \\ \mathcal{E} &= \frac{1}{2\beta_0} \sum_{y,\alpha} n_{y,\alpha}^2 \end{aligned} \quad (3.38)$$

with  $n_{y,\alpha}$  being integers satisfying the conservation condition

$$\sum_{\alpha} (n_{y,\alpha} - n_{y-\alpha,\alpha}) = 0 \quad (3.39)$$

The vacuum solution corresponds to  $n_{y,\alpha} = 0$ . Excited states are described by a closed loop on the lattice, such that at each site the  $n$  are conserved. We recognize the same set of loops which we had in the description of the global  $O(2)$  model, but the interpretation is now different. Each loop labels the quantum state in the present case. In the strong coupling limit the energy of this state is given by (3.38) and is proportional to the total length. As we consider the time propagation of such loops we obtain the world surfaces of the Euclidean approach. Physically, the loop is formed of Faraday flux lines (we see from (3.38) that  $n_{y,\alpha}$  is an eigenvalue of the electric field  $E_{y,\alpha}$ ). The condition (3.39)

is just the conservation of electric flux (which is true since we consider a theory without charges). Due to the angular nature of the vector potential the flux is quantized. In the zero approximation the shape of the loop remains unchanged by time development. As we include the second term in (3.34) two different effects arise. First of all, our closed flux line acquires quasimomentum and begins to move across the lattice. This is quite analogous to the case of global symmetries where we had point-like elementary excitations. The second effect is more tricky. As we perturb the state of a given shape with the cosine term in (3.34) this shape can change (for instance, the cosine can create a new small square, formed by the flux lines). Therefore the true quantum state is a superposition of closed flux lines with different shapes. In later chapters we shall develop a string theory so as to describe the continuum limit of such a picture. This continuum limit exists, provided that no phase transition in  $e_0^2$  takes place.

Let us clarify the relation of this picture to confinement of charges. This relation is quite simple. Suppose that we introduce two static opposite charges into our system. Then, we shall have a flux line which ends on these charges. The energy of such a state is proportional to the distance between the charges. If no phase transition takes place, this picture will remain true even for small coupling. In the next chapter it will be shown that for  $\mathcal{D} = 3$  it is indeed so, while for  $\mathcal{D} = 4$  there is a phase transition leading to a condensation of strings. After the condensation we obtain the Coulomb law instead of confinement.

Generalization of the above discussion to the Non-Abelian case does not present any difficulties. The basic variables in this case are matrices of  $SU(N)$  attached to the links of a  $\mathcal{D} - 1$ -dimensional lattice:  $B_{y,\alpha}$ . In the abelian case we had the electric field operator  $E$ , with commutation relations:

$$[iE_{y,\alpha}, e^{iA_{y,\beta}}] = i\delta_{y,y'}\delta_{\alpha,\beta}e^{iA_{y,\beta}}$$

$$iE_{y,\alpha} = \frac{\partial}{\partial A_{y,\alpha}} \Leftrightarrow i\dot{A}_{y,\alpha} = (e^{iA_{y,\alpha}})\dot{e}^{-iA_{y,\alpha}}$$
(3.40)

The Non-Abelian generalization of (3.40) is the following. We have two different electric fields, corresponding to left and right invariant forms of the matrix  $B_{y,\alpha}$ . Namely let us introduce:

$$L_{y,\alpha} = \dot{B}_{y,\alpha}B_{y,\alpha}^{-1} \quad R_{y,\alpha} = B_{y,\alpha}^{-1}\dot{B}_{y,\alpha}$$

$$R_{y,\alpha} = B_{y,\alpha}^{-1}L_{y,\alpha}B_{y,\alpha}$$

$$\text{Tr}(L_{y,\alpha}^2) = \text{Tr}(R_{y,\alpha}^2)$$
(3.41)

The commutation relations satisfied by the  $L$  and  $R$  operators are:

$$\begin{aligned} [L_{y,\alpha}^a, B_{y',\beta}] &= \delta_{y,y'} \delta_{\alpha,\beta} \lambda^a B_{y',\beta} \\ [R_{y,\alpha}^a, B_{y',\beta}] &= \delta_{y,y'} \delta_{\alpha,\beta} B_{y',\beta} \lambda^a \end{aligned} \quad (3.42)$$

(Here  $\{\lambda^a\}$  is a set of generators for  $SU(N)$  and  $iL = \lambda^a L^a$ ;  $iR = \lambda^a R^a$ .)

If the group is parametrized in some way, say like  $SU(2)$  by Euler angles, then it is easy to express  $L^a$  and  $R^a$  as differential operators with respect to these parameters. However, we shall not need these explicit expressions.

The Hamiltonian for the Non-Abelian case has the form (compare with (3.34)):

$$H = -\frac{1}{2\beta_0} \sum_{y,\alpha} \text{Tr} L_{y,\alpha}^2 - \frac{\beta_1}{2} \sum_{y,\alpha,\beta} [\text{Tr}(B_{y,\alpha} B_{y+\alpha,\beta} B_{y+\beta,\alpha}^{-1} B_{y,\alpha}^{-1})] \quad (3.43)$$

There are operators  $\Gamma_y^a$ , which commute with  $H$  and generate gauge transformation of (3.43):

$$\begin{aligned} \Gamma_y &= i\lambda^a \Gamma_y^a = \sum_{\alpha} (L_{y,\alpha} - R_{y-\alpha,\alpha}) \\ &= \sum_{\alpha} (L_{y,\alpha} - R_{y-\alpha,\alpha}^{-1} L_{y-\alpha,\alpha} B_{y-\alpha,\alpha}) \end{aligned} \quad (3.44)$$

$$[\Gamma_y^a, H] = 0$$

$$[\Gamma_y^a, \Gamma_{y'}^b] = \delta_{y,y'} f^{abc} \Gamma_y^c \quad (3.45)$$

( $f^{abc}$  being the structure constants of the group). The last equation is easily derived by combining the definitions with (3.42) and the Jacobi identity. In the continuum limit (3.44) gives the covariant divergence of the Non-Abelian electric field:

$$\begin{aligned} L_{y,\alpha} &\approx E_{y,\alpha} \\ B_{y,\alpha} &\approx I + A_{y,\alpha} \\ \Gamma_y &\approx \partial_{\alpha} E_{\alpha} + [A_{\alpha}, E_{\alpha}] \end{aligned} \quad (3.46)$$

Without external charges, the spectrum of our system is described by the Shrödinger equation with subsidiary condition:

$$H\Psi[B] = \mathcal{E}\Psi[B], \Gamma_y^a \Psi[B] = 0 \quad (3.47)$$

Since  $\Gamma_y^a$  is a generator of gauge transformations this last condition means that we have to choose a gauge invariant  $\Psi[B]$  from all possible solutions of the Shrödinger equation. In the strong coupling limit we neglect the last term in (3.43) and obtain a set of independent tops (for

the  $SU(2)$  group, taken as a representative example). Again the simplest gauge invariant excitation is a square formed by flux lines:

$$\begin{aligned} \Psi_1[B] &\sim \text{Tr}(B_{x,\alpha} B_{x+\alpha,\beta} B_{x+\beta,\alpha}^{-1} B_{x,\alpha}^{-1}) \\ \mathcal{E} &= \frac{4l(l+1)}{4\beta_0} \Big|_{l=1/2} = \frac{3}{4\beta_0} \end{aligned} \quad (3.48)$$

Just as in the Abelian case, this elementary excitation will become a superposition of different shapes in higher orders. The only physical difference from the Abelian case appears when we consider the interaction of external charges. Let us develop the corresponding formalism. There is a small subtlety in this procedure since although infinitely heavy charges are classical as far as their orbital properties are concerned their isotopic spins must be treated quantum mechanically. Therefore, it is not advisable just to add terms like  $\mathcal{L}_\mu^{\text{ext}} A_\mu$  to the lagrangian. A fool-proof procedure is to describe the charge by a quantum field  $\chi$  with isotopic spin  $I$ , having the Lagrangian:

$$\mathcal{L}_\chi = i\chi^+ \left( \frac{\partial}{\partial t} + A_0 \right) \chi - M\chi^+ \chi \quad (3.49)$$

The fact that  $\chi$  has no kinetic energy implies that the position of our charge is fixed. After passing to the Hamiltonian and applying the condition  $\chi^+ \chi = 1$ , we obtain the following natural prescription for describing static charges with isotopic spins  $I_1, I_2, \dots, I_N$  sitting at the points  $x_1, \dots, x_N$ . Consider the solution of the equations:

$$\begin{aligned} H\Psi(I_1, m_1, x_1; \dots; I_N, m_N, x_N) &= \mathcal{E}_{I_1 m_1 x_1; \dots; I_N m_N x_N} \Psi(\dots) \\ \Gamma^a(y) \Psi(I_1 m_1 x_1; \dots; I_N m_N x_N) &= - \left[ \sum_j \sum_{\{m'_j\}} (\lambda_{(I_j)}^a)_{m_j m'_j} \delta(y - y_j) \right] \Psi(\{I_j m'_j y_j\}) \end{aligned} \quad (3.50)$$

( $H$  is again (3.43).)

Then  $\mathcal{E}(\dots)$  is just the energy of the sector with  $N$  external charges. The second condition (3.50) means that instead of gauge invariant  $\Psi$  in the vacuum sector we have to consider  $\Psi$  which changes under the gauge transformations according to the rule:

$$\begin{aligned} \Psi_{m_1 \dots m_N}[B^\Omega] &= \sum_{m'_1 \dots m'_N} \mathcal{D}_{m_1 m'_1}^{I_1}(\Omega^{-1}(x_1)) \times \dots \\ &\quad \times \mathcal{D}_{m_N m'_N}^{I_N}(\Omega^{-1}(x_N)) \times \Psi_{m'_1 \dots m'_N}[B]; \\ B_{x,\alpha}^\Omega &= \Omega_x^{-1} B_{x,\alpha} \Omega_{x+\alpha} \end{aligned} \quad (3.51)$$

and  $\mathcal{D}_{mm}^I(\Omega)$  are standard representation matrices for the  $SU(2)$  group. The meaning of the rules (3.50) and (3.51) is quite transparent. They imply that static charges, the state of which is characterized by their isotopic spin  $I$  and its projection  $m$ , are sources of Non-Abelian electric flux.

With these prescriptions at hand we can demonstrate an important property specific to the Non-Abelian case mentioned above. Let us first consider the interaction of two charges with isotopic spins  $1/2$  placed at a distance  $R$  from each other. In the strong coupling limit the wave function of their state is given by:

$$\Psi_{(m_1 x_1, m_2 x_2)}[B] \sim \mathcal{D}_{m_1 m_2}^{1/2} \left( \prod_{L_{x_1, x_2}} B_{y, \alpha} \right) \quad (3.52)$$

(where we have taken a product of  $B$ -matrices along the straight line  $L_{x_1, x_2}$  connecting the charges). Since tops placed on this line, are excited (they have  $l = 1/2$ ) we obtain the result:

$$\mathcal{E}_{1/2}(x_1, x_2) \sim |x_1 - x_2| \quad (3.53)$$

just as in the Abelian case. Now, let us consider two charges with  $I = 1$ . One possible wave function will be the same as (3.52) with  $\mathcal{D}^{1/2}$  replaced by  $\mathcal{D}^1$ . However, here we have a state with much lower energy. To see how it is formed let us compare the first option, which could have been represented by the picture

$$\Psi^{(1)} = x_1, m_1 \longrightarrow x_2, m_2 \quad (3.54)$$

(here we draw the flux line connecting  $x_1$  and  $x_2$ ) with another one

$$\Psi^{(2)} x_1, m_1 \quad \square \quad x_2, m_2 = \Psi_{m_1}, \Psi_{m_2} \quad (3.55)$$

where the flux curls up. The explicit expression for, say, the left square is:

$$\Psi_m = \text{Tr}(\tau^m B_1 B_2 B_3^+ B_4^+) \quad (3.56)$$

$$\tau^{\pm \lambda} = \tau^x \pm i\tau^y, \tau^0 = \tau^z$$

( $\tau^{x,y,z}$  are Pauli matrices.)

It is trivial to check that under a gauge transformation (3.56) transforms in accordance with (3.51). At the same time the energy of this state does not depend on  $|x_1 - x_2|$  and therefore for large distances the state (3.55) is much more favourable than (3.54) and we have no confinement for the states with  $I = 1$ . Notice, that for  $I = 1/2$  the option (3.55) was absent, which was the reason for confinement. It is

quite easy to understand what has happened. The Non-Abelian gauge field describes gluons, having  $I = 1$ . When we added a source with  $I = 1$  it is energetically favourable for this source to be screened by an adjacent gluon (this process is shown in the picture (3.55)). If the source has  $I = 1/2$ , then in the strong coupling phase such screening is impossible and we have colour confinement. In principle another phase is possible in which even  $I = 1/2$  will be screened away by  $I = 1$  gluons. For this to happen we must have a cloud containing an infinite number of gluons. Then statistically they can screen our source. However such a cloud would have finite energy only if the constituent gluons are massless. Therefore, we expect that if there is an energy gap in the gauge theory then the effective cloud will consist of a finite number of gluons and will be unable to screen half-integer spin. One of the amusing consequences of this picture is that the string tension between sources with, say,  $I = 7/2$  is the same as for  $I = 1/2$  because three units of  $I$ -spin will be screened by gluons and only  $1/2$  will remain. All this is a manifestation of the fact that the electric flux is conserved only mod(1).

The question whether we indeed have a mass gap in the continuum limit ( $e_0^2 \rightarrow 0$ ) cannot be solved by the strong coupling expansion, because its presence in any order in  $1/e_0^2$  proves nothing. More subtle methods are needed. We shall describe some of them in the next chapter.



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

---

## CHAPTER 4

# Instantons in Abelian Systems

---

---

DOI: 10.1201/9780203755082-4

In the previous chapter we saw that the existence of a mass gap is the most important property of gauge and spin systems. In this chapter we shall discuss a specific mechanism for gap generation, which is especially important in Abelian systems and will also play some role in nonabelian ones.

### 4.1 Instantons in Quantum Mechanics and the Ising Model

Let us describe the symmetry properties in the quantum mechanical system described by the action (for imaginary time):

$$S = \int dt \left\{ \frac{1}{2} \dot{\varphi}^2 - \frac{1}{2} \mu^2 \varphi^2 + \frac{\lambda}{4} \varphi^4 + \frac{\mu^4}{4\lambda} \right\} \quad (4.1)$$

with  $\lambda \ll \mu^2$ . The only interest of this model for us is that it provides the simplest demonstration of a phenomenon present in more complicated systems. The point we intend to examine is that this model in any finite order of perturbation theory has apparently broken symmetry, whereas in reality the symmetry is restored. To see this we expand:

$$\varphi = \pm \frac{\mu}{\lambda^{1/2}} + \chi \quad (4.2)$$

Expanding the action near, say, the left-hand minimum we have:

$$S \approx \int dt \left\{ \frac{1}{2} \dot{\chi}^2 + \mu^2 \chi^2 + O(\lambda^{1/2}/\mu) \chi^3 \right\} \quad (4.3)$$

We see that we have (for small  $\lambda$ ) almost harmonic oscillations near the bottom of the left-hand well. The left-right symmetry  $\varphi \rightarrow -\varphi$  is broken and this will remain so in any finite order in  $\lambda$ . At the same time we know from quantum mechanics that the ground state of this theory is described by an even  $\psi$ -function and therefore is nondegenerate, with

restored symmetry. It is clear that restoration of symmetry occurs because a particle placed in the left well will (with finite probability) tunnel to the right one and back. Therefore if we wait long enough there will be equal probability of finding the particle in either of the wells. There exists an interesting way to describe tunnelling which we briefly discuss now. It is easy to see (by rescaling  $\varphi \rightarrow \lambda^{-1/2} \varphi$ ) that for small  $\lambda$ ,  $S$  is very large, being of the order  $\lambda^{-1}$ . That means that in the functional integral:

$$Z = \int \mathcal{D}\varphi e^{-S[\varphi]} \quad (4.4)$$

we can use a saddle point approximation. It is crucial that together with the trivial minimum  $\varphi = \pm \mu/\sqrt{\lambda}$ , the classical equations of motion for imaginary time:

$$-\frac{\delta S}{\delta \varphi(t)} = \ddot{\varphi} + \mu^2 \varphi - \lambda \varphi^3 = 0 \quad (4.5)$$

have a solution:

$$\varphi(t) = \frac{\mu}{\sqrt{\lambda}} \tanh\left(\frac{\mu(t - t_0)}{\sqrt{2}}\right) \quad (4.6)$$

( $t_0$  is an arbitrary constant). This solution (being a local minimum for  $S[\varphi]$ ) connects the left well at  $-\infty$  with the right one at  $+\infty$ . Substituting (4.6) into (4.1) we find a classical action:

$$S_{cl} = \frac{(2\sqrt{2})\mu^3}{3\lambda} \quad (4.7)$$

At first glance the contribution of the trajectory (4.6) to the functional integral has the factor

$$e^{-S_{cl}} = e^{-(2\sqrt{2})\mu^3/3\lambda} \quad (4.8)$$

and is negligible. However this is not so. The reason is that we have not one classical solution but a continuum set of them distinguished by the value of  $t_0$ . Therefore we have to expect that the contribution to  $Z$  has the form:

$$Z \sim e^{-(2\sqrt{2})\mu^3/3\lambda} \int dt_0 \quad (4.9)$$

This means that for relatively short periods of time,  $t < e^{-(2\sqrt{2})\mu^3/3\lambda}$ , the contribution of our trajectory is indeed irrelevant, but in the large time

limit it becomes very large. This is in complete agreement with the tunnelling interpretation of our solution, namely the characteristic time we have introduced, is just the tunnelling time and hence the time for restoration of symmetry.

Let us perform now a more quantitative analysis.

Let us expand our field near the classical solution  $\varphi_{cl}(t - t_0)$ :

$$\varphi(t) = \varphi_{cl}(t - t_0) + \sum_{n \neq 0} C_n \psi_n(t - t_0) \quad (4.10)$$

In (4.10) we have introduced functions  $\psi_n$  which are normal modes for the oscillations near  $\varphi_{cl}$ . They are to be found from

$$\int dt_1 \left. \frac{\delta^2 S_{cl}}{\delta\varphi(t)\delta\varphi(t_1)} \right|_{\varphi=\varphi_{cl}(t)} \psi_n(t_1) = \omega_n^2 \psi_n(t) \quad (4.11)$$

or

$$\ddot{\psi}_n + \mu^2 \psi_n - 3\lambda\varphi_{cl}^2 \psi_n = -\omega_n^2 \psi_n$$

In the complete set of functions  $\psi_n$  there exists  $\psi_0 \sim \dot{\varphi}_{cl}$  for which  $\omega_0^2 = 0$ . Its existence, being a consequence of translation invariance, is easily checked directly by differentiating equation (4.5) with respect to  $t$ . In the expansion (4.10) we did not include  $\psi_0$  in the sum, introducing instead the parameter  $t_0$ . The reason for this is that while fluctuation of the  $C'_n$  are small, bounded by the action  $S$ , this action does not depend on  $t_0$  and this degree of freedom has to be treated separately. In order to do this let us pass from the integration over  $\varphi(t)$  to the integration over  $C_n$  and  $t_0$ . The easiest way to find the corresponding Jacobian is to examine the metric in the functional space:

$$\begin{aligned} \|\delta\varphi\|^2 &= \int dt (\delta\varphi(t))^2 \approx (\delta t_0)^2 \int dt \dot{\varphi}_{cl}^2 \\ &+ \sum_{n \neq 0} (\delta C_n)^2 + O((\delta t_0)^3) \end{aligned} \quad (4.12)$$

Therefore:

$$\begin{aligned} \mathcal{D} \varphi(t) &= A \prod_{n \neq 0} dC_n dt_0 \\ A &= \left( \int dt \dot{\varphi}_{cl}^2 \right)^{+1/2} \end{aligned} \quad (4.13)$$

and with this knowledge we can proceed to calculate the one kink contribution to the correlation function:

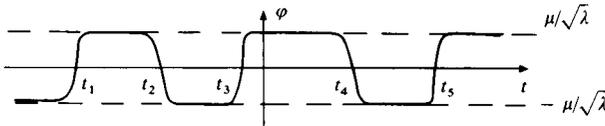
$$\begin{aligned}
 \langle \varphi(t_1)\varphi(t_2) \rangle &\approx \frac{\mu^2/\lambda + B e^{-S_{c1}} \int dt_0 \varphi_{c1}(t_1 - t_0)\varphi_{c1}(t_2 - t_0)}{1 + B e^{-S_{c1}} \int dt_0} \\
 &\approx \frac{\mu^2}{\lambda} + B e^{-S_{c1}} \int_{-\infty}^{+\infty} dt_0 (\varphi_{c1}(t_1 - t_0)\varphi_{c1}(t_2 - t_0) - \mu^2/\lambda) \\
 &B = A \frac{\int \prod d\xi_n e^{-\Sigma \omega_n^2 \xi_n^2}}{\int \prod d\xi_n e^{-\Sigma \omega_{n,0}^2 \xi_n^2}}
 \end{aligned} \tag{4.14}$$

( $\omega_{n,0}$  are eigenfrequencies for the trivial minima  $\varphi_{c1}^2 = \mu^2/\lambda$ ).

Substituting (4.6) into (4.14) we obtain:

$$\begin{aligned}
 \langle \varphi(t_1)\varphi(t_2) \rangle &= \frac{\mu^2}{\lambda} (1 - C e^{-S_{c1}} |t_1 - t_2|) \\
 C &= -B \int_{-\infty}^{+\infty} dx \left( \tanh \frac{\mu x}{\sqrt{2}} |t_1 - t_2| \tanh \frac{\mu(x-1)}{\sqrt{2}} |t_1 - t_2| - 1 \right) \\
 &\approx 2B \quad |t_1 - t_2| \gg \mu^{-1}
 \end{aligned} \tag{4.15}$$

As was expected, for large times the kink solution (4.6) gives a large contribution. Moreover, it is clear that we have to take into account multi-kink configurations, which in the “tunnelling” language correspond to trajectories travelling from left to right and back several times, for which  $\varphi(t)$  has the following structure:



There is no exact classical solution of such a kind, because there is a tendency for kink and anti-kink to annihilate. The attractive force between them is easily estimated. Since the tails of kinks are exponentially small:

$$\varphi_{c1}(t) \simeq \pm \frac{\mu}{\sqrt{\lambda}} (1 + O e^{-2\mu(t)/\sqrt{2}}) \quad |t| \gg 1 \tag{4.16}$$

a mere superposition of kink and anti-kink will have the action:

$$\begin{aligned} S^{(k, \bar{k})} &= S^{(k)} + S^{(\bar{k})} + O(e^{-2\mu|t_{12}|/\sqrt{2}}) \\ &= 2S^{(k)} + O(e^{-2\mu|t_{12}|/\sqrt{2}}) \end{aligned} \quad (4.17)$$

where  $t_{12}$  is the distance between  $k$  and  $\bar{k}$ . This result implies that the interaction between kinks is exponentially small. Since for small  $\lambda$  the average time distance between our objects is of the order of  $c^{-1} e^{2\sqrt{2}\mu^3/3\lambda} \gg \mu^{-1}$  the above-mentioned interaction can be neglected. Another simplification, possible for the same reason, is that the width of the kink can be neglected. We can approximate the general configuration by:

$$\varphi_{\text{cl}}^{(N)}(t) = \frac{\mu}{\sqrt{\lambda}} \prod_{j=1}^N \text{sgn}(\tau - \tau_j) \quad (4.18)$$

$$S_{\text{cl}}^{(N)} = N \frac{2\sqrt{2}\mu^3}{3\lambda} \quad (4.19)$$

$$\tau_1 \leq \tau_2 \leq \dots \leq \tau_N$$

The total effect on the correlation function is given by:

$$\langle \varphi(t_1)\varphi(t_2) \rangle = \frac{\mu^2}{\lambda} \sum_{N=0}^{\infty} (-C)^N e^{-S_{\text{cl}}^{(N)}} \int d\tau_1 \dots d\tau_N = \frac{\mu^2}{\lambda} e^{-\Delta|t_1 - t_2|} \quad (4.20)$$

$$\min(t_1, t_2) < \tau_1 < \tau_2 < \dots < \tau_N < \max(t_1, t_2), \quad \Delta = C e^{-2\sqrt{2}\mu^3/3\lambda}$$

We have found that tunnelling trajectories (which are also called instantons) remove the degeneracy of the ground state present on a perturbative level. The symmetry  $\varphi \rightarrow -\varphi$  gets restored and the system acquires a finite, though large, correlation length  $r_c \sim \Delta^{-1}$ . As we discussed in Chapter 1, in the limit of large correlation length the quantum theory with action (4.1) must be equivalent to the  $\mathcal{D} = 1$  Ising model.† This equivalence is quite obvious from the present consideration: the moment we replaced the  $\varphi$ -field by the step function (4.16) we actually started counting configurations of the Ising model described by the picture:



and the counting of kinks in (4.20) is exactly the counting of spin reversals in (1.18)

† This fact was realized long ago by Vaks and Larkin.

## 4.2 Instantons in the Global $O(2)$ Model

This model is described by the partition function:

$$Z = \int_{-\pi}^{\pi} \prod_{\mathbf{x}} \frac{d\varphi_{\mathbf{x}}}{2\pi} \exp(\beta \sum_{\mathbf{x}, \delta} (\cos(\varphi_{\mathbf{x}} - \varphi_{\mathbf{x}+\delta}) - 1)) \quad (4.21)$$

We shall consider its properties in the large- $\beta$  limit (weak coupling): the natural thing to do would be to expand the cosine in (4.21) and to write:

$$S = \frac{\beta}{2} \sum_{\mathbf{x}, \delta} (\varphi_{\mathbf{x}} - \varphi_{\mathbf{x}+\delta})^2 \approx \frac{\beta}{2} \int d^2x (\nabla\varphi)^2 \quad (4.22)$$

However, this expansion is not entirely correct. It restricts  $\varphi_{\mathbf{x}}$  to be very close to its neighbour  $\varphi_{\mathbf{x}+\delta}$ . This is physically reasonable since at large  $\beta$  neighbouring spins must be almost parallel. However, almost parallel spins do not always mean  $\varphi_{\mathbf{x}} \simeq \varphi_{\mathbf{x}+\delta}$  since we can have as well the situation:

$$\begin{aligned} \varphi_{\mathbf{x}} &\simeq \pi - \epsilon \\ \varphi_{\mathbf{x}+\delta} &\simeq -\pi + \epsilon \end{aligned} \quad \epsilon \ll 1 \quad (4.23)$$

The configuration (4.23), for which  $\varphi_{\mathbf{x}+\delta} - \varphi_{\mathbf{x}} = 2\epsilon - 2\pi$  must be just as important as configurations with  $\varphi_{\mathbf{x}+\delta} - \varphi_{\mathbf{x}} \sim \epsilon$ , but in the expansion (4.22), where we have lost the periodicity of the cosine, (4.23) is strongly suppressed. We must find the remedy for this unphysical situation. There are several ways of doing it. The most elegant one is to consider a continuum limit (4.22) but to allow  $\varphi$  to be a multivalued function, so that it has  $2\pi$  jumps at certain branch cuts. We shall return to this approach but first it is useful to work out the theory on a lattice, and then to see how this multivalued field arises.

Our aim is to retain the harmonic approximation for the  $\varphi$ -field, but to account for the configuration with (4.23). This aim can be achieved by replacing (4.21) by:

$$Z = \sum_{\{n_{\mathbf{x}, \delta}\}} \int_{-\pi}^{\pi} \prod_{\mathbf{x}} \frac{d\varphi_{\mathbf{x}}}{2\pi} \exp\left(-\frac{\beta}{2} \sum_{\mathbf{x}, \delta} (\varphi_{\mathbf{x}} - \varphi_{\mathbf{x}+\delta} + 2\pi n_{\mathbf{x}, \delta})^2\right) \quad (4.24)$$

where the  $n$  are arbitrary integers. Formally, we have replaced the function  $\exp(\beta(\cos \varphi - 1))$  by

$$\theta(\varphi) = \sum_{n=-\infty}^{+\infty} \exp\left(-\frac{\beta}{2}(\varphi - 2\pi n)^2\right) \quad (4.25)$$

In the large  $\beta$  limit when the only important property of the action was its periodicity and anharmonic terms  $\sim \varphi^4$  are irrelevant as was discussed in Chapter 1, the replacement of (4.21) by (4.25) is legitimate. With (4.24), which is periodic, we properly take into account the formally discontinuous configurations (4.23).

The partition function (4.24) can be transformed into a physically meaningful form. In order to do this let us take the case  $\mathcal{D} = 2$  and characterize the set of integers  $\{n_{\mathbf{x},\delta}\}$  by the integers  $q_{\mathbf{x}^*}$  (where  $\mathbf{x}^*$  are the centres of plaquettes) defined as

$$\begin{aligned} q_{\mathbf{x}^*} &= n_{\mathbf{x},1} + n_{\mathbf{x}+1,2} - n_{\mathbf{x}+2,1} - n_{\mathbf{x},2} \\ &= \sum_{\square} n_{\mathbf{x},\delta} \end{aligned} \quad (4.26)$$

In other words  $q_{\mathbf{x}^*}$  is a “field strength” created by the “vector potential”  $n_{\mathbf{x},\delta}$ . Any set  $n_{\mathbf{x},\delta}$  can be represented by:

$$n_{\mathbf{x},\delta} = m_{\mathbf{x}} - m_{\mathbf{x}+\delta} + \alpha_{\mathbf{x}} - \alpha_{\mathbf{x}+\delta} + \varepsilon_{\delta\gamma}(\phi_{\mathbf{x}^*} - \phi_{\mathbf{x}^*-\gamma}) \quad (4.27)$$

Here  $\varepsilon_{\delta\gamma}$  is the standard antisymmetric tensor, the  $m_{\mathbf{x}}$  are integers,  $|\alpha_{\mathbf{x}}| < 1$ , and  $\phi_{\mathbf{x}^*}$  satisfies the equations:

$$\Delta_{\mathbf{x}^*,\mathbf{x}^*} \phi_{\mathbf{x}^*} \equiv \sum_{\gamma} (4\phi_{\mathbf{x}^*} - \phi_{\mathbf{x}^*-\gamma} - \phi_{\mathbf{x}^*+\gamma}) = q_{\mathbf{x}^*} \quad (4.28)$$

The decomposition (4.27) splits  $n_{\mathbf{x},\delta}$  into longitudinal and transverse parts. The lattice Laplace equation (4.28) is obtained from (4.27) by forming the “field strength” (4.26), to which only  $\phi$  contributes. If we form the lattice divergence of  $n_{\mathbf{x},\delta}$  we get an equation which determines  $m_{\mathbf{x}}$  and  $\alpha_{\mathbf{x}}$  in terms of  $\{n_{\mathbf{x},\delta}\}$ .

Summation over  $\{n_{\mathbf{x},\delta}\}$  can be replaced by summations over  $\{m_{\mathbf{x}}\}$  and  $\{\alpha_{\mathbf{x}}\}$ . Substituting (4.27) into (4.24) and changing variables in each term by

$$\varphi_{\mathbf{x}} \rightarrow \varphi_{\mathbf{x}} - 2\pi(m_{\mathbf{x}} + \alpha_{\mathbf{x}}) \quad (4.29)$$

we obtain:†

$$\begin{aligned}
 Z &= \sum_{\{q_{\mathbf{x}_i}\}} \int_{-\infty}^{+\infty} \prod_{\mathbf{x}} \frac{d\varphi_{\mathbf{x}}}{2\pi} \\
 &\quad \times \exp\left(-\frac{\beta}{2}\left(\sum_{\mathbf{x},\delta}(\varphi_{\mathbf{x}} - \varphi_{\mathbf{x}+\delta})^2 + 2\pi^2 \sum_{\mathbf{x},\gamma}(\phi_{\mathbf{x}_*} - \phi_{\mathbf{x}_*-\gamma})^2\right)\right) \quad (4.30) \\
 &= Z_{\text{Gauss}} \sum_{\{q_{\mathbf{x}_i}\}} \exp\left(-\frac{\beta}{2} \sum_{\mathbf{x}_*,\mathbf{x}_i} \Delta_{\mathbf{x}_*,\mathbf{x}_i}^{-1} q_{\mathbf{x}_*} q_{\mathbf{x}_i} 2\pi^2\right) \\
 Z_{\text{Gauss}} &= \prod_{\mathbf{x}} \int_{-\infty}^{+\infty} \frac{d\varphi_{\mathbf{x}}}{2\pi} \exp\left(-\frac{\beta}{2} \sum_{\mathbf{x},\delta}(\varphi_{\mathbf{x}} - \varphi_{\mathbf{x}+\delta})^2\right)
 \end{aligned}$$

In deriving (4.30) we have used the fact that the replacement (4.29)

$$\int_{-\pi}^{\pi} d\varphi \Rightarrow \int_{-\pi+2\pi(m+\alpha)}^{\pi+2\pi(m+\alpha)} d\varphi$$

and summation on  $m$  are thus equivalent to replacing

$$\int_{-\pi}^{\pi} \frac{d\varphi}{2\pi} \Rightarrow \int_{-\infty}^{+\infty} \frac{d\varphi}{2\pi}$$

The formula (4.30) has a remarkable physical interpretation. It shows that in order to account for the periodicity of the action in the large  $\beta$  limit one has to introduce a set of vortices into the system, which interact according to the two-dimensional Coulomb law (the inverse Laplacian in (4.30)). Let us examine the correspondence between the distribution of  $\{q_{\mathbf{x}}\}$  and configuration of angles  $\{\varphi_{\mathbf{x}}\}$ . Take the case when only one vortex is present at  $\mathbf{x} = 0$ ,  $q_0 = 1$ . Take a large closed loop  $L$  on the lattice surrounding  $\mathbf{x} = 0$ , and examine  $B_{\mathbf{x},\delta} = \varphi_{\mathbf{x}} - \varphi_{\mathbf{x}+\delta}$  along this loop. From the definition it is clear that:

$$\oint_L B_{\mathbf{x},\delta} = 0 \quad (4.31)$$

---

† The whole set of arguments, leading to (4.30) first appeared in V. G. Berezinsky's Ph.D. Thesis.

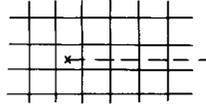
(Here we have denoted by  $\oint_L$  the sum of  $B_{\mathbf{x},\delta}$  along the loop; in the continuum limit it goes to an ordinary contour integral). As we noticed at the beginning of this section,  $|B_{\mathbf{x},\delta}| \ll 1 \pmod{2\pi}$  at large  $\beta$ . The set of integers  $\{n_{\mathbf{x},\delta}\}$  which we have introduced is defined so that:

$$-\pi < A_{\mathbf{x},\delta} = \varphi_{\mathbf{x}} - \varphi_{\mathbf{x}+\delta} + 2\pi n_{\mathbf{x},\delta} < \pi$$

and for large  $\beta$ ,  $|A_{\mathbf{x},\delta}| \ll 1$ . The quantity  $A_{\mathbf{x},\delta}$ , uniquely defined for a given configuration  $\{\varphi_{\mathbf{x}}\}$ , has a nonzero circulation, equal to the vorticity:

$$q_0 = \frac{1}{2\pi} \oint_L A_{\mathbf{x},\delta} \quad (4.32)$$

A simple example of  $\{n_{\mathbf{x},\delta}\}$  leading to a unit vortex can be constructed from the following picture:



(4.33)

All links except those intersected by the dashed line have  $n_{\mathbf{x},\delta} = 0$ . Intersected links have  $n_{\mathbf{x},\delta} = 1$ . This picture clearly satisfies the condition (4.32) and corresponds to the angles  $\varphi_{\mathbf{x}}$  having a  $2\pi$ -jump on the dashed line. The exact shape of this line is irrelevant because its change is just a gauge transformation of  $\{n_{\mathbf{x},\delta}\}$ . In the continuum limit this dashed line becomes a cut in the complex plane with the branch point at the position of the vortex.

In the large- $\beta$  limit in two dimensions, owing to the long range properties of the two-dimensional Coulomb force, vortices are combined into neutral dipoles, and the system (4.30) is dielectric. Such dipoles have very small influence on the correlation functions and are irrelevant at large  $\beta$ . At some critical  $\beta$  the dipoles dissociate and we get a plasma of vortices. We shall not investigate these phenomena here (see e.g. [3]). Instead, it is conceptually important to explain how (4.30) could have been obtained, directly from the continuum theory, and that it is precisely the instanton approximation to (4.21).

As we have said, the action in the continuum limit has the form:

$$S = \frac{\beta}{2} \int (\partial_\mu \varphi)^2 d^2 x \quad (4.34)$$

Classical minima of this action are defined from the equation:

$$\partial^2 \varphi(\mathbf{x}) = 0 \quad (4.35)$$

If  $\varphi(x) \xrightarrow{x \rightarrow \infty} 0$ , eq. (4.35) has only the zero continuous solution. Let us suppose now, that we have introduced a set of vortices, placed at the points  $\{x_a\}$  with strengths  $\{q_a\}$ . This means that we integrate over the multivalued fields  $\varphi$  which change by  $2\pi q_a$  as we go around the point  $x_a$ . In this sector we do have a nontrivial classical solution of (4.35). It is given by:

$$\begin{aligned}\varphi &= \sum_{a=1}^N q_a \operatorname{Im} \log(z - z_a) \\ z &= x_1 + ix_2\end{aligned}\tag{4.36}$$

The formula (4.36) can be viewed as a continuum approximation to the classical solution, minimizing the action:

$$S = \sum_{x, \delta} u(\varphi_x - \varphi_{x+\delta})\tag{4.37}$$

with any periodic  $u(\varphi)$ :

$$\begin{aligned}u(\varphi + 2\pi) &= u(\varphi) \\ u(\varphi) &\approx \left(\frac{\beta}{2}\right)\varphi^2, \quad \varphi \ll 1\end{aligned}\tag{4.38}$$

This solution far from the singularity is universal and described by (4.36). The structure near  $x_a$  (the core of the vortex) depends on the detailed form of  $u$ , but fortunately it appears to be irrelevant.

Substituting (4.36) into (4.34) we obtain:

$$S_{cl} = \frac{\beta}{2} \left\{ \sum_{a \neq b} q_a q_b 2\pi \log \frac{R}{|x_a - x_b|} + \operatorname{const.} \sum_a q_a^2 \right\}\tag{4.39}$$

( $R$  being the size of the system; the second term is the self-energy of the vortex).

Noticing that  $(1/2\pi) \log(R/|x|)$  is just the inverse Laplacian of a two-dimensional Coulomb energy we see that taking account of instantons for (4.21) leads to the continuum version of formula (4.30). We see that the obscure transformations leading to (4.30) serve a simple purpose—they take account of vortices.

However, in the present case, owing to the “confinement” of vortices they do not have any qualitative effects (at large  $\beta$ ). In the next section we shall consider the case when such effects are present.

Let us mention what happens for  $\mathcal{D} = 3$ . In this case we have, instead of point-like singularities, singular lines. The field  $\varphi$  has a  $2\pi$  jump as we

go around such vortex line. The vortex lines carry energy and are directly observable in  $^4\text{He}$ .

To conclude this section, we describe the continuum description of  $O(2)$  systems. It is given by the Lagrangian:

$$\mathcal{L} = |\partial_\mu \phi|^2 + v(|\phi|^2) \quad (4.40)$$

where, as we explained in Chapter 1,  $v(|\phi|^2)$  can be taken without loss of generality to be:

$$v(|\phi|^2) = \frac{\mu_0^4}{\lambda} - \mu_0^2 |\phi|^2 + \frac{\lambda_0}{2} |\phi|^4 \quad (4.41)$$

The relation of this theory to the one described by (4.21) is the same as that of the quantum mechanics of a double well to the  $\mathcal{D} = 1$  Ising model. Their long range properties are identical. This is most easily seen in the present case if we introduce variables:

$$\phi(x) = u(x)e^{i\theta(x)}$$

$$S = \int \{(\partial_\mu u)^2 + v(u^2) + u^2(\partial_\mu \theta)^2\} dx. \quad (4.42)$$

For small values of  $\lambda_0$  we have

$$u(x) = \frac{\mu_0}{\sqrt{\lambda_0}} \left( 1 + O\left(\frac{\sqrt{\lambda}}{\mu}\right) \right) \quad (4.43)$$

and we see that the fluctuations of the modulus  $u(x)$  are small and short ranged (they have a mass  $\mu_0$ ). At the same time, the field  $\theta(x)$  is massless and the only one contributing at very large distances. It is worthwhile demonstrating how the vortex contribution arises directly in (4.40). As before, the vortex arises as a nontrivial classical minimum for the action (4.40). In the two-dimensional case the equations have the form:

$$\partial^2 \phi - \mu_0^2 \phi + \lambda |\phi|^2 \phi = 0 \quad (4.44)$$

or

$$\partial^2 \phi_a - \mu_0^2 \phi_a + \frac{\lambda}{2} \left( \sum_b^2 \phi_b^2 \right) \phi_a = 0$$

(where  $\phi_{1,2}$  are defined by  $\phi = (\phi_1 + i\phi_2)/\sqrt{2}$ ).

This equation has  $O(2) \otimes O(2)$  invariance, one of the  $O(2)$  being rotation of  $x$ -space and the other rotation of the  $\phi$ . Let us look for a solution which breaks this  $O(2) \otimes O(2)$  but preserves the single  $O(2)$

symmetry under simultaneous rotations of  $x$  and  $\phi$  spaces. The most general ansatz with such a property is:

$$\begin{aligned}\phi_a &= u(r) \frac{x_a}{r}, \text{ or } \phi = u(r)e^{i\varphi} \\ r^2 &= x_1^2 + x_2^2, \varphi = \text{arctg}(x_2/x_1)\end{aligned}\tag{4.45}$$

From the above symmetry consideration it has to be consistent with (4.44), and direct substitution confirms that it indeed is. We obtain an equation for  $u$ :

$$u'' - \mu_0^2 u - \frac{2}{r^2} u + \lambda u^3 = 0\tag{4.46}$$

There exists a solution to (4.46) with the properties:

$$u(r) \xrightarrow{r \rightarrow 0} 0, \quad u(r) \xrightarrow{r \rightarrow \infty} \mu_0/\sqrt{\lambda}\tag{4.47}$$

The reasons for the existence and stability of such a solution are the following. Substitution of (4.45) into (4.42) gives:

$$S = \int_0^\infty 2\pi r \, dr \{ (du/dr)^2 + v(u^2) + u^2/r^2 \}\tag{4.48}$$

In order to avoid quadratic divergence at infinity we must have  $v(u^2(\infty)) = 0$  or  $u^2(\infty) = \mu_0^2/\lambda$ . Now, because of the last term in (4.48) it is desirable to have  $u(0) = 0$ . Interpolation between zero and  $\mu^2/\lambda$  should not be too fast because of the first term and should not be too slow because of the second one. So we expect that there is a unique function  $u(r)$  which minimizes  $S$ . This is indeed the case as can be rigorously shown, but we content ourselves with the above heuristic consideration.

We have also to consider stability with respect to variation of the phase  $\theta$ . It is clear from (4.42) that if it were possible to deform  $\theta(x)$  continuously from  $\theta(x) = \varphi$  to  $\theta = 0$ , the action (4.42) would decrease. However, such a deformation is not possible, as follows from the condition that  $\phi$  is single valued and that every  $\theta(\varphi)$  has to satisfy:

$$\theta(2\pi) - \theta(0) = 2\pi q\tag{4.49}$$

with integer  $q$ . The solution (4.45) corresponds to  $q = 1$ , and it cannot be deformed without violation of single-valuedness of  $\phi$  to the solution with  $q = 0$ .

This fact guarantees the phase stability of (4.45). Notice that very similar topological considerations could have been used to prove stability of the kink solution of the preceding section; it also can be seen that in the case of a complex field the kink solution is unstable, because the kink can be deformed to nothing by continuous phase rotations.

The action for a single vortex is logarithmically divergent at large distances, due to the last term in (4.48). However, if we consider a neutral superposition of vortices and anti-vortices the total action will be finite, and in the limit when all relative distances are larger than  $\mu_0^{-1}$  it will be given by (4.39) with  $\beta$  replaced by  $2\mu_0^2/\lambda$ .

Our conclusion is that in the infrared limit we have three equivalent descriptions of the  $O(2)$  systems, given by the actions (4.21), (4.24) and (4.30).

*Remarks* 1. The analogy between the  $\mathcal{D} = 1$  Ising model and double-well quantum mechanics was first noted by Vaks and Larkin. The theory of vortices in the planar  $O(2)$  magnet and the effective description by  $\theta$ -functions was suggested by V. L. Berezinsky in his PhD thesis (1970) and rediscovered by Koesterlitz and Thouless.

2. The  $O(2)$  model is quite relevant to several physical problems. First of all it describes by its definition planar magnetic systems with  $O(2)$  symmetries. Secondly since the action (4.40) is the second quantized hamiltonian for the interacting Bose gas, the model describes two-dimensional  $^4\text{He}$  films at nonzero temperature (the static long range properties). It is also operative for the theory of two dimensional crystals. This occurs for the following reasons. If we denote by  $u_\alpha(x)$  a displacement of the atom placed at the point  $x$  of the crystal then in the infrared limit we have (according to the theory of elasticity):

$$E = \int \left\{ \frac{\mu}{4} (\partial_\alpha u_\beta + \partial_\beta u_\alpha)^2 + \frac{\lambda}{2} (\partial_\lambda u_\lambda)^2 \right\} d^2x \quad (4.50)$$

( $\lambda, \mu$  are the so-called Lamé constants). In order to find a partition function we have to compute:

$$Z = \int \mathcal{D}u_\alpha(x) e^{-\beta E[u]} \quad (4.51)$$

This would have been an easy Gaussian integral if the fields were single valued. However, just as phases  $\{\varphi_x\}$  were defined (mod  $2\pi$ ), displacements  $u_\alpha(x)$  are defined (mod  $b_\alpha$ ) where  $b_\alpha$  is a lattice vector, since on a

lattice the energy must be a periodic function of displacements. Therefore, we have to account for jumps in  $u_\alpha(\mathbf{x})$ . If we have a branch point at  $\mathbf{x}_0$ , and a loop  $L$  surrounding this point then:

$$\oint_L \partial_\beta u_\alpha(\mathbf{x}) dx^\beta = b_\alpha \quad (4.52)$$

For such  $u_\alpha(\mathbf{x})$  it is said that we have a dislocation located at  $\mathbf{x}_0$ . We see that the properties of dislocations almost coincide with those of vortices. In particular, at some  $\beta$  we have a condensation of dislocations which can be interpreted as melting.

### 4.3. Compact QED ( $O(2)$ Gauge Model)

In this section we shall examine the case of Abelian gauge theories. This case is nontrivial, in spite of the fact that the naive continuum limit of the action is given by:

$$S = \frac{1}{4e_0^2} \int d^D \mathbf{x} (\hat{c}_\mu A_\nu - \hat{c}_\nu A_\mu)^2 \quad (4.53)$$

and describes apparently free photons. The nontriviality, as in the preceding sections, comes from the fact that the vector potential has certain angular properties which force us to account for the analogues of vortices or dislocations in the functional integral.

Before we do this, let us explain why the vector potential is supposed to be an angular variable.

*A priori*, we can define on a lattice two different models. The first one is (1.49) with the action:

$$S = \frac{1}{2e_0^2} \sum_{\mathbf{x}, \alpha\beta} (1 - \cos F_{\mathbf{x}, \alpha\beta}) \quad (4.54)$$

$$F_{\mathbf{x}, \alpha\beta} = A_{\mathbf{x}, \alpha} + A_{\mathbf{x}+\alpha, \beta} - A_{\mathbf{x}+\beta, \alpha} - A_{\mathbf{x}, \beta}$$

$$-\pi \leq A_{\mathbf{x}, \alpha} \leq \pi$$

The second option is:

$$S = \frac{1}{4e_0^2} \sum_{\mathbf{x}, \alpha\beta} F_{\mathbf{x}, \alpha\beta}^2; \quad -\infty \leq A_{\mathbf{x}, \alpha} \leq +\infty \quad (4.55)$$

In the naive continuum limit both of these actions lead to (4.53), but the physics of these models is different. An analogous situation arises in

ordinary quantum mechanics with the action:

$$S = \frac{1}{2} \int \dot{x}^2 dt \quad (4.56)$$

This action can describe a free particle on a line or a free particle on a circle. The spectrum is continuous in the first case and is discrete in the second. The difference arises because in the first case we integrate over continuous  $x(t)$  while in the second we should allow  $2\pi$ -jumps.

We have to decide on physical grounds what version of QED is realized in Nature. The main reason why we believe in the periodic (or compact) version of QED is based on the empirical fact of charge quantization. We know that the ratio of any two electric charges is a rational number. Let us show that this is a necessary consequence of compact QED, while in the noncompact version it would be an unexplained mystery.

The qualitative argument is that, as we saw in Chapter 3, the electric flux (which is an analogue of angular momentum) is quantized. Since charged particles are sources of electric flux, which according to Gauss' theorem must be equal to their charges, we conclude that possible charges are quantized as well.

To be more explicit, let us consider two charged fields,  $\psi_x$  with a unit charge and  $\chi_x$  with charge  $e$ . The Lagrangian for these fields is of the form:

$$\mathcal{L}' = \sum_{\mathbf{x}, \delta} (\psi_{\mathbf{x}+\delta}^\dagger e^{iA_{\mathbf{x},\delta}} \psi_{\mathbf{x}} + \chi_{\mathbf{x}+\delta}^\dagger e^{ieA_{\mathbf{x},\delta}} \chi_{\mathbf{x}}) + \text{c.c.} \quad (4.57)$$

The form of (4.57) is dictated by gauge invariance:

$$\begin{aligned} A_{\mathbf{x},\delta} &\rightarrow A_{\mathbf{x},\delta} + \alpha_{\mathbf{x}+\delta} - \alpha_{\mathbf{x}} \\ \psi_{\mathbf{x}} &\rightarrow e^{i\alpha_{\mathbf{x}}} \psi_{\mathbf{x}}, \quad \chi_{\mathbf{x}} \rightarrow e^{ie\alpha_{\mathbf{x}}} \chi_{\mathbf{x}} \end{aligned} \quad (4.58)$$

From (4.57) we see that  $\mathcal{L}'$  is well defined in our phase space only if  $e$  is integer, which is needed for periodicity in  $A_{\mathbf{x},\delta}$ . Notice also that the period of the free action (4.54) defines a natural unit of charge.

In noncompact QED the flux is continuous and there are no reasons for charge quantization. Another important thing concerning compactness is that the two above-mentioned options are present only in the abelian case. For any Non-Abelian group the fact of compactness or noncompactness can be seen from its Lie algebra. For instance if the gauge group is  $SU(2)$  we cannot formulate the noncompact version at all. If we consider the noncompact group  $SU(1, 1)$ , we get gauge quanta with negative norm. Therefore, when we consider QED as arising as a

subgroup of some Non-Abelian gauge theory we are necessarily dealing with the compact version.

After this explanation let us work out the theory of this version starting from  $\mathcal{D} = 3$ . Following the strategy of the preceding section we consider a partition function:

$$Z = \sum_{\{n_{x,\alpha\beta}\}} \int_{-\pi}^{\pi} \prod_{x,\alpha} dA_{x,\alpha} \exp \left\{ -\frac{1}{4e_0^2} \sum_{x,\alpha\beta} (F_{x,\alpha\beta} - 2\pi n_{x,\alpha\beta})^2 \right\} \quad (4.59)$$

which takes account of the periodicity of the action. For a given set of  $\{n_{x,\alpha\beta}\}$  we introduce numbers  $q_z$  (where  $z$  belongs to the centres of the cubes of our lattice) defined as the flux of  $n$  through a given cube  $\sigma_z$ :

$$\oint_{\sigma_z} n_{x,\alpha\beta} = q_z \quad (4.60)$$

(the notation in the l.h.s. of (4.60) means the sums of  $n_{x,\alpha\beta}$  corresponding to oriented plaquettes forming a cube with centre at  $z$ . Its continuum analogue will be an integral of the field  $n_{\alpha\beta}$  over a closed surface). Decomposing  $n_{x,\alpha\beta}$  as:

$$\begin{aligned} n_{x\alpha\beta} = & m_{xx} + m_{x+\alpha,\beta} - m_{x+\beta,\alpha} - m_{x,\beta} \\ & + \lambda_{x,\alpha} + \lambda_{x+\alpha,\beta} - \lambda_{x+\beta,\alpha} - \lambda_{x,\beta} \\ & + \epsilon_{\alpha\beta\gamma}(\phi_z - \phi_{z-\gamma}) \end{aligned} \quad (4.61)$$

We obtain from (4.60):

$$\Delta_{zz'} \phi_{z'} = q_z \quad (4.62)$$

( $\Delta_{zz'}$  is the lattice Laplace operator). Substitution of (4.61) and (4.62) into (4.59) gives:

$$\begin{aligned} Z = & Z_{\text{Gauss}} \sum_{\{q_z\}} \exp \left( -\frac{1}{4e_0^2} \sum_{z,z'} q_z \Delta_{zz'}^{-1} q_{z'} \cdot 8\pi^2 \right) \\ Z_{\text{Gauss}} = & \int_{-\pi}^{+\pi} \prod_{x,\alpha} dA_{x,\alpha} \exp \left( -\frac{1}{4e_0^2} \sum_{x,\alpha\beta} F_{x,\alpha\beta}^2 \right) \end{aligned} \quad (4.63)$$

We thus obtain the same Coulomb system as (4.30) but in three dimensions. As we shall see that makes a lot of physical difference. Before working this out let us explain the meaning of the "charges" which appeared in (4.63). They can be thought of as instanton solutions

for the action (4.54). In the infrared limit the classical equations take the form:

$$\begin{aligned}\partial_\mu F_{\mu\nu} &= -\text{curl } \mathbf{H} = 0 \\ \text{div } \mathbf{H} &= 0, \mathbf{H} = \text{curl } \mathbf{A} = \frac{1}{2}\epsilon_{\alpha\beta\gamma} F_{\beta\gamma}\end{aligned}\quad (4.64)$$

They do not have nonsingular solutions except for  $\mathbf{H} = 0$ . However, as before, since (4.64) was obtained from a periodic action we can allow certain  $\delta$ -like singularities in  $\mathbf{H}$  with strengths  $2\pi q$ . Such a solution (analogous to (4.36)) can be easily given:

$$H_\mu(\mathbf{x}) = \frac{q}{2} \frac{x_\mu}{|\mathbf{x}|^3} - 2\pi q \delta_{\mu 3} \delta(x_1) \delta(x_2) \theta(x_3) \quad (4.65)$$

( $\theta$  is the step function).

This solution represents a magnetic charge, sitting at the origin. Its magnetic flux  $2\pi q$  created by the first term is compensated by the ingoing flux along the third axis. In two-dimensions we had the analogue of (4.65):

$$A_\mu = \partial_\mu \varphi = q \epsilon_{\mu\nu} (x_\nu/x^2) - 2\pi q \theta(x_2) \delta(x_1) \quad (4.66)$$

where the second term arose from the discontinuity of the phase. Owing to the periodicity of the action, the second terms in both (4.66) and (4.65) do not contribute to physical quantities. Therefore, the general instanton configuration in our case is described by a set of magnetic charges with Coulomb interactions and has the action:

$$S_{\text{INST}} = \frac{1}{4e_0^2} \sum_{a \neq b} \frac{q_a q_b \cdot 2\pi}{|\mathbf{x}_a - \mathbf{x}_b|} + \frac{\text{const.}}{e_0^2} \sum_a q_a^2 \quad (4.67)$$

This is the same result as (4.63), provided that these charges are far apart. Let us show now that due to the disordering effect of instantons the system acquires a finite correlation length (the photon becomes massive). To show this we shall use the following functional representation for the instanton part of  $Z$ :

$$\begin{aligned}Z &= Z_{\text{Gauss}} \cdot Z_{\text{INST}} \\ Z_{\text{INST}} &= \sum_{N, \{q_a\}} \frac{\zeta^N}{N!} \int \prod_{j=1}^N d\mathbf{x}_j \exp\left(-\frac{\pi}{2e^2} \sum_{a \neq b} \frac{q_a q_b}{|\mathbf{x}_a - \mathbf{x}_b|}\right) \\ &= \int \mathcal{D}\chi(\mathbf{x}) \exp\left\{-\left(\frac{e}{2\pi}\right)^2 \int (\nabla\chi)^2\right\} \sum_N \sum_{\{q_a\}} \frac{\zeta^N}{N!} \int d\mathbf{x}_1 \dots d\mathbf{x}_N \\ &\quad \times \exp\left(i \sum_a q_a \chi(\mathbf{x}_a)\right) \\ &\approx \int \mathcal{D}\chi(\mathbf{x}) \exp\left\{-\left(\frac{e}{2\pi}\right)^2 \int ((\nabla\chi)^2 - M^2 \cos \chi)\right\} d\mathbf{x}\end{aligned}\quad (4.68)$$

Here:

$$\xi = e^{-\text{const}/e^2} \quad M^2 = \left(\frac{2\pi}{e}\right)^2 \xi$$

The formula (4.63) was obtained through the use of the general property of Gaussian integrals:

$$\int \prod_i dx_i \exp\left(-\frac{1}{2} \sum_{i,j} A_{ij} x_i x_j + i \sum b_i x_i\right) = (\text{Det } A)^{-1/2} \exp\left(-\sum_{i,j} A_{ij}^{-1} b_i b_j\right) \quad (4.69)$$

and also by taking into account only  $q_a = \pm 1$ . The last approximation will be justified when we check that for small  $e^2$  the monopoles are far apart (like the kinks in Section 4.1) and therefore monopoles with  $q > 1$  have a tendency to dissociate to monopoles with  $q = 1$ . Actually taking account of  $q > 1$  would lead to terms  $\sim \cos(q\chi)$  in (4.68). The functional integral (4.68) supplies us with a diagrammatic expansion for the monopole plasma. However, the effective nonlinearity in (4.68) is exponentially small, because the coefficient  $g$  of  $\chi^4$  in (4.63) is of the order of

$$g \sim \left(\frac{M}{e}\right)^2 \sim e^{-\text{const}/e^2} \ll 1 \quad (4.70)$$

This result could have been anticipated since it corresponds to the condition for validity of the Debye or mean field approximation. For this to hold it is necessary that in the Debye volume, of order  $M^{-3} \sim \exp(-3 \text{const.}/e^2)$ , there is a large number of particles, so that the fluctuations in the sum of their individual fields may be neglected. But according to the Boltzmann formula, the density of particles is given by

$$n \sim e^{-\text{const}/e^2} \quad (4.71)$$

Hence the criterion for the mean field approximation is

$$nM^{-3} \sim e^{\text{const}/e^2} \gg 1 \quad (4.72)$$

which is the same as before.

Now let us calculate certain correlation functions. We shall concern ourselves only with gauge-invariant quantities. As an intermediate step it is convenient to have an expression for the generating functional for the charge density of the plasma. After simply repeating the derivation of (4.68) we get:

$$\left\langle \exp\left(i \int \rho(x)\eta(x) dx\right) \right\rangle = Z[\eta(x)]/Z[0] \quad (4.73)$$

where

$$\rho(\mathbf{x}) = \sum_a q_a \delta(\mathbf{x} - \mathbf{x}_a)$$

$$Z[\eta] = \int \mathcal{D}\chi \exp \left\{ - \left( \frac{e}{2\pi} \right)^2 \int [(\nabla(\chi - \eta))^2 - M^2 \cos \chi] \right\} \quad (4.74)$$

The simplest correlation functions for our problem are those of the operator

$$H_\alpha(\mathbf{x}) = \frac{1}{2} \epsilon_{\alpha\beta\gamma} F_{\beta\gamma}(\mathbf{x}). \quad (4.75)$$

At large distances this is just the electromagnetic field strength.

In the quasiclassical approximation,  $H_\mu(\mathbf{x})$  is connected with the charge density as follows:

$$H_\mu(\mathbf{x}) = \frac{1}{2} \int d^3y \frac{(x-y)_\mu}{|\mathbf{x}-\mathbf{y}|^3} \rho(\mathbf{y})$$

$$H_\mu(\mathbf{x}) = \frac{2\pi i k_\mu}{k^2} \rho(k) \quad (4.76)$$

Using formula (4.74) we get:

$$\langle \rho(k) \rho(-k) \rangle = \left( \frac{e}{2\pi} \right)^2 k^2 - k^4 \langle \chi(k) \chi(-k) \rangle \left( \frac{e}{2\pi} \right)^4$$

$$\left( \frac{2\pi i}{e} \right)^N \langle \rho(k_1) \dots \rho(k_N) \rangle = \prod_{j=1}^N k_j^2 \langle \chi(k_1) \dots \chi(k_N) \rangle \quad (4.77)$$

Now the correlation function of the  $H$ -field is given by

$$\langle H_\mu(k) H_\nu(-k) \rangle = \langle H_\mu(k) H_\nu(-k) \rangle^{(0)} + (2\pi)^2 \frac{k_\mu k_\nu}{k^4} \langle \rho(k) \rho(-k) \rangle \quad (4.78)$$

The first term in (4.78) is the bare (that is without monopoles) Green function of the  $H$ -field. It has the form

$$\langle H_\mu H_\nu \rangle^{(0)} = e^2 \left( \delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2} \right) \quad (4.79)$$

Its singularity at  $k = 0$  reflects the existence of a massless photon in this approximation. Using (4.77) and the previous comment about the small coupling of the  $\chi$ -field we get

$$\langle \rho(k) \rho(-k) \rangle = \left( \frac{e}{2\pi} \right)^2 \left( k^2 - \frac{k^4}{M^2 + k^2} \right) = \left( \frac{e}{2\pi} \right)^2 \frac{M^2 k^2}{M^2 + k^2} \quad (4.80)$$

From formula (4.78)-(4.80) we obtain

$$\begin{aligned} \frac{1}{e^2} \langle H_\mu(k) H_\nu(-k) \rangle &= \delta_{\mu\nu} - \frac{k_\mu k_\nu}{k^2} + \frac{k_\mu k_\nu}{k^2} \frac{M^2}{M^2 + k^2} \\ &= \delta_{\mu\nu} - \frac{k_\mu k_\nu}{M^2 + k^2} \end{aligned} \quad (4.81)$$

This formula implies that there are no massless particles in the theory and instead we have a massive scalar particle with a small mass  $M$ . Analogously we find the result

$$\langle H_{\mu_1}(k_1) \dots H_{\mu_j}(k_j) \rangle_{\text{conn}} = e^j k_{\mu_1} \dots k_{\mu_j} \prod_j \frac{1}{k_j^2 + M^2} \quad \left( \sum_j k_j = 0 \right) \quad (4.82)$$

The qualitative explanation of the above result is the following. In our system there is a finite density of pseudoparticles with long range interaction and their random fields spoil the correlation. From the strong coupling expansion we know that the correlation length is nonzero also for  $e^2 \gg 1$ . It is reasonable to assume that there is no phase transition in this system and that the confinement regime continues to weak coupling. As we shall show in the next chapter, this is indeed true. Here we shall discuss another implication of the result.

First of all let us notice that since we showed that the  $O(2)$  system for  $\mathcal{D} = 3$  is disordered (has a mass gap) the same must be true for nonabelian systems. Indeed, let us take the case of  $SU(2)$  and constrain the  $B_{x,\alpha}$  so that they lie in  $O(2) \subset SU(2)$ . One should expect that this constraint increases the order in the system and if the constrained system is disordered the unconstrained one must be even more disordered and have an even larger gap. There is no doubt that this statement is correct but a rigorous proof has yet to be given.

As we turn to  $\mathcal{D} = 4$  the picture described above changes. Three-dimensional cubes embedded into four-dimensional space have four different orientations (analogously to the three different orientations for squares embedded in 3-space). Hence we expect that the number  $q$  in (4.60) for  $\mathcal{D} = 4$  will have a directional index  $\alpha: q = q_{x,\alpha}$ . To check this, let us write down the decomposition analogous to (4.61) for the numbers  $N_{x,\alpha\beta}$  in the case  $\mathcal{D} = 4$ :

$$N_{x,\alpha\beta} = \partial_\alpha m_\beta - \partial_\beta m_\alpha + \partial_\alpha \lambda_\beta - \partial_\beta \lambda_\alpha + \epsilon_{\alpha\beta\gamma\delta} \partial_\gamma \phi_\delta \quad (4.83)$$

Here we have used the notation  $\partial_\alpha$  for the lattice derivatives, e.g.:

$$\partial_\alpha m_\beta = m_{x,\beta} - m_{x-\alpha,\beta} \quad (4.84)$$

Taking the surface integral over the surface of the cube with direction  $\gamma$  we obtain the analogue of eq. (4.62)

$$-\partial^2\phi_\gamma + \partial_\gamma\partial_\delta\phi_\delta = q_\gamma \quad (4.85)$$

Here we have used a continuum notation in the hope that after previous discussions it will not be misleading. Formulas with explicit lattice notations would look rather clumsy for  $\mathcal{D} = 4$ .

From eq. (4.85) we see that  $q_\gamma$  are subjected to the constraint:

$$\partial_\gamma q_\gamma = \sum_\gamma (q_{\mathbf{x},\gamma} - q_{\mathbf{x}-\boldsymbol{\gamma},\gamma}) = 0 \quad (4.86)$$

The instanton part of the partition function has the form:

$$Z_{\text{INST}}^{\mathcal{D}=4} = \sum_{(q_{\mathbf{x},\gamma}; \partial_\gamma q_\gamma = 0)} \exp\left(-\frac{1}{4e_0^2} \sum_{\mathbf{x}, \mathbf{x}'} 2(2\pi)^2 q_{\mathbf{x},\gamma} \Delta_{\mathbf{x},\mathbf{x}'}^{-1} q_{\mathbf{x}',\gamma}\right) \quad (4.87)$$

Let us explain the meaning of these results, which could have been anticipated without the derivation given. The instanton for  $\mathcal{D} = 3$  was described by a magnetic monopole solution (4.65). Suppose, we add one more dimension (“time”). Then this solution, being point-like for  $\mathcal{D} = 3$  is represented by a line for  $\mathcal{D} = 4$  (the world-line of the point-like object). We can take the line to be arbitrarily curved. Then we shall have for each shape a different classical solution, minimizing locally the action (4.54). Since magnetic flux is conserved these lines have to be closed or infinite. The contribution to the partition function can be expressed as a sum over all possible magnetic flux lines, which have Coulomb attraction between them. This picture is quantitatively reflected in formulas (4.83), (4.86), (4.87).

It is easy to deduce that for large  $e_0^{-2}$  those flux lines have negligible influence on the system in the infrared limit. The reason is that lines of length  $L$  have action  $\sim L$  and their contribution to (4.87) is given by:

$$Z_L^{(1)} \sim e^{-cL/e_0^2(c')^L} \quad (4.88)$$

(Here  $(c')^L$  is the number of loops of the length  $L$ ). We conclude that for large enough  $e_0^{-2}$  magnetic loops are small and we have no disordering in the system, which is effectively described in this case by means of massless free photons. At the same time there is a critical coupling  $e_{\text{cr}}^2$  at which condensation of magnetic lines begins. It is to be expected that for  $e_0^2 > e_{\text{cr}}^2$  we obtain a strong coupling phase with confinement of charge, while for  $e_0^2 < e_{\text{cr}}^2$  we obtain the photon phase.

The picture described above can be understood in yet another way. Magnetic monopoles which were instantons for  $\mathcal{D} = 3$  become particles

for  $\mathcal{D} = 4$ , being minima of the potential energy for the hamiltonian (3.34). This point will be discussed in detail in Chapter 6. Here it suffices to realize that each time-independent classical solution with finite energy (a soliton) corresponds in the limit of small coupling to a stable particle. As we go to  $\mathcal{D} = 4$  the quantity which was action in  $\mathcal{D} = 3$  becomes energy.

The closed loops we discussed above are the world lines of monopole-antimonopole pairs. We see that at some coupling the vacuum becomes filled with a monopole condensate. Such a medium confines electric charges as is seen from the following physical analogy. Take a superconductor in which in the ground state we have a condensate of electrically charged fields (Cooper pairs). It is known that an external magnetic field can penetrate inside a superconductor only by forming a number of thin filaments, carrying quantized magnetic flux. If we imagine two magnetic charges inside the superconductor, we shall deduce from the Ginzburg-Landau equations that their magnetic flux should be concentrated inside such a filament, connecting them. Their interaction energy is proportional to the distance. If in this description we interchange words "electric" and "magnetic" we conclude that two electric charges in a medium formed by a monopole condensate are confined.

Can we have an Abelian theory which confines even at  $\mathcal{D} = 4$  for all couplings? We see that this equation is equivalent to finding a system with point-like finite action instantons. It is easy to give an example of such a system. Let us consider a "gauge field of the third rank", namely consider as primary variables not vector potentials  $A_{x,\alpha}$  but tensors  $F_{x,\alpha\beta}$  attached to plaquettes. Form a field strength, associated with cubes:

$$\phi_x = \epsilon_{\alpha\beta\gamma\delta} \partial_\beta F_{\gamma\delta} \quad (4.89)$$

and consider the action

$$S = \frac{1}{e_0^2} \sum_{x,\alpha} (1 - \cos \phi_{x,\alpha}) \quad (4.90)$$

Literal repetition of the above consideration would lead us to the following instanton part of the partition function:

$$Z_{\text{INST}}^{\mathcal{D}=4} = \sum_{(q_x)} \exp\left(-\frac{1}{2e_0^2} \sum_{x,x'} q_x \Delta_{xx'}^{-1} q_{x'} (2\pi)^2\right) \quad (4.91)$$

In this case we have a  $\mathcal{D} = 4$  plasma of point-like instantons. We see that the system will acquire a mass gap owing to Debye screening.

Following the tradition of previous sections let us explain how the systems we have discussed look in the continuum formulation. As we have already said, compact QED can be obtained from Non-Abelian gauge theory. To give an example, let us consider  $SU(2)$  theory, which contains three gauge bosons, and break the  $SU(2)$  so as to make two of these bosons massive. This is the so-called Georgi-Glashow model for unifying weak and electromagnetic interactions. All more complicated unifications follow the pattern we encounter in this model. Its action is given by:

$$S = \int dx \left\{ \frac{1}{2} (\nabla_\mu \Phi)^2 - \frac{1}{2} \mu_0^2 \Phi^2 + \frac{\lambda}{4} (\Phi^2)^2 + \frac{1}{4e^2} (F_{\mu\nu})^2 \right\} \quad (4.92)$$

where  $\Phi = (\varphi_1, \varphi_2, \varphi_3)$  is a triplet of isotopic vector fields,  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + A_\mu \times A_\nu$  is the Yang-Mills field strength, and  $\nabla_\mu \Phi = \partial_\mu \Phi + A_\mu \times \Phi$ . If we consider an expansion of the fields near the absolute minimum of the potential energy:  $\varphi_{1,2} = 0$ ,  $\varphi_3 = \mu_0/\sqrt{\lambda}$  we find as a result the following particle content: heavy charged  $W^\pm$  bosons

$$W_\mu^\pm = (A_\mu^1 \mp iA_\mu^2)/\sqrt{2} \quad \text{with} \quad m_W^2 = e^2 \mu_0^2/\lambda \quad (4.93)$$

a scalar field  $\sigma = \varphi_3 - \mu_0/\sqrt{\lambda}$  with a mass  $m_\sigma^2 = 2\mu_0^2$ , and the electromagnetic field  $A_\mu^3$  which is massless. We see that in the infrared limit this theory describes free photons. All heavy fields can be viewed in this limit as a regularization of this photon theory, which replaces the lattice regularization. The fact that it is a compact version of photon theory reveals itself in the existence of nontrivial instantons for the action (4.92). They have the form (for  $\mathcal{D} = 3$ ):

$$A_\mu^a = a(r) \epsilon_{\mu ab} \frac{X_b}{r}$$

$$\varphi^a = u(r) \frac{X_a}{r}$$

with

$$\begin{aligned} u(0) = 0 & \quad u(\infty) = \mu_0/\sqrt{\lambda} \\ a(0) = 0 & \quad a(r) \sim -1/r, \quad r \rightarrow \infty \end{aligned}$$

The analysis which will be given in Chapter 6 shows that this solution has a finite action and describes magnetic charge. For the present purposes it is enough to understand that since  $A_\mu^a \sim 1/r$  and  $F_{\mu\nu}^a \sim 1/r^2$  as  $r \rightarrow \infty$  the interaction of two such objects,  $\int F_{\mu\nu}^2 d^3x \sim r^3/r^4 \sim 1/r$

and we obtain a Coulomb force. Therefore all our considerations of compact QED in three dimensions work for the Georgi-Glashow model. In particular, in the weak coupling limit this model confines electric charges by a linear potential for  $\mathcal{D} = 3$  and describes an ordinary Coulomb electric force for  $\mathcal{D} = 4$ .

It would be quite interesting to have a continuum formulation for the compact theory of the third rank. Unfortunately this is not known, mainly because it is difficult to give a continuum Non-Abelian generalization for this case.

---

## CHAPTER 5

# Quark Confinement, Superfluidity, Elasticity. Criteria and Analogies

---

DOI: 10.1201/9780203755082-5

In the previous chapter we were mainly interested in the influence of instantons on the mass gap. Here we shall present more direct criteria for confinement by computation of the electric force between charges and the dielectric permeability of the vacuum. We shall also stress analogies with some phenomena in solid state physics.

Let us begin from the general expression for the static potential valid in both the Abelian and the Non-Abelian cases. The main idea is the following. We have seen in Chapter 3 that the eigenstates of a gauge Hamiltonian can be divided into sectors, containing static charges at points  $x_1, \dots, x_N$  with colour spins (for the  $SU(2)$  group as an example)  $I_1, \dots, I_N$ . We shall now show how to express the partition function:

$$Z(I_1 x_1, \dots, I_N x_N) = \sum_n e^{-\beta E_n(I_1 x_1, \dots, I_N x_N)} \quad (5.1)$$

in terms of a functional integral. Here  $E_n(I_1 x_1, \dots, I_N x_N)$  are the energy levels for the corresponding sector and  $\beta$  is the physical inverse temperature (and not an inverse coupling constant). We find it convenient to begin with nonzero temperature both for physical and technical reasons. The static potential (which depends on  $\beta$ ) is defined as the difference in free energy between the sector we are considering and the vacuum sector:

$$e^{-\beta V(I_1 x_1, \dots, I_N x_N, \beta)} = Z(I_1 x_1, \dots, I_N x_N) / Z_0 \quad (5.2)$$

Our first aim now is to express this quantity in terms of a functional integral over the gauge field. The derivation is based on the following formula of quantum mechanics:

$$\begin{aligned} Z(x, \tilde{x}, \beta) &\equiv \sum_n e^{-\beta E_n} \psi_n(x) \psi_n(\tilde{x}) \\ &= \int_{\substack{x(0)=x \\ x(\beta)=\tilde{x}}} \mathcal{D}x(t) \exp \left\{ - \int_0^\beta d\tau (\dot{x}^2 + v(x)) \right\} \end{aligned} \quad (5.3)$$

The generalization of this formula for the gauge system is:†

$$\begin{aligned}
 Z(\{\tilde{A}_n\}|\{A_n\}) &\equiv \sum_n e^{-\beta E_n} \Psi_n(\{A_n(x)\}) \Psi_n(\{\tilde{A}_n(x)\}) \\
 &= \int_{\substack{A_n(x,0)=A_n(x) \\ A_n(x,\beta)=\tilde{A}_n(x)}} \mathcal{D}A_n(x,\tau) \exp\left\{-\int_0^\beta d\tau \operatorname{Tr}(\dot{A}_n^2 + F_{nm}^2) dx^3\right\} \quad (5.4)
 \end{aligned}$$

Here the dot means time derivative, indices refer to space directions,  $F_{nm}$  is the magnetic part of the Yang–Mills field strength, and the sum over eigenstates includes them all. Let us first compute the partition function corresponding to the vacuum sector. In order to project it out of (5.4), we use the following comment. According to (3.37),

$$\Psi_n(A^\Omega) = \Psi_n(A) \quad \text{for the vacuum sector}$$

and

$$\begin{aligned}
 \Psi_n(A^\Omega) &= \prod_j \mathcal{D}^j(\Omega(x_j)) \Psi_n(A) \\
 (A_n^\Omega &= \Omega^{-1} A_n \Omega + \Omega^{-1} \partial_n \Omega) \quad (5.5)
 \end{aligned}$$

for the other sectors (where  $\mathcal{D}$  is the representation matrix for the  $SU(2)$  group). These functions  $\mathcal{D}$  have the property:

$$\int d\Omega \mathcal{D}^I(\Omega) = 0 \quad \text{for } I \neq 0 \quad (5.6)$$

(where the integral is taken over the  $SU(2)$  group measure). Using (5.5) and (5.6) we conclude that:

$$\begin{aligned}
 Z_0 &= \sum_{\substack{n \text{ vacuum} \\ \text{sector}}} e^{-\beta E_n} = \int \mathcal{D}\Omega(x) \int \mathcal{D}A_n(x) Z(\{A_n\}|\{A_n^\Omega\}) \\
 &= \int \mathcal{D}\Omega(x) \int_{A_n(x,\beta)=A_n^\Omega(x,0)} \mathcal{D}A_n(x,\tau) \\
 &\quad \times \exp\left\{-\operatorname{Tr} \int_0^\beta (\dot{A}_n^2 + F_{nm}^2) dx d\tau\right\} \\
 &\equiv \int \mathcal{D}\Omega(x) Z[\Omega(x)] \quad (5.7)
 \end{aligned}$$

---

† We have fixed the gauge by  $A_0(x) = 0$

We see that in order to project out gauge invariant states (vacuum sector) one has to integrate over gauge fields, rotated by the gauge transformation  $\Omega$ . The total (unprojected)  $Z$  would be given by the integral over strictly periodic  $A_n$ .

In order to project out the sector with a certain number of static charges we have to consider the integral:

$$Z(I_1 x_1, \dots, I_N x_N) = \int \mathcal{D}\Omega(x) \chi_{I_1}(\Omega(x_1)) \cdots \chi_{I_N}(\Omega(x_N)) Z[\Omega(x)] \quad (5.8)$$

where  $\chi_I(\Omega)$  is the character of the representation with spin  $I$ :

$$\chi_I(\Omega) = \sum_{m=-I}^{+I} \mathcal{D}_{mm}^I(\Omega) \quad (5.9)$$

This formula is easily checked by use of the transformation law (3.51) and standard orthonormality conditions for the representation matrix  $\mathcal{D}^I$ . For a general group  $G$  the formula will be the same, except that  $I$  will be replaced by the set of numbers characterizing the representation.

We have obtained the following result:

$$e^{-\beta V(I_1 x_1, \dots, I_N x_N; \beta)} = \langle \chi_{I_1}(\Omega(x_1)) \cdots \chi_{I_N}(\Omega(x_N)) \rangle \quad (5.10)$$

where the averaging must be performed with the “rotated” partition function  $Z[\Omega]$  as a weight. We see from this representation that a single static charge  $I$  would have an infinite energy if:

$$\langle \chi_I(\Omega(x)) \rangle = 0 \quad (5.11)$$

The linear potential between two charges would correspond to:

$$\langle \chi_I(\Omega(x_1)) \chi_I(\Omega(x_2)) \rangle \xrightarrow{x_{12} \rightarrow \infty} e^{-c|x_1 - x_2|} \quad (5.12)$$

Let us show now that there is a symmetry in the gauge systems which, if unbroken, leads to the condition (5.11) for half-integer  $I$ , and hence to charge confinement. This is the symmetry of the centre of the gauge group. In the case of  $SU(2)$  the centre consists of reflections  $\Omega \rightarrow -\Omega$ , while for  $SU(N)$  it is formed by transformations:

$$\Omega \rightarrow e^{2\pi i/N} \Omega \quad (5.13)$$

From the definition of  $Z[\Omega]$ :

$$\begin{aligned} Z[\Omega] &= \int \mathcal{D}A_n(x, \tau) \exp(-S(A_n)) \\ A_n(x, \beta) &= \Omega^{-1} A_n(x, 0) \Omega + \Omega^{-1} \partial_n \Omega \end{aligned} \quad (5.14)$$

we see that  $Z[\Omega]$  is indeed invariant under the transformations described. At the same time (for  $SU(2)$ ):

$$\begin{aligned}\chi_I(-\Omega) &= \chi_I(\Omega) & I \text{ integer} \\ \chi_I(-\Omega) &= -\chi_I(\Omega) & I \text{ half integer}\end{aligned}\tag{5.15}$$

We see, that in the case of unbroken symmetry (5.11) is true for half-integer  $I$ , only. We have already encountered this situation in Chapter 3, and explained that it reflects screening of integer  $I$ -spins by  $I = 1$  gluons.

Half-integer spins cannot be screened if  $\Omega \rightarrow -\Omega$  symmetry is unbroken. In the broken symmetry phase they are also screened thus giving  $\langle \chi_I \rangle \neq 0$ . It is a dynamical question to find out which phase is realized in our theory.

We shall see below how this question is solved for the cases analysed in the previous chapter. Before doing that, let us give another criterion for confinement, which is sometimes more convenient than the previous one. It is applicable only at zero temperature.

Let us consider a closed loop  $C$ , and associate with it a phase factor:

$$\Psi(C) = \text{P exp} \oint_C A_\mu dx^\mu\tag{5.16}$$

Here P exp means an ordered exponent defined as:

$$\text{P exp} \oint A_\mu dx^\mu = \lim_{\Delta x_j \rightarrow 0} \prod_j (1 + A_\mu(x_j) \Delta x_j^\mu)\tag{5.17}$$

(remember that  $A_\mu$  is a matrix, lying in the Lie algebra of our group and factors in (5.17) do not commute). Now the properties of gauge systems can be characterized by the behaviour of the following correlation function:

$$W_I(C) = \langle 0 | \chi_I(\Psi(C)) | 0 \rangle\tag{5.18}$$

Let us show that if

$$\begin{aligned}W_I(C) &\rightarrow \exp(-\text{const } \mathcal{A}_{\min}(C)) \\ (\mathcal{A}_{\min}(C) \text{ is the minimal area bounded by } C)\end{aligned}\tag{5.19}$$

for large enough loops, then the static potential between charges is linear. To prove this, let us consider a rectangle lying in the  $x^3, t$  plane,

and take the gauge  $A_0 = 0$ . We have then:

$$\begin{aligned}
 W_I(C_{RT}) &= \langle 0 | \chi_I \left( \text{P exp} \int_R^0 A_3(x^3, T) dx^3 \cdot \text{P exp} \int_0^R A_3(x^3, 0) dx^3 \right) | 0 \rangle \\
 &= \langle 0 | \mathcal{D}_{mm'}^I \left( \text{P exp} \int_R^0 A_3(x^3, T) dx^3 \right) \mathcal{D}_{m'm}^I \left( \text{P exp} \int_0^R A_3(x^3, 0) dx^3 \right) | 0 \rangle \\
 &= \langle 0 | \mathcal{D}_{mm'}^I \left( \text{P exp} \int_R^0 dx^3 A_3(x^3, 0) \right) e^{-\hat{\mathcal{H}}T} \\
 &\quad \times \mathcal{D}_{m'm}^I \left( \text{P exp} \int_0^R A_3(x^3, 0) dx^3 \right) | 0 \rangle = \sum_{n, m, m'} |(f_{m', m}^I)_n, 0|^2 e^{-E_n T}
 \end{aligned}$$

where

$$(f_{m', m}^I)_n, 0 = \langle n | \mathcal{D}_{m', m}^I \left( \text{P exp} \int_0^R A_3(x^3, 0) dx^3 \right) | 0 \rangle \quad (5.20)$$

$\hat{\mathcal{H}}$  is the Hamiltonian of the gauge system and a standard insertion of a complete set of states has been used. The crucial point of the derivation is that while the vacuum  $|0\rangle$  belongs to the gauge invariant sector the states  $|n\rangle$  belong to the sector with two static charges of colour spin  $I$ . This is because the P-exponent in (5.20) is not gauge invariant, but transforms as:

$$\text{P exp} \int_0^R A_3^a dx^3 = \Omega^{-1}(0) \left( \text{P exp} \int_0^R A_3 dx^3 \right) \Omega(R) \quad (5.21)$$

Therefore the states

$$|\Psi\rangle = \mathcal{D}_{mm'}^I \left( \text{P exp} \int_0^R A_3 dx^3 \right) | 0 \rangle$$

have a transformation law, which according to (3.51), corresponds to the two charge sector.

As we take the limit  $T \rightarrow \infty$  we find from (5.20):

$$W_I(C_{RT}) \xrightarrow{T \rightarrow \infty} e^{-E_0^{(I)}(R)T} \quad (5.22)$$

where  $E_0^{(I)}(R)$  is the minimal energy in the sector described. From this we see that the “area law” (5.19) corresponds to

$$E_0^{(I)}(R) \sim R \quad (5.23)$$

From the above, we expect the area law to be true in the confining phase for half-integer  $I$ , while for integers the decay of  $W_I(C)$  is much slower. Intuitively the criterion described can be understood as follows. Take a charge  $I$  around a closed loop  $C$  in space-time. The transition amplitude associated with this process is given by  $W_I(C)$ . On the other hand, this loop can be interpreted as the creation of a quark-anti-quark pair, propagation of this pair for a long time  $T$  and finally its annihilation. The  $T$ -dependence of the amplitude must be given by a factor  $\exp\{-iE(R)(-iT)\} = \exp(-E(R)T)$ , where  $E(R)$  is the interaction energy of the pair and  $(-iT)$  is the time of its existence.

Equating these two factors we again get (5.22).

Let us check now if we really have confinement in the models described in the previous chapter and compute the binding force between charges. We shall show that instanton contributions to the phase factor in the case of  $\mathcal{D} = 3$   $O(2)$  gauge theory lead to the area law.

The calculation is easily performed since

$$F(C) = \left\langle \exp\left(i \oint_{\partial S} A_\mu^3 dx^\mu\right) \right\rangle \approx \left\langle \exp\left(i \int_S H_\mu dS^\mu\right) \right\rangle \quad (5.24)$$

and we know from (4.76) that this can be written

$$F(C) = \left\langle \exp\left(i \int \eta(x)\rho(x) dx\right) \right\rangle \quad (5.25)$$

where

$$\eta(x) = \frac{1}{2} \int_S dS_y \cdot \frac{(x-y)}{|x-y|^3}$$

Since the field  $\eta(x)$  is strong enough we cannot neglect nonlinearities in (4.74). Rather  $F(C)$  is given by

$$F(C) = \exp\left\{-\left(\frac{e}{2\pi}\right)^2 \int [(\nabla(\chi_{cl} - \eta))^2 - M^2 \cos \chi_{cl}] d^3x\right\} \quad (5.26)$$

where  $\chi_{cl}$  is determined from the nonlinear Debye equation

$$2\nabla^2(\chi_{cl} - \eta) = M^2 \sin \chi_{cl} \quad (5.27)$$

Fluctuation corrections to this field are again exponentially small.

Let us assume that the contour  $C$  is planar and lies in the  $xy$  plane. Then, eq. (5.27) takes the form

$$\begin{aligned} \nabla^2 \chi_{cl} &= 2\pi \delta^1(z) \theta_S(xy) + \frac{1}{2} M^2 \sin \chi_{cl} \\ \theta_S(xy) &= \begin{cases} 1, & x, y \in S \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{5.28}$$

Far from the boundaries of the contour, eq. (5.28) is essentially one dimensional ( $\chi_{cl}$  depends only on  $Z$ ) and has the solution

$$\chi_{cl}(Z) = \begin{cases} 4 \arctan(e^{-MZ}) & Z > 0 \\ -4 \arctan(e^{MZ}) & Z < 0 \end{cases} \tag{5.29}$$

Substituting (5.29) into (5.26), we obtain

$$\begin{aligned} F(C) &= e^{-\gamma S} \\ E(R) &= \gamma R \end{aligned} \tag{5.30}$$

$$\gamma = \left(\frac{e}{2\pi}\right)^2 M \int_{-\infty}^{+\infty} dy ((\chi_{cl}'' - \eta'')(x-y) + M^2 \cos \chi_{cl}(y))$$

The result (5.30) implies that between two fixed charges there exists an electric string with energy density  $\gamma$ .

Now several comments are in order. First it is useful to understand the results we have obtained, using a special diagram technique. Namely, it is possible to represent

$$\begin{aligned} \langle H_\mu(x) H_\nu(y) \rangle &= \bullet \text{---} \bullet + \bullet \text{---} \circ \text{---} \bullet + \bullet \text{---} \circ \text{---} \text{---} \circ \text{---} \bullet \\ &+ \bullet \text{---} \circ \text{---} \text{---} \circ \text{---} \bullet + \dots \end{aligned} \tag{5.31}$$

Here we denote the free  $H$ -field operator by a solid line, pseudoparticles by open circles and the Coulomb interaction between pseudoparticles by dashed lines. It is possible to draw more complicated diagrams containing both solid and dashed lines, but all of them are small for small momenta and charges.

From (5.31) and (4.78) we see the crucial difference between pseudo-particle and instanton contributions to correlation functions. The second is purely transverse and the first is longitudinal, due to the fact that the quantity  $\partial_\mu H_\mu$  measures the density of topological charge. The existence of these two contributions makes possible the cancellation of singularities at zero momentum.

We have proved that the potential between two charges grows linearly. It is evident that there should exist an infinite resonance

spectrum in our theory, and it is tempting to find it directly by considering the correlation functions for operators of higher spin. Unfortunately in our approximation these correlation functions contain only scalar particle thresholds, and resonances should appear in higher order approximation, so the resonance problem remains to be solved even in our model.

An interesting property of the above formulas is that they lead to confinement of half-integer charges for which the Dirac quantization condition is true:

$$e \cdot g = 2\pi$$

(where  $g$  is the minimal charge of a magnetic pole). If we double the charge of the test particles, then it is easy to check that the contribution of monopoles to  $\exp\{i \int H_\mu dS^\mu\}$  will be negligible and we have no confinement of such charges. In the framework of the Georgi-Glashow model, described in Chapter 4, this result is quite natural since integer charges can be screened by  $W^\pm$  bosons, while for half-integers this is impossible. However, in the case of the lattice  $O(2)$  model the above conclusion is rather surprising. Thus, we have proved the anticipated result that  $O(2)$  gauge theories for  $\mathcal{D} = 3$  produce linear confinement of half-integer charges. For  $\mathcal{D} = 4$  it is not so for  $e_0^2 \ll 1$  since small instanton loops give very small contributions to the phase factors. This is easy to verify by calculations analogous to those above.

Let us now describe one more manifestation of confinement, which is of some interest because of analogies with solid state physics. Namely let us show that in the confining phase the dielectric constant is zero. In order to introduce this quantity in abelian systems let us consider an antisymmetric external field  $f_{\mu\nu}(x)$  coupled to our system in the following way:

$$\exp(-W[f_{\mu\nu}]) = \int \mathcal{D}A_\mu e^{-S[F_{\mu\nu} + f_{\mu\nu}]} \quad (5.32)$$

(where  $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ ; we take the liberty of using continuum notation while meaning a lattice field theory). From this definition it is clear that  $W[f]$  is invariant under gauge transformation of the "third kind":

$$f_{\mu\nu} \rightarrow f_{\mu\nu} + \partial_\mu \lambda_\nu - \partial_\nu \lambda_\mu \quad (5.33)$$

because it can be compensated in (5.32) by a change  $A_\mu \rightarrow A_\mu + \lambda_\mu$ . Therefore,  $W$  should depend on the following combinations of  $f_{\mu\nu}$ :

$$\phi_{\mu\nu\lambda} = \partial_\mu f_{\nu\lambda} + \partial_\nu f_{\lambda\mu} + \partial_\lambda f_{\mu\nu} \quad (5.34)$$

At first glance this implies that a constant external  $f_{\mu\nu}$  has no influence on the system. This is indeed so if we are in a phase with finite correlation length. However, when massless photons are present we have for constant  $f_{\mu\nu}$  (at first neglecting instanton effects):

$$e^{-W[f]} = \int \mathcal{D}A_\mu \exp\left(-\frac{1}{e_0^2} \int F_{\mu\nu}^2 dx\right) \exp\left(-\frac{1}{e_0^2} \int f_{\mu\nu}^2 dx\right) \quad (5.35)$$

$$W[f] - W[0] = \frac{1}{e_0^2} \int f_{\mu\nu}^2 dx$$

(because  $\int F_{\mu\nu} f_{\mu\nu} dx = f_{\mu\nu} \int F_{\mu\nu} dx = 0$ ). Accounting for small (in  $\mathcal{D} = 4$ ) instanton effects gives:

$$W[f] - W[0] = \frac{\kappa(e_0^2)}{e_0^2} \int f_{\mu\nu}^2 dx \quad (5.36)$$

where  $\kappa$  is some constant which has a natural interpretation as the dielectric constant of the vacuum. The apparent paradox between (5.34) and (5.36) is easily resolved if we notice that due to the gapless photon, terms like  $\int \phi_{\mu\nu\lambda} \partial^{-2} \phi_{\mu\nu\lambda} dx dx'$  (where  $\partial^{-2}$  is the inverse Laplacian) are present in the effective action. Such terms do not vanish for constant  $f_{\mu\nu}$  and produce the result (5.36). We can also say that while constant  $f_{\mu\nu}$  can be removed by the transformation:

$$A_\mu \rightarrow A_\mu + f_{\mu\nu} X^\nu \quad (5.37)$$

This transformation changes the boundary conditions which were  $A_\mu|_{\text{inf}} = 0$ . If the system has long range correlations, this change of boundary condition will change the partition function which thus becomes dependent on  $f_{\mu\nu}$ . As we know, for  $e_0^2 > e_{0,\text{crit}}^2$  in  $\mathcal{D} = 4$  we have a mass gap and a confining phase. From the above argument it is clear that at this point the dielectric constant  $\kappa$  becomes zero, and remains zero in all the confining region:

$$\kappa(e_0^2) = 0 \quad \text{for } e_0^2 > e_{0,\text{crit}}^2 \quad (5.38)$$

In the case  $\mathcal{D} = 3$ ,  $\kappa$  is zero for all values of the coupling. We conclude that the reaction to the homogeneous external antisymmetric field, described by the dielectric constant  $\kappa$  can serve as a confinement criterion; we have

$$\kappa = \begin{cases} \text{const} & \text{in the Coulomb phase} \\ 0 & \text{in the confining phase} \end{cases} \quad (5.39)$$

There exists an interesting analogy between  $\kappa$  in the gauge systems and certain quantities in global systems. Let us consider the response of the global  $O(2)$  system to the external vector field  $v_\mu$ :

$$e^{-W[v_\mu]} = \int \mathcal{D}\varphi e^{-S(\partial_\mu\varphi + v_\mu)} \quad (5.40)$$

Again  $W[v_\mu]$  is gauge invariant:

$$W[v_\mu + \partial_\mu\lambda] = W[v_\mu] \quad (5.41)$$

If we neglect vortices, for constant  $v_\mu$  we obtain:

$$W = \frac{\beta}{2} \rho_S(\beta) \int v^2 dx \quad (5.42)$$

( $\rho_S(\beta)$  being some function).

The constant  $v_\mu$  can be removed from the action by the transformation:

$$\varphi \rightarrow \varphi + v_\mu x^\mu \quad (5.43)$$

but then it appears in the boundary conditions. Formula (5.42) does not contradict gauge invariance since due to massless excitations in the system  $v_\mu(x)$  appears in the combination:

$$W \sim \int v_\mu \left( \delta^{\mu\nu} - \frac{\hat{c}_\mu \hat{c}_\nu}{\hat{c}^2} \right) v_\nu dx \quad (5.44)$$

After the phase transition, when a correlation length appears in the system, we must have:

$$\rho_S(\beta) = 0 \quad \text{for } \beta < \beta_{cr} \quad (5.45)$$

The analogy between  $\rho_S$  and  $\kappa$  is obvious. What is interesting is that if our  $O(2)$  system describes  ${}^4\text{He}$  then  $\rho_S(\beta)$  has an interpretation as the superfluid density. Therefore, the normal phase of  ${}^4\text{He}$  correspond to the confining phase of the gauge system, and the superfluid phase to the Coulomb phase. Let us show now that  $\rho_S$  is indeed the superfluid density. To do this we consider a second-quantized Hamiltonian, describing a Bose liquid:

$$\hat{\mathcal{H}} = \int \left( \frac{1}{2m} |\partial_a \Psi|^2 + \int \Psi^\dagger(x) \Psi(x) u(x-y) \Psi^\dagger(y) \Psi(y) dy \right) dx \quad (5.46)$$

where  $u$  is a shortrange interatomic interaction. Suppose that the walls, containing our liquid, move with velocity  $v$ . The question is whether the liquid will move with the same velocity (owing to friction with the

walls) or whether it will stay at rest (as in the case of a superfluid). Let us go to the coordinate system where the walls are at rest. This is achieved by a Galilean transformation

$$\Psi(x) \rightarrow e^{-imvx}\Psi(x) \quad (5.47)$$

after which

$$\begin{aligned} \langle \hat{\mathcal{H}} \rangle &\rightarrow \left\langle \int \frac{1}{2m} |(\partial_x - imv_x)\Psi|^2 dx + \dots \right\rangle \\ &= \langle \mathcal{H}_0 \rangle - v \langle \mathbf{P}_0 \rangle + Nmv^2/2 \end{aligned} \quad (5.48)$$

Here

$$\mathbf{P}_0 = \int \Psi^* \frac{1}{i} \frac{\partial}{\partial x} \Psi dx$$

is the total momentum. We see that if before the transformation the liquid was at rest ( $\mathbf{P}_0 = 0$ ,  $\mathcal{H}_0 = E_0$ ) we have to expect that the change (5.47) will add to the Hamiltonian a  $v^2$ -term. If the liquid moves then  $\mathcal{H}_0 = E_0 + Nmv^2/2$ ;  $\mathbf{P}_0 = +Nmv$  and the substitution (5.47) should add nothing to  $\langle \mathcal{H}_0 \rangle$ . If we recall now that the only infrared-important quantity in  ${}^4\text{He}$  is  $\varphi(x)$ , the phase of  $\Psi(x)$ , we conclude that the existence of superfluidity manifests itself in the response of our system to the transformation:

$$\varphi(x) \rightarrow \varphi - mvx \quad (5.49)$$

Normal liquid is indifferent to such a transformation, while the superfluid does react. We have also seen that such a response is possible exclusively because of the pole terms in (5.44). The dielectric constant in a gauge system was shown to be a precise analogue of the superfluid density  $\rho_S$ . This analogy can be extended further in order to include crystals which, as we saw, are described by the same kind of actions depending on  $u_\alpha(x)$ . In this case we have to examine the response of the system to a symmetric traceless field defined by:

$$u_\alpha(x) \rightarrow u_\alpha(x) + h_{\alpha\beta} x^\beta \quad (5.50)$$

( $h_{\alpha\alpha} = 0$ ). Again, nonzero response to  $h_{\alpha\beta}$  (which is a kind of external gravitational field, just as  $v_\mu$  could have been considered as an external electromagnetic field), is possible in the longranged phase, and is described by the shear modulus

$$W = \frac{\mu}{2} h_{\alpha\beta}^2 \quad (5.51)$$

After melting we have  $\mu = 0$ . All these nontrivial constants,  $\kappa$ ,  $\rho_S$ ,  $\mu$ , etc., can be called, following V. L. Berezinsky, transverse rigidities. They can be formally defined as residues at the massless poles of corresponding polarization operators and therefore exist only in the broken symmetry phase. A subtle point is that, as we mentioned in Chapter 1, in the case of gauge systems, and in  $\mathcal{D} = 2$  global systems, the order parameters are zero and there are no poles in the Green's functions. However, there are poles in the corresponding polarization operators and it is these poles which determine the physics of the systems.

In Non-Abelian cases with global symmetries it is quite straightforward to introduce an analogue of the superfluid density. If we take, for instance, the case of the  $n$ -field we can examine its response to an external triplet vector field:

$$e^{-W[\mathbf{v}_\mu]} = \int \mathcal{D}\mathbf{n}(x) \exp\left\{-\frac{\beta}{2} \int (\partial_\mu \mathbf{n} + \mathbf{v}_\mu \times \mathbf{n})^2 dx\right\} \quad (5.52)$$

Again we conclude that  $W[\mathbf{v}]$  has to possess Non-Abelian gauge invariance:

$$W[\mathbf{v}_\mu] = \tilde{W}[f_{\mu\nu}] \quad (5.53)$$

$$f_{\mu\nu} = \partial_\mu v_\nu - \partial_\nu v_\mu + v_\mu \times v_\nu$$

The “superfluid density” can be defined as the residue:

$$\tilde{W} \approx \frac{\beta}{2} \rho_S \int f_{\mu\nu} \frac{1}{\partial^2} f_{\mu\nu} dx \quad (5.54)$$

The Non-Abelian gauge theories present some difficulty in this respect. We have to introduce a gauge field of the “third kind”,  $f_{\mu\nu}$ , coupled to the Yang–Mills field so as to have Non-Abelian third kind gauge invariance. I do not know how to do this in a continuum theory, though on a lattice there are some possibilities. We still can introduce a purely static dielectric constant by means of integrating over  $A_\mu$  fields with boundary conditions:

$$A_\mu^a \xrightarrow{x \rightarrow \infty} f_{\mu\nu}^a x^\nu \quad (5.55)$$

and defining  $\kappa$  as the coefficient before  $(f_{\mu\nu}^a)^2$  in the effective action. Vanishing of this  $\kappa$  is a signal for quark confinement.

Ending this section, let us stress that the dielectric constants introduced above do not have a naive direct relation to static potentials. This is because they determine the response to *infinitesimal* fields, and static charges, being quantized, are necessarily finite.

---

## CHAPTER 6

# Topology of Gauge Fields and Related Problems

---

---

DOI: 10.1201/9780203755082-6

We have seen in the previous chapters, that in Abelian systems the problem of charge confinement is solved by instantons.

In Non-Abelian theories instanton solutions are also present. However, due to the large perturbative fluctuations, discussed in Chapter 2, it is difficult to judge whether they play a decisive role in forming a mass gap and a confining regime. In such theories we have a kind of instanton liquid which is difficult to treat. It is possible that due to some hidden symmetries, present in these systems, instantons may form a useful set of variables for an exact description of the system, but this has not yet been shown.

At the same time, due to the fact that instantons carry nontrivial topology (they describe configurations of the fields which can not be “disentangled”), some manifestations of instantons cannot be mixed up with perturbative fluctuations.

In this chapter we shall analyse topological properties of (mainly) nonabelian instantons and solitons and discuss some associated peculiar effects.

### 6.1 Instantons for $\mathcal{D} = 2$ , $N = 3$ $n$ -Fields

Let us find minima of the classical action for the  $n$ -field in the case  $N = 3$ . In order that this action be finite, we have to consider a boundary condition:

$$n(x) \xrightarrow{x \rightarrow \infty} n_0 \tag{6.1}$$

Therefore, since infinity can be viewed as one point, our  $x$ -space is topologically a sphere. Each configuration  $n(x)$  defines a map of such a sphere in  $x$ -space onto the sphere  $n^2 = 1$ , which in the case  $N = 3$  gives

$S^2 \rightarrow S^2$ . It is known that such maps can be classified by integers  $q$  which define the number of times the second sphere is covered by the first one. The simplest example of the  $q$ -map is described by the formulas:

$$\tilde{\vartheta} = \vartheta, \tilde{\varphi} = q\varphi \pmod{2\pi} \quad (6.2)$$

Here  $(\vartheta, \varphi)$  and  $(\tilde{\vartheta}, \tilde{\varphi})$  are polar and azimuthal angles for the first and the second sphere. In this formula  $q$  must be integer since otherwise the map would be discontinuous. For the general case the number of coverings (or topological charge) is defined by:

$$q = \frac{1}{4\pi} \int_0^{2\pi} d\varphi \int_0^\pi d\vartheta \sin \vartheta \frac{\partial(\tilde{\vartheta}, \tilde{\varphi})}{\partial(\vartheta, \varphi)} \quad (6.3)$$

$(\partial(\tilde{\vartheta}, \tilde{\varphi})/\partial(\vartheta, \varphi)) = (\partial\tilde{\vartheta}/\partial\vartheta)(\partial\tilde{\varphi}/\partial\varphi) - (\partial\tilde{\vartheta}/\partial\varphi)(\partial\tilde{\varphi}/\partial\vartheta)$  is the jacobian for the mapping under consideration). It is easy to check that (6.3) can be rewritten in a more invariant form. For the mapping  $\mathbf{n} = \mathbf{n}(x)$ ,  $\mathbf{n}^2 = 1$ :

$$q = \frac{1}{8\pi} \int d^2x \mathbf{n} \cdot [\partial_\mu \mathbf{n} \partial_\nu \mathbf{n}] \epsilon^{\mu\nu} \quad (6.4)$$

Here  $\epsilon^{\mu\nu}$  is the standard antisymmetric tensor. This formula is checked by the direct substitution of  $\mathbf{n} = (\cos \tilde{\vartheta}, \sin \tilde{\vartheta} \cos \tilde{\varphi}, \sin \tilde{\vartheta} \sin \tilde{\varphi})$  into (6.4) after which we obtain (6.3). We wish now to minimize the classical action in the sector with given  $q$ . This problem is simplified by the following trick. Let us consider the identity:

$$\begin{aligned} & \frac{1}{4} \int (\partial_\mu \mathbf{n} + \epsilon_{\mu\nu} [\mathbf{n} \times \partial_\nu \mathbf{n}])^2 d^2x \\ &= \frac{1}{2} \int (\partial_\mu \mathbf{n})^2 d^2x - \frac{1}{2} \int \epsilon_{\mu\nu} \mathbf{n} \cdot [\partial_\mu \mathbf{n} \times \partial_\nu \mathbf{n}] d^2x \end{aligned} \quad (6.5)$$

From (6.5) we conclude:

$$S = \frac{1}{2e_0^2} \int (\partial_\mu \mathbf{n})^2 d^2x = \frac{4\pi q}{e_0^2} + \frac{1}{4e_0^2} \int (\partial_\mu \mathbf{n} + \epsilon_{\mu\nu} [\mathbf{n} \times \partial_\nu \mathbf{n}])^2 d^2x \quad (6.6)$$

It follows from (6.6) that in order to find an absolute minimum for the  $\mathbf{n}$ -fields with the topological charge  $q$  one can avoid the problem of solving classical equations of motion, which are second order differential equations. Instead, first order equations can be considered which are in some sense the “square root” of the classical equations. We have:

$$\partial_\mu \mathbf{n} = -\epsilon_{\mu\nu} [\mathbf{n} \times \partial_\nu \mathbf{n}] \quad (6.7)$$

the so-called duality equation. If (6.7) is satisfied then, owing to (6.6), the action has the absolute minimum value equal to  $4\pi q/e_0^2$ .

Solution of (6.7) is easy. We introduce a complex field  $w$  by a stereographic projection:

$$\begin{aligned} n_1 + in_2 &= 2w/(1 + |w|^2) \\ n_3 &= (1 - |w|^2)/(1 + |w|^2) \end{aligned} \quad \mathbf{n}^2 = 1 \quad (6.8)$$

The substitution of (6.8) into (6.7) reduces this equation to:

$$\partial_z w \equiv (\partial_1 + i\partial_2)w = 0 \quad (6.9)$$

Therefore (6.7) are just Cauchy-Riemann equations for the function  $w$ . This function must be not only analytic, but also meromorphic, since otherwise  $\mathbf{n}$  would have branch cuts. Hence the most general solution has the form:

$$w(z) = \prod_{j=1}^q \frac{z - a_j}{z - b_j} \quad (6.10)$$

where we have normalized  $w$  by  $w(\infty) = 1$ , corresponding to  $\mathbf{n}(\infty)$  pointing in the  $x$ -direction. The integer  $q$  in (6.10) is just the topological number. This can be seen without explicit computation since it is clear from (6.10) that the inverse function  $z = z(w)$  is  $q$ -valued. That implies that the  $w$ -sphere is covered  $q$  times by the  $z$ -sphere. Of course, it is also not difficult to substitute (6.10) into (6.4) and to compute  $q$  explicitly.

The remarkable thing about the instantons (6.10) is that they do not interact classically. There are, however, anti-instantons

$$w = \prod_{j=1}^q \left( \frac{\bar{z} - a_j}{\bar{z} - b_j} \right) \quad (6.11)$$

having negative topological charge. The mixed configuration of instantons and anti-instantons is not a strict classical solution (as it was for kinks and anti-kinks in Chapter 4) and a dipole-dipole like interaction is present. Notice also, that the instantons (6.10) have a natural structure of dipoles with poles placed at  $a_j$  and  $b_j$ .

As we turn from this nice, clean classical mathematics to functional integrals we encounter a difficulty. Namely, while in the zeroth order approximation the one instanton contribution is proportional to  $e^{-S_{cl}} = e^{-4\pi/e_0^2}$ , it is quite obvious that quantum fluctuations should renormalize the bare coupling  $e_0^2$ . The resulting contribution can be found without explicit computation by the following argument. As we see from (6.10), the effective size of the instanton with parameters  $a$  and

$b$  is  $|a - b|$ . So, it is natural to expect that  $e_0^2$  in the above expression will be replaced by:

$$e^2(|a - b|) \simeq 2\pi/\log(\lambda|a - b|)^{-1} \quad (6.12)$$

obtained from (2.49) for  $N = 3$ . The instanton contribution  $Z^{(1)}$  has to contain an integral over the parameters  $a$  and  $b$ . The measure of integration must have both translational and scale invariance (since  $a$  and  $b$  in the classical solution are defined modulo their common scale). On these grounds we expect:

$$\begin{aligned} Z^{(1)} &\sim \int \frac{d^2a d^2b}{|a - b|^4} \exp(-4\pi/e^2(|a - b|)) \\ &= \lambda^2 \int \frac{d^2a d^2b}{|a - b|^2} \simeq \lambda^2 V \int \frac{d\rho}{\rho} \end{aligned} \quad (6.13)$$

( $V$  is the volume of the system and  $\rho = |a - b|$ ). We see that this integral diverges at large values of  $|a - b|$  but we learned in Chapter 2 that for  $|a - b| \gtrsim \lambda^{-1}$  the formula (6.12) is not applicable since  $e^2(|a - b|)$  is not small any more. For the same reason we cannot trust (6.13) in this region, since the WKB approximation used in its derivation is valid only for small couplings.

Nevertheless it makes sense to ask what happens after accounting for the multi-instanton solution. The tactic for this is to take a classical field (6.10) and to consider small fluctuations on it as background. After performing the Gaussian integral, we obtain a determinant of the corresponding quadratic form which will depend on  $\{a_j\}$  and  $\{b_j\}$ . The logarithm of this determinant can be considered as an interaction energy between instantons, induced by quantum fluctuations. Finally we shall have to integrate over the parameters  $\{a_j\}$  and  $\{b_j\}$ . This programme can be carried through quite explicitly, the main reason being that for any self-dual background the kernel of the above-mentioned quadratic form simplifies considerably. To show this let us recall that the quadratic part of the action is described by (2.47). In the case of  $O(3)$  we can introduce the complex notation:

$$\begin{aligned} \varphi &= \varphi_1 + i\varphi_2 \quad A_\mu = A_\mu^{12} \\ S^{\text{II}} &= \frac{1}{2e_0^2} \int \{ |(\partial_\mu + iA_\mu)\varphi|^2 - \frac{1}{2}(\varphi^+ \varphi) B_\mu^a B_\mu^a \} dx^2 \\ &= \frac{1}{2e_0^2} \int \{ (\partial_\mu + iA_\mu)\varphi|^2 + \frac{1}{2}\epsilon_{\mu\nu} F^{\mu\nu}(\varphi^+ \varphi) \} dx^2 \quad (F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu) \end{aligned} \quad (6.14)$$

where we have used self-duality (6.7):

$$B_\mu^a = \epsilon_{ab} \epsilon_{\mu\nu} B_\nu^b$$

and the consequence of (2.45):

$$F_{\mu\nu} = i(B_\mu^* B_\nu - B_\nu B_\mu^*) \quad (6.15)$$

The kernel of this quadratic form coincides with the square of the Dirac operator:

$$\begin{aligned} (\gamma_\mu(\partial_\mu + iA_\mu))^2 &= (\partial_\mu + iA_\mu)^2 + \frac{1}{2}\gamma_5 \epsilon_{\mu\nu} F_{\mu\nu} \\ \gamma_5 &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \end{aligned} \quad (6.16)$$

It is well known (and will be discussed in Chapter 8) that the last determinant is easily computed. Namely:

$$\log \det (\gamma^\mu(\partial_\mu + iA_\mu)) = \frac{1}{4\pi} \int (F_{\mu\nu} \frac{1}{\partial^2} F_{\mu\nu}) d^2x \quad (6.17)$$

In order to calculate the multi-instanton contribution much work is still needed. One has to take account properly of collective coordinates and to compute the integral (6.17). I do not know any simple way of doing this (which, I am sure, exists). So, referring the reader for further details to the original papers, let me give the final result:

$$\begin{aligned} Z^{(q)} &= \frac{\lambda^{2q}}{(q!)^2} \int d^2a_1 \dots d^2a_q d^2b_1 \dots d^2b_q \\ &\quad \times \prod_{i<j} |a_i - a_j|^2 \prod_{i<j} |b_i - b_j|^2 \prod_{i,j} |a_i - b_j|^{-2} \end{aligned} \quad (6.18)$$

Or, after summing over  $q$ :

$$\begin{aligned} Z_{\text{INST}} &= \sum_{q=0}^{\infty} \frac{\lambda^{2q}}{(q!)^2} \int \prod_j d^2a_j d^2b_j \\ &\quad \times \exp \left\{ \sum_{i<j} (\log |a_i - a_j|^2 + \log |b_i - b_j|^2) \right. \\ &\quad \left. - \sum_{i,j} \log |a_i - b_j|^2 \right\} \end{aligned} \quad (6.19)$$

This result is quite surprising. We see that each instanton behaves as if it is composed of a pair of opposite Coulomb charges, placed at  $a_j$  and  $b_j$ . Since the two dimensional Coulomb energy is given by  $(1/4\pi) \log |a - b|^2$ , the expression (6.19) is the partition function for the

plasma with inverse temperature  $\beta = 4\pi$ . As was discussed in Chapter 4, such a plasma has two different phases. For large  $\beta$  the charges form dipoles and the system is neutral, with long range correlations (no mass gap). At some critical  $\beta$  (which is known to be equal to  $8\pi$ ), dissociation of the dipoles occurs and for  $\beta < \beta_{cr} = 8\pi$  we have a plasma phase with Debye screening, and therefore a mass gap. We conclude that owing to quantum effects instantons “melt” and create a finite mass gap in the theory. Before discussing the validity of the approximations made, let us give a very useful representation for (6.19). It is based on the so-called bosonization formulas which we discuss later. If one considers a free massless Dirac field  $\psi = \begin{pmatrix} \psi_L \\ \psi_R \end{pmatrix}$  for  $\mathcal{D} = 2$  and introduces two operators:  $\sigma_+(x) = \psi_L^\dagger \psi_R$  and  $\sigma_- = \psi_R^\dagger \psi_L$  it can be shown that

$$\begin{aligned} & \langle \sigma_+(a_1) \dots \sigma_+(a_N) \sigma_-(b_1) \dots \sigma_-(b_M) \rangle \\ &= \prod_{i < j} |a_i - a_j|^2 |b_i - b_j|^2 \prod_{i, j} |a_i - b_j|^{-2} \delta_{N, M} \end{aligned} \quad (6.20)$$

From this formula it follows, that

$$\begin{aligned} Z_{\text{INST}} &= \int \mathcal{D}\psi(x) \mathcal{D}\bar{\psi}(x) \\ &\times \exp \left\{ - \int (\bar{\psi} i \gamma^\mu \partial_\mu \psi + \lambda \bar{\psi} \psi) \right\} \end{aligned} \quad (6.21)$$

(where  $\bar{\psi}\psi = \sigma_+(x) + \sigma_-(x)$  is a mass term).

We see that in this representation, expansion in instantons becomes a mass expansion. It is also obvious from (6.21) that

$$\log Z_{\text{INST}} = V \int \frac{d^2 p}{(2\pi)^2} \text{Tr} \log (\gamma^\mu p_\mu + \lambda) \quad (6.22)$$

Expansion in  $\lambda$  leads to more and more infrared singular terms, containing  $\int d^2 p/p^n$  but the sum (6.22) is well behaved.

We may draw important conclusions from the above computations, but to what extent are the results reliable? There are two sources of errors. First of all we have to include anti-instantons and take account of their (classical) interaction with instantons. This interaction is of dipole-dipole type. One can imagine that we shall have to introduce two sorts of massive fermions,  $\psi_1$ , and  $\psi_2$ , describing instantons and anti-instantons and to consider some kind of interaction for  $\psi_1$  and  $\psi_2$ . This approach† results in a completely integrable model containing the fermions just described. The problem with this approach is that the

† Due to Buchvostov and Lipatov (1980).

instanton–anti-instanton configuration can be unambiguously defined only for well-separated objects, because otherwise it is difficult to distinguish it from other fluctuations with zero topological charge. This ambiguity is closely connected with another source of errors, coming from the fact that the quantum fluctuations above were treated in one loop approximation. For each separate  $q$ -instanton contribution this would be disastrous because these terms are infrared divergent and the effective charge  $e^2$  becomes unboundedly large. [However, we see that after summation over  $q$  these divergences are cut off at the Debye length  $\lambda^{-1}$ .] At this length  $e^2 \sim 1$  and so are higher quantum corrections. That gives hope that at least qualitatively the system is properly described by the above approach. Moreover, it often happens in integrable systems that the one loop approximation turns out to be exact. So an optimistic view of the situation is the following. One has to introduce instantons, described by a massive Dirac field  $\psi_1$ , and anti-instantons described by  $\psi_2$ . Next one must find a certain extrapolation of the instanton–anti-instanton interaction to small distance. Then it might be hoped that there exists such an extrapolation that the resulting system describes the  $n$ -field exactly and not only in one loop approximation. Whether this is true can in principle be checked by the use of exact solutions, but up to now this has not yet been done.

So at present we do not know whether exact properties of the  $n$ -field can be formulated in terms of instantons, but this possibility seems to be open.

The same, and even more difficult, problems exist in nonabelian gauge theories which we describe in the next section. Before coming to that let us mention briefly what kind of instanton structure is present in other versions of chiral models. First of all,  $n$ -fields with the group  $O(N)$ ,  $N \geq 4$  do not have any nontrivial topology; that is to say any map  $S^2 \rightarrow S^{N-1}$  for  $N \geq 4$  is contractible. The reason for this is easy to understand if we consider the case of mapping  $S^1 \rightarrow S^2$ , that is a circle mapped on to an ordinary sphere, say to its equator. It is obvious, that by moving this circle to the north pole we can contract it to a point. For similar reasons any map  $S^2 \rightarrow S^{N-1}$  can be deformed to the trivial one. A slightly more complicated argument shows that the map of  $S^2$  onto any Non-Abelian Lie group  $G$ , described by the principal chiral field  $g(x)$  is also contractible. These theories do not have stable instantons. The chiral theories which do have them are described by the coset spaces  $G/H$  in which  $H$  contains  $U(1)$  as a factor. Let us explain how this comes about. In the above theories the fields can be represented by

$$\varphi_a(x) = g_{ab}(x)\varphi_b^{(0)}, \quad x \in S^2 \quad (6.23)$$

where  $g_{ab} \in G$  and  $\varphi_b^{(0)}$  is a constant field, invariant under  $H$ :

$$h_{ab}\varphi_b^{(0)} = \varphi_a^{(0)} \quad \text{for } h \in H \quad (6.24)$$

The matrix  $g_{ab}(x)$  in (6.23) need not be continuous. Let us consider a set of matrices  $g^{(N)}(x)$  defined in the northern hemisphere and  $g^{(S)}(x)$  defined in the southern one. Suppose on the equator we have the relation:

$$g^{(N)}(x) = g^{(S)}(x) \cdot h(x) \quad x \in \text{equator} = S^1 \quad h \in H \quad (6.25)$$

We see from (6.23) and (6.24) that in spite of the discontinuity in  $g(x)$  the field  $\varphi_a(x)$  is continuous and defines a map  $S^2 \rightarrow G/H$ . From (6.25) we deduce that these maps can be classified according to the maps of the equator to  $H: S^1 \rightarrow H$ . If  $H = U(1)$  this is just a  $S^1 \rightarrow S^1$  map, classified by the winding number. If  $H = U(1) \times \text{something}$ , then we can map  $S^1$  onto the first factor in  $H$ . In mathematical notation the statement we have proved is written as:

$$\pi_2(G/H) \simeq \pi_1(H) \quad \text{if } \pi_2(G) = 0 \quad (6.26)$$

(where  $\pi_k(M)$  is the  $k$ th homotopy group, elements of which are classes of nontrivial maps of  $S^k \rightarrow M$ ).

The most familiar example of a chiral theory with instanton structure is the so-called  $CP^{N-1}$ -model where the field belongs to the complex projective space:

$$\phi \in CP^{N-1} = \frac{SU(N)}{SU(N-1) \otimes U(1)} \quad (6.27)$$

(The case  $N = 2$  is the  $O(3)$   $n$ -field.) There are some interesting dynamics in this model. We shall discuss it in Chapter 8.

## 6.2 Instantons in Non-Abelian Gauge Theories

Non-Abelian gauge theories with any symmetry group  $G$  possess topologically nontrivial fields. This can be seen from the following consideration. In order that the Yang-Mills action be finite one must require that:

$$F_{\mu\nu}(x) \xrightarrow{x \rightarrow \infty} o(1/x^2) \quad (6.28)$$

From this fact we deduce:

$$A_\mu(x) \xrightarrow{x \rightarrow \infty} g^{-1}(x) \hat{\partial}_\mu g(x) + o(1/x) \quad (6.29)$$

where  $g(x) \in G$ .

If we bound our  $\mathcal{D} = 4$  Euclidean space by a large three-dimensional sphere  $S^3$  we obtain, according to (6.29), a map  $g(x): S^3 \rightarrow G$ . It is easy to see that all such maps are classified by the integers for any  $G$ . Let us prove this for  $G = SU(2)$ . Any matrix  $g$  in this case can be written as:

$$g = n_4 + i\mathbf{n} \cdot \boldsymbol{\tau} \quad (6.30)$$

$$n_4^2 + \mathbf{n}^2 = 1; \quad g^+ g = I \quad (6.31)$$

(here  $\boldsymbol{\tau}$  are the Pauli matrices). Therefore, elements of the  $SU(2)$  group are in one to one correspondence with points of the sphere  $S^3$  defined by the equation (6.31). The map of the  $S^3$  which bounds the  $x$ -space onto  $SU(2)$  is therefore just the map  $S^3 \rightarrow S^3$ . In this latter case all the arguments we had for  $S^2 \rightarrow S^2$  in Section 6.1 are applicable. We have an integer  $q$  which is equal to the number of coverings given by the integral of the Jacobian. The analogue of the formula (6.4) in the present case has the form:

$$\begin{aligned} q &= \frac{1}{12\pi^2} \int d^3x \epsilon^{abcd} \epsilon_{\mu\nu\lambda} (n^a \hat{\partial}_\mu n^b \hat{\partial}_\nu n^c \hat{\partial}_\lambda n^d) \\ &= \frac{1}{24\pi^2} \int d^3x \epsilon_{\mu\nu\lambda} \text{Tr}(L_\mu L_\nu L_\lambda) \end{aligned} \quad (6.32)$$

with

$$L_\mu(x) = g^{-1} \hat{\partial}_\mu g(x)$$

It is easily checked that the combination of  $n$  in the first equality is just the surface element of  $S^3$  and hence the first term (6.32) is the Jacobian for the transformation from the  $x$ -space (forming  $S^3$ ) to the  $n$ -space. The second equality can be checked by explicit computation or by realizing that the integrand in this formula is the only possible expression, having dimension 3 and invariant under  $G \otimes G$ :

$$g(x) \rightarrow ug(x)v \quad (6.33)$$

Therefore it must be proportional to an element of the group volume.

We have obtained the following classification of  $A_\mu$ . Take the asymptotic form of a given  $A_\mu$ -field at  $x \rightarrow \infty$ , and determine  $g(x)$  from (6.29). After that, compute  $q$  from (6.32). Gauge fields with different  $q$  cannot be continuously deformed to one another.

There exists a more convenient expression for  $q$  than (6.32). It is given by the formula:

$$\begin{aligned} q &= \frac{1}{32\pi^2} \int \epsilon^{\mu\nu\lambda\rho} \text{Tr}(F_{\mu\nu} F_{\lambda\rho}) d^4x \\ &= \frac{1}{8\pi^2} \int_x \text{Tr} F \wedge F \end{aligned} \quad (6.34)$$

Here in the second equality we have used a convenient notation adopted in mathematics. For each skew-symmetric  $p$ -rank tensor  $T^{\mu_1 \dots \mu_p}$  one can define the  $p$ -form:

$$T_p = \frac{1}{p!} T^{\mu_1 \dots \mu_p} dx_{\mu_1} \wedge \dots \wedge dx_{\mu_p}$$

where the “wedge” product  $\wedge$  is a skew-symmetric bilinear operation. Its main property is:

$$dx_i \wedge dx_j = -dx_j \wedge dx_i \quad (\text{i.e. } dx_i \wedge dx_i = 0)$$

Generally,

$$T_p \wedge T_q = (-1)^{pq} T_q \wedge T_p$$

The volume element of a space of dimension  $n$  can also be represented as an  $n$ -form:

$$dV = dx^1 \wedge \dots \wedge dx^n = \frac{1}{n!} \epsilon_{\mu_1 \dots \mu_n} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_n}$$

Below we will use the operator one-form

$$d = dx^\mu \frac{\partial}{\partial x^\mu}$$

that transforms  $p$ -forms into  $(p + 1)$ -forms:

$$dT_p = \frac{1}{p!} (\partial_\nu T_{\mu_1 \dots \mu_p}) dx^\nu \wedge dx^{\mu_1} \wedge \dots \wedge dx^{\mu_p}$$

From this definition the important property  $d^2 T_p = 0$  follows for arbitrary  $T_p$ . The main convenience of this notation is that we avoid writing tensor indices, thus saving a lot of ink.

Let us prove now that (6.34) is equivalent to (6.32). For this let us show first that the integrand in (6.34) is a total divergence. The easiest way of doing this is to consider a variation of the integrand under

variation of the field. We have:

$$\begin{aligned}
 \rho(x) &= \text{Tr}(F \wedge F) \\
 \delta\rho(x) &= 2 \text{Tr}(F \wedge \delta F) \\
 F &= dA + A \wedge A = \frac{1}{2} F_{\mu\nu} dx^\mu \wedge dx^\nu \\
 \delta F &= d\delta A + A \wedge \delta A + \delta A \wedge A = \nabla \cdot \delta A
 \end{aligned} \tag{6.35}$$

(where  $F$  is a 2-form;  $d$  is called the exterior derivative). For pedagogical reasons let us repeat (6.35) in the ordinary notation:

$$\begin{aligned}
 \rho(x) &= \frac{1}{4} \epsilon_{\mu\nu\lambda\rho} \text{Tr}(F_{\mu\nu} F_{\lambda\rho}) d^4x \\
 \delta\rho &= \frac{1}{2} \epsilon_{\mu\nu\lambda\rho} \text{Tr}(F_{\mu\nu} \delta F_{\lambda\rho}) d^4x \\
 F_{\mu\nu} &= \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu] \\
 \delta F_{\mu\nu} &= \partial_\mu \delta A_\nu - \partial_\nu \delta A_\mu + [A_\mu, \delta A_\nu] - (\mu \leftrightarrow \nu) = \nabla_\mu \delta A_\nu - \nabla_\nu \delta A_\mu
 \end{aligned} \tag{6.36}$$

From these equalities we obtain:

$$\begin{aligned}
 \delta\rho(x) &= 2 \text{Tr} F \wedge (d\delta A + A \wedge \delta A + \delta A \wedge A) \\
 &= 2 \text{Tr}(F \wedge d\delta A) + 2 \text{Tr}(F \wedge A - A \wedge F) \wedge \delta A \\
 &= 2 \text{Tr} F \wedge d\delta A + 2 \text{Tr} dF \wedge \delta A = 2d(\text{Tr} F \wedge \delta A) \\
 &= \epsilon^{\mu\nu\lambda\rho} \partial_\mu \text{Tr}(F_{\nu\lambda} \delta A_\rho) d^4x
 \end{aligned} \tag{6.37}$$

where we have used the Bianchi identity:

$$\begin{aligned}
 [\nabla, F] &= dF + A \wedge F - F \wedge A \\
 &= \frac{1}{2} (\partial_\nu F_{\lambda\rho} + [A_\nu, F_{\lambda\rho}]) dx^\nu \wedge dx^\lambda \wedge dx^\rho = 0
 \end{aligned} \tag{6.38}$$

We have:

$$\begin{aligned}
 \delta\rho(x) &= (\partial_\mu \delta \mathcal{K}^\mu(x)) d^4x \\
 \delta \mathcal{K}^\mu(x) &= \epsilon^{\mu\nu\lambda\rho} \text{Tr}(F_{\nu\lambda} \delta A_\rho)
 \end{aligned} \tag{6.39}$$

Now we have to obtain the current  $\mathcal{K}_\mu$  itself. To do this let us introduce a parameter  $\tau: 0 \leq \tau \leq 1$  and consider a family of gauge fields  $A_\mu(x, \tau) = \tau A_\mu(x)$ . According to (6.39) we have:

$$\begin{aligned}
 \frac{\partial \mathcal{K}^\mu(x, \tau)}{\partial \tau} &= \epsilon^{\mu\nu\lambda\rho} \text{Tr} \left( F_{\nu\lambda}(x, \tau) \frac{\partial A_\rho}{\partial \tau} \right) \\
 &= \epsilon^{\mu\nu\lambda\rho} \text{Tr} \{ (\tau(\partial_\nu A_\lambda - \partial_\lambda A_\nu) + \tau^2 [A_\nu, A_\lambda]) \cdot A_\rho \}
 \end{aligned} \tag{6.40}$$

Integrating (6.40) on  $\tau$  we obtain:

$$\begin{aligned}\mathcal{K}^\mu(x) &= \mathcal{K}^\mu(x, 1) \\ &= \frac{1}{2}\epsilon^{\mu\nu\lambda\rho} \text{Tr}((\partial_\nu A_\lambda - \partial_\lambda A_\nu + \frac{2}{3}[A_\nu, A_\lambda]) \cdot A_\rho)\end{aligned}\quad (6.41)$$

$$\begin{aligned}\frac{1}{2} \text{Tr}(F * F) &= \partial_\mu \mathcal{K}^\mu(x) \\ (*F_{\mu\nu} &= \frac{1}{2}\epsilon_{\mu\nu\lambda\rho} F^{\lambda\rho})\end{aligned}\quad (6.42)$$

Of course, after the answer (6.41) and (6.42) is known, it can be checked by direct computation.

Substitution of (6.42) into (6.34) gives:

$$q = \frac{1}{8\pi^2} \oint_{S^3} \mathcal{K}_\mu d^3\sigma^\mu \quad (6.43)$$

where we integrate over a large  $S^3$ . At these distances  $F_{\mu\nu} = 0$ , and (6.41) can be replaced by:

$$\begin{aligned}\mathcal{K}^\mu(x) &= \frac{1}{2}\epsilon^{\mu\nu\lambda\rho} \text{Tr}(\partial_\nu A_\lambda - \partial_\lambda A_\nu + \frac{2}{3}A_\nu A_\lambda) A_\rho \\ &\underset{x \rightarrow \infty}{\approx} -\frac{1}{3}\epsilon^{\mu\nu\lambda\rho} \text{Tr}(A_\nu A_\lambda A_\rho) \\ &\approx -\frac{1}{3}\epsilon^{\mu\nu\lambda\rho} \text{Tr}(L_\nu L_\lambda L_\rho)\end{aligned}\quad (6.44)$$

This proves the equivalence of (6.32) and (6.34).

Our aim now is to find an instanton solution with  $q = 1$ . As in the case of the  $n$ -field we can avoid solving the Yang-Mills equations themselves, by considering instead the “square root” of them. Let us use an identity:

$$\begin{aligned}S &= \frac{1}{4e_0^2} \int \text{Tr} F_{\mu\nu}^2 d^4x \\ &= \frac{1}{8e_0^2} \int \text{Tr}((F_{\mu\nu} - *F_{\mu\nu})^2) d^4x + \frac{1}{4e_0^2} \int \text{Tr}(F_{\mu\nu} *F_{\mu\nu}) d^4x \\ &= \frac{8\pi^2 q}{e_0^2} + \frac{1}{8e_0^2} \int \text{Tr}((F_{\mu\nu} - *F_{\mu\nu})^2) d^4x\end{aligned}\quad (6.45)$$

We see that if we find a solution of the “duality” equation

$$F_{\mu\nu} = *F_{\mu\nu} \quad (6.46)$$

then the action for a fixed  $q$  will be minimal. Actually, it is trivial to check that if the first-order equations (6.46) are satisfied, then the Yang-Mills equations

$$\nabla_\mu F^{\mu\nu} = 0 \quad (6.47)$$

will also be satisfied (the converse is not true). To check this, let us differentiate (6.46):

$$\nabla_\mu F^{\mu\nu} = \nabla_\mu {}^*F_{\mu\nu} \equiv 0 \quad (6.48)$$

(where the last equality is a consequence of the Bianchi identity:

$$\begin{aligned} \nabla_\alpha F_{\beta\gamma} + \nabla_\beta F_{\gamma\alpha} + \nabla_\gamma F_{\alpha\beta} &= 0 \\ \text{if } F_{\alpha\beta} &= \partial_\alpha A_\beta - \partial_\beta A_\alpha + [A_\alpha, A_\beta] \end{aligned} \quad (6.49)$$

Notice by the way that for constant  $A_\alpha$  this reduces to the Jacobi identity:

$$[A_\alpha, [A_\beta, A_\gamma]] + [A_\beta, [A_\gamma, A_\alpha]] + [A_\gamma, [A_\alpha, A_\beta]] = 0 \quad (6.50)$$

We see that the ‘‘duality’’ equations (6.46) are in some sense a four dimensional analogue of the Cauchy–Riemann equations (6.9). Their most surprising property is that they possess multi-instanton solutions. Before discussing them let us present a solution with  $q = 1$ . The ansatz for this solution can be found by the following trick.

Let us consider instead of the gauge group  $SU(2)$ , a group  $SU(2) \otimes SU(2) \simeq O(4)$ . Then equations (6.46) will have the symmetry group  $O(4) \otimes O(4)$ , where the first factor is space rotations, and the second isotopic rotations. We shall be looking for a solution which breaks  $O(4) \otimes O(4)$  but preserves the single  $O(4)$  formed by simultaneous rotations in  $x$ -space and isotopic space. After that we shall return to  $SU(2)$ .

Generators of  $O(4)$  are described by matrices  $I^{\alpha\beta}$ , skew-symmetric in  $(\alpha, \beta)$ , which represent a rotation in the  $(\alpha, \beta)$ -plane. Therefore, gauge fields for this group also have these indices:

$$A_\mu = A_\mu^{\alpha\beta}(x) \quad (6.51)$$

The most general  $O(4)$ -symmetric ansatz is given by:

$$A_\mu^{\alpha\beta}(x) = a(r^2)(x_\alpha \delta_{\mu\beta} - x_\beta \delta_{\mu\alpha}), \quad (r^2 = x_\mu x^\mu) \quad (6.52)$$

On symmetry grounds it must be compatible with the Yang–Mills and duality equations. The six fields  $A_\mu^{\alpha\beta}$  can be split into two sets, each of three fields,  $A_\mu^a$  and  $B_\mu^a$ , each corresponding to an  $SU(2)$ . This splitting is described by:

$$\begin{aligned} A_\mu^a &= \frac{1}{2}(A_\mu^{a0} + \frac{1}{2}\epsilon^{abc} A_\mu^{bc}) \equiv \frac{1}{4}\eta_{a\alpha\beta} A_\mu^{\alpha\beta} \\ B_\mu^a &= \frac{1}{2}(A_\mu^{0a} + \frac{1}{2}\epsilon^{abc} A_\mu^{bc}) \equiv \frac{1}{4}\eta_{a\alpha\beta} A_\mu^{\alpha\beta} \end{aligned} \quad (6.53)$$

(here  $\eta_{abc} = \epsilon_{abc}$ ,  $\eta_{ab0} = \delta_{ab}$ , etc.). The origin of this decomposition is based on a check that the commutation relations:

$$\begin{aligned} [I^{a\beta}, I^{\gamma\delta}] &= \delta^{\alpha\gamma} I^{\beta\delta} + \delta^{\beta\delta} I^{a\gamma} \\ &\quad - \delta^{\alpha\delta} I^{\beta\gamma} - \delta^{\beta\gamma} I^{a\delta} \end{aligned} \quad (6.54)$$

get decomposed as:

$$[X_a^\pm, X_b^\pm] = \epsilon_{abc} X_c^\pm, \quad [X_a^+, X_b^-] = 0$$

for

$$X_a^\pm = \frac{1}{2}(\frac{1}{2}\epsilon_{abc} I_{bc} \pm I_{a0})$$

We conclude from (6.52) and (6.53) that the ansatz:

$$A_\mu^a = -\frac{1}{2}\eta_{\alpha\mu\nu} x_\nu a(r^2) \quad (6.55)$$

is bound to be compatible with the duality equation. The substitution of (6.55) into (6.46) gives after some simple calculations:

$$\begin{aligned} A_\mu^a(x) &= \frac{2\eta_{\alpha\mu\nu}(x_\nu - a_\nu)}{(x - a)^2 + \rho^2} \\ F_{\mu\nu}^a(x) &= -\frac{4\eta_{\alpha\mu\nu}\rho^2}{((x - a)^2 + \rho^2)^2} \end{aligned} \quad (6.56)$$

with arbitrary scale parameter  $\rho$  and position parameter  $a_\mu$ .

This Non-Abelian instanton can be viewed as a magnetic dipole of size  $\rho$ . If we consider now the contribution of one instanton to the partition function  $Z$  we find, just as in case of  $n$ -field, several factors. First of all we have a factor  $e^{-S_{cl}} = e^{-8\pi^2/e_0^2}$  which gets replaced, after taking account of the one loop correction by  $e^{-8\pi^2/e^2(\rho)}$  where, for  $SU(N)$

$$e^2(\rho) = \frac{3}{11N} \frac{8\pi^2}{-\log(\lambda\rho)} \quad (6.57)$$

is an effective coupling for the size  $\rho$ . The contribution has to be integrated over  $\rho$  and  $a$ . The measure must be both scale and translationally invariant. The only combination with these properties is  $d^4R d\rho \rho^{-5}$ . We find from this consideration:

$$Z_{\text{INST}}^{(1)} \sim V \int \frac{d\rho}{\rho^5} e^{-8\pi^2/e^2(\rho)} = V \int \frac{d\rho}{\rho^5} \rho^{11N/3} \quad (6.58)$$

( $V$  is the 4-volume).

As happened in the case of the  $n$ -field, the instanton contribution has an infrared divergence. This implies that in the multi-instanton picture,

individual instantons tend to grow and to overlap. The naive dilute gas approximation is certainly inapplicable then, and we should expect something like dissociation of dipole-like instantons to their elementary constituents, as happened in the case of the  $n$ -field. However, even one loop computations on the multi-instanton background have not yet been performed, and nothing similar to the Coulomb plasma of the previous section has been discovered. This is connected partly with the fact that multi-instanton solutions have not been explicitly parametrized up to now. I expect many interesting surprises await us, even on the one loop level, in this hard problem.

There exists an interesting phenomenological description of the instanton liquid, which explains some qualitative features of hadrons. This approach<sup>†</sup> will not be discussed in this book.

So, our conclusion is that on the present level of understanding of instanton dynamics, we cannot obtain any exact dynamical statements concerning Non-Abelian gauge theory. In the case of  $n$ -fields the situation is slightly better, since we were able to demonstrate the appearance of the mass gap on a qualitative level. Even in this case one would like to have much deeper understanding of the situation. There are reasons to believe that some considerable progress will be achieved in the near future. In the case of gauge fields we have to pray for luck.

At the same time, the existence of fields with topological charge has a deep qualitative influence on the dynamical structure of the theory. We describe some of this in the next section.

### 6.3 Qualitative Effects of Instantons

The most dramatic manifestation of topological effects occurs when we take account of the interaction of massless Dirac fermions with instantons. We shall show in this section that instantons lead in this case to violation of some apparent conservation laws. Qualitatively, the effect can be described as follows. Let us examine an isospinor Dirac field  $\psi$  in the external Non-Abelian gauge field  $A_\mu$ . It is represented by the action:

$$S_\psi = \int d^4x \bar{\psi}(i\gamma^\mu(\partial_\mu + A_\mu))\psi \quad (6.59)$$

---

<sup>†</sup> Due to Callan and Gross (1979).

On a classical level, this action conserves an axial current:

$$\partial_\mu(i\bar{\psi}\gamma^\mu\gamma^5\psi) = 0 \quad (6.60)$$

However, if we consider the axial current which arises in the vacuum when we apply the external field  $A_\mu$ , then equation (6.60) will be untrue because of the so-called quantum anomaly. Let us show how this comes about. The partition function of fermions in the external field is given by:

$$Z[A] = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp\left(-\int \bar{\psi}i\gamma^\mu(\partial_\mu + A_\mu)\psi d^4x\right) \quad (6.61)$$

The induced axial current can be written as:

$$\begin{aligned} J_{\mu 5}(x, A) &= Z^{-1}[A] \times \int \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp\left(-\int \bar{\psi}i\gamma^\mu(\partial_\mu + A_\mu)\psi d^4x\right) i\bar{\psi}\gamma_\mu\gamma_5\psi \\ &= \text{diagram 1} + \text{diagram 2} + \text{diagram 3} + \dots \\ &= -i \text{Tr } \gamma_\mu\gamma_5 G(x, x; A) \end{aligned} \quad (6.62)$$

where  $G(x, x'; A)$  is a Green function for the Dirac operator in the field  $A_\mu$ . The definition (6.62) diverges because of singularities of the Green function at coincident points. It is necessary therefore to introduce a cut-off and to separate the divergent terms in  $J_{\mu 5}$ .

In order to perform this programme, let us express  $G(x, x'; A)$  in terms of eigenfunctions of the Dirac equation,  $\psi_n(x)$ :

$$i\gamma^\mu(\partial_\mu + A_\mu)\psi_n(x) = E_n\psi_n(x) \quad (6.63)$$

According to standard formulas, the Green function is given by:

$$G(x, x') = \sum_n \frac{\psi_n(x)\bar{\psi}_n(x')}{E_n} \quad (6.64)$$

In order to regularize (6.64) we intend to insert into sums over eigenstates a factor  $e^{-\epsilon E_n^2}$ , with  $\epsilon$  being of the order of  $\Lambda^{-2}$  and  $\Lambda$  being a momentum cut-off. The motivation for such a procedure is the following. High momentum divergences, or divergences for large  $n$  in (6.64) would not have arisen if we had worked with a theory on the lattice. While in this case the low lying  $E_n$  coincide with those of continuum theory, the higher  $E_n$  are not present at all because we have a finite number of degrees of freedom per unit volume. If we expect that the ultraviolet region produces only local effects, removed by renormalization, then we can imitate a lattice, which achieves this by means of a

natural cut-off, by the quenching factor  $e^{-\epsilon E_n^2}$ . The  $E_n^2$  is needed here, since  $E_n$  in (6.62) can be both positive and negative.

After these explanations let us compute the quantity:

$$J_{\mu 5}(x; A) = \sum_n \frac{i\bar{\psi}_n \gamma_\mu \gamma_5 \psi_n}{E_n} e^{-\epsilon E_n^2} \quad (6.65)$$

which is just the regularized axial current. We have:

$$\begin{aligned} \partial_\mu J_{\mu 5} &= \sum_n \frac{i \partial_\mu (\bar{\psi}_n \gamma_\mu \gamma_5 \psi_n)}{E_n} e^{-\epsilon E_n^2} \\ &= -2 \sum_n \bar{\psi}_n \gamma_5 \psi_n e^{-\epsilon E_n^2} = 2 \operatorname{tr} \langle x | \gamma_5 e^{-\epsilon \mathcal{D}} | x \rangle, \end{aligned} \quad (6.66)$$

$$(\mathcal{D} \equiv (i\gamma_\mu \nabla_\mu)^2)$$

Here we have used an identity, following straight from (6.63):

$$i \partial_\mu (\bar{\psi}_n \gamma_\mu \gamma_5 \psi_n) = -2E_n (\bar{\psi}_n \gamma_5 \psi_n)$$

The last term in (6.66) is easily calculable as  $\epsilon \rightarrow 0$ . We have:

$$\mathcal{D} = (i\gamma_\mu \nabla_\mu)^2 = -(\partial_\mu + A_\mu)^2 - \frac{1}{2} \sigma_{\mu\nu} F_{\mu\nu}; \quad \sigma_{\mu\nu} = \frac{1}{2} [\gamma_\mu, \gamma_\nu] \quad (6.67)$$

The most singular term in  $e^{-\epsilon \mathcal{D}}$  will come from  $\partial_\mu^2$  in (6.67). If we neglect all fields, then:

$$\langle x | e^{-\epsilon \mathcal{D}} | x \rangle = \int \frac{d^4 p}{(2\pi)^4} e^{-\epsilon p^2} = \frac{1}{16\pi^2 \epsilon^2} \quad (6.68)$$

Expansion in  $A_\mu$  and  $F_{\mu\nu}$  will give us less singular terms. These terms are important since for  $A_\mu = 0$ :

$$\operatorname{Tr}(\gamma_5 e^{-\epsilon \mathcal{D}}) = \frac{1}{16\pi^2 \epsilon^2} \operatorname{Tr} \gamma_5 = 0 \quad (6.69)$$

We see that it will be useless to expand  $(\partial_\mu + A_\mu)^2$  in terms of  $A_\mu$ , since  $\operatorname{Tr} \gamma_5 = 0$  will remain in our formulas. Expansion up to the first term in  $\sigma_{\mu\nu} F_{\mu\nu}$  also does not help since  $\operatorname{Tr} \gamma_5 \sigma_{\mu\nu} = 0$ . The first nonzero contribution arises from:

$$\operatorname{Tr}_x \gamma_5 e^{-\epsilon \mathcal{D}} \approx \langle x | e^{\epsilon \partial_\mu^2} | x \rangle \frac{\epsilon^2}{8} \operatorname{Tr}(\gamma_5 \sigma_{\mu\nu} \sigma_{\lambda\rho}) \cdot \operatorname{Tr}(F_{\mu\nu} F_{\lambda\rho}) \quad (6.70)$$

If we notice that:

$$\operatorname{Tr}(\gamma_5 \sigma_{\mu\nu} \sigma_{\lambda\rho}) = 4\epsilon_{\mu\nu\lambda\rho}$$

and compare (6.70) with (6.66) we get:

$$\begin{aligned} \partial_\mu J_{\mu 5}(x) &= \lim_{\epsilon \rightarrow 0} \sum_n \partial_\mu \frac{i\bar{\psi}_n \gamma_\mu \gamma_5 \psi_n}{E_n} e^{-\epsilon E_n x} \\ &= \frac{1}{8\pi^2} \text{Tr}(F_{\mu\nu} * F_{\mu\nu}) \end{aligned} \quad (6.71)$$

This “anomaly” relation and other similar ones have amazingly many consequences. Before coming to them, let us explain again how (6.71) arises.

We begin with a Lagrangian which conserves the axial current or, in other words it conserves separately the number of left- and right-handed fermions,  $\psi_{L,R} = \frac{1}{2}(1 \pm \gamma_5)\psi$ . Since this theory is divergent, we consider its regularized version, and then let the cut-off tend to infinity. After that we find a finite contribution (which is independent of the form of the cut-off) to the divergence of the axial current (6.71). This fact implies, that in the process of regularizing the theory we are bound to violate conservation of the axial current, and that in the limit  $\Lambda \rightarrow \infty$  a finite piece of this violation remains. Let us give some examples of this phenomenon. Take a nonrelativistic Fermi gas in one dimension in its ground state. This state is formed by particles with momenta  $-p_F < p < p_F$  where  $p_F$  is the Fermi-momentum. The spectrum of this system contains gapless excitations (if there is no superfluidity). These excitations are by particles and holes with momenta lying very near  $\pm p_F$ . The particle-hole excitation energy is given by:

$$E = \frac{(p_F + K_1)^2}{2} - \frac{(p_F - K_2)^2}{2} \approx p_F \cdot (K_1 - K_2) \quad (6.72)$$

where  $p_F + K_1$  is a particle momentum and  $p_F - K_2$  is a hole momentum. Therefore, all low lying excitations are described by two fields; the first,  $\psi_R(k)$  corresponding to the particles with  $p \simeq p_F + k$  (antiparticles for them are holes with  $p = p_F - k$ ) and  $\psi_L(k)$  corresponding to particles with  $p \simeq -p_F - k$ . These fields have a linear spectrum (according to (6.72) and satisfy the massless Dirac equation:

$$\begin{aligned} (\omega - k)\psi_L &= 0 \\ (\omega + k)\psi_R &= 0 \end{aligned} \quad (6.73)$$

The axial current is therefore conserved. However, we see that  $\psi_L(k)$  and  $\psi_R(k)$  are not independent fields in an exact sense—they become mixed as  $k \sim p_F$ . Therefore taking into account the cut-off region (in this case the cut-off is just the Fermi-momentum) leads to nonconserva-

tion of left- and right-handed particles separately. Whether this leads to a finite effect after removing the cut-off is another matter. It depends on the compensation between the small values of amplitudes involving  $p \sim p_F$  particles and the large number of particles in the Dirac sea.

Our computation which lead to (6.71) revealed that the compensation was exact in the  $\mathcal{D} = 4$  case. Although this computation was performed for a specific regularization of the theory, its final and finite result (6.71) can be shown to be regularization-independent. In more than one dimension, fermions need spin  $\sigma$  in order to be relativistic. In this case  $\omega = \sigma \mathbf{K}$ . Spinless fermions cannot be relativistic, from spin statistics theory.

We recognize in the right hand side of (6.71) the density of topological charge. As we have shown in Section 6.2 this can be written as the divergence of some current  $\mathcal{X}^\mu(x)$ . One might have thought, therefore, that the anomaly equation (6.71) does not break the conservation of axial charge, but rather redefines this current:  $J_{\mu 5} \rightarrow J_{\mu 5} - 4\pi^2 \mathcal{X}_\mu$ .

This is not the case. If we consider a field  $A_\mu$  with topological charge  $q$  and integrate equation (6.71) over the 4-volume, we obtain:

$$\Delta Q_5 = \int \hat{c}_\mu J_{\mu 5} d^4x = 2q \quad (6.74)$$

Here:

$$Q_5 = \int J_{05} d^3x = N_L - N_R$$

and we denote by  $\Delta Q_5$  the total change of  $Q_5$  for infinite time under the influence of  $A_\mu$ . Equation (6.74) shows that it is equal to twice the topological charge of  $A_\mu$ .

This amazing result means that there exists “compulsory” production of fermions and antifermions in topological nontrivial fields and that the numbers of left- and right-handed particles  $N_L$  and  $N_R$  necessarily change.

We derived all this using the Euclidean formalism, but there is no difficulty at all in understanding this effect in Minkowski space. To do this, let us recall the Minkowskian interpretation of instantons. As we have seen in Chapter 4, the instanton represents a tunnelling transition between two states divided by a barrier (recall the double-well example). In the real time formulation we have very many different trajectories connecting these two states. However, since the classical action does not have a corresponding extremum (since the transition is forbidden

classically) the interference among these trajectories is highly destructive and the amplitude is exponentially small. The instanton is a classical solution for imaginary time, which accounts for this smallness as described in Chapter 4.

Let us explain, after these reminiscences, what kind of states are connected by the instantons in the nonabelian gauge theories. In order to do this we can pass to the gauge  $A_0 = 0$  in the instanton solution, finding that the nonzero part of it,  $A_n(x, t)$ , has the property:

$$\begin{aligned} A_n(x, t) &\rightarrow 0 && t \rightarrow -\infty \\ A_n(x, t) &\rightarrow g^{-1}(x) \partial_n g(x) && t \rightarrow +\infty \end{aligned} \tag{6.75}$$

Here  $g(x)$  is a certain matrix, satisfying the condition  $g(\infty) = I$ . Therefore it maps  $x$ -space, which can be considered as  $S^3$  due to the last condition, onto the group  $G$ . This is just the same topologically nontrivial map, which was discussed in Section 6.2, only transcribed to the  $A_0 = 0$  gauge.

Now we can interpret the Yang-Mills instanton as follows. The configuration space in gauge theory is formed by all possible fields  $\{A_n(x)\}$ . Let us consider a transition in real time such that it lead from zero  $A_n$  at  $t = -\infty$  to  $g^{-1} \partial_n g$  with topologically nontrivial  $g(x)$  at  $t = +\infty$ . Due to this nontrivial  $g(x)$ , in the process of such a transition nonzero field strengths necessarily arise. (Otherwise the field  $g(x, t)$  in  $A_n(x, t) = g^{-1}(x, t) \partial_n g(x, t)$  would have interpolated between  $g(x)$  and  $I$ , which is impossible.)

The Minkowskian interpretation of (6.74) is that those field strengths are such that they lead to compulsory pair-creation. The word ‘‘compulsory’’ here means that the transition amplitude *without* pair creation is exactly zero.

It is interesting to see how these properties are fulfilled in terms of explicit expressions for the amplitudes.

The vacuum-vacuum amplitude, which as we said must be zero, is given by (6.61):

$$\begin{aligned} Z[A] &= \int \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp \left\{ - \int i\bar{\psi} \gamma_\mu (\partial_\mu + A_\mu) \psi \, d^4x \right. \\ &= \text{Det}(i\gamma_\mu (\partial_\mu + A_\mu)) \end{aligned} \tag{6.76}$$

(We obtain the determinant to the first power and not the inverse determinant because  $\psi$  are anticommuting variables, or in other words, fermionic loops come with a negative sign.)

The amplitude (6.76) may be seen to be zero because the Dirac operator, as we shall show in a moment, has zero eigenvalues in the topological fields, and the determinant, being a product of the eigenvalues, is zero.

The proof that the Dirac operator has zero eigenvalues is based on the Atiyah–Singer index theorem. We shall derive here a special case of this theorem sufficient for our purposes. To do this, let us notice that according to (6.71):

$$\int d^4x \lim_{\epsilon \rightarrow 0} \sum_n \bar{\psi}_n(x) \gamma_5 \psi_n(x) e^{-\epsilon E_n^2} = -q \tag{6.77}$$

The theorem follows from (6.77) after observation that all nonzero modes give zero contribution to (6.77). This happens because nonzero eigenvalues appear in pairs symmetric under reflection  $E_n \rightarrow -E_n$ ,  $\psi_n \rightarrow \gamma_5 \psi_n$ . (Check this from eq. (6.63).) The value of  $\bar{\psi}_n \gamma_5 \psi_n$  under this reflection changes its sign. This proves the cancellation.)

As far as zero eigenmodes are concerned, they can be, and really are, asymmetric. Since the equation

$$\gamma_\mu \nabla_\mu \psi^{(0)} = 0 \tag{6.78}$$

is  $\gamma_5$ -invariant,  $\psi^{(0)}$  may be purely left or right:

$$\gamma_5 \psi_{L,R}^{(0)} = \pm \psi_{L,R}^{(0)} \tag{6.79}$$

If we denote by  $n_{L,R}$  the number of corresponding zero modes we get a beautiful theorem:

$$n_R - n_L = q \tag{6.80}$$

This result shows that for  $q \neq 0$  we indeed have zero modes and thus that the vacuum–vacuum transition is zero. Moreover, we can easily compute the nonzero matrix elements for which the selection rule (6.74) is satisfied. Consider the Green functions instead of  $Z$ :

$$G(x_i, y_i) = Z^{-1} \int e^{-S_\psi} \mathcal{D}\psi \mathcal{D}\bar{\psi} (\psi(x_1) \dots \psi(x_N) \bar{\psi}(y_1) \dots \bar{\psi}(y_N)) \tag{6.81}$$

We find that this quantity is not well defined in the instanton field because  $Z = 0$ . This just means that the Green function usually represents a transition amplitude divided by the amplitude for the vacuum to remain unchanged. This is not possible in the instanton field

and we have to consider the absolute value of the particle creation amplitude, described by

$$\begin{aligned} Z \cdot G(x_i, y_i) &= F(x_i, y_i) \\ &= \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-S\psi(\psi(x_1) \dots \psi(x_N) \bar{\psi}(y_1) \dots \bar{\psi}(y_N))} \end{aligned} \quad (6.82)$$

Since the action is quadratic this amplitude can be computed by expansion in normal modes:

$$\psi(x) = \sum_{\alpha} C_{0\alpha} \psi_{0\alpha}(x) + \sum_{n \neq 0} C_n \psi_n(x) \quad (6.83)$$

Here  $\{\psi_{0\alpha}\}$  are the zero modes of the Dirac operator and the second term represents the nonzero modes' contribution. The key point in computing (6.82) lies in the Berezin rule:

$$\int dC = 0 \quad \text{and} \quad \int C dC = 1 \quad (6.84)$$

for anticommuting variables. Since

$$\mathcal{D}\psi \mathcal{D}\bar{\psi} = \prod_{\alpha} dC_{0\alpha} d\bar{C}_{0\alpha} \prod_{n \neq 0} dC_n d\bar{C}_n \quad (6.85)$$

and

$$S(\psi, \bar{\psi}) = \sum_{n \neq 0} E_n \bar{C}_n C_n \quad (6.86)$$

we must carefully collect in the integrand of (6.82) the terms containing a product of all the  $C_{0\alpha} \bar{C}_{0\alpha}$ . All other terms will give zero according to (6.84). Since each left-handed  $\psi_{0\alpha}$  gives a right-handed  $\bar{\psi}_{0\alpha}$  the amplitude will be nonzero, only if the selection rule (6.74) is satisfied. In this case it is proportional to the product of the corresponding zero mode eigenfunctions  $\psi_{0\alpha}(x)$ .

The above configuration treated  $A_{\mu}$  as an external field. It is quite obvious, however, that if we take a functional integral over  $A_{\mu}$ , including in it nontrivial topological fields, then we shall obtain (with an amplitude  $\sim e^{-8\pi^2/e_0^2}$ ) nonconservation of the axial current. This effect leads to important physical consequences for strong and weak interactions briefly described in the remarks to this chapter. Here we shall mention another aspect of the above result. Namely it shows that massless quarks tend to suppress instanton contribution, because

$Z[A] = 0$  in the instanton field. If we consider instanton-anti-instanton configurations then their contribution will be nonzero, due to the fact that the total topological charge is zero. But the effective action  $U(R_{12})$ , where  $R_{12}$  is the distance between our objects, must have the property

$$U(R_{12}) \xrightarrow{R_{12} \rightarrow \infty} \infty \quad (6.87)$$

This means, that exchange of a massless fermion pair leads to long-range forces between instantons and anti-instantons. The result of this may have several alternative consequences. The first one is that since (6.87) implies quenching of large fluctuations in the presence of massless fermions, the system loses the confining property and we would end up with massless gauge fields together with fermions. This option seems highly improbable to me on the basis of some analogies and some model considerations. However, I am not aware of any strict statements permitting us to reject it.

The second possibility, which in my opinion is realized in the theory, is the following. Due to the strong binding force between fermions the chiral symmetry gets spontaneously broken and as a result the fermions acquire a mass. After that has happened, the long range force between instantons and anti-instantons disappears, being screened by the fermionic mass term in the effective lagrangian. The only remaining effect of anomalous non-conservation will consist of giving a mass to the corresponding Goldstone boson.

There is also another improbable option, namely that instantons get confined but some other type of large fluctuations, not suppressed by fermions, disorder the system.

Unfortunately, at present we are unable to make a decisive choice between the options.

Let us discuss another qualitative phenomenon, arising because of the instantons. The Lagrangian density for Yang-Mills fields is conventionally taken to be  $(1/4e_0^2) \text{Tr}(F_{\mu\nu}^2)$ . The standard reason for this choice is that this is the only invariant expression of dimension 4. Any higher invariant terms like  $\text{Tr}(\nabla_\alpha F_{\mu\nu})^2$  will be irrelevant in the infrared region and can be omitted. This reason overlooks another invariant expression,  $\text{Tr} F_{\mu\nu} * F_{\mu\nu}$ , on the basis that it is a total divergence, which has no influence on the equations of motion. But, as we already know, instantons activate this total divergence. Therefore the most general Lagrangian of dimension 4 has the form

$$\mathcal{L} = -\frac{1}{4e_0^2} \text{Tr} F_{\mu\nu}^2 + \frac{i\theta}{16\pi^2} \text{Tr} F_{\mu\nu} * F_{\mu\nu} \quad (6.88)$$

for the Yang–Mills theory, and the analogous Lagrangian for the  $\mathbf{n}$ -field

$$\mathcal{L} = \frac{1}{2e_0^2} (\partial_\mu \mathbf{n})^2 + \frac{i\theta}{8\pi} \epsilon_{\mu\nu} \mathbf{n} [\partial_\mu \mathbf{n} \times \partial_\nu \mathbf{n}] \quad (6.89)$$

where  $\theta$  in both cases is a new coupling constant. Due to the presence of instantons, physical transition amplitudes will depend on  $\theta$ . For instance, the vacuum to vacuum amplitude will be

$$Z = \sum_{q=-\infty}^{+\infty} e^{i\theta q} Z_q \quad (6.90)$$

where by  $Z_q$  we denoted the functional integral over the fields with a fixed value of  $q$ .

The extra terms in the above expressions are purely imaginary in Euclidean space for the following reasons. We have to have a real action in Minkowski space. As we change  $t \rightarrow -it$  we have to change  $A_n \rightarrow A_n$  and  $A_0 \rightarrow iA_0$ , because  $A_0$  transforms as  $\partial/\partial t$ . Hence electric and magnetic fields change as:

$$\begin{aligned} E_n &\rightarrow iE_n \\ H_n &\rightarrow H_n \end{aligned} \quad (6.91)$$

The two terms in (6.88) change as:

$$\begin{aligned} \text{Tr } F_{\mu\nu}^2 &= E^2 - H^2 \rightarrow -(H^2 + E^2) \\ \text{Tr } F_{\mu\nu} * F_{\mu\nu} &= -2\mathbf{H} \cdot \mathbf{E} \rightarrow -2i\mathbf{H} \cdot \mathbf{E} \end{aligned} \quad (6.92)$$

So, in the Euclidean action (6.88) we obtain  $i$  in the second term when we transform the real Minkowskian action.

Topological “ $\theta$ -terms” in the action create a problem when we apply the theory to describe strong interactions. If they really do contribute to the physical amplitudes, then the whole theory loses invariance under time reversal (since  $\mathbf{H} \cdot \mathbf{E}$  is  $T$ -odd). This implies that for some reasons the coupling  $\theta$  must be zero or extremely small.

Certainly the existence of this problem somewhat depends on one’s personal philosophy. One can take the view that in the cut-off region, i.e. at the Planck length, we have a gauge Lagrangian which preserves  $T$ -invariance. Then  $\theta = 0$  from the very beginning and no problem of strong  $T$ -noninvariance arises. It is interesting, however, to consider another point of view, according to which there are no special symmetries at the Planck length and they persist in the low energy region only for dynamical reasons, namely because only renormalizable interactions contribute significantly for large scales. Accepting this view, we

have to consider all possible terms in the lagrangian, having dimensions four or less with coefficients having their natural magnitude. If so, we do have a strong  $T$ -violation problem. An attempt to resolve this problem led to an interesting suggestion [4]. It can be shown that if massless fermions with broken chiral symmetry are present in the system, then due to instanton effects Goldstone's massless particles obtain some mass (because of nonexact conservation of  $J_{\mu 5}$ ) and simultaneously the  $\theta$ -term gets absorbed after redefinition of Goldstone's field. This consideration predicted a light isoscalar boson (which would have been massless without instantons). Unfortunately this particle, called the axion, has not been found, and the strong  $\theta$ -problem remains open.

Let us mention briefly some other interesting effects of instantons together with the  $\theta$ -term. It appears that their presence induces a small electric charge  $\sim \theta$  for a magnetic monopole, and in this case the operator of electric charge differs from a gauge group generator by a small constant.

Another interesting thing is a probably rich phase structure of the theory as a function of  $\theta$ . In the case of  $n$ -fields, periodic  $\theta$ -dependence seems to have important consequences, explaining the quantized Hall effect in metals.

Let us stress that  $\theta$ -dependence of physical quantities is not to be taken for granted from what we have said about it. It is quite possible that due to dynamical effects, the contributions with  $q \neq 0$  may become absent in the infinite volume limit. Example of such a phenomenon is observed in the case of plasma where only states with total charge zero contribute in the thermodynamic limit (because of the enormous Coulomb energy of other states). However, this option seems unlikely in Non-Abelian gauge theory.

Another word of caution concerns the apparent periodicity in  $\theta$ . It might well happen that because of dissociation of instantons, states with noninteger topological charge will give finite contributions to the partition function. These configurations certainly have infinite action, but this can be compensated by the entropy. As a result, the formal decomposition (6.30) will be untrue and we shall now have a periodic  $\theta$ -dependence.

All these questions are purely dynamical and their solution requires some new concepts. In the next chapters we shall discuss what has been done in this direction.



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

---

## CHAPTER 7

# Analogies Between Gauge and Chiral Fields. Loop Dynamics

---

---

DOI: 10.1201/9780203755082-7

In the previous chapters we indicated some similarities between chiral fields and gauge fields. We saw that both theories—the one for  $\mathcal{D} = 2$  and the other for  $\mathcal{D} = 4$  share the property of asymptotic freedom. Their strong coupling expansions also looked similar, provided that we substitute point-like excitations of chiral fields by closed flux lines for the gauge fields. Recall also the similarity of the instanton structures of  $n$ -fields and Yang–Mills fields.

In this chapter we shall explore the reasons for these analogies. Roughly speaking, it appears that small pieces of the flux lines in the gauge case move as if they were point-like excitations of a chiral field. On the other hand, the gauge field itself can be considered as a chiral field defined on the space of all possible loops.

These ideas are not well developed. But they provide us with a bridge between gauge theories, chiral fields and the theory of strings (discussed in Chapter 9). They may also help in a search for hidden symmetries in many-dimensional systems. One cannot help feeling that many beautiful secrets are concealed in loop space.

This space is the main topic of discussion below.

### 7.1 Non-Abelian Phase Factor

In the previous chapter we have already been dealing with a field:

$$\begin{aligned}
\Psi(C) &= \text{P exp} \oint_C A_\mu dx^\mu \\
&= \text{P exp} \left( \int_0^{2\pi} ds A_\mu(x(s)) \frac{dx^\mu(s)}{ds} \right) \\
&= \lim_{\Delta x_j \rightarrow 0} \prod_j (1 + A_\mu(x_j) \Delta x_j^\mu) \\
&= \sum_{n=0}^{\infty} \int_0^{2\pi} ds_1 \int_0^{s_1} ds_2 \int_0^{s_2} ds_3 \cdots \int_0^{s_{n-1}} ds_n \frac{dx^\mu(s_n)}{ds_n} A_\mu(x(s_n)) \cdots \frac{dx^\mu}{ds_1} A_\mu(x(s_1))
\end{aligned} \tag{7.1}$$

(Here  $\text{P}$  is the Dyson ordering operation,  $C$  is a loop parametrized by  $x_\mu = x_\mu(s)$ ,  $0 \leq s \leq 2\pi$ ;  $A_\mu$  are matrices lying in the Lie algebra of the gauge group). For Abelian  $A_\mu$  it would have been possible to symmetrize the integrand, obtaining  $n$ -fold integration from 0 to  $2\pi$ . Then  $\Psi$  would be an ordinary exponential. For the Non-Abelian case the ordering is very important.

The most important property of (7.1) is its simple behaviour under a gauge transformation of  $A_\mu$ . We have:

$$\begin{aligned}
A_\mu(x) &\rightarrow \Omega^{-1}(x) A_\mu(x) \Omega(x) + \Omega^{-1} \partial_\mu \Omega \\
\Psi(C) &\rightarrow \Omega^{-1}(x(0)) \Psi(C) \Omega(x(2\pi))
\end{aligned} \tag{7.2}$$

(where  $x(0) = x(2\pi)$  is the beginning of the loop).

For an open path  $C_{xy}$  connecting points  $x$  and  $y$ , the transformation law has the form:

$$\begin{aligned}
\Psi(C_{xy}) &= \text{P exp} \int_{C_{xy}} A_\mu dx^\mu \\
\Psi(C_{xy}) &\rightarrow \Omega^{-1}(x) \Psi(C_{xy}) \Omega(y)
\end{aligned} \tag{7.3}$$

The lattice version of  $\Psi$  is given by the formula

$$\Psi(C_{xy}) = \prod_{(C_{xy})} B_{z,\alpha} \tag{7.4}$$

where  $\prod_{(C_{xy})}$  is the product of matrices  $B_{z,\alpha}$  attached to the links, forming the contour  $C_{xy}$ .

Formulas (7.2) and (7.3) are most easily checked in the lattice version (7.4), since under gauge transformation

$$B_{z,\alpha} \rightarrow \Omega_z^{-1} B_{z,\alpha} \Omega_{z+\alpha} \tag{7.5}$$

and the factors  $\Omega$  in (7.4) cancel each other except at the ends. The factors  $\Psi(C_{xy})$  can be considered as matrices of “parallel transport” of different quantities. This means the following. Suppose we have a field  $\varphi(y)$  at the point  $y$  which transforms according to the fundamental representation of the gauge group:

$$\varphi(y) \rightarrow \Omega^{-1}(y)\varphi(y) \tag{7.6}$$

Without the gauge fields, the quantities  $\varphi(y)$  cannot be compared with the field  $\varphi(x)$  which has a different transformation law (with the matrix  $\Omega^{-1}(x)$ ). But, having a gauge field at our disposal, we can perform a “parallel transport” of  $\varphi(y)$  to the point  $x$ , by defining the transported field “ $\varphi(x)$ ” as

$${}_{\parallel}\varphi(x) = \Psi(C_{yx})\varphi(y) \tag{7.7}$$

Now, the field “ $\varphi(x)$ ” has the same transformation law as  $\varphi(x)$  and we can consider a covariant change of the field  $\varphi$  as we move from  $x$  to  $y$  along  $C_{xy}$ :

$$\begin{aligned} \delta_{C_{xy}}^{\text{cov}}\varphi &= \varphi(x) - {}_{\parallel}\varphi(x) \\ &= \varphi(x) - \Psi(C_{xy})\varphi(y) \end{aligned} \tag{7.8}$$

For infinitesimal  $C_{xy}$  this change becomes independent of  $C_{xy}$ :

$$\begin{aligned} \delta^{\text{cov}}\varphi &= \varphi(x) - [1 + A_\mu(x)(y_\mu - x_\mu)]\varphi(y) \\ &\approx (\partial_\mu + A_\mu)\varphi(x)(x_\mu - y_\mu) \end{aligned} \tag{7.9}$$

thus defining the covariant derivative.

The role of the gauge field in this geometrical language is seen to permit comparison of the fields in different points. For that reason the gauge field is called a “connection” by mathematicians. A field strength  $F_{\mu\nu}$  can be defined through the field change after the parallel transport along a small closed loop:

$$\varphi(x) - {}_{\parallel}\varphi(x) \simeq \frac{1}{2}F_{\mu\nu}\sigma^{\mu\nu}\varphi(x) \tag{7.10}$$

( $\sigma^{\mu\nu}$  is the area of the loop), where

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

the mathematical name for  $F_{\mu\nu}$  is “curvature”.

We see that the phase factor  $\Psi(C)$  is an important object for differential geometry. On the other hand it has a well-defined physical meaning. Let us consider a quantum particle moving in the external gauge field  $A_\mu$ . For  $A_\mu = 0$  the transition amplitude from  $x$  to  $y$  is given by a functional integral over trajectories, connecting  $x$  and  $y$ , each entering with the weight  $e^{iS_0}$  ( $S_0$  being the classical action for a free particle on the trajectory). It is not difficult to show, that as we switch on  $A_\mu$ , the amplitude becomes  $e^{iS_0}\Psi(C_{xy})$  integrated over all trajectories  $C_{xy}$ . If our particle moves along a classical trajectory in space-time (but has quantized colour) then  $\Psi(C_{xy})$  defines its transition amplitude. This conforms with what we said about  $\Psi$  in Chapter 4.

The natural intention, which stems from both the mathematical and the physical importance of the phase factor, is to attempt to reformulate gauge theory in their terms.

As a first step, let us derive classical equations of motion. In the lattice version with the action

$$S = \frac{1}{e_0^2} \text{Tr} \sum_{x,\alpha,\beta} (B_{x,\alpha} B_{x+\alpha,\beta} B_{x+\beta,\alpha}^{-1} B_{x,\beta}^{-1}) \quad (7.11)$$

we can obtain classical equations of motion by taking a variation:

$$\begin{aligned} \delta B_{xx} &= \omega_{xx} B_{xx} \\ \delta B_{xx}^{-1} &= -B_{xx}^{-1} \delta B_{xx} B_{xx}^{-1} = -B_{xx}^{-1} \omega_{xx} \end{aligned} \quad (7.12)$$

(here  $\omega_{xx}$  is an arbitrary element of the Lie algebra of  $G$ , while  $B_{xx}$  is an element of  $G$  itself). The classical equations of motion are:

$$\frac{\partial S}{\partial \omega_{xx}} = 0 \quad (7.13)$$

Their explicit form has a useful graphical representation. To find it let us consider at first the variation of the first factor,  $B_{xx}$ , in (7.11). We have:

$$\begin{aligned} \delta^{(1)} S &\sim \text{Tr} \sum_{x,\alpha,\beta} \left( \omega_{xx} B_{x,\alpha} B_{x+\alpha,\beta} B_{x+\beta,\alpha}^{-1} B_{x,\beta}^{-1} \right) \\ &= \sum_{x,\alpha,\beta} \text{Tr} \left( \omega_{x,\alpha} \begin{array}{c} \square \\ \text{---} \beta \\ \text{---} \alpha \\ \text{---} \end{array} \right) \end{aligned} \quad (7.14)$$

where we draw a plaquette  $(x, \alpha, \beta)$  starting from the point  $x$ , and associate with each link the matrix  $B_{x,\alpha}$ . The product starts from the link  $(x, \alpha)$  and proceeds anti-clockwise.

In the second term

$$\delta^{(2)}S \sim \text{Tr} \sum_{x, \alpha, \beta} (B_{x\alpha} \omega_{x+\alpha, \beta} B_{x+\alpha, \beta} B_{x+\beta, \alpha}^{-1} B_{x, \beta}^{-1}) \quad (7.15)$$

we shall change  $x + \alpha \rightarrow x, (\alpha \rightleftharpoons \beta)$  and use cyclic invariance of the trace so as to have  $\omega$  in the first place. We perform similar transformations with the other terms. Recalling that for  $SU(N)$   $\omega$  is an arbitrary anti-Hermitian matrix we obtain the following graphical equations:

$$\sum_{\beta} \left\{ \begin{array}{c} \text{square with } \omega \text{ on top, } x \text{ on left, } \beta \text{ on right} \\ \text{square with } \omega \text{ on top, } -x \text{ on left, } -\beta \text{ on right} \end{array} - \text{h.c.} \right\} = 0 \quad (7.16)$$

(h.c., means Hermitian conjugation). The same equation can be rewritten as

$$\sum_{\beta} \left\{ \begin{array}{c} \text{square with } \omega \text{ on top, } \beta \text{ on right} \\ \text{square with } \omega \text{ on top, } \beta \text{ on left} \end{array} - \text{h.c.} \right\} = 0 \quad (7.17)$$

or in analytic notation:

$$\sum_{\beta} \{ (F_{x, x, \beta} - B_{x-\beta, \beta}^{-1} F_{x-\beta, x\beta} B_{x-\beta, \beta}) - \text{h.c.} \} = 0 \quad (7.18)$$

$$F_{x, x\beta} = B_{x, x} B_{x+\alpha, \beta} B_{x+\beta, x}^{-1} B_{x, \beta}^{-1}$$

The form (7.18) has the advantage of having a familiar continuum limit: if we set  $B_{x\alpha} \simeq 1 + A_x$  and the potentials are slowly varying, we have:

$$\hat{c}_{\beta} F_{x\beta} + [A_{\beta}, F_{x\beta}] = 0 \quad (7.19)$$

Let us transcribe the equation for  $B_{x, \alpha}$  to equations for the phase factor  $\Psi(C_{xy})$ . To do this we introduce the "current" defined by

$$\mathcal{F}_{\alpha}(s, C) = \Psi(C + \prod_{s\alpha}) \Psi^{-1}(C) \quad (7.20)$$

where we denote by  $C + \prod_{s\alpha}$  the contour obtained from  $C$  by replacing the link  $\#s$  by a segment shaped like the letter  $\Pi$  in the direction  $\alpha$ . It is supposed that  $\alpha$  is orthogonal to the direction of the original link. Graphically:

$$\mathcal{F}_{s\alpha}(C) = \begin{array}{c} \text{contour } C \text{ with a } \Pi \text{ segment on link } s \end{array} = \begin{array}{c} \text{contour } C \text{ with a } \Pi \text{ segment on link } s \end{array}$$

From this picture one sees that  $\mathcal{F}$  is nothing but the field strength, transported along  $C$  to the beginning of the loop,  $x$ . It is now easy to check that equations (7.17) imply that

$$\sum_x (\mathcal{F}_{sa}(C) - \mathcal{F}_{sa}(C - \Pi_{sa})) - \text{h.c.} = 0 \quad (7.21)$$

Before discussing the meaning of these relations, let us present their continuous version. Its form can be extracted from the preceding formulas. Namely, in this case the field  $\Psi(C) = \Psi[x(s)]$  is a functional depending on the shape of the contour but not on its parametrization. Therefore:

$$\Psi[x(s)] = \Psi[x(\alpha(s))] \quad (7.22)$$

and

$$\frac{dx_\mu(s)}{ds} \frac{\delta\Psi}{\delta x_\mu(s)} = 0 \quad (7.23)$$

Relation (7.23) follows from (7.22) if one takes  $\alpha(s) = s + \epsilon(s)$  with infinitesimal  $\epsilon(s)$ . The variational derivative of any functional  $f$  is defined through the relation:

$$\delta f = \int ds \delta x_\mu(s) \left( \frac{\delta f}{\delta x_\mu(s)} \right) \quad (7.24)$$

Let us compute  $\delta\Psi/\delta x_\mu$  using the definition (7.1). For that we shall use the following general relations:

$$\begin{aligned} \frac{d}{dt} \text{P exp} \int_0^t M d\tau &= \left( \text{P exp} \int_0^t M d\tau \right) M(t) \\ \delta \text{P exp} \int_0^{2\pi} M(\tau) d\tau &= \int_0^{2\pi} dt \left( \text{P exp} \int_0^t M(\tau_1) d\tau_1 \right) \delta M(t) \\ &\quad \times \left( \text{P exp} \int_t^{2\pi} M(\tau_2) d\tau_2 \right) \\ &\equiv \int_0^{2\pi} dt \text{P} \left( \delta M(t) \exp \int_0^{2\pi} M(\tau) d\tau \right) \\ \delta^2 \left( \text{P exp} \int_0^{2\pi} M(\tau) d\tau \right) &= \iint_0^{2\pi} dt_1 dt_2 \text{P} \left( \delta M(t_1) \delta M(t_2) \exp \int_0^{2\pi} M(\tau) d\tau \right) \end{aligned}$$

for an arbitrary matrix  $M(\tau)$ .

With the aid of these formulas we obtain:

$$\delta\Psi(C) = \int_0^{2\pi} P\left(F_{\mu\nu}(x(s)) \exp \int_0^{2\pi} A_\mu dx^\mu\right) \dot{x}_\nu \delta x_\mu(s) ds \quad (7.25)$$

where  $F_{\mu\nu}$  is the Yang-Mills field strength. We see that the continuum counterpart of the quantity (7.20) is given by:

$$\begin{aligned} \mathcal{F}_\mu(s, C) &= \frac{\delta\psi(C)}{\delta x_\mu(s)} \Psi^{-1}(C) \\ &= \dot{x}_\nu \left( P \exp \int_0^s A_\mu \dot{x}_\mu dt F_{\mu\nu}(x(s)) P \exp\left(-\int_0^s A_\mu \dot{x}_\mu dt\right) \right) \end{aligned} \quad (7.26)$$

By its definition it satisfies:

$$\begin{aligned} \dot{x}_\mu(s) \mathcal{F}_\mu(s, C) &= 0 \\ \frac{\delta\mathcal{F}_\mu(s, C)}{\delta x_\nu(s')} - \frac{\delta\mathcal{F}_\nu(s, C)}{\delta x_\mu(s')} + [\mathcal{F}_\mu(s, C), \mathcal{F}_\nu(s', C)] &= 0 \end{aligned} \quad (7.27)$$

The second relation expresses the fact that a gauge field having nonzero field-strength in ordinary space, in the loop space has zero curvature being thus a chiral field.

If we take a functional divergence of (7.26) we obtain:

$$\begin{aligned} \frac{\delta\mathcal{F}_\mu(s, C)}{\delta x_\mu(s)} &= \delta(0) \left( P \exp \int_0^s A_\mu \dot{x}_\mu dt \right. \\ &\quad \left. \times (\partial_\mu F_{\mu\nu} + [A_\mu, F_{\mu\nu}]) P \exp\left(-\int_0^s A_\mu \dot{x}_\mu dt\right) \right) \dot{x}_\nu \end{aligned} \quad (7.28)$$

From this relation we conclude that the Yang-Mills equations in terms of phase factors have the form:

$$\frac{\delta}{\delta x_\mu(s)} \left( \frac{\delta\Psi}{\delta x_\mu(s)} \Psi^{-1} \right) = 0 \quad (7.29)$$

To be sure (7.29) is just a continuum version of the lattice equations (7.18) and (7.21).

These formulas show a remarkable correspondence with equations of motion for chiral fields. Indeed, on the lattice the action

$$S = -\frac{1}{e_0^2} \sum_{\mathbf{x}, \mathbf{s}} \text{Tr}(g_{\mathbf{x}}^{-1} g_{\mathbf{x}+\mathbf{s}}) \quad (7.30)$$

leads to equations of motion (obtained by a variation  $\delta g_x = \omega_x g_x$ ):

$$\left. \begin{aligned} \sum_{\delta} (A_{x,\delta} - A_{x-\delta,\delta}) - \text{h.c.} &= 0 \\ A_{x,\delta} &= g_{x+\delta} g_x^{-1} \end{aligned} \right\} \quad (7.31)$$

or in a continuum version:

$$\begin{aligned} A_{\mu} &= (\partial_{\mu} g) g^{-1} \\ \partial_{\mu} A_{\nu} - \partial_{\nu} A_{\mu} + [A_{\mu}, A_{\nu}] &= 0 \\ \partial_{\mu} A_{\mu} &= 0 \end{aligned} \quad (7.32)$$

This is to be compared with (7.27) and (7.29).

Although we have derived this analogy from the classical equations of motion it persists on the quantum level as well. The classical equations of motion get transcribed in quantum theory into equations for correlations functions. The standard way to obtain these equations is to consider an infinitesimal change of variables in a functional integral (described by (7.12) in our case) and to apply the condition that the integral remains unchanged. In this way one easily obtains a Schwinger-like chain of equations in our loop space.

That will be done in the next section. Before coming to it let us sum up what we know about the connections between gauge fields and chiral fields. We have established above that gauge fields are chiral fields defined on the loop space. Unfortunately, up to now it has not been possible to extract any practical, dynamical information from this fact. This mainly has to do with the difficulties we experience in treating equations in loop space or, what is more or less the same, with string dynamics. These problems we shall discuss in Chapter 9. I anticipate an enormous progress in this field in the near future.

On the other hand, there are several, more pragmatic, similarities between the theories under discussion. Asymptotic freedom, instantons and also large  $N$  behaviour (see Chapter 8) and the ‘‘chiral’’ form of the duality equations (see Chapter 6) make chiral theories an excellent theoretical laboratory for the study of gauge fields.

In the next section we shall study the dynamics of loops in a very imperfect way which is all that is known to-day. Still, it will have important consequences for the large  $N$ -expansion, examined in Chapter 8.

There are many obvious but unanswered questions concerning our subject. For example, does the similarity of equations (7.31) and (7.21) imply any similarities in the strong coupling expansions for these theories? In particular, it seems probable that scattering small bits of

string (which, according to Chapter 3 are elementary excitations in the gauge case) is precisely the same as scattering of point-like excitations of a chiral field. This could be checked in the frame of the strong coupling expansion, but never has been. Another problem is to derive asymptotic freedom directly in the loop space, following the pattern we had for chiral fields in Chapter 2.

Another set of questions is the following. We derived the loop representation starting from local gauge fields. However, another view is possible (and seems even more natural to me). Let us suppose that the primary quantities are the fields on the loop space. Then we can consider gauge fields as something like Goldstone fields on the loop space, efficiently described by the chiral equations of motion (7.29). This point of view leads to many new options. For example we can consider, instead of  $\Psi$  fields which belong to a gauge group  $G$ , fields belonging to different coset spaces  $G/H$ . These seem to be interesting objects, though I do not know either their local representation or their role (if any) in Nature.

## 7.2 Quantum Theory of Loops

It appears more practical to use a set of equations slightly different from (7.29). It exploits the fact that, according to the transformation law (7.2), only  $\varphi(C) = \text{Tr } \Psi(C)$  is a gauge-invariant quantity. Therefore the only quantities which make sense in quantum theory are:

$$W(C) = \langle \varphi(C) \rangle; \quad W_2(C, \tilde{C}) = \langle \varphi(C)\varphi(\tilde{C}) \rangle \quad (7.33)$$

etc. Using formulas (7.24), (7.25) we derive:

$$\begin{aligned} \frac{\delta^2 W(C)}{\delta x_\mu(s) \delta x_\mu(s')} &= \left\langle \text{Tr } P(F_{\mu\nu}(x(s'))F_{\mu\lambda}(x(s')) \exp\left(\int A_\mu dx^\mu\right) \right\rangle \dot{x}_\nu(s) \dot{x}_\lambda(s') \\ &+ \delta(s-s') \dot{x}_\nu(s) \left\langle \text{Tr } P\left(\nabla_\mu F_{\mu\nu}(x(s)) \exp\left(\int A_\mu dx^\mu\right)\right) \right\rangle \end{aligned} \quad (7.34)$$

Let us now introduce a local derivative:

$$\frac{\partial^2}{\partial x^2(s)} \stackrel{\text{def}}{=} \lim_{\epsilon \rightarrow 0} \int_{-\epsilon}^{\epsilon} dt \frac{\delta^2}{\delta x_\mu(s+t/2) \delta x_\mu(s-t/2)} \quad (7.35)$$

which picks up  $\delta$ -like terms in the second functional derivative. If the quantum theory is regularized, then the first term in (7.34) does not contain singularities as  $s \rightarrow s'$ . From this we infer that:

$$\frac{\partial^2 \mathcal{W}(C)}{\partial x^2(s)} = \left\langle \text{Tr} \left( P \left( \nabla_\mu F_{\mu\nu}(x(s)) \exp \left( \int A_\mu \dot{x}_\mu ds \right) \right) \dot{x}_\nu \right) \right\rangle \quad (7.36)$$

For classical fields the r.h.s. of (7.36) would be zero. In the quantum case it is finite and calculable. To find it we consider the following (non-invariant) functional integral:

$$\int \mathcal{D}A_\mu \exp \left( \frac{1}{2e_0^2} \int \text{tr} F_{\mu\nu}^2 d^4x \right) \left( P \exp \left( \int_0^{2\pi} A_\mu dx^\mu \right) \right) \quad (7.37)$$

and change variables by  $A_\mu \rightarrow A_\mu + \delta A_\mu$ . There are two effects coming from this change, which must cancel each other. The first is the variation of the action, proportional to  $\nabla_\mu F_{\mu\nu}$ , while the second is the variation of the phase factor. The cancellation condition gives the identities:

$$\begin{aligned} \left\langle -\frac{1}{e_0^2} \nabla_\mu F_{\mu\nu}^a(z) \Psi(x, x) \right\rangle &= \oint_C dy_\nu \delta(z-y) \langle \Psi(x, y) \lambda^a \Psi(y, x) \rangle \\ \left\langle -\frac{1}{e_0^2} \text{Tr}(\nabla_\mu F_{\mu\nu}(z) \Psi(x, x)) \right\rangle &= \oint_C dy_\nu \delta(z-y) \langle \text{Tr} \lambda^a \Psi(x, y) \lambda^a \Psi(y, x) \rangle \end{aligned} \quad (7.38)$$

Here the index  $a$  labels the generators  $\lambda^a$  of our Lie algebra and  $\Psi(x, y)$  is a piece of the phase factor for the part of the contour  $C$  connecting  $x$  and  $y(x, y \in C)$ . For the  $SU(N)$  group, (7.38) is further simplified through the use of the identity:

$$\sum_a \lambda_{\alpha\beta}^a \lambda_{\gamma\delta}^a = \delta_{\alpha\delta} \delta_{\beta\gamma} - \frac{1}{N} \delta_{\alpha\beta} \delta_{\gamma\delta} \quad (7.39)$$

We have:

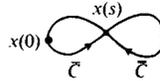
$$\begin{aligned} &\left\langle -\frac{1}{e_0^2} \text{Tr}(\nabla_\mu F_{\mu\nu}(z) \Psi(x, x)) \right\rangle \\ &= \oint_C dy_\nu \delta(z-y) \\ &\quad \times \left\{ \langle \text{Tr} \Psi(x, y) \text{Tr} \Psi(y, x) \rangle - \frac{1}{N} \langle \text{Tr} \Psi(x, y) \Psi(y, x) \rangle \right\} \end{aligned} \quad (7.40)$$

Substitution of (7.40) into (7.36) gives

$$\frac{\partial^2 W(C_{xx})}{\partial x^2(s)} = -e_0^2 \oint \delta(x(s) - y) \left\{ W(\bar{C})W(C) - \frac{1}{N} W(\bar{C}, C) \right\} \dot{x}_\mu(s) dy_\mu \quad (7.41)$$

which is the first equation in the Schwinger-like chain.

This equation requires some explanations. First of all,  $C$ ,  $\bar{C}$  and  $\bar{\bar{C}}$  are defined as following. The loop  $C_{xx}$  starts and terminates at the point  $x = x(0)$ . If this loop does not have a self-intersection at the point  $x(s)$  then the r.h.s. of (7.41) will be zero because of the  $\delta$ -function. If it does, then the point  $y$  splits the loop onto two closed contours  $\bar{C}$  and  $\bar{\bar{C}}$ :



$$\quad (7.42)$$

A very important point, not to be forgotten, is that eq. (7.41) was derived in an unrenormalized but regularized version of the theory. So the  $\delta$ -function in (7.41) must be somehow smeared, and  $\epsilon$ , entering in (7.35) must be taken much less than the smearing length. The equation itself corresponds to a particular cut-off of the gauge theory. It may be untrue for a different cut-off. All this is quite unpleasant, It would be much nicer to have an equation for the finite, renormalized  $W(C)$ .

Unfortunately this equation is not known, and we are unable at present to remove the scaffolding (the regularization) from our construction.

Nevertheless, equation (7.41) is meaningful. It reproduces perturbation theory and in the large  $N$  limit, presents a closed equation which sums up all planar Feynman diagrams (see the next chapter). Here we shall show how the first order of perturbation theory for  $W(C)$  arises from equation (7.41).

For this, let us consider the following ansatz for  $W(C)$ :

$$W(C) = 1 + \iint_{s_1 < s_2} \Gamma_{\mu\nu}(x_1, x_2) dx_1^\mu dx_2^\nu + \iiint_{s_1 < s_2 < s_3} \Gamma_{\mu\nu\lambda}(x_1, x_2, x_3) dx_1^\mu dx_2^\nu dx_3^\lambda + \dots \quad (7.43)$$

This ansatz is true for any  $W(C)$  which can be represented as an average of the loop factor (7.1), with:

$$\Gamma_{\mu\nu\lambda\dots}(x_1, x_2, x_3, \dots) = \langle \text{Tr } A_\mu(x_1) A_\nu(x_2) A_\lambda(x_3) \dots \rangle \quad (7.44)$$

In fact, it can be shown that this ansatz is true for more or less any functional  $W(C)$ , being an analogue of the Taylor expansion in loop space. To see how the perturbation theory arises let us find  $\partial^2 W_2(C)/\partial x^2(s)$  where  $W_2$  is the second term in (7.43). We have:

$$\begin{aligned} \frac{\delta W_2}{\delta x_\mu(s)} &= \int ds_1 ds_2 \dot{x}_\rho(s_2) \{ \dot{x}_\lambda(s_1) \partial_{1,\mu} \Gamma_{\lambda\rho}(x_1(s_1), x_2(s_2)) \delta(s-s_1) \\ &\quad + \delta_{\lambda\mu} \dot{\delta}(s_1-s) \Gamma_{\lambda\rho}(x(s_1), x(s_2)) \} \\ &= \dot{x}_\lambda(s) \int ds_2 \dot{x}_\rho(s_2) (\partial_{1,\mu} \Gamma_{\lambda\rho}(x_1(s), x_2(s_2)) \\ &\quad - \partial_{1,\lambda} \Gamma_{\mu\rho}(x_1(s), x(s_2))) \end{aligned} \quad (7.45)$$

Next we have to take the second derivative  $\delta/\delta x_\mu(s')$  and to pick out the terms containing  $\delta(s-s')$ . That gives:

$$\begin{aligned} \frac{\delta^2 W_2}{\delta x_\mu(s) \delta x_\mu(s')} &= \int (\partial_1^2 \Gamma_{\lambda\rho} - \partial_{1,\lambda} \partial_{1,\sigma} \Gamma_{\sigma\rho}) \delta(s-s') \dot{x}_\lambda(s) \dot{x}_\rho(s_2) ds_2^- \\ &\quad + \text{terms, which do not contain } \delta(s-s'). \end{aligned} \quad (7.46)$$

Thus:

$$\frac{\partial^2 W_2}{\partial x^2(s)} = \int \partial_1^2 \Gamma_{\lambda\rho}(x(s), x(t)) - \partial_{1,\lambda} \partial_{1,\sigma} \Gamma_{\sigma\rho}(x(s), x(t)) \dot{x}_\lambda(s) \dot{x}_\rho(t) dt \quad (7.47)$$

We must substitute this result into the r.h.s. of (7.41) replacing  $W(C)$  in the l.h.s. by 1. This gives the equation for  $\Gamma_{\mu\nu}$ :

$$\begin{aligned} \partial_1^2 \Gamma_{\lambda\rho}(x, x') - \partial_{1,\lambda} \partial_{1,\sigma} \Gamma_{\sigma\rho}(x, x') \\ = -e_0^2 \delta_{\lambda\rho} \delta(x-x') + \frac{\partial}{\partial x'_\rho} \phi_\lambda(x, x') \end{aligned} \quad (7.48)$$

Here  $\phi$  is an arbitrary function which is needed to make (7.48) solvable, and which it is possible to add, since

$$\oint dx'_\rho \frac{\partial}{\partial x'_\rho} \phi_\lambda = 0$$

The solution of (7.48) has the form:

$$\Gamma_{\lambda\rho}(x, x') = \frac{e_0^2}{4\pi^2} \frac{\delta_{\lambda\rho}}{(x-x')^2} + \frac{\partial h_\rho(x, x')}{\partial x_\lambda} \quad (7.49)$$

where  $h_p$  is arbitrary. This arbitrariness does not appear in  $W_2(C)$  itself which has the form:

$$W_2(C) = \frac{e_0^2}{8\pi^2} \oint \frac{dx_\mu dx'_\mu}{|x - x'|^2} = \text{---} \circlearrowleft_c \quad (7.50)$$

The formula (7.50) is precisely the first nontrivial perturbative contribution to  $W(C)$ .

In order to obtain higher orders this contribution has to be substituted into the r.h.s. of (7.41). After some complicated combinatorics (which can be found in A. A. Migdal (1977)) one finds, order by order, the standard Feynman diagrams contributing to  $W(C)$ . It is remarkable that no *a priori* gauge fixing is needed, since (7.41) is an equation for gauge invariant quantities. Ghost diagrams in the higher orders appear automatically in the process of iteration (A. A. Migdal (1977)).

So we conclude that in spite of some dubious operations performed in deriving (7.41), namely separating from the complete derivative  $\delta^2/\delta x_\mu(s)\delta x_\mu(s')$  the part containing  $\delta(s - s')$  only, which we called  $\partial^2/\partial x^2(s)$ , we did not lose any information. It appears that in the frame of the general ansatz (7.43), knowledge of  $\partial^2/\partial x^2(s)$  is sufficient for the reconstruction of  $W(C)$ . Again, we have to warn the reader that this is true only in an unrenormalized but regularized version of the theory in which the coefficients in (7.43) do not have singularities at coincident points. Renormalized quantities, while being finite, are singular in these cases and that makes our definition of  $\partial^2/\partial x^2(s)$  inoperative. I believe that there should exist some kind of renormalized equations but they have not yet been found.

What is the use of loop equations, like (7.41)? Their main purpose is to provide us with a description of gauge fields in terms of their natural elementary excitations. We have seen in Chapter 3, that in the confinement phase those excitations are closed strings. In general these strings interact with each other. In some cases, in particular in the large  $N$  limit, they must become free, as will be shown in Chapter 8.

The loop equation (7.41) is aimed at choosing from among possible free string theories the one describing or being described by the gauge fields. This task does not yet have a final solution, although considerable progress has been achieved.

Our next step will be to consider the large  $N$  approximation and to prove that in this limit particles in chiral theories and strings in gauge theories become free.



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

---

## CHAPTER 8

# The Large $N$ Expansion

---

---

DOI: 10.1201/9780203755082-8

The number  $N$  is the only free parameter both in chiral and gauge theories, because in the continuum limit the coupling constant disappears through dimensional transmutation (remember that the strong coupling expansion referred to the lattice theory and does not give a continuum theory in any finite order). It is very tempting therefore to try to make a  $1/N$  expansion. This will be done in this chapter. As before, our main intention will be to improve our qualitative understanding of dynamics by means of this approximation.

### 8.1 $O(N)$ $\sigma$ -Model

This case is the simplest one. Let us consider the functional integral in the form:

$$\begin{aligned} Z &= \int \mathcal{D}\mathbf{n}(x) \left( \prod_x \delta(\mathbf{n}(x)^2 - 1) \right) \exp\left(-\frac{1}{2g_0^2} \int (\partial_\mu \mathbf{n})^2 d^2x\right) \\ &= \int_{C-i\infty}^{C+i\infty} \mathcal{D}\lambda(x) \int \mathcal{D}\mathbf{n}(x) \exp\left(-\frac{1}{2g_0^2} \int ((\partial_\mu \mathbf{n})^2 + \lambda(\mathbf{n}^2 - 1)) d^2x\right) \end{aligned} \quad (8.1)$$

Here we have replaced the functional  $\delta$ -function by its integral representation, involving the Lagrange multiplier  $\lambda(x)$ :

$$\begin{aligned} \prod_x \delta(\mathbf{n}(x)^2 - 1) &= \prod_x \int_{C-i\infty}^{C+i\infty} d\lambda(x) \exp\{\lambda(x)(\mathbf{n}^2(x) - 1)\} \\ &= \int \mathcal{D}\lambda(x) \exp\left\{ \int \lambda(x)(\mathbf{n}^2 - 1) d^2x \right\} \end{aligned} \quad (8.2)$$

(The functions  $\lambda(x)$  in (8.2) and (8.1) differ by an irrelevant factor  $1/2g_0^2$ .) We have called the field  $\lambda(x)$  a ‘‘Lagrange multiplier’’ since in the classical limit, when we have to minimize the  $O(N)$  action with the constraint  $\mathbf{n}^2 = 1$ , it is just that.

The representation (8.1) is convenient, since the integral over  $\mathbf{n}$  is Gaussian and can be performed in a standard fashion. We have:

$$Z = \int_{C-i\infty}^{C+i\infty} \mathcal{D}\lambda(x) \exp \left\{ \frac{1}{2g_0^2} \int \lambda(x) d^2x - \frac{N}{2} \log \det \| -\partial^2 + \lambda(x) \| \right\} \quad (8.3)$$

where  $N$  is the number of components of the field  $\mathbf{n}$ . The factor  $N$  appears in (8.3) since each component enters independently in (8.1). The second term in (8.3) has the Feynman graph representation:

$$\log \det \| -\partial^2 + \lambda \| = \bigcirc + \begin{array}{c} \text{---} \\ | \\ \bigcirc \end{array} + \begin{array}{c} \text{---} \\ | \\ \bigcirc \\ | \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ | \\ \bigcirc \\ | \\ \text{---} \\ | \\ \text{---} \end{array} + \dots, \quad (8.4)$$

in which the wavy line corresponds to the external  $\lambda$ -field and the propagators of the  $\mathbf{n}$ -field, represented by solid lines, are  $1/p^2$ . This expansion is formal, since for  $\mathcal{D} = 2$  it is both infrared and ultraviolet divergent. The ultraviolet divergence has to be cut off by hand (or, better to say, by a lattice), while the infrared one will be taken care of by the theory itself.

If we take  $N$  to be large, then since it enters the exponent (8.3), we have reasons to expect that the saddle point approximation will be applicable. This is indeed the case. To show this let us first compute the variation of the action in (8.3) with respect to  $\lambda$ . We get:

$$\begin{aligned} \frac{1}{2g_0^2} &= \frac{N}{2} \frac{\delta}{\delta \lambda(x)} \log \det \| -\partial^2 + \lambda(x) \| \\ &= \frac{N}{2} G(x, x; \lambda) \end{aligned} \quad (8.5)$$

Here we have introduced the Green function:

$$G(x, x') = \langle x' | (-\partial^2 + \lambda)^{-1} | x \rangle$$

The last equality follows from the relation

$$\delta \log \det A = \delta \text{Tr} \log A = \text{Tr} A^{-1} \delta A \quad (8.6)$$

or else can be checked from (8.4). The meaning of equation (8.5) becomes transparent if we notice that according to (8.1):

$$\begin{aligned} \langle n_i(x)n_j(y) \rangle &= Z^{-1} \int \mathcal{D}\lambda \int \mathcal{D}\mathbf{n} \exp\left(-\frac{1}{2g_0^2} \int \{(\partial\mathbf{n})^2 + \lambda(\mathbf{n}^2 - 1)\} d^2x\right) \\ &\quad \times n_i(x)n_j(y) \\ &= g_0^2 \delta_{ij} \frac{\int \mathcal{D}\lambda e^W G(x, y; \lambda)}{\int \mathcal{D}\lambda e^W} \quad (8.7) \\ W &= \frac{1}{2g_0^2} \int \lambda(x) d^2x - \frac{N}{2} \text{Tr}_x \log(-\partial^2 + \lambda) \end{aligned}$$

If the  $\lambda$ -integral is to be approximated by the saddle point we obtain:

$$\langle n_i(x)n_j(y) \rangle = g_0^2 \delta_{ij} G(x, y; \lambda_*(z)) \quad (8.8)$$

where  $\lambda_*$  is the saddle point value of  $\lambda$ . From here we see that (8.5) is nothing but the condition  $\langle \mathbf{n}^2(x) \rangle = 1$ .

Let us now solve the equation (8.5). If we guess that its solution is homogeneous in  $x$ -space, we can verify this conjecture by passing to the momentum representation in (8.5):

$$\begin{aligned} G(x, x'; \lambda) &= \int \frac{d^{\mathcal{D}}p}{(2\pi)^{\mathcal{D}}} \frac{e^{ip(x-x')}}{p^2 + \lambda} \quad (8.9) \\ 1 &= N g_0^2 G(x, x; \lambda) = N g_0^2 \int \frac{d^{\mathcal{D}}p}{(2\pi)^{\mathcal{D}}} \frac{1}{p^2 + \lambda} \end{aligned}$$

As is to be expected, this equation reflects a qualitative difference between  $\mathcal{D} = 2$  and  $\mathcal{D} > 2$ . For  $\mathcal{D} = 2$  we have:

$$\begin{aligned} 1 &= \frac{N g_0^2}{4\pi} \log \frac{\Lambda^2}{\lambda} \quad (8.10) \\ \lambda &= \Lambda^2 \exp\left(-\frac{4\pi}{N g_0^2}\right) \end{aligned}$$

(where  $\Lambda$  is the momentum cut-off and  $\sqrt{\lambda}$  is a physical mass or inverse correlation length as is seen from (8.9) and (8.8)). This formula agrees with (2.49), the only difference being that  $N - 2$  in the "exact" formula (2.49) is replaced by  $N$  in (8.10). So, we see that for all values of  $g_0^2$  we have the same phase, with a finite correlation length. It would be quite easy to repeat the derivation in a lattice version of the theory and to see

explicitly how the strong coupling expansion connects without a phase transition to the weak coupling region, giving (8.10).

For  $\mathcal{D} > 2$ , the situation is different. We have a critical coupling  $g_{0,\text{cr}}^2$ , defined by:

$$1 = Ng_{0,\text{cr}}^2 \int \frac{d^{\mathcal{D}}p}{(2\pi)^{\mathcal{D}}p^2} \sim Ng_{0,\text{cr}}^2 \frac{\Lambda^{\mathcal{D}-2}}{\mathcal{D}-2} \quad (8.11)$$

If  $g_0^2 > g_{0,\text{cr}}^2$  then the equation (8.9) has a solution with  $\lambda > 0$  and we are in the strong coupling phase. The theory has a continuum limit as  $g_0^2 \rightarrow g_{0,\text{cr}}^2 + 0$  as can be seen by rewriting (8.9) using (8.11) as:

$$\begin{aligned} 1 &= Ng_0^2 \int \frac{d^{\mathcal{D}}p}{(2\pi)^{\mathcal{D}}p^2} - Ng_0^2 \int \frac{d^{\mathcal{D}}p}{(2\pi)^{\mathcal{D}}} \left( \frac{1}{p^2} - \frac{1}{p^2 + \lambda} \right) \\ &= \frac{g_0^2}{g_{0,\text{cr}}^2} - Ng_0^2 \int \frac{d^{\mathcal{D}}p}{(2\pi)^{\mathcal{D}}} \frac{\lambda}{p^2(p^2 + \lambda)} \end{aligned} \quad (8.12)$$

For  $2 < \mathcal{D} < 4$  the integral in (8.12) is convergent and proportional to  $\lambda^{(\mathcal{D}/2)-1}$ . Therefore:

$$m^2 \equiv \lambda = \text{const } \Lambda^2 \left( \frac{g_0^2 - g_{0,\text{cr}}^2}{g_{0,\text{cr}}^2} \right)^{2/(\mathcal{D}-2)} \quad (8.13)$$

We achieve the continuum limit by taking  $g_0^2 \rightarrow g_{0,\text{cr}}^2$  and  $\Lambda \rightarrow \infty$  so that  $m^2 = \lambda$  remains fixed. We see also, that while the value of  $g_{0,\text{cr}}^2$  depends on the cut-off and is not at all universal the ‘‘critical exponent’’  $2/(\mathcal{D} - 2)$  in (8.13) is independent of the short distance dynamics. As we explained in Chapter 1 this is quite a general situation.

If we take  $g_0^2 < g_{0,\text{cr}}^2$  something goes wrong with equation (8.9)—it does not have solutions any more. That means that we have lost the saddle point. To understand what happens in this case we have to recall that originally we had an integral over  $\lambda$ . If we introduce the Fourier expansion:

$$\lambda(x) = \sum_q \lambda_q e^{iqx} \quad (8.14)$$

then

$$\mathcal{D}\lambda(x) = \left( \prod_{q \neq 0} d\lambda_q \right) \cdot d\lambda_0 \quad (8.15)$$

In the large  $N$  limit all  $q \neq 0$  integrations are irrelevant (we shall prove this later in this Section) and the integral over  $\lambda$  has the form:

$$Z = \int_{C-i\infty}^{C+i\infty} d\lambda_0 \exp \left\{ V \left( \frac{\lambda_0}{2g_0^2} - \frac{N}{2} \int_{|p| < \Lambda} \frac{d^{\mathcal{D}}p}{(2\pi)^{\mathcal{D}}} \log(p^2 + \lambda_0) \right) \right\} \quad (8.16)$$

which is obtained by the substitution  $\lambda(x) = \lambda_0 = \text{const.}$  into (8.3);  $V$  here is the volume of the system. In the complex plane of the  $\lambda_0$  variable we have a cut from the logarithm in (8.16) which goes from  $-\infty$  to zero, and a saddle point when  $g_0^2 > g_{0,\text{cr}}^2$ . For  $g_0^2 < g_{0,\text{cr}}^2$  this saddle point disappears under the cut. Therefore, in order to obtain the dominant contribution to the integral, our contour for  $g_0^2 > g_{0,\text{cr}}^2$  must be taken to pass through the saddle point and the integrand is strongly concentrated near this saddle point. This situation we have analysed above.

For  $g_0^2 < g_{0,\text{cr}}^2$  the contour can be deformed so that it goes around the cut. In the limit of infinite volume the dominant contribution comes from the origin of the cut at  $\lambda_0 = 0$ . In order to estimate the most important values of  $\lambda_0$ , let us notice that the singular part of the integral in (8.16) is given by:

$$\int d^{\mathcal{D}}p \log(p^2 + \lambda) = \text{const. } \lambda^{\mathcal{D}/2} + \text{regular terms} \quad (8.17)$$

From this we conclude that the essential  $\lambda_0$  is defined by:

$$\begin{aligned} NV \lambda_0^{\mathcal{D}/2} &\sim 1 \\ \lambda_0 &\sim (NV)^{-2/\mathcal{D}} \rightarrow 0 \end{aligned} \quad (8.18)$$

The conclusion we reach is that for  $g_0^2 < g_{0,\text{cr}}^2$  the quantity  $\lambda_0$  has to be zero in the infinite volume limit, and we have massless particles in our system. That implies that the symmetry has been broken as we go to  $g_0^2 < g_{0,\text{cr}}^2$  and these massless particles are Goldstone's. We postpone direct verification of this fact and proceed to consider corrections to the naive saddle point picture.

We have to take into account that the field  $\lambda(x)$  acquires a vacuum expectation value in the phase with *unbroken*  $O(N)$  symmetry (the only existing one for  $\mathcal{D} \leq 2$ ). In the leading approximation it is given by (8.10). Successive approximations will change this value by an amount  $\sim 1/N$ . In order to develop a general formalism we must not fix  $\langle \lambda(x) \rangle = \lambda$  from the beginning, but keep it arbitrary while performing integrations over  $\{\lambda_{q \neq 0}\}$ . After that we shall obtain an effective action



The second term in (8.21) comes from the fluctuations of the Lagrange multiplier.

At this point we encounter some trouble. From (8.19) (8.20) we see that the propagator for the  $\beta$ -field is

$$\mathcal{D}(q^2) = \frac{1}{\Pi(q^2)} \xrightarrow{q \rightarrow \infty} \frac{2\pi q^2}{\log(q^2/\mu^2)} \quad (8.22)$$

This rising propagator introduces some power-like divergences in Feynman graphs and this is clearly unphysical. Let us locate the origin of this trouble and explain how to avoid it. First of all we see that these divergences persist even in the  $\mathcal{D} = 1$  system which is the quantum mechanics of the  $\mathbf{n}$ -field. Quantum mechanics is a finite theory, hence we have done something wrong.

To locate the difficulty let us consider instead of the  $\mathbf{n}$ -field with its rigid constraint  $\mathbf{n}^2 = 1$  a theory with a Lagrangian:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \mathbf{n})^2 + \frac{g}{4}(\mathbf{n}^2 - 1)^2 \quad (8.23)$$

which in the limit  $g = \infty$  goes over to the theory of the  $\mathbf{n}$ -field. If we introduce a  $\lambda$ -field, we replace (8.23) by the Lagrangian:

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \mathbf{n})^2 + \frac{1}{2}\lambda(\mathbf{n}^2 - 1) - \frac{\lambda^2}{2g} \quad (8.24)$$

and we shall have the same large  $N$  expansion as above except that  $\mathcal{D}(q^2)$  will be replaced by:

$$\tilde{\mathcal{D}}(q^2) = \frac{1}{\Pi(q^2) + (2/Ng)} \quad (8.25)$$

Since  $\Pi(q^2) \sim \log(q^2/\mu^2)/q^2$ , this new propagator does not rise at  $q^2 \rightarrow \infty$  and we shall have no power divergences. In particular, quantum mechanics will be finite, as it should be.

In order to understand the origin of the power divergences, let us stick to the case of quantum mechanics ( $\mathcal{D} = 1$ ). The potential:

$$V = \frac{g}{4}(\mathbf{n}^2 - 1)^2 \quad (8.26)$$

has the tendency, as  $g \rightarrow \infty$ , to confine our particle to the surface of the sphere. However, owing to the uncertainty principle, attempting such confinement leads to large kinetic energy. Therefore, the energy levels of the particle are shifted to infinity; that is, we have zero point energy of the order of  $\sqrt{g}$  and radial excitations of the same order. But, on top of

this large vacuum energy we have finite angular excitations, labelled by angular momentum. In the limit  $g \rightarrow \infty$  we have to subtract an infinite constant from the ground state energy, and after that the theory will be finite.

Returning to our problem we are led to expect that power divergences caused by (8.22) will be removed if we renormalize the mass of elementary excitations. A convenient procedure is the following. Let us introduce a quantity:

$$m^2 = G^{-1}(p=0) \quad (8.27)$$

and self-energy part:

$$-\Sigma(p) = \text{diagram 1} + \text{diagram 2} + \dots \quad (8.28)$$

The Green function is defined by:

$$G(p) = \frac{1}{p^2 + m^2 + \Sigma(p^2) - \Sigma(0)} \quad (8.29)$$

While  $\Sigma(p^2)$  contains power divergences, it is easy to see that  $\Sigma(p^2) - \Sigma(0)$  does not. Let us estimate as an example the first term in (8.28) (we consider  $p \gg m$ ):

$$\begin{aligned} \Sigma(p) - \Sigma(0) &= \frac{2}{N} \int \frac{d^2k}{(2\pi)^2 \Pi(k^2)} \left( \frac{1}{(p-k)^2} - \frac{1}{k^2} \right) \\ &\approx \frac{2}{N} \int_{\rho < |k| < \Lambda} \frac{k^2 d^2k}{2\pi \log(k^2/m^2)} \left( \frac{4(p \cdot k)^2}{k^6} - \frac{p^2}{k^4} \right) + \text{finite terms} \\ &= \frac{p^2}{N} \log\{(\log \Lambda^2/m^2)/(-\log m^2/p^2)\} + \text{finite part} \end{aligned} \quad (8.30)$$

We have obtained a strange double-log divergence. In order to interpret it, let us recall that the correlation function  $\langle n(0)n(R) \rangle$  must not be finite after charge renormalization only, since, according to Chapter 2, we have to renormalize the  $n$ -field itself. In this chapter we have obtained the result:

$$\langle n(0)n(R) \rangle = \left( 1 - \frac{N-2}{2\pi} g_0^2 \log R/a \right)^{(N-1)/(N-2)} \quad (8.31)$$

while in the leading  $1/N$  approximation:

$$\begin{aligned} \langle \mathbf{n}(0)\mathbf{n}(R) \rangle &= Ng_0^2 \int \frac{d^2p}{(2\pi)^2} \frac{e^{i\mathbf{p}R}}{p^2 + m^2} \\ &= \frac{Ng_0^2}{2\pi} \log \frac{1}{mR} \end{aligned} \quad (8.32)$$

Let us compare these expressions. Since

$$ma = \exp\left(-\frac{2\pi}{Ng_0^2}\right) \quad (8.33)$$

we have:

$$\frac{Ng_0^2}{2\pi} \log \frac{1}{mR} = 1 - \frac{N}{2\pi} g_0^2 \log \frac{R}{a} \quad (8.34)$$

We see that in the leading approximation (8.32) matches with (8.31), owing to the fact that the “anomalous dimension”  $(N-1)/(N-2) \rightarrow 1$ . As we expand (8.31) in  $1/N$ , in the next order we shall get:

$$\begin{aligned} \langle \mathbf{n}(0)\mathbf{n}(R) \rangle &= \left(1 - \frac{N-2}{2\pi} g_0^2 \log \frac{R}{a}\right) \left(1 - \frac{N-2}{2\pi} g_0^2 \log \frac{R}{a}\right)^{1/N-2} \\ &\simeq \left(1 - \frac{N-2}{2\pi} g_0^2 \log \frac{R}{a}\right) \left(1 + \frac{1}{N} \log\left(\frac{\log(1/mR)}{\log(1/ma)}\right) + \dots\right) \end{aligned} \quad (8.35)$$

Equation (8.35) demonstrates how double logarithms arise in the  $1/N$  expansion of the “exact” formula (8.31); and they are precisely those obtained in (8.30).

Finally, all that we have said implies that after mass renormalization (8.29) (which is essentially charge renormalization), and renormalization of the  $\mathbf{n}$ -field, all terms in the  $1/N$  expansion will be finite. This can be proved by standard methods, used in field theory.

It might have seemed strange that we needed to renormalize the field  $\mathbf{n}$ , the scale of which is defined by  $\mathbf{n}^2 = 1$ . As we see from (8.31) and (8.32) this happens because in  $\langle \mathbf{n}(0)\mathbf{n}(R) \rangle$  the constraint is satisfied only after we take  $R \lesssim a$  ( $a$ -being the lattice size) and finite isotropic correlations exists only at  $R \gg a$ . Therefore, the constraint could not be directly applied in the continuum limit. As we see from (8.32), the correlation of the original  $\mathbf{n}$ -field in this limit contains the cut-off dependent factor

$$\frac{Ng_0^2}{2\pi} \sim \frac{1}{\log(1/ma)}$$

The other way to express the same thing is to notice that after integrating out high frequencies of our field we obtain an effective action for block spins, which are normalized differently from the original ones.

To sum up, after proper renormalizations, the  $1/N$  expansion is perfectly defined. Its main advantage is that it explicitly describes continuum theory in the correct phase. In principle, there could be a phase transition as we decrease  $N$ , but in the case of the  $\mathbf{n}$ -field this does not happen. In the large  $N$  limit, the  $\mathcal{D} = 2$   $\mathbf{n}$ -field appeared to be a theory of free massive particles. The scattering amplitudes, which are easily computed (see the next chapter), are of the order  $1/N$ .

Unfortunately, this simple picture is not true, as far as principal chiral fields and  $\mathcal{D} = 4$  gauge fields are concerned. The first indication of the trouble comes from the formula (2.35) for the correlation function of the chiral field. In the limit  $N \rightarrow \infty$  it reads

$$\begin{aligned} \mathcal{D}(R) &= \left(1 - \frac{Ne_0^2}{4\pi} \log \frac{R}{a}\right)^{2(N^2-1)/N^2+1} \\ &\underset{N \rightarrow \infty}{\simeq} \left(1 - \frac{Ne_0^2}{4\pi} \log \frac{R}{a}\right)^3 \sim \log^3 \frac{1}{mR}, \quad mR \ll 1 \end{aligned} \quad (8.36)$$

while the free Green function should behave as  $\log(1/mR)$ . We see that in contrast with the  $\mathbf{n}$ -field, the field  $g(x)$  does not become free in the large  $N$  limit. In the next section we shall try to understand why.

## 8.2 The Principal Chiral Field For $SU(N)$

Let us start by attempting the previous trick with the Lagrange multiplier. In this case our variables are complex matrices  $g_{ab}(x)$ . The partition function can be written as:

$$\begin{aligned} Z &= \int \prod_{a,b} \mathcal{D}g_{ab} \mathcal{D}g_{a,b}^* \prod_{a,b} \delta\left(\sum_c g_{ac}g_{bc}^* - \delta_{ab}\right) \\ &\quad \times \exp\left\{-\frac{1}{e_0^2} \sum_{a,b} \int \partial_\mu g_{ab} \partial_\mu g_{ab}^* d^2x\right\} \\ &= \prod_{\substack{a < b \\ c < +i\infty}}^{c+i\infty} \int \mathcal{D}\lambda_{ab} \mathcal{D}\lambda_{ab}^* \exp\left(\frac{1}{e_0^2} \int_a \sum \lambda_{aa} d^2x\right) \int \mathcal{D}g_{ab} \mathcal{D}g_{ab}^* \\ &\quad \times \exp\left\{-\frac{1}{e_0^2} \sum_{a,b} \int |\partial_\mu g_{ab}|^2 d^2x - \frac{1}{e_0^2} \int \lambda_{ab} g_{ac} g_{bc}^* d^2x\right\} \end{aligned} \quad (8.37)$$

Here we have introduced a Lagrange multiplier, which is in this case a Hermitian matrix (because  $gg^+$  is Hermitian). It ensures that in (8.37) we integrate over unitary matrices only. So, strictly speaking, the model (8.37) describes the  $U(N)$  and not the  $SU(N)$  group. However, it is easy to see that in the product  $U(N) = U(1) \otimes SU(N)$  the field corresponding to the  $U(1)$  factor decouples and can be ignored.

As before, the Gaussian integral in (8.37) can be represented as a functional determinant. We have:

$$\begin{aligned} Z &= \int \prod_{a \leq b} d^2 \lambda_{ab} \exp \left\{ \frac{1}{e_0^2} \sum_a \lambda_{aa} - N \operatorname{Tr} \log(-\partial^2 \delta_{ab} + \lambda_{ab}) \right\} \\ &\equiv \int \prod_{a \leq b} \mathcal{D} \lambda_{ab} e^{W(\lambda)} \end{aligned} \tag{8.38}$$

( $N$  appears here owing to the fact that we had a sum over  $c$  in (8.37)).

It now seems very natural to follow the same strategy as in the case of the  $n$ -field. Here, however, it leads to trouble. Let us demonstrate this fact and discuss possible ways out (at present the only way to overcome this problem is to expand its exact solution in  $1/N$ —not very practical for generalizations).

It is natural to expect that  $\lambda_{ab}$  acquires a nonzero expectation value:

$$\langle \lambda_{ab} \rangle = \mu^2 \delta_{ab} \tag{8.39}$$

Therefore we expand:

$$\begin{aligned} \lambda_{ab} &= \mu^2 \delta_{ab} + \frac{i}{\sqrt{N}} v_{ab} \quad \sum_a \int v_{aa} d^2x = 0 \\ \frac{W}{V} &= \frac{1}{e_0^2} \sum_a \lambda_{aa} - N \operatorname{Tr} \log(-\partial^2 \delta_{ab} + \lambda_{ab}) \\ &= N \mu^2 / e_0^2 - N \operatorname{Tr} \log[(-\partial^2 + \mu^2) \delta_{ab} + i(N)^{-1/2} v_{ab}] \\ &= N \left( \frac{\mu^2}{e_0^2} - N \operatorname{Tr} \log(-\partial^2 + \mu^2) \right) - \frac{1}{2} \sum_q \Pi_{ab|cd}(q) v_{ab}(q) v_{cd}(-q) \\ &\quad + \frac{N^{-1/2}}{3} \sum_{\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3 = 0} \Gamma_{ab|cd|ef} v_{ab}(k_1) v_{cd}(k_2) v_{ef}(k_3) + \dots \end{aligned} \tag{8.40}$$

The value of  $\mu^2$  has to be determined from the minimum of the effective action or else from the unitarity condition

$$\langle \operatorname{Tr}(g^+ g) \rangle = N$$

Up to now everything has mimicked precisely the consideration of Section 8.1. Now we come to the crucial difference. Namely, in spite of the  $N^{-1/2}$ -factor before higher powers in  $v$ , the nonlinearity in (8.40) is absolutely important, because this smallness is compensated by summation over isotopic indices in Feynman diagrams. Let us prove this and simultaneously separate the relevant set of diagrams. According to (8.40):

$$\begin{aligned} \Pi_{ab|cd}(q) &= \int \frac{d^2k}{(2\pi)^2} \frac{1}{(k^2 + \mu^2)((k+q)^2 + \mu^2)} \\ &\quad \times \frac{1}{2}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}) \\ &\equiv \pi(q^2) \frac{1}{2}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}) \end{aligned} \tag{8.41}$$

(This follows from the matrix equation:

$$\begin{aligned} &\text{Tr} \log((-\hat{c}^2 + \mu^2)I + v) \\ &= \text{Tr} \log(-\hat{c}^2 + \mu^2) + \text{Tr} \log\left(I + \frac{1}{\mu^2 - \hat{c}^2} v\right) \\ &= \dots - \frac{1}{2} \text{Tr}\left(v \frac{1}{\mu^2 - \hat{c}^2} v \frac{1}{\mu^2 - \hat{c}^2}\right) + \dots \end{aligned} \tag{8.42}$$

where Tr is understood both in coordinate and isotopic indices).

The propagator for the  $v$ -field is given by:

$$\langle v_{ab}(q)v_{cd}(-q) \rangle = \frac{1}{2}(\delta_{ac}\delta_{bd} + \delta_{ad}\delta_{bc}) \frac{1}{\pi(q^2)} \tag{8.43}$$

The isotopic structure in (8.43) is conveniently represented by the picture

$$\langle v_{ab}v_{cd} \rangle = \frac{b-d}{a-c} + \frac{b}{a} \times \frac{d}{c} \tag{8.44}$$

where each line corresponds to a  $\delta$ -symbol. Let us now consider the nonlinear correction to this propagator, coming from the cubic term in (8.40). The structure of  $\Gamma$ , obtained in the same way as in (8.42) is:

$$\Gamma_{ab|cd|ef} = N^{-1/2} \left( \text{diagram} + \text{permutations} \right) \tag{8.45}$$

(here single lines are ordinary propagators  $\delta_{ab}/(k^2 + \mu^2)$ ). Therefore, the first correction to  $\langle v_{ab}v_{cd} \rangle$  contains a term which has the form:

$$\langle v_{ab}v_{cd} \rangle^{(1)} = N^{-1} \times \frac{b}{a} \text{diagram} \frac{c}{d} \tag{8.46}$$

We see, that because of the summation over  $e$ , the diagram in (8.46) acquires a factor  $N$  which cancels the  $N^{-1}$  in front.

It is easy to check that such a cancellation will take place in any planar graph (i.e. a graph which does not have overlapping lines). For instance consider the diagram:

(8.47)

It contains four cubic and one quartic vertices. According to (8.42) it will come with a factor  $(N^{-1/2})^4 N^{-1} = N^{-3}$ . At the same time we have three windows in (8.47) each carrying a free isotopic index, thus giving a factor  $N^3$ . The diagram doesn't contain any small factor as  $N \rightarrow \infty$ .

Nonplanar diagrams are small. For example:

(8.48)

This is easily checked by the above double line representation.

We have reached the conclusion that in order to find the large  $N$  limit for principal chiral fields one has to sum all planar diagrams for the field of the Lagrange multiplier.

Why is it that despite the large action  $W$  in (8.40), the saddle point approximation appears to be wrong? The reason is the following. The order of magnitude of the action  $W$  is  $N^2$  (one  $N$  entering in front of the trace, and the other because the trace itself is of the order of  $N$ ). However, the number of variables  $\lambda_{ab}$  is also  $\sim N^2$ . Therefore corrections to the effective action coming from the fluctuations of  $\lambda_{ab}$  are  $\sim N^2$ , i.e. of the same order as the leading term. In the case of the  $n$ -field we had the effective action  $\sim N$ , but the number of variables was one. That explains why we have a good saddle point approximation for  $n$ -fields and nothing similar for chiral ones. Together with the above discussion it also explains why the propagator has an extra  $\log^2(1/mR)$  at  $mR \ll 1$ , according to (8.36). One can think that it comes from exponentiating  $\log \log(1/mR)$  terms which are, according to (8.30), present in planar graphs.

At the same time, though the propagators are not free and infinitely many graphs contribute to them, the theory in the limit  $N \rightarrow \infty$  describes free particles. To show this, let us estimate the scattering amplitude  $\mathcal{F}$ :

$$\mathcal{F} \sim \overline{\square\square} + \overline{\square\square\square} + \dots \sim 1/N \tag{8.49}$$

This is an important observation. It means that in the large  $N$  limit the field  $g(x)$  is some complicated but almost local function of some free field, and that in fact we are dealing with a disguised free theory. If we were able to introduce this function explicitly, the perturbation theory in  $1/N$  might become tractable.

This task has not been solved yet, although an exact solution of the  $SU(N)$ -chiral field exists. The best we can do at the moment is to present some steps leading, so it seems, in the right direction. As the trouble was identified as the large number of integration variables, let us introduce the following decomposition:

$$\lambda = \omega \Lambda \omega^{-1} \quad \omega^+ = \omega^{-1} \quad (8.50)$$

with

$$\Lambda_{ab} = \Lambda_a \delta_{ab}$$

Substitution of (8.50) into (8.38) gives after some standard regauging:

$$\frac{W}{\bar{V}} = \frac{1}{e_0^2} \sum_{a=1}^N \Lambda_a - N \operatorname{Tr} \log [ -(\partial_\mu \delta_{ab} + A_\mu^{ab})^2 + \Lambda_a \delta_{ab} ] \quad (8.51)$$

$$A_\mu^{ab} = (\omega^{-1} \partial_\mu \omega)^{ab}$$

We have to replace the integral over  $\lambda_{ab}$  by one over  $\Lambda_a, \omega$ :

$$\prod_{a,b} \mathcal{D}\lambda_{ab} = \prod_a \mathcal{D}\Lambda_a \mathcal{D}A_\mu^{ab} \delta(F_{\mu\nu}^{ab}) \quad (8.52)$$

( $F_{\mu\nu}^{ab}$  being the Yang-Mills field strength).

Let us suppose now that we manage somehow to integrate over  $A_\mu^{ab}$  (it is here that the real difficulty lies). After that we obtain an effective action depending on  $\Lambda_a$ . This time the size of the action is  $N^2$ , while the number of variables is of the order of  $N$ . Hence, we have just to minimize over  $\Lambda_a$  and not to bother about their fluctuations.

The integration over  $A_\mu^{ab}$  is highly nontrivial, and we do not know how to perform it. A useful observation in this respect is that for equal  $\Lambda_a = \Lambda$  the action  $W$ , being gauge invariant, depends on  $F_{\mu\nu}$  only or, since  $F_{\mu\nu} = 0$ , does not depend on  $A_\mu$  at all. So, if we assume that in the vacuum we have an  $a$ -independent condensate of  $\{A_a\}$ , we deduce that we are indeed dealing with a free field theory. At the same time, the Green functions of  $g(x)$  depend on  $A_\mu$  and are nontrivial. So, we have arrived qualitatively at the expected picture. A quantitative check of these guesses has not yet been done. There are no doubts, however, that the mystery of the large  $N$  limit for chiral fields will soon be resolved.

### 8.3 The $CP^{N-1}$ $n$ -Model

Being unable to proceed further with principal chiral fields, we consider in this section another interesting model, namely the field which belongs to the coset space  $CP^{N-1} = SU(N)/SU(N-1) \otimes U(1)$ . It resembles the  $n$ -field in the respect that the large  $N$  expansion in this case is easy, with no planar graphs arising. On the other hand the model is topologically nontrivial and contains instantons. This gives the possibility to analyse effects of topological charge quantitatively. Another remarkable feature of this model, as we shall see, is the dynamical generation of gauge fields.

Complex projective space  $CP^{N-1}$  is defined by taking  $N$ -dimensional complex space and identifying in it the point  $(z_1, \dots, z_N)$  with the point  $(\lambda z_1, \dots, \lambda z_N)$  where  $\lambda \neq 0$  is an arbitrary complex number. By choice of  $\lambda$  we can parametrize the points of  $CP^{N-1}$  by a unit sphere:

$$|z|^2 = |z_1|^2 + \dots + |z_N|^2 = 1 \quad (8.53)$$

with the identification

$$z \sim e^{iy} z \quad (8.54)$$

The complex dimensionality of the resulting space is  $N - 1$ .

The Lagrangian of the  $z(x)$ -field must be invariant under the gauge group:

$$z(x) \rightarrow e^{iy(x)} z(x) \quad (8.55)$$

Only in this case does it describe a field belonging to  $CP^{N-1}$  and not to the  $N$ -dimensional complex sphere (8.53) (which would be the same as the  $2N$ -dimensional  $n$ -field). Such a Lagrangian can be written as

$$S = \frac{1}{e_0^2} \int d^2x |(\partial_\mu - iA_\mu)z|^2 \quad (8.56)$$

with  $A_\mu$  being a new independent field which transforms as

$$A_\mu \rightarrow A_\mu + \partial_\mu \gamma \quad (8.57)$$

It is quite obvious that (8.56) is invariant under simultaneous transformations (8.55) and (8.57). Also, since (8.56) does not contain a kinetic term for the  $A_\mu$ -field this field can be eliminated (at least classically) by minimizing  $S$ . We have:

$$\frac{\delta S}{\delta A_\mu} = \frac{\delta}{\delta A_\mu} (|\partial_\mu z|^2 + i(z^* \partial_\mu z - (\partial_\mu z^*) z) A_\mu + A_\mu^2) = 0 \quad (8.58)$$

From (8.58) we find:

$$A_\mu = -\frac{i}{2} (z^* \partial_\mu z - (\partial_\mu z^*) z) \quad (8.59)$$

This value of  $A_\mu$  can be substituted back into (8.56), providing us with a nonlinear Lagrangian which depends on the  $z$ -field only. In the case  $N = 2$  these Lagrangians describe the  $O(3)$   $\mathbf{n}$ -field, because  $CP^1$  is equivalent to a two-dimensional sphere. A simple way to see this is based on the relation:

$$\mathbf{n} = (z^\dagger \boldsymbol{\sigma} z) \quad z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \quad (8.60)$$

( $\boldsymbol{\sigma}$  are the Pauli matrices).

If  $z^\dagger z = 1$ , the  $\mathbf{n}$ -vector satisfies the condition:

$$\mathbf{n}^2 = 1 \quad (8.61)$$

These formulas define a topologically interesting projection of the sphere  $S^3$  in  $z$ -space onto a sphere  $S^2$  defined by the  $\mathbf{n}$ . It is called Hopf's bundle in mathematics. The aspect which is interesting for us now is that the Lagrangian (8.56) with the constraint (8.59) can be written as a Lagrangian for the  $\mathbf{n}$ -field, with  $\mathbf{n}$  given by the Hopf projection (8.60). This can be checked by a straightforward calculation, but instead one can argue that  $(\partial_\mu \mathbf{n})^2$  expressed in terms of  $z$  gives a nonlinear Lagrangian, with two derivatives only, which is invariant under the gauge transformation (8.55) (because  $\mathbf{n}$  does not change at all under this transformation). There is only one expression with such a property and hence

$$\frac{1}{4} (\partial_\mu \mathbf{n})^2 = |(\partial_\mu - iA_\mu)z|^2 \quad (8.62)$$

$$\mathbf{n} = z^\dagger \boldsymbol{\sigma} z; \quad A_\mu = -\frac{i}{2} [z^\dagger \partial_\mu z - (\partial_\mu z^\dagger) z].$$

Another useful relation concerns the density of topological charge. We have:

$$\frac{1}{2} \mathbf{n} [\partial_\mu \mathbf{n} \times \partial_\nu \mathbf{n}] = \partial_\mu A_\nu - \partial_\nu A_\mu \quad (8.63)$$

which again can be expected on the basis of counting derivatives, tensor properties and gauge invariance (and after that checked by direct computation).

Let us stress again that the option of introducing a local, gauge invariant field  $\mathbf{n}$  exists only in the case of  $CP^1$ ; for  $CP^{N-1}$  with  $N > 2$  it is not possible and we have to work with "charged" fields  $z$ . We shall investigate now the limit of large  $N$  which exhibits some interesting phenomena.

Let us write the partition function as:

$$\begin{aligned}
 Z &= \int \mathcal{D}^2 \mathbf{z}(x) \mathcal{D} A_\mu(x) \delta(z^\dagger z - 1) \exp\left(-\frac{1}{e_0^2} \int |(\partial_\mu - i A_\mu) \mathbf{z}|^2 d^2 x\right) \\
 &= \int \mathcal{D} \lambda(x) \mathcal{D}^2 \mathbf{z}(x) \mathcal{D} A_\mu(x) \exp\left\{\frac{\int \lambda d^2 x}{e_0^2} - \frac{1}{e_0^2} \int \left\{ |(\partial_\mu - i A_\mu) \mathbf{z}|^2 + \lambda |\mathbf{z}|^2 \right\} d^2 x\right\} \\
 &= \int \mathcal{D} \lambda(x) \mathcal{D} A_\mu(x) \exp\left\{\frac{1}{e_0^2} \int \lambda(x) d^2 x - N \operatorname{Tr} \log[-(\partial_\mu - i A_\mu)^2 + \lambda]\right\}
 \end{aligned} \tag{8.64}$$

In this case, in the large  $N$  limit the saddle point approximation works perfectly. In the vacuum:

$$\langle \lambda \rangle = m^2 \quad \langle A_\mu \rangle = 0 \tag{8.65}$$

$$1 = N e_0^2 \int \frac{d^2 p}{(2\pi)^2 (p^2 + m^2)}$$

At this point, a new phenomenon occurs. The field  $A_\mu$  originally had no kinetic energy and could have been eliminated from the Lagrangian. However, as we consider corrections to the  $1/N$  expansion using an expansion of the determinant in (8.64) near the saddle point (8.65), we obtain in the quadratic approximation:

$$W_{\text{II}} = \frac{N}{2} \left\{ \sum_q (\Pi(q) v(q) v(-q) + \Pi_{\mu\nu}(q) A_\mu(q) A_\nu(q)) \right\} \tag{8.66}$$

with:

$$\begin{aligned}
 \Pi(q) &= \text{---} \text{---} \text{---} \text{---} = \int \frac{d^2 k}{(2\pi)^2} \frac{1}{(k^2 + m^2)((k+q)^2 + m^2)} \\
 \Pi_{\mu\nu}(q) &= \text{---} \text{---} \text{---} \text{---} + \text{---} \text{---} \text{---} \text{---} = \\
 &= \int \frac{d^2 k}{(2\pi)^2} \frac{(2k+q)_\mu (2k+q)_\nu}{(k^2 + m^2)((k+q)^2 + m^2)} - 2\delta_{\mu\nu} \int \frac{d^2 k}{(2\pi)^2 (k^2 + m^2)}
 \end{aligned} \tag{8.67}$$

As could have been expected on the basis of gauge invariance,  $\Pi_{\mu\nu}$  satisfies the relation:

$$q_\mu \Pi_{\mu\nu}(q) = 0 \tag{8.68}$$

$$\Pi_{\mu\nu}(q) \underset{q \rightarrow 0}{\approx} \frac{\text{const.}}{m^2} (q^2 \delta_{\mu\nu} - q_\mu q_\nu)$$

On substituting it into (8.66) we find that for large wavelengths the effective action contains a term:

$$W_{\text{II}} \sim \frac{N}{m^2} \int d^2 x F_{\mu\nu}^2 + \dots \tag{8.69}$$



to construct such terms:

$$\begin{aligned}\Gamma_\mu(p_1, p_2; k) &\sim (\epsilon_{\alpha\beta} p_{1\alpha} p_{2\beta}) \epsilon_{\mu\nu} k_\nu \\ &= p_{1\mu}(p_2 k) - p_{2\mu}(p_1 k)\end{aligned}\quad (8.74)$$

If we substitute this into (8.72) we obtain:

$$\begin{aligned}\Gamma^{(1)}(p_1, p_2) &\sim \theta(\epsilon_{\alpha\beta} p_{1\alpha} p_{2\beta}) \frac{1}{k^2} \epsilon_{\mu\lambda} k_\lambda \epsilon_{\mu\nu} k_\nu \\ &= \theta \epsilon_{\alpha\beta} p_{1\alpha} p_{2\beta}\end{aligned}\quad (8.75)$$

Presence of the  $\epsilon_{\mu\nu}$ -containing terms in the two point function indicates breakdown of  $CP$ -invariance. As we have discussed before, the analogous phenomenon in QCD creates some unsolved problems.

Let us finally discuss to what extent the topological effects described above can be attributed to instantons. A naive estimate of the instanton contribution to  $Z$  would give a contribution of the order of  $e^{-N}$ . This happens because the instanton action for the  $CP^{N-1}$  action is finite and independent of  $N$ . At the same time, the coupling constant  $e_0^2$  scales as  $1/N$ . Hence we obtain an exponentially small contribution. However, this naive argument is wrong. The one loop calculation of the determinant near a multi-instanton configuration reduces the partition function of the  $CP^{N-1}$  model to one of some generalized plasma, in the same way as in  $CP^1$  (see Chapter 6). The correlation length which is established in this plasma is such that the entropy, coming from the fluctuations, compensates the smallness of the classical contribution. Roughly speaking the following happens. The one instanton contribution to the partition function has the form:

$$\begin{aligned}V^{-1} \log Z^{(1)} &\sim \int \frac{d\rho}{\rho^3} \exp\left(-\frac{2\pi N}{2\pi} \log(\mu\rho)^{-1}\right) \\ &= \int \frac{d\rho}{\rho^3} (\mu\rho)^N\end{aligned}\quad (8.76)$$

This formula is strictly correct only for  $\mu\rho \ll 1$ , thus indicating the exponential damping described above. However, if we consider the plasma and not a dilute gas approximation, the infrared divergence in (8.76) gets cut off at  $\rho \sim \mu^{-1}$ . It is true that at this point, where instantons dissociate (just as in the  $O(3)$  case of Chapter 6) the one loop approximation is not applicable, and we have to deal with a strong

coupling problem. By counting powers of  $N$ , we can easily show that multiloop corrections to the (8.76) would lead to the replacement:

$$\int \frac{d\rho}{\rho^3} (\mu\rho)^N \rightarrow \int \frac{d\rho}{\rho^3} (f(\mu\rho))^N$$

with some unknown  $f$ , which can be computed for small  $\mu\rho$  loop by loop. The use of the approximation  $f(x) \simeq x$  for  $x \sim 1$  is only qualitatively justified.

So, the final conclusion is the following. At all values of  $N$  we have nontrivial topological fluctuations of the fields. They can be efficiently described by the  $1/N$  approximation. There could be another, complementary, description in which they are represented as a collection of melted instantons. Unfortunately rigorous quantitative methods for the second description are not known at present. It seems that the one-loop WKB approximation is qualitatively sensible. But before the development of quantitative methods, the possibility of describing topological excitations in terms of instantons remains a semantic question.

#### 8.4 Non-Abelian Gauge Theory

This is the case most interesting for us. Its  $1/N$  properties are to some extent similar to those of principal chiral fields. It is convenient to consider not  $SU(N)$  but the  $U(N) = SU(N) \otimes U(1)$  case. Here again, the  $U(1)$  part trivially decouples. The field is described by anti-Hermitian matrices:

$$\begin{aligned} (A_\mu)^j_i &= -(A_\mu^*)^i_j; \quad i, j = 1, \dots, N \\ (F_{\mu\nu})^j_i &= \partial_\mu A^j_\nu - \partial_\nu A^j_\mu - A^k_{\mu i} A^j_{\nu k} + A^k_{\nu i} A^j_{\mu k} \\ S &= -\frac{N}{4e_0^2} \int \text{Tr} F_{\mu\nu}^2 d^4x \\ &= -\frac{N}{4e_0^2} \int (F_{\mu\nu})^j_i (F_{\mu\nu})^i_j d^4x \end{aligned} \tag{8.77}$$

Here we have changed the normalization for the bare coupling constant (in comparison with the previous chapters) so as to make the action have its natural scale  $N^2$ . One of the  $N$ -factors come from the trace; the scale is natural since the Lagrangian (8.77) describes  $N^2$  interacting gluons. Their vacuum fluctuations provide the energy or effective action  $\sim N^2$ , of the same scale as the classical action (8.77). Another check is

the formula (2.68) of asymptotic freedom. For rescaled coupling it is  $N$ -independent.

Just as before, in spite of the large factor in front of the action, the naive saddle point approximation does not work owing to the large number of fields, and the sum over all planar diagrams is needed. To see this one has to use a double line representation, as in Section 8.2:

$$\langle (A_\mu)^j_i (A_\nu)^l_k \rangle = \delta_i^l \delta_k^j \mathcal{D}_{\mu\nu}(x-y) \equiv \begin{array}{c} i \quad l \\ \leftarrow \quad \rightarrow \\ j \quad k \end{array} = \begin{array}{c} i \quad l \\ \text{~~~~~} \\ j \quad k \end{array} \quad (8.78)$$

$$\begin{array}{c} \text{Y} \\ \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array}, \dots$$

In a certain sense the gauge field is represented by a “quark” with index  $i$  and an antiquark with index  $j$ . Each “quark” line in (8.78) corresponds to a  $\delta$ -symbol. Of course the word “quark” here is just a way of expressing the fact that the adjoint representation can be obtained as a product of two conjugate fundamental ones. We have not yet introduced quarks as physical particles.

With this notation, the logarithm of the partition function is represented by planar Feynman graphs:

$$\log Z = \begin{array}{c} \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \quad \text{---} \\ \text{---} \quad \text{---} \end{array} + \dots \quad (8.79)$$

$$= N^2 f(e_0^2)$$

Let us check this. The first diagram contains the summation  $\delta_i^j \delta_k^l = N^2$ . The second one has three closed paths and hence the factor  $N^3$ , but it is also proportional to the coupling constant, which in our notation, gives a factor  $e_0^2/N$ . In the same way it is checked that any planar diagram has the same magnitude, while any nonplanar one is suppressed by  $1/N$ . This has an important topological interpretation. Take the first diagram in (8.79), and imagine that it is a picture of two disks, lying one upon the other. The orientation of each disk is defined by the corresponding arrow on its boundary. Let us now glue together the “quark” and “anti-quark” lines. As a result we shall obtain a topological sphere. Any planar diagram has this property—after gluing all the cuts we obtain a sphere. Now, according to (8.77), each vertex of a diagram carries a factor  $N$ , each propagator (or edge on our surface) contains  $N^{-1}$ , and each free face contains a closed loop giving  $N$ . Hence the total contribution is:

$$N^V - E + F \equiv N^X \quad (8.80)$$



Let us now suppose that we have a pole in the  $\langle \varepsilon\varepsilon \rangle$  correlation function. It describes some one-particle state,  $|r\rangle$ . According to the first of equation (8.83) we have:

$$\begin{aligned} \langle \varepsilon\varepsilon \rangle &= \text{---} \overbrace{\text{---}}^r \text{---} + \dots \sim N^0 \\ \langle 0|\varepsilon|r \rangle &\sim N^0 \end{aligned} \tag{8.84}$$

At the same time, the second of equations (8.83) gives:

$$\langle \varepsilon\varepsilon\varepsilon \rangle = \text{---} \overbrace{\text{---}}^r \text{---} + \text{---} \overbrace{\text{---}}^r \text{---} + \dots \sim N^{-1} \tag{8.85}$$

From this we conclude that:

$$\begin{aligned} \langle 0|\varepsilon|r_1 r_2 \rangle &= \text{---} \overbrace{\text{---}}^{r_1} \text{---} \overbrace{\text{---}}^{r_2} \text{---} \sim N^{-1} \\ g = \text{---} \overbrace{\text{---}}^{r_1} \text{---} \overbrace{\text{---}}^{r_2} \text{---} \overbrace{\text{---}}^{r_3} \text{---} &\sim N^{-1} \end{aligned} \tag{8.86}$$

These estimates show that we have no thresholds in the correlation functions as  $N = \infty$ , and that the width of resonant states scale as  $g^2 \sim N^{-2}$ . The same is the scale of the two particle scattering amplitude.

It is quite clear that the number of particles in the theory must be infinite and that their masses should increase. This follows from the representation:

$$\langle \varepsilon(p)\varepsilon(-p) \rangle = \sum_r \frac{g_r^2}{p^2 + m_r^2} \tag{8.87}$$

If we consider the limit  $p^2 \rightarrow \infty$ , then for a finite number of resonances we would get the behaviour  $\sim p^{-2}$  from (8.87). However, as we know from asymptotic freedom, the true result contains powers of  $\log^{-1}(p/m)$ . This is compatible with (8.87) only in the case of an infinite number of poles.

This result is very natural from the point of view of string representations. Indeed, as we have seen in Chapter 3, in the confining phase elementary excitations are formed from closed strings of electric flux. Such closed string have infinitely many vibrational modes (we shall study them in Chapter 10), each of which corresponds to a particle and produces a pole in (8.87). It is easy to give a crude estimate for the number of states of a given mass. For a string of length  $L$ , the number of its configurations increases as  $e^{cL}$ . The mass of a configuration is proportional to  $L$ . Hence

$$N(M) \sim e^{cM} \tag{8.88}$$

(where  $N(M)$  is the number of particles with mass  $M$ , and  $c$  is some constant).

In Chapter 10 we shall discuss these things in more detail.

Another important feature of the large  $N$  limit concerns the phase factors  $\psi(c)$ . Let us show that the field:

$$\varphi(C) = \frac{1}{N} \text{Tr} \left( \text{P exp} \oint_C A_\mu dx^\mu \right) \tag{8.89}$$

has the following decoupling property:†

$$\langle \varphi(C_1)\varphi(C_2) \rangle \underset{N \rightarrow \infty}{\simeq} \langle \varphi(C_1) \rangle \langle \varphi(C_2) \rangle \tag{8.90}$$

This is again immediately seen from the “double line” representation. We have

$$\langle \varphi(c) \rangle = 1 + \frac{1}{N} \text{diagram} + \dots = 1 + \frac{1}{N} \text{diagram} + \dots \sim N^0 \tag{8.91}$$

$$\langle \varphi(C_1)\varphi(C_2) \rangle - \langle \varphi(C_1) \rangle \langle \varphi(C_2) \rangle = \frac{1}{N^2} \text{diagram} + \dots = \frac{1}{N^2} \text{diagram} + \dots \sim N^{-1} \tag{8.92}$$

Equation (8.90) implies that the field  $\varphi(C)$  can be considered as a classical field in loop space, because its fluctuations  $\langle (\varphi(c) - \langle \varphi(c) \rangle)^2 \rangle$  are negligible. This does not imply, however, that  $A_\mu$  itself becomes classical in the large  $N$  limit.

This classical field  $W(C) = \langle \varphi(C) \rangle$  satisfies a closed nonlinear equation in loop space which follows from (7.41). We have:

$$\frac{\partial^2 W(C)}{\partial x^2(s)} = -e_0^2 \oint \delta(x(s) - y) \dot{x}_\mu(s) dy_\mu \times W(\bar{C})W(\bar{C}) \tag{8.93}$$

It can be shown by proceeding along the lines mentioned in Chapter 7, that perturbative solution of (8.93) gives all planar diagrams, as it should, in their unrenormalized form. However, the real destination of this equation is to help one to choose from among possible theories of free strings the one which describes large  $N$  gauge theory. Unfortunately this most important question is not settled yet. The main difficulty

† This property was discovered by A. A. Migdal (1977).

lies in the inadequacy of our understanding of string theory. This will be discussed in Chapter 10. Another problem is related to the following clumsy feature of equation (8.93). This equation is true only for a definite regularization of gauge theory and of string theory. Therefore it is possible that the "correct" string theory will not satisfy (8.93) just because it has been regularized in a different way. It makes attempts to solve (8.93) by the use of the string ansatz very difficult. Perhaps it will be easier to compare directly the small distance behaviour in the string theory with the one described by asymptotically free gluons. The situation with the loop equations, described above, is reminiscent of the one with Schrödinger equations for field theory. Although these functional equations are in principle correct for the regularized theory, and can be renormalized with some effort, their practical use up to now has been quite limited. Instead, it appears more convenient to study directly their solutions in the form of functional integrals.

In the next chapter we shall describe a similar approach to the theory of strings.



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

---

## CHAPTER 9

# Quantum Strings and Random Surfaces

---

---

DOI: 10.1201/9780203755082-9

We have seen in the previous chapters that while chiral fields in the large  $N$  limit describe free particles, gauge theories describe, or are described by, noninteracting strings. The notion of “free string” is far from being trivial, and actually, we have used this term before rather loosely. It is the purpose of this chapter to present the theory of strings. Unfortunately this theory is not yet completed. So, we shall give a review of existing results and indicate the possible directions of future development.

We begin our discussion with a very simple special case of infinitely short strings or, which is the same, point-like particles. After this warm-up we will proceed to our main case of interest—the general theory of strings.

### 9.1 Mathematical Preliminaries: Summation of Random Paths

The position of a point-like particle is described by a four-vector  $x_\mu$ . The meaningful question is: what is the amplitude  $G(x, x')$  for a particle to go from the point  $x$  to a point  $x'$ . As usual, this amplitude is given by the sum over all possible trajectories connecting the points  $x$  and  $x'$ . In Euclidean space the amplitude is given by:

$$G(x, x') = \sum_{(P_{xx'})} \exp\left(-\frac{S[P_{xx'}]}{\hbar}\right) \quad (9.1)$$

Here  $P_{xx'}$  are paths connecting  $x$  and  $x'$ , and  $S[P_{xx'}]$  is the classical action for the given path. As such, one takes the simplest invariant characteristic of the path—its length. So, we have the following expression for the action:

$$S[P_{xx'}] = m_0 L(P_{xx'}) \quad (9.2)$$

where  $m_0$  is some parameter (connected with the mass of the particle) and  $L$  is the length of the path  $P_{xx'}$ .

Our aim now is to define properly the sum in (9.1) and to proceed to its computation. Had we had a lattice in the  $x$  space, the definition of the sum would be easy, because on the lattice there are a finite number of  $P_{xx'}$  having a fixed length. So, one respectable way of doing things is to start on a lattice, compute the sum (9.1), and then take the mesh  $a \rightarrow 0$ , simultaneously choosing  $m_0(a)$  so as to make the amplitude  $G(x, x')$  finite. Since the problem of the random walk on a lattice is exactly solvable, the way described will readily give the desired answer. It has only one flaw (apart from the aesthetic one): there exists no generalization of this approach to the case of surfaces. Because of this we shall not take this standard route but try to develop continuum theory *ab initio*.

In the continuum theory the action (9.2) takes the form:

$$S = m_0 \int_0^1 d\tau \left[ \left( \frac{dx_\mu}{d\tau} \right)^2 \right]^{1/2} \quad (9.3)$$

(here we have parametrized the path by  $x_\mu = x_\mu(\tau)$  with  $x_\mu(0) = x_\mu$ ;  $x_\mu(1) = x'_\mu$ ). This action is invariant under the “gauge transformations” or “diffeomorphisms” or “reparametrizations”, given by

$$x_\mu(\tau) \rightarrow x_\mu(f(\tau)) \quad (9.4)$$

with the function  $f(\tau)$  satisfying the conditions

$$f(0) = 0 \quad f(1) = 1 \quad \frac{df(\tau)}{d\tau} > 0 \quad (9.5)$$

At this point we encounter two difficulties. First, we have to manage the integration of the action (9.3) containing a square root. Second, the measure of integration should be defined so as to count only  $x_\mu(\tau)$  modulo reparametrization (9.4) (it is obvious that two functions  $x_\mu(\tau)$  and  $x_\mu(f(\tau))$  describe the same single path). We shall now show how to resolve these difficulties which, to some extent, compensate each other.

We wish to compute the integral:

$$G(x, x') = \int \left( \frac{\mathcal{D}x(\tau)}{\mathcal{D}f(\tau)} \right) \exp \left( -m_0 \int_0^1 (\dot{x}^2(\tau))^{1/2} d\tau \right) \quad (9.6)$$

where we have denoted by the ratio  $\mathcal{D}x(\tau)/\mathcal{D}f(\tau)$  the measure on the coset space obtained from the space of all  $x_\mu(\tau)$  by identifying functions

connected by reparametrizations. In other words  $\mathcal{D}f(\tau)$  can be understood as the volume of the gauge (reparametrization) group. Let us rewrite (9.6) as

$$G(x, x') = \int \frac{\mathcal{D}h(\tau)}{\mathcal{D}f(\tau)} \exp\left(-m_0 \int_0^1 (h(\tau))^{1/2} d\tau\right) \times \int \mathcal{D}x(\tau) \delta(\dot{x}^2(\tau) - h(\tau)) \tag{9.7}$$

where we have introduced a “metric tensor”  $h(\tau)$  on the path and inserted a functional  $\delta$ -function into the integrand. Let us begin with computation of the second integral in (9.7) and use a Lagrange multiplier to define the  $\delta$ -function:

$$\begin{aligned} \mathcal{X}(x, x', h(\tau)) &\equiv \int \mathcal{D}x(\tau) \delta(\dot{x}^2(\tau) - h(\tau)) \\ &= \int_{c-i\infty}^{c+i\infty} \mathcal{D}\lambda(\tau) \exp\left(\int_0^1 \lambda(\tau) h(\tau) d\tau\right) \int \mathcal{D}x(\tau) \exp\left(-\int_0^1 \lambda(\tau) \dot{x}^2(\tau) d\tau\right) \end{aligned} \tag{9.8}$$

The action in (9.8) is invariant under reparametrizations, if we transform:

$$\begin{aligned} x(\tau) &\rightarrow x(f(\tau)) \\ h(\tau) &\rightarrow \left(\frac{df}{d\tau}\right)^2 h(f(\tau)) \\ \lambda(\tau) &\rightarrow \left(\frac{df}{d\tau}\right)^{-1} \lambda(f(\tau)) \end{aligned} \tag{9.9}$$

The measures of integration and the cut-off have to be defined so as to preserve (9.9). In particular, it will be quite a mistake to split the interval  $0 \leq \tau \leq 1$  into small equal pieces, because such a procedure violates gauge invariance. Instead the size of the mesh  $\Delta\tau$  has to be defined by:

$$h(\tau_j) (\Delta\tau_j)^2 = \epsilon^2 \tag{9.10}$$

(because  $h(\tau)$  is the metric tensor).

It is convenient to introduce instead of the “tensor”  $\lambda$  the “scalar” Lagrange multiplier,  $\alpha(\tau)$ :

$$\begin{aligned}\lambda(\tau) &\equiv \alpha(\tau)h(\tau)^{-1/2} \\ \alpha(\tau) &\rightarrow \alpha(f(\tau))\end{aligned}\tag{9.11}$$

So that:

$$\begin{aligned}\mathcal{K}(x, x', h(\tau)) &= \int \mathcal{D}\alpha(\tau) \exp \int_0^1 \alpha(\tau)(h(\tau))^{1/2} d\tau \\ &\times \int \mathcal{D}x(\tau) \exp \left\{ - \int_0^1 \alpha(\tau) \frac{\dot{x}^2(\tau)}{(h(\tau))^{1/2}} d\tau \right\}\end{aligned}\tag{9.12}$$

We shall show now that in the continuum limit  $\alpha(\tau)$  in (9.12) can be replaced by a constant  $\langle \alpha \rangle$ , just as happened in Chapter 8 with the  $\sigma$ -model in the large  $N$  limit. It is convenient to make one more (the last) change of variables in (9.12) by introducing instead of  $\tau$  the proper time  $t$ :

$$t = \int_0^\tau (h(\tau_1))^{1/2} d\tau_1; \quad T = t(1)\tag{9.13}$$

$$\begin{aligned}\mathcal{K}(x, x', T) &= \int \mathcal{D}\alpha(t) \exp \int_0^T \alpha(t) dt \\ &\times \int_{\substack{x(T)=x' \\ x(0)=x}} \mathcal{D}x(t) \exp \left( - \int_0^T \alpha(t) \dot{x}^2(t) dt \right)\end{aligned}\tag{9.14}$$

According to (9.10), the functional integral in (9.14) is defined in the conventional fashion, namely by splitting the interval  $[0, T]$  into equal pieces:

$$(\Delta t_j) = \varepsilon\tag{9.15}$$

It is not hard to compute the Gaussian integral in (9.14) with the regularization (9.15). Let us do it, though as we shall explain later, the

final answer could have been foreseen without this computation. We have:

$$\begin{aligned}
 & \int_{\substack{x(T)=x' \\ x(0)=x}} \mathcal{D}x(t) \exp\left[-\int_0^T \alpha(t) \dot{x}^2(t) dt\right] \\
 & \simeq \int \prod_{i=0}^{T/\varepsilon} dx_i \exp\left[-\sum_{i=0}^{T/\varepsilon} \alpha_i \frac{(x_{i+1} - x_i)^2}{\varepsilon}\right] \\
 & = \int \prod_{i=0}^{T/\varepsilon} dV_i \delta\left(\sum_{i=0}^{T/\varepsilon} V_i + x - x'\right) \exp\left[-\sum_{i=0}^{T/\varepsilon} \frac{\alpha_i V_i^2}{\varepsilon}\right] \\
 & = \int \frac{d\mathbf{p}}{(2\pi)^{\mathcal{D}}} \exp(i\mathbf{p}(x - x')) \prod_i \int dV_i \exp\left(i\mathbf{p} V_i - \frac{\alpha_i V_i^2}{\varepsilon}\right) \\
 & = \text{const.} \left(\prod_{i=0}^{T/\varepsilon} \alpha_i^{-\mathcal{D}/2}\right) \exp\left[-\frac{(x' - x)^2}{\varepsilon \left(\sum_i \alpha_i^{-1}\right)}\right] \left(\sum_i \varepsilon \alpha_i^{-1}\right)^{-\mathcal{D}/2} \quad (9.16)
 \end{aligned}$$

Next, we have to compute a seemingly complicated integral:

$$\mathcal{X}(x, x', T) \sim \int_{c-i\infty}^{c+i\infty} \left(\prod_{i=0}^{T/\varepsilon} d\alpha_i \alpha_i^{-\mathcal{D}/2} e^{\varepsilon \alpha_i}\right) \Phi(x - x'; \varepsilon \sum_i \alpha_i^{-1}) \quad (9.17)$$

where:

$$\Phi(R, \alpha) = \alpha^{-\mathcal{D}/2} \exp(-R^2/\alpha)$$

The major idea which permits us to compute (9.17) in the limit  $\varepsilon \rightarrow 0$  is the “law of large numbers”. Namely the interesting piece  $\Phi$  in the integrand depends only on  $\sum_i \alpha_i^{-1}$ . Each  $\alpha_i$  fluctuates near its mean value  $\langle \alpha \rangle$  and its fluctuation is  $\sim 1$ ; the average is defined by:

$$\begin{aligned}
 \overline{f(\alpha)} & = \int_{c-i\infty}^{c+i\infty} d\alpha \alpha^{-\mathcal{D}/2} e^{\varepsilon \alpha} f(\alpha) \left[ \int_{c-i\infty}^{c+i\infty} d\alpha \alpha^{-\mathcal{D}/2} e^{\varepsilon \alpha} \right]^{-1} \\
 \varepsilon^2 \overline{(\Delta \alpha)^2} & = \varepsilon^2 \overline{\alpha^2} - (\overline{\alpha})^2 \sim 1 \quad (9.18) \\
 \overline{\left(\Delta \frac{1}{\varepsilon \alpha}\right)^2} & \sim 1 \\
 \left[\overline{\Delta \left(\sum_i \frac{1}{\varepsilon \alpha_i}\right)^2}\right]^{1/2} & \sim \left(\frac{T}{\varepsilon}\right)^{1/2} \ll \overline{\sum_i \frac{1}{\varepsilon \alpha_i}} \sim \frac{T}{\varepsilon}
 \end{aligned}$$

As a result, in the limit  $\varepsilon \rightarrow 0$  the formula (9.17) gives:

$$\mathcal{H}(x, x', T) = \text{const. } T^{-g/2} \exp\left(-\frac{(x' - x)^2}{T} \langle \alpha \rangle + T \langle \alpha \rangle\right) \quad (9.19)$$

In the process of this derivation it became clear that we could do much better than performing the dumb discretization (9.16). If we look at (9.14), we discover that the fluctuations of  $\alpha(t)$  are short ranged. Indeed, if we write

$$\alpha(t) = \langle \alpha \rangle + \beta(t) \quad (9.20)$$

then the bilinear term in  $\beta$  in the effective action,  $W_{\text{II}}$ , obtained after Gaussian integration has the representation:

$$\begin{aligned} W_{\text{II}} &= -\frac{q}{\beta} \left( \begin{array}{c} q-k \\ \circlearrowleft \\ x \\ \circlearrowright \\ k \end{array} \right) \frac{q}{\beta} = \int \frac{dq}{2\pi} \beta(q) \beta(-q) \int \frac{dk}{2\pi} \frac{k \cdot (q-k) k \cdot (q-k)}{k^2 (q-k)^2 \langle \alpha \rangle^2} \\ &\simeq \varepsilon \sum_q \beta(q) \beta(-q). \end{aligned} \quad (9.21)$$

(Here we have the vertex of the  $\beta x x$ -interaction from the term  $\int \beta(t) \dot{x}^2(t) dt$  in the Lagrangian.) This computation shows that the field  $\beta$  does not acquire any kinetic energy, and therefore has a correlation in  $t$  of the order of  $\varepsilon$ . That in turn means that in all expressions containing a “large number” of  $\alpha$ , as in  $\int \alpha(t) \dot{x}^2(t) dt$  (provided  $x(t)$  is smooth) one can boldly replace  $\alpha(t)$  by its mean value  $\langle \alpha \rangle$ . This is an important lesson.

So far we have solved only one part of the problem with the result:

$$\begin{aligned} \mathcal{H}(x, x', h(\tau)) &= \int_{\substack{x(0)=x \\ x(1)=x'}} \mathcal{D}x(\tau) \delta(\dot{x}(\tau) - h(\tau)) \\ &= \text{const. } T^{-g/2} \exp\left(-\frac{(x - x')^2}{T} \langle \alpha \rangle + \langle \alpha \rangle T\right), \quad \langle \alpha \rangle \sim \frac{1}{\varepsilon} \\ & \quad T = \int_0^1 d\tau (h(\tau))^{1/2} \end{aligned} \quad (9.22)$$

Now we have to complete the integration in (9.7). Let us notice, first of all, that (9.22) depends on  $h(\tau)$  only through  $T$ . This of course is not an accident since  $T$  is the only invariant quantity in one-dimensional Riemannian geometry (things like scalar curvatures etc. are zero in this case). It is natural to expect that the integration over  $\mathcal{D}h(\tau)/\mathcal{D}f(\tau)$  in

(9.7) will become just the integration over  $T$ . This is indeed the case, but it is worthwhile for future generalizations to work out this fact in detail.

## 9.2 Measures in the Space of Metrics and Diffeomorphisms

In order to define an invariant measure for the  $h$  integration, we shall start from the definition of the metric in the space of metrics  $h$ , and then use the “metric tensor” in this functional space for finding the volume element. The only local expression for the “distance”  $\|\delta h\|$  between the metrics  $h(\tau)$  and  $h(\tau) + \delta h(\tau)$  has the form:

$$\|\delta h\|^2 = \int_0^1 d\tau (\delta h(\tau))^2 h^{-3/2}(\tau) \tag{9.23}$$

It is quite obvious that (9.23) is invariant under reparametrizations.

Let us notice now that any metric  $h(\tau)$  can be made  $\tau$ -independent by means of a properly chosen gauge transformation  $f(\tau)$ . Due to (9.9) this implies that any  $h$  can be represented as:

$$h(\tau) = T^2 \left( \frac{df}{d\tau} \right)^2 \tag{9.24}$$

with some  $T$  and  $f$ . Our strategy will be to pass from the integration over  $\mathcal{D}h(\tau)$  to integration over the new variables  $T$  and  $f$ . According to general rules:

$$\mathcal{D}h(\tau) = dT \mathcal{D}f(\tau) \times (\text{jacobian}) \tag{9.25}$$

If we manage to find the jacobian, our problem will be solved, since instead of  $\mathcal{D}h(\tau)/\mathcal{D}f(\tau)$  we shall obtain a properly defined integral over  $dT$ . The easiest way for solving this problem is to substitute the decomposition (9.24) into (9.23) so as to find the “distance” in terms of the new coordinates.

The computation is greatly simplified if we use a general geometrical formula, which we derive now in its  $n$ -dimensional form (needed for future applications). Suppose, that we have two metric tensors  $g_{ab}(\xi)$  and  $h_{ab}(\xi)$ , where  $\xi$  are coordinates of some  $n$ -dimensional Riemannian manifold. Let these two metrics be connected by the coordinate transformation (or diffeomorphism)  $\xi \rightarrow f(\xi)$ :

$$g = h^{(f)} \tag{9.26a}$$

or, explicitly:

$$g_{ab}(\xi) = \frac{\partial f^c(\xi)}{\partial \xi^a} \frac{\partial f^d(\xi)}{\partial \xi^b} h_{cd}(f(\xi)) \quad (9.26b)$$

We should like to prove the following relation, connecting small variations of  $g$ ,  $h$ , and  $f$ :

$$\begin{aligned} \delta g_{ab} &= [\delta h_{ab} + \nabla_a \omega_b + \nabla_b \omega_a]^{(f)} \\ &= \frac{\partial f^c}{\partial \xi^a} \frac{\partial f^d}{\partial \xi^b} (\delta h_{cd} + \nabla_c \omega_d + \nabla_d \omega_c)(f(\xi)) \end{aligned} \quad (9.27)$$

with

$$\omega^a(\xi) = \delta f^a(f^{-1}(\xi)) \quad (9.28)$$

Here by  $\nabla_a$  we mean the standard covariant derivative computed in the metric  $h$ , and the symbol  $f^{-1}$  means the inverse function. This formula (9.27) can be checked by direct computations, but it can be easily understood without that by the following consideration, based on the group properties of diffeomorphism. We can consider the transformation  $f + \delta f$  as the transformation  $f$ , followed by the infinitesimal transformation  $1 + \omega$ . Hence it is sufficient to compute the change of the metric tensor under the infinitesimal transformation  $\omega$ , which is given by the expression in brackets in (9.27). Let us also notice, that the analogous formula for gauge fields has the form:

$$\begin{aligned} A_\mu &= f^{-1} B_\mu f + f^{-1} \hat{c}_\mu f \\ \delta A_\mu &= f^{-1} [\delta B_\mu + \nabla_\mu \omega] f \\ \omega &= \delta f f^{-1}; \quad \nabla_\mu \omega = \hat{c}_\mu \omega + [B_\mu, \omega] \end{aligned} \quad (9.29)$$

(here  $f(x)$  is the field of unitary matrices, performing gauge transformations).

It is straightforward now to use these general geometrical formulas in our special case. Substitution of (9.24) into (9.23) gives:

$$\begin{aligned} \|\delta h\|^2 &= \int d\tau h^{-3/2}(\tau) \left( \frac{df}{d\tau} \right)^4 [T \delta T + \dot{\omega}]^2 \\ &= \int_0^1 d\tau \left( \frac{df}{d\tau} \right)^{-3} \left( \frac{df}{d\tau} \right)^4 T^{-3} (T \delta T + \dot{\omega}[f])^2 \\ &= \frac{(\delta T)^2}{T} + T^{-3} \int_0^1 df \dot{\omega}^2(f) \end{aligned} \quad (9.30)$$

Here  $\dot{\omega}(f) = d\omega/df$ , and the relation  $\omega(1) = \omega(0) = 0$  has been used (which follows from the conditions  $f(1) = 1$  and  $f(0) = 0$ ).

As the next step let us determine the measure and the metric in the space of diffeomorphisms. Just as in the case of ordinary groups there is only one metric invariant under left and right multiplication simultaneously. It is given by the following general formula:

$$\begin{aligned} \|\delta f\|^2 &= \int (h)^{1/2} h_{ab}(\xi) \omega^a(\xi) \omega^b(\xi) d^n \xi \\ \omega^a(\xi) &= \delta f^a(f^{-1}(\xi)) \end{aligned} \tag{9.31}$$

This formula is a consequence of two facts. Let us consider first “right multiplication”, namely the change:

$$\begin{aligned} f &\rightarrow f \circ \alpha \\ f(\xi) &\rightarrow f(\alpha(\xi)) \end{aligned} \tag{9.32}$$

Then we have:

$$\begin{aligned} f^{-1}(\xi) &\rightarrow \alpha^{-1}(f^{-1}(\xi)) \\ \delta f(\xi) &\rightarrow \delta f(\alpha(\xi)) \\ \omega(\xi) &\rightarrow \delta f(\alpha(\alpha^{-1}(f^{-1}(\xi)))) = \omega(\xi) \end{aligned} \tag{9.33}$$

So, the form  $\omega(\xi)$  is invariant under “right multiplication” of a diffeomorphism. If we look at “left multiplication”, we have:

$$\begin{aligned} f &\rightarrow \beta \circ f \\ f^a(\xi) &\rightarrow \beta^a(f(\xi)) \\ \delta f^a(\xi) &\rightarrow \frac{\partial \beta^a(f)}{\partial f^b} \delta f^b(\xi) \\ \delta f^a(f^{-1}(\xi)) &\rightarrow \frac{\partial \beta^a(\beta^{-1})}{\partial (\beta^{-1})^b} \delta f^b(f^{-1}(\beta^{-1})) \\ \omega^a(\xi) &\rightarrow \frac{\partial \beta^a(\beta^{-1})}{\partial (\beta^{-1})^b} \omega^b(\beta^{-1}(\xi)) \end{aligned} \tag{9.34}$$

Therefore  $\omega^a$  transforms as a standard covariant vector under the left multiplication. If we transform the metric  $h_{ab}$  simultaneously, the distance (9.31) will be invariant. It is obviously the only local expression with these two properties. This expression generalizes the well known double invariant Killing metrics for finite dimensional groups. In the

finite dimensional case, if  $f$  is a group element then the Killing metric is given by:

$$\|\delta f\|^2 = \text{Tr}(\omega^2), \quad \omega = \delta f \cdot f^{-1} \quad (9.35)$$

This metric is invariant under the change:

$$f \rightarrow \beta \cdot f \cdot \alpha; \quad \omega \rightarrow \beta \cdot \omega \cdot \beta^{-1} \quad (9.36)$$

(where dots mean the usual matrix multiplication.)

Diffeomorphisms, while forming an infinite dimensional group, behave very similar to the finite dimensional cases. The only crucial difference is that in order to define the analogue of the trace in (9.35) we have to use the metric of the manifold  $h_{ab}$ , as seen from (9.31).

Certainly, our manipulations with metrics in the functional spaces make sense only if some invariant regularization and renormalization is performed. This will be the case in all our future applications.

Returning to our problem we see that:

$$\|\delta f\|^2 = T^{-3} \int_0^1 \dot{\omega}^2(f) df \quad (9.37)$$

$$\omega = \delta f$$

After rescaling  $t = T \cdot f$ ,  $\omega \rightarrow T\varepsilon$  our expressions can be written as

$$\|\delta h\|^2 = \frac{(\delta T)^2}{T} + \int_0^T dt \varepsilon^2(t) \quad (9.38)$$

$$\|\delta f\|^2 = \int_0^T dt \varepsilon^2(t)$$

From (9.38) we deduce that:

$$\mathcal{D}h(\tau) = \frac{dT}{\sqrt{T}} \text{Det}^{1/2} \left( -\frac{d^2}{d\tau^2} \right) \mathcal{D}f(\tau) \quad (9.39)$$

$$\frac{\mathcal{D}h(\tau)}{\mathcal{D}f(\tau)} = \frac{dT}{\sqrt{T}} \text{Det}^{1/2} \left( -\frac{d^2}{d\tau^2} \right)$$

and hence we have found the desired jacobian. Now it is time to regularize and compute the determinant in (9.39). Its infiniteness reflects the formal nature of our manipulations with infinite dimensional measures. However, we know that the ultraviolet divergences, according to (9.15), have to be cut off by splitting the interval into equal small pieces, since our metric tensor on the interval  $[0, T]$  is equal to 1. There are many equivalent and more convenient regularizations. We shall use the following expression:

$$\log \text{Det}_R \left( -\frac{d^2}{dt^2} \right) = - \int_{\varepsilon^2}^{\infty} \frac{d\tau}{\tau} \sum_n e^{-\tau \lambda_n} \tag{9.40}$$

(where  $\lambda_n$  are the eigenvalues for  $-d^2/dt^2$  which are equal to:  $\lambda_n = \pi^2 n^2 T^{-2}$ ).

For the  $n$  for which  $\lambda_n \ll \varepsilon^{-2}$  the contribution to (9.40) is just  $\log(1/\lambda_n \varepsilon^2)$ . Harmonics with  $\pi n T^{-1} \gg \varepsilon^{-1}$  give negligible contribution to (9.40). But  $\pi n/T$  is just the wave vector corresponding to the eigenfunction  $\psi_n$  of  $-d^2/dt^2$ :

$$\psi_n \sim \sin(\pi n t/T) \tag{9.41}$$

Therefore, the definition (9.40) accounts correctly for the eigenmodes which vary slowly in the time intervals  $\sim \varepsilon$ , but cuts off higher modes. Hence  $\varepsilon$  plays the role of a lattice spacing. A little later we shall see that after renormalization we can take the limit  $\varepsilon \rightarrow 0$ , and that the concrete form of the cut-off is irrelevant.

To do the computation, we represent the sum in (9.40) in the form:

$$\begin{aligned} \sum_{n=1}^{\infty} \exp(-\pi^2 n^2 \tau/T^2) &= \frac{1}{2} \sum_{n=-\infty}^{+\infty} \exp(-\pi^2 n^2 \tau/T^2) - \frac{1}{2} \\ &= \frac{1}{2} \int_{-\infty}^{+\infty} dn \exp(-\pi^2 n^2 \tau/T^2) - \frac{1}{2} + O(\exp(-c/\tau)) \\ &= \frac{T}{2\sqrt{\pi\tau}} - \frac{1}{2} + O(\exp(-c/\tau)) \end{aligned} \tag{9.42}$$

(where the exponential smallness of the correction follows from the Poisson summation formula). We obtain:

$$\begin{aligned}
 & -\log \det_{\mathbf{R}} \left( -\frac{d^2}{dt^2} \right) \\
 &= \int_{\varepsilon^2}^{\infty} \frac{d\tau}{\tau} \sum_{n=1}^{\infty} \exp(-\pi^2 n^2 \tau / T^2) \\
 &= \int_{(\varepsilon/T)^2}^{\infty} \frac{dx}{x} \sum_{n=1}^{\infty} \exp(-\pi^2 n^2 x) \\
 &= \int_{(\varepsilon/T)^2}^1 \frac{dx}{x} \sum_{n=1}^{\infty} \exp(-\pi^2 n^2 x) + \int_1^{\infty} \frac{dx}{x} \sum_{n=1}^{\infty} \exp(-\pi^2 n^2 x) \\
 &= \int_{(\varepsilon/T)^2}^1 \frac{dx}{x} \left( \frac{1}{2\sqrt{(\pi x)}} - \frac{1}{2} \right) + \int_0^1 \frac{dx}{x} \left( \sum_{n=1}^{\infty} \exp(-\pi^2 n^2 x) - \frac{1}{2} - \frac{1}{2\sqrt{(\pi x)}} \right) \\
 &\quad + \int_1^{\infty} \frac{dx}{x} \sum_{n=1}^{\infty} \exp(-\pi^2 n^2 x) \tag{9.43}
 \end{aligned}$$

Now, the last two terms in the last formula are  $T$ -independent constants. Therefore:

$$-\log \det \left( -\frac{d^2}{dt^2} \right)_{\varepsilon \rightarrow 0} = \frac{T}{\varepsilon \sqrt{\pi}} - \log \frac{T}{\varepsilon} \tag{9.44}$$

If we substitute this expression into (9.39) we obtain:

$$\begin{aligned}
 \frac{\mathcal{D}h(\tau)}{\mathcal{D}f(\tau)} &= \text{const} \cdot \exp(-T/2\varepsilon\sqrt{\pi}) \frac{dT}{\sqrt{T}} \cdot T^{1/2} \\
 &= \text{const} \cdot \exp(-T/2\varepsilon\sqrt{\pi}) dT \tag{9.45}
 \end{aligned}$$

The divergences are condensed into the constant factor in front and the term  $\exp(-T/2\varepsilon\sqrt{\pi})$ .

Now we are ready to compute (9.6). Combining (9.45), (9.22) and (9.7) we obtain:

$$\begin{aligned}
 G(x, x') &= \text{const.} \int_0^\infty dT \exp(-(m_0 - \text{const.}/\varepsilon)T) \\
 &\quad \times T^{-\vartheta/2} \exp(-(x - x')^2/\varepsilon T) \\
 &= \text{const.} \int_0^\infty dT \exp(-\mu\varepsilon T) T^{-\vartheta/2} \exp(-(x - x')^2/\varepsilon T) \\
 &= \text{const.} \int \frac{d^{\vartheta}p}{p^2 + \mu} \exp(ip(x - x')), \\
 \mu &= \varepsilon^{-1}(m_0 - m_{0,\text{cr}}) \equiv \varepsilon^{-1} \left( m_0 - \frac{\text{const}}{\sqrt{\varepsilon}} \right)
 \end{aligned} \tag{9.46}$$

All our derivations make sense only if relevant values of  $T$  in the integral (9.46) are much larger than the cut-off (or “lattice spacing”)  $\varepsilon$ . Therefore, the integral (9.6) has a continuum limit only if we adjust  $m_0$  to be very close to the critical value, or, in other words we have to take the limit  $\varepsilon \rightarrow 0$  for the cut-off simultaneously with  $m_0(\varepsilon) \rightarrow m_{0,\text{cr}}$ . Then the terms like  $e^{T/\varepsilon}$  will be compensated and we obtain the continuum theory. This result is quite easy to understand from the lattice point of view. On a lattice we have:

$$G(x, x') = \sum_{L=0}^\infty N_L(x, x') e^{-m_0 L} \tag{9.47}$$

where  $N_L(x, x')$  is the number of paths of the length  $L$  connecting points  $x$  and  $x'$ . We know that for large  $L$ ,  $N_L(x, x') \sim (c)^L$  where  $c$  depends on the type of the lattice. We see, that for  $m_0 > \log c$ , the relevant paths in (9.47) have a length of the order of the lattice spacing and no continuum limit is possible. But as we approach  $m_0 \rightarrow m_{0,\text{cr}} = \log c$ , then a typical  $L \sim (m_0 - m_{0,\text{cr}})^{-1} \gg 1$  and the theory becomes continuous and lattice independent. Our description refers precisely to this limit.

We see from (9.46) that the correlation length or the physical mass have a nontrivial critical exponent:

$$\tau_{\text{corr}}^{-1} \sim m_{\text{phys}} = \sqrt{\mu} \sim (m_0 - m_{0,\text{cr}})^{1/2} \tag{9.48}$$

The procedure of renormalization in this case consists of expressing everything in terms of the physical mass and eliminating the irrelevant

constant factor in front of (9.46). After that we obtain a finite amplitude in the limit  $\varepsilon \rightarrow 0$ .

In view of the future generalizations we have to develop the same formalism as above for closed paths.

### 9.3 Closed Paths

In this case there are some technical differences in the integration over the metric field. Let us begin with the integration over  $x(\tau)$ , provided that the Lagrange multiplier is replaced by its average, which we shall choose to be  $\frac{1}{2}$  by a choice of scale. We have:

$$\mathcal{N}[h(\tau)] = \int_{x(0)=x(T)} \mathcal{D}x(\tau) \exp\left(-\frac{1}{2} \int_0^T \dot{x}^2 d\tau\right) \quad (9.49)$$

$$T = \int_0^1 h^{1/2}(\tau) d\tau$$

First of all we have to remove the trivial divergence in (9.49) connected with the fact that the integrand has translation invariance  $x_\mu \rightarrow x_\mu + c_\mu$  and so we have to fix some point of our loop. Such fixing is most conveniently performed by inserting the relation

$$\int dc \left( \frac{1}{T} \int_0^T \delta(x(\tau) - c) d\tau \right) = 1 \quad (9.50)$$

and omitting the  $\int dc = V$ . After that we have (with the choice  $c = 0$ ):

$$\mathcal{N}[h(\tau)] = \int \mathcal{D}x(\tau) \left( \frac{1}{T} \int_0^T \delta(x(\tau)) d\tau \right) \exp\left(-\frac{1}{2} \int_0^T \dot{x}^2 d\tau\right) \quad (9.51)$$

The  $\delta$ -function in (9.51) will exclude the dangerous zero mode of the operator  $-d^2/dt^2$ . If we expand

$$x(\tau) = \left( a_0 + \sum_{n \neq 0} a_n \exp(2\pi i n \tau / T) \right) \frac{1}{\sqrt{T}} \quad (9.52)$$

then

$$\int_0^T \dot{x}^2 d\tau = \sum_{n \neq 0} \lambda_n a_n^2, \quad \lambda_n = \frac{4\pi^2 n^2}{T^2}$$

$$\mathcal{D}x(\tau) = da_0 \prod_{n \neq 0} da_n$$

From this we derive:

$$\begin{aligned} \mathcal{X}[h(\tau)] &= \int d^{\mathcal{D}}a_0 \frac{1}{T} \int_0^T \delta\left(\frac{a_0}{\sqrt{T}} + \dots\right) d\tau \int \prod_{n \neq 0} da_n \exp\left(-\sum_{n \neq 0} \lambda_n a_n^2\right) \\ &= T^{\mathcal{D}/2} \left[ \det'\left(-\frac{d^2}{dt^2}\right) \right]^{-\mathcal{D}/2} \end{aligned} \tag{9.53}$$

(where  $\det'(x)$  is defined as the product of the nonzero eigenvalues).  
Using the representation of the preceding section we get:

$$\begin{aligned} -\log \text{Det}'\left(-\frac{d^2}{dt^2}\right) &= \int_{\varepsilon^2}^{\infty} \frac{ds}{s} \sum_{n \neq 0} \exp(-4\pi^2 n^2 s/T^2) \\ &= \int_{\varepsilon^2}^{\infty} \frac{ds}{s} \left( \sum_n \exp(-4\pi^2 n^2 s/T^2) - 1 \right) \\ &= \int_{(\varepsilon/T)^2}^{\infty} \frac{dx}{x} \left( \sum_n \exp(-4\pi^2 n^2 x) - 1 \right) \\ &= \int_{(\varepsilon/T)^2}^1 \frac{dx}{x} \left( \sum_n \exp(-4\pi^2 n^2 x) - 1 \right) \\ &\quad + \int_1^{\infty} \frac{dx}{x} \left( \sum_n \exp(-4\pi^2 n^2 x) - 1 \right) \\ &= \int_{(\varepsilon/T)^2}^1 \frac{dx}{x} \left\{ \frac{1}{2\sqrt{(\pi x)}} - 1 + O(\exp(-c/x)) \right\} + \text{const} \\ &= \frac{T}{\varepsilon\sqrt{\pi}} - 2 \log \frac{T}{\varepsilon} \end{aligned} \tag{9.54}$$

This gives the result:

$$\begin{aligned}\mathcal{N}[h(\tau)] &= T^{\mathcal{D}/2} T^{-\mathcal{D}} \exp(cT/\varepsilon) \\ &= T^{-\mathcal{D}/2} \exp(cT/\varepsilon)\end{aligned}\tag{9.55}$$

Now we have to find the measure  $\mathcal{D}h(\tau)/\mathcal{D}f(\tau)$ . Here again we must treat zero modes carefully. Since our parametrization space is a circle it now admits an isometry—a transformation which does not change the metric. This is just a translation  $\tau \rightarrow \tau + a$  and it must be excluded from our gauge group. This can be done in the same way as in the case of the  $x$  integration, by inserting the relation:

$$\int_0^T da \int_0^T \frac{d\tau}{T} \delta(f(\tau) - a) = 1\tag{9.56}$$

As a result the measure in  $T$  will be the same as for open paths and we get the following answer for the number of closed paths of length  $T$ :

$$dN(T) = \frac{dT}{T} T^{-\mathcal{D}/2} \exp(-cT/\varepsilon)\tag{9.57}$$

Of course, this formula could have been anticipated from (9.46). If we set  $x = x'$  we get the integrand in this formula to be  $T^{-\mathcal{D}/2} \exp(cT/\varepsilon)$ . The extra  $1/T$  in (9.57) follows from the fact that we should count paths with different starting points  $x$  as one path. Since for a path of the length  $T$  we have  $T$  different possibilities for a choice of the starting point we obtain a  $1/T$  factor.

Different physical quantities can be expressed in terms of the amplitudes for a path to pass through a prescribed set of points  $\{x_i\}$ . These amplitudes are obtained as expectation values of the following type:

$$\begin{aligned}F(x_1, \dots, x_N) &= \left\langle \prod_j \int_0^T d\tau_j \delta(x(\tau_j) - x_j) \right\rangle \\ &= \int_0^\infty \frac{dT}{T} \int \mathcal{D}x(\tau) \exp\left(-\frac{1}{2} \int \dot{x}^2 d\tau\right) \prod_j \int_0^T d\tau_j \delta(x(\tau_j) - x_j)\end{aligned}\tag{9.58}$$

In order to compute these integrals it is convenient to go to the momentum representation:

$$F(\mathbf{q}_1, \dots, \mathbf{q}_N) = \int_0^\infty \frac{dT}{T} \int \mathcal{D}x(\tau) \exp\left(-\frac{1}{2} \int_0^T \dot{x}^2 d\tau\right) \int_0^T \prod_j d\tau_j \exp\left(i \sum_j \mathbf{q}_j x(\tau_j)\right) \quad (9.59)$$

The integrals in (9.59) are Gaussian. According to the general rules, to find them we have first to solve the classical equations:

$$\ddot{x}_{cl} = -i \sum_j \mathbf{q}_j \delta(\tau - \tau_j) \quad (9.60)$$

and then to compute the classical action. We have:

$$x_{cl}(\tau) = i \sum_j \mathbf{q}_j \mathcal{D}(\tau|\tau_j) \quad (9.61)$$

where  $\mathcal{D}(\tau|\tau')$  is a Green function for the operator  $-d^2/d\tau^2$ . The substitution of (9.61) into the classical action gives:

$$\begin{aligned} S_{cl} &= \frac{1}{2} \int \dot{x}_{cl}^2 d\tau - i \sum_j \mathbf{q}_j x_{cl}(\tau_j) \\ &= \frac{1}{2} \sum_{(i,j)} \mathbf{q}_i \mathbf{q}_j \mathcal{D}(\tau_i|\tau_j) \end{aligned} \quad (9.62)$$

The Green function on a circle contains a zero mode contribution:

$$\mathcal{D}(\tau|\tau') = \sum_{n \neq 0} \frac{\exp i\sqrt{\lambda_n}(\tau - \tau')}{T \lambda_n} + C \quad (9.63)$$

(here  $\lambda_n = 4\pi^2 n^2/T^2$ , and  $C$  is the zero mode contribution which is arbitrary). However if momentum is conserved ( $\sum_j \mathbf{q}_j = 0$ ) this arbitrariness disappears from (9.62):

$$C \sum_{i,j} \mathbf{q}_i \mathbf{q}_j = C \left(\sum_j \mathbf{q}_j\right)^2 = 0 \quad (9.64)$$

The necessity of momentum conservation can be seen also from (9.60) by integrating it over  $\tau$ . Then periodicity of  $\dot{x}_{cl}$  requires  $\sum_j \mathbf{q}_j = 0$ .

We obtained the following expression:

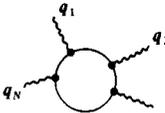
$$\begin{aligned} F(\mathbf{q}_1, \dots, \mathbf{q}_N) &= \int_0^\infty \frac{dT}{T} T^{-\mathcal{D}/2} e^{-m^2 T/2} \int_0^T \prod_j d\tau_j \\ &\times \exp\left\{-\frac{1}{2} \sum_{ij} \mathbf{q}_i \mathbf{q}_j \mathcal{D}(\tau_i|\tau_j)\right\} \end{aligned} \quad (9.65)$$



where we have used the standard representation:

$$\frac{1}{AB} = \int_0^1 \frac{dx}{(Ax + B(1-x))^2}$$

The derivation we have given serves as a check that all normalizations, measures etc. in the path integrals were correctly chosen because at the end it produced the standard answer for the amplitude (9.69). In the same way it can be shown that  $F(q_1, \dots, q_N)$  defined by (9.59) reduces to the standard diagram:

$$F(q_1, \dots, q_N) = \text{diagram} \tag{9.70}$$


In this section we have developed an unusual formalism which deals with path integrals but in the end all the answers have appeared to be much more easily described by the standard Feynman graphs. However, there are serious reasons for working with geometrical integrals directly. They lie in the fact that as we go from paths to surfaces and higher dimensional objects, geometrical functional integrals remain our only existing tool. In the next section we are going to show how this tool works in the case of string theory.

### 9.4 General Theory of Random Hypersurfaces

In the previous sections we discussed one dimensional curves randomly immersed into  $\mathcal{D}$ -dimensional Euclidean space. This problem was equivalent to the problems of Brownian motion and (after analytic continuation to Minkowski space) to the quantum theory of a free relativistic particle. It was hardly possible to get any new results in this field since it has been completely investigated for many years by classical mathematics and quantum physics. However, we have developed an approach which is readily generalizable to the case of an  $n$ -dimensional hypersurfaces immersed into  $\mathcal{D}$ -dimensional space.

This problem is of great interest for both physics and mathematics. At the same time, the case  $n > 1$  is incomparably harder than that of  $n = 1$ . Some incomplete success has been achieved for  $n = 2$  and will be discussed in later sections. Here we shall develop the general formalism for any  $n$  up to the furthest point possible at present.

Let us consider an immersion of the hypersurface described by the functions:

$$x_\mu(\xi^1, \dots, \xi^n) \equiv x_\mu(\xi) \quad \mu = 1, \dots, \mathcal{D} \quad (9.71)$$

The action must depend on  $x_\mu(\xi)$  in such a way that it is invariant under the transformation of diffeomorphisms:

$$\xi \rightarrow f(\xi)$$

or, explicitly:

$$\xi^a \rightarrow \tilde{\xi}^a = f^a(\xi^1, \dots, \xi^n) \quad a = 1, \dots, n \quad (9.72)$$

where the functions  $f^a$  have to satisfy the condition:

$$\frac{\partial(f^1, \dots, f^n)}{\partial(\xi^1, \dots, \xi^n)} \equiv \det \left\| \frac{\partial f^a}{\partial \xi^b} \right\| > 0 \quad (9.73)$$

This claim follows from the fact that  $x = x(\xi)$  and  $x = x(f(\xi))$  represent the same hypersurface, differently parametrized.

When we are looking for a continuous theory of surfaces, we have to start from actions containing the minimal number of derivatives. If a theory with such an action is renormalizable, then all the terms with higher derivatives are irrelevant. Sometimes in order to achieve renormalizability one has to include higher terms. These questions we shall discuss later.

A covariant expression with the minimal number of derivatives is just the hypervolume  $A$  of our hypersurface. Therefore, the simplest action which can be constructed is the following:

$$\begin{aligned} S[x(\xi)] &= m_0^n \int d^n \xi (h(\xi))^{1/2} \\ h(\xi) &\equiv \det \| h_{ab}(\xi) \| \\ h_{ab}(\xi) &= \partial_a x^\mu \partial_b x_\mu \end{aligned} \quad (9.74)$$

In principle, one can add some higher derivative terms like

$$\begin{aligned} S_1 &= c_1 \int d^n \xi R(h) h^{1/2} \\ S_2 &= c_2 \int d^n \xi (\Delta(h)x)^2 h^{1/2} \\ &\text{etc.} \end{aligned} \quad (9.75)$$

(Here  $R(h)$  is the scalar curvature computed with the metric  $h_{ab}$  and  $\Delta(h)$  is the corresponding Laplace operator). In each case special

analysis is needed in order to determine the relevance of those terms. It is reasonable, however, to begin with the simplest action (9.74) and to postpone the question whether renormalization produces higher derivative terms.

The amplitude for a surface with a given boundary can be written as:

$$\begin{aligned} G(c(s)) &= \int \frac{\mathcal{D}\mathbf{x}(\xi)}{\mathcal{D}f(\xi)} \exp\left(-m_0^n \int (\det\|\partial_a x \partial_b x\|)^{1/2} d^n \xi\right) \\ &= \int \frac{\mathcal{D}h_{ab}(\xi)}{\mathcal{D}f(\xi)} \exp\left(-m_0^n \int h^{1/2} d^n \xi\right) \int \mathcal{D}\mathbf{x}(\xi) \delta(\partial_a x \partial_b x - h_{ab}) \end{aligned} \quad (9.76)$$

Here the boundary of the surface  $\mathbf{x}(\xi)$  is an  $n - 1$ -dimensional hypersurface parametrized by the functions  $c_\mu(s_1, \dots, s_{n-1})$ . To give a precise formulation for the boundary condition to the functional integral (9.76) one has to postulate that the topology of the  $\xi$ -space is identical to the topology of the surfaces which we consider. This is the necessary condition under which we can consider  $x_\mu(\xi)$  as smooth functions. If the boundary in  $\xi$ -space is determined from the equations

$$\xi^a = \xi^a(s^1, \dots, s^{n-1}) \quad (9.77)$$

then the boundary conditions for the integral (9.76) are

$$\mathbf{x}(\xi(s)) = c(s) \quad (9.78)$$

The measure  $\mathcal{D}\mathbf{x}(\xi)/\mathcal{D}f(\xi)$  is, just as in the case of paths, the measure on a coset space obtained by factorising all functions  $\mathbf{x}(\xi)$  by the group of diffeomorphisms  $f(\xi)$ . In other words it is an invariant measure on the space of gauge orbits where the gauge transformations in this case are induced by the changes

$$\xi \rightarrow \tilde{\xi} = f(\xi). \quad (9.79)$$

For the invariance of our integral we have to restrict the possible  $f(\xi)$  not only by (9.73) but also by the condition that they do not move the boundary points:

$$f(\xi(s)) = \xi(s) \quad (9.80)$$

The result of integration,  $G[c(s)]$ , is invariant under the reparametrization of the boundary:

$$G[c(s)] = G[c(\alpha(s))] \quad (9.81)$$

(where:

$$\{\alpha^i(s^1, \dots, s^{n-1}), \quad i = 1, \dots, n - 1\}$$

is a diffeomorphism).

As we have done before, let us begin with the computation of the quantity

$$\mathcal{X}[c(s), h(\xi)] = \int \mathcal{D}x(\xi) \delta(\partial_a x \cdot \partial_b x - h_{ab}(\xi)) \quad (9.82)$$

which is equal to the number of immersions of a surface with fixed intrinsic metric.

The use of the Lagrange multiplier representation gives:

$$\begin{aligned} \mathcal{X}[c, h] &= \int_{A-i\infty}^{A+i\infty} \mathcal{D}\lambda^{ab} \exp\left(\int h^{1/2} \lambda^{ab} h_{ab} d^n \xi\right) \\ &\times \int \mathcal{D}x(\xi) \exp\left(-\int h^{1/2} \lambda^{ab} \partial_a x \cdot \partial_b x d^n \xi\right) \end{aligned} \quad (9.83)$$

The norms in functional spaces defining the volume elements in this functional integral are given by:

$$\begin{aligned} \|\delta x(\xi)\|^2 &= \int h^{1/2} (\delta x(\xi))^2 d^n \xi \\ \|\delta \lambda^{ab}\|^2 &= \int h^{1/2} (h_{aa'} h_{bb'} + c h_{ab} h_{a'b'}) \\ &\times \delta \lambda^{ab} \delta \lambda^{a'b'} d^n \xi \end{aligned} \quad (9.84)$$

In deriving these formulas we have used, apart from locality and general covariance, the claim that the distance in functional space has to be invariant under

$$\begin{aligned} x(\xi) &\rightarrow x(\xi) + a(\xi) \\ \lambda^{ab}(\xi) &\rightarrow \lambda^{ab}(\xi) + c^{ab}(\xi) \end{aligned} \quad (9.85)$$

where  $a$  and  $c^{ab}$  are arbitrary. The invariance (9.85) of the measure ensures the possibility of using the equations of motion for  $x$  and  $\lambda$ , which are derived just by the replacements (9.85) in the action.

It appears to be convenient to decompose

$$\lambda^{ab}(\xi) = \alpha(\xi) h^{ab} + f^{ab}(\xi) \quad (9.86)$$

with

$$h_{ab}(\xi) f^{ab}(\xi) = 0.$$

The integral (9.83) takes the form:

$$\begin{aligned} \mathcal{H}[c, h] = & \int \mathcal{D}\alpha(\xi) \exp\left(n \int \alpha(\xi) h^{1/2} d^n \xi\right) \\ & \times \int \mathcal{D}f^{ab}(\xi) \mathcal{D}x(\xi) \exp\left(- \int h^{1/2} (\alpha h^{ab} \partial_a x \partial_b x + f^{ab} \partial_a x \partial_b x) d^h \xi\right) \end{aligned} \quad (9.87)$$

In order to continue the calculation we have to conjecture (and check it later) that the correlation lengths of  $\alpha(\xi)$  and  $f^{ab}(\xi)$  are small in comparison with the size of our region. If this is true, then, as we have seen in case of paths, these quantities can be replaced by their mean values. On the grounds of general covariance we have:

$$\begin{aligned} \langle \alpha(\xi) \rangle &= \bar{\alpha} + c_1 R(\xi) + \dots \\ \langle f^{ab}(\xi) \rangle &= 0 \end{aligned} \quad (9.88a)$$

where  $\bar{\alpha}$  is an unknown constant, being of the order of  $\Lambda^n$  and  $R(\xi)$  is the scalar curvature, computed with the metric  $h_{ab}$ . Equation (9.88) reflects the fact that  $\alpha(\xi)$  is a scalar, while  $f^{ab}(\xi)$  is a traceless tensor.

We shall be interested in such metrics  $h_{ab}$  which are slowly varying on the cut-off scale, or, in other words with  $R(\xi) \ll \Lambda^2$ , and we can neglect the second term in (9.88).

After that we obtain the following expression for  $\mathcal{H}[c, h]$ :

$$\begin{aligned} \mathcal{H}[c, h] = & \exp\left(n\bar{\alpha} \int h^{1/2} d^n \xi\right) \\ & \times \int \mathcal{D}x(\xi) \exp\left(-\bar{\alpha} \int h^{1/2} h^{ab} \partial_a x \cdot \partial_b x d^n \xi\right) \end{aligned} \quad (9.88b)$$

Before going further, we have to check whether our conjectures concerning the “freezing” of the Lagrange multipliers are correct.

Let us begin with fluctuations of  $\alpha$ . Introducing:

$$\alpha(\xi) = \bar{\alpha}(1 + i\beta(\xi)) \quad (9.89)$$

we can easily find the quadratic term in the induced action for the  $\beta$ -fields. Taking for simplicity of estimate  $h_{ab} = \delta_{ab}$  we obtain

$$S_{II}(\beta) = \frac{1}{2} \int \beta(q)\beta(-q)B(q^2) \frac{d^n q}{(2\pi)^n} \quad (9.90)$$

with

$$B(q^2) = - \left\langle \bigcirc_k \right\rangle = \frac{1}{2} \int \frac{d^n k [k(k+q)]^2}{(2\pi)^n k^2 (k+q)^2} \quad (9.91)$$

where we have accounted for the fact that  $\beta$  interacts with the  $x$ -field through the Lagrangian:

$$\mathcal{L}_{\text{int}} = i\bar{\alpha}^2 \int \beta(\xi)(\partial_a x)^2 d^n \xi \quad (9.92)$$

A straightforward estimate of (9.91) gives:

$$B(q^2) = c\Lambda^n(1 + O(q^2/\Lambda^2)) \quad (9.93)$$

If the constant  $c$  in (9.93) does not vanish, then the correlation length for the  $\beta$ -field, which is determined by the singularities in  $q$ -space, is of the order of  $\Lambda^{-1}$ . Therefore, in the generic case  $c \neq 0$ , we can neglect the influence of  $\beta$ -fluctuations since:

$$\frac{(\overline{(\Delta\alpha)^2})^{1/2}}{\alpha} \sim (\overline{\beta^2})^{1/2} \sim (\Lambda A)^{-1/2} \ll 1 \quad (9.94)$$

where  $A$  is the volume of our object. In the presence of nontrivial  $h_{ab}$  our computation indicates that we have a term in the induced action

$$S_{\text{II}}(\beta) = \frac{c}{2} \Lambda^n \int \beta^2(\xi) h^{1/2}(\xi) d^n \xi + \dots \quad (9.95)$$

which suppresses fluctuations of  $\beta$ . Turning to the case of  $f$ -fluctuations, by similar arguments we find the most singular term in the corresponding action:

$$S_{\text{II}}[f] = d \cdot \Lambda^n (\bar{\alpha})^{-2} \int h^{1/2} h_{aa'} h_{bb'} f^{ab} f^{a'b'} d^n \xi \quad (9.96a)$$

which for  $d \neq 0$  indicates the irrelevance of  $f$ -fluctuations.

Let us notice at this point that the question of whether  $c$  and  $d$  may be taken to have generic values or whether one should apply the condition that either of them be zero is far from trivial. There could exist different continuum limits for string theory, the simplest one obtained without extra conditions on  $c$  and  $d$ , while the others require fine tuning of these constants. One has to decide what kind of possible continuum limits have the desired properties and correspond to gauge theory. At present, this question is unsolved, and we shall mainly investigate the generic continuum theory, keeping in mind other options.

In the generic case we have shown that the fluctuations of the Lagrange multiplier may be dropped, and we end up with the following

expression (after trivial change of scale):

$$\begin{aligned} \mathcal{X}[c(s), h_{ab}(\xi)] &= \exp\left(A \int h^{1/2} d^n \xi\right) \\ &\times \int_{\mathbf{x}(\xi(s)) = \mathbf{c}(s)} \mathcal{D}\mathbf{x}(\xi) \exp\left(-\int h^{1/2} h^{ab} \partial_a \mathbf{x} \cdot \partial_b \mathbf{x} d^n \xi\right) \end{aligned} \quad (9.96b)$$

The Green function for the contour  $c(s)$  is obtained by integration on  $h_{ab}$ :

$$\begin{aligned} G[c(s)] &= \int \frac{\mathcal{D}h_{ab}}{\mathcal{D}f} \exp\left(-\mu \int h^{1/2} d^n \xi\right) \\ &\times \int \mathcal{D}\mathbf{x}(\xi) \exp\left(-\int h^{1/2} h^{ab} \partial_a \mathbf{x} \cdot \partial_b \mathbf{x} d^n \xi\right) \end{aligned} \quad (9.97)$$

(where  $\mu$  is the critical parameter).

Since expressions like (9.97) will form the basis of our further discussion, it is worthwhile to present another derivation of it. Let us consider the integral:

$$\int \mathcal{D}h_{ab} \exp\left(-\mu \int h^{1/2} d^n \xi - \int h^{1/2} h^{ab} g_{ab} d^n \xi\right) \quad (9.98)$$

where  $g_{ab}(\xi)$  is some tensor. The integral in (9.98) is supposed to be covariantly regularized. That means that we can compute it by the following procedure. First find the saddle point of the action (9.98):

$$\begin{aligned} &\delta\left(\mu \int h^{1/2} d^n \xi + \int h^{1/2} h^{ab} g_{ab} d^n \xi\right) \\ &= \frac{1}{2} \int h^{1/2} h^{ab} (\mu \delta h_{ab}) d^n \xi \\ &+ \frac{1}{2} \int h^{1/2} (h^{cd} g_{cd} h^{ab} \delta h_{ab} + g_{cd} \delta h^{cd}) d^n \xi = 0 \end{aligned} \quad (9.99)$$

This equation gives the position of the saddle point and the value of the action:

$$h_{ab} \sim g_{ab}, S \sim \int g^{1/2} d^n \xi \quad (9.100)$$

If we consider small fluctuations near this saddle point, we notice that they are nonpropagating owing to the absence of derivatives in (9.98).

More precisely, this means that they have a correlation length proportional to the inverse cut-off. Therefore their leading correction to the effective action must be local, and on the basis of general covariance must have the form (9.100). This is again the demonstration of our general rule that fields with short range correlations can be replaced by their mean values. In our case covariance dictates that these mean values are given by (9.100).

So, by the  $h$ -integration of (9.98) we have recovered the action (9.74), provided that  $g_{ab} = \partial_a \mathbf{x} \cdot \partial_b \mathbf{x}$ . This proves the equivalence of (9.97) to (9.76). Again, as in our first derivation of this equivalence, we have assumed the generic situation, i.e. that no divergent constants are zero. Whether this continuum limit is what we are interested in must be investigated separately in each particular case. In the next section we shall show how to compute (9.97) in the case  $n = 2$  and what kind of physics is described by it.

## 9.5 Two-Dimensional Surfaces. Geometrical Introduction

For  $n = 2$  we can proceed much further with (9.97), by explicit computation of the functional integrals.

To do this we shall first present the necessary geometrical properties of the functionals involved in the game.

Let us consider first:

$$W = \int h^{1/2} h^{ab} \partial_a \mathbf{x} \cdot \partial_b \mathbf{x} \, d^2 \xi \quad (9.101)$$

where the function  $\mathbf{x}(\xi)$  satisfies the boundary condition:

$$\mathbf{x}(\xi(s)) = \mathbf{c}(s). \quad (9.102)$$

Variation of  $W$  with respect to  $h_{ab}$  gives as we have seen in the preceding section:

$$\delta W = \int h^{1/2} T_{ab}(\xi) \delta h^{ab}(\xi) \, d^2 \xi \quad (9.103)$$

where  $T_{ab}$  can be considered as an energy-momentum tensor for the  $\mathbf{x}$  field:

$$T_{ab} = \partial_a \mathbf{x} \cdot \partial_b \mathbf{x} - \frac{1}{2} h_{ab} h^{cd} \partial_c \mathbf{x} \cdot \partial_d \mathbf{x} \quad (9.104)$$

If we take  $h_{ab}$  so as to minimise  $W$ , that is apply the condition

$$T_{ab} = 0 \quad (9.105)$$

we obtain:

$$h_{ab}^{\text{extr}} = c(\xi) \partial_a \mathbf{x} \cdot \partial_b \mathbf{x} \tag{9.106}$$

where  $c(\xi)$  remains undetermined by (9.105). Substituting this into (9.101) we get:

$$W \Big|_{h_{ab} = h_{ab}^{\text{extr}}} = \int (\det \|\partial_a \mathbf{x} \cdot \partial_b \mathbf{x}\|)^{1/2} d^2 \xi \tag{9.107}$$

We conclude, that the problem of finding the minimal area, given by (9.107) can be reduced to two equations:

$$\begin{aligned} \partial_a (h^{1/2} h^{ab} \partial_b \mathbf{x}) &= 0 \\ T_{ab} = \partial_a \mathbf{x} \cdot \partial_b \mathbf{x} - \frac{1}{2} h_{ab} h^{cd} \partial_c \mathbf{x} \cdot \partial_d \mathbf{x} &= 0 \end{aligned} \tag{9.108}$$

The next important geometrical fact which will be extensively used below is the possibility of choosing a “conformal gauge” in which the metric tensor  $h_{ab}$  takes the form:

$$h_{ab}(\xi) = e^{\varphi(\xi)} \delta_{ab} \tag{9.109}$$

This extremely convenient gauge has some topological limitations. We shall discuss now both the derivation of (9.109) and these limitations.

The first naive argument which shows that (9.109) is possible is the following. The possibility of the choice (9.109) means that any metric  $h_{ab}$  can be given in the form:

$$g_{ab}(\xi) = (e^{\varphi(\xi)} \delta_{ab})^f = e^{\varphi(f(\xi))} \frac{\partial f^c}{\partial \xi^a} \frac{\partial f^c}{\partial \xi^b} \tag{9.110}$$

where  $\{f^a(\xi)\}$  defines the necessary coordinate transformation. Hence, the r.h.s. of (9.110) depends on 3 arbitrary functions  $f^1(\xi)$ ,  $f^2(\xi)$ ,  $\varphi(\xi)$ . But  $h_{ab}(\xi)$  also has three independent components. Therefore, the number of independent functions matches. However, this is not enough. We must show that the transformation (9.110) is nonsingular, i.e. the jacobian for passing to the  $(\varphi, f^a)$  variables is nonzero. To show this we shall consider a small variation of (9.110):

$$\delta g_{ab} = [\delta \varphi(\xi) h_{ab} + \nabla_a \omega_b + \nabla_b \omega_a]^f \tag{9.111}$$

where  $\omega^a = \delta f^a(f^{-1}(\xi))$  and we have used the equation (9.28). The nonsingular nature of the transformation (9.110) will be proved if for

any  $\delta g_{ab}$  we can find  $\delta\varphi$  and  $\omega$  such that (9.111) will hold. In other words, we must be able to solve the equation:

$$\delta\varphi(\xi)h_{ab} + \nabla_a\omega_b + \nabla_b\omega_a = \delta g_{ab}^{f^{-1}} \equiv \gamma_{ab} \quad (9.112)$$

or

$$(L\omega)_{ab} = \nabla_a\omega_b + \nabla_b\omega_a - h_{ab}\nabla^c\omega_c = \gamma_{ab} - \frac{1}{2}h_{ab}\gamma^c_c \quad (9.113)$$

which is obtained from (9.112) by subtracting the trace. The question, whether the conformal gauge is always accessible, is reduced now to the possibility of solving equation (9.113) which we shall rewrite symbolically:

$$L\omega = \gamma \quad (9.114)$$

Here we have denoted by  $L$  the differential operator, defined by (9.113) which takes vector fields into traceless tensors (notice that the number of independent components is the same). There exists a conjugate operator which acts in the opposite direction—transforming tensors into vectors. It is easy to realize that the equation (9.114) will be solvable if and only if the conjugate operator doesn't have zero modes. Indeed, let us multiply (9.114) by some tensor field  $f$ :

$$(f, L\omega) = (L^+f, \omega) = (f, \gamma) \quad (9.115)$$

where:

$$(L^+f_{ab}) = -\nabla^af_{ab}; \quad f^a_a = 0 \quad (9.116)$$

and scalar products are defined in a covariant way. We see that if  $f$  is a zero mode, i.e.

$$L^+f = 0 \quad (9.117)$$

then for such  $\gamma$  that  $(\gamma, f) \neq 0$  equation (9.114) is not solvable. Now, if zero modes are absent then from

$$L^+L\omega = L^+\gamma \quad (9.118)$$

it follows that

$$L\omega = \gamma \quad (9.119)$$

since otherwise  $L\omega - \gamma$  would be a zero mode.

The operator  $L^+L$  is a self-conjugate operator and (9.118) has a solution

$$\omega = \frac{1}{L^+L} L^+\gamma \quad (9.120)$$

provided that  $L$  doesn't have a zero mode either. If it has, then the solution of (9.118) still exists but is not unique. Namely in this case we have to define the Green function  $1/L^+L$  as the sum over nonzero modes only:

$$\langle \xi' | \frac{1}{L^+L} | \xi \rangle = \sum_{n \neq 0} \frac{\omega_n(\xi') \omega_n(\xi)}{E_n} \tag{9.121}$$

$$L^+L\omega_n = E_n\omega_n$$

and the general solution of (9.119) is:

$$\omega = \frac{1}{L^+L} (L^+\gamma) + \sum_{\alpha} c_{\alpha} \omega_{\alpha,0} \tag{9.122}$$

where  $\{\omega_{\alpha,0}\}$  is the set of zero modes:

$$L\omega_{\alpha,0} = 0 \tag{9.123}$$

and  $\{c_{\alpha}\}$  are arbitrary constants.

So, our conclusion is that zero modes of the operator  $L^+$  mean that the conformal gauge is not accessible, and zero modes of  $L$  that it is not unique.

The number of zero modes is regulated by the index theorem which we have mentioned in Chapter 6, and is closely connected with the topology of our manifold. We shall show this for the case of closed manifolds. The demonstration is based on the identity:

$$N_0(L) - N_0(L^+) = \text{Tr}(e^{-tL^+L} - e^{-tLL^+}) \tag{9.124}$$

(where  $N_0$  is the number of zero modes) which in turn follows from the coincidence of nonzero eigenvalues of the operators  $L^+L$  and  $LL^+$ . From the first order equations:

$$L\varphi = \varepsilon\chi$$

$$L^+\chi = \varepsilon\varphi \tag{9.125}$$

we deduce:

$$L^+L\varphi = \varepsilon L^+\chi = \varepsilon^2\varphi$$

$$LL^+\chi = \varepsilon L\varphi = \varepsilon^2\chi \tag{9.126}$$

Therefore the only contribution to (9.124) comes from the zero modes. The right hand side of (9.124) can be evaluated by taking  $t \rightarrow 0$ . In this

limit it is a local, invariant expression depending on the metric tensor  $h_{ab}$ . If the metric were Euclidean then

$$\begin{aligned} (L^+ L \omega)_a &= -\nabla^b (\nabla_a \omega_b + \nabla_b \omega_a - h_{ab} \nabla^c \omega_c) \\ &= -\partial^2 \omega_a \end{aligned} \quad (9.127)$$

( $h_{ab} = \delta_{ab}$ )

and

$$\langle \xi | e^{-tL^+ L} | \xi \rangle = 2 \int \frac{d^2 p}{(2\pi)^2} e^{-tp^2} = \frac{1}{2\pi t} \quad (9.128)$$

where the coefficient 2 comes from the number of the components of  $\omega$ . As  $t \rightarrow 0$  the external gravitational field does not have time to influence (9.128). As we see from this formula the characteristic intervals for the motion of a "particle" described by the wave equation (9.127) are  $\Delta \xi \sim 1/p \sim \sqrt{t}$  which is just a diffusion law. The curvature  $R(\xi)$  will have a considerable effect only if:

$$R(\xi)(\Delta \xi)^2 \sim tR(\xi) \sim 1 \quad (9.129)$$

This consideration makes plausible the statement that for  $t \rightarrow 0$  the expansion parameter in  $\langle \xi | e^{-tL^+ L} | \xi \rangle$  is  $t \cdot R$ . We expect that:

$$\begin{aligned} \langle \xi | e^{-tL^+ L} | \xi \rangle &= \frac{1}{2\pi t} \{1 + a_1 t R(\xi) + O(t^2)\} \\ \langle \xi | e^{-tLL^+} | \xi \rangle &= \frac{1}{2\pi t} \{1 + a_2 t R(\xi) + O(t^2)\} \end{aligned} \quad (9.130)$$

Later we shall derive these formulas explicitly and compute  $a_{1,2}$ . Now, substituting (9.130) into (9.124) we get

$$N_0(L) - N_0(L^+) = c \cdot \chi \quad (9.131)$$

where  $c = 2(a_1 - a_2)$  is still to be determined and

$$\chi = \frac{1}{4\pi} \int R \sqrt{h} d^2 \xi \quad (9.132)$$

is the Euler character of our manifold. Here, in order to find  $c$ , it is sufficient to consider a sphere as an example in which  $N_0(L)$  and  $N_0(L^+)$  can be explicitly found. Using formulas of Riemannian geometry one has for  $g_{ab} = e^\varphi \delta_{ab}$ :

$$\begin{aligned} \nabla_a \omega_b &= \partial_a \omega_b - \Gamma_{ab}^c \omega_c \\ \Gamma_{ab}^c &= \frac{1}{2} (\partial_a \varphi \delta_{bc} + \partial_b \varphi \delta_{ac} - \partial_c \varphi \delta_{ab}) \end{aligned} \quad (9.133)$$

Substituting these into equation (9.123) we reduce it to

$$\partial_z \varepsilon = 0 \tag{9.134}$$

Here we have introduced the notation:

$$\begin{aligned} \varepsilon &= \omega^1 + i\omega^2 = e^{-\varphi}(\omega_1 + i\omega_2) \\ z &= \xi^1 + i\xi^2, \quad \partial_{\bar{z}} = \frac{1}{2} \left( \frac{\partial}{\partial \xi^1} + i \frac{\partial}{\partial \xi^2} \right) \end{aligned} \tag{9.135}$$

At first glance there is a continuum number of solutions of (9.134):

$$\varepsilon = \varepsilon(z) \tag{9.136}$$

However, we have to consider only normalizable ones:

$$\begin{aligned} \|\omega_a\|^2 &= \int \sqrt{h_{ab}} \omega^a \omega^b d^2 \xi \\ &= \int e^{2\varphi} |\varepsilon|^2 d^2 z < \infty \end{aligned} \tag{9.137}$$

The standard metric for a sphere in stereographic coordinates is

$$ds^2 = \frac{dz d\bar{z}}{(1 + |z|^2)^2}; \quad e^\varphi \underset{|z| \rightarrow \infty}{\sim} |z|^{-4} \tag{9.138}$$

Therefore the only possible function (9.136) with a finite norm is:

$$\varepsilon(z) = \alpha + \beta z + \gamma z^2 \tag{9.139}$$

Therefore  $N_0(L) = \sigma$  (since  $\alpha, \beta, \gamma$  are complex). The geometrical meaning of these solutions is clear—they are transformations which leave the original metric conformal. Therefore (9.139) describes the two-dimensional conformal group  $O(2, 1)$ .

As far as  $L^+$  is concerned, the same reasoning leads to:

$$\begin{aligned} \nabla^a \gamma_{ab} &= \gamma^a_a = 0, \quad \text{or} \quad \partial_{\bar{z}} \phi = 0 \\ \phi &= e^{-\varphi}(\gamma_{11} - \gamma_{22} + 2i\gamma_{12}) \end{aligned} \tag{9.140}$$

and the norm is defined as:

$$\begin{aligned} \|\gamma\|^2 &= \int \sqrt{h} h^{aa'} h^{bb'} \gamma_{ab} \gamma_{a'b'} d^2 \xi \\ &= \int e^{-\varphi} |\phi|^2 d^2 z \end{aligned}$$

Since  $e^{-\varphi} \underset{z \rightarrow \infty}{\sim} |z|^4$ , we conclude that the operator  $L^+$  does not have normalizable zero modes.

Recalling at last that a sphere has Euler character  $\chi = 2$  we obtain:

$$N_0(L) - N_0(L^+) = 3\chi \quad (9.141)$$

Actually, this result can be strengthened. Let us show, that for  $\chi < 0$  (a sphere with more than one handle)  $N_0(L) = 0$ . For this we compute:

$$\begin{aligned} (L^+ L\omega)_a &= -\nabla^b(\nabla_a \omega_b + \nabla_b \omega_a - h_{ab} \nabla^c \omega_c) \\ &= -\nabla^2 \omega_a - [\nabla^b, \nabla_a] \omega_b \\ &= -(\nabla^2 + \frac{1}{2}R)\omega_a \end{aligned} \quad (9.142)$$

For the manifolds with  $\chi < 0$  one can consider a metric with constant  $R < 0$  everywhere. For such a metric we have:

$$(L\omega, L\omega) = (\nabla_b \omega_a, \nabla_b \omega_a) - \frac{1}{2}R(\omega_a, \omega_a) > 0 \quad (9.143)$$

(where scalar products are understood in the same way as above). From this inequality we conclude that  $N_0(L) = 0$ . This means that:

$$N_0(L^+) = -3\chi = 6g - 6 \quad (\text{for } g \geq 2) \quad (9.144)$$

where  $g$  is the number of handles. It is easy to check directly, that on a torus ( $g = 1, \chi = 0$ )  $N_0(L) = N_0(L^+) = 2$ .

So, we have found that on a sphere we can always introduce a conformal gauge, which is defined modulo  $SL(2, C)$  transformations (9.139), which require extra gauge fixing. In the case of manifolds with higher topologies we have topological obstructions for the conformal gauge. The best thing which can be done is the following choice of gauge:

$$h_{ab}(\xi) = e^{\varphi(\xi)} h_{ab}^{(0)}(\xi; \tau_1, \dots, \tau_{6g-6}) \quad (9.145)$$

where  $h_{ab}^{(0)}$  is some metric, which can be chosen to have constant negative curvature and which depends on  $6g - 6$  extra parameters. Integration over all metrics must include not only functional integration over  $\varphi(\xi)$  but also the  $6g - 6$  dimensional integral over  $\{\tau_i\}$ . The theory of such integrations is not well developed, but we shall explain here a qualitative meaning of these extra parameters.

Let us first consider a torus. It can be represented as a parallelogram in the  $\xi$ -plane for which opposite sides are identified. In general, this figure can be mapped by a conformal transformation onto any other, say onto a square. However, the identification of the opposite sides will be lost. If we insist on preserving the identifications, the only options for a conformal map onto another parallelogram will be rigid rotations and scale transformations. Hence each parallelogram representing the torus can be characterized by the two conformal invariants—the ratio

of its sides and the angle between them. They are just the extra parameters, described above. One of them can be interpreted as a “length” of the torus, while the other is the angle which determines some canonical directions on it.

To finish this section, let us discuss briefly what happens in the case of surfaces with a boundary, taking the topology of a disc as an example. The analysis again rests on equation (9.114) but in this case we have to think of boundary conditions. If we try to consider transformations  $\xi \rightarrow f(\xi)$  which do not change the boundary, i.e.:

$$f(\xi_0(s)) = \xi_0(s) \tag{9.146}$$

(where  $\xi = \xi_0(s)$  is the equation for the boundary) we shall find that this requires the boundary condition:

$$\omega^a(\xi(s)) = 0 \tag{9.147}$$

However, it is not possible for general  $\gamma$  in (9.114) to find such a solution because it is a first-order differential equation. Equation (9.118) could have been solved with conditions (9.147) since it is a standard Dirichlet problem, but its solution (9.120) will not satisfy (9.114). This means that we cannot reach the conformal gauge by transformations with the condition (9.146), and we have to weaken this condition.

What is possible, is to use the transformations which reparametrize the boundary:

$$f^a(\xi_0(s)) = \xi_0^a(\alpha(s)) \tag{9.148}$$

but leave the shape of the disc unchanged. In terms of  $\omega$ , that means that:

$$\begin{aligned} \omega_\perp(\xi_0(s)) &\equiv n_a(s)\omega^a(\xi_0(s)) = 0 \\ \omega_\parallel(\xi_0(s)) &\equiv t_a(s)\omega^a(\xi_0(s)) - \text{unconstrained} \end{aligned} \tag{9.149}$$

(where  $n$  and  $t$  are the tangent and normal vectors of the boundary).

In this case it is easy to solve equation (9.119) explicitly. According to (9.134) it takes the form:

$$\begin{aligned} \partial_{\bar{z}}\varepsilon &= \gamma(z, \bar{z}) \\ \varepsilon_\perp &= \text{Re}(e^{-is}\varepsilon(e^{is})) = 0 \end{aligned} \tag{9.150}$$

where we have parametrized the boundary of the disc by

$$z_0(s) = e^{is}, \quad 0 \leq s < 2\pi$$

The general solution of (9.150) is given by:

$$\varepsilon(z, z) = \partial_z \int G(z, w) \gamma(w) d^2w + f(z) \quad (9.151)$$

where  $f(z)$  must be adjusted so as to satisfy the boundary conditions and  $G(z, w)$  is a Green function for the Laplacian.

The solution of this problem always exists, but is it unique? To answer this question we have to look at the homogeneous equation:

$$\begin{aligned} \partial_{\bar{z}} \varepsilon &= 0 \\ \varepsilon_{\perp} &= \text{Re}(e^{-is} \varepsilon(e^{is})) = 0 \end{aligned} \quad (9.152)$$

It is easy to check that this has three linearly independent solutions:

$$\varepsilon^{(1)} = iz, \quad \varepsilon^{(2)} = 1 - z^2, \quad \varepsilon^{(3)} = i(1 + z^2) \quad (9.153)$$

which can be added to (9.151). These solutions are infinitesimal conformal transformations of the  $SL(2, R)$  group which map the unit disc onto itself. The finite version of these maps is given by:

$$z \rightarrow \tilde{z} = e^{i\alpha} \frac{z - a}{1 - \bar{a}z} \quad (9.154)$$

(where  $\alpha$  is a real phase and  $a$  a complex number).

Notice, that after the solution of (9.150) is found, then  $\omega_{\parallel}(\xi_0(s))$  is uniquely defined (modulo  $SL(2, R)$  transformations).

Our conclusion for the case of the unit disc is thus the following. The conformal gauge is accessible provided we include diffeomorphisms which reparametrize the boundary. This reparametrization, defined modulo  $SL(2, R)$  transformations, is determined by the original metric  $h_{ab}(\xi)$ . Since in our original formulation of the functional integral (9.76) we factored out the diffeomorphisms which become identical at the boundary, we have to expect that the integration will be reduced to the form:

$$\frac{\mathcal{D}h_{ab}(\xi)}{\mathcal{D}f(\xi)} = \mathcal{D}\varphi(\xi) \mathcal{D}\alpha(s) \times (\text{jacobian}) \quad (9.155)$$

We shall compute the jacobian later, by now it is important to realize that the integration over all metrics must include not only the  $\varphi$ -integration, but also integration over all possible reparametrizations  $\alpha(s)$ . In a certain sense  $\{\alpha(s)\}$  replaces the discrete set of parameters which we had for closed surfaces with complicated topology.

### 9.6 Computation of Functional Integrals

The aim of this section is to compute the integral:

$$G[c(s)] = \int \frac{\mathcal{D}h_{ab}(\xi)}{\mathcal{D}f(\xi)} \exp(-\mu \int h^{1/2} d^2\xi) \times \int_{x(\xi(s))=c(s)} \mathcal{D}x(\xi) \exp\left(-\int h^{1/2} h^{ab} \partial_a x \cdot \partial_b x d^2\xi\right) \quad (9.156)$$

to which the problem of random surfaces has been reduced.

We shall begin with the  $x$ -integration. The reason why it can be performed is easy to see. If we substitute  $h_{ab}$  in the conformal gauge  $h_{ab} = e^\varphi \delta_{ab}$  into (9.156) we see that the  $\varphi$ -dependence disappears, indicating conformal invariance of the action. However, since we are dealing with quantum theory, there is a conformal anomaly which will bring  $\varphi$ -dependence back. The anomaly is due to the regularization (which breaks conformal invariance) of (9.156) at small distances, which is needed to avoid divergences. Since it is an important point we shall derive it by several different approaches. First, let us consider the case of weak gravitational fields:

$$h_{ab} = \delta_{ab} + h_{ab}^{(1)} \quad (9.157)$$

In the case of infinite system, the induced action in quadratic approximation will be given by:

$$S_{II} = -\frac{1}{2} \int \frac{d^2q}{(2\pi)^2} h_{ab}^{(1)}(q) h_{cd}^{(1)}(-q) \Pi_{ab|cd}(q)$$

$$\Pi_{ab|cd}(q) = \text{---} \begin{array}{c} k+q \\ \circ \\ k \end{array} \text{---} = \frac{\mathcal{D}}{8} \int \frac{d^2k}{(2\pi)^2} \frac{t_{ab}(k, q) t_{cd}(k, q)}{k^2(k+q)^2} \quad (9.158)$$

$$t_{ab}(k, q) = k_a(k+q)_b + k_b(k+q)_a - \delta_{ab} k \cdot (k+q)$$

(the vertex  $t_{ab}$  is easily read off from (9.104)).

The integral (9.158) is divergent and must be regularized. This means that the integral itself defines only the part of  $\Pi(q)$  which has singularities in  $q$  or in other words, propagates in  $\xi$ -space, while the regularization adds to this expression arbitrary (*a priori*) polynomials in  $q$  or local expressions in  $\xi$ -space. The condition which fixes these local parts

uniquely is gauge invariance of the effective action. For the actual computation it is convenient to introduce complex notation:

$$\begin{aligned} k_{\pm} &= k_1 \pm ik_2 = |k|e^{\pm i\alpha} \\ k^2 &= k_+ k_- \end{aligned} \quad (9.159)$$

and to consider first the component:

$$\begin{aligned} \Pi_{++++} &= \frac{\mathcal{D}}{2} \int \frac{d^2k}{(2\pi)^2} \frac{k_+^2 (q_+ + k_+)^2}{k^2 (q_+ + k)^2} \\ &= \frac{\mathcal{D}}{2} \int \frac{d^2k}{(2\pi)^2} \frac{k_+ (q_+ + k_+)}{k_- (q_- + k_-)} = \mathcal{D} C(q^2) q_+^4 \end{aligned} \quad (9.160)$$

with as yet unknown  $C(q^2)$ ; we have used the  $O(2)$  invariance of the integral and the fact that  $1/k_-$  transforms as  $k_+$ . It is tempting to split the integral (9.160) into (+) and (-) factors since  $2\pi i d^2k = dk_+ dk_-$ . This is almost possible but requires some extra care. Oddly, one has to go to Minkowskian space. The propagators and momenta change as:

$$\begin{aligned} k_{\pm} &\rightarrow k_0 \pm k_1 \\ \frac{1}{k^2} &\rightarrow \frac{1}{k^2 + i\epsilon} \\ \frac{1}{k_-} &\rightarrow \frac{k_+}{k^2 + i\epsilon} = \frac{1}{k_- + i\epsilon \operatorname{sign} k_+} \end{aligned} \quad (9.161)$$

So our integral has the form:

$$\begin{aligned} \Pi_{++++} &= \int_{-\infty}^{\infty} dk_+ k_+ (q_+ + k_+) \frac{\mathcal{D}}{16i\pi^2} \\ &\times \int_{-\infty}^{+\infty} \frac{dk_-}{(k_- + i\epsilon \operatorname{sign} k_+) (q_- + k_- + i\epsilon \operatorname{sign}(q_+ + k_+))} \end{aligned} \quad (9.162)$$

If we take  $q_+ < 0$  then the singularities in  $k$ -space are on opposite sides of the real axis only if  $0 < k_+ < -q_+$ . Otherwise the  $k_-$ -integral gives zero. We have:

$$\begin{aligned} \Pi_{++++} &= -\frac{\mathcal{D}}{8} \frac{1}{q_-} \int_0^{-q_+} dk_+ k_+ (q_+ + k_+) \\ &= \frac{\mathcal{D}}{48\pi} \frac{q_+^3}{q_-}, \end{aligned} \quad (9.163)$$

which is the desired answer.

The finiteness of the integral (9.160) also allows one to use the following trick to compute it:

$$\begin{aligned}
 (9.160) &= \frac{\mathcal{D}}{2} \int \frac{d^2k}{(2\pi)^2} \frac{k_+(k_+ + q_+)}{k_-(k_- + q_-)} e^{-\epsilon k^2} \Big|_{\epsilon \rightarrow 0} \\
 &= \frac{\mathcal{D}}{2q_-} \int \frac{d^2k}{(2\pi)^2} e^{-\epsilon k^2} k_+(k_+ + q_+) \left( \frac{1}{k_-} - \frac{1}{k_- + q_-} \right) \Big|_{\epsilon \rightarrow 0} \\
 &= -\frac{\mathcal{D}}{2q_-} \int \frac{d^2k}{(2\pi)^2} e^{-\epsilon(k-q)^2} \frac{(k-q)_+ k_+}{k_-} \Big|_{\epsilon \rightarrow 0} \\
 &= -\frac{\mathcal{D}}{2q_-} \int \frac{d^2k}{(2\pi)^2} e^{-\epsilon k^2} e^{\epsilon k \cdot q_+} \frac{k_+^2 - k_+ q_+}{k_-} \Big|_{\epsilon \rightarrow 0} \\
 &= -\frac{\mathcal{D}}{2q_-} \int \frac{d^2k}{(2\pi)^2} e^{-\epsilon k^2} \left( \frac{1}{6} \epsilon^3 q_+^3 k^4 - \frac{1}{2} q_+^3 \epsilon^2 k^2 \right) \Big|_{\epsilon \rightarrow 0} \\
 &= \frac{\mathcal{D} q_+^3}{16\pi q_-} \int_0^\infty dt e^{-t} (t - t^2/3) = \frac{\mathcal{D} q_+^3}{48\pi q_-}
 \end{aligned}$$

Obviously this also implies:

$$\Pi_{-,-,-,-} = \frac{\mathcal{D}}{48\pi} \frac{q_-^3}{q_+} \quad (9.164)$$

Let us now consider the remaining component

$$\begin{aligned}
 \Pi_{+,-,-,-} &= \frac{\mathcal{D}}{2} \int \frac{d^2k}{k^2(q+k)^2} k_+(q+k)_+ k_-(q+k)_- \\
 &= \int d^2k = A \quad (9.165)
 \end{aligned}$$

where  $A$  is a quadratically divergent constant which we shall fix in a moment. So the general form of the induced action is the following:

$$\begin{aligned}
 S_{\Pi} &= -\frac{\mathcal{D}}{48\pi} \int \frac{d^2q}{(2\pi)^2} \left\{ \frac{1}{2} \frac{q_+^3}{q_-} h_{--}(q) h_{--}(-q) \right. \\
 &\quad + \frac{1}{2} \frac{q_-^3}{q_+} h_{++}(q) h_{++}(-q) + A q^2 h_{++}(q) h_{--}(-q) \\
 &\quad \left. + B q^2 h_{+-}(q) h_{+-}(-q) + C [h_{+-} q_-^2 h_{++} + q_+^2 h_{+-} h_{--}] \right\} \quad (9.166)
 \end{aligned}$$

where we have added another possible local counter-terms with some constants  $A, B$  and  $C$ . The constants  $A, B$  and  $C$  can be determined from

the condition that the induced action is generally covariant. The transformation law for  $h_{ab}$  is:

$$h_{ab} \rightarrow h_{ab} + \nabla_a \omega_b + \nabla_b \omega_a \quad (9.167)$$

which in the linearized approximation takes the form:

$$\begin{aligned} h_{++} &\rightarrow h_{++} + 2q_+ \omega_+(q) \\ h_{--} &\rightarrow h_{--} + 2q_- \omega_-(q) \\ h_{+-} &\rightarrow h_{+-} + q_+ \omega_-(q) + q_- \omega_+(q) \end{aligned} \quad (9.168)$$

An important point is that although the nonlocal part in (9.166) is not gauge invariant, its gauge variation appears to be local and hence can be compensated by the gauge variation of the local part. We have the only gauge invariant choice:

$$S_{II} = -\frac{\mathcal{D}}{48\pi} \int \frac{d^2q}{(2\pi)^2} R(q) \frac{1}{2q^2} R(-q) \quad (9.169)$$

where:

$$\begin{aligned} R(q) &= \frac{1}{4}(q_+^2 h_{--} + q_-^2 h_{++} - 2q^2 h_{+-}) \\ &= (q_\alpha q_\beta - q^2 \delta_{\alpha\beta}) h^{\alpha\beta} \end{aligned} \quad (9.170)$$

(it is easy to recognize in  $R(q)$  a linearized scalar curvature made of  $h_{ab}$ ).

In the conformal gauge:

$$h_{++} = h_{--} = 0, \quad h_{+-} \approx 2\varphi \quad (9.171)$$

and

$$S_{II} = -\frac{\mathcal{D}}{96\pi} \int \frac{d^2q}{(2\pi)^2} q^2 \varphi(q) \varphi(-q) \quad (9.172)$$

The dependence on  $\varphi$  which was originally absent in the action has appeared here through the following mechanism. The induced action contained  $\varphi$ -independent nonlocal terms which were not gauge invariant. We were forced to add a certain combination of local terms which restore the gauge invariance. These terms were  $\varphi$ -dependent. It is quite obvious that in higher orders the nonlocal part of the induced action will also be  $\varphi$ -independent. From this follows an important conclusion: in the conformal gauge the induced action is local. This can be considered as a manifestation of the fact that the original action was conformally invariant and the conformal anomaly arises only because of the small distance cut-off, the effect of which must necessarily be local.

The locality of the action, together with invariance properties, permits us to determine its exact form. To do this, let us notice that the conformally Euclidean metric

$$ds^2 = e^{\varphi(z, \bar{z})} dz d\bar{z} \tag{9.173}$$

preserves its form under analytic changes of coordinates:

$$\begin{aligned} z &\rightarrow w(z) \\ \bar{z} &\rightarrow \bar{w}(\bar{z}) \end{aligned} \tag{9.174}$$

$$ds^2 = e^{\varphi(z, \bar{z})} dz d\bar{z} \rightarrow e^{\varphi(z(w), \overline{z(w)})} \left| \frac{dz}{dw} \right|^2 dw d\bar{w}$$

Therefore, in the conformal gauge, the induced action must be a local function of  $\varphi$  invariant under the transformation:

$$\varphi \rightarrow \tilde{\varphi}(z, \bar{z}) = \tilde{\varphi}(w(z), \overline{w(z)}) + \log \left| \frac{dw}{dz} \right|^2 \tag{9.175}$$

where  $w(z)$  is an arbitrary analytic function. The invariance is correct only modulo boundary terms since the shape of the domain inevitably changes under the analytic map  $w$ . We shall postpone the discussion of boundary effects. As far as the bulk part of the action is concerned, the only option consistent with the above requirement is

$$S[\varphi] = A \int d^2z (2\partial_z \varphi \partial_{\bar{z}} \varphi + \mu^2 e^\varphi) \tag{9.176}$$

(where  $A$  is some constant).

We will call  $S$  the Liouville action.

The invariance (9.175) of (9.176) is quite obvious since

$$\partial_z \partial_{\bar{z}} \log \left| \frac{dw}{dz} \right|^2 = 0$$

Any other invariant expression would have a higher number of derivatives and so can be dropped in the continuum limit.

Comparing (9.176) with the approximate expression (9.172) one gets:

$$A = - \frac{\mathcal{D}}{48\pi} \tag{9.177}$$

Now, it remains to compute the Jacobian in (9.155), and the problem of the distribution of random surfaces in internal geometries will be

solved. The simplest way to find the jacobian is to consider the norm in the space of all metrics:

$$\|\delta h\|^2 = \int h^{1/2}(h^{aa'}h^{bb'} + ch^{ab}h^{a'b'})\delta h_{ab}\delta h_{a'b'} \quad (9.178)$$

and to substitute

$$h^{ab} = e^{-\varphi}\delta^{ab}; \quad h^{1/2} = e^\varphi$$

$$\delta h_{ab} = (\delta\varphi + \nabla^a \varepsilon_a)h_{ab} + (L\varepsilon)_{ab}$$

(where  $L$  was defined in (9.113).

We obtain:

$$\|\delta h\|^2 = \int e^\varphi(\delta\varphi + \nabla^a \varepsilon_a)^2(2 + 4c) d^2\xi + (L\varepsilon, L\varepsilon) \quad (9.179)$$

From here it follows:

$$\mathcal{D}h = \mathcal{D}\varphi\mathcal{D}\varepsilon \det^{1/2}(L^+L)$$

$$(L^+L\omega)_a = \nabla^b(\nabla_a\omega_b + \nabla_b\omega_a - h_{ab}\nabla^c\omega_c) \quad (9.180)$$

We perform now the computation of  $\det(L^+L)$  in a way similar to the one we used for the ordinary Laplacian. Namely, let us again take a weak gravitational field. The formal quadratic form, from which the operator (9.180) arises as a kernel, can be written as:

$$W = \int h^{1/2}h^{aa'}h^{bb'}\phi_{ab}\phi_{a'b'} d^2\xi$$

$$\phi_{ab} = \frac{1}{2}(\nabla_a\omega_b + \nabla_b\omega_a - h_{ab}\nabla^c\omega_c) \quad (9.181)$$

Now, we shall compute the polarization operator  $\Pi_{--}$  as in (9.164). For doing this it is sufficient to take  $h_{ab}$  in the form:

$$\|h_{ab}\| = \begin{pmatrix} h_{++} & 2 \\ 2 & 0 \end{pmatrix} \quad (9.182)$$

with small  $h_{++}$ . As we have seen in the scalar case, the general determinant can be reconstructed from this small  $h$ -field limit. Standard formulae for covariant derivatives give:

$$\begin{aligned} \phi_{++} &\approx \nabla_+\omega_+ - \frac{1}{4}(\nabla_+\omega_- + \nabla_-\omega_+)h_{++} \\ &\approx \partial_+\omega_+ + \frac{1}{4}(\partial_-h_{++})\omega_+ - \frac{1}{4}(\partial_+h_{++})\omega_- \\ &\quad - \frac{1}{4}h_{++}(\partial_+\omega_- + \partial_-\omega_+) \\ \phi_{--} &\approx \partial_-\omega_- \end{aligned} \quad (9.183)$$

Hence:

$$\begin{aligned}
 W = \int & [(\partial_+ \omega_+) (\partial_- \omega_-) + \frac{1}{4} (\partial_- h_{++}) (\omega_+ \partial_- \omega_-) \\
 & - \frac{1}{4} (\partial_+ h_{++}) \omega_- \partial_- \omega_- - \frac{1}{4} h_{++} (\partial_+ \omega_-) (\partial_- \omega_-) \\
 & - \frac{1}{4} h_{++} (\partial_- \omega_+) (\partial_- \omega_-)] d^2 \xi \tag{9.184}
 \end{aligned}$$

Since only the  $\langle \omega_+ \omega_- \rangle$  propagator is nonzero, the third and the fourth terms do not contribute.

We are left with

$$\begin{aligned}
 \det^{-1/2}(L^+ L) = \int \mathcal{D} \omega_{\pm} \exp \left\{ - \int d^2 \xi [ \partial_+ \omega_+ \partial_- \omega_- \right. \\
 \left. + \frac{1}{4} (\partial_- h_{++}) \omega_+ \partial_- \omega_- - \frac{1}{4} h_{++} (\partial_- \omega_+) \partial_- \omega_- ] \right\} \tag{9.185}
 \end{aligned}$$

From this it is easy to obtain the quadratic term of the effective action for  $h_{++}$ :

$$\begin{aligned}
 S_{II} = - \frac{1}{2} \int \frac{d^2 q}{(2\pi)^2} h_{++}(q) h_{++}(-q) \Pi^{gh}_{--}(q); \\
 \Pi^{gh}_{--}(q) = \begin{array}{c} k+q \\ \circlearrowleft \\ k \end{array} + \partial_- \begin{array}{c} k+q \\ \circlearrowleft \\ k \end{array} + \partial_- \begin{array}{c} k+q \\ \circlearrowleft \\ k \end{array} \partial_- \\
 = \int \frac{d^2 k}{(2\pi)^2} \left( \frac{1}{16} \frac{k_-^2 (k+q)_-^2}{k^2 (k+q)^2} + 2 \cdot \frac{1}{16} q_- \frac{k_-^2 (k+q)_-}{k^2 (k+q)^2} \right. \\
 \left. - \frac{1}{16} q_-^2 \frac{k_- (k+q)_-}{k^2 (k+q)^2} \right) \tag{9.186}
 \end{aligned}$$

A computation similar to (9.160)–(9.163) gives the result

$$\Pi^{gh}_{--}(q) = \frac{q_-^3}{16q_+} \left( \frac{1}{24\pi} + \frac{1}{4\pi} + \frac{1}{4\pi} \right) = \frac{26}{48\pi} \frac{q_-^3}{16q_+} \tag{9.187}$$

Substituting into (9.186) and comparing with (9.165) we conclude that the role of  $\mathcal{D}$  is now played by the number 26 and in the gauge (9.180)

$$\det^{1/2}(L^+ L) = \exp \left\{ - \frac{26}{48\pi} \int d^2 \bar{z} (2\partial_z \varphi \partial_{\bar{z}} \varphi + \bar{\mu}^2 e^{\varphi}) \right\} \tag{9.188}$$

Collecting now (9.176), (9.177), (9.180) and (9.188) we get the final expression for the distribution of the surfaces in internal metrics in (9.156):

$$Z^{\text{closed surf.}} = \int \mathcal{D} \varphi \exp \left\{ - \frac{26 - \mathcal{D}}{48\pi} \int d^2 \xi \left( \frac{1}{2} (\partial_a \varphi)^2 + m^2 e^{\varphi} \right) \right\} \tag{9.189}$$

### 9.7 Scattering Amplitudes

In the case of paths we have demonstrated how to compute Green functions and amplitudes i.e. the amplitudes for a path to pass through a set of given points. The answer was given by Feynman diagrams as it should be. What shall we obtain in the case of surface, now that we have learned how to compute the functional integral?

Let us consider a surface with pinned points  $\{x_i\}$ . The corresponding amplitude is given by:

$$G(x_1, \dots, x_N) = \langle \prod_i \int h^{1/2}(\xi_i) \delta(x_i - x(\xi_i)) d^2 \xi_i \rangle \quad (9.190)$$

(where the average is taken with respect to the random surface functional measure). By passing to the momentum representation, we rewrite (9.190) as:

$$G(p_1, \dots, p_N) = \langle \prod_i \int h^{1/2}(\xi_i) e^{i p_i \cdot x(\xi_i)} d^2 \xi_i \rangle \quad (9.191)$$

A nice thing about this formula is that the  $x$ -integration in it remains Gaussian and can be easily performed. Namely, we have to compute:

$$\begin{aligned} G(p_1, \dots, p_N) &= \int \frac{\mathcal{D}h_{ab}(\xi)}{\mathcal{D}f(\xi)} \exp\left(-\lambda_0 \int h^{1/2} d^2 \xi\right) \int \prod_j d^2 \xi_j h^{1/2}(\xi_j) \\ &\times \int \mathcal{D}x(\xi) \exp\left\{-\frac{1}{16\pi} \int h^{1/2} h^{ab} \partial_a x \cdot \partial_b x + i \sum_{j=1}^N p_j \cdot x(\xi_j)\right\} \end{aligned} \quad (9.192)$$

(The coefficient  $1/16\pi$  is a convenient  $x$ -normalization). Passing to the conformal gauge, and writing  $x = x_{cl} + y$  where  $x_{cl}$  satisfies

$$-\frac{1}{8\pi} \partial^2 x_{cl} = i \sum_j p_j \delta(\xi - \xi_j) \quad (9.193)$$

$$x_{cl} = i \sum_j \mathcal{D}(\xi | \xi_j) p_j = -i \sum_j p_j \log |\xi - \xi_j|^2$$

(Here  $\mathcal{D}(\xi | \xi')$  is the Green function of the Laplacian). We obtain:

$$\begin{aligned} G(p_1, \dots, p_N) &= \int \mathcal{D}\varphi(\xi) \exp\left[-\frac{1}{\gamma} \int ((\partial\varphi)^2 + \mu_0 e^\varphi) d^2 \xi\right] \\ &\times \int \prod_j d^2 \xi_j \exp(\varphi(\xi_j) - \frac{1}{2} p_j^2 \mathcal{D}(\xi_j | \xi_j)) \\ &\times \exp\left\{\sum_{i < j} 2 p_i \cdot p_j \log |\xi_i - \xi_j|\right\} \end{aligned} \quad (9.194)$$

The quantity  $\mathcal{D}(\xi|\xi)$  is singular and an explanation of its meaning will be given shortly below. Before doing this let us rewrite (9.194) in the form:

$$G(p_1, \dots, p_N) = \int \prod_j d^2\xi_j \prod_{i < j} |\xi_i - \xi_j|^{2p_i \cdot p_j} \times F(p_1^2, \dots, p_N^2; \xi_1, \dots, \xi_N) \tag{9.195}$$

where  $F$  is given by a Liouville functional integral. We have to expect that  $F$  has poles in  $p_j^2$  corresponding to the physical spectrum of the string. The on-shell scattering amplitude will be given by the residue at these poles. Denoting

$$\phi(\xi_1, \dots, \xi_N) = \text{res}_{p_j^2 = -m^2} F(p_1^2, \dots, p_N^2; \xi_1, \dots, \xi_N) \tag{9.196}$$

(where  $m^2$  is the position of the presumed pole and the space-time signature is chosen such that  $p^2 = p_i^2 - p_0^2 = -m^2$ ) we obtain an expression for the scattering amplitude  $\mathcal{A}(p_1, \dots, p_N)$ :

$$\mathcal{A}(p_1, \dots, p_N) = \int \prod_{j=1}^N d^2\xi_j \prod_{i < j} |\xi_i - \xi_j|^{2p_i \cdot p_j} \phi(\xi_1, \dots, \xi_N) \tag{9.197}$$

Even without knowledge of  $\phi$  this formula contains information, since  $\phi$  depends on fewer variables than  $A$ . But what is the meaning of  $\phi$ ? To be sure it is some correlation function of the Liouville field theory, and so in principle we have reduced the problem of finding the  $\mathcal{D}$ -dimensional scattering amplitude to two-dimensional theory. In order to find the kind of the correlation functions we need let us go back to (9.195) and try to decipher  $\mathcal{D}(\xi|\xi)$  in it. It is clear that we have to use a cut-off in this expression which does not destroy general covariance. A suitable choice is

$$\mathcal{D}(\xi|\xi) = \sum_n \frac{\psi_n^2(\xi)}{\lambda_n} e^{-\epsilon\lambda_n} \tag{9.198}$$

where  $\psi_n(\xi)$  and  $\lambda_n$  are eigenfunctions and eigenvalues of the laplacian. This choice is the same one which we used in defining determinants (which were regularized by the Schwinger proper time cut-off) and is equivalent to the use of Pauli-Villars regulators. There is a simple naive way to compute (9.198). It consists of the observation that:

$$\mathcal{D}(\xi|\xi) = -\log(\Delta\xi)_{\min}^2 \tag{9.199}$$

where  $(\Delta\xi)_{\min}^2$  is the cut-off in the  $\xi$ -space. However, this cut-off is  $\xi$ -dependent, since we have to fix the minimal invariant interval  $(\Delta s)_{\min}^2$ .

We have:

$$\begin{aligned}(\Delta s)_{\min}^2 &= e^{\varphi(\xi)}(\Delta \xi)_{\min}^2 = \epsilon \\ (\Delta \xi)_{\min}^2 &= \epsilon e^{-\varphi(\xi)}\end{aligned}\tag{9.200}$$

and on substituting this into (9.199) we get:

$$\mathcal{D}(\xi|\xi) = \log \frac{1}{\epsilon} + \varphi(\xi)\tag{9.201}$$

This formula looks suspicious, since the left-hand side must be scalar under conformal transformation:

$$\begin{aligned}\xi &\rightarrow f(\xi) \\ \mathcal{D}(\xi|\xi) &\rightarrow \mathcal{D}(f(\xi)|f(\xi))\end{aligned}\tag{9.202}$$

At the same time:

$$\varphi(\xi) \rightarrow \varphi(f(\xi)) + \log \left| \frac{df}{d\xi} \right|^2\tag{9.203}$$

This is slightly confusing. We can even strengthen this confusion by noticing that the Green function for non-coincident points,

$$\mathcal{D}(\xi_1|\xi_2) = -\log |\xi_1 - \xi_2|^2\tag{9.204}$$

is not scalar either, since, for example under the scale transformation  $\xi \rightarrow \lambda\xi$ , it changes. The origin of the trouble is that the Laplace operator on any compact manifold has a zero mode—the constant function. This zero mode, which is revealed as translation invariance  $x \rightarrow x + a$  in the integrand (9.192) must be carefully subtracted from the Green function. The Green function without zero mode  $\mathcal{D}'(\xi_1|\xi_2)$  is well defined and does not increase logarithmically as (9.204). It is easy to show that

$$\mathcal{D}'(\xi_1|\xi_2) = \mathcal{D}(\xi_1|\xi_2) + f(\xi_1) + f(\xi_2)\tag{9.205}$$

where  $f(\xi)$  is determined from the conditions that its Laplacian is constant and that it cancels the logarithmic increase at infinity. It is easy to give explicit formulas for  $f(\xi)$ , but they are not needed. Indeed:

$$\begin{aligned}\sum_{i,j} \mathbf{p}_i \cdot \mathbf{p}_j \mathcal{D}'(\xi_i|\xi_j) &= \sum_{i,j} \mathbf{p}_i \cdot \mathbf{p}_j \mathcal{D}(\xi_i|\xi_j) \\ &\quad + \sum_{i,j} \mathbf{p}_i \mathbf{p}_j (f(\xi_i) + f(\xi_j)) \\ &= \sum_{i,j} \mathbf{p}_i \cdot \mathbf{p}_j \mathcal{D}(\xi_i|\xi_j) + 2 \left[ \sum_i \mathbf{p}_i f(\xi_i) \right] \cdot \left( \sum_j \mathbf{p}_j \right) \\ &= \sum_{i,j} \mathbf{p}_i \cdot \mathbf{p}_j \mathcal{D}(\xi_i|\xi_j)\end{aligned}\tag{9.206}$$

where we have used momentum conservation. By its construction  $\mathcal{D}'$  has correct transformation properties, but we need not bother to replace  $\mathcal{D}$  by  $\mathcal{D}'$  in the previous formulas.

So, we obtain the following expression for the function  $\phi$  in (9.190):

$$\begin{aligned} \phi(\xi_1, \dots, \xi_N) &= \langle \psi_\Delta(\xi_1), \dots, \psi_\Delta(\xi_N) \rangle \\ \psi_\Delta &= e^{\Delta\varphi(\xi)}, \quad \Delta = 1 + m^2/2 \end{aligned} \tag{9.207}$$

The average in this formula is taken with respect to the Liouville action. The value of  $\Delta$  is determined from the condition that the function  $F(\mathbf{p}_1^2, \dots, \mathbf{p}_N^2; \xi_1, \dots, \xi_N)$  in (9.196) has a pole at  $\mathbf{p}_j^2 = -m^2$ . This correlation function can also be presented as a partition function of a punctured sphere on which the Liouville field  $\varphi(\xi)$  has parabolic singularities:

$$e^{\varphi(\xi)} \rightarrow \frac{1}{|\xi - \xi_i|^2 \log^2 |\xi - \xi_i|}$$

The next natural question which we shall address in the following section concerns the properties of the correlation functions (9.207) and their relation to the properties of the string scattering amplitudes (9.197).

### 9.8 Scattering Amplitudes and the Operator Product Expansion

The results of the preceding section can be summarized as follows. The amplitude is given by:

$$\begin{aligned} \mathcal{A}(\mathbf{p}_1, \dots, \mathbf{p}_N) &\sim \int \prod_{j=1}^N d^2\xi_j \langle V_{\mathbf{p}_1}(\xi_1) \dots V_{\mathbf{p}_N}(\xi_N) \rangle \\ V_{\mathbf{p}}(\xi) &= \psi_\Delta(\xi) : \exp(i\mathbf{p} \cdot \mathbf{x}(\xi)) : \end{aligned} \tag{9.208}$$

Here we have introduced a new notation,  $: \exp(i\mathbf{p} \cdot \mathbf{x}(\xi)) :$ . By it we mean that  $\mathbf{x}(\xi)$  is a free bosonic field and when computing averages of the product of different exponents, one has, while using the Wick theorem, to avoid pairing of  $x$  inside the symbol:  $\therefore$ . Therefore:

$$\begin{aligned} \langle : e^{i\mathbf{p}_1 \cdot \mathbf{x}(\xi_1)} : \dots : e^{i\mathbf{p}_N \cdot \mathbf{x}(\xi_N)} : \rangle \\ \stackrel{\text{def}}{=} \exp \left( 2 \sum_{i < j} \mathbf{p}_i \cdot \mathbf{p}_j \log |\xi_i - \xi_j| \right) \\ = \prod_{i < j} |\xi_i - \xi_j|^{2\mathbf{p}_i \cdot \mathbf{p}_j} \end{aligned} \tag{9.209}$$

From this we recover (9.197) (A helpful note: to compute such averages it is convenient to consider small  $p_i$  and to write

$$\begin{aligned} \left\langle \prod_j e^{i p_j \cdot x(\xi_j)} \right\rangle &\approx \left\langle \prod_j (1 + i p_j \cdot x(\xi_j)) \right\rangle \\ &= 1 - \sum_{i < j} p_i \cdot p_j \langle x(\xi_i) \cdot x(\xi_j) \rangle + O(p_i^4) \quad (9.210) \\ &\approx \exp\left(-\sum_{i < j} p_i \cdot p_j \langle x(\xi_i) \cdot x(\xi_j) \rangle\right) \end{aligned}$$

(because we know in advance that the result is Gaussian with respect to  $p_i$ ).

The correlation functions (9.209) have a remarkable property. They are covariant under the group of projective transformations  $SL(2, C)$ . Let us introduce complex variables  $z = \xi^1 + i\xi^2$  and  $\bar{z} = \xi^1 - i\xi^2$ , and check what happens with the product (9.207) if we change:

$$z \rightarrow \frac{\alpha z + \beta}{\gamma z + \delta} = w(z); \quad \alpha\delta - \gamma\beta = 1 \quad (9.211)$$

We have:

$$z_1 - z_2 \rightarrow \frac{z_1 - z_2}{(\gamma z_1 + \delta)(\gamma z_2 + \delta)}$$

And, hence:

$$\begin{aligned} \prod_{i < j} |\xi_i - \xi_j|^{2 p_i \cdot p_j} &= \prod_{i < j} ((z_i - z_j)(\bar{z}_i - \bar{z}_j))^{p_i \cdot p_j} \\ &\rightarrow \prod_{i < j} \left( \frac{1}{|\gamma z_i + \delta|^2 |\gamma z_j + \delta|^2} \right)^{p_i \cdot p_j} \prod_{i < j} |\xi_i - \xi_j|^{2 p_i \cdot p_j} \quad (9.212) \end{aligned}$$

The first factor has the structure which we have already encountered in (9.206), and we rewrite it as:

$$\begin{aligned} \prod_{i < j} \left( \left| \frac{dw}{dz}(z_i) \right| \left| \frac{dw}{dz}(z_j) \right| \right)^{p_i \cdot p_j} &= \exp\left( 2 \sum_{i < j} p_i \cdot p_j \log \left| \frac{dw}{dz}(z_i) \right| \right) \\ &= \prod_j \left| \frac{dw}{dz}(z_j) \right|^{-p_j^2} \quad (9.213) \end{aligned}$$

(where we have used momentum conservation  $\sum_j p_j = 0$ ). This result can be formalized as follows. Let us consider an operator:

$$U_p(z, \bar{z}) = :e^{i p \cdot x(z, \bar{z})}: \quad (9.214)$$

We have proved that its correlation functions are invariant under the transformation:

$$U_p(z, \bar{z}) \rightarrow \left(\frac{dw}{dz}\right)^{p/2} \left(\frac{\overline{dw}}{d\bar{z}}\right)^{p/2} U_p(w(z), \overline{w(z)}) \tag{9.215}$$

with

$$w(z) = \frac{\alpha z + \beta}{\gamma z + \delta}, \quad \alpha\delta - \gamma\beta = 1$$

We shall refer to such a situation by saying that the operator has anomalous dimension  $p^2$ , and that we are dealing with a conformal quantum field theory.

How did it happen, that starting from the field  $x$  which has dimension zero we obtained by exponentiating it a field with nonzero dimension? The “leak” of dimensionality occurs through the ultraviolet cut-off which we excluded by taking the normal product : :. Otherwise we would get coefficients containing  $\varepsilon^{-p^2}$  (where  $\varepsilon$  is the small distance cut-off) and the correlation functions are formally dimensionless, provided that we change the cut-off scale simultaneously. This scale transformation is not interesting. The physical rescaling does not change the cut-off. Generically, there could be no scale invariance at all in this case. But in many important cases, like the one we have discussed, it still exists although the transformation properties of the fields are renormalized.

Let us turn now to the first factor in (9.208). By construction it has invariance under the transformation (see (9.203))

$$\psi_\Delta(z, \bar{z}) \rightarrow \left(\frac{dw}{dz}\right)^\Delta \left(\frac{\overline{dw}}{d\bar{z}}\right)^\Delta \psi_\Delta(w(z), \overline{w(z)}) \tag{9.216}$$

This conformal invariance, as we have already noted, is a remnant of the general covariance in our gauge. The anomalous dimension  $\Delta$  in (9.207) is such that the vertex operator (9.208) has dimension  $2(\Delta + p^2/2) = 2$ . Therefore, the mass of the ground state of the string is determined by the anomalous dimension of the field  $\psi_\Delta(z)$ , the quantity being completely defined in terms of the Liouville field theory. Also, the integral (9.208) is invariant under  $SL(2, C)$  transformations, and hence, in order to avoid integrating over the infinite volume of  $SL(2, C)$  one has to factor this volume out of the integral. Practically, this consists of simply shifting three points  $\xi_1, \xi_2, \xi_3$  to  $0, 1, \infty$  by the use of  $SL(2, C)$ . Different choices of  $SL(2, C)$ —“gauge fixing” would lead to apparently

different but in fact coincident amplitudes  $\mathcal{A}$ . The matter is quite trivial, like fixing center of mass coordinates.

We shall now show an extremely important thing. Namely, the amplitude  $\mathcal{A}$  as function of  $s_{i_1, \dots, i_k} = (\mathbf{p}_{i_1} + \dots + \mathbf{p}_{i_k})^2$  has an infinite number of poles with factorized residues, by which we mean the following relation, satisfied by the set of amplitudes  $\mathcal{A}_N(\mathbf{p}_1, \dots, \mathbf{p}_N)$ :

$$\begin{aligned}
 \text{Diagram 1} &= \sum_r \text{Diagram 2} \\
 &= \sum_r \frac{\mathcal{A}_k^{(r)}(\mathbf{p}_{i_1}, \dots, \mathbf{p}_{i_k}) \mathcal{A}_{N-k}^{(r)}(\mathbf{p}_{i_{k+1}}, \dots, \mathbf{p}_{i_N})}{(\mathbf{p}_{i_1} + \dots + \mathbf{p}_{i_k})^2 + m_r^2} \quad (9.217)
 \end{aligned}$$

Here, we have denoted by  $m_r^2$  the mass of a resonance, among which the ground state with  $m_0^2 = 2(\Delta - 1)$  must be present. The  $\mathcal{A}_k^{(r)}$  are amplitudes involving  $k$ -ground state particles and one resonance. So we were slightly imprecise when talking about the set of  $\{A_N\}$ . In fact the factorization requirement must be imposed on the larger set of amplitudes, involving arbitrary number of excited particles. We have not given a formula like (9.208) for them yet, but this will be done in the course of the proof.

The factorization condition (9.217) is clearly necessary and obvious from the physical point of view. It reflects the fact that due to the short range of forces, separated interactions are independent. It is equally obvious from the geometrical point of view. A resonance is presented by a thin tube which is spanned in space-time by propagation of the closed string in the corresponding state. The amplitude is obtained from surfaces having topology of spheres with pinned points. In the kinematical region, described by (9.217) the relevant configurations of this sphere will be



thus representing a fusion of two (or more) closed strings into one and subsequent decay. This is just the pole diagram (9.217). The diagrams with branch cuts, which have several closed strings in the middle correspond to more complex topologies, spheres with handles. At the same time, algebraically, factorization is so terribly complicated that the proof of it must be either very simple or not at all. Fortunately, the first alternative is realized. We shall now show that the factorization

condition is equivalent to the operator product expansion of the underlying two dimensional field theory, and that there is a correspondence between 2d-operators and resonances.

The operator product expansion is a fundamental property inherent in any reasonable field theory. It states that two local operators in the theory, placed close together, can be viewed as a set of local operators. "Close together" here means that we are considering some multi-point correlation function, including our pair of operators, as well as many others, and the distance between the pair is smaller than all the other distances. Roughly speaking, when viewed from a large distance, our two points coalesce into one. Formally, if the field theory possesses a set of local operators  $\{O_n(\xi)\}$  the following relation holds:

$$O_n(\xi)O_m(0) \underset{\xi \rightarrow 0}{\sim} \sum_l C_{nm}^l(\xi)O_l(0) \tag{9.218}$$

which should be understood as a prescription to be inserted into any correlation function, say:

$$\begin{aligned} &\langle O_n(\xi)O_m(0)O_1(\xi_1)\dots O_N(\xi_N)\rangle \\ &\underset{\xi \rightarrow 0}{\sim} \sum_l C_{nm}^l(\xi)\langle O_l(0)O_1(\xi_1)\dots O_N(\xi_N)\rangle \end{aligned} \tag{9.219}$$

We shall also assume that it is possible to find a complete set of operators  $\{O_n\}$ , so that any quantum state can be generated by acting on the vacuum by their linear combination. This implies, that the relation (9.218), when the whole set  $\{O_n\}$  is inserted, becomes an exact relation. Now, what can we say about the "structure functions",  $C_{nm}^l(\xi)$ ? Suppose that we are dealing with conformal field theory, invariant under  $SL(2, C)$  transformations. Then any operator  $O_n$  is characterized by its anomalous dimension  $\Delta_n$ . This means that the correlation functions are invariant under the transformations:

$$O_n(\xi) \rightarrow \lambda^{\Delta_n}O_n(\lambda\xi) \tag{9.220}$$

In order that (9.218) be consistent with this, we have to require:

$$C_{nm}^l(\lambda\xi) = \lambda^{\Delta_l - \Delta_n - \Delta_m}C_{nm}^l(\xi) \tag{9.221}$$

For scalar operators that means simply:

$$C_{nm}^l(\xi) \sim |\xi|^{\Delta_l - \Delta_n - \Delta_m} \tag{9.222}$$

while in general some tensor structures appear. All this already looks similar to the factorization property, but a few more steps are needed to

establish an exact correspondence. First of all, let us work out explicitly the case of operators

$$U_{\mathbf{p}}(\xi) = :e^{i\mathbf{p} \cdot \mathbf{x}(\xi)}:$$

We have:

$$\begin{aligned} & :e^{i\mathbf{p}_1 \cdot \mathbf{x}(\xi)}: \quad :e^{i\mathbf{p}_2 \cdot \mathbf{x}(0)}: \\ & \approx \lim_{\xi \rightarrow 0} |\xi|^{(\mathbf{p}_1 + \mathbf{p}_2)^2 - \mathbf{p}_1^2 - \mathbf{p}_2^2} e^{i(\mathbf{p}_1 + \mathbf{p}_2) \cdot \mathbf{x}(0)} + \text{less singular terms} \quad (9.223) \end{aligned}$$

Here we have made use of the fact that the dimension of  $U_{\mathbf{p}}(\xi)$  is equal to  $\mathbf{p}^2$ . The formula (9.223) follows from the fact that operators in this case carry a conserved momentum  $\mathbf{p}$ . This can be easily checked from the expression for the correlation function (9.209) by allowing the points  $\xi_1$  and  $\xi_2$  to coalesce, and comparing with the correlation function containing  $U_{\mathbf{p}_1 + \mathbf{p}_2}(\xi_2)$ . This is a good exercise for a first acquaintance with operator products. Let us turn now to the scattering amplitude, and first consider the case  $\mathcal{D} = 26$  when the Liouville field is absent. The main point is that singularities of the amplitude  $\mathcal{A}$  as a function of  $\mathbf{p}_j$  arise from the regions in  $\xi$ -space where some points, say  $\xi_1$  and  $\xi_2$ , coincide. In this case a pole singularity develops. Indeed, we have:

$$\begin{aligned} \mathcal{A}_N(\mathbf{p}_1, \dots, \mathbf{p}_N) &= \int \prod_j d^2 \xi_j \langle U_{\mathbf{p}_1}(\xi_1) \dots U_{\mathbf{p}_N}(\xi_N) \rangle \\ &= \int d^2 \eta \int d^2 \xi_2 \dots d^2 \xi_N \langle U_{\mathbf{p}_1}(\xi_2 + \eta) U_{\mathbf{p}_2}(\xi_2) \dots U_{\mathbf{p}_N}(\xi_N) \rangle \\ &\approx \int_{\eta \rightarrow 0} d^2 \eta |\eta|^{(\mathbf{p}_1 + \mathbf{p}_2)^2 - \mathbf{p}_1^2 - \mathbf{p}_2^2} \\ &\quad \times \int d^2 \xi_2 \dots d^2 \xi_N \langle U_{\mathbf{p}_1 + \mathbf{p}_2}(\xi_2) \dots U_{\mathbf{p}_N}(\xi_N) \rangle \\ &= \left( \frac{1}{(\mathbf{p}_1 + \mathbf{p}_2)^2 - 2} \right) \mathcal{A}_{N-1}(\mathbf{p}_1 + \mathbf{p}_2, \mathbf{p}_3, \dots, \mathbf{p}_N) \\ &\quad + \text{terms, regular when } (\mathbf{p}_1 + \mathbf{p}_2)^2 = 2 \quad (9.224) \end{aligned}$$

(other poles in  $(\mathbf{p}_1 + \mathbf{p}_2)^2$  are also present and they will be considered below). We have used here the operator product expansion (9.223) and the fact that the anomalous dimension of  $U_{\mathbf{p}_j}(\xi)$  must be equal to 2 (to ensure projective invariance without the Liouville field), thus giving the mass of the ground state  $m_0^2 = -\mathbf{p}^2 = -2$  (the well known tachyon of

the standard dual model). We see from (9.224) that two tachyons may coalesce into a single one and the amplitude for such a process satisfies the factorization condition. The derivation of (9.224) shows a simple mechanism which transforms the operator product expansion in the underlying two dimensional field theory (the theory on the world sheet of the string) into a  $\mathcal{D}$ -dimensional factorization of poles. We have considered only the simplest example to demonstrate the idea and now it is necessary to go further.

Staying in  $\mathcal{D} = 26$ , we can look at the remaining poles of the amplitude. They will correspond to higher operators in the OPE. By the use of the exact version of (9.223):

$$\begin{aligned} & :e^{i\mathbf{p}_1 \cdot \mathbf{x}(\xi)}: :e^{i\mathbf{p}_2 \cdot \mathbf{x}(\xi)}: \\ & = |\xi|^{(\mathbf{p}_1 + \mathbf{p}_2)^2 - \mathbf{p}_1^2 - \mathbf{p}_2^2} :e^{i\mathbf{p}_1 \cdot \mathbf{x}(\xi) + i\mathbf{p}_2 \cdot \mathbf{x}(0)}: \end{aligned} \quad (9.225)$$

we find that the next term is given by

$$\begin{aligned} U_{\mathbf{p}_1}(\xi)U_{\mathbf{p}_2}(0) \approx & |\xi|^{(\mathbf{p}_1 + \mathbf{p}_2)^2 - \mathbf{p}_1^2 - \mathbf{p}_2^2} \left\{ U_{\mathbf{p}_1 + \mathbf{p}_2}(0) + :i\mathbf{p}_1 \cdot \xi^a \partial_a \mathbf{x}(0) e^{i(\mathbf{p}_1 + \mathbf{p}_2) \cdot \mathbf{x}(0)}: \right. \\ & \left. + : \left( \frac{i}{2} p_{1,\mu} \xi^a \xi^b \partial_a \partial_b x_\mu - \frac{1}{2} p_{1,\mu} p_{1,\nu} \xi^a \xi^b \partial_a x_\mu \partial_b x_\nu \right) e^{i(\mathbf{p}_1 + \mathbf{p}_2) \cdot \mathbf{x}(0)}: + \dots \right\} \end{aligned} \quad (9.226)$$

Substituting this into (9.224) we arrive at the general rule, which can be stated as follows.

The general operator, which appears in the OPE (9.225) has the form:

$$\begin{aligned} O_{m_1 \mu_1 \dots m_k \mu_k; n_1 \nu_1 \dots n_l \nu_l} \\ = : \partial_z^{m_1} x_{\mu_1} \dots \partial_z^{m_k} x_{\mu_k} \partial_z^{n_1} x_{\nu_1} \dots \partial_z^{n_l} x_{\nu_l} e^{i\mathbf{p} \cdot \mathbf{x}}: \end{aligned} \quad (9.227)$$

It has dimensionality:

$$\Delta = \mathbf{p}^2 + \sum_{j=1}^k m_j + \sum_{i=1}^l n_i$$

The pole corresponding to this contribution in (9.224) occurs when:

$$\Delta = 2 \quad (9.228)$$

or

$$m^2 = -\mathbf{p}^2 = \sum_j m_j + \sum_i n_i - 2 \quad (9.229)$$

A convenient notational trick is to introduce two sets of harmonic oscillators  $\{a_{m,\mu}^+, \tilde{a}_{n,\nu}^+\}$  with the ordinary commutation rules  $[a_{n,\mu}, a_{m,\nu}] = [\tilde{a}_{n,\mu}, \tilde{a}_{m,\nu}] = n\delta_{m,n}\delta_{\mu,\nu}$ ;  $[a, \tilde{a}] = 0$  and associate with the operator (9.227) a vector in their Fock space:

$$|0, p\rangle \leftrightarrow e^{ipx} \quad (9.230)$$

$$O_{m_1\mu_1\dots m_k\mu_k, n_1\nu_1\dots n_l\nu_l} \leftrightarrow a_{m_1\mu_1}^+ \dots a_{m_k\mu_k}^+ \tilde{a}_{n_1\nu_1}^+ \dots \tilde{a}_{n_l\nu_l}^+ |0, p\rangle$$

We see, that with these notations at hand we can present  $m^2$  as an eigenvalue of the operator:

$$m^2 = \sum_m (a_{m,\mu}^+ a_{m,\mu} + \tilde{a}_{m,\nu}^+ \tilde{a}_{m,\nu}) - 2 \quad (9.231)$$

The states of these oscillators thus can be interpreted as vibrational excitations of the string. Sometimes this representation is very useful.

An important point is that not all of the operators (9.227) contribute to the amplitude. The residues of some of the corresponding poles vanish. For example, it is clear that operators odd under the change of orientation  $z \leftrightarrow \bar{z}$  (such as  $(\partial_z x_\mu \partial_{\bar{z}} x_\nu - \partial_{\bar{z}} x_\mu \partial_z x_\nu) e^{ip \cdot x}$ ) will not contribute, since the residue, being proportional to

res  $\mathcal{A}_N(p_1 \dots p_N)$

$$\sim \int d^2z d^2z_3 \dots d^2z_N \langle O_{m_1\mu_1, \dots, n_1\nu_1, \dots}^{p_1, \dots, p_N}(z, \bar{z}) U_{p_3}(z_3, \bar{z}_3) \dots U_{p_N}(\bar{z}_N, z_N) \rangle \quad (9.232)$$

would vanish since all  $\{U_p\}$  are even under this transformation.

Also, the operators must have conformal spin zero i.e. under the rotation  $z \rightarrow e^{i\alpha}z$  they have to be scalars for the same reason as above. The first excited state of the string thus corresponds to the operator

$$O_{\mu\nu} = (\partial_z x_\mu \partial_{\bar{z}} x_\nu + \partial_{\bar{z}} x_\mu \partial_z x_\nu) e^{ip \cdot x} \quad (9.233)$$

and has

$$m^2 = -2 + 1 + 1 = 0$$

We shall analyse its spin content later.

There is another reason for the cancellation of residues. Consider an operator:

$$\begin{aligned} ip_\mu O_{\mu\nu} &= ip_\mu (\partial_z x_\mu \partial_{\bar{z}} x_\nu + \partial_{\bar{z}} x_\mu \partial_z x_\nu) e^{ip \cdot x} \\ &= \partial_z (\partial_{\bar{z}} x_\nu e^{ip \cdot x}) + \partial_{\bar{z}} (\partial_z x_\nu e^{ip \cdot x}) \end{aligned} \quad (9.234)$$

(where we have used the fact, that  $\partial_z \partial_{\bar{z}} x = 0$  since  $x$  is a free field. The same thing was implicitly used in (9.227)).

Owing to the fact that (9.234) is a total derivative its contribution to the residue (9.232) is equal to zero†, or in other words the amplitude for emission of the massless tensor state (9.233) is purely transverse. This is a very nice fact because otherwise in the Minkowski continuation of our theory we would inevitably have negative norm states, the time-like gravitons.

There are actually many more cancellations. To find them we have to exploit the conformal properties of the underlying two-dimensional theory.

### 9.9 The Energy–Momentum Tensor in Conformal Quantum Field Theory

There exists a remarkable mechanism which transforms the conformal Ward identities of the two-dimensional field theory of a string into certain transversality conditions, generalizing the decoupling of the states similar to (9.234). This mechanism will be important not only for  $\mathcal{D} = 26$ , but for all  $\mathcal{D}$  and for that reason we shall investigate it in detail.

Let us begin with the derivation of the Ward identities in the case when only the free  $x$ -field is present ( $\mathcal{D} = 26$ ). The energy-momentum tensor of such a field is given by:

$$T_{\alpha\beta}(\xi) = \partial_\alpha x \cdot \partial_\beta x - \frac{1}{2} \delta_{\alpha\beta} (\partial_\lambda x)^2 \tag{9.235}$$

and due to the equations of motion it is conserved:

$$\frac{\partial T_{\alpha\beta}}{\partial \xi^\alpha} = 0 \tag{9.236}$$

It is very helpful to use a complex notation and to introduce:

$$T(z, \bar{z}) = T_{11} - T_{22} + 2i T_{12} = \partial_z x \cdot \partial_z x \tag{9.237}$$

The conservation law now reads

$$\partial_{\bar{z}} T(z, \bar{z}) = 2 \partial_z x \cdot \partial_z \partial_{\bar{z}} x = 0 \tag{9.238}$$

implying that  $T$  is an analytic function of  $z$ . In the derivation of (9.238) we have used not only (9.236) but also the following important property of (9.235):

$$T_{\alpha\alpha} = 0 \tag{9.239}$$

---

† Because integrals are analytic in  $p_i \cdot p_j$  and there is always a region in the space of momenta  $p_j$  where all boundary terms vanish.

This property expresses conformal invariance of a theory in any number of dimensions. Indeed, generically, the energy momentum tensor determines the variation of the action  $S$  of any system under the change of variables:

$$\begin{aligned}\xi^\alpha &\rightarrow \xi^\alpha + \varepsilon^\alpha(\xi) \\ S &\rightarrow S + \delta S\end{aligned}\tag{9.240}$$

$$\delta S = \int T^{\alpha\beta}(\xi)(\partial_\alpha \varepsilon_\beta + \partial_\beta \varepsilon_\alpha) d^n \xi$$

Hence, the tracelessness (9.239) implies that the action is not changed for such  $\varepsilon_\alpha$  that

$$\partial_\alpha \varepsilon_\beta + \partial_\beta \varepsilon_\alpha \sim \delta_{\alpha\beta}\tag{9.241}$$

or

$$\partial_\alpha \varepsilon_\beta + \partial_\beta \varepsilon_\alpha - \frac{2}{n} \delta_{\alpha\beta} \partial_\lambda \varepsilon_\lambda = 0\tag{9.242}$$

( $n$  is the dimensionality of the  $\xi$ -space).

These conditions for  $n = 2$  are just the Cauchy-Riemann equations, satisfied by any analytic function. Thus we have an infinite dimensional conformal group in this case. At the same time, for  $n > 2$  the group is finite, and consists of the ordinary translations and rotations plus dilations and inversions.

We see that the formula (9.238) is a condensed expression of conformal invariance of the theory. It is not just conservation of  $T$  since generically one would also have terms in (9.238) containing  $\partial_z$  derivatives. Their absence is crucial for further discussion.

Let us show that correlation functions involving the operator  $T$  can be determined explicitly by the Ward identities.

We shall give first their formal derivation, and then in order to get a feeling of how they arise, analyze the special case of free fields.

It is convenient to treat the variables  $z$  and  $\bar{z}$  as independent and to concentrate on the variation of  $z$ . Suppose that we have a set of operators  $\{O_n\}$  which transform as

$$O_n(z) \rightarrow \left(\frac{df}{dz}\right)^{\Delta_n} O_n(f(z))$$

(9.243)

or

$$\delta_\varepsilon O_n(z) = \varepsilon(z) \partial_z O_n(z) + \Delta_n \varepsilon'(z) O_n(z)$$

where we have passed to infinitesimal transformations,  $z \rightarrow f(z) = z + \varepsilon(z)$ . Under the same transformations the action acquires a variation (9.240) which can be rewritten as:

$$\delta_\varepsilon S = \int d^2z T(z) \partial_{\bar{z}} \varepsilon \tag{9.244}$$

As is usual in field theory, the Ward identity arises when we change variables in the functional integral and require its value to be unchanged. It reads:

$$\begin{aligned} &\langle \delta_\varepsilon S O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &= \sum_{j=1}^k \langle O_{n_1}(z_1) \dots \delta_\varepsilon O_{n_j}(z_j) \dots O_{n_k}(z_k) \rangle \end{aligned} \tag{9.245}$$

where the left-hand side represents the variation of the factor  $e^{-S}$  in the Euclidean functional integral under the transformation (9.244) while the right-hand side is the variation of the operators  $O_n$  themselves.

Requiring that (9.245) holds for any  $\varepsilon$  and integrating by parts we obtain:

$$\begin{aligned} &\partial_{\bar{z}} \langle T(z) O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &= \sum_j \left\{ \delta(z - z_j) \frac{\partial}{\partial z_j} + \Delta_{n_j} \delta'(z - z_j) \right\} \\ &\quad \times \langle O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \end{aligned} \tag{9.246}$$

(here  $\delta$  is the two dimensional  $\delta$ -function).

A structure like (9.246) is typical for all Ward identities we encounter in field theory. What is unusual is that equation (9.246) can be integrated uniquely. To do this, let us use the following simple relations:

$$\begin{aligned} &\partial_z \partial_{\bar{z}} \log(|z - z'|^2) \\ &= \partial_z \partial_{\bar{z}} [\log(z - z') + \log(\bar{z} - \bar{z}')] \\ &= \partial_{\bar{z}} \frac{1}{z - z'} = -\pi \delta(z - z') \end{aligned} \tag{9.247}$$

Therefore, changing without further notice, the normalization of  $T$  by  $\pi$  we get:

$$\begin{aligned} &\langle T(z) O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &= \sum_{j=1}^k \left\{ \frac{\Delta_{n_j}}{(z - z_j)^2} + \frac{1}{z - z_j} \frac{\partial}{\partial z_j} \right\} \\ &\quad \times \langle O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \end{aligned} \tag{9.248}$$

The reason why the correlation functions involving  $T(z)$  are found explicitly lies in the fact that due to the tracelessness of  $T_{\alpha\beta}$ , the energy-momentum conservation law involves only  $\partial_{\bar{z}}$  and not  $\partial_z$ . Speaking less technically,  $T(z)$  generates some conformal transformation, and owing to invariance of the theory its result can be expressed in terms of variations of  $\{O_n\}$ . We shall also need Ward identities involving several  $T(z)$ . In order to find them we have to analyse how  $T(z)$  is transformed under the conformal group. Since  $T(z)$  is a second rank symmetric traceless tensor under  $SL(2, C)$  we expect that its general transformation law respects these properties. Since the parameter of a general transformation is a vector field  $\varepsilon(z)$  (remember, that we are using a complex notation) the most general variation of  $T(z)$  has the form:

$$\delta_\varepsilon T(z) = \varepsilon(z)\partial_z T + 2(\partial_z \varepsilon)T + \frac{c}{12} \partial_z^3 \varepsilon \quad (9.249)$$

(since under  $SL(2, C)$ ,  $T(z)$  transforms as  $(dz)^{-2}$  and  $\varepsilon(z)$  as  $dz$ ).

The constant  $c$  which appeared in (9.249) is a dynamical characteristic of the theory which plays an important role. For free fields we shall compute it in a moment, but first let us describe its general meaning.

Its appearance in (9.249) implies that under the conformal group  $T(z)$  transforms not as a primary operator with the variation given by (9.243). Perhaps it is now time to say that the primary operators are not the only ones in our complete set. Indeed, together with some primary  $O_n(z)$  we must have the derivative operator  $\partial_z O_n$ . Its transformation law is (we obtain this by differentiating (9.243)):

$$\begin{aligned} \delta_\varepsilon A_n &= \varepsilon \partial_z A_n + (\Delta_n + 1)(\partial_z \varepsilon)A_n \\ &\quad + \Delta_n (\partial_z^2 \varepsilon_n) O_n \\ A_n &\equiv \partial_z O_n \end{aligned} \quad (9.250)$$

It is clear that higher derivatives of  $O_n$  will contain higher derivatives of  $\varepsilon$  in their conformal variations. The prospects for the use of an operator algebra (9.218) with such an enormous variety of operators seem rather hopeless.

However, this is not so. We shall show that only the primary operators define the structure of the theory, while all others are obtained by simple rules. The energy-momentum tensor plays a crucial role in finding these rules and thus we go back to (9.249).

Using (9.244) and (9.245) we obtain:

$$\frac{1}{\pi} \int d^2u \partial_{\bar{u}} \varepsilon \langle T(u)T(z) \rangle = \langle \delta_\varepsilon T(z) \rangle = \frac{c}{12} \partial_z^3 \varepsilon(z) \tag{9.251}$$

$$\partial_{\bar{z}} \langle T(z)T(u) \rangle = \pi \frac{c}{12} \partial_z^3 \delta^{(2)}(z - u)$$

or

$$\langle T(z)T(u) \rangle = -\frac{c}{12} \partial_z^3 \left( \frac{1}{z - u} \right) = \frac{c}{2} (z - u)^{-4} \tag{9.252}$$

We observe that if it were not for the  $c$ -term in (9.249) the two-point function of the energy momentum tensor would be zero, which is incompatible with a positive norm of the space of states. For the case of a free  $x$ -field when  $T(z) = \frac{1}{2}(\partial_z x)^2$  elementary computation of the two point function shows that  $c = \mathcal{D}$  (where  $\mathcal{D}$  is the number of components of  $x$ ).

There is yet another geometrical interpretation of the constant  $c$  and its role in Ward identities. Let us multiply (9.248) by some meromorphic function  $\varepsilon(z)$  and integrate around the contour  $C$  which surrounds all the points  $\{z_j\}$  and inside which  $\varepsilon(z)$  is analytic. Then the equation (9.248) takes the form:

$$\begin{aligned} \langle T_\varepsilon O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle &= \delta_\varepsilon \langle O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ T_\varepsilon &= \oint_C dz \varepsilon(z) T(z) \end{aligned} \tag{9.253}$$

$$\delta_\varepsilon O_{n_k}(z_k) = \varepsilon(z_k) \partial_{z_k} O_{n_k}(z_k) + \Delta_{n_k} \varepsilon'(z_k) O_{n_k}(z_k)$$

so that  $T_\varepsilon$  generates conformal transformation in the space of different correlation functions. Let us examine the commutator of two such transformations. What kind of algebra should we expect? From the geometrical point of view, conformal transformations are generated by the operators

$$\hat{\delta}_\varepsilon = \varepsilon(z) \frac{\partial}{\partial z}$$

which form a Lie algebra, since

$$\begin{aligned} [\hat{\delta}_{\varepsilon_1}, \hat{\delta}_{\varepsilon_2}] &= \hat{\delta}_{[\varepsilon_1, \varepsilon_2]} \\ [\varepsilon_1, \varepsilon_2] &= \varepsilon_1 \varepsilon_2' - \varepsilon_1' \varepsilon_2 \end{aligned} \tag{9.254}$$

So, one would think that if we consider a correlation function:

$$\langle T_{\varepsilon_1} T_{\varepsilon_2} O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle$$

where the contour  $C_1$  surrounds all singularities of  $\varepsilon_2$  and  $\{z_j\}$  and subtract the same expression with 1 and 2 interchanged the correlation function:

$$\langle T_{[\varepsilon_1, \varepsilon_2]} O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle$$

will appear. This is not quite true because of the  $C$ -term. The correct formula can be derived from (9.249). We have, first of all:

$$[T_\varepsilon, T(z)] = \varepsilon \partial_z T + 2(\partial_z \varepsilon)T + \frac{c}{12} \varepsilon'''(z) \quad (9.255)$$

The meaning of the commutator in (9.255) is, as explained above, given by

$$\begin{aligned} & \langle [T_\varepsilon, T(z)] O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &= \left( \oint_{C_1} du - \oint_{C_2} du \right) \varepsilon(u) \langle T(u) T(z) O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \end{aligned} \quad (9.256)$$

where  $C_1$  includes the point  $z$  which lies outside of  $C_2$ . Also  $\varepsilon(u)$  is analytic inside both contours and  $\{z_j\}$  lie inside them. This definition of the commutator in Euclidean field theory will of course become equivalent to the one defined in the Hamiltonian formalism in Minkowski space. It is however independent of the choice of time axis. If we integrate (9.255) with some function  $\varepsilon_2$  around the loop  $C_2$ , we get

$$[T_{\varepsilon_1}, T_{\varepsilon_2}] = T_{[\varepsilon_1, \varepsilon_2]} + \frac{c}{24} \oint_C dz (\varepsilon_1''' \varepsilon_2 - \varepsilon_2''' \varepsilon_1) \quad (9.257)$$

where the contour  $C$  has to separate the points  $\{z_j\}$  from the singularities of  $\varepsilon_1$  and  $\varepsilon_2$ .

The algebra (9.257) (called the Virasoro algebra) is an extension of (9.254). It is easy to check that the last term in (9.257) is the only possible functional compatible with the Jacobi identity.

We see that due to (9.249) we can easily compute correlation functions containing any number of  $T$ . For example:

$$\begin{aligned} & \langle T(z) T(u) O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &= \frac{c}{2} (z-u)^{-4} \langle O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &+ \left\{ \frac{2}{(z-u)^2} + \frac{1}{z-u} \frac{\partial}{\partial u} + \sum_k \left( \frac{\Delta k}{(z-z_k)^2} + \frac{1}{z-z_k} \frac{\partial}{\partial z_k} \right) \right\} \\ &\times \langle T(u) O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \end{aligned} \quad (9.258)$$

There are always several ways to derive formulas like (9.258) since for example one can start not from  $T(z)$  but from  $T(u)$  in deriving Ward identities. One can check that the condition that the resulting formula is independent of the order of derivation is just the Jacobi identity for the algebra (9.257). The Ward identities (9.258) give important information concerning the operator algebra (9.218). Namely, they must be such that after fusing together any two  $O_n$  in the left hand side, we obtain a result consistent with the corresponding fusion of two points in the right hand side (which depends on  $\{z_{jj}\}$  explicitly). In other words the operator algebra must be conformally invariant.

In order to classify possible operators let us consider the operator product of  $T(z)$  and some primary operator  $\psi(0)$  with dimension  $\Delta$ . We have:

$$T(z + \zeta)\psi(z) = \frac{\Delta}{\zeta^2} \psi(z) + \frac{1}{\zeta} \partial_z \psi(z) + \psi_2(z) + \zeta \psi_3(z) + \dots \tag{9.259}$$

The first two terms in this formula are obtained by an immediate check with equation (9.248) while the operators  $\psi_k(z)$  with  $k \geq 2$  must be determined by more careful analysis of the same equation. For example, the field  $\psi_2(z)$  is defined through its correlation function with any others as follows. Let us write (9.248) in the form:

$$\begin{aligned} &\langle T(z + \zeta)\psi(z)O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &= \left( \frac{\Delta}{\zeta^2} + \frac{1}{\zeta} \frac{\partial}{\partial z} \right) \langle \psi(z)O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &+ \sum_k \left( \frac{\Delta_{n_k}}{(z + \zeta - z_k)^2} + \frac{1}{z + \zeta - z_k} \frac{\partial}{\partial z_k} \right) \\ &\times \langle \psi(z)O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \end{aligned} \tag{9.260}$$

Taking  $\zeta \rightarrow 0$  and comparing (9.260) and (9.259) we get:

$$\begin{aligned} &\langle \psi_2(z)O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \\ &= \left[ \sum_k \left( \frac{\Delta_{n_k}}{(z - z_k)^2} + \frac{1}{z - z_k} \frac{\partial}{\partial z_k} \right) \right] \\ &\times \langle \psi(z)O_{n_1}(z_1) \dots O_{n_k}(z_k) \rangle \end{aligned} \tag{9.261}$$

So, we see that there are an infinite number of fields  $\psi_k(z)$  associated with the primary field  $\psi(z)$ , with uniquely defined correlation functions (for  $k > 2$  we expand the r.h.s. to the corresponding power of  $\zeta$ ).

It is convenient to represent the fields  $\psi_k(z)$  in the form:

$$\begin{aligned}\psi_k(z) &= T_{-k}\psi(z) \\ T_{-k}(z) &= \frac{1}{2\pi i} \oint_{\zeta} \frac{d\zeta}{\zeta^{k+1}} T(z + \zeta)\end{aligned}\quad (9.262)$$

This notation tells us that the  $\psi_k(z)$  do not exhaust all the secondary operators which arise from one primary field  $\psi(z)$ . This is because there are other Ward identities to be satisfied, containing more  $T$ , like (9.258). The whole conformal family of  $\psi(z)$  is composed of the operators:

$$\psi_{k_1 \dots k_n; \bar{k}_1 \dots \bar{k}_n} = T_{-k_1} \dots T_{-k_n} \bar{T}_{-\bar{k}_1} \dots \bar{T}_{-\bar{k}_n} \psi(z, \bar{z}) \quad (9.263)$$

(where  $\bar{T}$  is the complex conjugate of  $T$ ). Again, all correlation functions of these operators are expressible through that of  $\psi(z)$ . Moreover, if in an operator product of two primary operators we have some operator  $\psi$  then together with it will appear the whole family  $\psi_{(k), (\bar{k})}$ , with coefficients calculable from the Ward identities. There exist simple algebraic prescriptions for doing this, but for our purposes they are not needed.

Generically, the field  $\psi_{(k), (\bar{k})}$  transforms under conformal transformations in a complicated inhomogeneous fashion, involving lower secondary fields (recall (9.250) and notice that  $\partial_z O_n = T_{-1} O_n$ ). However there are important special cases in which some secondary field  $\psi_{(k), (\bar{k})}$  becomes a primary one. This happens if for any  $n > 0$   $T_n \psi_{(k), (\bar{k})} = \bar{T}_n \psi_{(k), (\bar{k})} = 0$ . These conditions can be verified using the Virasoro algebra in the form:

$$[T_n, T_m] = (n - m)T_{n+m} + \frac{c}{12} n(n^2 - 1)\delta_{n+m, 0} \quad (9.264)$$

(and the same for  $\bar{T}$ ) that follows from (9.255), (9.262). In this case, owing to the homogeneity of the transformation law, we can eliminate this field from the operator algebra without spoiling conformal invariance (in these cases coefficients in the operator product expansion become undetermined). These cases of degeneracy will serve us in the next section to reduce the number of string states.

Let us give the simplest examples of this phenomenon. If we take an operator  $T_{-1} \phi_\Delta$ , where  $\phi_\Delta$  is some primary operator with dimension  $\Delta$ , the above degeneracy condition takes the form:

$$\begin{aligned}T_1 T_{-1} \phi &= 2T_0 \phi = 2\Delta \phi = 0 \\ \Delta &= 0\end{aligned}\quad (9.265)$$

In this case we say that the conformal family with  $\Delta = 0$  is degenerate at the first level. A less trivial example is degeneracy at the second level. Consider an operator:

$$\chi = (aT_{-2} + bT_{-1}^2)\phi \quad (9.266)$$

and apply the degeneracy conditions:

$$T_1\chi = T_2\chi = 0 \quad (9.267)$$

The conditions (9.267) are always sufficient because owing to the Virasoro algebra (9.264) any  $T_n$  can be obtained by combining  $T_1$  and  $T_2$ . Applying the first of these equations, and using

$$\begin{aligned} [T_1, T_{-2}] &= 3T_{-1} \\ [T_1, T_{-1}^2] &= 2(T_0T_{-1} + T_{-1}T_0) \\ &= 2T_{-1}(2T_0 + 1) \end{aligned}$$

we get

$$T_1\chi = (3a + 2(2\Delta + 1)b)T_{-1}\phi = 0 \quad (9.268)$$

Or, in other words the degenerate operator must have the form:

$$\chi = \left( T_{-2} - \frac{3}{2(2\Delta + 1)} T_{-1}^2 \right) \phi \quad (9.269)$$

This form, which satisfies (9.268) for any  $\Delta$ , is in fact just the projective invariant operator which can always be formed out of any primary field. The nontrivial condition, which implies covariance under the infinite conformal group is the second equation (9.267). Applying it to (9.266), and using the relations:

$$\begin{aligned} [T_2, T_{-2}] &= 4T_0 + \frac{c}{2} \\ [T_2, T_{-1}^2] &= [T_2, T_{-1}]T_{-1} + T_{-1}[T_2, T_{-1}] \\ &= 3(T_1T_{-1} + T_{-1}T_1) \\ &= 6T_0 + 6T_{-1}T_1 \end{aligned}$$

we obtain, together with (9.268):

$$\begin{aligned} 3a + 2(2\Delta + 1)b &= 0 \\ \left( 4\Delta + \frac{c}{2} \right) a + 6\Delta b &= 0 \end{aligned} \quad (9.270)$$

The condition of degeneracy is given by:

$$\begin{aligned} \begin{vmatrix} 3 & 2(2\Delta + 1) \\ 4\Delta + c/2 & 6\Delta \end{vmatrix} &= 0 \\ 8\Delta^2 + (c - 5)\Delta + c/2 &= 0 \\ \Delta &= \frac{5 - c \pm \sqrt{(c - 25)(c - 1)}}{16} \end{aligned} \quad (9.271)$$

This result shows, that for a fixed value of the central charge  $c$ , we have an operator family degenerate at the second level, if and only if the dimension  $\Delta$  of the original operator is given by (9.271). The natural question is what are the conditions for degeneracy at the  $N$ th level? The simple answer to this question is given by the so called Kac formula. The result is the following. There are a set of dimensionalities  $\Delta_{nm}$  characterized by two integers  $n$  and  $m$ . They lead to degeneracy at the level  $N = nm$ , provided that:

$$\begin{aligned} \Delta_{nm} &= \frac{c - 1}{24} + \frac{1}{8}(n\alpha_+ + m\alpha_-)^2 \\ \alpha_{\pm} &= \left(\frac{1 - c}{12}\right)^{1/2} \pm \left(\frac{25 - c}{12}\right)^{1/2} \end{aligned} \quad (9.272)$$

In the case  $n = 1$ ,  $m = 2$  the (9.272) is equivalent to (9.271).

Degenerate conformal families are very important. The reason is that their operator content is in a sense minimal, since many operators can be set to zero. In string theory, as we shall see, this implies decoupling of unwanted ghost states, while in statistical physics of phase transitions all known systems choose to be “minimal” theories or combinations thereof.

Armed with these results, we are ready to discuss the “no-ghost” properties of critical strings. By “no-ghost” we mean positivity of the norm of string states. Actually it is guaranteed by our construction of Liouville field theory, when  $\mathcal{D} \leq 26$ . But explicit check at  $\mathcal{D} = 26$  is very instructive.

## 9.10 Physical States of String Theory in the Critical Dimension

In Section 9.9 we presented a set of operators, composed out of the free field  $x(z, \bar{z})$ , which correspond to the particle states of the string. We also showed by explicit calculations, that some of these particles

decouple. Now we shall start the general analysis of the spectrum. The first (and the most important) statement will be the decoupling of all secondary operators. In other words particles correspond not to possible operators, but to the possible conformal families.

To demonstrate this let us examine the following Ward identity:

$$\begin{aligned}
 & \int d^2z_1 \dots d^2z_N \langle T(z) U_{p_1}(z_1) \dots U_{p_N}(z_N) \rangle \\
 &= \int \prod_j d^2z_j \left[ \sum_j \left( \frac{p_j^2/2}{(z-z_j)^2} + \frac{1}{z-z_j} \frac{\partial}{\partial z_j} \right) \right] \\
 & \quad \times \langle U_{p_1}(z_1) \dots U_{p_N}(z_N) \rangle \\
 &= \int \prod_j d^2z_j \sum_j \frac{\partial}{\partial z_j} \left( \frac{1}{z-z_j} \langle U_{p_1}(z_1) \dots U_{p_N}(z_N) \rangle \right) \\
 &= 0
 \end{aligned} \tag{9.273}$$

(where we have used the condition  $p_j^2 = 2$  which is needed for the conformal symmetry of the amplitudes). From this identity it follows that the operators which appear in the product  $T(z)U_p(z')$  will decouple from the amplitude. In other words, secondary operators of the type  $T_{-n}U_p(z)$  do not correspond to physical particles. By a similar argument, the same is true for the  $T_{-n_1} \dots T_{-n_k} U_p$ . If we use the analogy with oscillators (9.230) we can say that out of  $\mathcal{D}$  chains of oscillators  $\{a_{n,\mu}^+\}$  roughly  $\mathcal{D} - 1$  correspond to physical particles and one chain decouples. If we are lucky this might imply decoupling of the negative norm states created by  $\{a_{n,0}^+\}$ —at least the number of necessary decouplings is the same.

To see what really goes on we have to analyse several examples. To simplify notation we shall work with the case of “open strings” which is obtained by forgetting about the dependence on  $\bar{z}$  and considering only  $z$ -dependent operators. The closed string case is obtained by simple “doubling” of the operator set with  $\bar{z}$  dependent fields, since  $x(z, \bar{z}) = x_L(z) + x_R(\bar{z})$ . Examples of such doubling will be given later.

Let us look at the vector (“photon”) vertex, given by:

$$U_{p,\zeta} = :i\zeta^\mu \partial_z x_\mu e^{ip \cdot x} : \leftrightarrow \zeta^\mu(p) a_{1,\mu}^+ |0, p\rangle \tag{9.274}$$

(here  $\zeta_\mu(p)$  is the “photon” polarization). Let us first apply the condition that  $U_{p,\zeta}$  must be a primary operator, or:

$$T_n U_{p,\zeta} = 0, \quad n > 0 \tag{9.275}$$

In order to find this expression we must use the operator product:

$$\begin{aligned}
 T(z)U_{p,\zeta}(0) &= :-\frac{1}{2}(\partial_z x)^2(z):i\zeta^\mu \partial_z x_\mu e^{ip \cdot x(0)}: \\
 &= z^{-3}\zeta^\mu p_\mu: e^{ip \cdot x}: + \text{less singular terms} \quad (9.276)
 \end{aligned}$$

Recalling the definition of  $T_n$  we find:

$$\begin{aligned}
 T_1 U_{p,\zeta} &= \zeta^\mu p_\mu U_p \\
 T_n U_{p,\zeta} &= 0, \quad n > 1 \quad (9.277)
 \end{aligned}$$

We see that the condition for  $U_{p,\zeta}$  to be a primary operator is

$$p^\mu \zeta_\mu(p) = 0 \quad (9.278)$$

An equivalent way to derive (9.278) is to use the oscillator representation:

$$\begin{aligned}
 ix_\mu(z) &= q_\mu + p_\mu \log z + \sum_{n=1}^{\infty} \left( -\frac{a_{n,\mu}}{n} z^{-n} + \frac{a_{n,\mu}^+}{n} z^n \right) \\
 T(z) &= -:\frac{1}{2}(\partial_z x)^2: = \sum_n T_n z^{-n-2} \\
 T_n &= p^\mu a_{n,\mu} + \sum_{m=1}^{\infty} a_{m,\mu}^+ a_{m+n,\mu} \\
 &\quad + \frac{1}{2} \sum_{k=1}^{n-1} a_k a_{n-k}, \quad n > 0 \quad (9.279) \\
 T_{-n} &= T_n^+ \\
 T_0 &= \frac{1}{2} p^2 + \sum_{m=1}^{\infty} a_m^+ a_m
 \end{aligned}$$

Applying  $T_1$  to the state  $\zeta^\mu(p)a_{1,\mu}^+|0,p\rangle$  we obtain (9.278). This is a useful demonstration of the complete equivalence of the operator product formalism with the harmonic oscillator one.

Returning to (9.278) let us notice that in order that  $U_{p,\zeta}$  shall have dimension 1, we must have  $p^2 = 0$ . In this case the number of coupled states will not be  $\mathcal{D} - 1$  as follows from (9.278) but  $\mathcal{D} - 2$ . In order to show this, we notice that among the physical operators  $i\zeta^\mu \partial_z x_\mu \exp(ip \cdot x(z))$  with  $p^\mu \zeta_\mu(p) = 0$  there is one which is both physical and secondary. We have already shown that the family produced by  $: e^{ip \cdot x}:$  with  $p^2 = 0$  is degenerate and:

$$T_{-1}:e^{ip \cdot x}: = :ip^\mu \partial_z x_\mu e^{ip \cdot x}: \quad (9.280)$$

is, according to (9.265), physical for  $p^2 = 0$ . The same thing is seen from the fact that  $\zeta_\mu(p) = p_\mu$  satisfies:

$$p^\mu \zeta_\mu = p^2 = 0 \quad (9.281)$$

This is a standard feature of the photon: its polarization on shell can be shifted by a gauge transformation:

$$\zeta_\mu(p) \rightarrow \zeta_\mu(p) + \lambda p_\mu \quad (9.282)$$

By the use of this transformation we can make  $\zeta_0(p) = 0$  and then

$$p^\mu \zeta_\mu(p) = -p^i \zeta_i(p) = 0 \quad (9.283)$$

Hence we find that the photon has  $\mathcal{D} - 2$  possible polarizations. In our formalism this was the consequence of two facts. First, we had the condition for the vertex operator to transform properly under the conformal group:  $T_n U_{p,\zeta} = 0, n > 0$ . That has left  $\mathcal{D} - 1$  possible states. Second, one of these states was seen to be the secondary operator generated by  $U_p = :e^{ip \cdot x}$ : with  $p^2 = 0$ . As was shown at the beginning of this section all secondary operators decouple. Thus, we got  $\mathcal{D} - 2$ .

It is remarkable that a similar mechanism works at higher levels. Let us discuss the second one. The most general vertex will be given by:

$$\begin{aligned} U_p(s, \zeta) &= : (S_{\mu\nu} i \partial_z x^\mu i \partial_{\bar{z}} x^\nu + \zeta_\mu i \partial_z^2 x^\mu) e^{ip \cdot x} : \\ &\leftrightarrow (S_{\mu\nu} a_{1,\mu}^+ a_{1,\nu}^+ + \zeta_\mu a_{2,\mu}^+) |0, p\rangle \\ & \quad p^2 = -2 \end{aligned} \quad (9.284)$$

Let us see what forms of  $S_{\mu\nu}(p)$  and  $\zeta_\mu$  are left by the condition of conformal covariance:

$$T_n U_p(S, \zeta) = 0, \quad n > 0 \quad (9.285)$$

It is easier to work out in the oscillator representation. Applying (9.279) we find:

$$\begin{aligned} T_1(S_{\mu\nu} a_{1,\mu}^+ a_{1,\nu}^+ + \zeta_\mu a_{2,\mu}^+) |0, p\rangle \\ &= 2(p_\mu S_{\mu\nu} + \zeta_\nu) a_{1,\nu}^+ |0, p\rangle = 0 \\ T_2(S_{\mu\nu} a_{1,\mu}^+ a_{1,\nu}^+ + \zeta_\mu a_{2,\mu}^+) |0, p\rangle \\ &= (2p_\mu \zeta_\mu + S_{\lambda\lambda}) |0, p\rangle = 0 \end{aligned} \quad (9.286)$$

So, the conditions for physical polarizations are:

$$\begin{aligned} p_\mu \zeta_\mu + \frac{1}{2} S_{\lambda\lambda} &= 0 \\ p_\mu S_{\mu\nu} + \zeta_\nu &= 0, \quad p^2 = -2 \end{aligned} \quad (9.287)$$

Again, among these physical states some are secondary. This implies that there is some second level gauge invariance in (9.287). To find it, let us use formulas (9.269) and (9.271) for the case  $\Delta = -1$  (in this case the secondary operator will have  $\Delta = 1$ ). We conclude that the states

$$|a\rangle = (T_{-2} + \frac{3}{2} T_{-1}^2)|0, p\rangle, \quad p^2 = -2 \quad (9.288)$$

and

$$|b\rangle = T_{-1}\lambda_\mu(p)a_{1,\mu}^+|0, p\rangle, \quad p_\mu\lambda_\mu = 0$$

are physical if  $c = \mathcal{D} = 26$  in (9.272). The state  $|b\rangle$  corresponds to the following gauge transformation:

$$\begin{aligned} S_{\alpha\beta} &\rightarrow S_{\alpha\beta} + \frac{1}{2}(p_\alpha\lambda_\beta + p_\beta\lambda_\alpha) \\ \zeta_\alpha &\rightarrow \zeta_\alpha + \lambda_\alpha \end{aligned} \quad (9.289)$$

with  $p_\alpha\lambda_\alpha = 0$ . It is easy to see, that the conditions (9.287) are invariant under (9.289) provided that  $p^2 = -2$ . This first gauge invariance works for all  $\mathcal{D}$ . The second one, associated with the state  $|a\rangle$  is true only for  $\mathcal{D} = 26$ . After some computations, one finds:

$$\begin{aligned} S_{\alpha\beta} &\rightarrow S_{\alpha\beta} + \varepsilon(\delta_{\alpha\beta} + 3p_\alpha p_\beta) \\ \zeta_\alpha &\rightarrow \zeta_\alpha + 5\varepsilon p_\alpha \end{aligned} \quad (9.290)$$

In both cases, gauge transformations create states of zero norm, just as was the case with photons (recall that the state  $p_\mu a_{1,\mu}^+|0, p\rangle$  with  $p^2 = 0$  is of zero norm). The reason for zero norm is quite simple. If any state  $|f\rangle$  is physical:

$$T_n|f\rangle = 0, \quad \forall n > 0$$

and secondary:

$$|f\rangle = T_{-m}|g\rangle; \quad T_n(T_{-m}|g\rangle) = 0$$

then:

$$\langle f|f\rangle = \langle g|T_m T_{-m}|g\rangle = 0 \quad (9.291)$$

The meaning of this result is that at each level we have a certain gauge transformation, which does not change the physics, but creates zero norm states. It is also clear, that existence of such an invariance implies a considerable reduction of the physical spectrum. Let us now count the number of states remaining. Consider the states with  $p^2 = -m^2 = 2(1-n)$ . They are represented in the oscillator formalism as:

$$\begin{aligned} |f_n\rangle &= (\zeta_\mu a_{n,\mu}^+ + S_{\mu\nu} a_{n-1,\mu}^+ a_{1,\nu}^+ + \dots)|0, p\rangle \\ p^2/2 &= 1 - n \end{aligned} \quad (9.292)$$

The total number of these states is given by the number of ways in which the integer  $n$  can be split into a sum of positive integers. Indeed, if it were not for the space-time indices  $\mu$ , we would have the general state in the form:

$$|f\rangle = a_{i_1}^+ \dots a_{i_k}^+ |0, p\rangle$$

$$l_1 + \dots + l_k = n, \quad l_j > 0$$
(9.293)

The number of such states is given by:

$$N(n) = \frac{1}{2\pi i} \oint \frac{dx}{x^{n+1}} \prod_{k=1}^{\infty} \frac{1}{1-x^k}$$
(9.294)

To derive this, let us represent (9.293) as

$$|f\rangle = \dots (a_k^+)^{p_k} \dots (a_1^+)^{p_1} |0, p\rangle$$

$$\sum_{k=1}^{\infty} kp_k = n$$

Then

$$N(n) = \sum_{\{p_k\}} \delta_{n, \sum kp_k}$$

$$= \frac{1}{2\pi i} \oint \frac{dx}{x^{n+1}} \sum_{\{p_k\}} \prod_k x^{kp_k}$$

$$= \frac{1}{2\pi i} \oint \frac{dx}{x^{n+1}} \prod_k \left( \sum_{p_k} x^{kp_k} \right)$$

$$= \frac{1}{2\pi i} \oint \frac{dx}{x^{n+1}} \prod_k \frac{1}{1-x^k}$$

When we add  $\mathcal{D}$  space-time dimensions it is clear, that the number of states  $N_{\mathcal{D}}(n)$  at the  $n$ -th level is given by:

$$\sum_{n=0}^{\infty} N_{\mathcal{D}}(n)x^n = \left( \prod_{k=1}^{\infty} \frac{1}{1-x^k} \right)^{\mathcal{D}}$$
(9.295)

Now, not all states (9.292) are physical. We have to apply the condition  $T_n|f\rangle = 0$ . This results in equations for  $\zeta_{\mu}$ ,  $S_{\mu\nu}$ , etc. similar to (9.287). The number of such states will be generated by:

$$\left( \prod_{k=1}^{\infty} \frac{1}{1-x^k} \right)^{\mathcal{D}-1} = \sum_{n=0}^{\infty} N_{\mathcal{D}}^{\text{phys}}(n)x^n$$

$$N_{\mathcal{D}}^{\text{phys}}(n) = N_{\mathcal{D}-1}(n)$$
(9.296)

because effectively the  $T_n$ -constraints remove one component of each oscillator. However, this is not the end of the story because among the physical states there are some of zero norm. Suppose first that  $\mathcal{D} \neq 26$ . The only set of zero norm states is produced by:

$$|f\rangle = T_{-1}(\lambda_\mu a_{n-1,\mu}^+ + \chi_{\mu\nu} a_{n-2,\mu}^+ a_{1,\nu}^+ + \dots)|0, p\rangle$$

$$p^2/2 = 1 - n \quad (9.297)$$

This state is physical, when  $T_{-1}$  acts on a physical state of zero dimension. Therefore, the number of physical states with nonzero norm  $v_{\mathcal{D}}(n)$  is given by:

$$v_{\mathcal{D}}(n) = N_{\mathcal{D}-1}(n) - N_{\mathcal{D}-1}(n-1)$$

$$v(x) = \sum_n v_{\mathcal{D}}(n)x^n \quad (9.298)$$

$$= (1-x) \left( \prod_{k=1}^{\infty} \frac{1}{1-x^k} \right)^{\mathcal{D}-1}$$

In the critical dimension  $\mathcal{D} = 26$  many more states have zero norm. We have already observed this in the case of the second level. For the general case we have to use the Kac formula. Let us see when some conformal family with  $c = \mathcal{D} = 26$  has a physical secondary operator at some level, with dimension one. Using (9.272) we obtain:

$$\Delta_{nm} + nm = \frac{25}{24} + \frac{1}{8} (n\alpha_+ + m\alpha_-)^2$$

$$= \frac{25}{24} - \frac{1}{24} (3n - 2m)^2 = 1 \quad (9.299)$$

We see that degeneracy (or zero norm states) is possible provided that:

$$3n = 2m \pm 1 \quad (9.300)$$

and the corresponding level is given by:

$$N = n \cdot m = \frac{n(3n \pm 1)}{2} \quad (9.301)$$

The previously found state with  $N = 2$  has  $n = 1, m = 2$ . It is clear, that now, instead of (9.298) we will have:

$$v(x) = (1 - x - x^2 - \dots) \left( \prod_{k=1}^{\infty} \frac{1}{1-x^k} \right)^{25} \quad (9.302)$$

However, we should not simply subtract all levels given by (9.301). That would lead to overcounting for the following reason. Among the zero norm states on the second level, which as we said have the form:

$$|f\rangle = (T_{-2} + \frac{3}{2}T_{-1}^2)|\Delta = -1\rangle \tag{9.303}$$

there are some, for which the state  $|\Delta = -1\rangle$  is itself secondary. These cases are again found from the Kac formula:

$$\begin{aligned} \Delta_{nm} + nm &= \frac{25}{24} - \frac{1}{24} (3n - 2m)^2 = -1 \\ 3n - 2m &= \pm 7 \end{aligned} \tag{9.304}$$

which is satisfied, say, on the level 3, by  $n = 3, m = 1$ . Therefore, some of the states (9.303) will be of the form:

$$|f\rangle = (T_{-2} + \frac{3}{2}T_{-1}^2)(T_{-3} + \dots)|\Delta = -4\rangle \tag{9.305}$$

This can be considered as a degenerate state at the level 5. So, this state should *not* be counted as we go to the level 5, setting  $n = 2$  in (9.301). Another state which should not be counted is the following:

$$\begin{aligned} |f\rangle &= T_{-1}|\Delta = 0\rangle \\ |\Delta = 0\rangle &= \{T_{-4} + \dots\}|\Delta = -4\rangle \end{aligned} \tag{9.306}$$

because according to the Kac formula:

$$\Delta_{1,4} + 4 = \frac{25}{24} - \frac{1}{24} (3 - 8)^2 = 0$$

On the other hand, the Kac formula predicts only one secondary state at the level 5, and we have already counted two states. The conclusion is that these two states are in fact the same, and that we have counted it twice. Hence, to correct the situation, we must write instead of (9.302):

$$v(x) = (1 - x - x^2 + x^5 + \dots) \left( \prod_k \frac{1}{1 - x^k} \right)^{25} \tag{9.307}$$

It is not hard to do the same combinatorics† at each level, using the Kac formula repeatedly. The result of this counting is:

$$\begin{aligned} v(x) &= \left\{ 1 + \sum_{n=1}^{\infty} (-1)^n (x^{n(3n+1)/2} + x^{n(3n-1)/2}) \right\} \\ &\times \left( \prod_{k=1}^{\infty} \frac{1}{1 - x^k} \right)^{25} = \left( \prod_{k=1}^{\infty} \frac{1}{1 - x^k} \right)^{24} \end{aligned} \tag{9.308}$$

---

† Described in the Appendix to Ch. 9.10, p. 221.

Here we have used an identity due to Gauss for the first factor. This result is quite remarkable. It shows that among the 25 chains of oscillators one chain forms zero norm states and decouples. This implies that all the remaining oscillators have positive norms and there are no ghosts in our theory. The reason is roughly speaking that all negative norm components went to form zero norm states. Indeed, any zero norm state  $|z\rangle$  can be represented as

$$|z\rangle = |n\rangle + |p\rangle \quad (9.309)$$

where

$$\langle n|n\rangle = -1, \quad \langle p|p\rangle = 1$$

Any physical state  $|f\rangle$  has the property

$$\langle f|z\rangle = 0 \quad (9.310)$$

which implies:

$$\begin{aligned} |f\rangle &= |n\rangle + |\tilde{p}\rangle \\ \langle \tilde{p}|p\rangle &= 1 \end{aligned} \quad (9.311)$$

But, since  $|p\rangle$  and  $|\tilde{p}\rangle$  lie in the positive norm Hilbert state, the Cauchy inequality implies:

$$\langle \tilde{p}|\tilde{p}\rangle > \langle \tilde{p}|p\rangle = 1$$

Hence:

$$\langle f|f\rangle = \langle \tilde{p}|\tilde{p}\rangle + \langle n|n\rangle > 0 \quad (9.312)$$

When we have many negative norm states, it is sufficient for this “no-ghost” theorem that their number at each level should be equal to the number of zero norm states, which is the case, according to (9.308).

We have arrived at the conclusion, that for  $\mathcal{D} = 26$  we have 24 chains of oscillators with positive norms. This result could have been expected, since in this case the Liouville field decouples, and the only physical fields are those of the coordinates  $x_\mu(\xi)$  of the string. They would describe  $\mathcal{D} = 26$  chains of oscillators, but owing to the general coordinate transformations  $\xi \rightarrow f(\bar{\xi})$ , involving two arbitrary functions, two chains are unphysical. This is just what we have seen by explicit computation. For  $\mathcal{D} > 26$ , the Liouville field will have the wrong sign of the kinetic energy, which has two interpretations. The first one is that large gradients of the  $\varphi$ -field become important, so that the surface loses its continuum limit. The second, more formal, interpretation, is that the wrong sign of the  $\varphi$ -propagator implies negative norm states.

This can be confirmed by examining the norms of physical states for  $\mathcal{D} > 26$ : one chain of oscillators will become ghosts.

The main result of our analysis is that for  $\mathcal{D} = 26$  the simple Koba-Nielsen amplitude has only one flaw—a tachyon. In the next section we will show how this flaw is corrected in the case of fermionic strings with  $\mathcal{D} = 10$ . After that, we return to the noncritical strings and discuss many interesting possibilities which arise there.

*Appendix to 9.10*

Here we will briefly describe the combinatorics needed for the derivation of (9.308). Our task is to subtract the contribution of all zero-norm states from the partition function  $N^{\text{ph}}$  of physical states

$$N^{\text{ph}}(z) = \sum_{n=0}^{\infty} N^{\text{ph}}(n)z^n = \prod_{n=1}^{\infty} (1 - z^n)^{-25}$$

Here  $N^{\text{ph}}(n)$  denotes the number of physical states (primary fields) at level  $n$ , and we put  $\mathcal{D} = C = 26$ . There are two cases when the physical state at a given level has zero norm:

$$(*) \quad N = n \cdot m, \quad \Delta_{n,m} + n \cdot m = \Delta_{n,-m} = 1$$

where  $\Delta_{n,m} = (25 - (2n + 3m)^2)/24$ , as follows from (9.272) for  $c = 26$ . From (\*) we get  $(2n - 3m)^2 = 1$ , or

$$N \in \Lambda_{\mu}^{(-)} \equiv \left\{ \frac{m(3m + 1)}{2}, \quad m \in 2\mathbb{Z} + 1 \right\}$$

$$(**) \quad N = n \cdot m + k \cdot l, \quad \Delta_{n,m} + n \cdot m = \Delta_{k,l}, \quad \Delta_{k,l} + kl$$

From the second equality follows

$$\Delta_{n,m} = \Delta_{k,l} \quad \text{or} \quad k = n + 3q, \quad l = -m - 2q, \quad q \in \mathbb{Z}$$

The third relation gives  $(2k - 3l)^2 = (2n + 3m + 12q)^2 = 1$ , and we find

$$N = nm + kl = -3qm - 2qn - 6q^2 = -6q^2 + q(12q \pm 1) = q(6q \pm 1)$$

or

$$N \in \Lambda^{(+)} \equiv \left\{ \frac{p(3p + 1)}{2}, \quad p \in 2\mathbb{Z} \setminus \{0\} \right\}$$

### 9.11 Fermi Particles

There exists a beautiful extension of the bosonic string, the Neveu-Schwarz-Ramond string (NSR) which represents a string analogue of Dirac particles. We will start this section by describing the path integral representation for these particles and then generalize it to the string case.

Fermions can be described in many different ways. One of them, from which we start, is to replace an ordinary path  $x_\mu = x_\mu(t)$  by a super "surface". Namely to consider a function

$$\begin{aligned}\hat{x}_\mu &= \hat{x}_\mu(t, \theta) = x_\mu(t) + \theta\psi_\mu(t) \\ \theta^2 &= 0\end{aligned}\tag{9.313}$$

The superspace  $(t, \theta)$  possesses the symmetry:

$$\begin{aligned}\delta_\varepsilon t &= \varepsilon\theta \\ \delta_\varepsilon\theta &= \varepsilon, \quad \varepsilon^2 = 0; \quad \varepsilon\theta + \theta\varepsilon = 0\end{aligned}\tag{9.314}$$

The commutator of two such transformations is an ordinary shift in  $t$ :

$$\begin{aligned}[\delta_{\varepsilon_2}\delta_{\varepsilon_1} - \delta_{\varepsilon_1}\delta_{\varepsilon_2}]t &= \delta_{\varepsilon_2}(\varepsilon_1\theta) - \delta_{\varepsilon_1}(\varepsilon_2\theta) \\ &= \varepsilon_1\varepsilon_2 - \varepsilon_2\varepsilon_1 = 2\varepsilon_1\varepsilon_2 \\ [\delta_{\varepsilon_2}, \delta_{\varepsilon_1}]\theta &= 0\end{aligned}\tag{9.315}$$

We can define a covariant derivative  $\mathcal{D}$  by:

$$\begin{aligned}\varepsilon\mathcal{D} &= \delta_\varepsilon\theta\frac{\partial}{\partial\theta} + \delta_\varepsilon t\frac{\partial}{\partial t} = \varepsilon\left(\frac{\partial}{\partial\theta} + \theta\frac{\partial}{\partial t}\right) \\ \mathcal{D} &= \frac{\partial}{\partial\theta} + \theta\frac{\partial}{\partial t} \\ \mathcal{D}^2 &= \frac{\partial}{\partial t}\end{aligned}\tag{9.316}$$

The supercovariant action is given by:

$$\begin{aligned}S &= \frac{1}{2} \int dt d\theta \mathcal{D}^2 \hat{x}_\mu \mathcal{D} \hat{x}^\mu \\ &= \int dt \left( \frac{1}{2} \dot{x}^2 - \frac{1}{2} \psi_\mu \dot{\psi}_\mu \right)\end{aligned}\tag{9.317}$$

We shall see in a moment that the fields  $\psi_\mu$  will play the role of the  $\gamma$ -matrices in the usual approach. The action (9.317) is invariant under the supersymmetry:

$$\begin{aligned} \delta x_\mu &= \varepsilon \psi_\mu \\ \delta \psi_\mu &= \varepsilon \dot{x}_\mu \end{aligned} \tag{9.318}$$

The action (9.317) is only globally invariant, while the proper action should be invariant under general supercovariant transformations. This can be achieved by the introduction of the “einbein” field  $e(t)$  (as we have done in the bosonic case) and its superpartner—the “gravitino” field  $\chi(t)$ . While  $e(t)$  is coupled to the energy-momentum tensor, which in our case is  $\frac{1}{2}\dot{x}^2$ , the field  $\chi(t)$  must be coupled to the supercurrent. It is clear from (3.318) that this symmetry is associated with the conserved current:

$$s(t) = \psi_\mu \dot{x}_\mu; \quad \dot{s} = 0 \tag{9.319}$$

Therefore, we can expect that the covariant action will be:

$$S = \int \left\{ \frac{1}{2} (e^{-1} \dot{x}_\mu^2 - \psi_\mu \dot{\psi}_\mu) + \frac{1}{2e} \chi \psi_\mu \dot{x}_\mu \right\} dt \tag{9.320}$$

This is indeed the case. It is easy to check that (9.320) is invariant under:

$$\begin{aligned} \delta x_\mu(t) &= \alpha(t) \psi_\mu(t) \\ \delta \psi_\mu(t) &= \alpha(t) \left( \frac{\dot{x}_\mu}{e(t)} + \frac{1}{2e} \chi \psi_\mu \right) \\ \delta e(t) &= \alpha(t) \chi(t) \\ \delta \chi(t) &= -2\dot{\alpha}(t) \end{aligned} \tag{9.321}$$

We are almost ready to write a complete analogue of the bosonic functional integral. There is, however, a small problem to overcome. Namely, the “cosmological” term  $\int e(t) dt$ , which we added in the bosonic case, is not invariant under (9.321). Moreover, no local expression made of  $\chi$  and  $e$  will be invariant. The “super-length”, which is invariant, has the form:

$$L = \int_0^1 e(\tau) d\tau - \frac{1}{8} \int_0^1 d\tau_1 d\tau_2 \text{sign}(\tau_1 - \tau_2) \chi(\tau_1) \chi(\tau_2) \tag{9.322a}$$

We can avoid the nonlocal form (9.322a) at the price of introducing an extra field  $\psi_5$  with which (9.322a) can be replaced by

$$L = \int_0^1 e(\tau) d\tau - \int \psi_5 \dot{\psi}_5 d\tau + \int \psi_5 \chi d\tau \quad (9.322b)$$

$$(\delta\psi_5 = -\alpha)$$

Integrating out  $\psi_5$  returns us to (9.322a).

Now, the functional integral for a supersymmetric path is†

$$Z = \int \mathcal{D}x_\mu \mathcal{D}\psi_\mu \mathcal{D}\chi \mathcal{D}e \mathcal{D}\psi_5 \exp \left\{ - \int_0^1 \left[ \frac{\dot{x}^2}{e} - \psi_\mu \dot{\psi}_\mu - \psi_5 \dot{\psi}_5 + \frac{1}{e} (\chi \psi_\mu \dot{x}_\mu - m e \psi_5) + m^2 e \right] d\tau \right\} \quad (9.323)$$

For a free particle, which we are considering, this integral is easily computed, using the following gauge fixing:

$$\dot{e} = 0, \quad e = T, \quad \dot{\chi} = 0, \quad \chi = \theta \quad (9.324)$$

(where  $T$  and  $\theta$  should be integrated over, since gauge freedom is not sufficient to eliminate them completely). For an open path from  $(x_\mu, \psi_\mu)$  to  $(x'_\mu, \psi'_\mu)$  (the quantity  $\psi_\mu$  must not change since it satisfies a first order differential equation in contrast with  $x_\mu$ ) we have in the momentum representation (changing  $(1/T)\dot{x}_\mu \rightarrow ip_\mu$ ):

$$G(p) = \int_0^\infty dT e^{-m^2 T} \int d\theta e^{\theta(ip_\mu \psi_\mu - m\psi_5)} e^{-p^2 T}$$

$$= \frac{1}{p^2 + m^2} (ip_\mu \psi_\mu + m\psi_5) \quad (9.325)$$

If we change  $\psi_\mu \rightarrow \gamma_\mu$  and  $\psi_5 \rightarrow \gamma_5$  we obtain the propagator for Dirac particles.

The formulas for closed loops, similar to (9.65), also can be written. We shall not give a detailed derivation here. The answer replacing

---

† This form was found by Brink, Di Vecchia, How (1975) and anticipated by Bezezin and Mazinov (1974).

(9.65) is given by

$$\begin{aligned}
 F(q_1, \dots, q_N) &= \int_0^\infty dT T^{-\mathcal{D}/2} e^{-m^2 T} \int_{0=\tau_1 < \tau_2 < \dots < \tau_N < T} d\tau_2 d\theta_2 \dots d\tau_N d\theta_N \\
 &\times \exp \left[ -\frac{1}{2} \sum_{i \neq j} \mathbf{q}_i \cdot \mathbf{q}_j \mathcal{D}(\hat{s}_{ij}) \right]; \tag{9.65a} \\
 \mathcal{D}(s) &= \frac{s(T-s)}{T}
 \end{aligned}$$

where  $\hat{s}_{ij}$  is the superinvariant distance:

$$\hat{s}_{ij} = |\tau_i - \tau_j| - \theta_i \theta_j \operatorname{sign}(\tau_i - \tau_j) \tag{9.326}$$

One can easily check that, first,  $\hat{s}_{ij}$  is invariant under (9.314) and, second, the representation (9.65a) is equivalent to taking the trace of Dirac matrices. Thus, (9.65a) gives us the supersymmetric extension of the Feynman-Schwinger parametrization appropriate for fermions. Notice also, that the exponent in (9.65a) is just the average of the product of vertex operators

$$\begin{aligned}
 V_p(\tau, \theta) &= e^{i\mathbf{p} \cdot \hat{\mathbf{x}}(\tau, \theta)} = e^{i\mathbf{p} \cdot (\mathbf{x} + \theta\boldsymbol{\psi})} \\
 &= [1 + i\theta(\mathbf{p} \cdot \boldsymbol{\psi})] e^{i\mathbf{p} \cdot \mathbf{x}} \tag{9.327}
 \end{aligned}$$

Before we go to the case of strings, several more points are worth discussing.

First of all, the geometrical properties of the fermionic path are very different from those of the bosonic one. In particular, while in the bosonic case we have a Brownian path with its size  $R$  and length  $L$  connected by the diffusion law  $R^2 \sim L$ , in the Fermi case things are different, namely  $R^2 \sim L^2$ . To show this, let us recall our games with the Lagrange multiplier. In (9.12) we argued that in the action  $\int \lambda(t)(\dot{x}^2/e - e) dt$  for a bosonic path we can replace  $\lambda(t)$  by some constant because  $\lambda(t)$  develops a constant vacuum expectation value. A crude argument for that was based on a fact that the effective action for  $\lambda$  contains a term:

$$W[\lambda(t)] = \frac{1}{a} \int_0^1 e(t) dt \log \lambda(t) - \int_0^1 \lambda(t) e(t) dt \tag{9.328}$$

where  $a$  is the lattice cut-off. Because of the first term we obtain  $\langle \bar{\lambda} \rangle \sim 1/a$ . For the supersymmetric action the situation is different. The

first term in (9.328) coming from the zero-point energy of the  $x$ -field cancels with the contribution of the  $\psi$ -fields. As a result, we do not have a divergent contribution to  $W[\lambda]$ . Instead, the coefficient  $1/a$  in the first term can be shown to get replaced by  $1/L$ . Because of this, for the fermionic path we have:

$$\overline{R^2} \sim \frac{1}{\langle \lambda \rangle} L \sim L^2 \quad (9.329)$$

This result is responsible for the different critical behaviour of fermionic particles.

There is also another interesting form into which the action (9.320) can be recast. Namely, let us integrate out the  $\chi$ -field using (9.322a) to find its propagator:

$$\langle \chi(\tau)\chi(\tau') \rangle = \frac{\partial}{\partial \tau} \delta(\tau - \tau') \quad (9.330)$$

Performing the  $\chi$  integration, we obtain, instead of (9.322a):

$$Z = \int \mathcal{D}\Psi \mathcal{D}x \exp \left[ - \int_0^T (\dot{x}_\mu^2 - \psi_\mu \dot{\psi}_\mu) dt + \frac{1}{4} \int_0^T (\psi_\mu \dot{x}_\mu)' \psi_\nu \dot{x}_\nu dt \right] \quad (9.331)$$

The last factor has a remarkable interpretation. If we replace  $\psi_\mu$  by  $\gamma_\mu$  (which is the actual role of the  $\psi$  integral) we cast this factor into the form:

$$\phi[c] = P \exp \frac{1}{8} \int_0^T \omega_{\mu\nu}[x(\tau)][\gamma_\mu, \gamma_\nu] d\tau \quad (9.332)$$

where  $\omega_{\mu\nu}[x(\tau)]$  is the rotation of the tangent vector to the trajectory:

$$\omega_{\mu\nu}[x(\tau)] = \frac{1}{2}(\dot{x}_\mu \ddot{x}_\nu - \ddot{x}_\mu \dot{x}_\nu), \quad (\dot{x}^2 = 1) \quad (9.333)$$

This gives us a new understanding of the meaning of path integrals for fermions. In particular, in two dimensions the factor (9.332) reduces to  $(-1)^v$  where  $v$  is the total rotation of the tangent, known to be equal to the number of self-intersections of the path. Because of these oscillating factors, many complicated trajectories are suppressed, and this is the reason for (9.329).

To complete our understanding of fermions, let us discuss another representation, which combines Fermi and Bose particles together. That is, it is possible to have a functional integral which is supersym-

metric not only in the “inner” space  $(\theta, \tau)$  but also in the “outer” space  $x_\mu$ . The  $\mathcal{D}$ -dimensional supersymmetry is described by the following transformations: if we define the superspace as the one described by  $x_\mu$  and a  $\mathcal{D}$ -dimensional spinor  $\varphi$  we have:

$$\begin{aligned} \delta x_\mu &= \bar{\varepsilon} \gamma_\mu \varphi \\ \delta \varphi &= \varepsilon \end{aligned} \tag{9.334}$$

Now we can define a propagator in this superspace as an amplitude  $\langle \tilde{x} \tilde{\varphi} | x \varphi \rangle$ . It describes the propagation of a superfield  $\phi(x, \varphi)$  which contains both Fermi and Bose fields in itself. Our aim now is to give an action invariant under (9.334). This is not hard. Consider the expression:

$$S = \int_0^T dt \frac{1}{2} (\dot{x}_\mu - \bar{\varphi} \gamma_\mu \dot{\varphi})^2 \tag{9.335a}$$

Invariance under (9.334) is obvious, so we can postulate that the functional integral over  $x$  and  $\varphi$  is just the required propagator. What is less obvious is how this new representation is connected with the previous ones, say (9.331). Details of this connection have never been worked out. I shall give here only the general idea. Let us begin from (9.332). It is easy to see that we have the following formula:

$$\lim_{a \rightarrow 0} P \exp \left( \frac{1}{a} \int_c \gamma_\mu(\tau) \dot{x}_\mu(\tau) d\tau \right) = e^{L/a} \phi[c] \tag{9.335b}$$

This can be checked by use of the definition of the ordered product. Now we can replace the ordered product (modulo the divergent factor) by the functional integral:

$$\begin{aligned} &P \exp \left\{ \frac{1}{a} \int \gamma_\mu(\tau) \dot{x}_\mu(\tau) d\tau \right\} \\ &= \int \mathcal{D}\chi \exp \left[ \int \left( \dot{\chi} \chi + \frac{1}{a} \chi \gamma_\mu \dot{\chi} \right) d\tau \right] \\ &= \int \mathcal{D}\chi \exp \left[ \int (\chi \gamma_\mu \dot{\chi} + a \dot{\chi} \chi) d\tau \right] \end{aligned} \tag{9.336}$$

If we introduce the field  $\varphi$  by the relation  $\chi = \sqrt{(d/dt)}\varphi$  we shall find that the spinor factor (9.335) in the limit  $a \rightarrow 0$  can be considered as an

amplitude for the propagation of the  $\mathcal{D}$ -dimensional fermion trapped on the path:

$$\phi[c] = \int \mathcal{D}\varphi \exp\left(\int \varphi \gamma_\mu \dot{\varphi} \dot{x}_\mu d\tau\right) \quad (9.337)$$

This is almost (9.335). Presumably the quartic term is needed for proper regularization. But how come that the supersymmetric expression (9.335), describing bosons and fermions together, is obtained from the purely fermionic path integral? The answer to this question is simple. When we treat the functional integral (9.337) we find that it describes a second quantized Dirac particle, living on the path. In order to reproduce fermionic results in (9.336) we implicitly assumed that the occupation number for this particle is equal to one. If we remove this constraint and integrate over all possible  $\varphi$ -fields, we add to this sector another one, with the occupation number zero. In this other case we have no particle on the path and thus describe a boson.

There are obvious gaps in our derivation, but I believe they can be filled without damage to the general ideas outlined above. This would be a theme for an interesting investigation.

## 9.12 Fermionic Strings

We shall now describe the string analogue of a Dirac particle. Just as in the latter case we had a field  $\psi_\mu$  distributed on the trajectory, which eventually played the role of the Dirac matrices  $\gamma_\mu$ , in the case of strings we have to consider a field  $\psi_\mu(\xi)$  living on the world sheet, which should be a supersymmetric partner of the field  $x_\mu(\xi)$ . As before, it is convenient to start from the superspace, which now must have two fermionic directions. If we describe  $\xi$ -space by complex variables  $z$  and  $\bar{z}$ , each of them has its Fermi-partner  $\theta$  and  $\bar{\theta}$ . The supersymmetry transformation is given by:

$$\begin{aligned} \delta z &= -\varepsilon\theta, & \delta \bar{z} &= -\bar{\varepsilon}\bar{\theta} \\ \delta \theta &= \varepsilon, & \delta \bar{\theta} &= \bar{\varepsilon} \end{aligned} \quad (9.338)$$

We see that we have a direct product of one dimensional supergroups. The corresponding covariant derivatives take the form:

$$\mathcal{D} = \frac{\partial}{\partial\theta} + \theta \frac{\partial}{\partial z}, \quad \bar{\mathcal{D}} = \frac{\partial}{\partial\bar{\theta}} + \bar{\theta} \frac{\partial}{\partial \bar{z}} \quad (9.339)$$

The field  $x_\mu(\xi)$  should be replaced by the superfield  $\hat{x}_\mu(z, \bar{z}, \theta, \bar{\theta})$  which has the decomposition:

$$\hat{x}_\mu(z, \bar{z}, \theta, \bar{\theta}) = x_\mu(\xi) + \theta\psi_\mu + \bar{\theta}\bar{\psi}_\mu + \theta\bar{\theta}f_\mu \tag{9.340}$$

We see that now we have two types of fermi-fields  $\psi_\mu$  and  $\bar{\psi}_\mu$  (in Euclidean space  $\bar{\psi}_\mu = \psi_\mu^*$  while in Minkowski space they are two real independent fields, just as happens to the coordinates  $z$  and  $\bar{z}$  which in  $M$ -space become  $\xi^0 \pm \xi^1$ ). These two fields form a spinor of the two dimensional space and a vector in the external space. We have also a new world sheet scalar  $f_\mu$ . A supersymmetric action is easily formed out of this material:

$$\begin{aligned} S &= \frac{1}{2} \int \bar{\mathcal{D}}x_\mu \mathcal{D}x_\mu d\theta d\bar{\theta} d^2\xi \\ &= \frac{1}{2} \int ((\partial x_\mu)^2 - \psi_\mu \partial_{\bar{z}}\psi_\mu - \bar{\psi}_\mu \partial_z\bar{\psi}_\mu + f_\mu^2) d^2\xi \end{aligned} \tag{9.341}$$

This action has the following conserved quantities, associated with its supersymmetry:

$$\begin{aligned} \mathcal{J} + \theta T &= -\frac{1}{2} \mathcal{D}\hat{x}\mathcal{D}^2\hat{x} \\ T &= -\frac{1}{2}(\partial_z x_\mu)^2 + \frac{1}{2}\psi_\mu \partial_z\psi_\mu \\ \bar{T} &= -\frac{1}{2}(\partial_{\bar{z}} x_\mu)^2 + \frac{1}{2}\bar{\psi}_\mu \partial_{\bar{z}}\bar{\psi}_\mu \\ \mathcal{J} &= -\frac{1}{2}\psi_\mu \partial_z x_\mu \\ \bar{\mathcal{J}} &= -\frac{1}{2}\bar{\psi}_\mu \partial_{\bar{z}} x_\mu \\ \partial_{\bar{z}} T &= \partial_z \bar{T} = \partial_z \mathcal{J} = \partial_{\bar{z}} \bar{\mathcal{J}} = 0 \end{aligned} \tag{9.342}$$

Here we have the energy momentum tensor  $T \equiv T_{++}$  with conformal spin equal to two and the supercurrent  $\mathcal{J} \equiv \mathcal{J}_+^L$  with spin 3/2 (since the fermion  $\psi \equiv \psi_L$  has spin 1/2). It is to be remembered that by the “conformal spin” we mean the transformation law under  $z \rightarrow e^{i\alpha z}$ . If some quantity is multiplied by  $e^{-is\alpha}$  we say that it has spin  $s$ . It is clear, that in order to achieve local supersymmetry, we have to introduce into (9.341) two gauge fields, one with spin 2 coupled to  $T(\bar{T})$ , which we call the “graviton”, and another with spin 3/2, coupled to  $\mathcal{J}(\bar{\mathcal{J}})$  which we call the “gravitino”.

The action which has local supersymmetry can be found to be:

$$\begin{aligned} S &= \frac{1}{2} \int d^2\xi g^{1/2} [g^{ab} \partial_a x \cdot \partial_b x + \psi \cdot e^{a i}(\xi) \sigma^i \partial_a \psi \\ &\quad + (\chi_a \sigma^b \sigma^a \psi_\mu)(\partial_b x_\mu - \frac{1}{4}(\chi_b \psi_\mu))] \end{aligned} \tag{9.343}$$

One would like to derive (9.343) directly from the superspace formalism. Unfortunately, this becomes rather clumsy. It is much easier to check that (9.343) has the desired covariance. In this formula we denoted by  $\sigma^i$  the usual Pauli matrices,  $e^{ai}(\xi)$  is a "Zweibein" field connected with  $g_{ab}$  by  $g_{ab} = e_a^i(\xi)e_b^j(\xi)$ ,  $\chi_a$  is the gravitino field with vector index  $a$  and suppressed spinor index. In the limit of weak graviton and gravitino fields, the action (9.343) will go to (9.341) plus a graviton coupled to  $T(\bar{T})$  and a gravitino to  $\mathcal{S}(\bar{\mathcal{F}})$ . The last term in (9.343) is a correction to the supercurrent produced by the  $\chi$ -field. We have used a non-covariant derivative acting on  $\psi_\mu$  in the second term since for Majorana spinors the spin connection gives zero contribution.

The symmetry of (9.343) apart from the usual general covariance is described by the transformations:

$$\begin{aligned}\delta x_\mu(\xi) &= \varepsilon(\xi)\psi_\mu(\xi) \\ \delta\psi_\mu(\xi) &= \sigma^a[\partial_a x_\mu(\xi) - \frac{1}{2}(\chi_a\psi_\mu)]\varepsilon(\xi) \\ \delta\chi_a &= 2\hat{\nabla}_a\varepsilon(\xi) \\ \delta g_{ab} &= \varepsilon(\sigma_a\chi_b + \sigma_b\chi_a)\end{aligned}\tag{9.344}$$

with  $\varepsilon(\xi)$  being a Majorana spinor.

Just as in the purely bosonic case, the theory is greatly simplified by the choice of the conformal gauge. This gauge is given by:

$$\begin{aligned}g_{ab}(\xi) &= e^{\sigma(\xi)}\delta_{ab} \\ \chi_a(\xi) &= \sigma_a\chi(\xi)\end{aligned}\tag{9.345}$$

Let us discuss whether it is possible to reach this gauge by the use of general supercovariance. First of all, we count the number of independent functions in the gauge (9.345) which gives us a rough orientation of the situation. We have 3 components of  $g_{ab}$  and two spinors  $\chi_a$ , or in other words 3 bosonic functions and 4 fermionic ones. We replace them by one bosonic  $\varphi$  and two fermionic  $\chi$ . This is reasonable, since we have two extra bosonic functions, describing general covariant transformation from the gauge (9.345) to an arbitrary one, and two fermionic ones which enter in (9.344). All in all, the number of independent functions matches.

Now, we have to repeat the more detailed analysis, which we have already done in the bosonic case.

In order to decide on the accessibility of the gauge (9.345), let us consider a general variation of  $\chi_a$  and examine whether we can write it as some variation of  $\chi$  plus a supercovariant transformation. We have:

$$\begin{aligned}\delta\chi_a &= \sigma_a\delta\chi + 2\hat{\nabla}_a\varepsilon \\ &= \sigma_a(\delta\chi + \sigma^b\nabla_b\varepsilon) + 2(\hat{\nabla}_a\varepsilon - \frac{1}{2}\sigma_a\sigma^b\hat{\nabla}_b\varepsilon)\end{aligned}\tag{9.346}$$

The meaning of the last expression in the brackets is that it is “traceless” in the sense that being multiplied by  $\sigma^a$  it gives zero. All this is quite analogous to the bosonic case (9.113). Accessibility of the gauge depends on the possibility of solving the equation

$$(L_F \varepsilon) \equiv \hat{\nabla}_a \varepsilon - \frac{1}{2} \sigma_a \sigma^b \nabla_b \varepsilon = \varphi_a \tag{9.347}$$

where  $\varphi_a$  is an arbitrary spin 3/2 spinor with  $\sigma^a \varphi_a = 0$ . This is to be combined and compared with the bosonic part of this equation (9.113):

$$(L_B \omega) \equiv \nabla_a \omega_b + \nabla_b \omega_a - g_{ab} \nabla^c \omega_c = h_{ab} \tag{9.348}$$

with arbitrary traceless  $h_{ab}$ . As we shall show now, the ghost determinants are just those of the operators  $L_B$  and  $L_F$ . In the purely bosonic case we have already derived this result (remember that we found a  $\det(L_B^+ L_B)$  factor in the functional integral). It is almost obvious that for the fermionic case we have to add  $\det^{-1}(L_F^+ L_F)$ . Here we shall derive this result by a different (from the bosonic case) method, which seems quite useful to know.

Let us take a super-metric  $(g, \chi_a)$  in the superconformal gauge

$$\begin{aligned} g_{ab} &= \rho \delta_{ab} \\ \chi_{a,\alpha} &= (\sigma_a \chi)_\alpha \end{aligned} \tag{9.349}$$

( $\alpha$  is a spinor index) and consider the Faddeev-Popov equality:

$$1 = e^{-W(g, \chi_a)} \int \mathcal{D}\omega \mathcal{D}\varepsilon \delta(g - L_B \omega) \delta(\chi_a - L_F \varepsilon) \tag{9.350}$$

which serves as a definition of  $W$ . Inserting this into the functional integral for the partition function we get

$$Z = \int \mathcal{D}\rho \mathcal{D}\chi \mathcal{D}\psi_\mu \mathcal{D}x_\mu e^{-W(\rho, \chi)} e^{-S(\rho, \chi, \psi_\mu, x_\mu)} \tag{9.351}$$

where we have omitted integration over the supercovariance group  $\mathcal{D}\omega \mathcal{D}\varepsilon$ . Now, it is convenient to introduce ghost fields explicitly by representing:

$$\begin{aligned} e^{-W(g, \chi_a)} &= \det L_B \det^{-1} L_F \\ &= \int (\mathcal{D}\omega \mathcal{D}h)^{-1} (\mathcal{D}\varepsilon \mathcal{D}f)^{(+)} \exp\{- (h, L_B \omega) - (f, L_F \varepsilon)\} \end{aligned} \tag{9.352}$$

The signs  $(-)$  and  $(+)$  here indicate that we integrate over fields with reversed statistics:  $\omega$  and  $h$  are fermions, while  $\varepsilon$  and  $f$  are bosons.

Now, our aim is to express the determinants entering in these formulas in terms of the fields  $\chi$  and  $\rho$  from (9.349). In the bosonic case we have already performed this task. However, it will be convenient to repeat everything again, using a somewhat different method, in the present situation, in order to see the close connection between bosons and fermions.

First of all, let us rewrite the operators (9.348) and (9.347) in the conformal gauge. For (9.348) we have the standard formulas of Riemann geometry:

$$\begin{aligned}\nabla_a \omega_b &= \partial_a \omega_b - \Gamma_{ab}^c \omega_c \\ \Gamma_{ab}^c &= \frac{1}{2} g^{cd} (\partial_a g_{db} + \partial_b g_{da} - \partial_d g_{ab}) \\ &= \frac{1}{2} (\partial_a \varphi \delta_{bc} + \partial_b \varphi \delta_{ac} - \partial_c \varphi \delta_{ab}) \quad (\varphi = \log \rho)\end{aligned}\quad (9.353)$$

From this we derive:

$$\begin{aligned}\nabla_a \omega_b + \nabla_b \omega_a - g_{ab} \nabla^c \omega_c &= \partial_a \omega_b + \partial_b \omega_a - \delta_{ab} \partial_c \omega_c \\ &\quad - (\partial_a \varphi \omega_b + \partial_b \varphi \omega_a) + \delta_{ab} \partial_c \varphi \omega_c \\ &= e^\varphi (\partial_a \tilde{\omega}_b + \partial_b \tilde{\omega}_a - \delta_{ab} \partial_c \tilde{\omega}_c), \\ \tilde{\omega}_a &= e^{-\varphi} \omega_a\end{aligned}\quad (9.354)$$

When computing the conjugate operator we need to know:

$$\begin{aligned}(L_B^+ h) &= \nabla^a h_{ab} \\ &= g^{ac} (\partial_c h_{ab} - \Gamma_{ca}^e h_{eb} - \Gamma_{cb}^e h_{ae}) \\ &= e^{-\varphi} (\partial_a h_{ab})\end{aligned}\quad (9.355)$$

With fermions a similar thing happens. According to general rules:

$$\nabla_a \varepsilon = \partial^a \varepsilon + \frac{1}{4} \Omega_a^{[bc]} [\sigma_b, \sigma_c] \varepsilon \quad (9.356)$$

where the connection  $\Omega_a^{[bc]}$  is defined by:

$$\begin{aligned}\partial_a e_b^c + \Omega_a^{[cd]} e_b^d &= \Gamma_{ab}^f e_f^c \\ g_{ab} &= e_a^c e_b^c\end{aligned}\quad (9.357)$$

In the conformal gauge:

$$e_a^b = \rho^{1/2} \delta_{ab}$$

and we find:

$$\Omega_a^{[bc]} = -\frac{1}{4} (\partial_b \varphi \delta_{ac} - \partial_c \varphi \delta_{ab}) \quad (9.358)$$

From this:

$$\nabla_a \varepsilon - \frac{1}{2} \sigma_a \sigma^b \nabla_b \varepsilon = e^{\varphi/2} [\partial_a (e^{-\varphi/2} \varepsilon) - \sigma_a \sigma^b \partial_b (e^{-\varphi/2} \varepsilon)]$$

For the conjugate spin 3/2 field  $f_a$  we have:

$$\nabla_a f_b = \partial_a f_b - \Gamma_{ab}^c f_c + \frac{1}{4} \Omega_a^{[pq]} [\sigma^p, \sigma^q] f_b \tag{9.359}$$

In the conformal gauge:

$$\nabla^a f_a = \rho^{-1} \partial_a f_a \tag{9.360}$$

This extensive use of general formulas was not actually needed. The results (9.354), (9.355), (9.359), (9.360) could have been foreseen from the following reasons. We can define the covariant derivative in the conformal gauge as follows. Suppose that we have a field  $A(\xi^+, \xi^-)$  (where  $\xi^\pm = \xi^1 \pm i\xi^2$ ) which transforms under analytic transformations

$$\xi^+ \rightarrow f^+(\xi^+), \quad \xi^- \rightarrow f^-(\xi^-)$$

as follows

$$A(\xi^+, \xi^-) = \left( \frac{df^+}{d\xi^+} \right)^{\bar{\Delta}} \left( \frac{df^-}{d\xi^-} \right)^{\Delta} \tilde{A}(f^+, f^-) \tag{9.361}$$

thus having conformal weight  $(\Delta, \bar{\Delta})$ . At the same time the metric  $\rho(\xi^+, \xi^-)$  is assumed to be transformed as:

$$\rho(\xi^+, \xi^-) = \frac{df^+}{d\xi^+} \frac{df^-}{d\xi^-} \tilde{\rho}(f^+, f^-)$$

Let us try to define covariant derivatives of  $\mathcal{A}$ . This can be easily achieved:

$$\begin{aligned} \nabla_+ A &= (\partial_+ - \Delta(\partial_+ \varphi))A \\ &= \rho^\Delta \partial_+ (\rho^{-\Delta} A) \end{aligned} \tag{9.362}$$

It is trivial to check that if the field  $A$  has the transformation law (9.361) with some  $(\Delta, \bar{\Delta})$ , then  $\nabla_+ A$  has the transformation law with  $(\Delta + 1, \bar{\Delta})$ . Now, the results of the previous computations are becoming transparent. Indeed the field  $\omega_a$  is a vector field  $(\omega_+, \omega_-)$  with the weights  $(1, 0)$  and  $(0, 1)$  correspondingly. Hence:

$$\nabla_+ \omega_+ = \rho \partial_+ (\rho^{-1} \omega_+) \tag{9.363}$$

which coincides with (9.354). The field  $h_{++}$  has weight  $(2, 0)$ , hence:

$$\nabla^a h_{a+} = \rho^{-1} \partial_- h_{++} \quad (\text{comp. (9.355)})$$

The spinor field  $\varepsilon_\pm$  has the weights  $(1/2, 0)$  and  $(0, 1/2)$ , while the gravitino field  $f_+^L (f^R)$  has  $(3/2, 0)$  and  $(0, 3/2)$ . That gives (9.359) and (9.360). Let us notice also, that the tracelessness of  $h_{ab}$  means that

$h_{+-} = 0$  and we have two components  $h_{++}$  and  $h_{--}$  only, while the condition  $\sigma^a f_a = 0$  implies also that we have only two components  $f_+^L$  and  $f_-^R$  of the vector-spinor  $f_a^\alpha$ , which correspond to spin 3/2, while the spin 1/2 component is removed. We must remember that all our computations of covariant derivatives have been done on the background where we have some metric  $\rho$  but no gravitino field  $\chi$ . In the final results we shall be able to restore  $\chi$ -dependence by the use of supersymmetry. It should be noted, however, that it is easy to work with  $\chi$  present. All we have to do is to consider a superspace generalization of (9.361) and (9.362).

To put things together, we see that the ghost contribution to the functional integral is defined by the determinants, which arise from the following eigenvalue problems (recall (9.352)):

$$\begin{aligned} \rho \partial_+(\rho^{-1}\omega_+) &= iEh_{++} \\ \partial^+ h_{++} \equiv g^{+-} \partial_- h_{++} &= \rho^{-1} \partial_- h_{++} = iE\omega_+ \end{aligned} \quad (9.364)$$

for bosons, and

$$\begin{aligned} \rho^{1/2} \partial_+(\rho^{-1/2}\epsilon_+) &= iEf_+ \\ \partial^+ f_+ &= \rho^{-1} \partial_- f_+ = iE\epsilon_+ \end{aligned} \quad (9.365)$$

for fermions. If we apply  $\partial_-$  to the first equations of each pair we get:

$$\begin{aligned} -\rho^{-2} \partial_-(\rho \partial_+ \bar{u}) &= E^2 \bar{u} \\ -\rho^{-3/2} \partial_-(\rho^{1/2} \partial_+ \bar{v}) &= E^2 \bar{v} \\ \bar{u}_{(0,-1)} \equiv \rho^{-1}\omega_+, \quad \bar{v}_{(0,-1/2)} \equiv \rho^{-1/2}\epsilon_+ \end{aligned} \quad (9.366)$$

The general operator acting on tensors with conformal spin  $(0, j)$  would be

$$\mathcal{L}_j = \rho^{j-1} \partial_-(\rho^{-j} \partial_+) \quad (9.367)$$

For  $j = 0$  this is the scalar laplacian: for  $j = -1/2$  and  $-1$  these are ghost operators, which we have just now derived. We shall need one more value,  $j = 1/2$ , corresponding to Dirac fermions. Indeed, Dirac fermions transform as:

$$\psi_\pm \sim (d\xi^\pm)^{-1/2} \sim \begin{pmatrix} 1/2, 0 \\ 0, 1/2 \end{pmatrix} \quad (9.368)$$

and hence the Dirac eigenvalue problem is:

$$\begin{aligned} \rho^{-1/2} \partial_- \psi_+ &= iE\psi_- \\ \rho^{-1/2} \partial_+ \psi_- &= iE\psi_+ \end{aligned} \quad (9.369)$$

Of course the same equation could have been derived in a long way, starting from the standard Dirac equation with spinorial connection:

$$\sigma^a \nabla_a \psi = iE\psi \tag{9.370}$$

We have already described this way above in the case of the ghosts and will not repeat it.

Our task now is to compute the determinants. All of them must be local functions of  $\rho$ , since under naive manipulations the  $\rho$  dependence drops out from all expressions. Indeed the action for the field  $u$  in (9.366) and (9.367) which leads to the correct eigenvalue problem has the form:

$$S_j = \int \rho^{-j} \partial_+ \bar{u} \partial_- u \, d^2\xi$$

$$N_j = \int \rho^{1-j} \bar{u} u \, d^2\xi \tag{9.371}$$

(where  $N_j$  is the norm in the  $u$ -space). If we try to use perturbation theory, setting  $\rho = 1 + \varphi$  we get:

$$W = \text{---} \overset{k+q}{\circlearrowleft} \text{---} \underset{k}{\circlearrowright} \text{---} \underset{q}{\circlearrowright}$$

$$\sim j^2 \varphi(q) \varphi(-q) \int \frac{d^2k \, k_+ (k+q)_- (k_-(k+q)_+)}{k^2 (k+q)^2}$$

$$= j^2 \varphi(q) \varphi(-q) \int d^2k \tag{9.372}$$

Which is ill-defined. This means that we have failed to compute the effective action and demonstrated only that it does not have an imaginary part in  $q$ -space, being a local function of  $\varphi$ . The same is true to all orders in  $\varphi$ .

There are several ways out of this problem. One is to use Pauli-Villars regularization of all loops. This amounts to adding to the action  $S_j$  the regulating action  $\tilde{S}_j$ :

$$\tilde{S}_j = \int (\rho^{-j} \partial_+ \bar{u} \partial_- \bar{u} + M^2 \rho^{1-j} \bar{u} \bar{u}) \, d^2\xi \tag{9.373}$$

where loops of  $\bar{u}$ -fields enter with negative signs, and cancel the divergences of  $S_j$ . The construction of the action  $\tilde{S}_j$  ensures that the regularization preserves conformal symmetry. The  $\varphi$ -dependence can easily be extracted from the loop containing the  $\bar{u}$ -field, and then the general argument, based on conformal invariance, would lead us to the

Liouville action with fixed coupling (this is the only conformally invariant expression of the required dimensionality).

We shall choose here a different approach to the problem, which seems more meaningful. Let us generalize the expression for the ghost lagrangian (9.352) which, without external field, has the form:

$$S_{gh}^{(B)} = \int (h_{++} \partial_- \omega^+ + h_{--} \partial_+ \omega^-) d^2 \xi; \quad \omega^\pm = \rho^{-1} \omega_\mp \quad (9.374)$$

In this bosonic part, the fields  $\omega$  and  $h$  have the following transformation law:

$$\begin{aligned} \omega^+ &\sim (-1, 0) & \omega^- &\sim (0, -1) \\ h_{++} &\sim (2, 0) & h_{--} &\sim (0, 2) \end{aligned}$$

In general we could consider the case:

$$S_j = \int (b^+ \partial_- a_+ + b^- \partial_+ a_-) d^2 \xi \quad (9.375)$$

with

$$\begin{aligned} a_+ &\sim (j, 0) & a_- &\sim (0, j) \\ b^+ &\sim (1-j, 0) & b^- &\sim (0, 1-j) \end{aligned} \quad (9.376)$$

Our tactics now will be to couple to (9.375) an external gravitational field not in the conformal gauge, in which the effective action is nonpropagating, but in the gauge in which we have only one component of the graviton, say  $g_{++}$ . As we have already seen, this leads to a finite expression containing  $g_{++}$  which can be generalized to a covariant expression.

First of all, we have to find an expression for the covariant derivative in this gauge. The metric has the form:

$$\begin{aligned} ds^2 &= d\xi^+ d\xi^- + g_{++} (d\xi^+)^2 \\ &= d\xi^+ (d\xi^- + g_{++} d\xi^+) \end{aligned} \quad (9.377)$$

It is easy now to find the coupling of the  $g_{++}$ -field to the  $a$  and  $b$ -fields, or, in other words, the energy-momentum tensor of the  $(a, b)$ -system. The idea is to find the change of the action under the coordinate

transformations:

$$\begin{aligned}\xi^+ &\rightarrow \xi^+ \\ \xi^- &\rightarrow f^-(\xi^+, \xi^-)\end{aligned}\tag{9.378}$$

The fields are transformed as:

$$\begin{aligned}a_-(\xi^+, \xi^-) &\rightarrow \left(\frac{\partial f^-}{\partial \xi^-}\right)^j a_-(\xi^+, f^-(\xi^+, \xi^-)) \\ b_-(\xi^+, \xi^-) &\rightarrow \left(\frac{\partial f^-}{\partial \xi^-}\right)^{1-j} b_-(\xi^+, f^-(\xi^+, \xi^-))\end{aligned}\tag{9.379}$$

The new action (or better to say the “minus” part of it) changes to

$$\begin{aligned}\tilde{S} &= \int d\xi^+ df^- \left(\frac{\partial f^-}{\partial \xi^-}\right)^{-1} \cdot \left(\frac{\partial f^-}{\partial \xi^-}\right)^{1-j} b(\xi^+, f^-) \\ &\quad \cdot \frac{\partial}{\partial \xi^+} \left(\frac{\partial f^-}{\partial \xi^-}\right)^j \cdot a(\xi^+, f^-(\xi^+, \xi^-)) \\ &= \int d\xi^+ df^- b(\xi^+ f^-) \frac{\partial}{\partial \xi^+} a(\xi^+, f^-) \\ &\quad + \int d\xi^+ df^- (ba)^j \cdot \frac{\partial}{\partial \xi^+} \log \left(\frac{\partial f^-}{\partial \xi^-}\right) \\ &\quad + \int d\xi^+ df^- \left(b \frac{\partial}{\partial f^-} a\right) \partial_+ f^-\end{aligned}\tag{9.380}$$

Now let us take

$$f^- = \xi^- + \varepsilon_+(\xi^+ \xi^-)\tag{9.381}$$

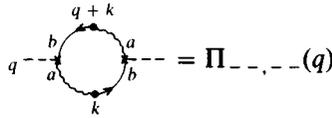
with infinitesimal  $\varepsilon_+$ . We get:

$$\tilde{S} - S = \int ((b \partial_- a) - j \partial_- (ba)) \partial_+ \varepsilon_+$$

From here we can read off the energy-momentum tensor which couples to  $g_{++}$ :

$$T_{--} = (b \partial_- a) - j \partial_- (ba) = (1 - j) b \partial_- a + j a \partial_- b\tag{9.382}$$

Now we can compute the second order response to the field  $g_{++}$ . It is described by the diagram:



$$S_{\text{eff}} = -\frac{1}{2} \int \Pi_{--}(q) g_{++}(q) g_{++}(-q) \frac{d^2q}{(2\pi)^2}$$

$$\begin{aligned} \Pi_{--}(q) &= -\frac{1}{16} \int \frac{d^2k [j(k_+ + q)_- + (1-j)k_-] [jk_- + (1-j)(k + q)_-]}{(k_+ + i\epsilon \text{sign } k_-)(k_+ + q_+ + i\epsilon \text{sign } (k + q)_-)} \\ &= -\frac{1 + 6j(j-1)}{24\pi} \frac{q_-^3}{16q_+} \end{aligned} \tag{9.383}$$

(we use here the correlations:

$$\langle a_{\pm}^{(j)}(q) b_{\pm}^{(1-j)}(-q) \rangle = (q_+ + i\epsilon \text{sign } q_-)^{-1}$$

The arguments, identical to those in section (9.6) now give the following result for the determinants:

$$\begin{aligned} &\log \det \| -\rho^{j-1} \hat{c}_-(\rho^{-j} \hat{c}_+) \| \\ &= -\frac{1 + 6j(j-1)}{24\pi} \int \left( \frac{1}{2} (\hat{c}_a \varphi)^2 + \mu^2 e^\varphi \right) d^2\xi \\ &\rho = e^\varphi \end{aligned} \tag{9.384}$$

This result is equivalent to the computation of the central charge for the Virasoro algebra of the energy-momentum tensor (9.381) which is just equal to:

$$c_j = \pm (1 + 6j(j-1)) \cdot 2 \tag{9.385}$$

where  $\pm$  refers to commuting or anticommuting fields (recall, that fermion loops must be supplied with a minus sign). Indeed, the central charge is just the expectation of two energy-momentum tensors  $\langle T_{--} T_{--} \rangle$  which we have computed.

One more step is needed before we come to the physical answer. We have computed all determinants under the condition that the gravitino field  $\chi = 0$ . We must now restore dependence on it in  $\det \mathcal{L}_j$ . This can be done without any further computations, since there is only one supersymmetric extension of the Liouville action (9.384). In order to find it we can consider the superfield:

$$\phi = \varphi + \theta_+ \chi_+ + \theta_- \chi_- + \theta_+ \theta_- f \tag{9.386}$$

and the action:

$$S = \int (\frac{1}{2} \mathcal{D}_+ \phi \mathcal{D}_- \phi + 2i\mu e^{\phi/2}) d\theta_- d\theta_+ d^2\xi$$

$$\mathcal{D}_\pm = \frac{\partial}{\partial \theta^\pm} + \theta^\pm \frac{\partial}{\partial \xi^\pm} \tag{9.387}$$

which is explicitly supersymmetric. Simple computation gives:

$$e^{\phi/2} = e^{\phi/2} e^{(\theta_+ \chi_+ + \theta_- \chi_- + \theta_+ \theta_- f)/2}$$

$$= e^{\phi/2} (1 + \frac{1}{2} \theta_+ \chi_+ + \frac{1}{2} \theta_- \chi_- + \frac{1}{2} \theta_+ \theta_- (f - \frac{1}{2} \chi_+ \chi_-))$$

$$\int d^2\theta d^2\xi e^{\phi/2} = \int d^2\xi \frac{1}{2} e^{\phi/2} (f - \frac{1}{2} \chi_+ \chi_-) \tag{9.388}$$

$$\int \frac{1}{2} \mathcal{D}_+ \phi \mathcal{D}_- \phi d^2\theta d^2\xi = \int (\frac{1}{2} (\partial\phi)^2 - \frac{1}{2} \chi_+ \partial_- \chi_+ - \frac{1}{2} \chi_- \partial_+ \chi_- + \frac{1}{2} f^2) d^2\xi$$

Taking the two terms together and eliminating the  $f$ -field by taking the minimum of the action with respect to it we obtain:

$$S = \int \{ \frac{1}{2} (\partial\phi)^2 + \frac{1}{2} \chi \not{\partial} \chi + \frac{1}{2} \mu e^{\phi/2} \bar{\chi} \omega + \mu^2 e^\phi \} d^2\xi \tag{9.389}$$

This action should replace the Liouville one for the case when the  $\chi$ -field is nonzero. It could have been obtained explicitly by coupling the gravitino to supercurrents and considering superconformal anomalies. As we see, this is not needed, since the superconformal extension of the bosonic part of the action is unique.

Now we are ready to write down the partition function for the superstring case. Let us find the value for the super-Liouville coupling constant, or, which is the same, the central charge  $c$ . We have:

$$(j = 0) \quad c^{(x)} = \mathcal{D} \quad (j = -1) \quad c_{gh}^{(x)} = -26$$

$$(j = \frac{1}{2}) \quad c^{(\psi)} = \mathcal{D}/2 \quad (j = -\frac{1}{2}) \quad c_{gh}^{(\psi)} = 11$$

$$c_{tot} = \frac{3}{2} \mathcal{D} - 15 = \frac{3}{2} (\mathcal{D} - 10)$$

(here the values of  $c$  for  $\psi$  and  $x$  are half those in (9.385) because (9.385) is written for a complex field). So, our final result for the effective action of a fermionic string, which we obtain by integrating out the  $x$  and  $\psi$  fields, is given by:

$$S = \frac{10 - \mathcal{D}}{32\pi} \int (\frac{1}{2} (\partial\phi)^2 + \frac{1}{2} \bar{\chi} \not{\partial} \chi + \frac{1}{2} \mu e^{\phi/2} \bar{\chi} \chi + \mu^2 e^\phi) d^2\xi \tag{9.391}$$

As in the bosonic case, this action describes longitudinal motions of the string. In the critical dimension,  $\mathcal{D}_{\text{cr}} = 10$ , they are absent.

So far, we have been dealing with surfaces of spherical topology without external lines. As in the bosonic case, in order to obtain scattering amplitudes, one has to puncture the sphere, and in order to find corrections to the tree amplitude one has to sum over topologies. We shall start with the first task, which in the fermionic case has some unusual features.

### 9.13 Vertex Operators

In the case of the bosonic string we have punctured its world surface by introducing factors:

$$\begin{aligned} V(x) &= \int \delta(x - x(\xi)) g^{1/2} d^2\xi \\ V(p) &= \int d^2\xi V_p(\xi) \\ V_p(\xi) &= e^{ip \cdot x(\xi)} (g(\xi))^{1/2} \end{aligned} \quad (9.392)$$

into the functional integral. Our task now is to find a supersymmetric generalization of these formulas.

Let us begin with the properties of the  $x$  fields. The Green function for the supersymmetric laplacian  $\bar{\mathcal{D}}\mathcal{D}$  is given by:

$$\begin{aligned} &\langle \hat{x}_\mu(\xi_1, \theta_1) \hat{x}_\nu(\xi_2, \theta_2) \rangle \\ &= -\frac{\delta_{\mu\nu}}{4\pi} \log(z_1 - z_2 - \theta_1 \theta_2)(\bar{z}_1 - \bar{z}_2 - \bar{\theta}_1 \bar{\theta}_2) \end{aligned} \quad (9.393)$$

An easy check of this formula is obtained by using the component representation: this formula is equivalent to:

$$\begin{aligned} \langle x_\mu(\xi_1) x_\nu(\xi_2) \rangle &= -\frac{\delta_{\mu\nu}}{4\pi} \log|z_1 - z_2|^2 \\ \langle \psi_{\mu L}(\xi_1) \psi_{\nu L}(\xi_2) \rangle &= -\frac{\delta_{\mu\nu}}{4\pi} \frac{1}{z_1 - z_2} \\ \langle \psi_{\mu R}(\xi_1) \psi_{\nu R}(\xi_2) \rangle &= -\frac{\delta_{\mu\nu}}{4\pi} \frac{1}{\bar{z}_1 - \bar{z}_2} \\ \langle f_\mu(\xi_1) f_\nu(\xi_2) \rangle &= \delta(\xi_1 - \xi_2) \delta_{\mu\nu} \end{aligned} \quad (9.394)$$

The reason why the particular combination  $\hat{z}_{12} = z_1 - z_2 - \theta_1\theta_2$  has appeared in (9.393) is also simple: it is invariant under supersymmetry:

$$\begin{aligned} \delta\hat{z}_{12} &= \delta z_1 - \delta z_2 - \delta\theta_1\theta_2 + \delta\theta_2\theta_1 = 0 \\ (\delta z_1 &= -\varepsilon\theta_1, \delta z_2 = -\varepsilon\theta_2, \delta\theta_1 = \delta\theta_2 = \varepsilon) \end{aligned}$$

Let us now find the expression for the supersymmetric vertex operator, beginning with the simplest case of the critical dimension  $D = 10$ . In this case, according to the preceding section we expect superconformal invariance of the theory, i.e. the fields  $\varphi$  and  $\chi$  should not participate on the mass shell. The natural guess for  $V(p)$  will be:

$$\begin{aligned} V(p) &= \int d^2\xi d^2\theta :e^{i p \cdot \hat{z}(\xi, \theta)}: \\ &= \int d^2\xi \int d^2\theta :e^{i p \cdot x} \{1 - i p \cdot \psi_L \theta \\ &\quad - i p \cdot \psi_R \bar{\theta} - \bar{\theta} \theta [(p \cdot \psi_L)(p \cdot \psi_R) - i p \cdot f]\}: \end{aligned} \tag{9.395}$$

Setting  $f = 0$  and integrating on  $\theta$  we obtain:

$$V(p) = \int d^2\xi : (p_\mu \psi_{\mu, L})(p_\nu \psi_{\nu, R}) e^{i p \cdot x(\xi)}: \tag{9.396}$$

As we have noted in the bosonic case, the dimension of  $:e^{i p \cdot x}$ : is equal to  $p^2$ . Since the free fermions,  $\psi_{\mu, L}$  and  $\psi_{\mu, R}$ , have dimension  $1/2$ , we obtain the following mass shell condition:

$$\Delta = p^2 + 1 = 2, \quad p^2 = 1 \tag{9.397}$$

So, the vertex (9.396) still describes a tachyon, but a “better” one than in the bosonic case (which has  $p^2 = 2$ ). If we consider an open string, which amounts to taking only the “left” part of all fields and coordinates, we get:

$$\begin{aligned} V_{\text{open}}(p) &= \int dz : p_\mu \psi_\mu(z) e^{i p \cdot x(z)}: \\ p^2 &= \frac{1}{2} \end{aligned} \tag{9.398}$$

We can be confident that the “old” bosonic tachyon will not appear in collisions of the new ones, because  $:e^{i p \cdot x}$ : is not supersymmetric. Still it is instructive to see this fact explicitly. Let us consider the operator product:

$$\begin{aligned} &: p_\mu \psi_\mu(z) e^{i p \cdot x(z)} : : k_\mu \psi_\mu(0) e^{i k \cdot x(0)} : \\ &= (p \cdot k) z^{2p \cdot k - 1} : e^{i(p+k) \cdot x(0)} : + \dots \end{aligned} \tag{9.399}$$

It seems at first that the bosonic tachyon has appeared, since we have obtained its vertex operator  $:e^{i\mathbf{p}\cdot\mathbf{x}}:$  on the r.h.s. of (9.399). However, its appearance is illusory. Indeed, the fusion of two tachyons of two fermionic strings (9.398) with  $p^2 = 1/2$  would produce a bosonic tachyon with  $p^2 = 1$ , according to (9.399), if  $p^2 = k^2 = 1/2$  and  $(\mathbf{p} + \mathbf{k})^2 = 1$ . That implies that  $\mathbf{p}\cdot\mathbf{k} = 0$  and hence the residue of the corresponding pole in all amplitudes is zero. Hence, the true ground state of the open fermionic string has  $p^2 = 1/2$ .

It is possible in a consistent manner to eliminate even this tachyon. Let us distinguish vertex operators by their parity under  $\psi_\mu \rightarrow -\psi_\mu$  (we are talking about open strings for simplicity). The tachyon state (9.398) is odd. The fusion of two tachyons will give even states. In order to find them we have to investigate the next terms in (9.399): the simplest one will be the supersymmetric extension of the operator  $:i\partial_z x_\mu e^{i\mathbf{p}\cdot\mathbf{x}}:$  of the bosonic string theory, which describes a massless vector state of the open string. This extension is easy to find by using superfields. Let us consider the expression:

$$\begin{aligned}\Gamma_\mu(\mathbf{p}) &= \int dz d\theta :i\mathcal{D}\hat{x}_\mu e^{i\mathbf{p}\cdot\mathbf{x}}: \\ &= \int dz d\theta :i(\psi_\mu + \theta \partial_z x_\mu)e^{i\mathbf{p}\cdot\mathbf{x}}(1 - i\mathbf{p}\cdot\psi\theta): \\ &= \int dz :(i\partial_z x_\mu + p_\nu\psi_\mu\psi_\nu)e^{i\mathbf{p}\cdot\mathbf{x}}: \quad (9.400)\end{aligned}$$

This vertex operator describes a vector massless state again, since the dimensionality of the integrand is given by:

$$\Delta = p^2 + 1 = 1; p^2 = 0 \quad (9.401)$$

By its construction, our vertex is supersymmetric. It is even under reflection  $\psi_\mu \rightarrow -\psi_\mu$ . Under fusion of two such particles the tachyon (9.398) cannot appear because it is odd, and the ‘‘bosonic’’ tachyon  $:e^{i\mathbf{p}\cdot\mathbf{x}}:$  cannot appear since it is not supersymmetric. To be more precise, in the operator product of two operators (9.400), the operator  $:e^{i\mathbf{p}\cdot\mathbf{x}}:$  does appear, but on mass shell the coefficient in front of it is equal to zero, just as happens in (9.399).

We come to the conclusion that the fermionic string when restricted to the sector even under  $\psi$ -reflection has no tachyons for  $\mathcal{D} = 10$ , and its ground state is the zero mass vector particle.

In the closed string case the vertex operator is given by a direct product of two  $\Gamma_\mu$  from (9.400), one depending on  $z$  and the other on  $\bar{z}$ . It describes a massless tensor particle, which we discuss a little later.

Massless vectors or tensors can interact without producing ghosts (and we know that the ghosts are absent for  $\mathcal{D} \leq 10$  either from (9.390) or from direct counting of states which we have done in the bosonic case), only if their effective action is gauge invariant, being of Yang–Mills type for vectors and Einstein’s for tensors. We shall postpone explicit computation of this effective action, and point out here only one manifestation of the gauge structure. In the operator product:

$$\Gamma_\mu(\mathbf{p}, z)\Gamma_\nu(\mathbf{k}, 0) \sim z^{2p \cdot k - 1} t_{\mu\nu\lambda}(\mathbf{p}, \mathbf{k})\Gamma_\lambda(\mathbf{p} + \mathbf{k}, 0) \tag{9.402}$$

an explicit computation shows that the quantity  $t_{\mu\nu\lambda}(\mathbf{p}, \mathbf{k})$  coincides with the triple Yang–Mills vertex. This is an important observation, since we already know that structure constants in operator products are just the residues of the poles in scattering amplitudes. A little later we shall describe an efficient method for their computation.

Up to now, we have spoken of the fermionic string but, clearly, all particles obtained from the vertex operators (9.400) are bosons.

It is most important that apart from them, the spectrum of the theory contains fermions appearing as soliton excitations of the string. In terms of functional integrals, solitons appear owing to special boundary conditions applied on the fields. In our case, anomalous boundary conditions have the following meaning. We are dealing with the world sheet which is a sphere with omitted points, where the external particles are injected. On such a punctured sphere we are free to chose different spinor structures for the fields  $\psi_\mu$ ; that is we can consider double-valued fields, which change sign when going round the injection point.

In order to understand why nontrivial spinor structures are related to space-time fermions, we have to take a step back and return to the case of particles. Let us recall, that we described spinors by the anti-commuting fields  $\psi_\mu$  with the action:

$$S = \int \psi_\mu \dot{\psi}_\mu d\tau \tag{9.403}$$

Why does this action describe a space-time spinor (while the variables  $\psi_\mu$  are vectors)? To answer this question let us pass to the Hamiltonian formalism. Assuming that the dimensionality of space-time is even we can choose half of  $\psi_\mu$  to be coordinates, while the other half are conjugate momenta. So, let us introduce complex fields:

$$\phi_a = \psi_a + i\psi_{a+\mathcal{D}/2} \quad a = 1, \dots, \mathcal{D}/2$$

Then the action takes the form:

$$S = \int \phi_a^+ \overset{\leftrightarrow}{\frac{\partial}{\partial t}} \phi_a dt \tag{9.404}$$

This is the standard form describing  $\mathcal{D}/2$  harmonic oscillators. The Hamiltonian in this case is equal to zero, while the states are formed by the “vacuum”

$$\phi_a|0\rangle = 0 \quad (9.405)$$

together with the “excited” states:

$$|f\rangle = \phi_{a_1}^+ \cdots \phi_{a_n}^+ |0\rangle \quad (9.406)$$

Since our operators are anticommuting,

$$[\phi_a, \phi_b^+]_+ = \delta_{ab}, \quad \phi_a^2 = (\phi_a^+)^2 = 0 \quad (9.407)$$

the number of independent states in (9.406) is  $2^{\mathcal{D}/2}$  (each  $a$  can be either empty or occupied). All these states have zero energy and describe a single particle with  $2^{\mathcal{D}/2}$  states. This is just the degeneracy of the spinor representation of the  $O(\mathcal{D})$  group, and in fact, what we have done is just to construct such a representation. Of course, to complete the job, one has to construct the  $O(\mathcal{D})$  generators out of  $\phi_a$ , and show that the representation is nondegenerate. This is not hard, but we stop here, since our task was just to see the connection between  $\psi_\mu$  and spinors.

Returning to the string case, let us look at the  $\psi_\mu$  part of the string action on a cylinder:

$$S = \int \left( \psi_{\mu,L} \left( \frac{\partial}{\partial \tau} - i \frac{\partial}{\partial \sigma} \right) \psi_{\mu,L} + \psi_{\mu,R} \left( \frac{\partial}{\partial \tau} + i \frac{\partial}{\partial \sigma} \right) \psi_{\mu,R} \right) d\tau d\sigma \quad (9.408)$$

Two spinor structures correspond to  $\psi_\mu$  being periodic or anti-periodic as we go around the cylinder. In the periodic case (which is called the Ramond sector)  $\psi_\mu$  has a  $\sigma$ -independent component, so that the action (9.408) is reduced (for each  $\psi_{\mu,L}$  and  $\psi_{\mu,R}$ ) to (9.403). We see that as we integrate over the periodic case, the ground state of the string is a space-time spinor. In the antiperiodic case we still have the states (9.406), but now the Hamiltonian is nonzero (due to  $i\psi \cdot \partial_\sigma \psi$ ) and these states are not degenerate. Actually, in this case the ground state is bosonic (in the space-time sense).

Now we have to return to our punctured sphere and the operator algebra. This is achieved by a conformal transformation which makes an annulus out of the cylinder. Then, the limit of an infinitely long cylinder will correspond to the limit, when the internal circle of the annulus shrinks to a point while the external one goes to infinity. That is, we shall get a sphere with an omitted point. Our task is to find what happens to different spinor structures under this transformation.

In order to solve this problem we notice that the conformal map we are talking about is just

$$\begin{aligned}
 w &= \log z = \tau + i\sigma \\
 \psi_{\mu L}(z) &= \left(\frac{dw}{dz}\right)^{1/2} \psi_{\mu L}(w(z)) \\
 \psi_{\mu R}(\bar{z}) &= \left(\frac{d\bar{w}}{d\bar{z}}\right)^{1/2} \psi_{\mu R}(\bar{w}(\bar{z}))
 \end{aligned}
 \tag{9.409}$$

(here  $z$  belongs to the punctured sphere and  $w$  to the cylinder; we have recalled that the conformal spin of  $\psi_\mu$  is  $1/2$ ).

The square root in (9.409) is rather important. It implies that  $\psi_\mu$  periodic on the cylinder (Ramond sector) becomes antiperiodic as we go to the  $z$ -plane. Vice versa, fields  $\psi_\mu$  which were antiperiodic on the cylinder (Neveu-Schwarz sector) transform into fields, which are univalued in the  $z$ -plane.

So, we arrive at the following rule. When we integrate over fields which are univalued in the whole  $z$ -plane, we describe bosonic states of the string which comprise what is called the Neveu-Schwarz sector. If we permit  $\psi_\mu$ -fields, which change sign after going round the injection point, then we have injected a space-time fermion.

How do we perform efficiently the functional integrals for non-trivial spinor structures? In the case of point singularities which we are discussing, the answer to this question is through the introduction of the so-called spin operators.

Since we know that the ground state in the Ramond sector is a space-time spinor, we expect that it corresponds to an operator  $S_a(z, \bar{z})$ , where  $a = 1, \dots, \mathcal{D}/2$  is the spinor index. An operator product with  $\psi_\mu(z)$  must be given by:

$$\psi_\mu(z)S_a(0) = (2z)^{-1/2}(\gamma_\mu)_{ab}S_b(0) + \dots
 \tag{9.410}$$

The factor  $z^{-1/2}$  reflects the non-trivial spinor structure at  $z = 0$  induced by  $S_a(0)$  and the coefficients  $(\gamma_\mu)_{ab}$  are just  $\mathcal{D}$ -dimensional  $\gamma$ -matrices. The fermionic state  $|a\rangle$  of the string ( $a$  being a spinor index) is given by:

$$|a\rangle = S_a(0)|0\rangle
 \tag{9.411}$$

These spin operators have a very transparent physical meaning. Let us recall that free Majorana fermions in two dimensions are equivalent to the two dimensional Ising model, in the sense that their partition functions are equal. A NRS-string can be imagined as a bosonic one,

each piece of which carries a  $\mathcal{D}$ -dimensional spin operator. Now, by a Jordan-Wigner transformation this system is transformed into the usual Majorana fermions on the world surface. On the other hand, since each piece of the string carries spin  $1/2$ , we can have two sectors, depending on whether we have an even or an odd number of spins. These are the Neveu-Schwarz and Ramond sectors respectively.

The spin operator, which interchanges these two sectors, can be expressed by Jordan-Wigner transformations through the fermionic field  $\psi_\mu$ . This representation, involving exponents of bi-linears of  $\psi_\mu$  is rather clumsy, but fortunately is not needed since we shall show how to compute  $S_a$ -correlations directly.

Consider first one fermionic field  $\psi$  and let us compute the correlation function:

$$\mathcal{K} = \langle \psi(z_1)\psi(z_2)S(x)S(y) \rangle \quad (9.412)$$

(This is precisely the case of the Ising model.)

It appears that the analytic properties of  $\mathcal{K}$  are sufficient for its determination.

We know from (9.410) that it is an analytic function of  $z_1$  and  $z_2$ , which has square root branch points at  $z_{1,2} = x, y$ . Also because of the operator product relation

$$\psi(z_1)\psi(z_2) \simeq z_{12}^{-1}I + \dots \quad (9.413)$$

it must have a simple pole as  $z_1 \rightarrow z_2$ . Hence, we can write:

$$\mathcal{K} = \frac{P(z_1, z_2, x, y)}{(z_1 - z_2)\{(z_1 - x)(z_1 - y)(z_2 - x)(z_2 - y)\}^{1/2}} \quad (9.414)$$

where  $P$  must be a polynomial in  $z_1$  and  $z_2$ . This polynomial must be arranged so that the residue at the pole  $z_1 = z_2$  is independent of  $z_1$  (because of the unit operator in (9.413)). Hence:

$$P(z_1, z_1, x, y) = Q(x, y)(z_1 - x)(z_1 - y) \quad (9.415)$$

Also, because of

$$S(x)S(y) \simeq |x - y|^{-4\Delta}I + \dots \quad (9.416)$$

we have the requirement:

$$P(z_1, z_2, x, y) = (z_1 - x)(z_2 - x) \quad x \rightarrow y$$

The only possible form for the correlation function is therefore:

$$\mathcal{K} = \frac{1}{2} \frac{(z_1 - x)(z_2 - y) + (z_1 - y)(z_2 - x)}{(z_1 - z_2)\{(z_1 - x)(z_1 - y)(z_2 - x)(z_2 - y)\}^{1/2}} \frac{1}{|x - y|^{4\Delta}} \quad (9.417)$$

Now we are ready to find  $\Delta$ .

The idea is the following. We first use the operator product

$$\psi(z_1)\psi(z_2) \underset{z_{12} \rightarrow 0}{\simeq} z_{12}^{-1} I + f z_{12} T(z_2) + \dots \quad (9.418)$$

where the constant  $f$  can be expressed through the dimensionality of  $\psi$ , i.e.  $1/2$ . To find  $f$  we can substitute (9.418) into:

$$\begin{aligned} \langle \psi(z)\psi(0)\psi(u)\psi(v) \rangle &= \frac{1}{z} \frac{1}{u-v} + \frac{1}{z-u} \frac{1}{v} - \frac{1}{z-v} \frac{1}{u} \\ &\underset{z \rightarrow 0}{\approx} \frac{1}{z} \frac{1}{u-v} + z \frac{u-v}{(uv)^2} + O(z^2) \end{aligned} \quad (9.419)$$

The first term is the contribution of the unit operator while the second one comes from the energy-momentum tensor. The conformal Ward identity fixes the normalization of  $T$ :

$$\langle T(0)\psi(u)\psi(v) \rangle = \frac{1}{2} \frac{u-v}{u^2 v^2} \frac{1}{2} \quad (9.420)$$

We see that in (9.418)  $f = 2$ . Of course, this could have been foreseen, because we are dealing with free fermions for which

$$T(z) = -\frac{1}{2} \psi \partial_z \psi \quad (9.421)$$

and there is nothing more in our derivation than checking this fact. But, in general, the method of operator products is applicable far beyond free field situations.

The second step in finding  $\Delta$  is to substitute (9.418) into  $\langle \psi\psi SS \rangle$ . Expanding (9.417) in  $z_{12}$  we obtain:

$$\begin{aligned} \mathcal{K} &\underset{z_{12} \rightarrow 0}{=} \frac{1}{z_{12}} \frac{1}{|x-y|^{4\Delta}} \\ &+ \frac{z_{12}}{8} \frac{(x-y)^2}{(z_2-x)^2(z_2-y)^2} \frac{1}{|x-y|^{4\Delta}} \end{aligned} \quad (9.422)$$

The first term here again comes from the unit operator, while the second is the energy-momentum contribution. Combining (9.418) and (9.402) we get:

$$\langle T(z)S(x)S(y) \rangle = \frac{1}{16} \frac{(x-y)^2}{(z-x)^2(z-y)^2} \frac{1}{|x-y|^{4\Delta}} \quad (9.423)$$

This is compatible with the conformal Ward identity (which, in particular, says that the residue of the second order pole at  $z \rightarrow x$  is equal to  $\Delta$ ) provided that we chose  $\Delta = 1/16$ . We see that conformal algebra indeed determines the dimensionality of the spin operator, as well as its correlation functions.

There is an important representation for (9.412). Take its  $y \rightarrow \infty$ ,  $x \rightarrow 0$  limit, which can be achieved by a projective transformation. We obtain:

$$\langle S(\infty)|\psi(z_1)\psi(z_2)|S(0) \rangle = \frac{1}{2} \frac{z_1 + z_2}{(z_1 z_2)^{1/2}} \frac{1}{z_1 - z_2} \quad (9.424)$$

This formula has an interesting interpretation. Namely, since we have introduced spins at 0 and  $\infty$  the fields  $\psi(z)$  have become double valued in the  $z$ -plane with a cut from 0 to  $\infty$ . We can expect that they have the following mode expansion:

$$\psi(z) = \frac{1}{z^{1/2}} \left( \psi_0 + \sum_{n=1}^{\infty} (b_n z^{-n} + b_n^+ z^n) \right) \quad (9.425)$$

and proper anticommutations of  $\psi$  requires

$$\begin{aligned} \{\psi_0 \psi_0\} &= 2\psi_0^2 = 1 \\ \{b_n, b_m^+\} &= \delta_{nm} \end{aligned} \quad (9.426)$$

Let us now compute the correlation function:

$$\langle \psi(z_1)\psi(z_2) \rangle = \frac{1}{(z_1 z_2)^{1/2}} \left\{ \frac{1}{2} + \sum_{n=1}^{\infty} \left( \frac{z_2}{z_1} \right)^n \right\} = \frac{1}{2} \frac{z_1 + z_2}{(z_1 z_2)^{1/2}} \frac{1}{z_1 - z_2} \quad (9.427)$$

which is just what we got from (9.412). In the case when spin operators and the cut are absent, the mode expansion would be

$$\psi(z) = z^{-1/2} \left[ \sum_{n=1/2}^{\infty} (b_n z^{-n} + b_n^+ z^n) \right] \quad (9.428)$$

and

$$\langle \psi(z_1)\psi(z_2) \rangle = \frac{1}{(z_1 z_2)^{1/2}} \sum_{n=1/2}^{\infty} \left( \frac{z_2}{z_1} \right)^n = \frac{1}{z_1 - z_2}$$

as it should be.

We see that the role of spin operators is to shift us from half-integer modes in (9.428) to integer ones in (9.425). Moreover, the dimensionality of the spin operator,  $\Delta = 1/16$  has a natural interpretation as the change of zero-point energies in the presence of the spin. Indeed, the following relation is true:

$$\begin{aligned} \Delta &= -\frac{1}{2} \left( \sum_{n=1}^{\infty} n - \sum_{n=1/2}^{\infty} n \right) \\ &= \frac{1}{24} + \frac{1}{48} = \frac{1}{16} \end{aligned} \tag{9.429}$$

where we have used the following formula for the regularized sums:

$$\begin{aligned} \sum_{n=0}^{\infty} (n-j) - \sum_{n=0}^{\infty} n &= -\frac{1}{2}j(j+1) \\ \sum_n \omega_n &\stackrel{\text{def}}{=} \lim_{\epsilon \rightarrow 0} \sum_n \omega_n e^{-\epsilon \omega_n} \end{aligned} \tag{9.430}$$

In the  $\mathcal{D}$  dimensional case the only difference which we encounter is that we now have the dimensionality of the spin:

$$\Delta = \mathcal{D}/16 \tag{9.431}$$

since *all*  $\psi_\mu$  change their zero-point fluctuations, and also instead of (9.425) we have:

$$\psi_\mu = z^{-1/2} \left\{ \frac{\gamma_\mu}{2^{1/2}} + \gamma_{\mathcal{D}+1} \sum_{n=1}^{\infty} (b_n z^{-n} + b_n^+ z^n) \right\} \tag{9.432}$$

where  $\gamma_\mu$  and  $\gamma_{\mathcal{D}+1}$  are the usual Dirac matrices.

Using these formulas it is easy to compute correlation functions, involving two spin operators. Of course the mode expansion method is inadequate when more spins are involved. In these cases methods based on operator algebra have to be used. It is not hard to construct any correlator of spins.

Let us recapitulate. We have started with the supersymmetric (on the world sheet) action containing  $x_\mu$  and  $\psi_\mu$  fields. Loosely speaking we had  $\gamma$ -matrices ( $\psi_\mu$ ) distributed on the world sheet. We have shown, that there exist, apart from the ordinary, bosonic sector of such a string, another, “soliton” sector described by antiperiodic boundary conditions for  $\psi_\mu$  at the injection points. As a result, each injection point where the “soliton” is concentrated can be described by a spin operator, which is a spinor in space time. This implies that the string has spinorial excitations, apart from the bosonic ones.

In principle it is possible to describe this string not by the  $\psi_\mu$  fields, but by the spin fields  $S_a$  on the world sheet. In the case of particles this corresponds to transition from the action (9.317) to (9.335). Unfortunately, since the  $S_a$  have anomalous dimension  $\mathcal{D}/16$  they are far from free fields, and no practically useful description has yet been found. I am sure it exists, however. The situation is somewhat easier for  $\mathcal{D} = 10$ , since then the light-cone gauge is possible and the effective dimensionality of the spin becomes  $(\mathcal{D} - 2)/16 = 1/2$ . In this case, as can be expected, the spin fields are free and the fermionic string can be described by the noncovariant Green-Schwarz action.

In the covariant formalism we have to construct the vertex operator for emission of a space-time fermion. It must be something like

$$V_{p,a} \sim S_a(\xi) e^{i p \cdot x(\xi)} \quad (9.433)$$

where  $a$  is a spinor index. However, (9.433) is not good by itself, since it does not have the required dimension 1 for the vertex. The reason is that, while  $S_a$  has changed the spinor structure for the  $\psi$ -fields, it is necessary to do the same job for the gravitino and ghost fields, which enter into the functional integral. There are formal constructions for completing this task, but an appealing derivation from first principles is still lacking. Still we can compute, if we like, amplitudes containing two fermions. For this we use bosonic emission vertices

$$V_{p\mu} = : (i\partial_z x_\mu + p_\nu \psi_\mu \psi_\nu) e^{i p \cdot x} :$$

and presume that  $\psi_\mu$  has the mode decomposition (9.432). This just means that we have spin operators at 0 and  $\infty$  and examine the scattering of bosons (in arbitrary number) by one fermion. For several fermions one has either to use the algebraic construction for the fermionic vertex alluded to above or to pass to the light-cone gauge. Hopefully, the situation will improve soon.

So in the critical dimension ( $\mathcal{D} = 10$ ) the fermionic string has the following properties. Its ground state is massless (being vector for the open string case and tensor for the closed one) and supersymmetric in the space-time sense, because the number of states in the Ramond sector is equal to that in the Neveu-Schwarz sector.

This point must be clarified, since certain projections in both sectors are needed to reach this conclusion. We have already mentioned them in passing, but now we will discuss their meaning.

When we worked in the bosonic sector, without spin operators, we considered only vertices which contain an even number of  $\psi_\mu$ , or in

other words, we projected onto the sector even under  $\psi_\mu \rightarrow -\psi_\mu$ . This projection eliminated the tachyon with its vertex  $:(p_s z_s)e^{i p \cdot x_s}$ ; also it is clearly self-consistent. But what is its true origin?

Let us discuss first, how to make this projection in terms of functional integrals. If we are interested in the partition function:

$$Z = \text{Tr} e^{-\beta H_{NS}} = Z^F \cdot Z^B = \int \mathcal{D}\psi \exp\left(-\int \psi \cdot \not{\partial} \psi d\sigma d\tau\right) \cdot Z^B \quad (9.434)$$

(where  $H_{NS}$  is the Neveu-Schwarz hamiltonian), then the integral in (9.434) must be formulated on a torus, with time periodicity  $\beta$  and space periodicity  $2\pi$ . The usual boundary conditions will be antiperiodic in both directions; we have already explained that space antiperiodicity implies consideration of NS-excitations, while the time antiperiodicity is the standard thing needed for the description of the partition functions of fermions. We can write symbolically:

$$Z^{(F)} = \text{Diagram} \quad (9.435)$$

Let us consider now the projection. It is easy to check that taking *periodic* fermions in the time direction amounts to computing  $\text{Tr} (-1)^F \exp(-\beta H)$ , where  $F$  is the fermion number. Hence, the projection onto even  $F$  will be achieved if we take:

$$Z = \text{Diagram}_+ + \text{Diagram}_- \quad (9.436)$$

Adding the Ramond sector, we arrive at the simple rule. In order to describe the projected NSR model we have to sum over all possible spin structures. This gives the supersymmetric string theory. Presumably, for higher genus surfaces the prescription must be the same.

We do not have a good derivation of this fact. However, some explanation can be given. Let us show that only under the above prescription is it possible to treat the system in terms of spin operators. For that matter, take an Ising model on a surface with high genus. We know that usually this model can be replaced by free fermions. Is this still true? In fermionization of Ising spins a crucial role is played by Kramers-Wannier duality (see the next chapter). Fermionic lines are essentially the boundaries of drops containing reversed spins. However, if the surface is homologically nontrivial, there are closed paths which do not form boundaries of anything. We must ensure that fermionic trajectories corresponding to these paths do not contribute. The way to achieve this is just to sum over spin structures, since then each

homologically nontrivial path will be cancelled by one of the opposite spin structure.

We conclude, that the global introduction of spin operators requires summation over spin structures. Space-time sypersymmetry must also be explainable along the same lines, but it is not at the moment clear how to do it.

---

## CHAPTER 10

# Attempt at a Synthesis

---

---

DOI: 10.1201/9780203755082-10

What is the use of string theories? What is their connection with the physical problems described in the other chapters? There is no final answer to these questions. In this chapter I shall describe reasons to believe that string theories are important and perhaps present the master-key to the most fundamental problems.

Three kinds of questions will be in the centre of our attention. We begin with the project of grand unification on the string basis, then describe the 3-D Ising model and conclude with 4-D chromodynamics.

### 10.1 Long Wave Oscillations of Strings in Critical Dimensions

We have seen in the previous chapter that string theories in critical dimension contain massless spin 2 particles in their spectrum (and their superpartners for superstrings). Since the only consistent interaction for these particles is the gravitational one, these string theories must contain within themselves a theory of gravity. While in the low energy limit this theory will necessarily be Einstein-like, for high energies it will have good regular behaviour because strings are extended objects.

The basic idea of unified theories is to presume that all our low-energy world arises from massless string excitations which include super-gravitons, gauge fields and matter fields. All massive modes of strings are presumed to be very heavy (of the order of the Planck mass), and therefore unobservable.

Let us explain how this idea can be implemented.

First of all, why and when do we have massless excitations in the string theory? To find the partial answer to this question, let us recall that an ordinary string has the action

$$S = \int g^{1/2} g^{ab} \partial_a x \partial_b x d^2 \xi \quad (10.1)$$

which was obtained from the Nambu action

$$\begin{aligned} S &= \int (\det \partial_a \cdot x \partial_b x)^{1/2} d^2 \xi \\ &= \int \{g^{1/2} + \lambda^{ab}(\partial_a x \cdot \partial_b x - g_{ab})\} d^2 \xi \end{aligned} \quad (10.2)$$

by presuming condensation of the Lagrange multiplier,  $\lambda^{ab}$ :

$$\langle \lambda^{ab} \rangle = \text{const. } g^{1/2} g^{ab} \quad (10.3)$$

Therefore, in the phase described by (10.3) we have to interpret  $g_{ab}$  in (10.1) as an induced metric  $\partial_a x \cdot \partial_b x$ .

Let us consider a scale transformation in space-time:  $x \rightarrow \lambda x$ . It is equivalent to a Weyl transformation on the world sheet:

$$g_{ab}(\xi) \rightarrow \lambda^2 g_{ab}(\xi) \quad (10.4)$$

We have shown in the previous chapter that if the dimension  $\mathcal{D}$  is critical ( $\mathcal{D} = 26$  for the bosonic case and  $\mathcal{D} = 10$  for the fermionic one) then the Liouville mode decouples and the theory is Weyl invariant.

The only way to ensure scale invariance in space-time is to have a massless dilaton in the string spectrum. It is indeed there. Perhaps, along the same lines one can explain the graviton as well.

With all these massless modes at hand, one has to be able to compute their effective action in the low energy approximation. In principle, since we know the rules for computing the  $S$ -matrix through the averages of vertex operators it is straightforward to go to the low energy limit in these formulas. However, this direct way is very cumbersome and not very illuminating. There exists an interesting alternative which we describe now.

Let us consider the string theory in a curved background:

$$S = \frac{1}{2} \int \gamma_{\mu\nu}(x(\xi)) g^{1/2} g^{ab}(\xi) \partial_a x^\mu \partial_b x^\nu d^2 \xi \quad (10.5)$$

where  $\gamma_{\mu\nu}(x)$  is some fixed metric tensor of the  $\mathcal{D}$ -dimensional space. We are going to show, that the conditions for conformal invariance of the action (10.5) give equations for the  $\gamma_{\mu\nu}$  (and some other fields) which coincide precisely with the ones obtained from the  $S$ -matrix. Therefore, the problem of the low energy limit reduces to the computation of the  $\beta$ -functions for (10.5) treated as a nonlinear  $\sigma$ -model.

Let us assume first, that the world sheet is flat,  $g_{ab} = \delta_{ab}$ . Then, the action

$$S = \frac{1}{2} \int \gamma_{\mu\nu}(x(\xi)) \delta^{ab} \partial_a x^\mu \partial_b x^\nu d^2 \xi \quad (10.6)$$

is renormalizable in the following sense. Suppose, that we integrate over fast components of the  $\mathbf{x}$ -field, with wave vectors lying between  $\Lambda$  and  $\bar{\Lambda}$ . As we will show now, the result will again be an action of the form (10.6) but with a renormalized metric tensor  $\tilde{\gamma}_{\mu\nu}(\mathbf{x})$ . So, the set of all possible  $\gamma_{\mu\nu}(\mathbf{x})$  plays the role of the coupling constants in more ordinary theories. In particular the nonlinear  $\sigma$ -models, considered in the beginning of the book, correspond to the special case when  $\gamma_{\mu\nu}(\mathbf{x})$  are taken to have constant curvature (this constraint is reproduced under renormalization). The coupling constants which we have introduced before are nothing but the values of these constant curvatures.

The computation begins, as usual, with the decomposition into fast  $\mathbf{y}$  and slow  $\mathbf{x}_0$  parts of the field:

$$\mathbf{x}(\xi) = \mathbf{x}_0(\xi) + \mathbf{y}(\xi) \tag{10.7}$$

Substituting (10.7) into (10.6) we obtain:

$$S(\mathbf{x}) = S(\mathbf{x}_0) + S_{II}(\mathbf{x}_0, \mathbf{y}) \tag{10.8}$$

where  $S_{II}(\mathbf{x}_0, \mathbf{y})$  is quadratic in  $\mathbf{y}$ . Linear terms in  $\mathbf{y}$  are absent, because they are of the form:

$$S_I(\mathbf{x}_0, \mathbf{y}) = \int \mathbf{y} \frac{\delta S}{\delta \mathbf{x}_0} d^2\xi \tag{10.9}$$

and while  $\delta S/\delta \mathbf{x}_0$  contains wave vectors  $\leq \bar{\Lambda}$ , wave vectors of  $\mathbf{y}$  lie in the range from  $\bar{\Lambda}$  to  $\Lambda$ . Hence, with our accuracy, the integral (10.9) is zero.

As far as  $S_{II}$  is concerned it has the form:

$$\begin{aligned} S_{II} = & \int \left\{ \frac{1}{2} \gamma_{\mu\nu}(\mathbf{x}_0(\xi)) \partial y^\mu \partial y^\nu \right. \\ & + \partial_\alpha \gamma_{\mu\nu}(\mathbf{x}_0) \partial_\alpha x_0^\mu y^\alpha \partial_\alpha y^\nu \\ & \left. + \frac{1}{4} \partial_\alpha \partial_\beta \gamma_{\mu\nu}(\mathbf{x}_0) \partial_\alpha x_0^\mu \partial_\beta x_0^\nu y^\alpha y^\beta \right\} d^2\xi \end{aligned} \tag{10.10}$$

In principle it is not hard to compute logarithmic corrections directly from (10.10). However, since the expected answer must be covariant in  $x$ -space, it is appropriate to recast (10.10) into an explicitly covariant form. Again, there is a "brute force" way of doing it, but it is more reasonable to perform the computations by using a slightly advanced notation.

Let us notice first that if

$$(A, B) = \int d^2\xi \gamma_{\mu\nu}(x(\xi)) A^\mu(x(\xi)) B^\nu(x(\xi)) \tag{10.11}$$

then its variation under  $x(\xi) \rightarrow x(\xi) + y(\xi)$  is given by:

$$\delta(A, B) = (\nabla_y A, B) + (A, \nabla_y B) \quad (10.12)$$

where we have introduced a covariant derivative in the  $y$ -direction:

$$(\nabla_y A)^\mu = y^\lambda \left( \frac{\partial A^\mu}{\partial x^\lambda} + \Gamma_{\lambda\rho}^\mu A^\rho \right) \quad (10.13)$$

Using these relations, we compute first the first variation,  $S_1$  of the action  $S$ :

$$\begin{aligned} S_1 &= (v_a, \nabla_y v_a) \\ v_a^\mu &= \partial_a x^\mu \end{aligned} \quad (10.14)$$

Owing to the symmetry of  $\Gamma_{\lambda\rho}^\mu$  (absence of torsion) we have:

$$(\nabla_y v_a)^\mu = (\nabla_a y)^\mu = \partial_a y^\mu + \Gamma_{\lambda\rho}^\mu \partial_a x^\lambda y^\rho \quad (10.15)$$

Varying once more, we find:

$$\begin{aligned} S_{II} &= \delta(v_a, \nabla_a y) \\ &= (\nabla_a y, \nabla_a y) + (v_a, \nabla_y \nabla_a y) \\ &= (\nabla_a y, \nabla_a y) + (v_a, [\nabla_y, \nabla_a] y) \end{aligned}$$

(where we have set to zero the  $\nabla_a v_a$  term, which is the classical equation of motion).

The commutator of two covariant derivatives is the curvature tensor  $R$ :

$$\begin{aligned} (v_a, [\nabla_y, \nabla_a] y) &= (v_a, R(y, v_a) y) \\ &= \int R_{[\alpha\lambda], [\beta\sigma]} y^\alpha \partial_a x^\lambda y^\beta \partial_a x^\sigma d^2 \xi \end{aligned} \quad (10.16)$$

So, going back to the usual notation, we have transformed (10.10) into the form:

$$\begin{aligned} S_{II} &= \int \frac{1}{2} (\gamma_{\mu\nu}(x_0(\xi))) \nabla_a y^\mu \nabla_b y^\nu \\ &\quad + R_{\alpha\mu\beta\nu}(x_0(\xi)) y^\alpha y^\beta \partial_a x_0^\mu \partial_b x_0^\nu g^{ab} d^2 \xi \\ \nabla_a y^\mu &= \partial_a y^\mu + (\omega_a)^\mu_\nu y^\nu \\ (\omega_a)^\mu_\nu &= \Gamma_{\nu\lambda}^\mu \partial_a x_0^\lambda \end{aligned} \quad (10.17)$$

(where we have reintroduced the omitted metric tensor of the world sheet).

Notice, that in the case when  $\gamma_{\mu\nu}$  describes a sphere, (10.17) is precisely (2.40) derived for the  $n$ -field.

It is straightforward now to compute the cut-off dependence of the one loop partition function.

First of all, the logarithmic correction to the effective action arise from the second term in (10.17):

$$\begin{aligned}
 S_{\text{eff}}^{(1)} &= \int \frac{1}{2} R_{\alpha\mu\beta\nu}(x_0(\xi)) \partial_a x_0^\mu \partial_b x_0^\nu g^{1/2} g^{ab} \langle y^\alpha y^\beta \rangle d^2 \xi \\
 &= \int \frac{1}{2} R_{\mu\nu}(x_0(\xi)) \partial_a x_0^\mu \partial_b x_0^\nu g^{1/2} g^{ab} d^2 \xi \int_{\frac{\Lambda}{\bar{\Lambda}}}^{\Lambda} \frac{d^2 k}{(2\pi)^2 k^2} \quad (10.18)
 \end{aligned}$$

As a result, we have renormalization of  $\gamma_{\mu\nu}(x)$  in the form:

$$\delta\gamma_{\mu\nu}(x) = \frac{1}{2\pi} \log \frac{\Lambda}{\bar{\Lambda}} R_{\mu\nu}[\gamma] \quad (10.19)$$

If the world sheet were flat, there would have been no other one loop divergences. Counter-terms dependent on  $\omega_a$  do not arise, since they can appear only through

$$\phi_{ab}^2 = (\partial_a \omega_b - \partial_b \omega_a + [\omega_a, \omega_b])^2$$

which is a dimension 4 operator.

However, in the case of a curved world sheet there is another type of counter-term which must be taken into account (as noticed by Fradkin and Tseytlin). It is clear, that we have a dimension 2 object composed of the external field  $g_{ab}$ —the curvature of the world sheet,  $R(\xi)$ . So we can expect counter terms, which do not depend on the derivatives of  $x$ , but have the form:

$$\bar{S} = \int \phi(x(\xi)) R(\xi) g^{1/2} d^2 \xi \quad (10.20)$$

This kind of the logarithmic divergence already appears in flat space,  $\gamma_{\mu\nu} = \delta_{\mu\nu}$ , since the determinant of the Laplacian contains the term:

$$\text{Tr} \log \Delta \sim \log \Delta \int R g^{1/2} d^2 \xi \quad (10.21)$$

When we switch on  $\gamma_{\mu\nu}$  the coefficient in (10.21) gets modified. This modification begins at the two loop order, for which we have to expand the action (10.17) further. To give an idea of what happens, let us look at the term (one of many) which appears in the expansion:

$$S_{\text{IV}} \sim \int g^{1/2} g^{ab} R_{\alpha\mu\beta\nu}(x_0(\xi)) y^\alpha y^\beta \partial_a y^\mu \partial_b y^\nu d^2 \xi \quad (10.22)$$

Averaging this term in  $y$  and using the formulas:

$$\begin{aligned} \langle y^\alpha y^\beta \rangle &= \gamma^{\alpha\beta}(x_0(\xi)) \left( \frac{1}{2\pi} \log \Lambda + \text{finite part} \right) \\ \langle \partial_a y^\alpha \partial_a y^\beta \rangle &= \gamma^{\alpha\beta}(x_0(\xi)) [k_1 \Lambda^2 + k_2 e^{-\varphi} \partial^2 \varphi(\xi)] \\ &\quad (g_{ab} = e^\varphi \delta_{ab}) \end{aligned} \quad (10.23)$$

we obtain a logarithmically divergent contribution to the effective action:

$$W \sim \log \Lambda \int R(x_0(\xi)) (g(\xi))^{1/2} d^2 \xi \quad (10.24)$$

Quadratic divergence renormalizes the cosmological constant and does not concern the massless sector.

As a result of these computations one has to add to (10.19) a renormalization equation for the  $\phi$ -field, since we have shown that this term is needed for the overall renormalizability. We have in the lowest order ( $\mathcal{D} = 26$ ):

$$\begin{aligned} \delta \gamma_{\mu\nu} &= \frac{1}{2\pi} \log \frac{\Lambda}{\bar{\Lambda}} (R_{\mu\nu}[\gamma] + c \nabla_\mu \nabla_\nu \phi) \\ \delta \phi &= (c_1 R(x) + c_2 (\nabla_\alpha \phi(x))^2 \\ &\quad + c_3 \nabla^2 \phi(x)) \log \frac{\Lambda}{\bar{\Lambda}} \end{aligned} \quad (10.25)$$

The origin of the  $\phi$ -dependent terms is simple: as we expand (10.20) in  $y$ , we get:

$$\begin{aligned} S_{\text{II}} &= \int \frac{1}{2} (\nabla_\alpha \nabla_\beta \phi(x_0(\xi))) R(\xi) y^\alpha y^\beta g^{1/2} d^2 \xi \\ &\quad + \int (\nabla_\alpha \phi(x_0(\xi))) R(\xi) y^\alpha(\xi) g^{1/2} d^2 \xi \end{aligned} \quad (10.26)$$

Using (10.23) and taking the second term in the second order we get the structure (10.25).

It is not hard to compute all the coefficients, after which one obtains the renormalization group equations:

$$\begin{aligned} \dot{\gamma}_{\mu\nu} &\equiv \frac{\partial \gamma_{\mu\nu}}{\partial \log \Lambda} = -\frac{1}{2\pi} R_{\mu\nu}(x) - \nabla_\mu \nabla_\nu \phi \\ \dot{\phi} &= (\nabla \phi)^2 - \frac{1}{4\pi} \nabla^2 \phi - \frac{1}{16\pi^2} R(x) \end{aligned} \quad (10.27)$$

These equations can be written in the form:

$$\begin{aligned} \dot{\gamma}_{\mu\nu} &= 8\pi e^{8\pi\phi} \frac{\delta\Gamma}{\delta\gamma^{\mu\nu}} \\ \dot{\phi} &= \frac{1}{8\pi} e^{8\pi\phi} \frac{\delta\Gamma}{\delta\phi} \end{aligned} \tag{10.28}$$

$$\Gamma = \int e^{-8\pi\phi} \left[ \frac{1}{4\pi} \nabla^2\phi - (\nabla\phi)^2 + \frac{1}{16\pi^2} R(\gamma) \right] \gamma^{1/2} d^{26}x$$

A remarkable fact which we are now going to discuss is that this effective action  $\Gamma$  is just the one which could be obtained by taking the low energy limit of the string  $S$ -matrix. Moreover, if we proceed to higher terms in the  $\beta$ -functions they would correspond to the higher terms of the low energy expansion.

In order to clarify this amazing connection between the properties of the world sheet of the string and the space-time into which it is embedded, let us consider this problem in a slightly different context.

Let us suppose that we have a string, on the world sheet of which "lives" some conformal field theory. Let us assume also, that in the set of all operators of this conformal field theory there are operators  $\{u_n(\xi)\}$  with dimension two. If the resulting theory has zero total central charge (after summing contributions from the coordinate of the string, the ghosts and the extra conformal theory), the vertex operators of the form

$$v_p^{(n)} = u_n(\xi) e^{ip \cdot x(\xi)} \tag{10.29}$$

describe scalar massless particles. We are going to demonstrate that the effective action for these particles is related to the  $\beta$ -functions.

Let us look at the perturbed lagrangian:

$$\mathcal{L} = \mathcal{L}_0 - \sum_n \lambda_n u_n(\xi) \tag{10.30}$$

The operators  $u_n$  are supposed to satisfy the operator product expansion:

$$u_n(\xi)u_m(0) = \frac{\delta_{nm}}{|\xi|^4} I + \frac{1}{|\xi|^2} f_{nmi} u_i(0) + \text{less singular terms} \tag{10.31}$$

(where  $f_{nmi}$  are the symmetric structure constants of the operator algebra which define the normalization of the three-point functions:

$$\langle u_n(\xi_1)u_m(\xi_2)u_l(\xi_3) \rangle = \frac{f_{nmi}}{|\xi_1 - \xi_2|^2 |\xi_2 - \xi_3|^2 |\xi_3 - \xi_1|^2} ) \tag{10.32}$$

Now, let us examine the perturbation theory for the partition function:

$$Z = \left\langle \exp \int \sum_n \lambda_n u_n(\xi) d^2 \xi \right\rangle \quad (10.33)$$

When we expand this expression in  $\lambda$  we get different products of  $u$ -operators, integrated over  $\xi$ -space. These integrals are divergent because of the singularities in the operator product (10.31). Therefore, we have to introduce a cut-off. The renormalization groups tells us how we should change  $\lambda_n$  under a given change of the cut-off, so that  $Z$  remains unchanged.

In order to find these conditions, we shall first of all discard all quadratic divergences, coming from the first term in (10.31). Formal justification of this lies in the fact that these divergences are absent if we use a dimensional regularization scheme. The real reason is that they are purely short-distance effect, contributing to  $Z$  but not to any connected Green function. On the contrary, logarithmic divergences, coming from the second term in (10.31) are physically important and we are now going to analyse them. Let us take the  $N$ th term in the expansion for  $Z$ :

$$Z^{(N)} = \int d^2 \xi_1 \dots d^2 \xi_N \langle u_{n_1}(\xi_1) \dots u_{n_N}(\xi_N) \rangle \lambda_{n_1} \dots \lambda_{n_N} \quad (10.34)$$

As, say,  $\xi_1 \rightarrow \xi_2$  this integral becomes divergent. Using (10.31) we find:

$$\begin{aligned} Z^{(N)} &\simeq f_{n_1 n_2 m} \lambda_{n_1} \lambda_{n_2} \dots \lambda_{n_N} 2\pi \log\left(\frac{1}{a}\right) \\ &\times \int d^2 \xi_2 \dots d^2 \xi_N \langle u_m(\xi_2) u_{n_3}(\xi_3) \dots u_{n_N}(\xi_N) \rangle \\ &+ \text{finite terms} \end{aligned} \quad (10.35)$$

where  $a$  is an ultraviolet cut-off. We derive from (10.35) that the change in  $a$ :  $a \rightarrow \tilde{a}$  can be compensated by a change of  $\lambda_n$  if we choose:

$$\lambda_n \rightarrow \tilde{\lambda}_n = \lambda_n + f_{nml} \lambda_m \lambda_l 2\pi \log\left(\frac{\tilde{a}}{a}\right) \quad (10.36)$$

In other words we have computed (approximately), how  $\lambda_n$  must depend on  $a$  so as to keep  $Z$   $a$ -independent. The answer is contained in the Gell-Mann-Low equation:

$$\frac{d\lambda_n}{d(2\pi \log(1/a))} \equiv \beta_n(\lambda) = -f_{nml} \lambda_m \lambda_l + O(\lambda^3) \quad (10.37)$$

We see from this formula that the  $\beta$ -function can be expressed in terms of the structure constants of the operator algebra. In higher orders in  $\lambda$  we would need structure constants of operator products involving several  $u$ -operators.

Now let us return to the string theory. Scattering amplitudes in the critical dimensionality are constructed out of simple vertex operators. If a typical string resonance is described by the vertex operator  $V_p(\xi)$ , where  $p$  is the momentum, then on mass shell this operator has dimension 2 and the scattering amplitude is given by:

$$\mathcal{A}(p_1, \dots, p_N) = \int d^2\xi_1 \dots d^2\xi_N \langle V_{p_1}(\xi_1) \dots V_{p_N}(\xi_N) \rangle \quad (10.38)$$

This is almost the same expression as (10.34). The main difference is that the short distance singularities governed by the operator product expansion reveal themselves in the form of poles of  $\mathcal{A}$  instead of logarithmic divergences. The residues of these poles are precisely the same structure constants which determine the  $\beta$ -functions. Let us consider the effective action for these massless particles,  $\Gamma(\phi_1, \dots, \phi_n)$ , where  $\{\phi_k\}$  are the fields corresponding to them. We presume that  $\{\phi_k\}$  are space-time independent, so that the particles carry zero momentum. The expansion of  $\Gamma$  in  $\phi$  starts with the cubic term (since the quadratic mass terms are equal to zero):

$$-\Gamma(\{\phi\}) \simeq \phi_n \phi_m \phi_l \langle V_0^{(n)}(0) V_0^{(m)}(1) V_0^{(l)}(\infty) \rangle + \dots \\ = f_{nml} \phi_n \phi_m \phi_l + O(\phi^4) \quad (10.39)$$

We have recalled in the derivation of this formula what was always implicitly assumed in the Koba-Nielsen integrals, that three integrations out of  $N$  are consumed by factoring out the  $SL(2, C)$  group. So the three point function does not have any integration at all.

We see from this formula that to this order the  $\beta$ -function is connected with the effective action:†

$$\beta_n(\phi) = \frac{\partial \Gamma}{\partial \phi_n}$$

and, therefore, the condition for stability  $\partial \Gamma / \partial \phi_n = 0$  coincides with the condition of conformal invariance  $\beta_n(\phi) = 0$ .

---

† Further conclusions arose in discussions with A. Zamolodchikov and partially use his results.

What happens in the next order? The effective action has the form:

$$-\Gamma(\{\phi\}) = f_{nmi} \phi_n \phi_m \phi_i + \phi_n \phi_m \phi_l \phi_k \left\{ \text{f.p.} \int d^2\xi \langle u_n(0) u_m(1) u_l(\infty) u_k(\xi) \rangle \right\} + O(\phi^5) \quad (10.40)$$

where the symbol f.p. means the finite part of the logarithmically divergent integral. Why have we taken it? The reason is that the four point function contains a pole term (which looks like a logarithmic divergence in (10.40)) and a contact term produced by the exchange of massive states. In the effective action the pole term must be omitted by definition (and generated back when solving the equations of motion). Thus we have (10.40), and analogous formulas to all orders in  $\phi$ .

In the fourth order it is easy to check that the conditions for conformal invariance  $\beta_n = 0$  again coincide with the equations of motion  $\partial\Gamma/\partial\phi_n = 0$ . It seems very likely that this is true to all orders. Indeed, the  $\beta$ -function is defined as the coefficient before the first power of the logarithm in the expansion of the renormalized  $\phi_n$ . As we have seen from the previous discussion, renormalization in the order  $N$  is obtained by replacing the  $N$  operators  $u(\xi)$  in  $Z$  in (10.34) by a single one, using the operator algebra:

$$u_{n_1}(\xi_1) \dots u_{n_N}(0) \underset{\xi_k \rightarrow 0}{\approx} C_{n_1 \dots n_N}^m(\xi_1, \dots, \xi_{N-1}) u_m(0) \quad (10.41)$$

The renormalization of  $\phi^m$  is given by:

$$\delta\phi^m = \int d^2\xi_1 \dots d^2\xi_{N-1} C_{n_1 \dots n_N}^m(\xi_1, \dots, \xi_{N-1}) \phi^{n_1} \dots \phi^{n_N} \quad (10.42)$$

This integral contains multiple logarithmic divergences, appearing because the fusion of any two  $\xi_a$  and  $\xi_b$  creates a factor  $|\xi_{ab}|^{-2}$ . If we subtract all these terms, then the only divergence comes from the overall scale integration. This is by definition the  $\beta$ -function. (In dimensional regularization we have to separate the first order pole from (10.42).)

Now, let us compare this expression with  $\partial\Gamma/\partial\phi^m$ :

$$-\frac{\partial\Gamma}{\partial\phi^m} = \sum \phi^{n_1} \dots \phi^{n_N} \text{f.p.} \int |R|^4 \langle u_m(R) u_{n_1}(1) u_{n_2}(0) \times u_{n_3}(\xi_3) \dots u_{n_N}(\xi_N) \rangle d^2\xi_3 \dots d^2\xi_N |_{R \rightarrow \infty} \quad (10.43)$$

In this integral we can at first consider all  $|\xi_j| \ll R$ . Then we can apply the operator product expansion (10.41) and get:

$$\int \langle u_m(R)u_{n_1}(1)u_{n_2}(0)\dots u_{n_N}(\xi_N) \rangle d^2\xi_3\dots d^2\xi_N$$

$$\simeq \int C_{n_1\dots n_N}^l(1, 0, \xi_3, \dots, \xi_N) d^2\xi_3\dots d^2\xi_N \langle u_m(R)u_l(0) \rangle_{\text{eff}} \quad (10.44)$$

The finite part of the first factor in this formula is precisely equal to the coefficient of the first power of the logarithm in (10.42), i.e. to the  $\beta$ -function. The second factor would be proportional to  $\delta_{ml}$  if it had not been for the contributions of  $\xi_j \sim R$  in (10.44). These contributions lead to

$$\langle u_m(R)u_l(0) \rangle_{\text{eff}} = \frac{1}{|R|^4} \left\{ g_{ml}(\phi) + A_{ml}(\phi) \log \frac{|R|}{a} + \dots \right\} \quad (10.45)$$

Separating the finite part we obtain the desired relation:

$$\frac{\partial \Gamma(\phi)}{\partial \phi^m} = g_{ml}(\phi) \beta^l(\phi) \quad (10.46)$$

from which it follows that the equations of motion are equivalent to the condition of conformal invariance. We have not given a complete proof of these formulas (for the reason that it has not been completed) but it seems not too difficult to obtain this proof along the lines sketched above.

The meaning of these results is quite transparent. They imply that if the conformal field theory, which lives on the world sheet is unstable under perturbations, then the string theory is unstable under condensation of the fields corresponding to these perturbations. If the two-dimensional theory tends to a fixed point, defined by  $\beta^n(\phi) = 0$ , then the original string theory will tend (after condensation of the fields) to a new string theory, with a spectrum defined by the dimensionalities of the operators at the fixed point. Stability in the sense of the renormalization group turns out to be stability in the sense of the mass spectrum. In particular, the mass matrix given by the second derivatives of  $\Gamma$ , according to (10.46) coincides with the matrix of anomalous dimensions, given by derivatives of  $\beta$ , as it should be.

The only question which we have not settled yet is the central charge of the resulting theory.

To investigate this problem let us introduce  $S(\xi)$  the trace of the energy-momentum tensor, and consider the pair correlation function in

momentum space,  $\langle S(k)S(-k) \rangle$ . This defines the reaction of the system to a weak external gravitational field in the conformal gauge. In the purely conformal theory this reaction is given by the Liouville action, while in a renormalizable field theory we have:

$$\langle S(k)S(-k) \rangle = C(\phi^1(k), \dots, \phi^n(k))k^2 \equiv C(k)k^2 \quad (10.47)$$

where we have introduced running (along the renormalization group trajectories) coupling constants. On the other hand, since a change of the cut-off is coupled to  $S(\xi)$  and owing to renormalizability it can be compensated by a change of couplings, we have the well-known identity:

$$S(\xi) = \sum_m \frac{\partial \phi^m}{\partial \log \Lambda} u_m = \sum_m 2\pi \beta^m(\phi) u_m \quad (10.48)$$

If we recall that:

$$\begin{aligned} \langle u_m(0)u_n(R) \rangle &= \frac{g_{mn}(\phi)}{|R|^4} + \dots \\ \langle u_m(k)u_n(-k) \rangle &= g_{mn}(\phi) \frac{1}{2\pi} \left( \log \frac{|k|}{\Lambda} \right) k^2 + \dots \end{aligned} \quad (10.49)$$

we find that the term proportional to the first power of the logarithm in  $C(k)$  is given by:

$$C(k^2) \simeq C - g_{mn}(\phi) \beta^m(\phi) \beta^n(\phi) 2\pi \log \frac{\Lambda}{|k|} \quad (10.50)$$

On the other hand, from (10.47):

$$\begin{aligned} C(k^2) &= C(\phi^1(k) \dots \phi^n(k)) \\ &\simeq C\left(\phi^1 - 2\pi \beta^1(\phi) \log \frac{\Lambda}{k}, \dots\right) \\ &\simeq C(\phi) - \beta^j \frac{\partial C}{\partial \phi^j} 2\pi \log \frac{\Lambda}{|k|} \end{aligned} \quad (10.51)$$

Therefore, we get:

$$\beta^n \left( \frac{\partial C}{\partial \phi^n} - \frac{\partial \Gamma}{\partial \phi^n} \right) = 0 \quad (10.52)$$

From this equation it follows, that along the renormalization group trajectory we have:

$$C(k^2) = c + \Gamma(\phi^1(k) \dots \phi^n(k)) \quad (10.53)$$

We see from this formula, that if the perturbation drives the system to a new fixed point, the central charge  $C(k^2)$  tends to some new constant value. Incidentally, since the system relaxes towards smaller value of the effective action  $\Gamma$ , this new central charge must be smaller than the original one (Zamolodchikov theorem).

The situation is rather amusing. Suppose that we start from a string theory in the critical dimension. Let us assume also, that there are massless modes which tend to condense, leading us to a new string theory. We see, that generically this new theory will have a noncritical central charge. The implication is that the Liouville field, originally absent, must somehow appear. The only exception to this rule is the case when all  $\beta$ -functions are identically zero.

By the way, what goes wrong if we consider formally the dual amplitudes for noncritical central charge without the Liouville mode? The answer to this question is that nothing will be wrong at the tree level (provided that  $C$  is less than critical—there will be ghosts otherwise). However, as we consider unitarity corrections to the amplitudes, or higher topologies, then we encounter unphysical singularities in momentum space.

So, we have traced the connection between the  $\beta$ -functions and the effective actions for the massless modes of the strings. Applying this conclusion to the case of the  $\sigma$ -model with the metric  $\gamma_{\mu\nu}(x)$  and the dilaton  $\phi(x)$ , we see that the low energy expansion of the  $S$ -matrix elements can be obtained by the loop expansion of the  $\beta$ -functions. Moreover, we see that the central charge is given by:

$$C_{\text{tot}} = \mathcal{D} - 26 + \Gamma(\gamma_{\mu\nu}(x), \phi(x))$$

$$S = \int \Gamma(\gamma_{\mu\nu}(x), \phi(x)) (\det \gamma)^{1/2} d^D x$$

(here  $S$  is the action which generates the  $S$ -matrix) and, according to previous formulas, the fields which minimize  $S$  have the property:

$$\Gamma(\gamma_{\mu}^{\text{cl}}(x), \phi^{\text{cl}}(x)) = \text{const.} \leq 0$$

Let us explain now the qualitative origin of these results. We have considered a string in the external fields and then minimized its effective action. This seems to be a bizzare procedure. The true meaning of it is the following. The string contains interacting massless modes. Let us suppose that they tend to form nonzero condensates. Then the picture of the random world-surface can be viewed as something like a sphere on which branching polymers, consisting of the propagators of these massless modes, grow. What we have really done, has been to prescribe

some amplitude for each plant and to define it self-consistently. We can say that the string is floating in its own condensate.

In principle there are corrections to the effective action, coming from higher topologies of the world sheet. Their structure has at this time not yet been investigated.

## 10.2 Possible Applications of Critical Strings

A unique property of the critical strings is that they describe gravitons. Let us discuss possible ways to utilize this fact. First of all, the bosonic strings, which we discussed in the previous section, are not totally consistent, since they contain tachyons. This flaw is absent in the case of fermionic strings, provided that we sum over spin structures on the world sheet (meaning, as we showed in the previous chapter, that we include space-time fermions into the spectrum).

According to the ideas of Sherk and Schwarz, such  $\mathcal{D} = 10$  Fermi-strings can describe our world; provided that we compactify somehow 6 out of the 10 dimensions.

In this section we shall briefly describe the picture of the world which arises in this way.

We begin from the  $\mathcal{D} = 10$  NSR string. As we know, it does not contain a tachyon and its ground states are massless. These ground states contain gravitons, dilatons, antisymmetric tensor fields, and their super-partners (constructed out of spin operators). The conjecture to be made is that the gravitons condense in such a way that the vacuum expectation value of  $\gamma_{\mu\nu}(x)$  has the form:

$$\begin{aligned}\langle \gamma_{mn}(x) \rangle &= \delta_{mn}, \quad m, n = 1, \dots, 4 \\ \langle \gamma_{\mu\nu}(x) \rangle &= G_{\mu\nu}(x^5, \dots, x^{10}), \quad \mu, \nu = 5, \dots, 10\end{aligned}$$

where the  $G_{\mu\nu}$  are to be determined by dynamics. Geometrically speaking this formula means that the  $\mathcal{D} = 10$  space has the structure  $M^4 \times K^6$  where  $M^4$  is Euclidean space and  $K^6$  is some curved and presumably compact manifold. At the moment we do not know what was the reason for this condensation, and, in particular, what or who has fixed  $\mathcal{D} = 4$  for our space. But, provided that it has happened, we can easily derive consistency conditions on  $G_{\mu\nu}$ .

Let us change notation slightly, and write  $x^m$ ,  $m = 5, \dots, 10$  as  $y^m$ ,  $m = 1, \dots, 6$  presuming that now  $\mu, \nu = 1, \dots, 4$ . The string action takes

the form:

$$S = \int g^{1/2} [g^{ab}(\partial_a x^\mu \partial_b x^\mu + G_{mn}(y) \partial_a y^m \partial_b y^n) + \epsilon^{ab} B_{mn}(y) \partial_a y^m \partial_b y^n] d^2 \xi + \text{fermionic part}$$

In this formula we have added the antisymmetric tensor field  $B_{mn}$ , which we have forgotten before. It is one of the massless modes of the string. In principle, it should have been taken into account in the  $\beta$ -function considerations. However, due to parity considerations, the  $B$ -modes can be omitted from the theory at the tree level, so we did not really make a mistake. In another version of the string theory they can be included. It is easy to realize, that these two options correspond to our decision, whether or not to include into consideration non-orientable surfaces. Their inclusion projects out the  $B$ -modes, since they couple through the tensor  $\epsilon_{ab}$  which is not defined in the nonorientable case.

The first question to pose is under what conditions does the above action include  $\mathcal{D} = 4$  gravity, among other things. As we have seen, these conditions can be obtained by requiring the theory to be conformally invariant with total central charge being zero.

First of all, this implies that the  $\sigma$ -model, associated with the  $y$ -space must be conformally invariant, that is its  $\beta$ -function must be zero. This gives us equations for the metric of the compact space  $K$ , and the  $B$ -field in this space. When the typical size of  $K$  is larger than the string parameter  $M^{-1}$ , we can use the low energy expansion, which essentially gives the Einstein equations for  $G_{mn}(y)$ .

If we insist on starting with the  $\mathcal{D} = 10$  superstring, then in order to preserve the central charge intact, the effective action must stay unchanged along the renormalization group trajectory, or, in other words, there must be a sequence of  $G_{mn}(y)$  which connect flat  $y$ -space with the desired space  $K$ , such that the  $\beta$ -function is zero for all these  $G_{mn}(y)$ . To prove the necessity of this condition, it suffices to notice that according to (10.52):

$$\frac{\partial \Gamma}{\partial (2\pi \log \Lambda)} = -\beta^n(\varphi) \frac{\partial \Gamma}{\partial \varphi^n} = -g_{nm}(\varphi) \beta^n \beta^m \leq 0$$

Hence if  $\beta^n \neq 0$  along the trajectory, but reaches a fixed point, the central charge at this fixed point will be smaller than the original one (Zamolodchikov's theorem).

This condition would be impossible to satisfy without supersymmetry, since the arguments of the Chapter 2 show that any compact

bosonic  $\sigma$ -model in the strong coupling limit must have a finite correlation length. When superpartners are included, the question becomes more complicated and the answer is not known in general. There are arguments to the extent that if  $K$  is Ricci-flat and has Kähler structure, then the supersymmetric  $\sigma$ -model has zero  $\beta$ -function perturbatively. If so, then we have an exact solution of the string equations in the tree approximation, and are in a position to start string perturbation theory. If it is not so, then one can either hope that string loop corrections will correct the effective action in such a way that  $K$  eventually will be a minimum, or we have to look for some other solutions.

At this point it is worth noticing, that the effect of the compactified dimensions is described by a conformal field theory on the world sheet, such that the total central charge is zero. In principle it is not necessary to insist that this field theory is just a  $\sigma$ -model of the six dimensional manifold. Any superconformal field theory will do, provided that it has the correct central charge, no world sheet supersymmetry breaking and stability (by stability we mean the absence of operators with dimensions less than two, so it is just stability in the sense of the renormalization group). Whether we should call these extra fields, living on the world sheet, "extra dimensions" is actually a matter of convention.

The properties, which we have listed are sufficient for having consistent four-dimensional gravity with some matter fields, including fermions. One has to impose further conditions, that some of these fermions are chiral, and also that there should be Yang-Mills fields in the theory.

One way of getting Yang-Mills fields is to exploit an old idea of Kaluza and Klein and to presume that  $K$  has some symmetry group acting on it, represented by Killing deformations  $\kappa^m$  which do not change the metric tensor  $G_{mn}$ . The vectors  $\kappa^m(y)$  satisfy the equation:

$$\nabla_m \kappa_n^{(A)} + \nabla_n \kappa_m^{(A)} = 0$$

(where  $A$  labels the generators of the symmetry group).

In the Kaluza-Klein approach, the vector fields were essentially the mixed components  $(m, \mu)$  of the metric tensor. In the string context this corresponds to the vertex operator for the Yang-Mills particle:

$$\Gamma_{p\mu}^{(A)} = \int d^2\xi g^{1/2} g^{ab} \partial_a x_\mu \partial_b y^m \kappa_m^A(y(\xi)) + \text{fermionic part}$$

One can show, that due to the Killing condition, this vertex describes massless vectors.

There are two problems in implementing this idea to the real world. First, the minimal symmetry group must be  $SU(3) \otimes SU(2) \otimes U(1)$  in order to include known interactions. Such a group can act only on manifolds of sufficiently large dimension. Namely, we must have one dimension for  $U(1)$  two dimensions for  $SU(2)$  (which can act on the two-dimensional sphere  $S^2$ ) and four dimensions for  $SU(3)$  (which acts naturally on  $\mathbb{C}\mathbb{P}^2$ ). Adding four space-time dimensions we get the minimal number of dimensions equal to eleven (the number obtained by Witten) which is larger than the critical dimension of the fermionic string.

A possible way out of this problem would be to find a  $\sigma$ -model which has a  $\beta$ -function with an isolated zero. In this case, as we have seen, the resulting central charge is less than the number of  $y$ -dimensions. Hence it is conceivably possible to begin with  $\mathcal{D} > 10$  (perhaps  $\mathcal{D} = 11$  is the best) and to descend to  $\mathcal{D} = 4$  from there. One example of isolated zeroes is given by the Wess-Zumino  $\sigma$ -model, which has a nonzero  $B_{mn}$ -background. In this case it is hard to preserve  $\mathcal{D} = 4$  supersymmetry, however. Other examples are  $\sigma$ -models with a  $\theta$ -term at  $\theta = \pi$ , but they are poorly investigated at the moment.

The second problem in the Kaluza-Klein approach is that it is rather hard to get chiral fermions (observed in nature), roughly because the field  $G_{mn}(y)$ , having even charge parity, does not distinguish opposite chiralities. Perhaps  $B_{mn}$  will help here as well, but no concrete solution has yet been found.

Another approach to the question of vector mesons and chiral fermions rests on the heterotic version of the fermionic string (developed by Gross, Martinec, Harvey, Rohm) which we shall briefly describe now. The major idea of the construction is to have world sheet supersymmetry acting only among left-moving particles. The lagrangian with such a property has the form:

$$\mathcal{L} = \frac{1}{2}g^{1/2}g^{ab}\partial_a x_\mu\partial_b x_\mu + \psi_{\mu+}\partial_-\psi_{\mu+} + \chi_-\psi_{\mu+}\partial_+x_\mu + \mathcal{L}_-$$

Here  $\psi_{\mu+}$  are the left-moving superpartners of  $x_\mu$ , and  $\chi_-$  the right moving gravitino; the term  $\mathcal{L}_-$  we shall discuss in a moment. The difference between the heterotic and the NSR Lagrangian is that in the latter case we have  $\psi_{\mu-}$ -fields as well and  $\chi_+$ , so that it is CP-symmetric.

The main property of the heterotic Lagrangian is CP-asymmetry on the world sheet. It can be formally obtained from the NSR-string by setting  $\chi_+ = \psi_{\mu-} = 0$ .

Now, repeating the analysis of the previous chapter we would find that generally speaking this theory is anomalous. When we take  $g_{ab} \simeq \delta_{ab} + h_{ab}$ , we find that in general the effective action for  $h_{ab}$  is of the form:

$$W \simeq c_+ \frac{K_+^3}{K_-} h_{--}^2 + c_- \frac{K_-^3}{K_+} h_{++}^2 + \text{possible local terms}$$

where  $c_+$  and  $c_-$  are the central charges of the left and right parts of the theory. We have seen in Chapter 9 that for  $c_+ = c_-$  the local term can be arranged so that  $W$  is gauge invariant. For  $c_+ \neq c_-$  this is not possible any more. Hence, if we want to maintain general covariance we have to equalize  $c_+$  and  $c_-$ .

To do this we recall that we have ghosts with  $c_+ = -15$  (for the supersymmetric part of the theory) and  $c_- = -26$  for the ordinary part. Therefore:

$$c_+^{\text{tot}} = \frac{3}{2}\mathcal{D} - 15 + \tilde{c}_+$$

$$c_-^{\text{tot}} = \mathcal{D} - 26 + \tilde{c}_-$$

We conclude that  $\tilde{\mathcal{L}}_-$  must describe a left-supersymmetric field theory on the world sheet with  $\tilde{c}_- - \tilde{c}_+ = \mathcal{D}/2 + 11$ . The critical dimension is given by  $\mathcal{D}_{cr} = \frac{3}{2}(10 - \mathcal{D})$ . The simplest choice would be to take  $\mathcal{D} = 10$ ,  $\tilde{c}_+ = 0$ ,  $\tilde{c}_- = 16$  by adding 32 right moving fermions on the world sheet. Then:

$$\tilde{\mathcal{L}} = \sum_{A=1}^{32} \chi^A_- \partial_+ \chi^A_+$$

Vector fields appear naturally through the vector vertex:

$$\Gamma_{p\mu}^{[AB]} = (i \partial_+ x_\mu + p_\nu \psi_{\mu+} \psi_{\nu+}) \chi^A_- \chi^B_- e^{ip \cdot x}$$

Right currents  $\chi^A_- \chi^B_-$  form the current algebra for the  $SO(32)$  group. We can in principle reduce this symmetry by choosing different spinor structures for different  $A$ , in which case it will be impossible to form conserved charges by integrating currents over the string. However, possible choices are strongly limited by an important condition of invariance under “large” diffeomorphisms or, which is the same, under the modular group.

Let us describe these large diffeomorphisms and show that lack of invariance under their action on the world sheet would manifest itself as gauge and gravitational anomalies in the space-time. Conversely, the condition of modular invariance is equivalent to the condition of cancellation of the above anomalies.

In the case of the torus, we saw that a conformal gauge can be chosen, provided that we represent this torus as a parallelogram, defined by vectors  $\omega_1 = (1, 0)$  and  $\omega_2 = (\text{Re } \tau, \text{Im } \tau)$  where  $\tau$  is a complex parameter. Integration over all metrics was reduced to integration over the Liouville field (absent in critical dimension and the integral  $\int d^2\tau$ . All integrands as functions of  $\tau$  must be invariant under a specific discrete group which we describe now. Therefore, to avoid repetitions one has to integrate only over the fundamental region in  $\tau$ -space.

Let us take a torus and cut it along a circle, making a cylinder out of it. Then twist one of the boundaries of this cylinder through the angle  $2\pi$  and glue it back, obtaining a torus again. Clearly, this is a diffeomorphism of the torus which is not homotopic to the identity. For example a large circle of the torus will wind once around it after such a diffeomorphism.

When we describe the torus by the parallelogram  $(\omega_1, \omega_2)$ , the effect of the two possible operations described above is the change:

$$(\omega_1, \omega_2) \rightarrow (\omega_1, \omega_1 + \omega_2)$$

$$(\omega_1, \omega_2) \rightarrow (\omega_1 + \omega_2, \omega_2)$$

The first transformation is simply the shift  $\tau \rightarrow \tau + 1$ . To find the effect of the second one we have to return the parallelogram to its original position. That will give us another transformation  $\tau \rightarrow 1 + 1/\tau$ . These two transformations generate the modular group. We can also describe this group as

$$\tilde{\omega}_i = A_{ik} \omega_k \quad i, k = 1, 2$$

where  $A_{ik}$  is a matrix with integer elements and with  $|\det A| = 1$ . Therefore, our group is  $SL(2, \mathbb{Z})$ . Similar considerations for arbitrary topology lead to the symplectic groups with integer coefficients, but we shall not need it here.

Our point will be to show that the lack of modular invariance on the world sheet must lead to space-time anomalies.

The main idea is that we have to check the nonpropagation of the spurious states on the torus. In the massless sectors, as we have seen, these spurious states are just the longitudinal vectors and gravitons. Hence, the absence of gauge and gravitational anomalies is equivalent to the above-mentioned nonpropagation.

At the tree level we have checked this property by showing that owing to the Ward identity  $\langle T(z) \rangle$  on the world sheet is zero.

In the case of the torus, Ward identities are more tricky and here modular invariance comes into play.

Let us derive Ward identities for the case of the torus. Let us recall, that if we make an infinitesimal coordinate transformation  $\epsilon(z, \bar{z})$ , then the action change by the amount:

$$\delta S = \int T(z) \partial_{\bar{z}} \epsilon(z, \bar{z}) d^2 z$$

while the fields transform as:

$$\delta \phi = \left( \epsilon(z, \bar{z}) \frac{\partial}{\partial z} + \Delta(\partial_z \epsilon) \right) \phi$$

If we choose  $\epsilon$  to be a solution of:

$$\partial_{\bar{z}} \epsilon = \delta^{(2)}(z - w)$$

we obtain the desired Ward identity. The main difference from the previous formulas is that  $\epsilon$  is not strictly periodic and thus changes the fundamental region. Namely, the solution of the above equation is given by the Weierstrass  $\zeta$  function:

$$\epsilon(z) = \zeta(z - w); \quad \zeta(z) = \sum_{\omega \in L} \left( \frac{1}{z - \omega} + \frac{1}{\omega} + \frac{z}{\omega^2} \right)$$

where  $L$  is a lattice formed by  $(\omega_1, \omega_2)$ . It is well known that:

$$\zeta(z + \omega_\alpha) = \zeta(z) + \eta_\alpha, \quad \alpha = 1, 2$$

and that the constants  $\eta_\alpha$  satisfy the relation:

$$\eta_1 \omega_2 - \eta_2 \omega_1 = 2/\pi i$$

It is easy to check, using these relations, that the above conformal transformation induces a constant change of  $\tau = \omega_2/\omega_1$ . The corresponding Ward identity is given by:

$$\begin{aligned} & \langle T(z) \phi(z_1) \cdots \phi(z_N) \rangle \\ &= \frac{\partial}{\partial \tau} \langle \phi(z_1) \cdots \phi(z_N) \rangle \\ &+ \left\{ \sum_k \left( -\Delta \zeta'(z - z_k) + \zeta(z - z_k) \frac{\partial}{\partial z_k} \right) \right\} \langle \phi(z_1) \cdots \phi(z_N) \rangle \end{aligned}$$

It should be stressed, that in this formula the averages are understood without dividing them by the partition function, which itself depends on  $\tau$ .

The role of modular invariance is evident by now. When we check the decoupling of the longitudinal states we have to integrate the Ward

identity on  $z_j$  and on  $\tau$ , restricting the  $\tau$  integral to the fundamental region. If the partition function violates modular invariance the term  $\partial/\partial\tau$  will give a nonzero contribution, coming from the boundaries. Thus a modular anomaly on the world sheet will manifest itself as a gauge and gravitational anomaly in space-time.

The conclusion is that in heterotic string theories one has to check modular invariance, which severely limits the possibilities. On the other hand, in ordinary closed strings modular invariance is automatic, since non-chiral determinants can be regularized in a manifestly covariant way. Such theories, even when they contain chiral fermions in space-time, are safe from anomalies.

Let us summarize. We see that we now have a consistent description of quantum gravity and supergravity, and we have many options for including chiral fermions and other particles into the scheme. This approach explains the enigma of generations, the strange copying which started with the muon. The number of copies now is determined by the number of zero modes of the generalized Dirac operators, which in turn is fixed by the topology of the background manifold. Actually it is hard to obtain a small enough number in this approach.

At the same time, we cannot explain at the moment why compactification took place, and most importantly, why the dimensionality of our world is equal to four. Until these problems are solved, we cannot be quite sure that we are on the right track. Still, there are many favourable signs.

### 10.3 The Three-Dimensional Ising Model†

It is well known that most of the phase transitions of the second kind occurring in Nature are equivalent, as far as critical behaviour is concerned, to the three dimensional Ising model. After Onsager's remarkable success with the two dimensional version of this model, many people have tried to find its solution in 3d (private communication). No such solution has been found and we have a strong feeling now that the model is not exactly solvable. At the same time, the total solution is not actually needed, because only the critical behaviour of this theory presents a universal interest. By this we mean that the

---

† Some of the results of this section have been obtained in collaboration with V. Dotsenko (1978).

critical properties of the Ising model (such as the singularity in the specific heat) are shared by many statistical systems in Nature. However if we are far from the phase transition point no lessons for other systems can be extracted.

In the present section we shall try to argue that in the continuum limit (i.e. in the vicinity of the phase transition point) the three-dimensional Ising model can be reduced to an exactly solvable system—the supersymmetric string. This result should allow in principle the determination of the critical behaviour, but it is still an unsolved problem.

In order to explain our ideas we have to recall the situation in the 2d Ising model, solved by Onsager. There are two kinds of basic variables in this model. First are spin variables  $\sigma_x$  (or order parameters) taking the values  $\pm 1$ . Second, there exist disorder variables,  $\mu_x$  first introduced by Kadanoff and Ceva which are defined as the endpoints of dislocation lines. These two sets of variables are dual to each other in the same way as electric and magnetic charges. It is remarkable that although equations of motion for  $\sigma_x$  and  $\mu_x$  separately are complicated, the product variable  $\psi = \sigma\mu$  satisfies a linear equation. Moreover, in the continuum limit (near the phase transition point) this linear equation is reduced to the two-dimensional Euclidean Dirac equation, implying that the  $\psi$ -variable describes a fermionic excitation. So, we can say that the two-dimensional Ising model is equivalent to a system of relativistic, noninteracting Fermi particles.

Let us turn now to the three-dimensional case. The basic difference here is connected with the nature of the disorder variables. The dislocation lines (surrounding the regions of the reversed spins) of the two-dimensional model are now replaced by dislocation surfaces. The boundary of these surfaces is now the argument on which the  $\mu$  variable depends. So, instead of variables  $\mu(x)$  of the 2d model we have now the contour variable  $\mu(C)$  (here  $C$  is a closed loop). Again, in order to obtain simple equations, we have to form the product of  $\mu(C)$  and  $\prod_i \sigma(x_i)$  where  $x_i$  are the points adjacent to the loop  $C$ . If we draw small normals to the loop  $C$ , connecting it with the points  $\{x_i\}$ , we obtain a string with pseudospins living on it, an object which resembles barbed wire. These “barbed wire” variables satisfy a linear equation in the loop space. The meaning of this equation is that each small piece of the barbed wire propagates exactly as an Ising fermion of the two dimensional model, if we consider the plane, orthogonal to the piece under consideration. Just as a Fermi particle propagator can be represented by a sum over all possible paths with a certain fermionic structure on

them, described in Section 10.3.1, the solution of the loop equation is found to be equal to a sum over surfaces with extra fermionic structure. This is described in Section 10.3.2.

The strategy which follows will be to find a continuous string model, such that the string moves piecewise as a free Dirac fermion. The natural candidate is the Neveu–Schwarz–Ramond string.

### 10.3.1 The Dirac Equation in the Two-Dimensional Ising Model

The two-dimensional Ising model is described by the partition function:

$$Z = \sum_{\{\sigma_x\}} e^{-\beta \mathcal{E}[\sigma_x]} \quad (10.54)$$

$$\mathcal{E}[\sigma_x] = - \sum_{x, \delta} \sigma_x \sigma_{x+\delta}$$

Here  $\beta$  is the inverse temperature, the points  $\{x\}$  belong to the two dimensional rectangular lattice,  $\sigma_x = \pm 1$ , and  $\{\delta\}$  are the two possible unit lattice vectors. Important physical information is condensed in the correlation functions

$$G_N(x_1, \dots, x_N) = \langle \sigma(x_1) \cdots \sigma(x_N) \rangle \quad (10.55)$$

defined in the obvious way. It is possible to obtain a Schwinger–Dyson chain of equations for  $G_N(x_1, \dots, x_N)$ . However, this way of proceeding would obscure the exact solvability of the model. The appropriate way is to introduce the so-called disorder variables. They are defined as follows. Consider a point  $x_*$  of the dual lattice (formed by the centres of the faces of the original lattice) and draw some path  $P$  on the dual lattice, leading from  $x_*$  to infinity. Change the sign of  $\beta$  on all bonds intersected by the path. Define the distorted partition function,  $\tilde{Z}(x_*, P)$ . Then, the disorder variable  $\mu(x_*)$  (defined through its Green functions) is given by:

$$\langle \mu(x_*) \rangle = \frac{\tilde{Z}(x_*, P)}{Z} \quad (10.56)$$

Analogously one defines

$$\langle \mu(x_{1*}) \mu(x_{2*}) \cdots \mu(x_{N*}) \rangle \quad (10.57)$$

Now, a simple argument shows that (10.57) does not depend on the paths leading from infinity to  $x_{k*}$ , but only on  $x_{k*}$  themselves. To show this, consider a closed path which surrounds some two-dimensional region, (a “drop”) and change the sign of the coupling on all the bonds

intersected by the path. Consider the modified partition function  $Z(P)$ . It is obvious now that  $Z(P) = Z$ , since to any given configuration contributing to  $Z(P)$  corresponds a configuration of the same energy contributing to  $Z$ . The latter configuration is obtained from the former one by reversing all the spins lying inside the drop. From this relation it follows that two different choices of the path in (10.56) and (10.57) give the same result. Let us notice parenthetically that such a simple definition of the disorder variable holds true for a lattice with simple topology (such that each closed loop bounds a drop). Otherwise there are different variables, classified by the first homology group.

In order to obtain linear equations let us consider the variable  $\psi$  which is formed by the product of the order parameter  $\sigma_x$  and the adjacent disorder variable. At each point of the original lattice  $x$ , we have four adjacent points of the dual lattice  $x_*^{(a)} = x + e_a$  where the four vectors  $e_a$  have length  $1/\sqrt{2}$  and are directed along the diagonals of the original lattice. Let us consider the four component object:

$$\psi_a(x) = \sigma(x)\mu(x + e_a), \quad a = 1, 2, 3, 4 \quad (10.58)$$

The "tail" necessary for the definition of  $\mu$  is supposed to go horizontally from  $x + e_a$  to the left infinity. Now, we have a simple identity:

$$\begin{aligned} \langle \psi_1(x) \rangle &= \langle \sigma(x) \prod_{n=0}^{\infty} \exp\{-2\beta(\sigma_{x-n\delta_1}\sigma_{x-n\delta_1+\delta_2})\} \rangle \\ \begin{array}{l} \delta_2 \\ \delta_1 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \end{array} &= \langle \sigma(x) \prod_{n=1}^{\infty} \exp\{-2\beta(\sigma_{x-n\delta_1}\sigma_{x-n\delta_1+\delta_2})\} \exp\{-2\beta\sigma_x\sigma_{x+\delta_2}\} \rangle \\ &= \langle \sigma(x)\mu(x + e_2) \rangle (\cosh(2\beta) - \sinh(2\beta)(\sigma_x\sigma_{x+\delta_2})) \\ &= \langle \psi_2(x) \rangle \cosh(2\beta) - \langle \psi_3(x + \delta_2) \rangle \sinh(2\beta) \end{array} \quad (10.59)$$

In the derivation of equation (10.59) all we have used was the definition of  $\mu$  (the product  $\prod_{n=0}^{\infty}$  represents the change from  $\beta$  to  $-\beta$  along the intersected bonds), the fact that  $\sigma_x^2 = 1$  and last but not least the possibility to turn the tail of  $\mu(x_*)$  if it does not intersect a spin variable. Proceeding in the same fashion we get:

$$\begin{aligned} \langle \psi_a(x) \rangle &= \cosh(2\beta)\langle \psi_{a+1}(x) \rangle \\ &\quad - \sinh(2\beta)\langle \psi_{a+2}(x + \delta_{a+1}) \rangle \end{aligned} \quad (10.60)$$

Here  $a = 1, 2, 3, 4$ . It is clear that  $\langle \psi_{a+4} \rangle$  and  $\langle \psi_a \rangle$  are essentially the same object but they are not identical. Namely  $\langle \psi_{a+4} \rangle$  is obtained by  $2\pi$  rotation of the arrow  $e_a$ . However, we must remember the horizontal tail attached to the end of the arrow. In the process of the rotation it

intersects once the point  $\mathbf{x}$ , where the spin  $\sigma_{\mathbf{x}}$  is placed. As a result the object  $\psi$  changes its sign (recall that we proved tail independence for a pure  $\mu$  correlator, but if there are  $\sigma$  present and the tail crosses  $\sigma$  the correlator changes sign). We conclude that the equation (10.60) must be augmented by the condition:

$$\psi_{a+4}(\mathbf{x}) = -\psi_a(\mathbf{x}) \quad (10.61)$$

The phase transition point is determined from (10.60) by looking for an  $\mathbf{x}$ -independent solution; substituting:

$$\psi_a \sim e^{\pm i\pi a/4} \quad (10.62)$$

in order to guarantee (10.61) we obtain from (10.60)

$$\sinh(2\beta_c) = 1, \cosh(2\beta_c) = \sqrt{2} \quad (10.63)$$

The other possible solution

$$\psi_a \sim e^{\pm 3i\pi a/4} \quad (10.64)$$

would lead to unphysical  $\beta_c$ . The two dependences (10.62) and (10.64) correspond to the spin 1/2 and spin 3/2 parts of the wave function. Near the point (10.63) the mode with spin 1/2 becomes soft and the mode 3/2 remains hard. The equation (10.60) is invariant under  $\pi/2$ -rotations with simultaneous rotation of the spin index  $a$ :

$$\psi_a(\mathbf{x}) \rightarrow \psi_{a+1}(\Gamma\mathbf{x}) \quad (10.65)$$

( $\Gamma$  is  $\pi/2$  rotation).

We may expect therefore that in the continuum limit, when only spin 1/2 propagates, we shall obtain the Dirac equation. This is indeed the case as is seen by expanding

$$\begin{aligned} \psi_a(\mathbf{x}) = & u_+(\mathbf{x})e^{i\pi(a+1/2)/4} + u_-(\mathbf{x})e^{-i\pi(a+1/2)/4} \\ & + v_+(\mathbf{x})e^{3i\pi(a+1/2)/4} + v_-(\mathbf{x})e^{-3i\pi(a+1/2)/4} \end{aligned} \quad (10.66)$$

Substituting (10.66) into (10.60) and neglecting the  $v$ -terms, we obtain (using the identity  $u_+(\mathbf{x}) = \frac{1}{4} \sum_a e^{-\pi i(a+1/2)/4} \psi_a$ ):

$$(\partial_1 + i\partial_2)u_+ = imu_- \quad (10.67)$$

$$(\partial_1 - i\partial_2)u_- = imu_+$$

$$m \sim \frac{\beta - \beta_c}{\beta_c}$$

Equation (10.67) is precisely the two-dimensional Dirac equation with the spinor  $u(\mathbf{x})$  transforming under rotations:

$$u(\mathbf{x}) \rightarrow e^{i\varphi/2} u(\Gamma(\varphi)\mathbf{x}) \quad (10.68)$$

( $\Gamma(\varphi)$  is a rotation with the angle  $\varphi$ .) The formula (10.68) is the continuum analogue of (10.65). Up to now we have dealt with the single field average  $\langle \psi_a(\mathbf{x}) \rangle$ . If we have to consider more complicated Green functions:

$$F_{a_1, \dots, a_l}(\mathbf{x}_1, \dots, \mathbf{x}_l) = \langle \psi_{a_1}(\mathbf{x}_1) \dots \psi_{a_l}(\mathbf{x}_l) \rangle \quad (10.69)$$

we find that with respect to each argument they satisfy the same equation (10.60) with the condition (10.61) but on the right hand side of this equation we shall have contact terms as usual. These terms reduce to the standard  $\delta$ -functions in the continuum limit. Let us demonstrate finally how to find the critical singularity in the specific heat. We have for the average energy density:

$$\begin{aligned} \langle E \rangle &= -2 \langle \sigma_x \sigma_{x+\delta_1} \rangle \\ &= -2 \langle \psi_1(\mathbf{x}) \psi_2(\mathbf{x} + \delta_1) \rangle \\ &= \langle u(\mathbf{x}) \bar{u}(\mathbf{x}) \rangle \\ &= \text{Tr} \int_{|\mathbf{p}| \ll 1} \frac{d^2 p}{(2\pi)^2} \frac{1}{m + i\hat{p}} \\ &= 2m \int \frac{d^2 p}{(2\pi)^2 (p^2 + m^2)} \sim m \log \frac{1}{m} \end{aligned} \quad (10.70)$$

where we have used the standard Dirac propagator. For the specific heat we get:

$$C \sim -\frac{\partial \langle E \rangle}{\partial \beta} \sim -\log \left| \frac{\beta - \beta_c}{\beta_c} \right|$$

which is Onsager's famous result.

### 10.3.2 The Three-Dimensional Case. The Loop Equation

The three-dimensional Ising model is defined again by equation (10.54) only now  $\mathbf{x}$  belongs to the three-dimensional cubic lattice. The order variable  $\sigma(\mathbf{x})$  is defined as before. A slight modification is needed for the  $\mu$ -variables since now the "drops" with reversed spins are three dimensional and their boundaries are formed by two-dimensional

surfaces. Let us introduce the dual lattice which is formed by the centres of the cubes of the original one. Consider a closed loop  $C$  on the dual lattice and form an arbitrary surface  $S_C$  (on this lattice) bounded by this loop. Define the distorted partition function  $\tilde{Z}(C, S_C)$  by changing the signs of all couplings on the original links intersected by  $S_C$ . Then  $\mu(C)$  is defined by:

$$\langle \mu(C) \rangle = \frac{\tilde{Z}(C, S_C)}{Z} \tag{10.71}$$

The definition (10.71) is a straightforward generalization of (10.56). Here again it appears that  $\langle \mu(C) \rangle$  does not depend on the choice of the surface  $S_C$ , because a closed surface of dislocation does not change  $Z$  (it is possible to reverse spins in the drop, bounded by  $S$ ). Therefore we can form loop Green functions like:

$$\langle \mu(C_1) \dots \mu(C_N) \rangle \tag{10.72}$$

If also  $\sigma(x)$  are present then these functions are defined up to a sign which is physically irrelevant.

As in the 2d case, we shall obtain simple equations not for  $\langle \mu(C) \rangle$  or  $\langle \sigma(x) \rangle$  but for certain mixed objects which will be called "fermionic string" or "barbed wire".

Let us supply the middle of each link(s) of  $C$  with one of the vectors  $e_{a_s}$  ( $a_s = 1, 2, 3, 4$ ) which lie in the plane orthogonal to the link under consideration. Let us consider the following object:

$$\Psi_{a_1, \dots, a_L}(C_L) = \mu(C_L) \prod_{s=1}^L \sigma(x_s + e_{a_s}) \tag{10.73}$$

Here  $L$  is the length of the loop  $C$  and  $x_s$  is the middle of the link  $s$ . In order to obtain an equation most closely resembling equations (10.67) it is convenient to notice that the average  $\langle \Psi_{a_1, \dots, a_L}(C_L) \rangle$  can be computed in two ways: first with the definition of the average by (10.54) and  $\mu$  given above, but also the dual way of computation is possible. By that we mean that we can define a dual Ising model by introducing variables  $\mu_{x, \alpha} = \pm 1$  attached to the links of the dual lattice and considering the partition function:

$$Z = \sum_{\mu_{x, \alpha}} \exp\left(\tilde{\beta} \sum_P \mu(\partial P)\right) \tag{10.74}$$

Here we have denoted by  $P$  plaquettes of the dual lattice and by  $\mu(\partial P)$  the product of  $\mu_{x, \alpha}$  around the plaquette. The dual temperature  $\tilde{\beta}$  is given by:

$$e^{-2\tilde{\beta}} = \tanh \beta \tag{10.75}$$

Kramers-Wannier duality ensures that the partition functions (10.74) and (10.54) are the same. The average of  $\sigma(x)$  variables (which are disorder variables with respect to (10.74)) is defined as follows. Attach an infinite path to the point  $x$  and change the sign of  $\tilde{\beta}$  on all plaquettes intersected by the path. Then:

$$\langle \sigma(x) \rangle = \frac{\tilde{Z}(x, P_x)}{Z} \quad (10.76)$$

After short meditation one concludes that (10.75), just as (10.56), does not depend on the choice of the path  $P_x$ . Indeed, let us consider a short closed path intersecting a bunch of four plaquettes having one common link. Changing the sign of the spin  $\mu$  on this link we get a new configuration which is equivalent to the one in the model without the path. Therefore for this small closed path the modified partition function coincides with the old one. Since any large path can be composed from small ones, we conclude that a closed loop of dislocations does not change the system.

We are in a position now to obtain the desired equation for  $\Psi_{a_1, \dots, a_2}(C)$ . Let us use the identities

$$\begin{aligned} e^{-2\tilde{\beta}\mu(\partial P)} &= \cosh(2\tilde{\beta}) - \sinh(2\tilde{\beta})\mu(\partial P) \\ \mu^2(\partial P) &= 1 \end{aligned} \quad (10.77)$$

and imagine a tail  $S$  attached to  $\sigma(x_s + e_{a_s})$ , which gives a contribution:

$$\sigma(x_s + e_{a_s}) \sim \prod_s e^{-2\tilde{\beta}\mu(\partial P)}$$

(where the product goes along all the plaquettes intersected by the tail  $S$ ). Using (10.77) (which is analogous to 10.59) we obtain:

$$\begin{aligned} \Psi_{a_1, \dots, a_2}(C) &= \cosh(2\beta)\Psi_{a_1, \dots, a_s+1, a_s+1, \dots, a_L}(C) \\ &\quad - \sinh(2\beta)\Psi_{a_1, \dots, a_s, a_s, a_s+2, a_s+3, \dots, a_L}(C + \Pi_{a_s}) \end{aligned} \quad (10.78)$$

Here the loop  $C + \Pi_{a_s}$  is obtained from the original loop by removing the link  $s$  and attaching instead the letter  $\Pi$  oriented in the direction  $a_s$ . The loop  $C + \Pi_{a_s}$  has length equal to  $L + 2$ , and on the two extra links we place indices  $a'_s$  and  $a'_s + 3$ , so that the corresponding  $\sigma$  are both placed in middle of the plaquette  $\Pi$ . Since  $\sigma^2 = 1$  this does not affect our relations.

We observe now a remarkable analogy between equation (10.78) and equation (10.60). Indeed (10.78) implies that if we concentrate on some link  $s$  then it propagates in the plane orthogonal to itself exactly as a

particle excitation in the two-dimensional Ising model. That does not mean of course that the 3d Ising string can be considered as a collection of noninteracting 2d Ising particles. The difference is that in the process of development the string does not break (as a consequence of gauge invariance) and hence when some link moves it creates extra portions of string, necessary for continuity. But this continuity is the only source of interaction. Such a system will be called a free string.

Let us now summarize our conclusions. We have proved that the 2d Ising model is equivalent to the problem of free particle propagation. This particle carries internal index  $a$ , or, more geometrically, an arrow  $e_a$ , which will be finally identified with spin. In the process of propagation of the particle the spin rotates due to spin-orbit interaction, but the total angular momentum is conserved. The classical vector  $e_a$  on the lattice does not correspond to any definite value of the spin but, according to (10.66) and (10.67) only the spin 1/2 part of the wave function has long range correlations in the critical region. It is not surprising therefore that as a result we obtained the Dirac equation.

Turning now to the 3d case we discovered that the Ising model is described by the propagation of a closed string with internal degrees of freedom distributed on the links. These degrees of freedom are precisely the same as in the previous case, so we can say that we have a spin density distributed along the string. Our major result was the conclusion that the string moves piecewise as a 2d-Ising particle, and there is only implicit interaction following from continuity.

The most difficult problem now is to find a continuum limit for the equation (10.78). In the 2d case such a problem was solved trivially, by solving first the lattice equation (10.60) and taking the limit  $\beta \rightarrow \beta_{cr}$ .

Unfortunately the equation (10.78) on the lattice is completely hopeless. The best thing we can do is to guess on physical grounds what kind of system it describes in the critical region. We shall do this by looking at things the other way around. Namely we discuss a continuum string model, which behave piecewise as a free Dirac particle, and has a very good chance to describe the critical region of equation (10.78).

Let us consider the NSR-string, the wave functional of which is annihilated by the supercurrent and energy momentum tensor. The supercurrent condition can be presented as:

$$\psi_{\mu}^{(\pm)}(\dot{x}_{\mu} \pm x'_{\mu})|\phi\rangle = 0$$

If we replace  $\dot{x}_{\mu}$  by  $(1/i)(\delta/\delta x_{\mu})$  and introduce:

$$\gamma_{\mu}^{(1,2)} = \psi_{\mu}^{(+)} \pm \psi_{\mu}^{(-)}$$

we obtain an equation:

$$\left( \frac{1}{i} \gamma_\mu^{(1)}(s) \frac{\delta\phi}{\delta x_\mu(s)} + \gamma_\mu^{(2)}(s) \frac{dx_\mu}{ds} \right) \phi = 0 \quad (10.79)$$

where

$$\{\gamma_\mu^{(a)}(s), \gamma_\nu^{(b)}(s')\} = 2\delta^{ab}\delta_{\mu\nu}\delta(s-s')$$

Now, if we consider a very short piece of the string for which  $|dx/ds|\Delta s = \epsilon$ , then we can replace  $\Delta s(\delta\phi/\delta x_\mu(s))$  by an ordinary derivative:

$$\Delta s \frac{\delta\phi}{\delta x_\mu(s)} \Rightarrow \frac{\partial\phi}{\partial x_\mu}$$

and  $\gamma_\mu^{(1)}(s)$  by an ordinary  $\gamma$ -matrix. The term

$$\gamma_\mu^{(2)}(s) \frac{dx_\mu}{ds} \Delta s$$

anticommutes with

$$\gamma_\mu^{(1)} \frac{\delta}{\delta x_\mu(s)}$$

and in the "short string" limit can be replaced by  $M\gamma_5$  with some  $M$ . As a result, the above equation becomes an ordinary Dirac equation

$$\left( \frac{1}{i} \gamma_\mu \frac{\partial}{\partial x_\mu} + M\gamma_5 \right) \phi = 0 \quad (10.80)$$

The same conclusion could have been reached by the mode expansion in the Ramond sector and by noticing that in the short string limit only zero modes are relevant (since other eigenvalues tend to infinity). We conclude, that the NSR string moves piecewise as a collection of Dirac particles connected only by continuity. This is the same picture which we derived for the 3D Ising model. So, perhaps these two strings coincide.

Needless to say, we have not proved it. But the intuitive arguments given above make very tempting the problem of finding the critical exponents of the NSR string, and comparing them with Ising ones.

This problem has not been solved yet. In the next section we shall describe a general approach to it, together with a preliminary classification of strings.

### 10.4 Extrinsic Geometry of Strings

In this section we shall discuss critical exponents of strings—a subject intimately connected with their extrinsic geometry. Again, this problem is not solved and we can only show an approach to it, and describe several options, which we have in string theory.

The most interesting critical exponent is defined as follows. We start from the Nambu-Goto action (or its fermionic extension):

$$\begin{aligned} \mu_0 A(S) &= \mu_0 \int g^{1/2} d^2\xi \\ g_{ab} &= \partial_a \mathbf{x} \cdot \partial_b \mathbf{x} \end{aligned} \tag{10.81}$$

and try to choose the bare surface tension  $\mu_0$  in such a way that the physical surface tension  $\mu$  can be defined by:

$$G(C) = \sum_{S_C} \exp\left(-\mu_0 \int g^{1/2} d^2\xi\right) \underset{C \rightarrow \infty}{\approx} \exp(-\mu A_{\min}(C)) \tag{10.82}$$

(here  $C$  is a boundary loop and  $A_{\min}(C)$  is an area of the minimal surface bounded by this loop). The quantity  $\mu(\mu_0)$  is by definition the physical surface tension. At the critical value of  $\mu_0$ ,  $\mu_{0cr}$  we expect that:

$$\mu \sim (\mu_0 - \mu_{0cr})^\alpha$$

where  $\alpha$  is the critical exponent to be determined. One can relate many interesting quantities to this exponent.

What determines  $\alpha$ ? The answer is different for bosonic and fermionic strings. Let us begin with the former. We have to ask ourselves at first, whether the Nambu term in the action, which is a kind of a cosmological term is the only relevant one in the continuum limit. One would expect that the Einstein term can be important as well, since the Newton constant is dimensionless in two dimensions. It is commonly known, however, that in this case the Einstein term is just the Euler character of the manifold which is presumed to be fixed. Nevertheless, for string theory a dimensionless term in the action does exist. It is formed out of the extrinsic curvature of the surface and is defined as follows. Let us introduce the second fundamental form  $K_{ab}^i$  given by the equation:

$$\begin{aligned} \partial_a \partial_b \mathbf{x} &= \Gamma_{ab}^c \partial_c \mathbf{x} + K_{ab}^i \mathbf{n}_i \\ (\mathbf{n}_i \cdot \mathbf{n}_j) &= \delta_{ij}; (\mathbf{n}_i \cdot \partial_a \mathbf{x}) = 0 \\ i &= 1, \dots, \mathcal{D} - 2 \end{aligned} \tag{10.83}$$

(here  $\Gamma_{ab}^c$  is the usual Christoffel symbol,  $n_i$  normal vectors). Then the intrinsic curvature  $R$  is connected with  $K$  by

$$R = (K^{ia})^2 - K_b^{ia} K_a^{ib} \quad (10.84)$$

and is indeed a total divergence. However the separate terms in (10.84) are not. Therefore we can write the following generalization of the Nambu-Goto action:

$$S = \mu_0 \int \xi g^{1/2} d^2 + \frac{1}{\alpha_0} \int K_b^{ia} K_a^{ib} g^{1/2} d^2 \xi \quad (10.85)$$

It is easy to check that the second term is the only possible one (up to total divergences) invariant under scale transformation  $x \rightarrow \lambda x$ .

Adding this new term to the action is not a caprice. Its influence in the infrared region determines the phase structure in the string theory. So, if we want to compute the critical behaviour of random surfaces and their geometrical and physical characteristics, it is absolutely necessary to include this term in the action.

Our first goal will be to investigate the relevance of extrinsic curvature in the continuum limit. Let us notice, that (10.85) can be rewritten in other forms (modulo total divergences):

$$\begin{aligned} & \int K_b^{ia} K_a^{ib} g^{1/2} d^2 \xi \\ &= \int g^{1/2} g^{ab} \partial_a t_{\mu\nu} \partial_b t_{\mu\nu} d^2 \xi \\ &= \int g^{1/2} (\Delta(g)x)^2 d^2 \xi \\ &= \int g^{1/2} g^{ab} (\nabla_a n_i) \cdot (\nabla_b n_i) d^2 \xi \end{aligned} \quad (10.86)$$

Here:

$$\begin{aligned} \Delta(g)x &= -\frac{g^{-1/2}}{g} \partial_a g^{1/2} g^{ab} \partial_b x \\ t_{\mu\nu} &= \frac{g^{-1/2}}{g} \epsilon^{ab} \partial_a x_\mu \partial_b x_\nu \\ \nabla_a n_i &= \partial_a n_i + \mathcal{A}_a^{ik} n_k = -K_a^{ib} \partial_b x \end{aligned}$$

The last form is especially interesting. It shows that we are dealing with a Grassman  $\sigma$ -model, associated with the surface. The order parameter lies in the homogeneous space

$$G_{2,\mathcal{D}} = \frac{SO(\mathcal{D})}{SO(2) \otimes SO(\mathcal{D} - 2)} \tag{10.87}$$

This is not an ordinary  $\sigma$ -model, because not every field in  $G_{2,\mathcal{D}}$  forms tangent planes to some surface. A certain integrability condition must be satisfied. Still, the analogy with the  $\sigma$ -model will be quite useful. Namely, it permits one to compute the  $\beta$ -function for the coupling  $\alpha$  which determines its scale dependence. We shall not describe this calculation here, but rather discuss its result and implications.

One finds for the momentum dependence of  $\alpha(p)$ :

$$\alpha(p) = \frac{\alpha_0}{1 - \frac{\mathcal{D}}{2} \cdot \frac{\alpha_0}{2\pi} \cdot \log \frac{\Lambda}{p}} \tag{10.88}$$

This formula makes clear, first of all, that our Grassmanian  $\sigma$ -model is not an ordinary one. In the latter case the coefficient before the logarithm would be  $D - 2$  (recall the  $n$ -field) instead of  $D/2$ . This difference comes from the integrability constraint on the Grassman fields—they have to form tangent planes to some surface.

Of course, the behaviour described by (10.88) is true only until  $\alpha(p)$  becomes large. What happens then? There are several possibilities. First of all, if the  $\beta$ -function has no zeros, then  $\alpha(p)$  continues to increase as we go to the infrared region. That means that the term  $(1/\alpha) \int K^2 g^{1/2} d^2\xi$  becomes irrelevant, since  $\alpha \rightarrow \infty$ . To describe the same thing in a different language, let us introduce the Lagrange multiplier:

$$S = \frac{1}{\alpha} \int K^2 g^{1/2} d^2\xi + \int (\lambda^{ab} (\partial_a x \cdot \partial_b x - g_{ab}) + \mu_0 g^{1/2}) d^2\xi$$

We have seen in the previous chapters that asymptotic freedom in the  $\sigma$ -models leads to the condensation of the Lagrange multiplier, or, which is the same, to creation of mass for the  $n$ -fields. In our case this means the following. The effective action for  $\lambda^{ab}$  develops a minimum so that

$$\langle \lambda^{ab} \rangle = \bar{\lambda} g^{1/2} g^{ab} \tag{10.89}$$

where dependence of  $g_{ab}$  is fixed by general covariance, while the value of  $\bar{\lambda}$  is just the position of the logarithmic pole in (10.88) (since  $\bar{\lambda}$  represents an infrared cut-off):

$$\bar{\lambda} \sim \Lambda e^{-4\pi/\mathcal{G}\alpha_0} \quad (10.90)$$

If we do not have a special fine tuning, then  $\alpha_0 \sim 1$  and  $\bar{\lambda} \sim \Lambda$ , so in the continuum limit we have the effective action

$$S = \bar{\lambda} \int (g^{1/2} g^{ab} \partial_a x \cdot \partial_b x + \mu_0 g^{1/2}) d^2 \xi \quad (10.91)$$

which we discussed in the previous chapters.

We see that in this case the critical exponent  $\alpha = 0$ , because the change of  $\mu_0$  does not have any influence on  $\bar{\lambda}$ . An attempt to make a small surface tension in this case will be ruined by the violent infrared fluctuations described by (10.91). From the geometrical point of view, the Lagrange multiplier  $\bar{\lambda}$  plays the role of the inverse correlation length for the normals of the surface. In the regime described this correlation length is of the order of the cut-off. The surface is terribly creased. Perhaps the bosonic tachyon is related to this creasing.

For QCD and for Ising models, creased strings with nonvanishing surface tension are unacceptable. How can this undesirable property be avoided?

It is clear, that in the purely bosonic case we have to find a version of the theory with the  $\beta$ -function having a zero at some point  $\alpha_*$ . If we succeed, then a generic  $\alpha_0$  will be attracted to  $\alpha_*$ , the correlation length will be infinite (without a Nambu term in the action) and we will have a scale invariant theory with anomalous dimensions, one of which will determine the critical exponent. The creasing will be avoided in this case.

In four-dimensions there is a good candidate for all that. In this case we have a specific  $\theta$ -term which can be added to the action. At  $\theta = \pi$  there are reasons to expect a scale-invariant theory. The term we are talking about is the algebraic number of self-intersections,  $v(S)$  for our two-dimensional surface  $S$ , immersed into four-dimensional space. Analytically, it is given by:

$$v(S) = \frac{1}{4\pi} \int d^2 \xi g^{1/2} g^{ab} \epsilon^{\mu\nu\lambda\rho} \partial_a t_{\mu\nu} \partial_b t_{\lambda\rho} \quad (10.92)$$

The partition function is given by:

$$Z = \sum_{(S)} e^{-\mu_0 A(S)} (-1)^{v(S)} \quad (10.93)$$

Perhaps, this string lies in the same universality class as the string of large  $N$  QCD. This is a problem for the future.

In the case of Fermi strings dependence on extrinsic geometry comes from the fermionic contributions on the world sheet which are perhaps more relevant than the  $K^2$  term. If so, the calculation of critical exponents will be easier in this case. This is also a problem for the future.



**Taylor & Francis**

Taylor & Francis Group

<http://taylorandfrancis.com>

---

# Index

---

- $A_x$ -quanta scattering 18
- abelian case 16
- abelian gauge theories 62
- abelian systems 49, 80
- abelian theory 70
- $(a,b)$ -system 236
- accessible conformal gauge 184
- action saddle point 175
- action variation 120, 204
- adjacent disorder variable 276
- adjoint representation 145
- $a$ -independent condensate 138
- algebraic self-intersection number 286
- almost harmonic oscillations 49
- almost local function 138
- almost parallel spins 54
- amplitude poles 200
- analytic coordinate changes 189
- anharmonic terms 55
- anomalous boundary conditions 243
- anomalous dimension 133, 197, 199, 250, 286
  - matrix 263
- anomaly relation 102
- anticommuting fields 238, 243
- anticommuting variables 106
- antifermions 103
- antihermitian matrices 144
- anti-instantons 87, 90
- antiparticles 31
- antiperiodic boundary conditions 249
- antiquark 145
- antisymmetric tensor fields 266-7
- area law 78
- asymptotic freedom 19, 24, 32, 111, 118, 119, 145, 147, 285
- Atiyah-Singer index theorem 105
- average energy density 278
- average fluctuations 142
- axial current 100
  - divergence 102
- axion 109
  
- $B$ -modes 267
- barbed wire 274, 279
- bare charge 22
- bare coupling 33, 87, 144
- bare propagator 8
- Berezin rule 106
- Bianchi identity 95, 97
- binding force 78, 107
- block spin 10, 134
- Boltzman formula 66
- Bose gas 61
- Bose liquid 82
- bosonic emission vertices 250
- bosonic ground state 244
- bosonic sector 249, 250
- bosonic string 222, 266, 283
- bosonic tachyon 242, 286
- bosonization formulas 90
- bosons 234
- boundary condition 81, 85, 171, 183, 243, 251
- boundary loop 283
- broken symmetry 12, 49, 129
  - phase 9, 76, 84

- Brownian path 225
- $\beta$  expansion 34, 41
- $\beta$ -field 130, 173
- $\beta$ -fluctuations 174
- $\beta$ -function 24, 254, 259, 261, 265, 267, 285, 286
- $\beta$ -limit 19, 38, 54–5
  
- CP-symmetry 269
- C-term 207
- cancellation of residues 202
- cancellation of singularities 79
- canonical directions 183
- Cauchy inequality 220
- Cauchy–Riemann equations 87, 204
- central charge 238–9, 263, 265
- characteristic intervals 180
- characteristic momentum 13
- characteristic time 51
- characteristic wavelength 18
- charge 64, 78
  - confinement 43, 75
  - density 66
  - quantization 63
  - renormalization 32, 132
- charged fields 63
- charged particles 30
- chiral equations of motion 119
- chiral fermions 268, 269, 273
- chiral field 19, 40, 111, 117, 134, 137–8
- chiral models 91
- chiral symmetry 107
- chiral theories 123
- Christoffel symbol 284
- classical action 21, 50, 85, 167
- classical equation 65, 114, 167
- classical field 88, 148
- classical solution 69
- classical vector 281
- closed loop 70, 112, 224, 274, 276, 279
- closed nonlinear loop space
  - equations 148
- closed path 34, 164, 166, 275
- closed string 123, 147, 198, 213, 242, 273, 281
- closed surface 40, 279
- collective coordinates 89
- color-electric strings 146
- color singlets 146
- colour confinement 47
- colour spin 73, 77
- combinatorics 219
- commutation relations 43, 98
- commutation rules 202
- commutator 207, 222
- commuting fields 238
- compactification 266
- compact Lie group 14, 17
- compact QED 62, 63, 71
- complex field 61
- complex notation 88
- complex plane cut 57
- compulsory pair-creation 104
- compulsory particle production 103
- condensation 254, 285
- configuration space 104
- confinement of charge 69
- confining phase 78, 81, 82, 147
- confining property 146
- conformal anomaly 185, 188
- conformal field theory 199, 259, 263, 268
- conformal gauge 177, 178, 182, 183, 185, 188, 230, 232, 264, 271
- conformal group 206, 215
- conformal invariance 185, 197, 210, 235, 254, 262, 263
- conformally invariant operator algebra 209
- conformal map 182, 245
- conformal quantum field theory 197, 203
- conformal spin 202, 229, 234
- conformal symmetry 213, 235
- conformal transformation 207, 210, 244, 272
- conformal Ward identity 203, 247, 248
- conformal weight 233
- conjugate momenta 243
- conjugate operator 232
- conjugate spin 233
- conserved current 12, 223
- conserved quantities 229
- conserved quantum numbers 41
- constant curvatures 255
- constrained system 68
- contact term 262, 278
- continuous complex field 11
- continuous phase rotations 61

- continuous symmetries 13
- continuum analogue 278
- continuum field 10
- continuum limit 9, 15, 16, 17, 27, 44, 47, 57, 62, 115, 125, 128, 154, 163, 189, 274, 277, 278, 281, 283, 284, 286
- continuum notation 69, 80
- continuum string model 281
- continuum theory 16, 57, 134, 152, 163
- contour variable 274
- Cooper pairs 70
- coordinate transformation 157, 177
- correlation function 6, 24, 28, 33, 37, 52, 53, 57, 66, 79, 118, 132, 134, 146, 193, 196, 199, 204, 206, 207, 248, 249, 275
- correlation length 7, 15, 82, 143, 174, 176, 286
- coset space 29, 91, 119, 139, 153, 171
- cosmological constant 258
- cosmological term 223, 283
- Coulomb 56-7, 65, 69, 72, 79, 82, 89, 109, 142
- coupling constant 108, 125, 145, 239, 255
- covariance 173, 176, 270, 286
- covariant derivative 112, 222, 233, 236
- covariant divergence 44
- covariantly regularized integral 175
- covariant regularization 273
- creased strings 286
- critical behavior 273
- critical coupling 69
- critical dimension 212, 241, 253, 265, 269, 270
- critical exponent 128, 282-3, 286
- critical singularity 278
- critical strings 212, 266
- crystals 83
- current algebra 270
- curvature 112, 256
- curved background 254
- curved compact manifold 266
- cut-off 100, 173, 197, 260, 286
- $\chi$ -dependence 234
- $\chi$ -field 67, 226, 230
- $\chi$ -variation 230
- $\mathcal{D}$ -dimensional space 254
- $\mathcal{D}$ -dimensional supersymmetry 227
- $\mathcal{D}$ -1-dimensional lattice 38, 41, 43
- Debye approximation 66
- Debye screening 70, 90
- definite regularization 149
- degeneracy 39, 40, 53, 210-2, 218
- degenerate conformal families 212
- diagrammatic expansion 34
- diagram technique 79
- diamagnetic part 31
- dielectric constant 80, 81, 83
- dielectric permeability of the vacuum 73
- diffeomorphism 152, 157, 158, 159, 170, 184, 270
- differential geometry 114
- diffusion law 180
- dilation 204, 254
- dilaton  $\phi(x)$  265
- dilatons 266
- dimensional regularization 260, 262
- dimensional transmutation 23, 39, 125
- dimensionless term 283
- dipole-dipole interaction 87, 90
- dipole dissociation 90
- Dirac eigenvalue problem 234
- Dirac equation 102, 274, 275, 277, 278, 281-2
- Dirac fermions 99, 234
- Dirac matrices 249
- Dirac operator 89, 100, 105, 273
- Dirac particle 222, 224, 228, 281-2
- Dirichlet boundary conditions 20
- Dirichlet problem 183
- disguised free theory 138
- dislocation 62, 274
- disordering effect 65
- disorder variables 274, 275, 280
- distorted partition function 275, 279
- divergent factor 227
- double invariant Killing metrics 159
- double line representation 145, 148
- double-valued fields 243, 248
- dual amplitudes 265
- duality equation 87, 96, 118
- dual lattice 275-6, 279
- dual temperature 279
- dynamical generation 139
- dynamics of loops 118
- Dyson equation 8
- Dyson ordering operation 112

- $\delta$ -functions 278
- $\delta$ -like singularities 65
  
- effective action 19, 22, 31, 81, 107, 129, 134, 137, 141, 144, 156, 225, 235, 243, 257, 258, 259, 261, 265, 267, 268, 270, 286
- effective Lagrangian 4, 17, 107
- effective nonlinearity 66, 130
- eigenfrequencies 52
- eigenfunction 161
- eigenmodes 105, 161
- eigenstates 36, 73, 74
- eigenvalue 39, 234, 282
- einbein field 223
- Einstein equations 267
- elasticity 61, 73
- electrically charged fields 70
- electric field operator 43
- electric flux 42, 63, 147
- electric string 79
- electromagnetic field 142
  - strength 67
- electromagnetic interactions
  - unification 71
- elementary excitation 33–4, 36, 39, 123, 132
- energy–momentum conservation law 206
- energy–momentum tensor 176, 203, 206, 223, 229, 236, 247, 263, 281
- entropy 143
- Euclidean model 36
- Euclidean space 108, 151
- Euler angles 40
- Euler character 146, 180, 283
- excitation hopping 35, 38
- excitation spectrum 39
- excited states 42, 244
- expectation values 166
- explicitly covariant form 255
- exponential damping 143
- exterior derivative 95
- external field 265
  - gravitational 180, 236
  - magnetic 30, 70
  - Non-Abelian gauge 99
  - triplet vector 84
- extra conformal theory 259
- extrinsic curvature 283
- extrinsic geometry dependence 287
- extrinsic string geometry 283
- $\langle \epsilon\epsilon \rangle$  correlation function 147
  
- Faddeev–Popov equality 231
- Faraday flux lines 42
- fast components 255
- fast field 255
- fast fluctuations 27
- fermi-fields 229
- Fermi-momentum 102
- fermion boson scattering 250
- fermionic string 221, 228, 242, 250, 266, 269, 279, 283,
- fermions 90, 102, 103, 232, 234, 241, 245, 247, 268
- Fermi particle 222, 274
  - propagator 274
- Fermi-partner 228
- ferromagnetic phase 7
- Feynman diagram 21, 121, 123, 136, 142, 168
- Feynman gauge 30
- Feynman graph 126, 131, 145
- Feynman–Schwinger parametrization
  - supersymmetric extension 225
- $f$ -field 239
- $f$ -fluctuations 174
- field strength 14, 55, 70, 116
- finite angular excitations 132
- finite correlation length 65, 268
- finite dimensional groups 159
- finite gap 36
- fixed coupling 236
- fixed metric tensor 254
- flat  $y$ -space 267
- fluctuations 25, 88, 137, 143–4, 175, 249, 286
- flux sources 37
- Fock space vector 202
- four-dimensional gravity 268
- Fourier expansion 128
- free bosonic field 195
- free energy 31, 73
- free field 138, 204, 206, 212
- free particle 63, 137, 224
  - propagation 281
- free photons 17, 71
- free string 123, 148, 151, 281

- functional integral 62, 125, 205, 227, 231, 234, 243, 245, 250, 251
  - computation 185
- functional space 51, 157, 160, 172
- gapless excitations 12, 102
- gapless photon 81
- gauge anomaly 270, 273
- gauge bosons 71
- gauge field 70, 73, 99, 107, 111, 118, 139, 145, 158, 229, 253
  - topology 85
- gauge fixing 197, 224
- gauge group 5, 14, 112
  - generator 109
  - invariance 139
- gauge invariance 17, 29, 40, 63, 140, 216, 281
  - excitation 45
  - Green functions 32
  - quantity 15, 66, 119
  - states (vacuum sector) 75
- gauge symmetries 14, 40
- gauge system 40, 49, 77, 82
- gauge theory 80, 104
  - cut-off 121
- gauge transformation 44, 45, 46, 57, 112, 152, 157, 171, 216
- Gaussian integral 66, 88, 135, 154
  - solution 167
- Gell-Mann-Low equation 24, 25, 260
- general operator 201, 234
- general relativity 5
- general supercovariance 230
- generators 268
- geometrical integrals 169
- Georgi-Glashow model 71, 80
- ghost fields 231, 250
- ghost Lagrangian 236
- ghost operators 234
- ghosts 259, 265, 270
- Ginzburg-Landau equations 70
- globally invariant action 223
- global  $O(2)$  model 42, 54
- global  $O(2)$  symmetry 11
- global symmetry 5, 11, 27, 34, 37, 40, 84
- gluon 30, 47, 76, 144, 146
- Goldstone boson 13, 39, 107
- Goldstone field 16
- Goldstone particles 109, 129
- Goldstone's theorem 12
- grand unification 253
- Grassman fields 285
- gravitational anomaly 270, 273
- gravitational field 5
- gravitino 229, 236, 239, 269
  - field 223, 234, 238, 250
- graviton 203, 229, 254, 266, 271
  - condensation 266
- Green function 8, 67, 100, 126, 132, 134, 138, 142, 167, 175, 179, 184, 194, 240, 260, 275, 278, 279
- ground state 15, 35, 38, 49, 53, 70, 242
  - wave function 9
- $\gamma$ -function 25
- $\gamma$ -matrix 223, 282
- H*-field 67
- Haar measure 14
- hadrons 99
- half-integer charges 80
- half-integer spin 47, 76
- Hamiltonian 39, 44, 77
  - formalism 208, 243
  - formulation 34
  - language 41
  - version 36, 38
- harmonic approximation 54
- harmonic oscillator 202, 214, 244
- heavy fields 71
- hermitian conjugation 115
- hermitian matrix 135
- heterotic fermionic string 269
- heterotic Lagrangian 269
- heterotic string theories 273
- Hilbert space 35
- homogeneous space 285
- homologically nontrivial path 252
- homotopy group 92
- Hopf's bundle 140
- hypercube 10
- hypervolume 170
- imaginary time 38, 49
- inaccessible conformal gauge 179
- independent tops 44
- index theorem 179
- induced action 173, 174, 185, 188

- infinite dimensional conformal group 204
- infinite dimensional measures 161
- infinitely heavy charges 45
- infinitely short strings 151
- infinite path 280
- infinitesimal field response 84
- infinitesimal transformation 184, 205
- infinite volume 129
- infrared divergence 98, 126, 143
- infrared limit 61, 65, 69, 71
- infrared region 32, 107, 284, 285
- infrared singular terms 90
- infrared-strong interaction 13
- inhomogeneous transformation 210
- injection points 249
- inner space supersymmetry 226
- instanton 9, 16, 17, 49, 53, 54, 58, 64, 65, 69, 70, 78, 79, 80, 87, 88, 90, 99, 109, 111, 118, 139, 143, 144
- instanton-anti-instanton configuration 91, 107
- integer mode shift 249
- integrability condition 285
- integrability constraint 285
- integration 184, 271
- interacting massless modes 265
- interaction energy 70
- internal degrees of freedom 281
- internal index 281
- intrinsic curvature 284
- invariance 59, 146, 186, 270
  - property 16-7
- invariant measure 157, 171
- inverse correlation length 26, 127, 286
- inverse cut-off 176
- inverse Laplacian 56, 58, 81
- inverse temperature 275
- Ising chain 6
- Ising mode 34
- Ising model 5, 15, 34, 49, 245, 253, 273-5, 278, 279, 282, 286
- Ising spins 251
- isometry 166
- isospinor Dirac field 99
- isotopic factor 31
- isotopic rotations 97
- isotopic spin 39, 45
- isotopic structure 136
- isotopic vector fields 71
- Jordan-Wigner transformation 246
- Kac formula 212, 218
- Kähler structure 268
- Kaluza-Klein approach 268, 269
- kernel 190
- $k$ -ground state particles 198
- Killing deformations 268
- kinematical power counting 146
- Koba-Nielsen amplitude 213, 221
- Koba-Nielsen integrals 261
- Kramers-Wannier duality 15, 251, 280
- $k$ -space singularities 186
- Lagrange multiplier 125, 134, 153, 164, 225, 254, 285, 286
  - fluctuations 131
  - freezing 173
  - representation 172
- Lagrangian 19, 38, 39, 45, 59, 63, 102, 109, 131, 139-40
- Lamé constants 61
- Landau diamagnetism 30
- Laplace operator 170, 194
- large fluctuation quenching 107
- large  $N$  expansion 125, 131, 139
- large  $N$  limit 121, 123, 129, 134, 137, 138, 140, 142, 148
- lattice 34, 64, 69, 112, 272, 276, 281
  - closed loop 42
  - cut-off 225
  - derivatives 68
  - divergence 55
  - equations 117
  - field theory 80, 125
  - graph 37
  - Laplace equation 55
  - Laplace operator 64
  - models 62
  - spacing 10, 22, 161
  - vector 61
- left rotations 40
- left-supersymmetric field theory 270
- Lie algebra 21, 63, 76, 112, 114, 207
  - generators 120
- light-cone gauge 250
- light isoscalar boson 109

- linear confinement 80
- linear equations 276
- linearized approximation 188
- linearized scalar curvature 188
- link 34, 43, 112, 279, 281
- Liouville action 195, 236, 238, 264
- Liouville field 200, 265, 271
  - decoupling 220, 254
  - theory 193, 197
- Liouville functional integral 193
- local derivative 119
- local invariant expression 180
- local supersymmetry 229
- local symmetry transformations 5
- logarithmic correction 255, 257
- logarithmic divergence 257, 260, 261
- logarithmic pole 286
- longitudinal vectors 271
- long-range correlations 90, 281
- long wave oscillations 253
- loop dynamics 111
- loop equation 123, 275, 278
- loop expansion 265
- loop factor 121
- loop point fixing 164
- loop space 111, 117, 118, 119
- Lorentz rotations 5
- low energy expansion 259, 265
- low energy limit 22, 253, 259
- lower secondary fields 210
- $\lambda$ -field 126, 131
  
- M*-space 229
- magnetic charge 65, 71
- magnetic dipole 98
- magnetic flux 65, 69
- magnetic line condensation 69
- magnetic loops 69
- magnetic monopole 109
  - solution 69
- magnetic pole 80
- Majorana spinors 230
- many-dimensional systems 111
- mapping 91
- mass creation 285
- mass gap 34, 38, 47, 49, 81, 90, 99
- massive Dirac field 91
- massive photon 65
  - massive states exchange 262
  - massive string modes 253
- massless electromagnetic field 71
- massless excitations 82, 253
- massless fermion pair exchange 107
- massless free photons 69
- massless Goldstone modes 16
- massless ground state 266
- massless modes 254, 265, 267
- massless particles 16, 19, 129, 259
- massless photon 67, 81
- massless pole residues 84
- massless quarks 106
- massless sector 258, 271
- massless string excitations 253
- massless tensor particle 242
- massless tensor state 203
- massless vector 242, 268
- mass matrix 263
- mass renormalization 133
- mass shell 241
- mass spectrum 263
- matter fields 253, 268
- melting 62, 84
- meromorphic function 207
- metric field integration 164
- metric tensor 153, 157, 161, 180, 255, 268
- Minkowskian action 108
- Minkowskian theory 20
- Minkowskian time 38
- Minkowski space 31, 103, 208, 229
- mixed components ( $m, \mu$ ) 268
- mode decomposition 250
- mode expansion 248–9, 282
- modified partition function 276
- modular anomaly 273
- modular group 270, 271
- modular invariance 271, 272
- moment of inertia 39
- momentum 41, 186, 261
  - conservation 167, 195, 196
  - cut-off 21, 127
  - dependence 22
  - representation 12, 167
- monopole 66
  - condensate 70
- monopole–antimonopole pairs 70
- multi-instanton configuration 143
- multi-instanton picture 98

- multi-instanton solution 88, 97
- multi-kink configurations 52
- multiloop corrections 144
- multiple logarithmic divergences 262
- multivalued fields 58
- muon 273
- $\mu$ -correlator 277
- $\mu$ -variables 278
  
- NRS-string 245
- NS-excitations 251
- NSR-Lagrangian 269
- NSR-string 269, 281-2
- Nambu action 254
- Nambu-Goto action 283, 284
- Nambu term 283, 286
- narrow resonances 146
- natural charge unit 63
- natural scale  $N_2$  144
- $n$ -dimensional Riemannian manifold 157
- negative norm components 220
- negative norm state 203
  - decoupling 213
- negligible fluctuations 148
- neighbouring spins 54
- neutral dipoles 57
- neutral objects 142
- neutral operator 142
- neutral pairs 142
- neutral vortex anti-vortex
  - superposition 61
- Neveu-Schwarz Hamiltonian 251
- Neveu-Schwarz-Ramond string 222, 275
- Neveu-Schwarz sector 245, 246, 250
- Newton constant 283
- new world sheet scalar 229
- $n$ -field 27, 39, 84, 91, 92, 99, 109, 111, 131, 134, 137, 140, 256, 285
  - propagator 12
- $n$ -flux 37
- no ghost theorem 220
- Non-Abelian case 14, 39, 44, 84
- Non-Abelian electric field 44
- Non-Abelian gauge field 29, 47
- Non-Abelian gauge invariance 84
- Non-Abelian gauge theory 17, 32, 71, 92, 104, 109, 144
  - subgroup 64
- Non-Abelian global symmetries 13
- Non-Abelian instanton 85, 98
- Non-Abelian models 14
- Non-Abelian phase factor 112
- Non-Abelian systems 68
- non-chiral determinants 273
- noncovariant Green-Schwarz action 250
- noninteger topological charge 109
- nonlinear correction 136
- nonlinear Debye equation 78
- nonorientable surfaces 267
- nonperturbative effects 16
- nonperturbative methods 33
- nonplanar diagram 137, 145
- nonrelativistic Fermi gas 102
- nonsingular solutions 65
- nonsingular transformation 177
- nontrivial spinor structures 243, 245
- nonvanishing surface tension 286
- nonzero circulation 57
- nonzero condensates 265
- nonzero field strength 117
- nonzero temperature 61
- normalization 144, 247, 259
- normal mode 51
  - expansion 106
- normal phase 82
  
- odd state 242
- one dimensional supergroups 228
- one-instanton contribution 143
- one kink contribution 52
- one-loop calculation 143
- one loop corrections 21
- one loop divergences 257
- one-loop WKB approximation 144
- one-particle state 147
- Onsager's specific heat 278
- on-shell amplitude 20, 193
- open string 213, 241
- operator 202
  - algebra 7, 259, 261
  - of higher spin 80
- operator product 261
  - expansion 195, 199, 200, 261
  - formalism 214
- orbital motion 30
- orbital properties 45

- order parameter 15, 274, 276, 285
- ordinary space 117
- orthonormality conditions 75
- oscillating factors 226
- oscillator representation 215
  
- pair correlation function 263
- parallel transport 112
- paramagnetic effect 32
- paramagnetic phase 7
- parametrization space 166
- parametrized group 44
- parity 242
  - considerations 267
- particle 16, 31, 39, 151
  - condensate 36
  - creation amplitude 106
  - density 66
  - excitation propagation 281
  - motion 180
  - physics 5
  - state 212
- particle-hole excitation 102
- partition function 11, 14, 37, 40, 54, 55,
  - 61, 64, 69, 70, 73, 75, 81, 89, 98, 100,
  - 134, 141, 143, 221, 231, 239, 245, 251,
  - 257, 260, 272, 275, 279, 286
- Pauli matrices 35, 46, 93, 140, 230
- Pauli paramagnetism 30
- Pauli-Villars regularization 235
- Pauli-Villars regulators 193
- perturbation 265
  - induced field condensations 263
  - theory 23, 36, 49, 121, 130, 138, 235,
  - 260
- perturbative interaction 16, 17
- perturbed Lagrangian 259
- $p$ -form 94
- phase factor 78, 80, 114, 117, 148
  - variation 120
- phase structure 15, 109, 284
- phase transition 7, 12, 13, 16, 34, 39, 82,
  - 134, 212, 273
  - point 36, 274, 277
  - point fluctuations 13
- photon emission vertex 142
- photon phase 69
- photon polarization 213
- photon propagator 142
- photon vertex 213
- physical states 212, 221
- physical surface tension 283
- planar diagrams 148
  - summation 137, 145
- planar graphs 137
- planar  $O(2)$  magnet 61
- Planck length 10, 108
- Planck mass 253
- plaquette 14, 40, 55, 64, 70, 114, 279
- plasma 57, 66, 90, 109
- point-like excitation 111
  - scattering 119
- point-like solution 69
- point singularities 245
- Poisson summation formula 162
- polarization 215
  - operator 31, 84, 190
- pole 147, 193
  - residue 242, 261
  - singularity 200
  - term 83, 262
- positive norm 207
  - Hilbert state 220
- power divergences 131-2
- powerlike singularity 12
- primary fields 221
- primary operator 206, 213
- principal bundle 19
- propagator 145, 186, 226, 265
  - of the  $n$ -field 126
- pseudoparticles 68, 79
- pseudospin inhabited string 274
- punctured sphere 243, 244
- $\pi/2$ -rotations 277
- $\phi$ -dependent terms 258
- $\phi$ -field 258
- $\phi(2)$  gauge systems 15
- $\phi$ -dependence 185, 235
- $\phi$ -field 54, 228
- $\phi$ -independent nonlocal terms 188
- $\phi$ -integration 184
- $\phi$ -propagator 220
- $\psi$ -fields 226, 228, 269
- $\Psi$ -functional 20
- $\psi$ -reflection 242
- $\psi$ -variable 274

- $q$ -map 86
- $q$ -space singularities 174
- quadratically divergent constant 187
- quadratic approximation 141, 185
- quadratic divergence 60, 260
- quantized color 114
- quantized flux 43, 70
- quantized Hall effect 109
- quantum effects 90, 142
- quantum electrodynamics 5
- quantum field 19, 45
- quantum fluctuations 20, 87, 88
- quantum gravity 273
- quantum mechanics 49, 61, 63, 131
- quantum state 42, 199
- quantum theory 42, 53, 185
  - of loops 119
- quark 145
  - confinement 73
- quark-anti-quark pair 78
- quasimomentum 36, 39, 43
- quenching factor 101
  
- radial excitations 131
- radiative correction 13
- Ramond sector 244, 245-6, 250, 251, 282
- random fields 68
- random surfaces 151, 169, 265
- regularization 71, 154, 185
- renormalization 9, 100, 134, 160, 257, 262
  - equation 258
  - group 19, 25, 27, 258, 260, 263, 264, 267, 268
- renormalized coupling constant 21
- reparametrization 152
  - group volume 153
  - invariance 157
- representation matrix 75
- rescaled coupling 145
- rescaling 160, 197
- resonance 80, 147, 198
- reversed spins 274, 278
- reversed statistics 231
- Riemannian geometry 156
- rising propagator 131
- rotation 204, 276
- rotationally invariant theory 9
- running coupling constants 264
- $SL(2, R)$  group 184
- $SL(2, C)$  transformations 197
- $S$ -matrix 254
- $S$ -matrix elements 265
- $SU(N)$  41, 43, 134
  - chiral field 138
- $SU(2)$  40, 45, 68, 75, 93, 97
  - theory 71
- saddle point 128
  - approximation 50, 126, 137, 141, 145
  - picture corrections 129
- scalar field 16, 71
- scalar laplacian 234
- scalar particle 32, 68
- scale invariance 88
- scale invariant theory 286
- scale transformation invariance 284
- scattering amplitude 13, 39, 134, 137, 147, 192, 200, 240, 243, 261
- Schwinger-Dyson chain 275
- Schwinger-like chain 118, 121
- screening 30, 47, 76
- secondary operator 210, 215
  - decoupling 213
- second quantized Hamiltonian 61, 82
- self-duality 89
- shear modulus 83
- short distance dynamics 128
- short distance singularities 261
- short-range correlations 176
- short-range inter-atomic interaction 82
- short string limit 282
- Shrödinger equation 44, 142
- Shrödinger perturbation theory 36
- Shrödinger wave function 20
- simplectic group 271
- singularity 67, 100, 208, 260
- singular lines 58
- slow field 255
- small coupling 43, 67, 88
- small distance behavior 149
- small distance cut-off 188
- small distance instanton-anti-instanton
  - interaction extrapolation 91
- solid state physics 73
- soliton 70, 85, 249
  - excitations 243
- space antiperiodicity 251
- space measures 157

- space rotations 97
- space-time 78, 198, 270, 273
  - anomalies 271
  - embedded string world sheet 259
  - fermion 243, 245, 250, 266
  - invariance 254
  - scale transformation 254
  - spinor 244
  - supersymmetry 252
- specific heat 278
  - singularity 274
- spectral analysis 212
- spectrum 146, 193, 253, 266
- spin 75, 277, 281
  - connection 230
  - content 202
  - flip probability 6
  - reversal 53, 276
  - structures 266
  - structure sum 252
  - systems 49
  - term 32
  - variable 274, 276
- spin operator 245, 246, 249, 250, 251, 266
  - dimensionality 248
- spinor 229, 278
  - factor 227
  - field 233
  - index 231
  - state degeneracy 244
  - structures 243–4, 270
- spin-orbit interaction 281
- spinorical excitations 249
- spontaneous magnetization 7
- stability 261, 263, 268
- stable particles 146
- standard covariant vector 159
- state counting 219
- state decoupling generalization 203
- static charges 46, 73
- static dielectric constant 84
- static potential 76
- statistical systems 274
- stereographic coordinates 181
- stereographic projection 87
- string 193, 201, 202, 203, 228, 243–4, 265–8, 287
  - distributed spin density 281
  - dynamics 118
  - ground state 197, 244
  - representations 147
  - resonance 261
  - $S$ -matrix 259
  - scattering amplitudes 195
  - spectrum 254
  - states 210
  - tension 47
  - theory 111, 212, 253, 261, 263
- strong  $CP$  violation 142
- strong coupling expansion 33, 47, 111, 118, 125
- strong coupling limit 33, 37, 42, 44, 46
- strong coupling phase 69, 128
- strong coupling problem 144
- strong coupling region 39, 40
- strong interaction 5, 18, 39, 106, 108
- strong  $\theta$ -problem 109
- strong  $T$ -violation 109
- structure constants 243, 261
- superconductor 70
- superconformal anomalies 239
- superconformal field theory 268
- superconformal gauge 231
- superconformal invariance 241
- supercovariance group 231
- supercovariant action 222
- supercovariant transformation 230
- supercurrent 223, 229, 230, 239, 281
- superfield 229, 238, 242
  - propagation 227
- superfluidity 73
- super-gravitons 253
- supergravity 273
- superinvariant distance 225
- super-length 223
- superpartner 223, 253, 266, 268, 269
- superspace 222, 228, 230
- superstring 239, 253
- supersurface 222
- supersymmetric action 225, 229, 249
- supersymmetric extension 238, 242
- supersymmetric partner 228
- supersymmetric path 224
- supersymmetric string 274
  - theory 251
- supersymmetry 223, 227, 267
  - transformation 228

- surface tension 283, 286
- symmetric structure constants 259
- symmetric tops 40
- symmetric traceless field 83
- symmetry 49–50, 60, 75, 108, 222
  - breaking 7, 13, 16, 107
  - group 40, 268–9
- $\sigma$ -independent component 244
- $\sigma$ -model 125, 254–5, 267, 285
  
- tachyon 200, 221, 241, 266
  - eliminating projection 250
  - fusion 242
- tail independence 277
- Taylor expansion in loop space 122
- tensor 70, 140, 199
- thermodynamic limit 109
- three-point functions 259
- time 69
  - anisotropy 38, 41
  - antiperiodicity 251
  - propagation 42
- topological charge 79, 86–7, 99, 103, 139, 140, 142
- topological properties 85
- topology 276
- torus 182, 271
- total angular momentum 39, 281
- total divergence 107, 284
- transformation invariance 197
- transformation law 75, 77, 112, 119, 229, 233, 236
- transition amplitude 78, 108, 114
- translational invariance 88
- translations 204
- transverse rigidities 84
- tree amplitude corrections 240
- tree level 267, 271
- tunnelling 50, 53, 103
- two-dimensional crystals 61
- two-dimensional field theory 203
- two-dimensional rectangular lattice 275
- two-dimensional surfaces 176
- two-dimensional theory 203
- two-point function 168, 207
- two point fusion 209
- $\theta$ -dependence 109
  - $\theta$ -functions 61
  - $\theta$ -term 108, 286
  
- $\tilde{u}$ -field 235
- ultraviolet divergence 126, 161
- ultraviolet region 100
- uncertainty principle 131
- unconstrained system 68
- unified theories 253
- unit lattice vectors 275
- universal correlation length 23
- universality class 12, 287
- universality properties 7
- unrenormalized regularized theory 121
- unstable kink solution 61
- $u$ -operators 260, 261
- $u$ -space 235
  
- vacuum energy 142
- vacuum expectation value 129, 266
- vacuum sector 45
- vacuum-vacuum amplitude 104
- vector field 32, 233, 268, 270
- vector mesons 269
- vector potential 14, 41, 55, 62
- vector vertex 213, 270
- vertex 145, 241
  - operator 197, 225, 240, 241–2, 250, 259, 261, 268
  - operator averaging 254
  - operator transformation 215
- $v$ -field propagator 136
- vibrational excitations 202
- vibrational modes 147
- Virasoro algebra 208, 210, 238
- volume element 157, 172
- vortex confinement 58
- vortex core 58
- vortices 16, 56, 61, 62
  
- Ward identity 12, 205, 209, 213, 271, 272
- wave equation 180
- wave function 35, 46, 277, 281
- weak coupling 54, 68, 72
- weak gravitational field 185, 190, 264

- weak interaction 5, 106
  - unification 71
- weakly coupled rotators 39
- Weierstrass  $\zeta$  function 272
- Wess-Zumino  $\sigma$ -model 269
- Weyl invariant theory 254
- Weyl transformation 254
- Wick theorem 195
- world lines 34, 70
- world sheet 201, 228, 249, 250, 257, 263, 266, 268, 270, 271, 273, 287
  - supersymmetry 269
  - supersymmetry breaking 268
- world surface 42, 146, 246
  - puncturing 240
  
- $x$ -field 174, 226, 240, 255
- $x$ -integration 185
- $x$ -space 255
  
- Yang-Mills action 17, 92, 243
- Yang-Mills equation 96
  - Yang-Mills field 5, 84, 111, 268
    - strength 29, 71, 74, 117, 138
  - Yang-Mills instanton 104
  - Yang-Mills particle 268
  - Yang-Mills vertex 243
  - $y$ -space 267
  
- Zamolodchikov's theorem 265, 267
- zero approximation 36
- zero charge 30, 259, 267, 268
- zero curvature 117
- zero mass vector particle 242
- zero mode 164, 167, 178, 273, 282
- zero momentum particles 261
- zero norm states 216, 218, 220
- zero-point energy 226, 249
- zeroth order approximation 87
- $z$ -field 140
  - quanta confinement 142
- $z$ -space 140
- Zweibein field 230
- $\zeta$ -plane 182
- $\zeta$ -space 228, 260
  - propagation 185
  - topology 171