AF

UNIVERSITÄT LEIPZIG

Naturwissenschaftlich-Theoretisches

Zentrum

CERN LIBRARIES, GENEVA

SCAN-9606064

Introduction to Gauge Theory

by

Y. M. Makeenko

Institute of Theoretical and
Experimental Physics
117 259 Moscow, Russia

and

The Niels Bohr Institute
2100 Copenhagen, Denmark

sw 9624

Preprint-Nr.20/96

Introduction to Gauge Theory

Lectures at Leipzig University 1995

by

Y. M. Makeenko*

Institute of Theoretical and Experimental Physics, 117259 Moscow, Russia

and

The Niels Bohr Institute, 2100 Copenhagen, Denmark

Abstract

The lecture notes contain a pedagogical introduction to modern methods of studying gauge theories. The main attention is paid to non-Abelian phase factors = Wilson loops, lattice gauge theories and $1/N$ -expansion of quantum chromodynamics.

CONTENTS

2

Contents

1	Introduction	5
2	Phase factors	7
2.1	Propagators in external field	7
2.2	Gauge invariance	10
2.3	Non-Abelian phase factors (definition)	12
2.4	Non-Abelian phase factors (properties)	15
2.5	Aharonov-Bohm effect	17
3	Gauge fields on a lattice	19
3.1	Sites, links, plaquettes and all that	19
3.2	Lattice formulation	21
3.3	The Haar measure	25
3.4	Wilson loops	27
3.5	Strong coupling expansion	30
3.6	Area law and confinement	32
3.7	Asymptotic scaling	34
4	Lattice methods	38
4.1	Phase transitions	38
4.2	Mean-field method	41
4.3	Mean-field method (variational)	43
4.4	Lattice renormalization group	45
4.5	Monte Carlo method	47
4.6	Some Monte Carlo results	50
5	Fermions on a lattice	52
5.1	Chiral fermions	52
5.2	Fermion doubling	53
5.3	Kogut-Susskind fermions	57
5.4	Wilson fermions	58
5.5	Quark condensate	61
6	$1/N$-expansion of vector models	64
6.1	Four-Fermi theory	64
6.2	Bubble graphs as zeroth order in $1/N$	67
6.3	Functional methods for φ^4 theory	72
6.4	Nonlinear sigma model	75
6.5	Large- N factorization in vector models	78

*E-mail addresses: makeenko@vxitep.itep.ru makeenko@nbi.dk

7 Multicolor QCD	79
7.1 Index graphs or fatgraphs	79
7.2 Planar and non-planar graphs	83
7.3 Planar and non-planar graphs (the boundaries)	84
7.4 Topological expansion and quark loops	89
7.5 t' Hooft versus Veneziano limits	92
7.6 Large- N_c factorization	95
7.7 The master field	98
7.8 $1/N_c$ as semiclassical expansion	100
8 QCD in loop space	103
8.1 Observables in terms of Wilson loops	103
8.2 Schwinger-Dyson equations for Wilson loop	105
8.3 Path and area derivatives	107
8.4 Loop equations	110
8.5 Relation to planar diagrams	112
8.6 Loop-space Laplacian and regularization	113
8.7 Survey of non-perturbative solutions	117

List of Figures

1	Discretization of continuous trajectory $z_\mu(t)$	13
2	Rectangular loop in the μ, ν -plane	16
3	Principal scheme of the Aharonov-Bohm experiment	17
4	Two-dimensional lattice with periodic boundary conditions	20
5	A link of a lattice	20
6	A plaquette of a lattice	21
7	Description of continuum configurations by lattices	22
8	Oriented boundary of a plaquette	23
9	Rectangular loop of the size $R \times T$	28
10	Boundaries of plaquettes with opposite orientations	31
11	Filling of a loop with plaquettes	32
12	Lines of force between static quarks	33
13	Dependence of the string tension on $1/g^2$	35
14	Specific energy for a first-order phase transition	39
15	Specific heat for a second-order phase transition	40
16	Graphic representation of the self-consistency condition	41
17	Solutions of the mean-field self-consistency equation	42
18	Lattice renormalization group transformation	45
19	Monte Carlo data [33] for $\Delta\beta$	47
20	Monte Carlo data [22] for the string tension	50
21	Equipotential lines at different values of β	51
22	Altering signs of ψ_x on a lattice	55

23	Momentum dependence of G^{-1} for lattice fermions	56
24	Momentum dependence of G^{-1} for lattice bosons	57
25	A path made out of the string bits	61
26	Dependence of π -meson mass on the quark mass	62
27	Monte Carlo data [44] for quark condensate	62
28	One-loop diagrams for the four-vertex	66
29	One-loop diagrams for the propagator of ψ -field	66
30	Bubble diagram which survives the large- N limit	67
31	Diagrams of the $1/N$ -expansion	68
32	Double-line representation of a one-loop diagram for gluon propagator	81
33	Double-line representation of a four-loop diagram	83
34	Double-line representation of a non-planar diagram	83
35	Generic double-line index diagram	85
36	Planar and non-planar parts of three-gluon vertex	86
37	Connected and disconnected planar graphs	87
38	Graphic derivation of Eq. (7.31)	88
39	Diagrams for gluon propagator which involve quark loop	89
40	Generic diagram in the index space	90
41	Diagrams with one quark loop in the Veneziano limit	93
42	Diagram with two quark loops in the Veneziano limit	94
43	Demonstration of the large- N_c factorization in perturbation theory	96
44	Contours in sum over paths representing observables	104
45	Examples of smooth contour and contour with cusp	105
46	Contours $C_{y,x}$ and $C_{x,y}$ entering loop equation	107
47	Graphic representation of Eq. (8.41)	113
48	Planar diagrams for $W(C)$ to order λ^2	113
49	Contours $C_{y,x}r_{xy}$ and $C_{x,y}r_{yx}$ in the regularized loop equation	116

1 Introduction

Modern theory of fundamental interactions is Gauge Theory. Examples of gauge theories include:

- Einstein's theory of gravitation (1916);
- Weyl's formulation of electromagnetism (1919);
- Yang-Mills fields (1954):
 electroweak theory (1967),
 quantum chromodynamics (1972).

The key point in gauge theories is played by the principle of local gauge invariance. The interactions are mediated by the gauge fields (photons, gluons, gravitons, ...) whose properties are conceptually different from those of the matter fields (leptons, quarks, ...). For this reason, gauge theories admit their own methods of non-perturbative investigation, which are not available in the standard quantum field theory (say, of self-interacting scalar or spinor fields).

These lecture notes contain a pedagogical introduction to modern methods of studying gauge theories. The main attention is paid to non-Abelian phase factors = Wilson loops, lattice gauge theories and $1/N$ -expansion of quantum chromodynamics.

We first introduce, in Section 2, Abelian and non-Abelian phase factors and show why they are observable in quantum theory. These phase factors, which are also referred to as the Wilson loops, are used throughout these lecture notes.

Sections 3-5 are devoted to lattice gauge theories which were proposed, in their modern form, in 1974 by Wilson [1] in connection with the problem of quark confinement in quantum chromodynamics (QCD).

Lattice gauge theories are a non-perturbative regularization of a gauge theory. The lattice formulation is a nontrivial definition of a gauge theory beyond perturbation theory. The problem of non-perturbative quantization of gauge fields is solved in a simple and elegant way on a lattice.

The use of the lattice formulation clarifies an analogy between quantum field theory and statistical mechanics. It offers a possibility to apply the non-perturbative methods, such as the strong coupling expansion or the numerical Monte Carlo method, to quantum chromodynamics and to other gauge theories, which provide an evidence for quark confinement.

However, the lattice in QCD is no more than an auxiliary tool to obtain results for the continuum limit. In order to pass to the continuum, the lattice spacing should be many times smaller than a characteristic scale of the strong interaction.

We describe the lattice formulation of gauge theories, and return to the continuum formulation from time to time when discussing the lattice approach. The point is that some ideas, *e.g.* about a possibility of reformulating gauge theories in terms of gauge invariant variables, which were originally introduced by Wilson [1] on a lattice, are applicable for the continuum theory as well.

Sections 6-8 are devoted to the method of the $1/N$ -expansion which is used in many physical problems, especially when fluctuations of scales of different orders of magnitude are

essential, so that there is no small parameter which could simplify a study. A typical example is QCD where the effective coupling, describing strong interaction at a given distance, becomes large at large distances so that the interaction really becomes strong.

't Hooft [2] proposed in 1974 to use the dimensionality of the gauge group $SU(N_c)$ as such a parameter, considering the number of colors, $N_c = 3$ in the real world, as a large number and performing an expansion in $1/N_c$. The motivation was an expansion in the inverse number of field-components N in statistical mechanics where it is known as the $1/N$ -expansion, and is a standard method for non-perturbative investigations.

The expansion of QCD in the inverse number of colors rearranges diagrams of perturbation theory in a way which is consistent with a string picture of strong interaction, whose phenomenological consequences agree with experiment. The accuracy of the leading-order term, which is often called multicolor QCD or large- N QCD, is expected to be of the order of the ratios of meson widths to their masses, *i.e.* about 10-15%.

While QCD is simplified in the large- N_c limit, it is still not yet solved. Generically, it is a problem of infinite matrices, rather than of infinite vectors as in the theory of second-order phase transitions in statistical mechanics.

We show how the $1/N$ -expansion works for the $O(N)$ -vector models and describe some applications to the four-Fermi interaction, the φ^4 theory and the nonlinear sigma model. Then we concentrate on multicolor QCD.

I am indebted to Martin Grtler for his help in preparing these lecture notes.

2 Phase factors

Modern theories of fundamental interactions are gauge theories. The principle of local gauge invariance was introduced by H. Weyl for electromagnetic interaction in analogy with general covariance in Einstein's theory of gravitation. An extension to non-Abelian gauge groups was given by Yang and Mills [3].

A crucial role in gauge theories is played by the phase factor which is associated with a parallel transport in an external gauge field. The phase factors are observable in quantum theory, in contrast to the classical theory. For the electromagnetic field, this is known as the Aharonov–Bohm effect.

We first consider in this Section propagators in an external electromagnetic field and then introduce the matrix notation for the non-Abelian gauge fields and proper non-Abelian phase factors. Finally, we discuss the relation between observables in classical and quantum theories.

2.1 Propagators in external field

Let us consider a (quantum) particle propagating in a classical electromagnetic field. The standard way of introducing an external electromagnetic field is to substitute the (operator of the) four-momentum $\mathbf{p}^\mu$ by

$$\mathbf{p}^\mu \longrightarrow \mathbf{p}^\mu - eA^\mu(x). \quad (2.1)$$

We shall deal throughout these lecture notes with the Euclidean space where $x_4 = it$ and

$$\mathbf{p}_\mu = -i\partial_\mu \equiv \left(-i\frac{\partial}{\partial x}, -i\frac{\partial}{\partial x_4}\right) \quad \boxed{\text{Euclidean space}} \quad (2.2)$$

for a free particle. In the presence of an electromagnetic field, ∂_μ needs to be replaced by the covariant derivative

$$\partial_\mu \longrightarrow \nabla_\mu = \partial_\mu - ieA_\mu(x). \quad (2.3)$$

The path-integral representation for the propagator of a scalar particle in an external electromagnetic field is due to Feynman [4] and reads

$$\begin{aligned} G(x, y; A) &\equiv \left\langle y \left| \frac{1}{-\nabla_\mu^2 + m^2} \right| x \right\rangle \\ &= \frac{1}{2} \int_0^\infty d\tau e^{-\frac{1}{2}\tau m^2} \int_{\substack{z(0)=x \\ z(\tau)=y}} Dz(t) e^{-\frac{1}{2} \int_0^\tau dt \dot{z}^2(t) + ie \int_0^\tau dt \dot{z}^\mu(t) A_\mu(z(t))}. \end{aligned} \quad (2.4)$$

Here the path integration is over all trajectories $z_\mu(t)$ between the initial point x and the final point y . The parameter τ plays the role of a proper time during which the particle propagates from x to y . Notice that the exponent is just the classical (Euclidean) action, in

the proper-time parametrization, of a particle in the presence of an external electromagnetic field.

We can alternatively rewrite Eq. (2.4) as

$$G(x, y; A) = \sum'_{\Gamma_{yx}} e^{ie \int_{\Gamma_{yx}} dz^\mu A_\mu(z)}, \quad (2.5)$$

where we have included the free action in the definition of the sum over trajectories:

$$\sum'_{\Gamma_{yx}} = \sum_{\Gamma_{yx}} e^{-S_{\text{free}}[\Gamma_{yx}]}, \quad (2.6)$$

and rewritten the (parametric invariant) integral over dt as the contour integral over

$$dz^\mu = dt \dot{z}^\mu(t) \quad (2.7)$$

along the trajectory Γ_{yx} .

Here Γ_{yx} stands for a trajectory as a geometric object, while $z_\mu(t)$ is a function which describes a given trajectory in some parametrization t . This function (but not the geometric object itself) depends on the choice of parametrization and changes under the *reparametrization transformation*

$$t \rightarrow \sigma(t), \quad \frac{d\sigma}{dt} \geq 0, \quad (2.8)$$

with σ being a new parameter. A convenient parametrization is via the proper length of Γ_{yx} .

The meaning of Eq. (2.5) is that the transition amplitude of a quantum particle in a classical electromagnetic field is the sum over paths of the Abelian *phase factor*

$$U[\Gamma_{yx}] = e^{ie \int_{\Gamma_{yx}} dz^\mu A_\mu(z)}. \quad (2.9)$$

Under the gauge transformation

$$A_\mu(z) \xrightarrow{g, t} A_\mu(z) + \frac{1}{e} \partial_\mu \alpha(z), \quad (2.10)$$

the Abelian phase factor transforms as

$$U[\Gamma_{yx}] \xrightarrow{g, t} e^{i\alpha(y)} U[\Gamma_{yx}] e^{-i\alpha(x)}. \quad (2.11)$$

Noting that a wave function at the point x is transformed under the gauge transformation (2.10) as

$$\varphi(x) \xrightarrow{g, t} e^{i\alpha(x)} \varphi(x), \quad (2.12)$$

we conclude that the phase factor is transformed as the product $\varphi(y)\varphi^\dagger(x)$:

$$U[\Gamma_{yx}] \xrightarrow{g, t} \varphi(y)\varphi^\dagger(x), \quad (2.13)$$

where “...” means literally that “transforms as ...”.

As a consequence of Eqs. (2.11) and (2.12), a wave function at the point x transforms like one at the point y after multiplication by the phase factor:

$$U[\Gamma_{yx}] \varphi(x) \xrightarrow{\text{g.t.}} \varphi(y), \quad (2.14)$$

and analogously

$$\varphi^\dagger(y) U[\Gamma_{yx}] \xrightarrow{\text{g.t.}} \varphi^\dagger(x). \quad (2.15)$$

Equations (2.14) and (2.15) show that the phase factor plays the role of a *parallel transporter* in an electromagnetic field, and that in order to compare phases of a wave function at points x and y , one should first make a parallel transport along some contour Γ_{yx} . The result is, generally speaking, Γ dependent except when $A_\mu(z)$ is a pure gauge. The sufficient and necessary condition for the phase factor to be Γ independent is the vanishing of the field strength, $F_{\mu\nu}(z)$, which is a consequence of the Stokes theorem when applied to the Abelian phase factor¹.

Below we shall deal with determinants of various operators. Analogous to Eq. (2.4), one gets

$$\ln \det \nabla_\mu^2 = \frac{1}{2} \int_0^\infty \frac{d\tau}{\tau} \text{Sp} e^{\frac{1}{2}\tau \nabla_\mu^2} = \frac{1}{2} \int_0^\infty \frac{d\tau}{\tau} \int_{z(0)=z(\tau)} Dz(t) e^{-\frac{1}{2} \int_0^\tau dt \dot{z}^2(t) + ie \oint_\Gamma dz^\mu A_\mu(z)}, \quad (2.16)$$

where the path integral goes over trajectories which are closed due to the boundary condition $z_\mu(0) = z_\mu(\tau)$. Equation (2.16) can be derived by using the formula

$$\ln \det \mathbf{D} = \text{Sp} \ln \mathbf{D}, \quad (2.17)$$

which relates the determinant and the trace of a Hermitean operator (or a matrix) $\mathbf{D}$.

The phase factor for a closed contour Γ enters Eq. (2.16). It describes a parallel transportation along a closed loop, and is gauge invariant as a consequence of Eq. (2.11):

$$e^{ie \oint_\Gamma dz^\mu A_\mu(z)} \xrightarrow{\text{g.t.}} e^{ie \oint_\Gamma dz^\mu A_\mu(z)}. \quad (2.18)$$

This quantity, which plays a crucial role in modern formulations of gauge theories, will be discussed in more detail below in this Section.

Remark on analogy with statistical mechanics

A formula of the type (2.16), which represents the trace of an operator via path integral over closed trajectories, is known as the Feynman–Kac formula. The terminology comes

¹Strictly speaking, this statement holds for the case when Γ can be chosen everywhere in space-time, i.e. which is simply connected. However, there exist situations when Γ can not penetrate into some regions of space like for the Aharonov–Bohm experiment which is discussed below in Subsect. 2.5.

from statistical mechanics where the partition function (or equivalently the statistical sum) is given by the Boltzmann formula

$$Z = \text{Sp} e^{-\beta \mathbf{H}} \quad (2.19)$$

(with β being the inverse temperature and $\mathbf{H}$ being the Hamiltonian) whose path-integral representation is of the type in Eq. (2.16). The expression which is integrated on the RHS of Eq. (2.16) over $d\tau/\tau$ is associated, in the statistical-mechanical language, with the partition function of a closed elastic string, whose energy is proportional to its length, and interacts with an external electromagnetic field. This shows an analogy between Euclidean quantum mechanics in d -dimensions and statistical mechanics in d (spatial)- and 1 (temporal)-dimensions whose time-dependence disappears, since nothing depends on time at equilibrium.

2.2 Gauge invariance

The principle of local gauge invariance deals with the gauge transformation of a matter field ψ , which reads

$$\psi(x) \xrightarrow{\text{g.t.}} \psi'(x) = \Omega(x) \psi(x). \quad (2.20)$$

Here $\Omega(x) \in G$ with G being a semisimple Lie group which is called the gauge group ($G = \text{SU}(3)$ for QCD). Equation (2.20) means that ψ belongs to the fundamental representation of G .

We restrict ourselves to a unitary gauge group when

$$\Omega^{-1}(x) = \Omega^\dagger(x), \quad (2.21)$$

while an extension to other Lie groups is obvious. Then we have

$$\psi^\dagger(x) \xrightarrow{\text{g.t.}} \psi'^\dagger(x) = \psi^\dagger(x) \Omega^\dagger(x). \quad (2.22)$$

In analogy with QCD, the gauge group $G = \text{SU}(N_c)$ is usually associated with *color* while the proper index of ψ is called the color index.

The gauge transformation (2.20) of the matter field ψ can be compensated by a transformation of the non-Abelian gauge field A_μ which belongs to the adjoint representation of G :

$$A_\mu(x) \xrightarrow{\text{g.t.}} A'_\mu(x) = \Omega(x) A_\mu(x) \Omega^\dagger(x) - \Omega(x) \partial_\mu \Omega^\dagger(x). \quad (2.23)$$

We have introduced in Eq. (2.23) the anti-Hermitean matrix $A_\mu(x)$ with the elements

$$[A_\mu(x)]^{ij} = ig \sum_a A_\mu^a(x) [t^a]^{ij}. \quad (2.24)$$

Here $[t^a]^{ij}$ are the generators of G ($a = 1, \dots, N_c^2 - 1$ for $\text{SU}(N_c)$) which are normalized so that

$$\text{tr} t^a t^b = \frac{1}{2} \delta^{ab}, \quad (2.25)$$

where tr is the trace over the matrix indices i and j , while g is the gauge coupling constant. This normalization is due to historical reasons, in particular

$$t^a = \frac{\sigma^a}{2} \quad (2.26)$$

for the $\text{SU}(2)$ group, with σ^a being the Pauli matrices.

Equation (2.24) can be inverted to give

$$A_\mu^a(x) = \frac{2}{ig} \text{tr} \mathcal{A}_\mu(x) t^a. \quad (2.27)$$

Substituting

$$\Omega(x) = e^{i\alpha(x)}, \quad (2.28)$$

we get for an infinitesimal α :

$$\delta \mathcal{A}_\mu(x) \stackrel{\text{g.t.}}{=} i \nabla_\mu^{\text{adj}} \alpha(x). \quad (2.29)$$

Here

$$\nabla_\mu^{\text{adj}} \alpha \equiv \partial_\mu \alpha - [\mathcal{A}_\mu, \alpha] \quad (2.30)$$

is the covariant derivative in the adjoint representation of G while

$$\nabla_\mu^{\text{fun}} \psi \equiv \partial_\mu \psi - \mathcal{A}_\mu \psi \quad (2.31)$$

is the one in the fundamental representation. It is evident that

$$\nabla_\mu^{\text{adj}} B(x) = [\nabla_\mu^{\text{fun}}, B(x)] \quad (2.32)$$

where $B(x)$ is a matrix-valued function of x .

The QCD action reads in the matrix notation as

$$S[\mathcal{A}, \psi, \bar{\psi}] = \int d^4x \left[\bar{\psi} \gamma_\mu (\partial_\mu - \mathcal{A}_\mu) \psi + m \bar{\psi} \psi - \frac{1}{2g^2} \text{tr} \mathcal{F}_{\mu\nu}^2 \right], \quad (2.33)$$

where

$$\mathcal{F}_{\mu\nu} = \partial_\mu \mathcal{A}_\nu - \partial_\nu \mathcal{A}_\mu - [\mathcal{A}_\mu, \mathcal{A}_\nu] \quad (2.34)$$

is the (anti-Hermitean) matrix of the non-Abelian field strength.

The action (2.33) is manifestly invariant under the local gauge transformation (2.20), (2.23) since

$$\mathcal{F}_{\mu\nu}(x) \stackrel{\text{g.t.}}{\rightarrow} \Omega(x) \mathcal{F}_{\mu\nu}(x) \Omega^\dagger(x) \quad (2.35)$$

or

$$\delta \mathcal{F}_{\mu\nu}(x) \stackrel{\text{g.t.}}{=} -i [\mathcal{F}_{\mu\nu}(x), \alpha(x)] \quad (2.36)$$

for the infinitesimal gauge transformation.

The non-Abelian Maxwell equation and the Bianchi identity can be written in the matrix notation, respectively, as

$$\nabla_\mu^{\text{adj}} \mathcal{F}_{\mu\nu} = 0 \quad (2.37)$$

and

$$\nabla_\mu^{\text{adj}} \tilde{\mathcal{F}}_{\mu\nu} = 0, \quad (2.38)$$

where the dual tensor is defined by

$$\tilde{\mathcal{F}}_{\mu\nu} = \frac{1}{2} \epsilon_{\mu\nu\lambda\rho} \mathcal{F}_{\lambda\rho}. \quad (2.39)$$

Remark of first- and second-order formalisms

Rewriting Eq. (2.34) as

$$\mathcal{F}_{\mu\nu} = -[\nabla_\mu^{\text{fun}}, \nabla_\nu^{\text{fun}}] \quad (2.40)$$

and using Eq. (2.32), we represent the Bianchi identity as

$$\epsilon_{\mu\nu\lambda\rho} [\nabla_\mu^{\text{fun}}, [\nabla_\nu^{\text{fun}}, \nabla_\lambda^{\text{fun}}]] = 0 \quad (2.41)$$

which is obviously satisfied due to the Jacobi identity.

We have thus shown the well-known fact that the Bianchi identity is explicitly satisfied in the second-order formalism, where $\mathcal{F}_{\mu\nu}$ is expressed via $\mathcal{A}_\mu$ by virtue of Eq. (2.34). On the contrary, $\mathcal{A}_\mu$ and $\mathcal{F}_{\mu\nu}$ are considered as independent variables in the first-order formalism, where both equations (2.37) and (2.38) are essential. The concept of the first- and second-order formalisms comes from the theory of gravity.

2.3 Non-Abelian phase factors (definition)

In order to compare the phases of wave functions at distinct points, one needs a non-Abelian extension of the parallel transporter which was considered in Subsect. 2.1. The proper extension of the Abelian formula (2.9) is written as

$$U[\Gamma_{yx}] = \mathbf{P} e^{\int_{\Gamma_{yx}} dz^\mu \mathcal{A}_\mu(z)}. \quad (2.42)$$

While the matrices $\mathcal{A}_\mu(z)$ do not commute, the path-ordered exponential on the RHS of Eq. (2.42) is unambiguously defined by the path-ordering prescription which is characterized

Fig. 1: Discretization of a continuous trajectory $z_\mu(t)$ by the line connecting $M+1$ points $z_0 = x, z_1, \dots, z_{M-1}, z_M = y$.

by the path-ordering operator $\mathbf{P}$. This is obvious after rewriting the phase factor in an equivalent form

$$\mathbf{P} e^{\int_{\Gamma_{yx}} dz^\mu \mathcal{A}_\mu(z)} = \mathbf{P} e^{\int_0^\tau dt \dot{z}^\mu(t) \mathcal{A}_\mu(z(t))}, \quad (2.43)$$

where the ordering is with respect to the parameter t .

Therefore, the path-ordered exponential in Eq. (2.42) is to be understood as

$$U[\Gamma_{yx}] = \prod_{t=0}^\tau [1 + dt \dot{z}^\mu(t) \mathcal{A}_\mu(z(t))], \quad (2.44)$$

where the matrices are ordered from the right to the left with increasing t .² Using Eq. (2.7), Eq. (2.44) can also be written as

$$U[\Gamma_{yx}] = \prod_x^y [1 + dz^\mu \mathcal{A}_\mu(z)]. \quad (2.45)$$

If the contour Γ_{yx} is discretized by the line connecting $M+1$ points $z_0 = x, z_1, \dots, z_{M-1}, z_M = y$ (with $z_i - z_{i-1} \sim \epsilon$) as is shown in Fig. 1, then the non-Abelian phase factor is approximated by

$$U[\Gamma_{yx}] = \lim_{M \rightarrow \infty} \prod_{i=1}^M \left[1 + (z_i - z_{i-1})^\mu \mathcal{A}_\mu \left(\frac{z_i + z_{i-1}}{2} \right) \right], \quad (2.46)$$

which obviously reproduces (2.45) in the limit $\epsilon \rightarrow 0$.

Notice that the non-Abelian phase factor (2.42) is, by construction, an element of the gauge group G itself, while $\mathcal{A}_\mu$ belongs to the Lie algebra of G .

Let us use the notation

$$\int_x^y dz^\mu \dots \equiv \int_{\Gamma_{yx}} dz^\mu \dots \quad (2.47)$$

²Sometimes the phase factor is defined by a similar formula but with the inverse order of multipliers. Our definition by Eq. (2.44) is exactly equivalent to the Dyson's definition of the $\mathbf{P}$ product which can be seen by choosing the contour Γ_{yx} to coincide with the temporal axis.

for the integral along the contour Γ_{yx} . By expanding the definition (2.42) of the non-Abelian phase factor in $\mathcal{A}_\mu$, we then write down an explicit formula

$$\mathbf{P} e^{\int_x^y dz^\mu \mathcal{A}_\mu(z)} = \sum_{k=0}^{\infty} \int_x^y dz_1^{\mu_1} \int_{z_1}^y dz_2^{\mu_2} \dots \int_{z_{k-1}}^y dz_k^{\mu_k} \mathcal{A}_{\mu_k}(z_k) \dots \mathcal{A}_{\mu_2}(z_2) \mathcal{A}_{\mu_1}(z_1). \quad (2.48)$$

For the Abelian case, when $\mathcal{A}(z_i)$ commute, Eq. (2.48) results in

$$\int_x^y dz_1^{\mu_1} \int_{z_1}^y dz_2^{\mu_2} \dots \int_{z_{k-1}}^y dz_k^{\mu_k} \mathcal{A}_{\mu_k}(z_k) \dots \mathcal{A}_{\mu_2}(z_2) \mathcal{A}_{\mu_1}(z_1) = \frac{1}{k!} \left(\int_x^y dz^\mu \mathcal{A}_\mu(z) \right)^k \quad (2.49)$$

so that the Abelian exponential of the contour integral is reproduced.

The non-Abelian phase factor emerges in a non-Abelian extension of Eq. (2.5), when the covariant derivative is in the fundamental representation of the gauge group. It reads

$$G(x, y; \mathcal{A}) \equiv \left\langle y \left| \frac{1}{-\nabla_\mu^{\text{fun}} \nabla_\mu^{\text{fun}} + m^2} \right| x \right\rangle = \sum'_{\Gamma_{yx}} \mathbf{P} e^{\int_{\Gamma_{yx}} dz^\mu \mathcal{A}_\mu(z)}, \quad (2.50)$$

where Σ' is defined by Eq. (2.6).

In order to write down an analogous formula for the adjoint representation, let us introduce

$$\nabla_\mu^{ab} = \partial_\mu \delta^{ab} - g f^{abc} A_\mu^c \quad (2.51)$$

with f^{abc} being the structure constants of the gauge group. Then the Green function $G^{ab}(x, y; \mathcal{A})$ obeys

$$\left(-\nabla_\mu^{ac} \nabla_\mu^{cb} + m^2 \delta^{ab} \right) G^{bd}(x, y; \mathcal{A}) = \delta^{ad} \delta^{(d)}(x - y), \quad (2.52)$$

and we get

$$G^{ab}(x, y; \mathcal{A}) = \sum'_{\Gamma_{yx}} 2 \text{tr} t^b U[\Gamma_{yx}] t^a U^\dagger[\Gamma_{yx}], \quad (2.53)$$

where $U[\Gamma_{yx}]$ is given by Eq. (2.42).

Since matrices are rearranged in an inverse order under Hermitean conjugation, one has³

$$U^\dagger[\Gamma_{yx}] = U[\Gamma_{xy}]. \quad (2.54)$$

In particular, the phase factors obey the *backtracking* condition

$$U[\Gamma_{yx}] U[\Gamma_{xy}] = 1. \quad (2.55)$$

We have chosen $\mathcal{A}_\mu$ in the discretized phase factor (2.46) at the center on the i -th interval in order to satisfy Eq. (2.55) at finite ϵ .

³The notation Γ_{yx} means that the contour is oriented from x to y while Γ_{xy} stands for the opposite orientation from y to x . In the path-ordered product (2.45), these two contours result in the opposite orders of multiplying the matrices.

2.4 Non-Abelian phase factors (properties)

Under the gauge transformation (2.23) the non-Abelian phase factor (2.42) transforms as

$$U[\Gamma_{yx}] \xrightarrow{g,t} \Omega(y) U[\Gamma_{yx}] \Omega^\dagger(x). \quad (2.56)$$

This formula stems from the fact that

$$\begin{aligned} [1 + dz^\mu \mathcal{A}_\mu(z)] &\xrightarrow{g,t} [1 + dz^\mu \mathcal{A}'_\mu(z)] \\ &= \Omega(z + dz) [1 + dz^\mu \mathcal{A}_\mu(z)] \Omega^\dagger(z) \end{aligned} \quad (2.57)$$

under the gauge transformation, which can be proven by substituting Eq. (2.23), so that $\Omega^\dagger(z)$ and $\Omega(z)$ cancel in the definition (2.45) at the intermediate point z .

One of the consequences of Eq. (2.56) is that $\psi(x)$, transported by the matrix $U[\Gamma_{yx}]$ to the point y , transforms under the gauge transformation as $\psi(y)$:

$$U[\Gamma_{yx}] \psi(x) \xrightarrow{g,t} \psi(y), \quad (2.58)$$

and, analogously,

$$\bar{\psi}(y) U[\Gamma_{yx}] \xrightarrow{g,t} \bar{\psi}(x). \quad (2.59)$$

Therefore, $U[\Gamma_{yx}]$ is, indeed, a parallel transporter.

It follows from these formulas that $\bar{\psi}(y) U[\Gamma_{yx}] \psi(x)$ is gauge invariant:

$$\bar{\psi}(y) U[\Gamma_{yx}] \psi(x) \xrightarrow{g,t} \bar{\psi}(y) U[\Gamma_{yx}] \psi(x). \quad (2.60)$$

Another consequence of Eq. (2.56) is that the trace of the phase factor for a closed contour Γ is gauge invariant:

$$\text{tr} \mathbf{P} e^{\oint_\Gamma dz^\mu \mathcal{A}_\mu(z)} \xrightarrow{g,t} \text{tr} \mathbf{P} e^{\oint_\Gamma dz^\mu \mathcal{A}_\mu(z)}. \quad (2.61)$$

These properties of the non-Abelian phase factor are quite similar to those of the Abelian one which was considered in Subsect. 2.1.

The partial derivative of $U[\Gamma_{yx}]$ with respect to either the initial point x or the final point y can easily be calculated. It is convenient to start from Eq. (2.46). Then only $(z_1 - x)$ in the last element of the product should be differentiated with respect to x or $(y - z_{M-1})$ in the first element of the product should be differentiated with respect to y . As $\epsilon \rightarrow 0$, we get

$$\begin{aligned} \frac{\partial}{\partial x_\mu} \mathbf{P} e^{\int_x^y dz^\mu \mathcal{A}_\mu(z)} &= -\mathbf{P} e^{\int_x^y dz^\mu \mathcal{A}_\mu(z)} \mathcal{A}_\mu(x), \\ \frac{\partial}{\partial y_\mu} \mathbf{P} e^{\int_x^y dz^\mu \mathcal{A}_\mu(z)} &= \mathcal{A}_\mu(y) \mathbf{P} e^{\int_x^y dz^\mu \mathcal{A}_\mu(z)}. \end{aligned} \quad (2.62)$$

These formulas are exactly the same as if one were to just differentiate the lower and upper limit in the path-ordered integral keeping in mind the ordering of matrices.

Fig. 2: The rectangular loop δC_{zz} , which is added to the contour Γ_{yx} at the intermediate point z in the μ, ν -plane.

One can rewrite Eq. (2.62) via the covariant derivatives as

$$\nabla_\mu^{\text{fun}}(y) U[\Gamma_{yx}] = 0, \quad U[\Gamma_{yx}] \overleftarrow{\nabla}_\mu^{\text{fun}}(x) = 0. \quad (2.63)$$

where $\overleftarrow{\nabla}_\mu^{\text{fun}} = \overleftarrow{\partial}_\mu + \mathcal{A}_\mu(x)$. It is the property of the parallel transporter which is annihilated by the covariant derivative.

The sufficient and necessary condition for the phase factor to be independent on a local variation of the path is the vanishing of $\mathcal{F}_{\mu\nu}$. In order to show this, let us add to Γ_{yx} at the point $z \in \Gamma_{yx}$ an infinitesimal rectangular loop δC_{zz} which lies in the μ, ν -plane, as is depicted in Fig. 2. Then the variation of the phase factor is

$$\delta U[\Gamma_{yx}] \equiv U[\Gamma_{yz} \delta C_{zz} \Gamma_{zx}] - U[\Gamma_{yx}] = U[\Gamma_{yz}] \mathcal{F}_{\mu\nu}(z) U[\Gamma_{zx}] \delta\sigma_{\mu\nu}(z). \quad (2.64)$$

where $\delta\sigma_{\mu\nu}(z) = dz_\mu \wedge dz_\nu$ and the symbol $\wedge$ implies antisymmetrization. We can rewrite Eq. (2.64) as

$$\delta U[\Gamma_{yx}] = \mathbf{P} U[\Gamma_{yx}] \mathcal{F}_{\mu\nu}(z) \delta\sigma_{\mu\nu}(z) \quad (2.65)$$

since the P-product will automatically put $\mathcal{F}_{\mu\nu}(z)$ at the point z of the contour Γ_{yx} .

Remark on analogy with differential geometry

The formulas of the type of Eq. (2.64) are well-known in differential geometry where the parallel transport around a small closed contour determines the curvature. Therefore, $\mathcal{F}_{\mu\nu}$ in Yang-Mills theory is the proper curvature in an internal color space while $\mathcal{A}_\mu$ is the connection.

A historical remark

An analog of the phase factors was first introduced by Weyl [5] in his attempt to describe gravitational and electromagnetic interaction of an electron on equal footing. What he did is associated in the modern language with the scale rather than gauge transformation,

Fig. 3: Principal scheme of the experiment which demonstrates the Aharonov–Bohm effect. Electrons do not pass inside the solenoid where the magnetic field is concentrated. Nevertheless, a phase difference arises between the electron beams passing through the two slits. The interference picture changes with the value of electric current.

i.e. the vector-potential was not multiplied by i as in Eq. (2.9). This explains the term “gauge invariance” — gauging literally means fixing a scale. The factor of i was inserted by London [6] after the creation of quantum mechanics and recognition of the fact that electromagnetic interaction corresponds to the freedom of a choice of the phase of a wave function and not with a scale transformation. However, the terminology was left.

2.5 Aharonov–Bohm effect

The simplest example of a gauge field is the electromagnetic field whose transversal components describe photons. Otherwise, the longitudinal components of the vector-potential, which are changeable under the gauge transformation, are related to gauging the phase of a wave function, *i.e.* permit to compare its values at different space-time points when an electron is placed in an external electromagnetic field.

As is well-known in quantum mechanics, the wave-function phase itself is unobservable. Only the phase differences are observable, *e.g.* via interference phenomena. For the electron in an electromagnetic field, the current (gauged) value of the phase of the wave function ψ at the point y is related, as is discussed in Subsect. 2.1, to its value at some reference point x by the parallel transport which is given by Eq. (2.14). Therefore, the phase difference depends on the value of the phase factor for a given path Γ_{yx} along which the parallel transport is performed.

It is essential that the phase factors are observable in quantum theory, contrary to classical theory. This is seen in the Aharonov–Bohm effect. The corresponding experiment is depicted schematically in Fig. 3.

It allows one to measure the phase difference between electrons passing through the two slits and, therefore, going across opposite sides of the solenoid. The fine point is that magnetic field is nonvanishing only inside the solenoid where electrons do not penetrate. Hence the electrons pass throughout the region of space where the magnetic field strength vanishes! Nevertheless, the vector potential A_μ itself does not vanish which results in observable consequences.

The probability amplitude for an electron to propagate from the source at the point x to the point y at the interference plane is given by the Minkowski-space analog of Eq. (2.5):

$$\Psi(x, y) = \sum_{\Gamma_{yx}^+} e^{ie \int_{\Gamma_{yx}^+} dz^\mu A_\mu(z)} + \sum_{\Gamma_{yx}^-} e^{ie \int_{\Gamma_{yx}^-} dz^\mu A_\mu(z)}, \quad (2.66)$$

where the contour Γ_{yx}^+ passes through the upper slit while the contour Γ_{yx}^- passes through the lower one.

The intensity of the interference pattern is given by $|\Psi(x, y)|^2$ which contains, in particular, the term which is proportional to

$$e^{ie \int_{\Gamma_{yx}^+} dz^\mu A_\mu(z)} e^{-ie \int_{\Gamma_{yx}^-} dz^\mu A_\mu(z)} = e^{ie \oint_\Gamma dz^\mu A_\mu(z)}, \quad (2.67)$$

where the closed contour Γ is composed from Γ_{yx}^+ and Γ_{yx}^- . This is nothing but the phase factor associated with a parallel transport along the closed contour Γ .

For the given process this phase factor does not depend on the shape of Γ_{yx}^+ and Γ_{yx}^- . Applying the Stokes theorem, one gets

$$e^{ie \oint_\Gamma d\xi^\mu A_\mu} = e^{ie \int d\sigma^{\mu\nu} F_{\mu\nu}} = e^{ie HS}, \quad (2.68)$$

where HS is the magnetic flow through the solenoid. Therefore, the interference picture changes when H changes⁴.

Remark on quantum vs. classical observables

A moral from the Aharonov–Bohm experiment is that the phase factors are observable in quantum theory while in classical theory only the electric and magnetic field strengths are observable. The vector potential plays, in classical theory, only an auxiliary role to determine the field strength.

For the non-Abelian gauge group $G = \text{SU}(N_c)$, a quark can alter its color under the parallel transport so the non-Abelian phase factor (2.42) is a unitary $N_c \times N_c$ matrix. A non-Abelian analogue of the quantity, which is measurable in the Aharonov–Bohm experiment, is the trace of the matrix of the parallel transport along a closed path. It is gauge invariant according to Eq. (2.61).

It looks suggestive to reformulate gauge theories entirely in terms of these observable quantities. How to do that will be explained in Sect. 8.

⁴A detailed computation of the interference picture for the Aharonov–Bohm experiment is contained, *e.g.*, in the review by Kobe [7].

3 Gauge fields on a lattice

The modern formulation of non-Abelian lattice gauge theories is due to Wilson [1]. Independently, gauge theories were discussed on a lattice by Wegner [8] as a gauge invariant extension of the Ising model and in an unpublished work by A. Polyakov in 1974 which deals mostly with Abelian theories.

Placing gauge fields on a lattice provides, firstly, a non-perturbative regularization of ultraviolet divergencies. Secondly, the lattice formulation of QCD possesses some non-perturbative terms in addition to perturbation theory. A result of this is that one has a nontrivial definition of QCD beyond perturbation theory which guarantees confinement of quarks.

The lattice formulation of gauge theories deals with phase-factor like quantities, which are elements of the gauge group, and are natural variables for quantum gauge theories.

The gauge group on the lattice is compact therefore offering the possibility of non-perturbative quantization of gauge theories without fixing the gauge. The lattice quantization of gauge theories is performed in a way to preserve the compactness of the gauge group.

The continuum limit of lattice gauge theories is reproduced when the lattice spacing is many times smaller than a characteristic scale. This is achieved when the non-Abelian coupling constant tends to zero as it follows from the renormalization-group equation.

We consider in this Section the Euclidean formulation of lattice gauge theories. We first introduce the lattice terminology and discuss the action of lattice gauge theory at the classical level. Then, we quantize gauge fields on the lattice by the path-integral method, where the integration is over the invariant group measure. We explain Wilson's criterion of confinement and demonstrate it by calculations in the strong-coupling limit. Finally, we discuss how to pass to the continuum limit of lattice gauge theories.

3.1 Sites, links, plaquettes and all that

The first step in constructing a lattice gauge theory is to approximate the continuous space by a discrete set of points — a *lattice*. In the Euclidean formulation, the lattice is introduced along all four coordinates, while the time is left continuous in the Hamiltonian approach⁵. We shall discuss only the Euclidean formulation of lattice gauge theories.

The lattice is defined as a set of points of the d -dimensional Euclidean space with the coordinates

$$x_\mu = n_\mu a, \quad (3.1)$$

where the components of the vector

$$n_\mu = (n_1, n_2, \dots, n_d) \quad (3.2)$$

are integer numbers. The points (3.1) are called the lattice *sites*.

⁵A Hamiltonian formulation of lattice gauge theories was developed by Kogut and Susskind [9].

Fig. 4: Two-dimensional lattice with periodic boundary conditions. The sites labelled by the same numbers are identified. The lattice spacing equals a while the spatial size of the lattice corresponds to $L_1 = 6$ and $L_2 = 4$.

Fig. 5: A link of a lattice. The link connects the sites x and $x + a\hat{\mu}$.

The dimensional constant a , which is equal to the distance between the neighboring sites, is called the *lattice spacing*. Dimensional quantities are usually measured in the units of a , putting therefore $a = 1$.

A two-dimensional lattice is depicted in Fig. 4. A four-dimensional lattice for which the distances between sites are the same in all directions (as for the lattice in Fig. 4) is called the *hypercubic* lattice.

The next concept is the *link* of a lattice. A link is a line which connects two neighboring sites. A link is usually denoted by the letter l and is characterized by the coordinate x of its starting point and its direction $\mu = 1, \dots, d$:

$$l = \{x, \mu\}. \quad (3.3)$$

The link l connects sites with the coordinates x and $x + a\hat{\mu}$, where $\hat{\mu}$ is a unit vector along the μ -direction, as shown in Fig. 5. The lengths of all links are equal to a for a hypercubic lattice.

The elementary square enclosed by four links is called the *plaquette*. A plaquette p is specified by the coordinate x of a site and by the two directions μ and ν along which it is built on:

$$p = \{x; \mu, \nu\}. \quad (3.4)$$

A plaquette is depicted in Fig. 6. The set of four links which bound the plaquette p is

Fig. 6: A plaquette of a lattice. The plaquette boundary is made of four links.

denoted as ∂p .

If the spatial size of the lattice is infinite, then the number of dynamical degrees of freedom is also infinite (but enumerable). In order to limit the number of degrees of freedom, one deals with a lattice which has finite size $L_1 \times L_2 \cdots \times L_d$ in all directions (see Fig. 4).

Usually, one imposes *periodic boundary conditions* to reduce finite-size effects which are due to the finite extent of the lattice. In other words, one identifies pairs of sites which lie on parallel bounding hyperplanes. Usually the sites with the coordinates $(0, n_2, \dots, n_d)$ and $(L_1, n_2, \dots, n_d)$ are identified and similarly along other axes.

The numbers of sites, links and plaquettes for a symmetric ($L_1 = L_2 = \dots = L_d = L$) hypercubic lattice with periodic boundary conditions are

$$N_s = L^d, \quad N_l = dL^d, \quad N_p = \frac{d(d-1)}{2}L^d, \quad (3.5)$$

respectively.

3.2 Lattice formulation

The next step is to describe how matter fields and gauge fields are defined on a lattice.

A matter field, say a quark field, is attributed to the lattice sites. One can just think that a continuous field $\varphi(x)$ is approximated by its values at the lattice sites⁶:

$$\varphi(x) \Rightarrow \varphi_x. \quad (3.6)$$

It is clear that, in order for the lattice field φ_x to be a good approximation of a continuous field configuration $\varphi(x)$, the lattice spacing should be much smaller than a characteristic size of a given configuration. This is explained in Fig. 7.

The gauge field is attributed to the links of the lattice:

$$\mathcal{A}_\mu(x) \Rightarrow U_{x,\mu}. \quad (3.7)$$

It looks natural since a link is characterized by a coordinate and a direction (see Eq. (3.3)) — the same as $\mathcal{A}_\mu(x)$. Sometimes the notation $U_\mu(x)$ is alternatively used for $U_{x,\mu}$.

⁶We write arguments of functions which are defined on a lattice as subscripts keeping the standard notation for their continuum counterparts.

Fig. 7: Description of continuum field configurations by lattices a) “coarse” and b) “fine”. The lattice a) can represent the given continuum field configuration very roughly, while the lattice b) has the spacing which is small enough.

The link variable $U_{x,\mu}$ can be viewed as

$$U_{x,\mu} = \mathbf{P} e^{\int_x^{x+a\hat{\mu}} dz^\mu \mathcal{A}_\mu(z)}, \quad (3.8)$$

where the integral is along the link $\{x, \mu\}$. As $a \rightarrow 0$, this yields

$$U_{x,\mu} \rightarrow e^{a\mathcal{A}_\mu(x)} \quad (3.9)$$

so that $U_{x,\mu}$ is expressed via the exponential of the μ -th component of the vector potential, say, at the center of the link to agree with Eq. (2.46).

Since the path-ordered integral in Eq. (3.8) depends on the orientation, the concept of the orientation of a given link arises. The same link, which connects the points x and $x+a\hat{\mu}$, can be written either as $\{x, \mu\}$ or as $\{x+a\hat{\mu}, -\mu\}$. The orientation is positive for $\mu > 0$ in the former case (*i.e.* the same as the direction of the coordinate axis) and is negative in the latter case.

We have assigned the link variable $U_{x,\mu}$ to links with positive orientations. The U -matrices which are assigned to links with negative orientations are given by

$$U_{x+a\hat{\mu}, -\mu} = U_{x,\mu}^\dagger. \quad (3.10)$$

This is a one-link analog of Eq. (2.54).

It is clear from the relation (3.8) between the lattice and continuum gauge variables how to construct lattice analogues of the continuum phase factors — one should construct the contours from the links of the lattice.

An important role in the lattice formulation is played by the phase factor for the simplest closed contour on the lattice: the (oriented) boundary of a plaquette, as is shown in Fig. 8. The plaquette variable is composed from the link variables (3.8) as

$$U_{\partial p} = U_{x,\nu}^\dagger U_{x+a\hat{\nu},\mu}^\dagger U_{x+a\hat{\mu},\nu} U_{x,\mu}. \quad (3.11)$$

The link variable transforms under the gauge transformation, according to Eq. (2.56), as

$$U_{x,\mu} \xrightarrow{g,t} \Omega_{x+a\hat{\mu}} U_{x,\mu} \Omega_x^\dagger, \quad (3.12)$$

Fig. 8: A contour in the form of an oriented boundary of a plaquette.

where the matrix Ω_x is equal to the value of $\Omega(x)$ at the lattice site x . This defines the lattice gauge transformation.

The plaquette variable transforms under the lattice gauge transformation as

$$U_{\partial p} \xrightarrow{\text{g.t.}} \Omega_x U_{\partial p} \Omega_x^\dagger. \quad (3.13)$$

Therefore, its trace over the color indices is gauge invariant:

$$\text{tr} U_{\partial p} \xrightarrow{\text{g.t.}} \text{tr} U_{\partial p}. \quad (3.14)$$

The invariance of the trace under the lattice gauge transformation is used in constructing an action of a lattice gauge theory. The simplest (Wilson) action is

$$S_{\text{lat}}[U] = \sum_p \left(1 - \frac{1}{N_c} \text{Re tr} U_{\partial p} \right). \quad (3.15)$$

The summation is over all the elementary plaquettes of the lattice (*i.e.* over all x , μ , and ν), regardless of their orientations.

Since a reversal of the orientation of the plaquette boundary results, according to Eq. (2.54), in complex conjugation:

$$\text{tr} U_{\partial p} \xrightarrow{\text{reorient}} \text{tr} U_{\partial p}^\dagger = (\text{tr} U_{\partial p})^*, \quad (3.16)$$

one can rewrite the action (3.15) in the equivalent form

$$S_{\text{lat}}[U] = \frac{1}{2} \sum_{\text{oriented } p} \left(1 - \frac{1}{N_c} \text{tr} U_{\partial p} \right), \quad (3.17)$$

where the sum is also over the two possible orientations of the boundary of a given plaquette.

In the limit $a \rightarrow 0$, the lattice action (3.15) becomes in $d = 4$ the action of a continuum gauge theory. In order to show this, let us first note that

$$U_{\partial p} \rightarrow \exp \left[a^2 \mathcal{F}_{\mu\nu}(x) + \mathcal{O}(a^3) \right], \quad (3.18)$$

where $\mathcal{F}_{\mu\nu}(x)$ is defined by Eq. (2.34).

In the Abelian theory, the expansion (3.18) is easily found from the Stokes theorem. The commutator of $\mathcal{A}_\mu(x)$ and $\mathcal{A}_\nu(x)$, which arises in the non-Abelian case, complements the field strength to the non-Abelian one, as is ensured by the gauge invariance.

The transition to the continuum limit is performed by virtue of

$$a^4 \sum_p \xrightarrow{a \rightarrow 0} \frac{1}{2} \int d^4x \sum_{\mu, \nu}. \quad (3.19)$$

Expanding the exponential on the RHS of Eq. (3.18) in a , we get

$$S_{\text{lat}} \xrightarrow{a \rightarrow 0} -\frac{1}{4N_c} \int d^4x \sum_{\mu, \nu} \text{tr} \mathcal{F}_{\mu\nu}^2(x), \quad (3.20)$$

which coincides modulo a factor with the action of the continuum gauge theory.

Remark on the naive continuum limit

The limit $a \rightarrow 0$, when Eqs. (3.9) and (3.18) hold reproducing the continuum action (3.20), is called the *naive continuum limit*. It is assumed in the naive continuum limit that $\mathcal{A}_\mu(x)$ is weakly fluctuating at neighboring lattice links. Fluctuations of the order of $1/a$ are not taken into account, since discontinuities of the vector potential in the continuum theory are usually associated with an infinite action.

Another subtlety with the naive continuum limit is that next order in a terms of the expansion of the lattice action (3.15), say the term $\propto a^2 \text{tr} \mathcal{F}^3$, are associated with non-renormalizable interactions and the smallness of a^2 can be compensated, in principle, by quadratic divergencies.

The actual continuum limit of lattice gauge theories is in fact very similar to the naive one modulo some finite renormalizations of the gauge coupling constant. The large fluctuations of $\mathcal{A}_\mu(x)$ of the order of $1/a$ become frozen when passing to the continuum limit. How to pass to the continuum limit of lattice gauge theories is explained in Subsect. 3.7 below.

Remark on ambiguities of the lattice action

The Wilson action (3.15) is the simplest one which reproduces the continuum action in the naive continuum limit. One can alternatively use characters of $U_{\partial p}$ in other representations of $\text{SU}(N_c)$, *e.g.* in the adjoint representation

$$\chi_{\text{adj}}(U) = |\text{tr} U|^2 - 1, \quad (3.21)$$

to construct the lattice action.

The adjoint-representation lattice action reads

$$S_{\text{adj}}[U] = \sum_p \left(1 - \frac{1}{N_c^2} |\text{tr} U_{\partial p}|^2 \right). \quad (3.22)$$

The naive continuum limit will be the same as for the Wilson action (3.15).

Moreover, one can define the lattice action as a mixture of the fundamental and adjoint representations [10, 11]:

$$S_{\text{mixed}}[U] = \sum_p \left(1 - \frac{1}{N_c} \text{Re tr } U_{\partial p} \right) + \frac{\beta_A}{2\beta} \sum_p \left(1 - \frac{1}{N_c^2} |\text{tr } U_{\partial p}|^2 \right). \quad (3.23)$$

The ratio β_A/β is a constant ~ 1 which does not affect the continuum limit. This action is called the *mixed action*.

The lattice action (3.22) for $N_c = 2$ is associated with the action of the $\text{SO}(3)$ lattice gauge theory. Since algebras of the $\text{SU}(2)$ and $\text{SO}(3)$ groups coincide, these two gauge theories coincide in the continuum while they differ on the lattice.

One more possibility is to use the phase factor associated, say, with the boundary of two plaquettes having a common link, or the phase factors for more complicated closed contours of finite size on the lattice. These actions will also reproduce, in the naive continuum limit, the action of the continuum gauge theory.

The independence of the continuum limit of lattice gauge theories on the choice of lattice actions is called the *universality*. We shall say more about this in the next Section when discussing renormalization group on the lattice.

3.3 The Haar measure

The partition function of a pure⁷ lattice gauge theory is defined by

$$Z(\beta) = \int \prod_{x,\mu} dU_{x,\mu} e^{-\beta S[U]}, \quad (3.24)$$

where the action is given by Eq. (3.15).

This is an analog of a partition function in statistical mechanics at an inverse temperature β given by

$$\beta = \frac{2N_c}{g^2}. \quad (3.25)$$

This formula results from Eq. (3.20).

A subtle question is what is the measure $dU_{x,\mu}$ in Eq. (3.24). To preserve the gauge invariance at finite lattice spacing, the integration is over the *Haar measure* which is an invariant group measure. Invariance of the Haar measure under multiplication by an arbitrary group element from the left or from the right:

$$dU = d(\Omega U) = d(U \Omega'), \quad (3.26)$$

guarantees the gauge invariance of the partition function (3.24).

This invariance of the Haar measure is crucial for the Wilson formulation of lattice gauge theories.

⁷Here “pure” means without matter fields.

It is instructive to present an explicit expression for the Haar measure in the case of the $\text{SU}(2)$ gauge group. An element of $\text{SU}(2)$ can be parametrized by the unit four-vector a_μ ($a_\mu^2 = 1$) as

$$U = a_4 \mathbf{1} + i \vec{a} \vec{\sigma} \quad (3.27)$$

with $\vec{\sigma}$ being the Pauli matrices. The Haar measure for $\text{SU}(2)$ then reads

$$dU = \frac{1}{\pi^2} \prod_{\mu=1}^4 da_\mu \delta^{(1)}(a_\mu^2 - 1), \quad (3.28)$$

since $\det U = a_\mu^2$.

An element of $\text{SU}(2)$ can be alternatively parametrized by a unit three-vector $\vec{n}$ ($\vec{n}^2 = 1$) and an angle φ ($\varphi \in [0, 2\pi]$):

$$U = e^{i\varphi \vec{n} \vec{\sigma}/2} = \cos \frac{\varphi}{2} + i \vec{n} \vec{\sigma} \sin \frac{\varphi}{2}. \quad (3.29)$$

The geometric meaning of this parametrization is simple: the element (3.29) is associated with a rotation through the angle φ around the $\vec{n}$ -axis. The Haar measure for the $\text{SU}(2)$ group then is

$$dU = \frac{d^2 \vec{n}}{4\pi} \frac{d\varphi}{\pi} \sin^2 \frac{\varphi}{2}. \quad (3.30)$$

This formula can be obtained from Eq. (3.28) by integrating over $|\vec{a}|$.

For the $\text{U}(N_c)$ group, the Haar measure can be represented as a multiple integral over the elements of a general complex matrix U :

$$\int dU \dots = \int_{-\infty}^{\infty} \prod_{i,j} d\text{Re } U_{ij} d\text{Im } U_{ij} \delta^{(N_c^2)}(U U^\dagger - 1) \dots \quad (3.31)$$

The integral in this formula goes over unrestricted U_{ij} while the delta-function restricts U to be unitary.

The partition function (3.24) characterizes vacuum effects in the quantum theory. Physical quantities are given by the averages of the type

$$\langle F[U] \rangle = Z^{-1}(\beta) \int \prod_{x,\mu} dU_{x,\mu} e^{-\beta S[U]} F[U], \quad (3.32)$$

where $F[U]$ is a gauge invariant functional of the link variable $U_{x,\mu}$. The averages (3.32) become the corresponding expectation values in the continuum theory as $a \rightarrow 0$ and β is related to g^2 by Eq. (3.25).

Remark on the lattice quantization

On a lattice of finite size, the integral over the gauge group in Eq. (3.32) is finite since the integration is over a compact group manifold, in contrast to the continuum case, where the volume of the gauge group is infinite. Therefore, the expression (3.32) is a constructive method for calculating averages of gauge invariant quantities, though the gauge is not fixed.

The gauge can be fixed on the lattice in the standard way by the Faddeev–Popov method. This procedure involves extracting a (finite) common factor, which equals to the volume of the gauge group, from the numerator and denominator on the RHS of Eq. (3.32). Therefore, the averages of gauge invariant quantities coincide for a fixed and unfixed gauge, while the average of a functional which is not gauge invariant vanishes when the gauge is not fixed.

The fixing of gauge is convenient (though not necessary) for calculations in a lattice perturbation theory. A Lorentz gauge can not be fixed, however, outside perturbation theory because of Gribov copies [12]. In contrast, the lattice path integral (3.32) with an unfixed gauge is a method of non-perturbative quantization.

A price for the compactness of the group manifold on the lattice is the presence of fluctuations $\mathcal{A}_\mu(x) \sim 1/a$ which do not occur in the continuum (say, the values of the vector potential A_μ and $A_\mu + 2\pi/ae$ are identified for the Abelian U(1) group). However, these fluctuations become unimportant when passing to the continuum limit.

3.4 Wilson loops

As is already mentioned in Subsect. 3.2, lattice phase factors are associated with contours which are drawn on the lattice.

In order to write down an explicit representation of the phase factor on the lattice via the link variables, let us specify the (lattice) contour C by its initial point x and by the directions (some of which may be negative) of the links which the contour is built of:

$$C = \{x; \mu_1, \dots, \mu_n\}. \quad (3.33)$$

Then the lattice phase factor U_C reads as

$$U_C = U_{x+a\hat{\mu}_1+\dots+a\hat{\mu}_{n-1}, \mu_n} \cdots U_{x+a\hat{\mu}_1, \mu_2} U_{x, \mu_1}. \quad (3.34)$$

For the links with a negative direction it is again convenient to use Eq. (3.10).

A closed contour has $\hat{\mu}_1 + \dots + \hat{\mu}_n = 0$. The trace of the phase factor for a closed contour, which is gauge invariant, is called the *Wilson loop*.

The average of the Wilson loop is determined by the general formula (3.32) to be

$$W_C \equiv \left\langle \frac{1}{N_c} \text{tr} U_C \right\rangle = Z^{-1}(\beta) \int \prod_{x, \mu} dU_{x, \mu} e^{-\beta S[U]} \frac{1}{N_c} \text{tr} U_C. \quad (3.35)$$

This average is often called the Wilson loop average.

A very important role in lattice gauge theories is played by the averages of the Wilson loops associated with rectangular contours. Such a contour lying in the x, t plane is depicted in Fig. 9.

Fig. 9: Rectangular loop of the size $R \times T$.

This Wilson loop average is related for $T \gg R$ to the energy of the interaction of the static (*i.e.* infinitely heavy) quarks which are separated by the distance R by the formula

$$W_{R \times T} \stackrel{T \gg R}{\sim} e^{-E_0(R)T}. \quad (3.36)$$

Equation (3.36) can be derived by fixing the axial gauge $\mathcal{A}_4 = 0$ where $U_{x,4} = 1$, so that only vertical segments of the rectangle in Fig. 9 contribute to $U_{R \times T}$. Denoting

$$\Psi_{ij}(t) \equiv \left[\mathbf{P} e^{\int_0^t dz_1 \mathcal{A}_1(z_1, \dots, t)} \right]_{ij}, \quad (3.37)$$

we then have

$$W_{R \times T} = \left\langle \frac{1}{N_c} \text{tr} \Psi_{ij}(0) \Psi_{ij}^\dagger(T) \right\rangle. \quad (3.38)$$

Inserting in the last equation a sum over a complete set of intermediate states

$$\sum_n |n\rangle \langle n| = 1, \quad (3.39)$$

we get

$$W_{R \times T} = \sum_n \frac{1}{N} \text{tr} \langle \Psi_{ij}(0) | n \rangle \langle n | \Psi_{ij}^\dagger(T) \rangle = \sum_n \frac{1}{N} |\langle \Psi_{ij}(0) | n \rangle|^2 e^{-E_n T} \quad (3.40)$$

where E_n is the energy of the state $|n\rangle$. As $T \rightarrow \infty$, only the ground state with the lowest energy survives in the sum over states and we finally find

$$W_{R \times T} \xrightarrow{\text{large } T} e^{-E_0 T}, \quad (3.41)$$

which results in Eq. (3.36).

Notice that nothing relies in this derivation on the lattice. Therefore, Eq. (3.36) holds for a rectangular loop in a continuum theory as well.

Equation (3.36) can also be understood as follows. Let us consider the Abelian case when the interaction is described by the Coulomb law. The contour integral can then be rewritten as the integral over the whole space

$$e \oint_C dz^\mu A_\mu(z) = \int d^d x J^\mu(x) A_\mu(x), \quad (3.42)$$

where

$$J^\mu(x) = e \oint_C dz^\mu \delta^{(d)}(x - z) \quad (3.43)$$

is a four-vector current of a classical particle moving along the trajectory C which is described by the function $z_\mu(t)$.

It is clear that

$$-\ln W(C) = -\ln \left\langle e^{i \int d^4x J^\mu(x) A_\mu(x)} \right\rangle \quad (3.44)$$

determines a change of the action of the classical particle due to electromagnetic interaction in accordance with Eq. (3.36).

A similar interpretation of Eq. (3.36) in the non-Abelian case is somewhat more complicated. For a heavy particle moving along some trajectory in space-time, color degrees of freedom are quantum and easily respond to changes of the gauge field $A_\mu(x)$, which interacts with them. Let us suppose that a quark and an antiquark are created at the same space-time point in some color state. Then this state must be singlet with respect to color (or *colorless*) since the average over the gauge field would vanish otherwise. When quarks are going away, their color changes from one point to another simultaneously with changing the color of the gauge field, in order for the system of the quarks plus the gauge field to remain colorless. Therefore, the averaging over the gauge field leads to an averaging over fluctuations of quark color degrees of freedom. $E_0(R)$ in Eq. (3.36) is associated with the interaction energy averaged over color in this way.

Remark on mass renormalization

By definition, $E_0(R)$ in Eq. (3.36) includes a renormalization of the mass of a heavy quark due to the interaction with the gauge field and which is thus independent of R . To the first order in g^2 , it is the same as in QED and reads

$$\Delta E_{\text{mass}} = \frac{g^2}{4\pi a} \frac{N_c^2 - 1}{2N_c} \quad (3.45)$$

as $a \rightarrow 0$.

The potential energy of the interaction between the static quarks is therefore defined as the difference

$$E(R) = E_0(R) - \Delta E_{\text{mass}}. \quad (3.46)$$

If $g^2/4\pi a$ in ΔE_{mass} did not become infinite as $a \rightarrow 0$, the term resulting from the mass renormalization would not have to be subtracted, since it simply changes the reference level for the potential energy.

3.5 Strong coupling expansion

We already mentioned in Subsect. 3.3 that the path integral (3.32) can be calculated by the lattice perturbation theory in g^2 . As was pointed out by Wilson [1], there exists an alternative way of evaluating the same quantity on a lattice by an expansion in $1/g^2$ or in β since they are related by Eq. (3.25). This expansion is called the *strong coupling* expansion. It is an analog of the high temperature expansion in statistical mechanics since β is an analog of an inverse temperature.

In order to perform the strong coupling expansion, we expand the exponential of the lattice action, say in Eq. (3.35), in β . Then the problem is to calculate the integrals over the unitary group of the form

$$I_{j_1 \dots j_m, l_1 \dots l_n}^{i_1 \dots i_m, k_1 \dots k_n} = \int dU U_{j_1}^{i_1} \dots U_{j_m}^{i_m} U_{l_1}^{k_1} \dots U_{l_n}^{k_n}. \quad (3.47)$$

where the Haar measure (given for $SU(2)$ by Eq. (3.28)) is normalized by

$$\int dU = 1. \quad (3.48)$$

It is clear from general arguments that the integral (3.47) is nonvanishing only if $n = m \pmod{N_c}$, i.e. only if $n = m + kN_c$ where k is integer.

For the simplest case $m = n = 1$, the answer can easily be found by using the unitarity of U 's and the orthogonality relation:

$$\int dU U_j^i U_l^{\dagger k} = \frac{1}{N_c} \delta_l^i \delta_j^k. \quad (3.49)$$

The simplest Wilson loop average, which is nonvanishing in the strong coupling expansion, is the one for the loop which coincides with the boundary of a plaquette (see Fig. 8). It is called the plaquette average and is denoted as

$$W_{\partial p} = \left\langle \frac{1}{N_c} \text{tr} U_{\partial p} \right\rangle. \quad (3.50)$$

In order to calculate the plaquette average to order β , it is sufficient to retain only the terms $\mathcal{O}(\beta)$ in the expansion of the exponentials in Eq. (3.35):

$$W_{\partial p} = \frac{\int \prod_{x,\mu} dU_{x,\mu} \left(1 + \beta \sum_{\nu'} \frac{1}{N_c} \text{Re tr} U_{\partial p'} \right) \frac{1}{N_c} \text{tr} U_{\partial p}}{\int \prod_{x,\mu} dU_{x,\mu} \left(1 + \beta \sum_{\nu'} \frac{1}{N_c} \text{Re tr} U_{\partial p'} \right)} + \mathcal{O}(\beta^2). \quad (3.51)$$

The group integration can then be performed remembering that

$$\int dU_{x,\mu} [U_{x,\mu}]_j^i [U_{y,\nu}]_l^k = \frac{1}{N_c} \delta_{xy} \delta_{\mu\nu} \delta_l^i \delta_j^k \quad (3.52)$$

at different links.

Fig. 10: Boundaries of the plaquettes p and p' with the opposite orientations: ∂p and $\partial p'$, respectively.

Using this property of the group integral in Eq. (3.51), we immediately see that the denominator is equal to one (each link is encountered no more than once), while the only nonvanishing contribution in the numerator is from the plaquette p' , which coincides with p but has the opposite orientation as is depicted in Fig. 10.

It is convenient to use the graphic notations⁸ for Eq. (3.49) at each link of ∂p :

$$\begin{array}{c} i \\ \leftarrow \\ \text{---} \\ \rightarrow \\ j \\ k \end{array} = \frac{1}{N_c} \times \left(\begin{array}{c} i \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ j \\ k \end{array} \right), \quad (3.53)$$

where the semicircles are associated with the Kronecker deltas:

$$\begin{array}{c} i \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ \text{---} \\ j \\ k \end{array} = \delta_i^j. \quad (3.54)$$

This notation is convenient since the lines which denote the deltas in the last equation can be associated with propagation of the color indices. Analogously a closed line represents the contracted delta, which is summed up over the color indices,

$$\bigcirc = \delta_i^i = N_c. \quad (3.55)$$

Using the graphic representation (3.53) for each of the four links depicted in Fig. 10, we get

$$\int \prod_{x,\mu} dU_{x,\mu} \text{tr} U_{\partial p} \text{tr} U_{\partial p'}^\dagger = \frac{1}{N_c^4} \times \begin{array}{cc} \bigcirc & \bigcirc \\ \bigcirc & \bigcirc \end{array} = 1, \quad (3.56)$$

where the contracted Kronecker deltas are associated with the four sites of the plaquette.

The final answer for the plaquette average is

$$\begin{aligned} W_{\partial p} &= \frac{\beta}{2N_c^2} && \text{for } \text{SU}(N_c) \text{ with } N_c \geq 3, \\ W_{\partial p} &= \frac{\beta}{4} && \text{for } \text{SU}(2). \end{aligned} \quad (3.57)$$

⁸A calculation of more complicated group integrals (3.47) in the graphic notations is discussed in the lectures by Wilson [13] and in the Chapter 8 of the book by Creutz [14]. An alternative way of calculating the group integrals by the character expansion is described in the review by Drouffe and Zuber [15].

Fig. 11: Filling of a loop with elementary plaquettes.

The result for $\text{SU}(2)$ differs by a factor of $1/2$ because $\text{tr} U_{\partial p}$ is real for $\text{SU}(2)$ so that the orientation of the plaquettes can be ignored.

The graphic representation (3.53) is useful for evaluating the leading order of the strong coupling expansion for more complicated loops. According to Eq. (3.52), a nonvanishing result emerges only when plaquettes, coming from the expansion of the exponentials of Eq. (3.35) in β , completely cover a surface enclosed by the given loop C as is depicted in Fig. 11. In this case each link is encountered twice (or never), once in the positive direction and once in the negative direction, so that all the group integrals are nonvanishing. The leading order in β corresponds to filling a *minimal surface*, whose area takes on the smallest possible value. This yields

$$W_C = (W_{\partial p})^{A_{\min}(C)}, \quad (3.58)$$

where $W_{\partial p}$ is given by Eq. (3.57) and $A_{\min}(C)$ is the area (in units of a^2) of the minimal surface.

For the rectangular loop, which is depicted in Fig. 9, the minimal surface is just a piece of the plane bounded by the rectangle. Therefore, we get

$$W_{R \times T} = (W_{\partial p})^{RT} \quad (3.59)$$

to the leading order in β .

More complicated surfaces, which do not lie in the plane of the rectangle, will give the contribution to W_C of the order of $\beta^{A_{\text{rea}}}$. They are suppressed at small β since their areas are larger than $A_{\min}$.

3.6 Area law and confinement

The exponential dependence of the Wilson loop average on the area of the minimal surface (as in Eq. (3.58)) is called the *area law*. It is customarily assumed that if an area law holds for loops of large area in pure gluodynamics (i.e. in the pure $\text{SU}(3)$ gauge theory) then quarks are confined. In other words, there are no physical $|in\rangle$ or $|out\rangle$ quark states. This is the essence of Wilson's *confinement criterion*. The argument is that physical amplitudes, e.g. the polarization operator, do not have quark singularities when the Wilson criterion is satisfied. I refer the reader to the well-written original paper by Wilson [1], where this point is clarified.

Fig. 12: Lines of force between static quarks for a) linear and b) Coulomb interaction potentials. For the linear potential the lines of force are contracted into a tube, while they are distributed over the whole space for the Coulomb one.

Another, somewhat oversimplified, justification for the Wilson criterion is based on the relationship (3.36) between the Wilson loop average and the potential energy of interaction between static quarks. When the area law

$$W(C) \xrightarrow{\text{large } C} e^{-K A_{\min}(C)} \quad (3.60)$$

holds for large loops, the potential energy is a linear function of the distance between the quarks:

$$E(R) = KR. \quad (3.61)$$

The coefficient K in these formulas is called the *string tension* because the gluon field between quarks contracts to a tube or string, whose energy is proportional to its length, as is depicted in Fig. 12a. The value of K is the energy of the string per unit length. This string is stretched with the distance between quarks and prevents them from moving apart to macroscopic distances.

Equation (3.59) gives

$$K = \frac{1}{a^2} \ln \frac{2N_c^2}{\beta} = \frac{1}{a^2} \ln (N_c g^2) \quad (3.62)$$

for the string tension to the leading order of the strong coupling expansion. Next orders of the strong coupling expansion results in corrections in β to this formula.

Therefore, confinement holds in the lattice gauge theory to any order of the strong coupling expansion.

Remark on the perimeter law

For the Coulomb potential

$$E(R) = -\frac{g^2}{4\pi R} \frac{N_c^2 - 1}{2N_c}, \quad (3.63)$$

the gauge field between quarks would be distributed over the whole space as is depicted in Fig. 12b. The Wilson loop average would have the behavior

$$W(C) \xrightarrow{\text{large } C} e^{-\text{const } L(C)}, \quad (3.64)$$

where $L(C)$ stands for the length (or perimeter) of the closed contour C .

This behavior of the Wilson loops is called the *perimeter law*. To each order of perturbation theory, it is the perimeter law (3.64), rather than the area law (3.60), that holds for the Wilson loop averages. A perimeter law corresponds to a potential which can not confine quarks.

Remark on the Creutz ratio

To distinguish between the area and perimeter law behavior of the Wilson loop averages, Creutz [16] proposed to consider the ratio

$$\chi(I, J) = -\ln \frac{W_{I \times J} W_{(I-1) \times (J-1)}}{W_{(I-1) \times J} W_{I \times (J-1)}}, \quad (3.65)$$

where $W_{I \times J}$ is as before the average of a rectangular Wilson loop of the size $I \times J$. The exponentials of the perimeter, which is equal to

$$L(I \times J) = 2I + 2J, \quad (3.66)$$

cancel out in the ratio (3.65). In particular, the mass renormalization (3.45) cancels out, which is essential for the continuum limit.

The Creutz ratio (3.65) has a meaning of a force of interaction between quarks, which can be seen by stretching the rectangle along the “temporal” axis (as is illustrated by Fig. 9). If the area law (3.60) holds for asymptotically large I and J , then

$$\chi(I, J) \xrightarrow{\text{large } I, J} a^2 K, \quad (3.67)$$

i.e. does not depend on I or J and coincides with the string tension. This property of the Creutz ratio was used for numerical calculations of the string tension.

3.7 Asymptotic scaling

Equation (3.62) establishes the relationship between values of the lattice spacing a and the coupling g^2 as follows. Let us set K to be equal to its experimental value⁹

$$K = (400 \text{ MeV})^2 \approx 1 \text{ GeV/fm}. \quad (3.68)$$

⁹This value results from the string model of hadrons where the slope of the Regge trajectory α' and the string tension K are related by $K = 1/2\pi\alpha'$. This formula holds even for a classical string. The slope $\alpha' = 1 \text{ GeV}^{-2}$ say from the $\rho - A_2 - g$ trajectory. A similar value of K is found from the description of mesons made out of heavy quarks by a nonrelativistic potential model.

Fig. 13: Dependence of the string tension on $1/g^2$. The strong coupling formula (3.62) holds for small $1/g^2$. The asymptotic scaling formula (3.69) sets in for large $1/g^2$. Both formulas are not applicable in the intermediate region of $1/g^2 \sim 1$ which is depicted by the dashed line.

Then the renormalizability prescribes that variations of a , which plays the role of a lattice cutoff, and of the bare charge g^2 should be done simultaneously in order that K does not change.

Given Eq. (3.62), this procedure calls for $a \rightarrow \infty$ as $g^2 \rightarrow \infty$. In other words, the lattice spacing is large in the strong coupling limit, compared to 1 fm — the typical scale of the strong interaction. This is a situation of the type shown in Fig. 7a. Such a coarse lattice can not describe correctly the continuum limit and, in particular, the rotational symmetry.

In order to pass to the continuum, the lattice spacing a should be decreased to have a picture like that in Fig. 7b. Equation (3.62) shows that a decreases with decreasing g^2 . However, this formula ceases to be applicable in the intermediate region of $g^2 \sim 1$ and, therefore, $a \sim 1$ fm.

The recipe for further decreasing of a is the same as in the strong coupling region: further decreasing of g^2 . While no analytic formulas are available at intermediate values of g^2 , the expected relation between a and g^2 for small g^2 is predicted by the known two-loop Gell-Mann-Low function of QCD.

For pure SU(3) gluodynamics, Eq. (3.62) is replaced at small g^2 by

$$K = \text{const} \cdot \frac{1}{a^2} \left(\frac{16\pi^2}{11g^2} \right)^{\frac{102}{121}} e^{-\frac{16\pi^2}{11g^2}}. \quad (3.69)$$

The exponential dependence of K on $1/g^2$ is called *asymptotic scaling*. An asymptotic scaling sets in for some value of $1/g^2$ as is depicted in Fig. 13. For such values of g^2 , where an asymptotic scaling holds, the lattice gauge theory enjoys a continuum limit.

The knowledge of the two asymptotics says nothing about the behavior of $a^2 K$ in the intermediate region of $g^2 \sim 1$. There can be either a smooth transition between these two

regimes or a phase transition. Numerical methods were introduced to study this problem. Some of them are described in the next Section.

Remark on dimensional transmutation

The QCD action (2.33) does not contain a dimensional parameter of the order of hundreds MeV. The masses of the light quarks are of the order of a few MeV and can be disregarded. The only parameter of the action is the dimensionless bare coupling constant g^2 . At the classical level, there is no way to get the dimensional parameter of the order of hundreds MeV.

In quantum theory, there is always a dimensional cutoff (like a for the lattice regularization). The renormalizability says that a and g^2 are not independent but are related by the Gell-Mann-Low equation

$$-\frac{a}{da} \frac{dg^2}{g^2} = B(g^2) \quad (3.70)$$

whose RHS involves the Gell-Mann-Low function $B(g^2)$ which is a function solely of g^2 — the bare charge. Equation (3.70) can be integrated to give the integration constant

$$\Lambda_{\text{QCD}} = \frac{1}{a} \exp \left[- \int \frac{dg^2}{B(g^2)} \right]. \quad (3.71)$$

Up to now there was no difference between QCD and QED. The difference stems from the fact that the Gell-Mann-Low function $B(g^2)$ is positive for QED and negative for QCD. In QED $e^2(a)$ increases with decreasing a while in QCD $g^2(a)$ decreases with decreasing a . The latter behavior of the coupling constant is called *asymptotic freedom*. In both cases the Gell-Mann-Low function vanishes when the coupling constant tends to zero. Such values of coupling constants where the Gell-Mann-Low function vanishes are called the *fixed points*. Since the infrared behavior of e^2 in QED is interchangeable with the ultraviolet behavior of g^2 in QCD, the origin is an infrared-stable fixed point in QED and an ultraviolet-stable fixed point in QCD. In QED the fine-structure constant ($\approx 1/137$) is measurable in experiments while in QCD the constant Λ_{QCD} is measurable.

This phenomenon of an appearance of a dimensional parameter in QCD, which remains finite in the limit of vanishing cutoff, is called *dimensional transmutation*. All observable dimensional quantities, say the string tension or hadron masses, are proportional to the corresponding powers of Λ_{QCD} . Therefore, their dimensionless ratios, say the ratio of $\sqrt{K}$ to hadron masses, are universal numbers which do not depend on g^2 . The goal of a non-perturbative approach in QCD is to calculate these numbers but not the overall dimensional parameter.

Remark on second-order phase transition

It is usually said in statistical physics that continuum limits of a lattice system are reached at the points of second-order phase transitions when the correlation length becomes infinite

in lattice units. This statement is in perfect agreement with what is said above about the continuum limit of lattice gauge theories.

A correlation length is given by a formula similar to Eq. (3.71):

$$\xi \sim \Lambda_{\text{QCD}}^{-1} = a \exp \left[\int \frac{dg^2}{B(g^2)} \right]. \quad (3.72)$$

The only chance for the RHS to diverge is to have a zero of the Gell-Mann–Low function $B(g^2)$ at some fixed point $g^2 = g_*^2$. Therefore, the bare coupling should approach the fixed-point value g_*^2 to describe the continuum.

As we have discussed, $B(0) = 0$ for a non-Abelian gauge theory so that $g_*^2 = 0$ is a fixed-point value of the coupling constant. Therefore, the continuum limit is associated with $g^2 \rightarrow 0$ as is said above.

4 Lattice methods

Analytic calculations of observables in the non-Abelian lattice gauge theories are available only in the strong coupling regime $g^2 \rightarrow \infty$, while one needs $g^2 \rightarrow 0$ for the continuum limit. When g^2 is decreased, the lattice systems can undergo phase transitions as it often happens in statistical mechanics.

To look for phase transitions, the mean-field method was first applied to lattice gauge theories [1, 17]. It turned out to be useful for studying the first-order phase transitions which very often happen in lattice gauge systems but do not affect the continuum limit.

The second-order phase transitions are better described by the lattice renormalization group method. The approximate Migdal–Kadanoff recursion relations [18, 19] were the first implementation of the renormalization group transformation on a lattice, which indicated the absence of a second-order phase transition in the non-Abelian lattice gauge theories and, therefore, quark confinement.

A most powerful method for practical non-perturbative calculations of observables in lattice gauge theories is the numerical Monte Carlo method. This method simulates statistical processes in a lattice gauge system and is often called for this reason the numerical simulation. The idea of applying it to lattice gauge theories is due to Wilson [20] while the practical implementation was done by Creutz, Jacobs and Rebbi [21] for Abelian gauge groups and by Creutz [22, 16] for the SU(2) and SU(3) groups.

We briefly describe in this Section the mean-field method, the lattice renormalization group method and the Monte Carlo method. Few results of the Monte Carlo simulations will be discussed.

4.1 Phase transitions

As is pointed out in the previous Section, analytic calculations of the string tension are available only in the strong coupling regime $g^2 \rightarrow \infty$, while one needs $g^2 \rightarrow 0$ for the continuum limit. A question arises what happens with lattice systems when g^2 is decreased. In particular, does an actual picture of the dependence of the string tension on g^2 look like that in Fig. 13?

We know from statistical mechanics that lattice systems can undergo phase transitions with changing the parameters, say the temperature, which completely alter macroscopic properties. A simplest example is that of a first-order phase transition which occurs in a teapot.

First-order phase transitions very often happen in the lattice gauge theories. They are usually seen as a discontinuity in the β (or $1/g^2$)-dependence of the plaquette average (3.50) as is depicted in Fig. 14. The form of $W_{\text{op}}(\beta)$ at small β is given to the leading order of the strong coupling expansion by Eq. (3.57), while that at large β is prescribed by the lattice perturbation theory¹⁰ to be

$$W_{\text{op}}(\beta) = 1 - \frac{d_G}{\beta d} + \mathcal{O}(\beta^{-2}), \quad (4.1)$$

¹⁰It is often called, for obvious reasons, the weak coupling expansion.

Fig. 14: Typical β -dependence of the plaquette average for a first-order phase transition which occurs at $\beta = \beta_*$.

where d_G is the dimensionality of the gauge group ($d_G = N_c^2 - 1$ for $SU(N_c)$, $d_G = N_c^2$ for $U(N_c)$) and d is the dimensionality of the lattice as before.

This behavior of the plaquette average is quite analogous to a dependence of the internal energy per unit volume (called the specific energy) in statistical systems. In order to see an analogy between the specific energy and $(1 - W_{\partial p})$, let us remember that β is an analog of the inverse temperature and rewrite Eq. (3.50) as

$$W_{\partial p}(\beta) = 1 + \frac{1}{N_p} \frac{\partial}{\partial \beta} \ln Z(\beta), \quad (4.2)$$

where the partition function is given by Eq. (3.24) and the number of plaquettes N_p is an analog of the volume of a statistical system.

The first-order phase transitions of the type in Fig. 14 are usually harmless and are not associated with deconfinement. They are related with dynamics of some lattice degrees of freedom (say, with large fluctuations of the link variable $U_{x,\mu}$ which occur independently at adjacent links) which do not affect the continuum limit and are called *lattice artifacts*. Moreover, these lattice degrees of freedom become frozen for $\beta > \beta_*$ which is necessary for the continuum limit to exist. A brief description of some lattice artifacts and of the related phase transitions can be found in the Section 3 of the review article [23].

Another possibility for a lattice system is to undergo a second-order phase transition in analogy with spin systems. In this case $W_{\partial p}$ is continuous but the derivative $\partial W_{\partial p} / \partial \beta$ becomes infinite at the critical point $\beta = \beta_*$ as is depicted in Fig. 15. Given Eq. (4.2), this derivative is to be considered as an analog of the specific heat of statistical systems. Its behavior at small and large β is governed by Eqs. (3.57) and (4.1), respectively.

Differentiating Eq. (3.50) with respect to β , the derivative $\partial W_{\partial p} / \partial \beta$ can be expressed via the sum of the connected correlators:

$$\frac{\partial W_{\partial p}}{\partial \beta} = \frac{1}{2} \sum_{\text{oriented } p'} \left\langle \frac{1}{N_c} \text{tr} U_{\partial p} \frac{1}{N_c} \text{tr} U_{\partial p'} \right\rangle_{\text{conn}}. \quad (4.3)$$

This formula shows also that $\partial W_{\partial p} / \partial \beta$ is positive definite, since the RHS can be rewritten

Fig. 15: Typical β -dependence of $\partial W_{\partial p} / \partial \beta$ for a second-order phase transition which occurs at $\beta = \beta_*$.

using translational invariance as

$$\begin{aligned} & \frac{1}{2} \sum_{\text{oriented } p'} \left\langle \frac{1}{N_c} \text{tr} U_{\partial p} \frac{1}{N_c} \text{tr} U_{\partial p'} \right\rangle_{\text{conn}} \\ &= \frac{1}{4N_p} \left\langle \left(\sum_{\text{oriented } p} \frac{1}{N_c} \text{tr} U_{\partial p} \right)^2 \right\rangle - \frac{1}{4N_p} \left(\left\langle \sum_{\text{oriented } p} \frac{1}{N_c} \text{tr} U_{\partial p} \right\rangle \right)^2 \geq 0, \end{aligned} \quad (4.4)$$

where the equality is possible only for a Gaussian averaging, *i.e.* for a free theory. This repeats the standard proof of the positivity of specific heat in statistical mechanics.

Since each term of the sum in Eq. (4.3) is finite (remember that a trace of a unitary matrix takes on values between $-N_c$ and N_c), the only possibility for the RHS to diverge is for the sum over plaquettes p' to diverge. This is possible only when long-range (in the units of the lattice spacing) correlations are essential or, in other words, the correlation length is infinite. Thus, we have once again reproduced the argument that the continuum limit of lattice theories is reached at the points of second-order phase transitions.

Such a second-order phase transition seems to occur in compact QED (*i.e.* the $U(1)$ lattice gauge theory with fermions) at $e_*^2 \sim 1$. It is associated there with deconfinement of electrons. Electrons are confined for $e^2 > e_*^2$, similar to quarks in the lattice QCD, and are liberated for $e^2 < e_*^2$. The interaction potential looks like that of Fig. 12b for $e^2 < e_*^2$ and like that of Fig. 12a in the confinement region $e^2 > e_*^2$.¹¹ In order to reach the continuum limit with deconfined electrons, the bare charge e^2 should be chosen slightly below the critical value. Then the renormalized physical charge can be made as small as the experimental value ($\alpha = 1/137$) according to the renormalization group arguments which are presented in the Remarks to Subsect. 3.7.

The nature of the phase transition in a four-dimensional compact $U(1)$ lattice gauge theory without fermions was investigated by numerical methods. While the very first paper [24] indicated that the phase transition is of second order, some more advanced recent investigations say that it may be weakly first order (see, *e.g.*, the talks at the Lattice Conference [25]).

¹¹The latter statement is not quite correct due to the possibility of creating electron-positron pairs out of vacuum.

Fig. 16: Graphic representation of the self-consistency condition (4.7). The link variables are replaced by $m \cdot 1$ at all links except for a given one denoted by the bold line.

Part 2.7). Anyway, we need fermions which usually weaken a phase transition that happens in a pure lattice gauge theory.

There are no indications that a second-order phase transition occurs in non-Abelian pure lattice gauge theories at intermediate values of β . This very strongly supports a behavior of the string tension of the type depicted in Fig. 13. The second-order phase transition occurs in four dimensions at $\beta = \infty$ (or $g^2 = 0$) according to the general arguments of Subsect. 3.7, which is necessary for the continuum limit to exist.

Remark on confinement in $4 + \epsilon$ dimensions

In $4 + \epsilon$ dimensions ($\epsilon > 0$), a second-order deconfining phase transition always occurs in non-Abelian pure lattice gauge theories at some finite value of $\beta < \infty$ (or $g^2 > 0$). The case of $\epsilon \ll 1$ can be considered analogous to the ϵ -expansion in statistical mechanics [26]. An ultraviolet-stable fixed point now exists at $g_*^2 \sim \epsilon$ since the theory is asymptotically free in $d = 4$. This phase transition is associated with deconfinement quite analogously to compact QED in $d = 4$.

4.2 Mean-field method

The idea to apply the mean-field method, which is widely used in statistical systems, to study phase transitions in the lattice gauge theories was proposed by Wilson [1] and first implemented for Abelian theories by Balian, Drouffe and Itzykson [17]. A mean field usually works well when there are many neighbor degrees of freedom, interacting with a given one.

In the simplest version of the mean-field method, the link variable $U_{x,\mu}$ is replaced by the mean-field value $m \cdot 1$ everywhere but at a given link (see Fig. 16) at which the self-consistency condition

$$\langle [U_{x,\mu}]^{\nu} \rangle_0 = m \delta^{\nu} \quad (4.5)$$

is imposed.

Fig. 17: Typical behavior of the solutions of the self-consistency equation (4.7). The only solution with $m = 0$ exists for $\beta < \beta_*$. Two more solutions appear for $\beta > \beta_*$. The one depicted by the dashed line is unstable. The actual value of m versus β is depicted by the bold lines. A first-order phase transition is associated with $\beta = \beta_*$.

The average on the LHS of Eq. (4.5) is calculated with the action which is obtained from (3.15) by the substitution of $m \cdot 1$ for all the link variables (or their Hermitean conjugate) except at the given link. Since the given link enters $2(d-1)$ plaquettes, the average on the LHS of Eq. (4.5) is to be calculated with the action

$$S_0[U] = 2(d-1)m^3 \text{Re tr } U_{x,\mu} + \text{const.} \quad (4.6)$$

Therefore, the self-consistency condition (4.5) can be written by the substitution of the mean-field ansatz into the lattice partition function (3.24) as

$$\frac{\int dU e^{\beta N_c \text{Re tr } U \frac{1}{N_c} \text{tr } U}}{\int dU e^{\beta N_c \text{Re tr } U}} = m \quad (4.7)$$

with

$$\tilde{\beta} = 2(d-1)m^3 \frac{\beta}{N_c^2}. \quad (4.8)$$

The meaning of Eq. (4.7) is very simple: the average of the trace of the link variable at the given link should coincide with m which is substituted for all other links of the lattice.

In order to verify whether the self-consistency condition (4.7) admits nontrivial solutions, one should first calculate the group integral on the LHS and then solve the self-consistency equation for m versus β . A typical behavior of the solution is depicted in Fig. 17. For all values of β , there exists a trivial solution $m = 0$ which is associated with no mean field. At some value β_* , two more solutions of the self-consistency equation appear. The upper of them is associated with positive specific heat while the lower one corresponds to negative one. This can be seen by noting that

$$W_{\partial p} = m^4 \quad (4.9)$$

in the mean-field approximation which follows from the substitution of the link-variables in the definition (3.50) by the mean-field values. This nontrivial solution is preferred for $\beta > \beta_*$.

since the partition function for it is larger (or the free energy is smaller) than for the $m = 0$ solution. The value of β_* is often associated with the point of a first-order phase transition.

The mean-field method in such a simple form was first applied to non-Abelian lattice gauge theories in Ref. [27, 28]. For the cases when a first-order phase transition occurs (say, for the $SU(N_c)$ groups with $N_c > 3$ or for the $SO(3)$ group), agreement with numerically calculated positions of the phase transitions is remarkable.

4.3 Mean-field method (variational)

There are some puzzles with the simplest mean-field ansatz described above. First of all, the average value of the link variable $U_{x,\mu}$ in a lattice gauge theory must vanish due to the gauge invariance (remember that $U_{x,\mu}$ changes under the gauge transformation according to Eq. (3.12) while the action and the measure are gauge invariant). This is in accordance with Elitzur's theorem [29] which says that a local gauge symmetry can not be spontaneously broken, so that any order parameter for phase transitions in lattice gauge theories must be gauge invariant.

A way out is to reformulate the mean-field method in lattice gauge theories as a variational method [17] which is similar to that proposed by R. Peierls in thirties. It is based on Jensen's inequality¹²

$$\langle e^F \rangle_0 \geq e^{\langle F \rangle_0} \quad (4.10)$$

which is due to the convexity of the exponential function, where $\langle \dots \rangle_0$ means averaging with respect to a trial action.

Let us choose the trial partition function

$$Z_0 = \int \prod_{x,\mu} dU_{x,\mu} e^{\beta N_c \sum_{x,\mu} \text{Re tr } U_{x,\mu}} \quad (4.11)$$

as a product of one-link integrals. Adding and subtracting the trial action, we write down the following bound on the partition function (3.24):

$$Z \geq Z_0 e^{\left\langle \frac{\beta}{N_c} \sum_p \text{Re tr } U_{\partial p} - \beta N_c \sum_{x,\mu} \text{Re tr } U_{x,\mu} \right\rangle_0}. \quad (4.12)$$

Here $\langle \dots \rangle_0$ means averaging with respect to the same action as in Eq. (4.11).

Since the expression which is averaged in the exponent in Eq. (4.12) is linear in each of the link variables, it can be calculated via the one-matrix integral given by the LHS of Eq. (4.7). Therefore, we get

$$\left\langle \frac{\beta}{N_c} \sum_p \text{Re tr } U_{\partial p} - \beta N_c \sum_{x,\mu} \text{Re tr } U_{x,\mu} \right\rangle_0 = \beta N_p m^4 - \bar{\beta} N_c^2 N_l m, \quad (4.13)$$

where Eq. (4.9) has been used.

¹²More detail can be found, e.g., in the book [30].

The idea of the variational mean-field method is to fix $\bar{\beta}$ from the condition for the trial ansatz (4.11) to give the best approximation to Z in the given class. Calculating the derivative of the RHS of Eq. (4.12) with respect to $\bar{\beta}$ and taking into account that m depends on $\bar{\beta}$ according to Eq. (4.7), we find the maximum at $\bar{\beta}$ given by Eq. (4.8), which reproduces the simplest version of the mean-field method described above.

To restore Elitzur's theorem, a more sophisticated trial ansatz [31] can be considered:

$$Z_0 = \int \prod_{x,\mu} dU_{x,\mu} e^{N_c \sum_{x,\mu} \text{Re tr } B_{x,\mu}^\dagger U_{x,\mu}}, \quad (4.14)$$

where we choose $B_{x,\mu}$ to be an arbitrary complex $N_c \times N_c$ matrix. Now the best approximation is reached for

$$B_{x,\mu} = \bar{\beta} \Omega_x \Omega_{x+a\hat{\mu}}^\dagger, \quad (4.15)$$

where $\bar{\beta}$ is given by exactly the same equation as before, while $\Omega_x \in SU(N_c)$ but is arbitrary otherwise. Now $\langle U_{x,\mu}^{ij} \rangle_0$ vanishes after summing over equivalent maxima which results in integrations over $d\Omega_x$.

For the ansatz (4.14), the analog of Eq. (4.13) reads

$$\left\langle \frac{\beta}{N_c} \sum_p \text{Re tr } U_{\partial p} - N_c \sum_{x,\mu} \text{Re tr } B_{x,\mu}^\dagger U_{x,\mu} \right\rangle_0 = \frac{\beta}{N_c} \sum_p \text{Re tr } M_{\partial p} - N_c \sum_{x,\mu} \text{Re tr } B_{x,\mu}^\dagger M_{x,\mu}, \quad (4.16)$$

where we have denoted

$$M_{x,\mu}^{ij} = \frac{\int \prod_{x,\mu} dU_{x,\mu} e^{N_c \sum_{x,\mu} \text{Re tr } B_{x,\mu}^\dagger U_{x,\mu}} U_{x,\mu}^{ij}}{\int \prod_{x,\mu} dU_{x,\mu} e^{N_c \sum_{x,\mu} \text{Re tr } B_{x,\mu}^\dagger U_{x,\mu}}}. \quad (4.17)$$

The inequality (4.12) takes the form

$$Z \geq Z_0 e^{\frac{\beta}{N_c} \sum_p \text{Re tr } M_{\partial p} - N_c \sum_{x,\mu} \text{Re tr } B_{x,\mu}^\dagger M_{x,\mu}}. \quad (4.18)$$

$B_{x,\mu}$ can now be determined by maximizing with respect to $B_{x,\mu}$ and taking into account Eq. (4.17).

It is easy to see that if $B_{x,\mu} = \bar{\beta} \cdot 1$ is a solution as before, then (4.15) is also a solution. Therefore, we get

$$\langle U_{x,\mu}^{ij} \rangle_0 = m \int d\Omega_x \Omega_{x+a\hat{\mu}} d\Omega_x \Omega_{x+a\hat{\mu}}^\dagger \Omega_x^\dagger = 0, \quad (4.19)$$

where the integration over Ω takes into account different equivalent maxima. Thus, all gauge invariant quantities for the ansatz (4.14) are the same as for the ansatz (4.11) while gauge noninvariant quantities now vanish in agreement with Elitzur's theorem.

Fig. 18: Lattice renormalization group transformation (4.20). The thin lines of the old lattice a) represent links on which integration is performed. The new lattice b) has the lattice spacing $2a$ but the same spatial extent La .

Remark on the criterion for phase transition

Another puzzle with the simplest mean-field method is why the point of the first-order phase transition is chosen as is explained in Fig. 17 but not when the free energy of both phases coincide (the standard Maxwell rule in statistical physics). Perhaps, the criterion of Fig. 17 should be chosen if a barrier between two phase is impenetrable which happens at large N_c or if quantum fluctuations are not taken into account as for the simplest mean field. The mean-field calculations of Ref. [32], which take into account fluctuations around the mean-field solution (4.15), agree for the Maxwell-rule criterion with numerical data. These results are reviewed in Ref. [15].

4.4 Lattice renormalization group

While the mean-field method is useful for studying the first-order phase transitions, the second-order phase transitions in lattice statistical systems are better described by the renormalization group method (see, *e.g.*, the review by Wilson and Kogut [26]). The idea to apply a similar method to lattice gauge theories is due to Migdal [18].

A simple renormalization group transformation in lattice gauge theories is associated with doubling of the lattice spacing a . One has originally a lattice depicted in Fig. 18a. The lattice renormalization group transformation consists in integrating over the link variables $U_{x,\mu}$ on the links shown by the thin lines which results in a lattice with spacing $2a$,

$$a \xrightarrow{\text{r.g.}} 2a, \quad (4.20)$$

which is depicted in Fig. 18b. The space size of the lattice is L before the transformation and becomes $L/2$ after the transformation,

$$L \xrightarrow{\text{r.g.}} \frac{L}{2}, \quad (4.21)$$

so that the lattice extent is $L \cdot a$ in both cases which is expected to reduce an influence of finite size effects on the transformation.

The Wilson action on the lattice of Fig. 18a becomes a more general one under the renormalization group transformation:

$$S[U] = \sum_p \beta \frac{1}{N_c} \text{tr} U_p \xrightarrow{\text{r.g.}} S'[U] = \sum_p \beta'_1 \frac{1}{N_c} \text{tr} U_p + \sum_{p_2} \beta'_2 \frac{1}{N_c} \text{tr} U_{p_2} + \sum_{p_3} \beta'_3 \frac{1}{N_c} \text{tr} U_{p_3} + \dots \quad (4.22)$$

The new action $S'[U]$ is not necessarily a single-plaquette action and can involve traces of the Wilson loops for boundaries of double plaquettes, triple plaquettes and so on.

The new action would be the same as the old one only at a fixed point. This usually happens after the renormalization group transformation is applied several times when the lattice theory does have a fixed point. The resulting action is then associated with an action of a continuum theory.

The great success of non-Abelian lattice gauge theories with the Wilson action in describing the continuum limit even at a relatively small spatial extent or, which is the same, at relatively large g^2 and a , is because it is not far away from the fixed-point action of the renormalization group. The proper numerical results are presented below in this Subsection.

If both actions $S[U]$ and $S'[U]$ are the single-plaquette Wilson actions, then

$$\beta \xrightarrow{\text{r.g.}} \beta' = \beta - \Delta\beta \quad (4.23)$$

under the renormalization group transformation on the lattice.

Since the Gell-Mann–Low function $B(g^2)$ in the continuum is known, $\Delta\beta$ versus β is determined by the equation

$$\int_{\beta-\Delta\beta}^{\beta} \frac{dx}{x^{3/2}} B^{-1}\left(\frac{6}{x}\right) = -\frac{2\ln 2}{\sqrt{6}}. \quad (4.24)$$

Here the $\ln 2$ on the RHS is due to Eq. (4.20) and the relation (3.25) between β and g^2 is used with $N_c = 3$.

For the pure SU(3) gauge theory, we get from Eq. (4.24)

$$\Delta\beta = 0,579 + \frac{0,204}{\beta} + O\left(\frac{1}{\beta^2}\right) \quad (4.25)$$

at asymptotically large β .

One can integrate over the thin links in Fig. 18a either approximately or numerically. The following procedure for an approximate integration is known as Migdal–Kadanoff recursion relations.

Let us expand the exponential of the old action in characters

$$e^{-S[U]} = \sum_{\tau} f_{\tau} d_{\tau} \chi_{\tau}(U) \quad (4.26)$$

Fig. 19: Monte Carlo data by Akemi *et al.* [33] for $\Delta\beta$. The errorbars represent statistical errors. The solid line represents the asymptotics (4.25).

where $d_r = \chi_r(1)$ is the dimension of a given representation r and f_r are the coefficients which depend on the form of $S[U]$.

Migdal [18] proposed to approximate the new action, which appear after

$$a \Rightarrow \rho a, \quad (4.27)$$

by the formula

$$e^{-S'[U']} = \left[\sum_r f_r \rho^2 d_r \chi_r(U') \right]^{\rho^d - 2}, \quad (4.28)$$

which is exact in $d = 2$ dimensions. Kadanoff [19] slightly modified the recursion relation (4.28).

The study of the Migdal-Kadanoff recursion relations was historically the first argument that second-order phase transitions do not occur in the non-Abelian lattice gauge theory when g^2 is decreased. Moreover, these relations in $d = 4$ are the same as for spin systems (with the same symmetry group) in $d = 2$ where this phenomenon is known. A disadvantage of the method is that it is difficult to estimate its accuracy.

A final answer to the question of whether or not a second-order phase transition occurs in the non-Abelian lattice gauge theory was done by the numerical integration. This is known as the Wilson Monte Carlo renormalization group. Some typical results [33] for the $\Delta\beta$, which is defined by Eq. (4.23), versus β are depicted in Fig. 19. The solid line represents the asymptotics (4.25). The agreement confirms that the continuum limit is reached already at these values of β , while the deviation of the Monte Carlo data from the asymptotics for smaller values of β is due to lattice non-perturbative effects.

4.5 Monte Carlo method

The idea of the Monte Carlo method is to calculate the partition function (3.24) and the averages (3.32) for arbitrary values of β numerically, using the fact that the multiplicity of the integral is large. For a $L \times L \times L \times L$ lattice in four dimensions, a typical multiplicity

of the integral is as large as $4 \cdot (N_c^2 - 1) \cdot L^4$ ($\sim 10^7$ for $L = 24$). It is hopeless to calculate such an integral exactly. On the contrary, the larger the multiplicity the better the Monte Carlo method works.

As usual, the Monte Carlo method is applied not to sequential integrals over $U_{x,\mu}$ at each link but rather to the multiple integral as a whole, which can be viewed as a sum over states of a statistical system.

A state is identified with a gauge field configuration which is described by the values of the link variables at all the links of the lattice:

$$C = \{U_{x,\mu}^i, \dots, \dots, \dots\}. \quad (4.29)$$

There are as many positions in this row as the multiplicity of the integral. Then the sequential integral can be represented as

$$\int \prod_{x,\mu} dU_{x,\mu} \dots = \sum_C \dots \quad (4.30)$$

The averages (3.32) can be rewritten as

$$\langle F(C) \rangle = \frac{\sum_C e^{-\beta S(C)} F(C)}{\sum_C e^{-\beta S(C)}}, \quad (4.31)$$

where $S(C)$ and $F(C)$ are the values of F and S for the given configuration C .

The task of Monte Carlo calculations is not to sum over all possible configurations whose number is infinite but rather to construct an ensemble, say, of N configurations

$$E = \{C_1, \dots, C_N\} \quad (4.32)$$

such that for a given configuration C_n to encounter with the Boltzmann probability

$$P_{\text{Bol}}(C_n) = Z^{-1}(\beta) e^{-\beta S(C_n)}. \quad (4.33)$$

Such a sample of configurations is called the *equilibrium ensemble*.

Given an equilibrium ensemble, the averages (4.31) take the form of the arithmetic mean

$$\langle F[U] \rangle = \frac{1}{N} \sum_{n=1}^N F(C_n) \quad (4.34)$$

because each configuration “weighs” already as much as is required. In particular, the Wilson loop average for a rectangular contour is given by

$$W_{R \times T} = \frac{1}{N} \sum_{n=1}^N \frac{1}{N_c} \text{tr} U_{R \times T}(C_n). \quad (4.35)$$

If all configurations in the equilibrium ensemble are independent, then the RHS of Eq. (4.35) will approximate the exact value of $W_{R \times T}$ with an accuracy $\sim \sqrt{N}$.

An analogy of this method of calculating averages with statistical physics is obvious. The equilibrium ensemble simulates actual states of a statistical system while the index n describes a time evolution.

A crucial point in the Monte Carlo method is to construct the equilibrium ensemble. It is not simple to do that because the Boltzmann probability is not known at the outset. A way out is to establish a random process for which each new configuration in the sequence (4.32) is obtained from the previous one by a definite algorithm but stochastically. In other words, the random process is completely determined by the probability $P(C_{n-1} \rightarrow C_n)$ for a transition from a state C_{n-1} to a state C_n and does not depend on the history of the system, *i.e.*

$$P(C_{n-1} \rightarrow C_n) = P(C_{n-1}, C_n). \quad (4.36)$$

Such a random process is known as the *Markov process*.

The transition probability $P(C, C')$ should be chosen in such a way to provide the Boltzmann distribution (4.33). This is ensured if $P(C, C')$ satisfies the detailed balance condition

$$e^{-\beta S(C)} P(C, C') = e^{-\beta S(C')} P(C', C). \quad (4.37)$$

Then

- 1) an equilibrium sequence of states will transform into another equilibrium sequence,
- 2) a nonequilibrium sequence will approach an equilibrium one when moving through the Markov chain.

These statements can be proven using the detailed balance condition (4.37).

Specific Monte Carlo algorithms differ in the choice of the transition probability $P(C, C')$ while the detailed balance condition (4.37) is always satisfied. The two most popular algorithms, which act at one link, are

Heat bath algorithm: A new link variable $U'_{x,\mu}$ is selected randomly from the group manifold with a probability given by the Boltzmann factor $P(U'_{x,\mu}) \propto e^{-\beta S(C')}$. Then this procedure is repeated for the next link and so on until the whole lattice is passed. This can be imagined as if a reservoir with a temperature $1/\beta$ touches each link of the lattice in succession. It is clear from physical intuition that the system will be brought to the thermodynamic equilibrium sooner or later.

Metropolis algorithm: This algorithm is used in statistical physics since the fifties and consists of several steps.

- a) A trial new link variable $U'_{x,\mu}$ is selected (*e.g.*, randomly on the group manifold).
- b) The difference between the action for this trial configuration and that for the old one is calculated: $\Delta S = S(C') - S(C)$.
- c) A random number $r \in [0, 1]$ is generated.
- d) If $e^{-\beta \Delta S} > r$, then $U'_{x,\mu}$ is accepted. Otherwise, $U'_{x,\mu}$ is rejected and the old value $U_{x,\mu}$ is kept.
- e) All this is repeated for the next links.

Fig. 20: Monte Carlo data by Creutz [22] for the string tension in the SU(2) pure lattice gauge theory.

A new configuration C' , which is obtained by applying either Monte Carlo algorithm to each link of the lattice once (this procedure is often called the Monte Carlo sweep), will be strongly correlated with the old one C . This is because the lattice action depends not only on the variable at the given link but also on those at the neighboring link which form plaquettes with the given one. In order for C' to become independent on C , this procedure should be repeated many times or special tricks to reduce the correlations should be implemented. Then this new configuration can be added to the equilibrium ensemble (4.32) as C_n .

More detail about the Monte Carlo algorithms as well as their practical implementation in lattice gauge theories can be found in the review [34] and the book by Creutz [14].

4.6 Some Monte Carlo results

The first Monte Carlo calculation in non-Abelian lattice gauge theories, which is relevant for the continuum limit, was that by Creutz [22] who evaluated the string tension for the SU(2) gauge group. His result is reproduced in Fig. 20 and looks very much like what is expected in Fig. 13. This calculation was the first demonstration that the continuum limit sets in for relatively large $g^2 \approx 1.9$ ($\beta \approx 2.2$) and that results for the continuum can therefore be extracted from relatively small lattices.

The restoration of rotational symmetry for these values of g^2 was explicitly demonstrated by Land and Rebbi [35]. They calculated equipotential surfaces for the interaction between static quarks. In the strong coupling region $g^2 \rightarrow \infty$, they appear as in Fig. 21a since the interaction potential is given by

$$E(x, y, z) = K(|x| + |y| + |z|) \quad (4.38)$$

because the distance between the quarks is measured along the lattice. This is associated

Fig. 21: Behavior of equipotential lines at different values of β : a) the strong coupling limit $\beta = 0$, b) $\beta = 2$, c) $\beta = 2.25$. Figs. b) and c), taken from the paper by Lang and Rebbi [35], show how the rotational symmetry is restored as β is increased.

with a cubic symmetry on the lattice (*i.e.* rotations through an angle which is a multiple of $\pi/2$ around each axis and translations by a multiple of the lattice spacing along each axis) rather than with the Poincaré group. The rotational symmetry must be restored in the continuum limit.

The Monte Carlo data by Land and Rebbi [35] shown in Figs. 21b, c demonstrate the restoration of rotational symmetry when passing from $\beta = 2$ (Fig. 21b) to $\beta = 2.25$ (Fig. 21c).

The old Monte Carlo calculations played an important role for developing the method. A dramatic improvement of the Monte Carlo technology in lattice gauge theories happened during the last ten years. The Proceedings of the Lattice Conference [25] are very useful, as is already mentioned, for a today's look at the subject.

5 Fermions on a lattice

It turned out to be most difficult in the lattice approach to QCD to deal with fermions. Putting fermions on a lattice is an ambiguous procedure since the cubic symmetry of a lattice is less restrictive than the continuous Lorentz group.

The simplest chiral-invariant formulations of lattice fermions lead to a doubling of fermionic degrees of freedom, as was first noted by Wilson [13], and describe from 16 to 4 relativistic continuum fermions, depending on the formulation. One half of them has a positive axial charge and the other half has a negative one, so that the chiral anomaly cancels. There is a no-go theorem which says that the fermionic doubling is always present under natural assumptions about a lattice gauge theory.

A practical way out is to choose the fermionic lattice action to be explicitly noninvariant under the chiral transformation and to have, by tuning a mass of the lattice fermion, 1 relativistic fermion in the continuum and the masses of the doublers to be of the order of the inverse lattice spacing. The chiral anomaly is recovered in this way.

We consider in this Section various formulations of lattice fermions and the doubling problem. We discuss briefly the results on spontaneous breaking of chiral symmetry in QCD.

5.1 Chiral fermions

The quark fields are generically matter fields, whose gauge transformation in the continuum is given by Eqs. (2.20) and (2.22), and can be put on a lattice according to Eq. (3.6). Then the lattice gauge transformation is

$$\begin{aligned}\psi_x &\xrightarrow{g,t} \psi'_x = \Omega_x \psi_x, \\ \bar{\psi}_x &\xrightarrow{g,t} \bar{\psi}'_x = \bar{\psi}_x \Omega_x^\dagger.\end{aligned}\quad (5.1)$$

The lattice analog of the QCD action (2.33) reads¹³

$$S[U, \bar{\psi}, \psi] = \beta S_{\text{lat}}[U] + M \sum_x \bar{\psi}_x \psi_x + \frac{1}{2} \sum_{x,\mu>0} [\bar{\psi}_x \gamma_\mu U_{x,\mu}^\dagger \psi_{x+\hat{\mu}} - \bar{\psi}_{x+\hat{\mu}} \gamma_\mu U_{x,\mu} \psi_x]. \quad (5.2)$$

The first term on the RHS is the pure gauge lattice action (3.15). The second term is a quark mass term on a lattice. The sum in the third term is over all lattice links (*i.e.* over all sites x and positive directions μ). This action is Hermitian and invariant under the lattice gauge transformation (3.12), (5.1) at a finite lattice spacing.

The partition function of lattice QCD with fermions is defined by

$$Z(\beta, M) = \int \prod_{x,\mu} dU_{x,\mu} \prod_x d\bar{\psi}_x d\psi_x e^{-S[U,\bar{\psi},\psi]}, \quad (5.3)$$

¹³A standard formula differs from this one by an interchange of U and $U^\dagger$ due to the inverse ordering of matrices in the phase factors. It does not matter how to define $U_{x,\mu}$ since the Haar measure is invariant under Hermitian conjugation.

where the action is given by Eq. (5.2). The integration over $U_{x,\mu}$ is as in Eq. (3.24), and the integral over the quark field is the Grassmann one. The averages are defined by

$$\langle F[U, \bar{\psi}, \psi] \rangle = Z^{-1}(\beta, M) \int \prod_{x,\mu} dU_{x,\mu} \prod_x d\bar{\psi}_x d\psi_x e^{-S[U, \bar{\psi}, \psi]} F[U, \bar{\psi}, \psi], \quad (5.4)$$

which extends Eq. (3.32) to the case of fermions. Since both the action and the measure in Eq. (5.4) are gauge invariant at finite lattice spacing, a nonvanishing result is only when the integrand, $F[U, \bar{\psi}, \psi]$, is gauge invariant as well.

In order to show how the lattice action (5.2) reproduces (2.33) in the naive continuum limit $a \rightarrow 0$, let us assume that the lattice quark field ψ_x slowly varies from site to site and substitute

$$\begin{aligned} \psi_x &\rightarrow a^{3/2} \psi(x), \\ \psi_{x+a\hat{\mu}} &\rightarrow a^{3/2} (\psi(x) + a\partial_\mu \psi(x)) \end{aligned} \quad (5.5)$$

in $d = 4$. Here $\psi(x)$ is a continuum quark field and the power of a is due to dimensional consideration (remember that ψ_x is dimensionless).

Equation (5.5) together with Eq. (3.9) yields

$$\bar{\psi}_x \gamma_\mu U_{x,\mu}^\dagger \psi_{x+a\hat{\mu}} \rightarrow a^3 \bar{\psi} \gamma_\mu \psi + a^4 \bar{\psi} \nabla_\mu^{\text{fun}} \gamma_\mu \psi + \mathcal{O}(a^5), \quad (5.6)$$

where there is no summation over μ in the second term as before. The first term cancels when substituted into Eq. (5.2) while the second one reproduces the fermionic part of the continuum action. The mass term is also reproduced if $M = am$.

The fermionic lattice action (5.2) was proposed in Ref. [1]. For $M = 0$ it is invariant under the global chiral transformation

$$\psi_x \xrightarrow{c,t} e^{i\alpha\gamma_5} \psi_x, \quad \bar{\psi}_x \xrightarrow{c,t} \bar{\psi}_x e^{i\alpha\gamma_5}. \quad (5.7)$$

For this reason, these lattice fermions are called *chiral fermions*. Since the lattice action is both gauge and chiral invariant, there is no Adler–Bell–Jackiw anomaly according to the general arguments.

5.2 Fermion doubling

As is pointed out at the end of the previous Subsection, the lattice fermionic action (5.2) is both gauge and chiral invariant (for $M = 0$) so that there is no axial anomaly in the continuum. Since the anomaly is present for one continuum fermion, this suggests that the action (5.2) is associated with more than one species of continuum fermions.

In order to verify this explicitly, let us calculate the poles of the lattice fermionic propagator.

As usual, it is easier to work with the Fourier image of ψ_x :

$$\psi_k = a^{5/2} \sum_x \psi_x e^{-ikx}. \quad (5.8)$$

The free fermionic action then reads

$$S_0[\bar{\psi}, \psi] = \int_{-\pi/a}^{\pi/a} \frac{d^4 k}{(2\pi)^4} \bar{\psi}_k G^{-1}(k) \psi_k \quad (5.9)$$

with

$$G^{-1}(k) = \frac{1}{a} \sum_{\mu=1}^4 i\gamma_\mu \sin k_\mu a \quad (5.10)$$

for $M = 0$.

In the naive continuum limit, the sin in Eq. (5.10) can be expanded in the power series in a , which results in the free (inverse) continuum propagator

$$G^{-1}(k) \rightarrow i \sum_{\mu=1}^4 \gamma_\mu k_\mu = i \hat{k}. \quad (5.11)$$

The Lorentz invariance has been restored after summing over μ .

When passing from the lattice expression (5.10) to the continuum one (5.11), it was saliently assumed that the momentum k_μ is not of order of $1/a$ because otherwise the sin can not be expanded in a . The doubling of relativistic continuum fermionic states occurs exactly for this reason.

To find the poles of the propagator, let us return to Minkowski space substituting $k_4 = iE$ with E being the energy. The poles are then determined by the dispersion law

$$\sinh^2 Ea = \sum_{\mu=1}^3 \sin^2 p_\mu a. \quad (5.12)$$

Let us look for solutions of Eq. (5.12) with positive energy $E > 0$ (solutions with negative energy are associated as usual with antiparticles). Suppose that a particle moves along the z -axis so that components of the four-momentum

$$p^{(1)} = (E, 0, 0, p_z) \quad (5.13)$$

are related by

$$\sinh Ea = \sin p_z a, \quad (5.14)$$

which follows from the substitution of (5.13) into the dispersion law.

Since the sin is a periodic function, the four-vector

$$p^{(2)} = \left(E, 0, 0, \frac{\pi}{a} - p_z \right) \quad (5.15)$$

is also a solution of Eq. (5.14) if (5.13) is. Quite analogously, the four-vectors

$$\begin{aligned} p^{(3)} &= \left(E, \frac{\pi}{a}, 0, p_z \right), \\ &\dots, \\ p^{(8)} &= \left(E, \frac{\pi}{a}, \frac{\pi}{a}, \frac{\pi}{a} - p_z \right), \end{aligned} \quad (5.16)$$

Fig. 22: Altering signs of ψ_x on a lattice along a) one axis and b) two axes.

which are obtained from $p^{(1)}$ and $p^{(2)}$ by changing zeros for π/a , also satisfy Eq. (5.14). Therefore, a quark state with the energy E is eightfold degenerate.

The quark states with the four-momenta $p^{(1)}, \dots, p^{(8)}$ are different states. Their wave functions equal

$$\Psi^{(j)}(t, x, y, z) \propto \exp(iEt - ip_x^{(j)}x - ip_y^{(j)}y - ip_z^{(j)}z). \quad (5.17)$$

The wave function in the state with the momentum $p^{(3)}$ differs, say, from the wave function in the state $p^{(1)}$ by an extra factor $(-1)^{x/a}$. In other words, it strongly changes as $a \rightarrow 0$ with one step along the lattice in the x -direction. One more step returns the initial value.

For such functions, the naive continuum limit of the lattice action (5.2) is as good as for the slowly varying functions when Eq. (5.5) holds. In order to see that, let us rewrite the action (5.2) as

$$S[U, \bar{\psi}, \psi] = \beta S_{\text{lat}}[U] + M \sum_x \bar{\psi}_x \psi_x + \frac{1}{2} \sum_{x, \mu} \left[\bar{\psi}_x \gamma_\mu (U_{x, \mu}^\dagger \psi_{x+a\hat{\mu}} - U_{x, -\mu}^\dagger \psi_{x-a\hat{\mu}}) \right]. \quad (5.18)$$

Even if ψ_x has opposite signs at neighboring lattice sites along the μ -axis, as is illustrated by Fig. 22a, *i.e.*

$$\psi_{x+a\hat{\mu}} \rightarrow -\psi_x, \quad (5.19)$$

then the difference $\psi_{x+a\hat{\mu}} - \psi_{x-a\hat{\mu}}$ on the RHS of Eq. (5.18) is still of the correct order in a :

$$\psi_x \rightarrow a^{3/2} \psi(x), \quad \psi_{x+a\hat{\mu}} - \psi_{x-a\hat{\mu}} \rightarrow -2a^{5/2} \partial_\mu \psi(x), \quad (5.20)$$

so that the continuum fermionic action is reproduced except for the sign of the γ_μ -matrix which is opposite to that in Eq. (2.33).

This extra minus sign can be absorbed in the redefinition of the continuum fermionic field $\psi(x) \rightarrow i\gamma_\mu \gamma_5 \psi(x)$, which changes its chirality. Therefore, the axial charge of the doublers is opposite. Analogously, four of the eight doublers have positive axial charge and four others have negative one dependently on whether the sign of ψ_x alters at neighboring sites along

Fig. 23: Momentum dependence of G^{-1} for the chiral lattice fermions. The periodicity leads to an extra zero at $k_4 = \pi/a$.

even or odd number of axes (see Fig. 22). In Euclidean space the doubling occurs also along the temporal axis, so the number of doublers is equal to $2^d = 16$: 8 of them with positive and 8 with negative axial charge. This explains why the chiral anomaly cancels.

The fermion doubling is intimately related [36] to the invariance of the lattice action (5.2) under the transformations

$$\psi_x \rightarrow T_\alpha \psi_x, \quad \bar{\psi}_x \rightarrow \bar{\psi}_x T_\alpha^\dagger \quad (5.21)$$

with

$$T_\alpha = I, \quad i\gamma_\mu \gamma_5 (-1)^{x_\mu/a}, \quad i\gamma_\mu \gamma_\nu (-1)^{(x_\mu + x_\nu)/a} \quad (\mu > \nu), \\ \gamma_4 (-1)^{(x+y+z)/a}, \dots, \gamma_1 (-1)^{(y+z+t)/a}, \gamma_5 (-1)^{(x+y+z+t)/a}. \quad (5.22)$$

There are all together $1 + 4 + 6 + 4 + 1 = 16$ independent transformations which tie the continuum fermion doublers.

It is worth noting that the mass term in Eq. (5.2) is not γ_5 invariant but does not remove the fermionic doubling.

One might think of removing the doubling problem by modifying the expression for the inverse lattice propagator $G^{-1}(k)$ in the free fermionic lattice action (5.9), for instance, by adding next-to-neighbor terms. It is easy to see that it does not help if the function $G^{-1}(k)$ is periodic as it should be on a lattice. A typical form of $G^{-1}(k)$ as a function of, say, k_4 is depicted in Fig. 23. The behavior around $k_4 = 0$ is prescribed by Eq. (5.11) and is just a straight line with a positive slope. Therefore, $G^{-1}(k)$ will have inevitably another zero at $k_4 = \pi/a$ due to periodicity.

This is the difference between the fermionic and bosonic cases. For bosons $G^{-1}(k)$ is quadratic in k_4 near $k_4 = 0$ rather than linear as for fermions. A typical behavior of $G^{-1}(k)$ for bosons is shown in Fig. 24. There is no doubling of states in the bosonic case.

Remark on the Nielsen–Ninomiya theorem

A general proof of the theorem which says that there is no way to avoid the fermion doubling under natural assumptions about the structure of a lattice gauge theory was given by Nielsen

Fig. 24: Momentum dependence of G^{-1} for the lattice bosons. No doubling of states is associated with this behavior.

and Ninomiya [37]. It is formulated sometimes as an absence of neutrinos on a lattice. In other words, this is a no-go theorem for putting theories with unequal number of left- and right-handed massless Weyl particles on a lattice, such as in the standard electroweak theory.

A way to bypass the Nielsen–Ninomiya theorem is, say, to choose a fermionic lattice action which is highly nonlocal. Then it is possible to replace $\sin k_\mu a$ in Eq. (5.10) by $k_\mu a$ itself to get an expression which is similar to the continuum propagator (5.11). However, such a nonlocal modification is useless for practical calculations. Some other modern approaches to resolving the fermion doubling problem can be found in the Proceedings of the Lattice Conference [25].

5.3 Kogut–Susskind fermions

The number of continuum fermion species is not necessarily equal to 16. It can be reduced down to 4 by the trick which was proposed for the Hamiltonian formulation in Refs. [9, 38] and elaborated for the Euclidean formulation in Refs. [39, 40].

Let us substitute

$$\psi_x = \gamma_1^{x/a} \gamma_2^{y/a} \gamma_3^{z/a} \gamma_4^{t/a} \phi_x \quad (5.23)$$

into the free fermionic action (5.9). Then it takes the form

$$S_0[\bar{\psi}, \psi] = \frac{1}{2} \sum_x \sum_{\mu > 0} \sum_{\mu} \eta_{x,\mu} \left([\phi_x^\dagger]^i [\phi_{x+a\hat{\mu}}]^i - [\phi_{x+a\hat{\mu}}^\dagger]^i [\phi_x]^i \right), \quad (5.24)$$

which is diagonal with respect to the spinor indices since

$$\eta_{x,\mu} = (-1)^{(x_1 + \dots + x_{\mu-1})/a}, \quad (5.25)$$

or explicitly

$$\begin{aligned} \eta_{x,\mu} &= 1 & \text{for } \mu = 1, \\ \eta_{x,\mu} &= (-1)^{x_1/a} & \text{for } \mu = 2, \\ &\dots & \\ \eta_{x,\mu} &= (-1)^{(x_1 + \dots + x_{\mu-1})/a} & \text{for } \mu = d, \end{aligned} \quad (5.26)$$

does not depend on spinor indices.

The idea is to leave only one component of ϕ_x^i in order to reduce the degeneracy:

$$\phi_x^i = \begin{pmatrix} \chi_x \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \bar{\phi}_x^i = \begin{pmatrix} \bar{\chi}_x \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (5.27)$$

These lattice fermions are known as the *staggered fermions*, since $\eta_{x,\mu}$ is staggering from one lattice site to another. They are more often called the *Kogut–Susskind fermions* due to their relation to those of Refs. [9, 38].

The action of the Kogut–Susskind fermions is

$$S[U, \bar{\psi}, \psi] = \beta S_{\text{Lat}}[U] + M \sum_x \bar{\chi}_x \chi_x + \frac{1}{2} \sum_{x,\mu > 0} \eta_{x,\mu} \left(\bar{\chi}_x U_{x,\mu}^\dagger \chi_{x+a\hat{\mu}} - \bar{\chi}_{x+a\hat{\mu}} U_{x,\mu} \chi_x \right). \quad (5.28)$$

It describes $2^{d/2} = 4$ species for complex χ_x or $2^{d/2-1} = 2$ species for Majorana χ_x . Components of a continuum bispinor are distributed in this approach over four lattice sites.

There is no axial anomaly for the Kogut–Susskind fermions as with the chiral fermions.

Remark on four generations

It is suggestive to identify four species of Kogut–Susskind fermions with four generations of quarks and leptons (see, e.g., Ref. [41]). Remember that one of the motivations to add the fourth generation to the standard model is to cancel the anomaly. However, there are problems with this idea concerning the splitting of fermion masses for the four generations.

5.4 Wilson fermions

The chiral lattice fermions were proposed by Wilson [1]. Soon after that he recognized [13] the problem of fermion doubling and proposed a lattice fermionic action which describes only one relativistic fermion in the continuum. The latter fermions are called *Wilson fermions*.

The lattice action for the Wilson fermions reads

$$S[U, \bar{\psi}, \psi] = \beta S_{\text{Lat}}[U] + M \sum_x \bar{\psi}_x \psi_x - \frac{1}{2} \sum_{x,\mu > 0} \left[\bar{\psi}_x (1 - \gamma_\mu) U_{x,\mu}^\dagger \psi_{x+a\hat{\mu}} + \bar{\psi}_{x+a\hat{\mu}} (1 + \gamma_\mu) U_{x,\mu} \psi_x \right]. \quad (5.29)$$

The difference between this action and the action (5.2) for chiral fermions is due to the projector operators $(1 \pm \gamma_\mu)$ which pick only one fermionic state.

Substituting the expansion (5.5) in the action (5.29), we get, in the naive continuum limit, the continuum fermionic action (2.33) with the mass being

$$m = \frac{M - 4}{a}. \quad (5.30)$$

Therefore, the Wilson lattice fermions describe a relativistic fermion of the mass m in the continuum when

$$M \rightarrow 4 + ma. \quad (5.31)$$

In order to see that there are no other relativistic fermion states in the limit (5.31), let us consider the fermionic propagator which is given by

$$G^{-1}(k) = M - \frac{1}{2} \sum_{\mu=1}^4 \left[(1 - \gamma_{\mu}) e^{ik_{\mu}a} + (1 + \gamma_{\mu}) e^{-ik_{\mu}a} \right]. \quad (5.32)$$

Introducing the Minkowski-space energy $E = -ik_4$, we get the following dispersion law

$$\cosh Ea = \frac{1 + \left(M - \sum_{\mu=1}^3 \cos p_{\mu}a \right)^2 + \sum_{\mu=1}^3 \sin^2 p_{\mu}a}{2 \left(M - \sum_{\mu=1}^3 \cos p_{\mu}a \right)}. \quad (5.33)$$

Let a particle be at rest, i.e. $p_1 = p_2 = p_3 = 0$ and $E = m > 0$. Then Eq. (5.33) reduces for $ma \ll 1$ to the relation (5.31). It is easy to show that a particle at rest is the only solution to Eq. (5.33) whose energy is finite as $a \rightarrow 0$.

The difference between the dispersion laws for the chiral and Wilson fermions is because the function on the RHS of Eq. (5.33) is no longer periodic. It reduces for $a \rightarrow 0$ and $M \rightarrow 4$ to a usual relation

$$E = \sqrt{\vec{p}^2 + m^2} \quad (5.34)$$

between the energy and momentum of a relativistic particle.

For $M \approx 4$, we can replace the LHS of Eq. (5.33) by 1 and substitute $M = 4$ on the RHS. Then the dispersion law (5.33) reduces to the equation for spatial components of the four-momentum:

$$\left(3 - \sum_{\mu=1}^3 \cos p_{\mu}a \right)^2 + \left(3 - \sum_{\mu=1}^3 \cos^2 p_{\mu}a \right) = 0, \quad (5.35)$$

whose unique solution is $p_1 = p_2 = p_3 = 0$ since both terms on the LHS are non-negative.

It is instructive to discuss what happens with the fermion doublers under the change of $\pm\gamma_{\mu}$ by $(1 \pm \gamma_{\mu})$ in the lattice fermionic action. Let us consider one of such states, say, that with $p_1 = \pi/a$, $p_2 = p_3 = 0$. Its energy is determined by Eq. (5.33) to be $\sim 1/a$ so that this state is inessential as $a \rightarrow 0$.

The chiral anomaly is correctly recovered by the Wilson fermions. The 15 states of the mass $\sim 1/a$ play thereat a role of regulators which result in the anomaly as $a \rightarrow 0$.

Remark on backtrackings for Wilson fermions

Another way to understand why the doubling problem is removed for the Wilson fermions is to consider how they propagate on a lattice. The projectors

$$P_{\mu}^{\pm} = \frac{1 \pm \gamma_{\mu}}{2} \quad \boxed{\text{Wilson fermions}} \quad (5.36)$$

restrict the propagation of the Wilson fermions. One half of the states propagates only in the positive directions and the other half propagates only in the negative ones. In particular, there are no backtrackings in the (lattice) sum over paths since

$$P_{\mu}^{+} P_{\mu}^{-} = 0. \quad (5.37)$$

This removes the doubling.

Remark on lattice fermion propagator

The fermion propagator in an external Yang-Mills field can be represented on a lattice as a sum over lattice paths by expanding in $1/M$. Let us first rescale the fermion field, absorbing the parameter M in front of the mass term. The fermionic part of the new action reads

$$S_{\psi} = \sum_x \bar{\psi}_x \psi_x - \kappa \sum_{x,\mu>0} \left[\bar{\psi}_x P_{\mu}^{-} U_{x,\mu}^{\dagger} \psi_{x+a\hat{\mu}} + \bar{\psi}_{x+a\hat{\mu}} P_{\mu}^{+} U_{x,\mu} \psi_x \right], \quad (5.38)$$

where $\kappa = 1/M$ is usually called the hopping parameter. The large mass expansion in $1/M$ is now represented as the hopping parameter expansion in κ .

It is convenient to depict each of the two terms in the square brackets in Eq. (5.38) by a string bit as in Fig. 5 with the quark fields at the ends and the gauge variable at the link. The first term corresponds to the negative direction of the link, and the second term corresponds to the positive direction. Substituting Eq. (5.38) into the definition (5.4) and expanding the exponential in κ , we get a combination of terms constructed from the string bits. A nonvanishing contribution to the quark propagator

$$G_{mn}^{ij}(x, y; U) = \left\langle \psi_m^i(x) \bar{\psi}_n^j(y) \right\rangle_{\psi}, \quad (5.39)$$

where i, j and m, n represent, respectively, color and spinor indices, emerges when the links form a path Γ_{yx} that connects x and y on the lattice as is depicted in Fig. 25. Otherwise, the average over ψ and $\bar{\psi}$ vanishes due to the rules of integration over Grassmann variables.

Therefore, we get

$$G_{mn}^{ij}(x, y; U) = \sum_{\Gamma_{yx}} \frac{1}{M^{L(\Gamma)+1}} U^{\Gamma} [\Gamma_{yx}] \left[\prod_{\Gamma_{yx}} P_{\mu}^{\pm} \right]_{mn}, \quad (5.40)$$

where P_{μ}^{+} or P_{μ}^{-} are associated with the positive or negative direction of a given link $\in \Gamma_{yx}$. For the Wilson fermions, they are given by Eq. (5.36), while

$$P_{\mu}^{\pm} = \pm \frac{\gamma_{\mu}}{2} \quad \boxed{\text{chiral fermions}} \quad (5.41)$$

for chiral fermions. The sum in Eq. (5.40) runs over all the paths between x and y on the lattice, while $L(\Gamma)$ stands for the length of the path Γ_{yx} in the lattice units. Eq. (5.40) is a lattice analog of the continuum formula (2.50) for the case of fermions.

Fig. 25: A path Γ_{yx} made out of the string bits, which leads to a nonvanishing term of the hopping parameter expansion for the quark propagator (5.39) on a lattice. Each site involves at least two quark fields (depicted by the circles). Otherwise the Grassmann integral at a given site vanishes.

Remark on the induced action

The integral over fermions in Eq. (5.3) can also be represented as a sum over closed paths on a lattice by expanding in $1/M$. The calculation is analogous to that of the previous Remark. The result reads

$$\int \prod_x d\bar{\psi}_x d\psi_x e^{-S_\psi} = e^{-S_{\text{ind}}[U]} \quad (5.42)$$

with

$$S_{\text{ind}}[U] = - \sum_{\Gamma} \frac{\text{tr } U[\Gamma]}{L(\Gamma) M^{L(\Gamma)}} \text{sp} \prod_{\mu} P_{\mu}^{\pm}, \quad (5.43)$$

where the combinatoric factor $1/L(\Gamma)$ is due to identity of L links forming the closed contour Γ , and the minus sign is because of fermions.

Equation (5.43) defines an effective action of a pure lattice gauge theory, which is nonlocal since involves arbitrary large loops. However, it can be made the single-plaquette lattice action (3.15) by introducing many flavors of lattice fermions [42, 43].

5.5 Quark condensate

The lattice action (5.29) is not invariant under the chiral transformation. Therefore, the chiral symmetry is explicitly broken for the Wilson fermions.

Nevertheless, one expects a restoration of chiral symmetry as $a \rightarrow 0$ when the relativistic fermion is massless (say, for $M = 4$ in the free case) while heavy states with $m \sim 1/a$ play the role of regulators. For the interacting theory, this restoration happens at some value $M = M_*$ which is no longer equal to 4. A signal of this restoration is the vanishing of the mass of the π -meson (as is illustrated by Fig. 26). $m_{\pi} = 0$ is usually associated with the fact that the chiral symmetry is realized in a spontaneously broken phase and the π -meson is the corresponding Goldstone boson.

For the chiral or Kogut-Susskind fermions with $M = 0$, the lattice action is invariant under the global chiral transformation (5.7). The order parameter for breaking the chiral symmetry is

$$\bar{\psi}\psi \xrightarrow{\text{c.t.}} \bar{\psi} e^{2i\alpha\gamma_5} \psi, \quad (5.44)$$

Fig. 26: Dependence of the π -meson mass on the lattice quark mass M . At $M = M_*$ the π -meson becomes massless and the chiral symmetry is restored.

Fig. 27: Monte Carlo data by Hamber and Parisi [44] for the quark condensate in the quenched approximation.

which is not invariant under the chiral transformation. Therefore, the average of $\bar{\psi}\psi$ must vanish if the symmetry is not broken spontaneously¹⁴. Such spontaneous breaking results in

$$\langle \bar{\psi}\psi \rangle \neq 0. \quad (5.45)$$

This nonvanishing value of the average of $\bar{\psi}_x\psi_x$ does not depend on x due to translational invariance and is called the *quark condensate*.

The spontaneous breaking of the chiral symmetry in QCD was demonstrated by the Monte Carlo calculations of the quark condensate. This quantity has a dimension of $[\text{mass}]^3$ and should depend on g^2 at small g^2 as is prescribed by the asymptotic scaling. The Monte Carlo data for the quark condensate from the pioneering paper by Hamber and Parisi [44] are shown in Fig. 27. Its agreement with asymptotic scaling demonstrates that the chiral symmetry is spontaneously broken in the continuum QCD.

¹⁴Spontaneous symmetry breaking usually occurs when the vacuum state is not invariant under the symmetry transformation.

Remark on Monte Carlo with fermions

Monte Carlo simulations with quarks are much more difficult than in a pure gauge theory. Integrating over the quark fields, one is left with the determinant, say for the Kogut–Susskind fermions, of the matrix

$$D[U] = M \delta_{x,y} + \frac{1}{2} \sum_{\mu>0} \left[\eta_{x,\mu} U_{x,\mu}^\dagger \delta_{x(x+a\hat{\mu})} - \eta_{x,\mu} U_{x,\mu} \delta_{x(x-a\hat{\mu})} \right] \quad (5.46)$$

for a given configuration of the gluon field $U_{x,\mu}$. This results in a pure gauge-field problem with the effective action given by

$$e^{-\beta S_{eff}[U]} = \det D[U] e^{-\beta S[U]}. \quad (5.47)$$

The matrix that appears in this determinant has at least $N_c L^4 \times N_c L^4$ elements, and is to be calculated at each Monte Carlo upgrading of $U_{x,\mu}$.

Several methods are proposed to manage the quark determinant approximately. The simplest one is not to take it into account at all. This approximation is known as the *quenched* approximation when only valence quarks are considered, while the effects of virtual quark loops are disregarded. Recently a progress in the full theory has been achieved using some tricks for evaluating the quark determinants (see Ref. [25] for a review of the subject).

6 1/N-expansion of vector models

The simplest models, which become solvable in the limit of a large number of field components, deal with a field which has N components forming an $O(N)$ vector in an internal symmetry space. A model of this kind was first considered by Stanley [45] in statistical mechanics and is known as the spherical model. The extension to quantum field theory was done by Wilson [46] both for the four-Fermi and φ^4 theories.

In the framework of perturbation theory, the four-Fermi interaction is renormalizable only in $d = 2$ dimensions and is non-renormalizable for $d > 2$. The $1/N$ -expansion resums perturbation-theory diagrams after which the four-Fermi interaction becomes renormalizable to each order in $1/N$ for $2 \leq d < 4$. An analogous expansion exists for the nonlinear $O(N)$ sigma model. The φ^4 theory remains trivial in $d = 4$ to each order of the $1/N$ -expansion while has a nontrivial infrared-stable fixed point for $2 < d < 4$.

The $1/N$ expansion of the vector models is associated with a resummation of Feynman diagrams. A very simple class of diagrams — the bubble graphs — survives to the leading order in $1/N$. This is why the large- N limit of the vector models is solvable. Alternatively, the large- N solution is nothing but a saddle-point solution in the path-integral approach. The existence of the saddle point is due to the fact that N is large. This is to be distinguished from a perturbation-theory saddle point which is due to the fact that the coupling constant is small. Taking into account fluctuations around the saddle-point results in the $1/N$ -expansion of the vector models.

We begin this Section with a description of the $1/N$ -expansion of the N -component four-Fermi theory analyzing the bubble graphs. Then we introduce functional methods and construct the $1/N$ -expansion of the $O(N)$ -symmetric φ^4 theory and nonlinear sigma model. At the end we discuss the factorization in the $O(N)$ vector models at large N .

6.1 Four-Fermi theory

The action of the $O(N)$ -symmetric four-Fermi theory in a d -dimensional Euclidean space¹⁵ is defined by

$$S[\bar{\psi}, \psi] = \int d^d x \left(\bar{\psi} \hat{\partial} \psi + m \bar{\psi} \psi - \frac{G}{2} (\bar{\psi} \psi)^2 \right). \quad (6.1)$$

Here $\hat{\partial} = \gamma_\mu \partial_\mu$ and

$$\psi = (\psi_1, \dots, \psi_N) \quad (6.2)$$

is a spinor field which forms an N -component vector in an internal-symmetry space so that

$$\bar{\psi} \psi = \sum_{i=1}^N \bar{\psi}_i \psi_i. \quad (6.3)$$

¹⁵In $d = 2$ this model was studied in the large- N limit in Ref. [47] and is often called the Gross–Neveu model.

The dimension of the four-Fermi coupling constant G is

$$\dim[G] = m^{2-d}. \quad (6.4)$$

For this reason, the perturbation theory for the four-Fermi interaction is renormalizable in $d = 2$ but is non-renormalizable for $d > 2$ (and, in particular, in $d = 4$). This is why the old Fermi theory of weak interactions was replaced by the modern electroweak theory, where the interaction is mediated by the $W^\pm$ and Z bosons.

The action (6.1) can be equivalently rewritten as¹⁶

$$S[\bar{\psi}, \psi, \chi] = \int d^d x \left(\bar{\psi} \hat{\partial} \psi + m \bar{\psi} \psi - \chi \bar{\psi} \psi + \frac{\chi^2}{2G} \right), \quad (6.5)$$

where χ is an auxiliary field. The two forms of the action, (6.1) and (6.5), are equivalent due to the equation of motion which reads in the operator notation as

$$\chi = G : \bar{\psi} \psi :, \quad (6.6)$$

where $:$ stands for the normal ordering of operators. Equation (6.6) can be derived by varying the action (6.5) with respect to χ .

In the path-integral quantization, where the partition function is defined by

$$Z = \int D\chi D\bar{\psi} D\psi e^{-S[\bar{\psi}, \psi, \chi]} \quad (6.7)$$

with $S[\bar{\psi}, \psi, \chi]$ given by Eq. (6.5), the action (6.1) appears after performing the Gaussian integral over χ . Therefore, one alternatively gets

$$Z = \int D\bar{\psi} D\psi e^{-S[\bar{\psi}, \psi]} \quad (6.8)$$

with $S[\bar{\psi}, \psi]$ given by Eq. (6.1).

The perturbative expansion of the $O(N)$ -symmetric four-Fermi theory can be conveniently represented using the formulation (6.5) via the auxiliary field χ . Then the diagrams are of the type of those in Yukawa theory, and resemble the ones for QED with $\bar{\psi}$ and ψ being an analog of the electron-positron field and χ being an analog of the photon field. However, the auxiliary field $\chi(x)$ does not propagate, since it follows from the action (6.5) that

$$D_0(x-y) \equiv \langle \chi(x) \chi(y) \rangle_{\text{Gauss}} = G \delta^{(d)}(x-y) \quad (6.9)$$

or

$$D_0(p) \equiv \langle \chi(-p) \chi(p) \rangle_{\text{Gauss}} = G \quad (6.10)$$

¹⁶The introduction of the auxiliary field is often referred to as the Hubbard Stratonovich transformation in analogy with statistical mechanics.

Fig. 28: Diagrams of the second order of perturbation theory for the four-Fermi vertex. The diagram b) involves the sum over the $O(N)$ indices.

Fig. 29: One-loop diagrams for the propagator of the ψ -field. The diagram b) involves the sum over the $O(N)$ indices.

in momentum space.

It is convenient to represent the four-Fermi vertex as the sum of two terms

$$\text{Diagram 1} = \text{Diagram 2} - \text{Diagram 3} \quad (6.11)$$

where the empty space inside the vertex is associated with the propagator (6.9) (or (6.10) in momentum space). The relative minus sign makes the vertex antisymmetric in both incoming and outgoing fermions as is prescribed by the Fermi statistics.

The diagrams that contribute to second order in G for the four-Fermi vertex are depicted, in these notations, in Fig. 28. The $O(N)$ indices propagate through the solid lines so that the closed line in the diagram in Fig. 28b corresponds to the sum over the $O(N)$ indices which results in a factor of N . Analogous one-loop diagrams for the propagator of the ψ -field are depicted in Fig. 29.

Fig. 30: Bubble diagram which survives the large- N limit of the $O(N)$ vector models.**Remark on four-Fermi theory in $d = 2$**

The one-loop Gell-Mann-Low function of the four-Fermi theory in $d = 2$ is given by the diagrams in Fig. 28 which are logarithmically divergent in $d = 2$. Noting that the diagrams in Fig. 29 do not contribute to the wave-function renormalization of the ψ -field, which emerges to the next order in G , one gets

$$\mathcal{B}(G) = -\frac{(N-1)G^2}{2\pi}. \quad (6.12)$$

The four-Fermi theory in 2 dimensions is asymptotically free as was first noted by Anselm [48] and rediscovered in Ref. [47].

The vanishing of the one-loop Gell-Mann-Low function in the Gross-Neveu model for $N = 1$ is related to the same phenomenon in the Thirring model. The latter model is associated with the vector-like interaction $(\bar{\psi}\gamma_\mu\psi)^2$ of one species of fermions with γ_μ being the γ -matrices in 2 dimensions. Since a bispinor has in $d = 2$ only two components ψ_1 and ψ_2 , both the vector-like and the scalar-like interaction (6.1) for $N = 1$ reduce to $\bar{\psi}_1\psi_1\bar{\psi}_2\psi_2$ since the square of a Grassmann variable vanishes. Therefore, these two models coincide. For the Thirring model, the vanishing of the Gell-Mann-Low function for any G was shown by Johnson [49] to all loops.

6.2 Bubble graphs as zeroth order in $1/N$

The perturbation-theory expansion of the $O(N)$ -symmetric four-Fermi theory contains, in particular, the diagrams of the type depicted in Fig. 30 which are called *bubble graphs*. Since each bubble has a factor of N , the contribution of the n -bubble graph is $\propto G^{n+1}N^n$ which is of the order of

$$G^{n+1}N^n \sim G \quad (6.13)$$

as $N \rightarrow \infty$ since

$$G \sim \frac{1}{N}. \quad (6.14)$$

Therefore, all the bubble graphs are essential to the leading order in $1/N$.

Let us denote

$$\text{wavy line} = G + \dots + G^2 \text{ (bubble)} + G^{n+1} \text{ (n loops)} + \dots \quad (6.15)$$

Fig. 31: Some diagrams of the $1/N$ -expansion for the $O(N)$ four-Fermi theory. The wavy line represents the (infinite) sum of the bubble graphs (6.15).

In fact the wavy line is nothing but the propagator D of the χ field with the bubble corrections included. The first term G on the RHS of Eq. (6.15) is nothing but the free propagator (6.10).

Summing the geometric series of the fermion-loop chains on the RHS of Eq. (6.15), one gets analytically¹⁷

$$D^{-1}(p) = \frac{1}{G} - N \int \frac{d^d k}{(2\pi)^d} \frac{\text{sp}(\hat{k} + im)(\hat{k} + \hat{p} + im)}{(k^2 + m^2)((k+p)^2 + m^2)}. \quad (6.16)$$

This determines the exact propagator of the χ field at large N . It is $\mathcal{O}(N^{-1})$ since the coupling G is included in the definition of the propagator.

The idea is now to change the order of summation of diagrams of perturbation theory using $1/N$ rather than G as the expansion parameter. Therefore, the zeroth-order propagator of the expansion in $1/N$ is defined as the sum over the bubble graphs (6.15) which is given by Eq. (6.16). Some of the diagrams of the new expansion for the four-Fermi vertex are depicted in Fig. 31. The first diagram is proportional to G while the second and third ones are proportional to G^2 or G^3 , respectively, and therefore are of order $\mathcal{O}(N^{-1})$ or $\mathcal{O}(N^{-2})$ with respect to the first diagram. The perturbation theory is thus rearranged as the $1/N$ -expansion.

The general structure of the $1/N$ -expansion is the same for all vector models, say, for the N -component φ^4 theory which is considered in the next Subsection.

The main advantage of the expansion in $1/N$ for the four-Fermi interaction, over the perturbation theory, is that it is renormalizable in $d < 4$ while the perturbation-theory expansion in G is renormalizable only in $d = 2$. Moreover, the $1/N$ -expansion of the four-Fermi theory in $2 < d < 4$ demonstrates [46] an existence of an ultraviolet-stable fixed point, i.e. a nontrivial zero of the Gell-Mann-Low function.

Remark on renormalizability of four-Fermi theory

The $1/N$ -expansion of the four-Fermi theory is renormalizable in $2 \leq d < 4$ (but not in $d = 4$). In order to demonstrate renormalizability, let us analyze indices of the diagrams of

¹⁷Recall that the free Euclidean fermionic propagator is given by $S_0(p) = (i\hat{p} + m)^{-1}$ due to Eqs. (6.5), (6.7) and the additional minus sign is associated with the fermion loop.

the 1/N-expansion.

First of all, we get rid of an ultraviolet divergence of the integral over the d -momentum k in Eq. (6.16). The divergent part of the integral is proportional to Λ^{d-2} (logarithmically divergent in $d = 2$) with Λ being an ultraviolet cutoff. It can be cancelled by choosing

$$G = \frac{g^2}{N} \Lambda^{2-d}, \quad (6.17)$$

where g^2 is a proper dimensionless constant which is not necessarily positive since the four-Fermi theory is stable with either sign of G . The power of Λ in Eq. (6.17) is consistent with the dimension of G . This prescription works for $2 < d < 4$ where there is only one divergent term while another divergency $\propto p^2 \ln \Lambda$ emerges additionally in $d = 4$. This is why the consideration is not applicable in $d = 4$.

The propagator $D(p)$ is therefore finite, and behaves at large momenta $|p| \gg m$ as

$$D(p) \propto \frac{1}{|p|^{d-2}}. \quad (6.18)$$

The standard power-counting arguments then show that the only divergent diagrams appear in the propagators of the ψ and χ fields, and in the $\bar{\psi}\chi\psi$ three-vertex. These divergencies can be removed by a renormalization of the coupling g , mass, and wave functions of ψ and χ , which is a property of renormalizable theory. For more detail about the renormalizability of the 1/N-expansion for the four-Fermi theory in $2 \leq d < 4$, see Ref. [50].

Remark on scale invariance at fixed point

The renormalization group says that

$$\mu = \Lambda \exp \left[- \int \frac{dg^2}{B(g^2)} \right], \quad (6.19)$$

which is the same as Eq. (3.72) since the correlation length $\sim \mu$. If B has a nontrivial fixed point g_*^2 near which

$$B(g^2) = b(g^2 - g_*^2) \quad (6.20)$$

with $b < 0$, then the substitution into Eq. (6.19) gives

$$g^2 = g_*^2 + \left(\frac{\mu}{\Lambda} \right)^{-b}. \quad (6.21)$$

Therefore, the approach to the critical point is power-like rather than logarithmic as for the case of $g_*^2 = 0$ when

$$B(g^2) = bg^4. \quad (6.22)$$

The latter behavior of B results, after the substitution into Eq. (6.19), in the logarithmic dependence

$$g^2 = \frac{1}{b \ln \frac{\mu}{\Lambda}} \quad (6.23)$$

when $b < 0$ which is associated with asymptotic freedom.

If g is chosen exactly at the critical point g_* , then the renormalization-group equations

$$\frac{\mu d \ln \Gamma_i}{d\mu} = \gamma_i(g^2), \quad (6.24)$$

where Γ_i stands generically either for vertices or for inverse propagators, possess the scale invariant solutions

$$\Gamma_i \propto \mu^{\gamma_i(g^2)}. \quad (6.25)$$

This shows the relation between scale invariance and the vanishing of the Gell-Mann-Low function.

For the four-Fermi theory in $d = 3$, Eq. (6.25) yields

$$S(p) = \frac{1}{i\bar{p}} \left(\frac{p^2}{\mu^2} \right)^{\gamma_\psi}, \quad (6.26)$$

$$D(p) = \frac{8}{N|p|} \left(\frac{p^2}{\mu^2} \right)^{\gamma_\chi}, \quad (6.27)$$

$$\Gamma(p_1, p_2) = \left(\frac{\mu^2}{p_1^2} \right)^{\gamma_\psi} f \left(\frac{p_2^2}{p_1^2}, \frac{p_1 p_2}{p_1^2} \right), \quad (6.28)$$

where f is an arbitrary function of the dimensionless ratios which is not determined by scale invariance. The indices here obey the relation

$$\gamma_\psi = \gamma_\psi + \frac{1}{2} \gamma_\chi \quad (6.29)$$

which guarantees scale invariance.

Remark on conformal invariance at fixed point

Scale invariance implies, in a renormalizable quantum field theory, more general conformal invariance as is first pointed out in Refs. [51, 52]. The conformal group in a d -dimensional space-time has $(d+1)(d+2)/2$ parameters as is illustrated by Table 1. More about the conformal group can be found in the lecture by Jackiw [53].

A heuristic proof [51] of the fact that scale invariance implies conformal one is based on the explicit form of the dilatation D_μ and conformal K_μ^α currents, which are associated with the dilatation and the special conformal transformation, via the energy-momentum tensor $\theta_{\mu\nu}$:

$$D_\mu = x_\nu \theta_{\mu\nu} \quad (6.30)$$

Group	Transformations	# Parameters
Lorentz	$\frac{d(d-1)}{2}$ rotations	$\frac{d(d-1)}{2}$
Poincaré	+ d translations	$\frac{d(d+1)}{2}$
Weyl	+ 1 dilatation	$\frac{d^2+d+2}{2}$
Conformal	+ d special conformal	$\frac{(d+1)(d+2)}{2}$

Table 1: Contents and the number of parameters of groups of space-time symmetry.

and

$$K_\mu^\alpha = (2x_\nu x^\alpha - x^2 \delta_\nu^\alpha) \theta_{\mu\nu}. \quad (6.31)$$

Differentiating, we get

$$\partial_\mu D_\mu = \theta_{\mu\mu}, \quad (6.32)$$

and

$$\partial_\mu K_\mu^\alpha = 2x^\alpha \theta_{\mu\mu}. \quad (6.33)$$

Therefore, both the dilatation and conformal currents vanish simultaneously when $\theta_{\mu\nu}$ is traceless which is provided, in turn, by the vanishing of the Gell-Mann–Low function.

Conformal invariance completely fixes three-vertices as was first shown by Polyakov [54] for scalar theories. The proper formula for the four-Fermi theory reads [55]

$$\Gamma(p_1, p_2) \propto \int \frac{d^d k}{(2\pi)^d} \frac{\hat{k} + \hat{p}_1}{[(k + p_1)^2]^{1+\gamma_\lambda/2}} \frac{\hat{k} + \hat{p}_2}{[(k + p_2)^2]^{1+\gamma_\lambda/2}} \frac{1}{|k|^{d-2+2\gamma_\phi-\gamma_\lambda/2}}, \quad (6.34)$$

where the indices are related by Eq. (6.29) but can be arbitrary otherwise¹⁸.

Equation (6.34), which results from conformal invariance, unambiguously fixes the function f in Eq. (6.28). In contrast to infinite-dimensional conformal symmetry in $d = 2$, the conformal group in $d > 2$ is less restrictive. It fixes only the tree-point vertex while, say, the four-point vertex remains an unknown function of two variables.

¹⁸The only restriction $\gamma_\phi \geq 0$ is imposed by the Källén–Lehmann representation of the propagator while there is no such restriction on γ_λ since it is a composite field.

Remark on broken scale invariance

Scale (and conformal) invariance at a fixed point $g = g_*$ holds only for large momenta $|p| \gg m$. For smaller values of momenta, scale invariance is broken by masses. In fact, any dimensional parameter breaks scale invariance. If the bare coupling g is chosen in the vicinity of g_* according to Eq. (6.21), then scale invariance holds even in the massless case only for $|p| \gg \mu$ while it is broken if $|p| \lesssim \mu$.

6.3 Functional methods for φ^4 theory

The large- N solution of the $O(N)$ vector models, which is given by the sum of the bubble graphs, can alternatively be obtained by evaluating the path integral at large N by the saddle-point method. We shall restrict ourselves to the scalar $O(N)$ -symmetric φ^4 theory while the analysis of the four-Fermi theory is quite analogous.

The action of the $O(N)$ -symmetric φ^4 theory reads

$$S[\varphi^a] = \int d^d x \left[\frac{1}{2} (\partial_\mu \varphi^a)^2 + \frac{1}{2} m^2 \varphi^a \varphi^a + \frac{\lambda}{8} (\varphi^a \varphi^a)^2 \right] \quad (6.35)$$

where

$$\varphi^a = (\varphi^1, \dots, \varphi^N). \quad (6.36)$$

The coupling λ in the action (6.35) must be positive for the theory to be well defined. The vertices of Feynman diagrams are associated with $-\lambda$.

Introducing the auxiliary field $\chi(x)$ as in Subsect. 6.1, the action (6.35) can be rewritten as

$$S[\varphi^a, \chi] = \int d^d x \left[\frac{1}{2} \varphi^a (-\partial_\mu^2 + m^2 + \chi) \varphi^a - \frac{\chi^2}{2\lambda} \right]. \quad (6.37)$$

The two forms are equivalent due to the equation of motion

$$\chi = \frac{\lambda}{2} : \varphi^a \varphi^a :. \quad (6.38)$$

In other words χ is again a composite field.

The correlators of φ 's and χ 's are determined by the generating functional

$$Z[J^a, K] = \int_{\dagger} D\chi(x) \int D\varphi^a(x) e^{-S[\varphi^a, \chi] + \int d^d x J^a(x) \varphi^a(x) + \int d^d x K(x) \chi(x)}, \quad (6.39)$$

which is a functional of the sources J^a and K for the fields φ^a and χ .

To make the path integral over $\chi(x)$ in Eq. (6.39) convergent, we integrate at each point x over a contour which is parallel to imaginary axis. This is specific to the Euclidean formulation. The propagator of the χ -field in the Gaussian approximation reads

$$D_0(p) = \langle \chi(-p) \chi(p) \rangle_{\text{Gauss}} = -\lambda. \quad (6.40)$$

which reproduces the four-boson vertex of perturbation theory.

Since the integral over φ^a is Gaussian, it can be expressed via the Green function

$$G(x, y; \chi) = \left\langle y \left| \frac{1}{-\partial_\mu^2 + m^2 + \chi} \right| x \right\rangle \quad (6.41)$$

as

$$Z[J^a, K] = \int_{\dagger} D\chi(x) e^{\int d^d x \frac{\chi^2}{2\lambda} + \frac{1}{2} \int d^d x d^d y J^a(x) G(x, y; \chi) J^a(y) + \int d^d x K(x) \chi(x) - \frac{N}{2} \text{Sp} \ln G^{-1}[\chi]}. \quad (6.42)$$

We have used here the obvious notation

$$G^{-1}[\chi] = -\partial_\mu^2 + m^2 + \chi. \quad (6.43)$$

It will also be convenient to use the short-hand notation

$$g \circ f = \langle g | f \rangle \equiv \int d^d x f(x) g(x). \quad (6.44)$$

Then, Eq. (6.42) can be rewritten as

$$Z[J^a, K] = \int_{\dagger} D\chi(x) e^{\frac{\lambda \chi^2}{2\lambda} + \frac{1}{2} J^a \circ G[\chi] \circ J^a + K \circ \chi - \frac{N}{2} \text{Sp} \ln G^{-1}[\chi]}. \quad (6.45)$$

The exponent in Eq. (6.45) is $\mathcal{O}(N)$ at large N so the path integral can be evaluated as $N \rightarrow \infty$ by the saddle-point method. The saddle-point field configuration

$$\chi(x) = \chi_{\text{sp}}(x) \quad (6.46)$$

is determined (implicitly) by the saddle-point equation

$$\chi_{\text{sp}}(x) - \frac{\lambda N}{2} G(x, x; \chi_{\text{sp}}) + \frac{\lambda}{2} J^a \circ G(\cdot, x; \chi_{\text{sp}}) G(x, \cdot; \chi_{\text{sp}}) \circ J^a + \lambda K(x) = 0. \quad (6.47)$$

If $K \sim 1/\lambda$, each term here is $\mathcal{O}(1)$ since

$$\lambda \sim \frac{1}{N} \quad (6.48)$$

in analogy with Eq. (6.14).

When the sources J^a and K vanish so that the last two terms on the LHS of Eq. (6.47) equal zero, this equation reduces to

$$\chi_{\text{sp}} - \frac{\lambda N}{2} G(x, x; \chi_{\text{sp}}) = 0. \quad (6.49)$$

Its solution is x independent due to translational invariance and can be parametrized as

$$\chi_{\text{sp}} = m_{\text{R}}^2 - m^2 \quad (6.50)$$

where m and m_{R} are the bare and renormalized mass, respectively. Equation (6.49) then reduces to the standard formula [46]

$$m^2 = m_{\text{R}}^2 - \frac{\lambda N}{2} \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2 + m_{\text{R}}^2)} \quad (6.51)$$

for the mass renormalization at large N .

To take into account fluctuations around the saddle point, we expand

$$\chi(x) = \chi_{\text{sp}} + \delta\chi(x) \quad (6.52)$$

where

$$\delta\chi(x) \sim \sqrt{\lambda} \sim N^{-1/2}. \quad (6.53)$$

The Gaussian integration over $\delta\chi(x)$ determines the pre-exponential factor in (6.45).

To construct the $1/N$ expansion of the generating functional (6.45), it is convenient to use the generating functional for connected Green's functions. It is usually denoted by $W[J^a, K]$ and is related to the partition function (6.39) by

$$Z[J^a, K] = e^{W[J^a, K]}. \quad (6.54)$$

Then we get

$$W[J^a, K] = \frac{1}{2\lambda} \chi_{\text{sp}} \circ \chi_{\text{sp}} - \frac{N}{2} \text{Sp} \ln G^{-1}[\chi_{\text{sp}}] + \frac{1}{2} J^a \circ G[\chi_{\text{sp}}] \circ J^a + K \circ \chi_{\text{sp}} - \frac{1}{2} \text{Sp} \ln(\lambda D^{-1}[\chi_{\text{sp}}]) + \mathcal{O}(N^{-1}), \quad (6.55)$$

where

$$D^{-1}(x, y; \chi) = -\frac{1}{\lambda} \delta^{(d)}(x - y) - \frac{N}{2} G(x, y; \chi) G(y, x; \chi) + J^a \circ G(\cdot, x; \chi) G(x, y; \chi) G(y, \cdot; \chi) \circ J^a. \quad (6.56)$$

This operator emerges when integrating over the Gaussian fluctuations around the saddle point. The corresponding (last) term on the RHS of Eq. (6.55) is associated with the pre-exponential factor and, therefore, is ~ 1 .

The next terms of the $1/N$ expansion can be calculated in a systematic way by substituting (6.52) in Eq. (6.45) and performing the perturbative expansion in $\delta\chi$.

If the sources J^a and K vanish so that the saddle-point value χ_{sp} is given by the constant (6.50), then the RHS of Eq. (6.56) simplifies to

$$D^{-1}(x, y; \chi_{\text{sp}}) = -\frac{1}{\lambda} \delta^{(d)}(x - y) - \frac{N}{2} G(x, y; \chi_{\text{sp}}) G(y, x; \chi_{\text{sp}}). \quad (6.57)$$

Remembering the definition (6.41) of G and passing to the momentum-space representation, we get

$$D^{-1}(p) = -\frac{1}{\lambda} - \frac{N}{2} \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2 + m_{\text{R}}^2)((k + p)^2 + m_{\text{R}}^2)}. \quad (6.58)$$

The sign of the first term on the RHS is consistent with Eq. (6.40).

Equation (6.58) is an analog of Eq. (6.16) in the fermionic case and can be alternatively obtained by summing bubble graphs of the type in Fig. 30 for

$$D(p) = \langle \chi(-p)\chi(p) \rangle. \quad (6.59)$$

The extra symmetry factor 1/2 in Eq. (6.58) is the usual combinatoric one for bosons. Therefore, the large- N saddle-point calculation of the propagator (6.59) results precisely in the zeroth-order of the 1/ N -expansion.

We see from Eq. (6.55) the difference between perturbation theory and the 1/ N -expansion. The perturbation theory in λ can be constructed as an expansion (6.52) around the saddle point χ_{sp} given again by Eq. (6.47), with the omitted second term on the LHS, which is now justified by the fact that λ is small (even for $N \sim 1$). The second term on the RHS of Eq. (6.55), which is associated with a one-loop diagram, appears in perturbation theory as a result of Gaussian fluctuations around this saddle-point.

Remark on Gell-Mann-Low function of φ^4 theory

The one-loop Gell-Mann-Low function of the $O(N)$ -symmetric φ^4 theory in $d = 4$ is given by diagrams similar to those of Fig. 28, though now the arrows are not essential since the field is real. The diagrams are logarithmically divergent in 4 dimensions. Each diagram contribute with the positive sign while the diagram of Fig. 28b has a combinatoric factor of 1/2. The diagrams of Fig. 29 result in a mass renormalization and there is no wave-function renormalization of the φ -field in one loop so that one gets

$$\mathcal{B}(\lambda) = \frac{(N+8)\lambda^2}{16\pi^2}. \quad (6.60)$$

The positive sign in this formula is the same as for QED and is associated with “triviality” of the φ^4 theory in 4 dimensions. It is also worth noting that the coefficient $(N+8)$ is large even for $N = 1$.

6.4 Nonlinear sigma model

The nonlinear $O(N)$ sigma model¹⁹ in 2 Euclidean dimensions is defined by the partition function

$$Z = \int D\vec{n} \delta\left(\vec{n}^2 - \frac{1}{g^2}\right) e^{-\frac{1}{2} \int d^2x (\partial_\mu \vec{n})^2} \quad (6.61)$$

where

$$\vec{n} = (n_1, \dots, n_N) \quad (6.62)$$

¹⁹The name comes from elementary particle physics where a nonlinear sigma model in 4 dimensions is used as an effective Lagrangian for describing low-energy scattering of the Goldstone π -mesons.

is an $O(N)$ vector. While the action in Eq. (6.61) is pure Gaussian, the model is not free due to the constraint

$$\vec{n}^2(x) = \frac{1}{g^2}, \quad (6.63)$$

which is imposed on the $\vec{n}$ field via the (functional) delta-function.

The sigma model in $d = 2$ is sometimes considered as a toy model for QCD since it possesses:

- 1) asymptotic freedom [56];
- 2) instantons for $N = 3$ [57].

The action in Eq. (6.61) is $\sim N$ as $N \rightarrow \infty$ but the entropy, *i.e.* a contribution from the measure of integration, is also $\sim N$ so that a straightforward saddle point is not applicable.

To overcome this difficulty, we proceed as in the previous Subsection, introducing an auxiliary field $u(x)$, which is ~ 1 as $N \rightarrow \infty$, and rewrite the partition function (6.61) as

$$Z \propto \int_{\dagger} Du(x) \int D\vec{n}(x) e^{-\frac{1}{2} \int d^2x \left[(\partial_\mu \vec{n})^2 - u \left(\vec{n}^2 - \frac{1}{g^2} \right) \right]}, \quad (6.64)$$

where the contour of integration over $u(x)$ is parallel to imaginary axis.

Doing the Gaussian integration over $\vec{n}$, we get

$$Z \propto \int_{\dagger} Du(x) e^{-\frac{N}{2} \text{Sp} \ln(-\partial_\mu^2 + u(x)) + \frac{1}{2g^2} \int d^2x u(x)}. \quad (6.65)$$

The first term in the exponent is as before nothing but the sum of one-loop diagrams in 2 dimensions

$$\frac{N}{2} \text{Sp} \ln(-\partial_\mu^2 + u(x)) = \sum_n \frac{1}{n} \cdot \text{Diagram} \quad (6.66)$$

where the auxiliary field u is denoted again by the wavy line. Equation (6.65) looks very much like Eq. (6.45) if we put there $J^a = K = 0$. The difference is that the exponent in (6.65) involves the term which is linear in u , while the analogous term in (6.45) is quadratic in χ .

Now the path integral over $u(x)$ in Eq. (6.65) is a typical saddle-point one: the action $\sim N$ while the entropy ~ 1 since only one integration over u is left. The saddle-point equation for the nonlinear sigma model

$$\frac{1}{g^2} - NG(x, x; u_{sp}) = 0 \quad (6.67)$$

is quite analogous to Eq. (6.49) for the φ^4 theory while G is defined by

$$G(x, y; u) = \left\langle y \left| \frac{1}{-\partial_\mu^2 + u} \right| x \right\rangle, \quad (6.68)$$

which is an analog of Eq. (6.41).

The coupling g^2 in Eq. (6.67) is $\sim 1/N$ as is prescribed by the constraint (6.63) which involves a sum over N terms on the LHS. This guarantees that a solution to Eq. (6.67) exists. Next orders of the $1/N$ -expansion for the 2-dimensional sigma model can be constructed analogously to the previous Subsection.

The $1/N$ -expansion of the 2-dimensional nonlinear sigma model has many advantages over perturbation theory, which is usually constructed solving explicitly the constraint (6.63), say, choosing

$$n_N = \frac{1}{g} \sqrt{1 - g^2 \sum_{a=1}^{N-1} n_a^2} \quad (6.69)$$

and expanding the square root in g^2 . Only $N - 1$ dynamical degrees of freedom are left so that the $O(N)$ -symmetry is broken in perturbation theory down to $O(N - 1)$. The particles in perturbation theory are massless (like Goldstone bosons) and it suffers from infrared divergencies.

On the contrary, the solution to Eq. (6.67) has the form

$$u_{\text{sp}} = m_{\text{R}}^2 \equiv \Lambda^2 e^{-\frac{4\pi}{Ng^2}}. \quad (6.70)$$

where Λ is an ultraviolet cutoff. Therefore, all N particles acquires the same mass m_{R} in the $1/N$ -expansion so that the $O(N)$ symmetry is restored. This appearance of mass is due to dimensional transmutation which says in this case that the parameter m_{R} rather than the renormalized coupling constant g_{R}^2 is observable. The emergence of the mass cures the infrared problem.

In order to show that (6.70) is a solution to Eq. (6.67), we substitute a translationally invariant ansatz

$$u_{\text{sp}}(x) \equiv m_{\text{R}}^2. \quad (6.71)$$

Then Eq. (6.67) in the momentum space reads

$$\frac{1}{g^2} = N \int^{\Lambda} \frac{d^2 k}{(2\pi)^2} \frac{1}{k^2 + m_{\text{R}}^2} = \frac{N}{4\pi} \int_0^{\Lambda^2} \frac{dk^2}{k^2 + m_{\text{R}}^2} = \frac{N}{4\pi} \ln \frac{\Lambda^2}{m_{\text{R}}^2}. \quad (6.72)$$

The exponentiation results in Eq. (6.70).

Equation (6.72) relates the bare coupling g^2 and the cutoff Λ and allows us to calculate the Gell-Mann–Low function yielding

$$\mathcal{B}(g^2) \equiv \frac{\Lambda}{d\Lambda} \frac{dg^2}{g^2} = -\frac{Ng^4}{2\pi}. \quad (6.73)$$

The analogous one-loop perturbation-theory formula for any N reads [56]

$$\mathcal{B}(g^2) = -\frac{(N-2)g^4}{2\pi}. \quad (6.74)$$

Thus, the sigma-model is asymptotically free in 2-dimensions for $N > 2$ which is the origin of the dimensional transmutation. There is no asymptotic freedom for $N = 2$ since $O(2)$ is Abelian.

6.5 Large- N factorization in vector models

The fact that a path integral has a saddle point at large N implies a very important feature of large- N theories — the factorization. It is a general property of the large- N limit and holds not only for the $O(N)$ vector models. However, it is useful to illustrate it by these solvable examples.

The factorization at large N holds for averages of *singlet* operators, for example

$$\langle u(x_1) \dots u(x_k) \rangle \equiv Z^{-1} \int_{\mathbb{T}} Du e^{-\frac{N}{2} \text{Sp} \ln[-\partial_{\mu}^2 + u] + \frac{1}{2g^2} \int d^2 x u} u(x_1) \dots u(x_k) \quad (6.75)$$

in the 2-dimensional sigma model.

Since the path integral has a saddle point at some²⁰

$$u(x) = u_{\text{sp}}(x), \quad (6.76)$$

we get to the leading order in $1/N$:

$$\langle u(x_1) \dots u(x_k) \rangle = u_{\text{sp}}(x_1) \dots u_{\text{sp}}(x_k) + \mathcal{O}\left(\frac{1}{N}\right), \quad (6.77)$$

which can be written in the factorized form

$$\langle u(x_1) \dots u(x_k) \rangle = \langle u(x_1) \rangle \dots \langle u(x_k) \rangle + \mathcal{O}\left(\frac{1}{N}\right). \quad (6.78)$$

Therefore, u becomes “classical” as $N \rightarrow \infty$ in the sense of the $1/N$ -expansion. This is an analog of the WKB-expansion in $\hbar = 1/N$. “Quantum” corrections are suppressed as $1/N$.

We shall return to discussing the large- N factorization in the next Section when considering the large- N limit of QCD.

²⁰This saddle point is, in fact, x -independent due to translational invariance.

7 Multicolor QCD

The method of the $1/N$ -expansion can be applied to QCD. This was done by 't Hooft [2] using the inverse number of colors for the gauge group $SU(N_c)$ as an expansion parameter.

For a $SU(N_c)$ gauge theory without virtual quark loops, the expansion goes in $1/N_c^2$ and rearranges diagrams of perturbation theory according to their topology. The leading order in $1/N_c^2$ is given by planar diagrams, which have a topology of a sphere, while the expansion in $1/N_c^2$ plays the role of a topological expansion. This reminds an expansion in the string coupling constant in string models of the strong interaction, which also has a topological character.

Virtual quark loops can be easily incorporated in the $1/N_c$ -expansion. One distinguishes between the 't Hooft limit when the number of quark flavors N_f is fixed as $N_c \rightarrow \infty$ and the Veneziano limit [58] when the ratio N_f/N_c is fixed as $N_c \rightarrow \infty$. Virtual quark loops are suppressed in the 't Hooft limit as $1/N_c$ and lead in the Veneziano limit to the same topological expansion as dual-resonance models of strong interaction.

The simplification of QCD in the large- N_c limit is due to the fact that the number of planar graphs grows with the number of vertices only exponentially rather than factorially as do the total number of graphs. Correlators of gauge invariant operators factorize in the large- N_c limit which looks like the leading-order term of a “semiclassical” WKB-expansion in $1/N_c$.

We begin this Section with a description of the double-line representation of diagrams of QCD perturbation theory and rearrange it as the topological expansion in $1/N_c$. Then we discuss some properties of the $1/N_c$ -expansion for a generic matrix-valued field.

7.1 Index graphs or fatgraphs

In order to describe the $1/N_c$ -expansion of QCD, whose extension to N_c colors has already been considered in Subsect. 2.2, it is convenient to use the matrix-field representation (2.24). In this Section we shall use a slightly different definition

$$[A_\mu(x)]^{ij} = \sum_a A_\mu^a(x) [t^a]^{ij}, \quad (7.1)$$

which is similar to that used by 't Hooft [2] and differs from (2.24) by a factor of ig

$$A_\mu^{ij}(x) = ig A_\mu^{ij}(x). \quad (7.2)$$

The matrix (7.1) is Hermitean.

The propagator of the matrix field $A^{ij}(x)$, in this notation, takes on the form

$$\langle A_\mu^{ij}(x) A_\nu^{kl}(y) \rangle_{\text{Gauss}} = \frac{1}{2} \left(\delta^{il} \delta^{kj} - \frac{1}{N_c} \delta^{ij} \delta^{kl} \right) D_{\mu\nu}(x-y), \quad (7.3)$$

where we have assumed, as usual, a gauge fixing to define the propagator in perturbation theory. For instance, one has

$$D_{\mu\nu}(x-y) = \frac{1}{4\pi^2} \frac{\delta_{\mu\nu}}{(x-y)^2} \quad (7.4)$$

in the Feynman gauge.

Equation (7.3) can be derived from the standard formula

$$\langle A_\mu^a(x) A_\nu^b(y) \rangle_{\text{Gauss}} = \delta^{ab} D_{\mu\nu}(x-y) \quad (7.5)$$

multiplying by the generators of the $SU(N_c)$ gauge group according to the definition (2.24) and using the completeness condition

$$\sum_{a=1}^{N_c^2-1} (t^a)^{ij} (t^a)^{kl} = \frac{1}{2} \left(\delta^{il} \delta^{kj} - \frac{1}{N_c} \delta^{ij} \delta^{kl} \right) \quad \boxed{\text{for } SU(N_c)}, \quad (7.6)$$

where the factor of $1/2$ is due to the normalization (2.25).

We concentrate in this Section only on the structure of diagrams in the index space, *i.e.* the space of the indices associated with the $SU(N_c)$ group. We shall not consider, in most cases, space-time structures of diagrams which are prescribed by Feynman's rules.

Omitting at large N_c the second term in parentheses on the RHS of Eq. (7.3), we depict the propagator by the double line

$$\langle A_\mu^{ij}(x) A_\nu^{kl}(y) \rangle_{\text{Gauss}} \propto \delta^{il} \delta^{kj} = \begin{array}{c} i \longrightarrow l \\ j \longleftarrow k \end{array} \quad (7.7)$$

Each line represents the Kronecker delta-symbol and has orientation which is indicated by arrows. This notation is obviously consistent with the space-time structure of the propagator which describes a propagation from x to y .

The arrows are due to the fact that the matrix A_μ^{ij} is Hermitean and its off-diagonal components are complex conjugate. The independent fields are, say, the complex fields A_μ^{ij} for $i > j$ and the diagonal real fields A_μ^{ii} . The arrow represents the direction of the propagation of the complex field A_μ^{ij} for $i > j$ while the complex-conjugate one, $A_\mu^{ji} = (A_\mu^{ij})^*$, propagates in the opposite direction. For the real fields A_μ^{ii} , the arrows are not essential.

The double-line notation (7.7) looks similar to that of Subsect. 3.5. The reason for that is deep: double lines appear generically in all models describing *matrix* fields in contrast to *vector* (in internal symmetry space) fields whose propagators are depicted by single lines as in the previous Section.

The three-gluon vertex, which is generated by the action (2.33), is depicted in the double-line notations as

$$\begin{array}{c} \text{Diagram 1: } \begin{array}{c} \text{Line 1 (1)} \\ \text{Line 2 (2)} \\ \text{Line 3 (3)} \end{array} - \begin{array}{c} \text{Line 1 (1)} \\ \text{Line 2 (2)} \\ \text{Line 3 (3)} \end{array} \propto g (\delta^{11} \delta^{22} \delta^{33} \delta^{12} \delta^{23} \delta^{31}) \end{array} \quad (7.8)$$

where the subscripts 1, 2 or 3 refer to each of the three gluons. The relative minus sign is due to the commutator in the cubic in A term in the action (2.33). The color part of the three-vertex is antisymmetric under interchanging the gluons. The space-time structure, which reads in the momentum space as

$$\gamma_{\mu_1 \mu_2 \mu_3}(p_1, p_2, p_3) = \delta_{\mu_1 \mu_2} (p_1 - p_2)_{\mu_3} + \delta_{\mu_2 \mu_3} (p_2 - p_3)_{\mu_1} + \delta_{\mu_1 \mu_3} (p_3 - p_1)_{\mu_2}, \quad (7.9)$$

Fig. 32: Double-line representation of a one-loop diagram for the gluon propagator. The sum over the N_c indices is associated with the closed index line. The contribution of this diagram is $\sim g^2 N_c \sim 1$.

is antisymmetric as well. We consider all three gluons as incoming so that their momenta obey $p_1 + p_2 + p_3 = 0$. The full vertex is symmetric as is prescribed by Bose statistics.

The color structure in Eq. (7.8) can alternatively be obtained by multiplying the standard vertex

$$\Gamma_{\mu_1 \mu_2 \mu_3}^{a_1 a_2 a_3}(p_1, p_2, p_3) = f^{a_1 a_2 a_3} \gamma_{\mu_1 \mu_2 \mu_3}(p_1, p_2, p_3) \quad (7.10)$$

by $(t^{a_1})^{i_1 j_1} (t^{a_2})^{i_2 j_2} (t^{a_3})^{i_3 j_3}$, with f^{abc} being the structure constants of the $SU(N_c)$ group, and using the formula

$$f^{a_1 a_2 a_3} (t^{a_1})^{i_1 j_1} (t^{a_2})^{i_2 j_2} (t^{a_3})^{i_3 j_3} = \frac{1}{2} (\delta^{i_1 j_3} \delta^{i_2 j_1} \delta^{i_3 j_2} - \delta^{i_1 j_2} \delta^{i_3 j_1} \delta^{i_2 j_3}), \quad (7.11)$$

which is a consequence of the completeness condition (7.6).

The four-gluon vertex involves six terms — each of them is depicted by a cross — which differ by interchanging of the color indices. We depict the color structure of the four-gluon vertex for simplicity in the case when $i_1 = j_2 = i$, $i_2 = j_3 = j$, $i_3 = j_4 = k$, $i_4 = j_1 = l$ but i, j, k, l take on different values. Then only the following term is left

$$\begin{array}{c} l \\ \downarrow \\ i \text{---} \text{---} \downarrow \\ \downarrow \text{---} \downarrow \\ k \end{array} \propto g^2 \quad (7.12)$$

and there are no deltas on the RHS since the color structure is fixed. In other words, we pick up only one color structure by equaling indices pairwise.

The diagrams of perturbation theory can now be completely rewritten in the double-line notation [2]. The simplest one which describes the one-loop correction to the gluon propagator is depicted in Fig. 32.²¹ This diagram involves two three-gluon vertices and a sum over the N_c indices which is associated with the closed index line analogous to Eq. (3.55). Therefore, the contribution of this diagram is $\sim g^2 N_c$.

In order for the large- N_c limit to be nontrivial, the bare coupling constant g^2 should satisfy

$$g^2 \sim \frac{1}{N_c}. \quad (7.13)$$

²¹Here and in the most figures below the arrows of the index lines are omitted for simplicity.

This dependence on N_c is similar to Eqs. (6.14) and (6.48) for the vector models and is prescribed by the asymptotic-freedom formula

$$g^2 = \frac{24\pi^2}{11N_c \ln(\Lambda/\Lambda_{QCD})} \quad (7.14)$$

of the pure $SU(N_c)$ gauge theory.

Thus, the contribution of the diagram of Fig. 32 is of order

$$\text{Fig. 32} \sim g^2 N_c \sim 1 \quad (7.15)$$

in the large- N_c limit.

The double lines of the diagram in Fig. 32 can be viewed as bounding a piece of a plane. Therefore, these lines represent a two-dimensional object rather than a one-dimensional one as the single lines do in vector models. These double-line graphs are often called in mathematics the *band* graphs or *fatgraphs*. We shall see below their connection with Riemann surfaces.

Remark on the $U(N_c)$ gauge group

As is said above, the second term in the parentheses on the RHS of Eq. (7.6) can be omitted at large N_c . Such a completeness condition emerges for the $U(N_c)$ group whose generators T^A ($A = 1, \dots, N_c^2$) are

$$T^A = \left(t^a, \frac{1}{\sqrt{2N}}\right), \quad \text{tr } T^A T^B = \frac{1}{2} \delta^{AB}. \quad (7.16)$$

They obey the completeness condition

$$\sum_{A=1}^{N_c^2} (T^A)^{ij} (T^A)^{kl} = \frac{1}{2} \delta^{ij} \delta^{kl} \quad \boxed{\text{for } U(N_c)}. \quad (7.17)$$

The point is that elements of both the $SU(N_c)$ group and the $U(N_c)$ group can be represented in the form

$$U = e^{iB}, \quad (7.18)$$

where B is a general Hermitean matrix for $U(N_c)$ and a traceless Hermitean matrix for $SU(N_c)$.

Therefore, the double-line representation of perturbation-theory diagrams which is described in this Section holds, strictly speaking, only for the $U(N_c)$ gauge group. However, the large- N_c limit of both the $U(N_c)$ group and the $SU(N_c)$ group is the same.

Fig. 33: Double-line representation of a four-loop diagram for the gluon propagator. The sum over the N_c indices is associated with each of the four closed index lines whose number is equal to the number of loops. The contribution of this diagram is $\sim g^8 N_c^4 \sim 1$.

Fig. 34: Double-line representation of a three-loop non-planar diagram for the gluon propagator. The diagram has six three-gluon vertices but only one closed index line (while three loops!). The order of this diagram is $\sim g^6 N_c \sim 1/N_c^2$.

7.2 Planar and non-planar graphs

The double-line representation of perturbation theory diagrams in the index space is very convenient to estimate their orders in $1/N_c$. Each three- or four-gluon vertex contributes a factor of g or g^2 , respectively. Each closed index line contributes a factor of N_c . The order of g in $1/N_c$ is given by Eq. (7.13).

Let us consider a typical diagram for the gluon propagator depicted in Fig. 33. It has eight three-gluon vertices and four closed index lines which coincides with the number of loops. Therefore, the order of this diagram in $1/N_c$ is

$$\text{Fig. 33} \sim (g^2 N_c)^4 \sim 1. \quad (7.19)$$

The diagrams of the type in Fig. 33, which can be drawn on a sheet of a paper without crossing any lines, are called the *planar* diagrams. For such diagrams, an adding of a loop inevitably results in adding of two three-gluon (or one four-gluon) vertex. A planar diagram with n_2 loops has n_2 closed index lines. It is of order

$$n_2\text{-loop planar diagram} \sim (g^2 N_c)^{n_2} \sim 1, \quad (7.20)$$

so that all planar diagrams survive in the large- N_c limit.

Let us now consider a non-planar diagram of the type depicted in Fig. 34. This diagram is a three-loop one and has six three-gluon vertices. The crossing of the two lines in the middle does not correspond to a four-gluon vertex and is merely due to the fact that the diagram cannot be drawn on a sheet of a paper without crossing the lines. The diagram has

only one closed index line. The order of this diagram in $1/N_c$ is

$$\text{Fig. 34} \sim g^6 N_c \sim \frac{1}{N_c^2}. \quad (7.21)$$

It is therefore suppressed at large N_c by $1/N_c^2$.

The non-planar diagram in Fig. 34 can be drawn without line-crossing on a surface with one handle which is usually called in mathematics a torus or the surface of genus one. A plane is then equivalent to a sphere and has genus zero. Adding a handle to a surface produces a *hole* according to mathematical terminology. A general Riemann surface with h holes has genus h .

The above evaluations of the order of the diagrams in Figs. 32–34 can now be described by the unique formula

$$\text{genus-}h \text{ diagram} \sim \left(\frac{1}{N_c^2} \right)^{\text{genus}}. \quad (7.22)$$

Thus, the expansion in $1/N_c$ rearranges perturbation-theory diagrams according to their topology [2]. For this reason, it is referred to as the *topological expansion* or the *genus expansion*. The general proof of Eq. (7.22) for an arbitrary diagram is given in Subsect. 7.4.

Only planar diagrams, which are associated with genus zero, survive in the large- N_c limit. This class of diagrams is an analog of the bubble graphs in the vector models. However, the problem of summing the planar graphs is much more complicated than that of summing the bubble graphs. Nevertheless, it is simpler than the problem of summing all the graphs, since the number of the planar graphs with n_0 vertices grows at large n_0 exponentially [59, 60]

$$\#_p(n_0) \equiv \# \text{ of planar graphs} \sim \text{const}^{n_0}, \quad (7.23)$$

while the number of all the graphs grows with n_0 factorially. There is no dependence in Eq. (7.23) on the number of external lines of a planar graph which is assumed to be much less than n_0 .

It is instructive to see the difference between the planar diagrams and, for instance, the ladder diagrams which describe e^+e^- elastic scattering in QED. Let the ladder has n rungs. Then there are $n!$ ladder diagrams, but only one of them is planar. This simple example shows why the number of planar graphs is much smaller than the number of all graphs, most of which are non-planar.

In the rest of these lecture notes, we shall discuss what is known about solving the problem of summing the planar graphs.

7.3 Planar and non-planar graphs (the boundaries)

Equation (7.22) holds, strictly speaking, only for the gluon propagator while the contribution of all planar diagrams to a connected n -point Green's function is $\sim g^{n-2}$ which is its natural order in $1/N_c$. Say, the three-gluon Green's function is $\sim g$, the four-gluon one is $\sim g^2$ and so on. In order to make contributions of all planar diagrams to be of the same order ~ 1 in

Fig. 35: Generic index diagram with $n_0 = 10$ vertices, $n_1 = 10$ gluon propagators, $n_2 = 4$ closed index lines, and $B = 1$ boundary. The color indices of the external lines are contracted by the Kronecker deltas (represented by the single lines) in a cyclic order. The extra factor of $1/N_c$ is due to the normalization (7.24). Its order in $1/N_c$ is $\sim 1/N_c^2$ in accord with Eq. (7.22).

the large- N_c limit, independently of the number of external lines, it is convenient to contract the Kronecker deltas associated with external lines.

Let us do this in a cyclic order as is depicted in Fig. 35 for a generic connected diagram with three external gluon lines. The extra deltas which are added to contract the color indices are depicted by the single lines. They can be viewed as a *boundary* of the given diagram. The actual size of the boundary is not essential — it can be shrunk to a point. Then a bounded piece of a plane will be topologically equivalent to a sphere with a puncture. I shall prefer to draw planar diagrams in a plane with an extended boundary (boundaries) rather than in a sphere with a puncture (punctures).

It is clear from the graphic representation that the diagram in Fig. 35 is associated with the trace over the color indices of the three-point Green's function

$$G_{\mu_1 \mu_2 \mu_3}^{(3)}(x_1, x_2, x_3) \equiv \frac{(ig)^3}{N_c} \langle \text{tr} [A_{\mu_1}(x_1) A_{\mu_2}(x_2) A_{\mu_3}(x_3)] \rangle. \quad (7.24)$$

We have introduced here the factor g^3/N_c to make G_3 of $\mathcal{O}(1)$ in the large- N_c limit and inserted i^3 for later convenience. Therefore, the contribution of the diagram in Fig. 35 having one boundary should be divided by N_c , while the factor g^3 is naturally associated with three extra vertices which appear after the contraction of color indices.

The extension of Eq. (7.24) to multi-point Green's functions is obvious:

$$G_{\mu_1 \dots \mu_n}^{(n)}(x_1, \dots, x_n) \equiv \frac{(ig)^n}{N_c} \langle \text{tr} [A_{\mu_1}(x_1) \dots A_{\mu_n}(x_n)] \rangle. \quad (7.25)$$

The factor $1/N_c$, which normalizes the trace, provides the natural normalization $G^{(0)} = 1$ of the averages.

Though the two terms in the index-space representation (7.8) of the three-gluon vertex looks very similar, their fate in the topological expansion is quite different. When the color

Fig. 36: Planar a) and non-planar b) contributions of the two color structures in Eq. (7.8) for three-gluon vertex to $G^{(3)}$ in the lowest order of perturbation theory.

indices are contracted anti-clockwise, the first term leads to the planar contributions to $G^{(3)}$, the simplest of which is depicted in Fig. 36a. The anti-clockwise contraction of the color indices in the second term leads to a non-planar graph in Fig. 36b which can be drawn without crossing of lines only on a torus. Therefore, the two color structures of the three-gluon vertex contribute to different orders of the topological expansion. The same is true for the four-gluon vertex.

Remark on oriented Riemann surfaces

Each line of an index graph of the type depicted in Fig. 35 is oriented. This orientation continues along a closed index line while the pairs of index lines of each double line have opposite orientations. The overall orientation of the lines is prescribed by the orientation of the external boundary which we choose to be, say, anti-clockwise like in Fig. 36a. Since the lines are oriented, the faces of the Riemann surface associated with a given graph are oriented too — all in the same way — anti-clockwise. Vice versa, such an orientation of the Riemann surfaces unambiguously fixes the orientation of all the index lines. This is the reason why we omit the arrows associated with the orientation of the index lines: their directions are obvious.

Remark on cyclic-ordered Green's functions

The cyclic-ordered Green's functions (7.25) naturally arise in the expansion of the trace of the non-Abelian phase factor for a closed contour (see Eq. (2.48)). One gets

$$\left\langle \frac{1}{N_c} \text{tr} \mathbf{P} e^{ig \oint_{\Gamma} dx^\mu A_\mu(x)} \right\rangle = \sum_{n=0}^{\infty} \oint_{\Gamma} dx_1^{\mu_1} \int_{x_1}^{x_1} dx_2^{\mu_2} \dots \int_{x_1}^{x_{n-1}} dx_n^{\mu_n} G_{\mu_1 \dots \mu_n}^{(n)}(x_1, \dots, x_n). \quad (7.26)$$

The reason is because the ordering along a closed path implies the cyclic-ordering in the index space.

Fig. 37: Example of a) connected and b) disconnected planar graphs.

Remark on generating functionals for planar graphs

By connected or disconnected planar graphs we mean, respectively, the graphs which were connected or disconnected before the contraction of the color indices as is illustrated by Fig. 37. The graph in Fig. 37a is connected planar while the graph in Fig. 37b is disconnected planar.

The usual relation (6.54) between the generating functionals $W[J]$ and $Z[J]$ for connected graphs and all graphs does not hold for the planar graphs. The reason is that an exponentiation of such a connected planar diagram for the cyclic-ordered Green's functions (7.25) can give disconnected non-planar ones.

The generating functionals for all and connected planar graphs can be constructed [61] by means of introducing non-commutative sources $\mathbf{j}_\mu(x)$. "Non-commutative" means that there is no way to transform $\mathbf{j}_{\mu_1}(x_1)\mathbf{j}_{\mu_2}(x_2)$ into $\mathbf{j}_{\mu_2}(x_2)\mathbf{j}_{\mu_1}(x_1)$. This non-commutativity of the sources reflects the cyclic-ordered structure of the Green's functions (7.25) which possess only cyclic symmetry.

Using the short-hand notations (6.44) where the symbol $\circ$ includes the sum over the d -vector (or whatever available) indices except for the color ones:

$$\mathbf{j} \circ \mathcal{A} \equiv \sum_\mu \int d^d x \mathbf{j}_\mu(x) \mathcal{A}_\mu(x), \quad (7.27)$$

we write down the definitions of the generating functionals for all planar and connected planar graphs, respectively, as

$$Z[\mathbf{j}] \equiv \sum_{n=0}^{\infty} \left\langle \frac{1}{N_c} \text{tr} (\mathbf{j} \circ \mathcal{A})^n \right\rangle \quad (7.28)$$

and

$$W[\mathbf{j}] \equiv \sum_{n=0}^{\infty} \left\langle \frac{1}{N_c} \text{tr} (\mathbf{j} \circ \mathcal{A})^n \right\rangle_{\text{conn}}. \quad (7.29)$$

The planar contribution to the Green's functions (7.25) and their connected counterparts can be obtained, respectively, from the generating functionals $Z[\mathbf{j}]$ and $W[\mathbf{j}]$ applying the non-commutative derivative which is defined by

$$\frac{\delta}{\delta \mathbf{j}_\mu(x)} \mathbf{j}_\nu(y) f(\mathbf{j}) = \delta_{\mu\nu} \delta^{(d)}(x-y) f(\mathbf{j}), \quad (7.30)$$

Fig. 38: Graphic derivation of Eq. (7.31): $Z[\mathbf{j}]$ is denoted by an empty box, $W[\mathbf{j}]$ is denoted by a shadow box, $\mathbf{j}$ is denoted by a filled circle. By picking a leftmost external line of a planar graph, we end up with a connected planar graph, whose remaining external lines are somewhere to the right interspersed by disconnected planar graphs. It is evident that $\mathbf{j}Z[\mathbf{j}]$ plays the role of a new source for the connected planar graph. If we instead pick up the rightmost external line, we get the inverse order $Z[\mathbf{j}]\mathbf{j}$, which results in Eq. (7.32).

where f is an arbitrary function of $\mathbf{j}$'s. In other words the derivative picks up only the leftmost variable.

The relation which replaces Eq. (6.54) for planar graphs is

$$Z[\mathbf{j}] = W[\mathbf{j}Z[\mathbf{j}]], \quad (7.31)$$

while the cyclic symmetry says

$$W[\mathbf{j}Z[\mathbf{j}]] = W[Z[\mathbf{j}]\mathbf{j}]. \quad (7.32)$$

A graphic derivation of Eqs. (7.31) and (7.32) is given in Fig. 38. In other words, given $W[\mathbf{j}]$, one should construct an inverse function as the solution to the equation

$$\mathbf{j}_\mu(x) = \mathbf{J}_\mu(x) W[\mathbf{j}], \quad (7.33)$$

after which Eq. (7.31) says

$$Z[\mathbf{j}] = W[\mathbf{J}]. \quad (7.34)$$

Equation (7.33) can be iteratively solved in the Gaussian case where only $G^{(2)}$ is nonvanishing which yields

$$W[\mathbf{j}] = 1 - g^2 \mathbf{j} \circ D \circ \mathbf{j}. \quad (7.35)$$

Here the propagator D is given by Eq. (7.4). Using Eq. (7.31), we get explicitly

$$Z[\mathbf{j}] = 1 - g^2 \int d^d x d^d y D_{\mu\nu}(x-y) \mathbf{j}_\mu(x) Z[\mathbf{j}] \mathbf{j}_\nu(y) Z[\mathbf{j}]. \quad (7.36)$$

While this equation for $Z[\mathbf{j}]$ is quadratic, its solution can be written only as a continued fraction due to the non-commutative nature of the variables. In order to find it, we rewrite Eq. (7.36) as

$$Z[\mathbf{j}] = \frac{1}{1 + g^2 \int d^d x d^d y D_{\mu\nu}(x-y) \mathbf{j}_\mu(x) Z[\mathbf{j}] \mathbf{j}_\nu(y)}, \quad (7.37)$$

Fig. 39: Diagrams for the gluon propagator with a quark loop which is represented by the single lines. The diagram a) involves one quark loop and has no closed index lines so that its order is $\sim g^2 \sim 1/N_c$. The diagram b) involves three loops one of which is a quark loop. Its order is $\sim g^6 N_c^2 \sim 1/N_c$.

whose iterative solution reads [61]

$$Z[j] = \frac{1}{1 + g^2 j \frac{1}{1 + g^2 j \frac{1}{1 + g^2 j \frac{1}{\vdots}}}} \circ D \circ j. \quad (7.38)$$

More about this approach to the generating functionals for planar graphs can be found in Ref. [62].

7.4 Topological expansion and quark loops

It is easy to incorporate quarks in the topological expansion. A quark field belongs to the fundamental representation of the gauge group $SU(N_c)$ and its propagator is represented by a single line

$$\langle \psi_i \bar{\psi}_j \rangle \propto \delta_{ij} = i \longrightarrow j. \quad (7.39)$$

The arrow indicates, as usual, the direction of propagation of a (complex) field ψ . We shall omit these arrows for simplicity.

The diagram for the gluon propagator which involves one quark loop is depicted in Fig. 39a. It has two three gluon vertices and no closed index lines so that its order in $1/N_c$ is

$$\text{Fig. 39a} \sim g^2 \sim \frac{1}{N_c}. \quad (7.40)$$

Analogously, the order of a more complicated tree-loop diagram in Fig. 39b, which involves one quark loop and two closed index lines, is

$$\text{Fig. 39b} \sim g^6 N_c^2 \sim \frac{1}{N_c}. \quad (7.41)$$

It is evident from this consideration that quark loops are not accompanied by closed index lines. One should add a closed index line for each quark loop in order for a given

Fig. 40: Generic diagram in the index space which has $L = 1$ quark loop and $B = 1$ loop associated with the external boundary. Its order in $1/N_c$ is described by Eq. (7.50).

diagram with L quark loops to have the same double-line representation as for pure gluon diagrams. Therefore, given Eq. (7.22), diagrams with L quark loops are suppressed at large N_c by

$$L \text{ quark loops} \sim \left(\frac{1}{N_c} \right)^{L+2 \text{ genus}}. \quad (7.42)$$

The single-line representation of the quark loops is similar to the one of the external boundary in Fig. 35. Moreover, such a diagram emerges when one calculates perturbative gluon corrections to the vacuum expectation value of the quark operator

$$O = \frac{1}{N_c} \bar{\psi} \psi, \quad (7.43)$$

where the factor of $1/N_c$ is introduced to make it $\mathcal{O}(1)$ in the large- N_c limit. Therefore, the external boundary can be viewed as a single line associated with valence quarks. The difference between virtual quark loops and external boundaries is that each of the latter comes along with the factor of $1/N_c$ due to the definitions (7.25) and (7.43).

In order to prove Eqs. (7.22) and its quark counterpart (7.42), let us consider a generic diagram in the index space which has $n_0^{(3)}$ three-point vertices (either three-gluon or quark-gluon ones), $n_0^{(4)}$ four-gluon vertices, n_1 propagators (either gluon or quark ones), n_2 closed index lines, L virtual quark loops and B external boundaries. A typical such diagram is depicted in Fig. 40. Its order in $1/N_c$ is

$$\frac{1}{N_c^B} g^{n_0^{(3)} + 2n_0^{(4)}} N_c^{n_2} \sim N_c^{n_2 - n_0^{(3)}/2 - n_0^{(4)} - B} \quad (7.44)$$

as is already explained. The extra factor of $1/N_c^B$ is due to the extra normalization factor of $1/N_c$ in operators associated with external boundaries.

The number of propagators and vertices are related by

$$2n_1 = 3n_0^{(3)} + 4n_0^{(4)}, \quad (7.45)$$

since three- and four-point vertices emit three or four propagators, respectively, and each propagator connects two vertices. Using the relation (7.45), we rewrite the RHS of (7.44) as

$$N_c^{n_2 - n_0^{(3)}/2 - n_0^{(4)} - B} = N_c^{n_2 - n_1 + n_0 - B}, \quad (7.46)$$

where the total number of vertices

$$n_0 = n_0^{(3)} + n_0^{(4)} \quad (7.47)$$

is introduced.

The exponent on the RHS of Eq. (7.46) can be expressed via the Euler characteristics χ of a given graph of genus h . Let us first mention that a proper Riemann surface, which is associated with a given graph, is open and has $B + L$ boundaries (represented by single lines). This surface can be closed by attaching a cap to each boundary. The single lines then become double lines together with the lines of the boundary of each cap. We have already considered this procedure when deducing Eq. (7.42) from Eq. (7.22).

The number of faces for a closed Riemann surface constructed in such a manner is $n_2 + L + B$, while the number of edges and vertices are n_1 and n_0 , respectively. Euler's theorem says that

$$\chi \equiv 2 - 2h = n_2 + L + B - n_1 + n_0. \quad (7.48)$$

Therefore the RHS of Eq. (7.46) can be rewritten as

$$N_c^{n_2 - n_1 + n_0 - B} = N_c^{2 - 2h - L - 2B}. \quad (7.49)$$

We have thus proven that the order in $1/N_c$ of a generic graph does not depend on its order in the coupling constant and is completely expressed via the genus h and the number of virtual quark loops L and external boundaries B by

$$\text{generic graph} \sim \left(\frac{1}{N_c}\right)^{2h + L + 2(B-1)}. \quad (7.50)$$

For $B = 1$, we recover Eqs. (7.22) and (7.42).

Remark on the order of gauge action

We see from Eq. (7.25) that the natural variables for the large- N_c limit are the calligraphic matrices $\mathcal{A}_\mu$ which include the extra factor of ig with respect to A_μ (see Eq. (7.2)). For these matrices

$$\frac{1}{N_c} \langle \text{tr} [\mathcal{A}_{\mu_1}(x_1) \cdots \mathcal{A}_{\mu_n}(x_n)] \rangle = G_{\mu_1 \cdots \mu_n}^{(n)}(x_1, \dots, x_n) \quad (7.51)$$

so that they are $\mathcal{O}(1)$ in the large- N_c limit since the trace is $\mathcal{O}(N_c)$.

In these variables, the gluon part of the QCD action (2.33) takes on the simple form

$$S = -\frac{1}{2g^2} \int d^d x \text{tr} \mathcal{F}_{\mu\nu}^2(x). \quad (7.52)$$

Since g^2 in this formula is $\sim 1/N_c$ and the trace is $\sim N_c$, the action is $\mathcal{O}(N_c^2)$ at large N_c .

This result can be anticipated from the free theory because the kinetic part of the action involves the sum over $N_c^2 - 1$ free gluons. Therefore, the non-Abelian field strength is ~ 1 for $g^2 \sim 1/N_c$.

The fact that the action is $\mathcal{O}(N_c^2)$ in the large- N_c limit is a generic property of the models describing matrix fields. It will be crucial for developing saddle-point approaches at large N_c which are considered below.

The formula (7.50) can be easily rederived using the calligraphic notation. The propagator of the $\mathcal{A}$ -field is $\sim g^2$ while both three- and four-gluon vertices are now $\sim g^{-2}$ as a consequence of Eq. (7.52). The contribution of a generic graph is now of the order

$$(g^2)^{n_1 - n_0} N_c^{n_2 - B} \sim N_c^{n_2 - n_1 + n_0 - B} \quad (7.53)$$

for $g^2 \sim 1/N_c$. This coincides with the RHS of Eq. (7.46) which results in Eq. (7.50).

7.5 't Hooft versus Veneziano limits

In QCD there are several species or flavors of quarks (u -, d -, s - and so on). We denote the number of flavors by N_f and associate a Greek letter α or β with a flavor index of the quark field.

The quark propagator then has the Kronecker delta with respect to the flavor indices in addition to Eq. (7.39):

$$\langle \psi_i^\alpha \bar{\psi}_j^\beta \rangle \propto \delta^{\alpha\beta} \delta_{ij}. \quad (7.54)$$

Their contraction results in

$$\sum_{\alpha=1}^{N_f} \delta_{\alpha\alpha} = N_f. \quad (7.55)$$

Therefore, an extra factor of N_f corresponds to each closed quark loop for the N_f flavors.

The limit when N_f is fixed as $N_c \rightarrow \infty$, as was considered in the original paper by 't Hooft [2], is called the 't Hooft limit. Only valence quarks are left in the 't Hooft limit. Hence, it is associated with the quenched approximation which was discussed in the Remark to Subsect. 5.5. In order for a meson to decay into other mesons built out of quarks, say for a ρ -meson to decay into a pair of π -mesons, a quark-antiquark pair must be produced out of the vacuum. Consequently, the ratios of meson widths to their masses are

$$\frac{\Gamma_{\text{total}}}{M} \sim \frac{N_f}{N_c} \quad (7.56)$$

in the 't Hooft limit. The ratio on the LHS of Eq. (7.56) is experimentally 10–15% for the ρ -meson. A hope to solve QCD in the 't Hooft limit is a hope to describe QCD with this accuracy.

An alternative large- N_c limit of QCD when $N_f \sim N_c$ as $N_c \rightarrow \infty$ was proposed by Veneziano [58]. Some diagrams for the gluon propagator, which involve one quark loop, are

Fig. 41: Diagrams with quark loops in the Veneziano limit. The color and flavor indices of a quark loop which are represented by the solid and dashed single lines, respectively. The diagram a) is $\sim g^2 N_f \sim N_f/N_c$. The diagram b) is $\sim g^6 N_c^2 N_f \sim N_f/N_c$.

depicted in Fig. 41. The dashed single line represents propagation of the flavor index. Each closed loop of the dashed line is associated with the factor N_f according to Eq. (7.55). This is analogous to the vector models which exactly describe the $O(N_f)$ flavor symmetry in these notations.

The diagrams in Fig. 41 contribute, respectively,

$$\text{Fig. 41a} \sim g^2 N_f \sim \frac{N_f}{N_c} \quad (7.57)$$

and

$$\text{Fig. 41b} \sim g^6 N_c^2 N_f \sim \frac{N_f}{N_c} \quad (7.58)$$

in the limit (7.13). Likewise, a more general diagram with L quark loops will contribute

$$L \text{ quark loops} \sim \left(\frac{N_f}{N_c}\right)^L \left(\frac{1}{N_c^2}\right)^{\text{genus}}. \quad (7.59)$$

This formula obviously follows from Eq. (7.42) since each quark loop results in N_f .

We see from Eq. (7.59) that quark loops are not suppressed at large N_c in the Veneziano limit

$$N_f \sim N_c \rightarrow \infty \quad (7.60)$$

if the diagram is planar. Furthermore, the representation of a flavored quark by one solid and one dashed line is obviously similar to the double-line representation of a gluon. All what is said above about the topological expansion of pure gluodynamics holds for QCD with quarks in the Veneziano limit.

It is the Veneziano limit (7.60) that is related to the hadronic topological expansion in the dual-resonance models. Hadrons can have, in the Veneziano limit, finite widths according to Eq. (7.56). I refer the reader to the original paper by Veneziano [58] for more detail.

There is an alternative way to show why virtual quarks are suppressed in the 't Hooft limit and survive in the Veneziano limit. Let us integrate over the quark fields which yields

$$\int D\bar{\psi} D\psi e^{-\int d^4x (\bar{\psi} \hat{C} \psi + m \bar{\psi} \psi)} = e^{\text{Sp ln}(\hat{C} + m)}. \quad (7.61)$$

Fig. 42: Diagram with two quark loops in the Veneziano limit. The diagram is $\sim g^6 N_c N_f^2 \sim (N_f/N_c)^2$.

The trace in the exponent involves summation both over color and flavor indices, so that

$$\text{Sp ln}(\hat{C} + m) \sim N_c N_f. \quad (7.62)$$

The order in N_c of the pure gluon action is $\mathcal{O}(N_c^2)$ as is already discussed. Hence, the quark contribution to the action is $\sim N_f/N_c$ in comparison to the gluon one. The quark determinant can be disregarded in the 't Hooft limit, but is essential in the Veneziano limit.

The consideration of the previous paragraph also explains why each quark loop contributes a factor $\sim N_f/N_c$. The exponent on the RHS of Eq. (7.61) is associated with one-loop diagrams. A diagram with L quark loops corresponds to the L -th term of the expansion of the exponential. This explains the factor $(N_f/N_c)^L$ in Eq. (7.59). A diagram with two quark loops, which appears to the second order of this expansion, is depicted in Fig. 42.

Remark on asymptotic freedom in the Veneziano limit

Though the number of flavors becomes large in the Veneziano limit, this does not mean that asymptotic freedom is lost. The leading-order coefficient of the $\mathcal{B}$ -function of QCD with N_c colors and N_f flavors reads

$$b = -\frac{11}{3}N_c + \frac{2}{3}N_f. \quad (7.63)$$

It is still negative if $N_f/N_c < 11/2$ in the Veneziano limit.

Remark on phenomenology of multicolor QCD

While $N_c = 3$ in the real world, there are phenomenological indications that $1/N_c$ may be considered as a small parameter. We have already mentioned some of them in the text – the simplest one is that the ratio of the ρ -meson width to its mass, which is $\sim 1/N_c$, is small. Considering $1/N_c$ as a small parameter immediately leads to qualitative phenomenological consequences which are preserved by the planar diagrams associated with multicolor QCD, but are violated by the non-planar diagrams.

The most important consequence is the relation of the $1/N_c$ -expansion to the topological expansion in the dual-resonance model of hadrons. Vast properties of hadrons are explained by the dual-resonance model. A very clear physical picture behind this model is that hadrons are excitations of a string with quarks at the ends.

I shall briefly list some consequences of multicolor QCD:

- 1) The “naive” quark model of hadrons emerges at $N_c = \infty$. Hadrons are built out of (valence or constituent) quark and antiquark $q\bar{q}$, while exotic states like $qq\bar{q}\bar{q}$ do not appear.
- 2) The partial width of decay of the ϕ -meson, which is built out of $s\bar{s}$ (the strange quark and antiquark), into K^+K^- is $\sim 1/N_c$, while that into $\pi^+\pi^-\pi^0$ is $\sim 1/N_c^2$. This explains Zweig’s rule. The masses of the ρ - and ω -mesons are degenerate at $N_c = \infty$.
- 3) The coupling constant of meson-meson interaction is small at large N_c .
- 4) The widths of glueballs are $\sim 1/N_c^2$, *i.e.* they should be even narrower than mesons built out of quarks. The glueballs do not interact or mix with mesons at $N_c = \infty$.

All these hadron properties (except the last one) approximately agree with experiment, and were well-known even before 1974 when multicolor QCD was introduced. Glueballs are not yet detected experimentally (maybe because of their property listed in the item 4).

7.6 Large- N_c factorization

The vacuum expectation values of several colorless or white operators, which are singlets with respect to the gauge group, factorize in the large- N_c limit of QCD (or other matrix models). This property is similar to that already discussed in Subsect. 6.5 for the vector models.

The simplest gauge-invariant operator in a pure $SU(N_c)$ gauge theory is the square of the non-Abelian field strength:

$$O(x) = \frac{1}{N_c^2} \text{tr} F_{\mu\nu}^2(x). \quad (7.64)$$

The normalizing factor provides the natural normalization

$$\left\langle \frac{1}{N_c^2} \text{tr} F_{\mu\nu}^2(x) \right\rangle = \left\langle \frac{1}{2N_c^2} F_{\mu\nu}^a(x) F_{\mu\nu}^a(x) \right\rangle \sim 1. \quad (7.65)$$

In order to verify the factorization in the large- N_c limit, let us consider the index space diagrams for the average of two colorless operators $O(x_1)$ and $O(x_2)$, which are depicted in Fig. 43.

The graph in Fig. 43a represents the zeroth order of perturbation theory. It involves four closed index lines (the factor N_c^4) and the normalization factor $1/N_c^4$ according to the definition (7.64). Its contribution is

$$\text{Fig. 43a} \sim \frac{1}{N_c^2} N_c^2 \times \frac{1}{N_c^2} N_c^2 \sim 1, \quad (7.66)$$

i.e. $\mathcal{O}(1)$ in accord with the general estimate (7.65).

The graph in Fig. 43b involves a gluon line which is emitted and absorbed by the same operator $O(x_1)$. It has five closed index lines (the factor N_c^5), the normalization factor $1/N_c^4$, and g^2 due to two three-gluon vertices. Its contribution is

$$\text{Fig. 43b} \sim g^2 N_c \sim 1, \quad (7.67)$$

Fig. 43: Demonstration of the large- N_c factorization to the lowest orders of perturbation theory. The closed double line represents the average of the operator (7.64) to zeroth order in g . The diagrams a) and b), which are associated with the factorized part of the average on the LHS of Eq. (7.70), are $\mathcal{O}(1)$. The diagrams c) and d), which would violate the factorization, are suppressed by $1/N_c^2$.

i.e. also $\mathcal{O}(1)$ in the limit (7.13).

The graph in Fig. 43c is of the same type as the graph in Fig. 43a, but the double lines now connect two different operators. It has two closed index lines (the factor N_c^2) and the normalization factor $1/N_c^4$, so that its contribution

$$\text{Fig. 43c} \sim \frac{1}{N_c^2} \quad (7.68)$$

is suppressed by $1/N_c^2$.

The graph in Fig. 43d, which is of the same order in the coupling constant as the graph in Fig. 43b, involves only three closed index lines (the factor N_c^3) and is of order $1/N_c^2$:

$$\text{Fig. 43d} \sim g^2 \frac{1}{N_c} \sim \frac{1}{N_c^2}. \quad (7.69)$$

Therefore, it is suppressed by $1/N_c^2$ in the large- N_c limit. For this graph, the gluon line is emitted and absorbed by different operators $O(x_1)$ and $O(x_2)$.

This lowest-order example illustrates the general property that only (planar) diagrams with gluon lines emitted and absorbed by the same operators survive as $N_c \rightarrow \infty$. Hence, correlations between the colorless operators $O(x_1)$ and $O(x_2)$ are of order $1/N_c^2$, so that the factorization property holds as $N_c \rightarrow \infty$:

$$\left\langle \frac{1}{N_c^2} \text{tr} F^2(x_1) \frac{1}{N_c^2} \text{tr} F^2(x_2) \right\rangle = \left\langle \frac{1}{N_c^2} \text{tr} F^2(x_1) \right\rangle \left\langle \frac{1}{N_c^2} \text{tr} F^2(x_2) \right\rangle + \mathcal{O}\left(\frac{1}{N_c^2}\right). \quad (7.70)$$

For a general set of gauge-invariant operators $O_1, \dots, O_n$, the factorization property can be represented by

$$\langle O_1 \cdots O_n \rangle = \langle O_1 \rangle \cdots \langle O_n \rangle + \mathcal{O}\left(\frac{1}{N_c^2}\right). \quad (7.71)$$

This is analogous to Eq. (6.78) for the vector models.

The factorization in large- N_c QCD was first discovered by A.A. Migdal in the late seventies. An important observation that the factorization implies a semiclassical nature of the large- N_c limit of QCD was done by Witten [63]. We shall discuss this in the next two Subsections.

The factorization property also holds for gauge-invariant operators constructed from quarks like in Eq. (7.43). For the case of several flavors N_f , we normalize these quark operators by

$$O_\Gamma = \frac{1}{N_f N_c} \bar{\psi} \Gamma \psi. \quad (7.72)$$

Here Γ stands for one of the combination of the gamma-matrices:

$$\Gamma = 1, \gamma_5, \gamma_\mu, i\gamma_\mu \gamma_5, \Sigma_{\mu\nu} = \frac{1}{2i} [\gamma_\mu, \gamma_\nu], \dots \quad (7.73)$$

Estimation of the order in $1/N_c$ of diagrams for the average of two quark operators (7.72) is analogous to that for the pure gluon graphs in Fig. 43.

Thus, the factorization of the gauge-invariant quark operators holds both in the 't Hooft and Veneziano limits:

$$\langle O_{\Gamma_1} \dots O_{\Gamma_n} \rangle = \langle O_{\Gamma_1} \rangle \dots \langle O_{\Gamma_n} \rangle + \mathcal{O}\left(\frac{1}{N_f N_c}\right). \quad (7.74)$$

The non-factorized part, which is associated with connected diagrams, is $\sim 1/N_c$ in the 't Hooft limit. This leads, in particular, to the coupling constant of meson-meson interaction of order $1/N_c$, clarifying the property of multicolor QCD listed above. The Veneziano limit is analogous to pure gluodynamics as is already mentioned.

It is worth noting that the factorization can be seen (at all orders of perturbation theory) from Eq. (7.50) for the contribution of a generic connected graph of genus h with B external boundaries which are precisely associated with the quark operators O_Γ , as is explained in Subsect. 7.4. The diagrams with gluon lines emitted and absorbed by the same operator like in Fig. 43b are products of diagrams having only one boundary. Hence, their contribution is of order 1. Otherwise, the diagrams with gluon lines emitted and absorbed by two different operators like in Fig. 43d have two boundaries. According to Eq. (7.50), their contribution is suppressed by $1/N_c^2$. Alternatively, the diagrams like in Fig. 43c (including its planar dressing by gluons) have one boundary.²² Their contribution is $\mathcal{O}(1)$ times $1/(N_f N_c)$ coming from the normalization of the operator (7.72). This proves the factorization property (7.74) at all orders of perturbation theory.

Remark on factorization beyond perturbation theory

The large- N_c factorization can be also verified at all orders of the strong coupling expansion in the $SU(N_c)$ lattice gauge theory. A non-perturbative proof of the factorization will be given in the next Section by using quantum equations of motion (the loop equations).

²²In the dual-resonance model, they are associated with the meson-meson interaction due to an exchange of constituent quarks.

7.7 The master field

The large- N_c factorization in QCD assumes that gauge-invariant objects behave as c -numbers, rather than as operators. Likewise the vector models, this suggests that the path integral is dominated by a saddle point.

We already saw in Subsect. 6.5 that the factorization in the vector models does not mean that the fundamental field itself, for instance $\vec{n}$ in the sigma-model, becomes "classical". It is the case, instead, for a singlet composite field.

We are now going to apply a similar idea to the Yang-Mills theory whose partition function reads

$$Z = \int D A_\mu^a e^{-\int d^4x \frac{1}{4} F_{\mu\nu}^a F_{\mu\nu}^a}. \quad (7.75)$$

The action, $\sim N_c^2$, is large as $N_c \rightarrow \infty$, but the "entropy" is also $\sim N_c^2$ due to the $N_c^2 - 1$ integrations over A_μ^a :

$$D A_\mu^a \sim e^{N_c^2}. \quad (7.76)$$

Consequently, the saddle-point equation of the large- N_c Yang-Mills theory is *not* the classical one which reads

$$\frac{\delta S}{\delta A_\mu^a} = (\nabla_\nu F_{\mu\nu})^a = 0. \quad (7.77)$$

The idea is to rewrite the path integral over A_μ for the Yang-Mills theory as that over a colorless composite field $\Phi[A]$, likewise it was done in Subsect. 6.4 for the sigma-model. The expected new path-integral representation of the partition function (7.75) would be something like

$$Z \propto \int D\Phi \frac{1}{\partial A_\mu^a} e^{-N_c^2 S[\Phi]}. \quad (7.78)$$

The Jacobian

$$\frac{\partial \Phi[A]}{\partial A_\mu^a} \equiv e^{-N_c^2 J[\Phi]} \quad (7.79)$$

in Eq. (7.78) is related to the old entropy factor, so that $J[\Phi] \sim 1$ in the large- N_c limit.

The original partition function (7.75) can be then rewritten as

$$Z \propto \int D\Phi e^{N_c^2 J[\Phi] - N_c^2 S[\Phi]}, \quad (7.80)$$

where $S[\Phi]$ represents the Yang-Mills action in the new variables. The new "entropy" factor $D\Phi$ is $\mathcal{O}(1)$ because the variable $\Phi[A]$ is a color singlet. The large parameter N_c enters Eq. (7.80) only in the exponent. Therefore, the saddle-point equation can be immediately written:

$$\frac{\delta S}{\delta \Phi} = \frac{\delta J}{\delta \Phi}. \quad (7.81)$$

Remembering that Φ is a functional of A_μ : $\Phi \equiv \Phi[A]$, we rewrite the saddle-point equation (7.81) as

$$\frac{\delta S}{\delta A_\nu^a} = (\nabla_\mu F_{\mu\nu})^a = \frac{\delta J}{\delta A_\nu^a}. \quad (7.82)$$

It differs from the classical Yang–Mills equation (7.77) by the term on the RHS coming from the Jacobian (7.79).

Given $J[\Phi]$ which depends on the precise form of the variable $\Phi[A]$, Eq. (7.82) has a solution

$$A_\mu(x) = A_\mu^{\text{cl}}(x). \quad (7.83)$$

Let us first assume that there exists only one solution to Eq. (7.82). Then the path integral is saturated by a single configuration (7.83), so that the vacuum expectation values of gauge-invariant operators are given by their values at this configuration:

$$\langle O \rangle = O(A_\mu^{\text{cl}}(x)). \quad (7.84)$$

The factorization property (7.71) will obviously be satisfied.

An existence of such a classical field configuration in multicolor QCD was conjectured by Witten [63]. It was discussed in the lectures by Coleman [64] who called it the *master field*. Equation (7.82) which determines the master field is often referred to as the master-field equation.

A subtle point with the master field is that a solution to Eq. (7.82) is determined only up to a gauge transformation. To preserve gauge invariance, it is more reasonable to speak about the whole gauge orbit as a solution of Eq. (7.82). However, this will not change Eq. (7.84) since the operator O is gauge invariant.

The conjecture about an existence of the master field has surprisingly rich consequences. Since vacuum expectation values are Poincaré invariant, the RHS of Eq. (7.84) does. This implies that $A_\mu^{\text{cl}}(x)$ must itself be Poincaré invariant up to a gauge transformation: a change of $A_\mu^{\text{cl}}(x)$ under translations or rotations can be compensated by a gauge transformation. Moreover, there must exist a gauge in which $A_\mu^{\text{cl}}(x)$ is space-time independent:

$$A_\mu^{\text{cl}}(x) = A_\mu^{\text{cl}}(0). \quad (7.85)$$

In this gauge, rotations must be equivalent to a global gauge transformation, so that $A_\mu^{\text{cl}}(0)$ transforms as a Lorentz vector.

In fact, the idea about such a master field in multicolor QCD may be just wrong as was pointed out by Haan [65]. The conjecture about an existence of only one solution to the master-field equation (7.82) seems to be somewhat incorrect. If several solutions exist, one needs an additional averaging over these solutions. This is a very delicate matter, since this additional averaging must still preserve the factorization property. One might better think about this situation as if $A_\mu^{\text{cl}}(0)$ would be an operator in some Hilbert space rather than a c-valued function. This is simply because $A_\mu^{\text{cl}}(0)$ is, in the matrix notation (7.1), a $N_c \times N_c$

matrix which becomes, as $N_c \rightarrow \infty$, an infinite matrix, or an operator in Hilbert space. Such an operator-valued master field is sometimes called the master field in the *weak* sense, while the above conjecture about a single classical configuration of the gauge field, which saturates the path integral, is called the master field in the *strong* sense.

The concept of the master field is rather vague until a precise form of the composite field $\Phi[A]$, and consequently the Jacobian $\Phi[A]$ that enters Eq. (7.82), is not defined. However, what is important is that the master field (in the weak sense) is space-time independent. This looks like a simplification of the problem of solving large- N_c QCD. A Hilbert space, in which the operator $A_\mu^{\text{cl}}(0)$ acts, should be specified by $\Phi[A]$. We shall consider in the next Subsection a realization of these ideas for the case of $\Phi[A]$ given by the trace of the non-Abelian phase factor for closed contours.

Remark on non-commutative probability theory

An adequate mathematical language for describing the master field in multicolor QCD (and, generically, in matrix models at large N_c) was found by I. Singer in 1994. It is based on the concept of free random variables of non-commutative probability theory, introduced by Voiculescu [66]. How to describe the master field in this language and some other applications of non-commutative free random variables to the problems of planar quantum field theory are discussed in Refs. [67, 68].

7.8 $1/N_c$ as semiclassical expansion

A natural candidate for the composite operator $\Phi[A]$ from the previous Subsection is given by the trace of the non-Abelian phase factor for closed contours — the Wilson loop. It is labeled by the loop C in the same sense as the field $A_\mu(x)$ is labeled by the point x , so we shall use the notation

$$\Phi(C) \equiv \Phi[A] = \frac{1}{N_c} \text{tr} \mathbf{P} e^{ig \oint_C dx^\mu A_\mu(x)}. \quad (7.86)$$

Nobody up to now managed to reformulate QCD at finite N_c in terms of $\Phi(C)$ in the language of path integral. This is due to the fact that self-intersecting loops are not independent (they are related by the so-called Mandelstam relations [69]), and the Jacobian is huge. The reformulation was done [70] in the language of Schwinger–Dyson or loop equations which will be described in the next Section.

Schwinger–Dyson equations are a convenient way of performing the semiclassical expansion, which is an alternative to the path integral. Let us illustrate an idea how to do this by an example of the φ^3 theory whose Schwinger–Dyson equations read

$$\begin{aligned} & (-\partial^2 + m^2) \langle \varphi(x) \varphi(x_2) \dots \varphi(x_n) \rangle + \frac{\lambda}{2} \langle \varphi^2(x) \varphi(x_2) \dots \varphi(x_n) \rangle \\ &= \hbar \sum_{j=2}^n \delta^{(d)}(x - x_j) \langle \varphi(x_2) \dots \varphi(x_j) \dots \varphi(x_n) \rangle, \end{aligned} \quad (7.87)$$

where $\varphi(x_j)$ means that the corresponding term $\varphi(x_j)$ is missing in the product.

The RHS of Eq. (7.87) is proportional to the Planck's constant $\hbar$. In the semiclassical limit $\hbar \rightarrow 0$, we get

$$\left(-\partial_1^2 + m^2\right) \langle \varphi(x_1) \dots \varphi(x_n) \rangle + \frac{\lambda}{2} \langle \varphi^2(x_1) \dots \varphi(x_n) \rangle = 0, \quad (7.88)$$

whose solution is of the factorized form

$$\langle \varphi(x_1) \dots \varphi(x_n) \rangle = \langle \varphi(x_1) \rangle \dots \langle \varphi(x_n) \rangle + \mathcal{O}(\hbar) \quad (7.89)$$

provided that

$$\langle \varphi(x) \rangle \equiv \varphi_{\text{cl}}(x) \quad (7.90)$$

obeys

$$\left(-\partial^2 + m^2\right) \varphi_{\text{cl}}(x) + \frac{\lambda}{2} \varphi_{\text{cl}}^2(x) = 0. \quad (7.91)$$

Equation (7.91) is nothing but the classical equation of motion for the φ^3 theory, which specifies extrema of the action. Thus, we have reproduced, using the Schwinger–Dyson equations, the well-known fact that the path integral is dominated by a classical trajectory as $\hbar \rightarrow 0$. It is also clear how to perform the semiclassical expansion in $\hbar$ in the language of the Schwinger–Dyson equations: one should solve Eq. (7.87) by iterations.

The reformulation of multicolor QCD in terms of the loop functionals $\Phi(C)$ is, in a sense, a realization of the idea of the master field in the weak sense, when the master field acts as an operator in the space of loops. The loop equation of the next Section will be a sort of the master-field equation in the loop space.

Remark on the large- N_c limit as statistical averaging

There is yet another, pure statistical, explanation why the large- N_c limit is a “semiclassical” limit for the collective variables $\Phi(C)$. The matrix $U^{ij}[C_{xx}]$, that describes the parallel transport along a closed contour C_{xx} , can be reduced by the gauge transformation to

$$U[C_{xx}] = \Omega[C_{xx}] \text{diag} \left(e^{ig\alpha_1(C)}, \dots, e^{ig\alpha_{N_c}(C)} \right) \Omega^\dagger[C_{xx}]. \quad (7.92)$$

Then $\Phi(C)$ reads

$$\Phi(C) = \frac{1}{N_c} \sum_{j=1}^{N_c} e^{ig\alpha_j(C)}. \quad (7.93)$$

The phases $\alpha_j(C)$ are gauge invariant and normalized so that $\alpha_j(C) \sim 1$ as $N_c \rightarrow \infty$. For simplicity we omit below all the indices (including space ones) except color.

The commutator of Φ 's can be estimated using the representation (7.93). Since

$$[\alpha_i, \alpha_j] \propto \delta_{ij}. \quad (7.94)$$

one gets

$$[\Phi(C), \Phi(C')] \sim g^2 \frac{1}{N_c} \sim \frac{1}{N_c^2} \quad (7.95)$$

in the limit (7.13), *i.e.* the commutator can be neglected as $N_c \rightarrow \infty$, and the field $\Phi(C)$ becomes classical.

Note that the commutator (7.95) is of order $1/N_c^2$. One factor $1/N_c$ is because of g in the definition (7.93) of $\Phi(C)$, while the other has a deep reason. Let us image the summation over j in Eq. (7.93) as some statistical averaging. It is well-known in statistics that such averages weakly fluctuate as $N_c \rightarrow \infty$, so that the dispersion is of order $1/N_c$. It is the factor which emerges in the commutator (7.95).

We see that the factorization is valid only for the gauge-invariant quantities which involve the averaging over the color indices, like that in Eq. (7.93). There is no reason to expect factorization for gauge invariants which do not involve this averaging, for instance for the phases $\alpha_j(C)$. Moreover, their commutator (7.94) is ~ 1 , so that $\alpha_j(C)$'s strongly fluctuate even at $N_c = \infty$. An explicit example of such strongly fluctuating gauge-invariant quantities was first constructed in Ref. [65].

The résumé to this Remark is that the factorization is due to the additional statistical averaging in the large- N_c limit. There is no reason to assume an existence of the master field in the strong sense in order to explain the factorization.

8 QCD in loop space

QCD can be entirely reformulated in terms of the colorless composite field $\Phi(C)$ — the trace of the Wilson loop for closed contours. This fact involves two main steps:

- i) All the observables are expressed via $\Phi(C)$.
- ii) Dynamics is entirely reformulated in terms of $\Phi(C)$.

This approach is especially useful in the large- N_c limit where everything is expressed via vacuum expectation value of $\Phi(C)$ — the Wilson loop average. Observables are given by summing the Wilson loop average over paths with the same weight as in free theory. The Wilson loop average obeys itself a close functional equation — the loop equation.

We begin this Section with presenting the formulas which relate observables to the Wilson loops. Then we translate quantum equation of motion of Yang–Mills theory into loop space. We derive the closed equation for the Wilson loop average as $N_c \rightarrow \infty$ and discuss its various properties, including a non-perturbative regularization. Finally, we briefly comment on what is known about solutions of the loop equation.

8.1 Observables in terms of Wilson loops

All observables in QCD can be expressed via the Wilson loops $\Phi(C)$ defined by Eq. (7.86). This property was first advocated by Wilson [1] on a lattice. Calculation of QCD observables can be divided in two steps:

- 1) Calculation of the Wilson loop averages for arbitrary contours.
- 2) Summation of the Wilson loop averages over the contours with some weight depending on a given observable.

The appropriate formulas for the continuum theory can be found in Ref. [71]. At finite N_c , observables are expressed via the n -loop averages

$$W_n(C_1, \dots, C_n) = \langle \Phi(C_1) \cdots \Phi(C_n) \rangle, \quad (8.1)$$

which are analogous to the n -point Green functions.

Great simplifications occur in these formulas at $N_c = \infty$, when all observables are expressed only via the one-loop average

$$W(C) = \langle \Phi(C) \rangle \equiv \left\langle \frac{1}{N_c} \text{tr} \mathbf{P} e^{ig \oint_C dx^\mu A_\mu} \right\rangle. \quad (8.2)$$

This is associated with the quenched approximation discussed in the Remark to Subsect. 5.5.

For example, the average of the product of two colorless quark vector currents (7.72) is given at large N_c by

$$\langle \bar{\psi} \gamma_\mu \psi(x_1) \bar{\psi} \gamma_\nu \psi(x_2) \rangle = \sum_{C \ni x_1, x_2} J_{\mu\nu}(C) \langle \Phi(C) \rangle, \quad (8.3)$$

where the sum runs over contours C passing through the points x_1 and x_2 as is depicted in Fig. 44a. An analogous formula for the (connected) correlators of three quark scalar currents

Fig. 44: Contours in the sum over paths representing observables: a) in Eq. (8.3) and b) in Eq. (8.4). The contour a) passes two nailed points x_1 and x_2 . The contour b) passes three nailed points x_1 , x_2 , and x_3 .

reads

$$\langle \bar{\psi} \psi(x_1) \bar{\psi} \psi(x_2) \bar{\psi} \psi(x_3) \rangle_{\text{conn}} = \sum_{C \ni x_1, x_2, x_3} J(C) \langle \Phi(C) \rangle. \quad (8.4)$$

where the sum runs over contours C passing through the three points x_1 , x_2 , and x_3 as is depicted in Fig. 44b. A general (connected) correlator of n quark currents is given by a similar formula with C passing through n points $x_1, \dots, x_n$ (some of them may coincide).

The weights $J_{\mu\nu}(C)$ in Eq. (8.3) and $J(C)$ in Eq. (8.4) are completely determined by free theory. If quarks were scalars rather than spinors, then we would get

$$J(C) = e^{-\frac{1}{2}m^2\tau - \frac{1}{2}\int_0^\tau dt \dot{x}_\mu^2(t)} \boxed{\text{scalar quarks}}. \quad (8.5)$$

Using the notation (2.6), we can rewrite Eq. (8.4) for scalar quarks as

$$\langle \psi^\dagger \psi(x_1) \psi^\dagger \psi(x_2) \psi^\dagger \psi(x_3) \rangle_{\text{conn}} = \sum'_{C \ni x_1, x_2, x_3} \langle \Phi(C) \rangle. \quad (8.6)$$

Therefore, we get the sum over paths of the Wilson loop, likewise in Subsects. 2.1 and 2.3.

For spinor quarks, an additional disentangling of the gamma-matrices is needed. This can be done in terms of a path integral over the momentum variable, with $k_\mu(t)$ ($0 \leq t \leq \tau$) being an appropriate trajectory. The result reads²³

$$J(C) = \int Dk(t) \text{sp} \mathbf{P} e^{-\int_0^\tau dt [ik_\mu(t)(\dot{x}_\mu(t) - \gamma_\mu(t)) + m]} \quad (8.7)$$

and

$$J_{\mu\nu}(C) = \int Dk(t) \text{sp} \mathbf{P} \left\{ \gamma_\mu(t_1) \gamma_\nu(t_2) e^{-\int_0^\tau dt [ik_\mu(t)(\dot{x}_\mu(t) - \gamma_\mu(t)) + m]} \right\}, \quad (8.8)$$

where the values t_1 and t_2 of the parameter t are associated with the points x_1 and x_2 in Eq. (8.3), and the symbol of P-ordering puts the matrices γ_μ and γ_ν at a proper order.

²³See, e.g., Ref. [72].

Fig. 45: Examples of a) smooth contour and b) contour with a cusp. The tangent vector to the contour jumps through angle γ at the cusp.

Remark on renormalization of Wilson loops

Perturbation theory for $W(C)$ can be obtained by expanding the path-ordered exponential in the definition (8.2) in g (see Eq. (7.26)) and averaging over the gluon field A_μ . Because of ultraviolet divergencies, we need a (gauge invariant) regularization. After such a regularization introduced, the Wilson loop average for a smooth contour C of the type in Fig. 45a reads

$$W(C) = e^{-g^2 \frac{(N_c^2-1)}{8\pi N_c} \frac{L(C)}{a}} W_{\text{ren}}(C), \quad (8.9)$$

where a is the cutoff, $L(C)$ is the length of C , and $W_{\text{ren}}(C)$ is finite when expressed via the renormalized charge g_R . The exponential factor is due to the renormalization of the mass of a heavy test quark, which was already discussed in the Remark to Subsect 3.4. This factor does not emerge in the dimensional regularization where $d = 4 - \varepsilon$. The multiplicative renormalization of the smooth Wilson loop was shown in Refs. [73, 74, 75].

If the contour C has a cusp (or cusps) but no self-intersections as is illustrated by Fig. 45b, then $W(C)$ is still multiplicatively renormalizable [76]:

$$W(C) = Z(\gamma) W_{\text{ren}}(C), \quad (8.10)$$

while the (divergent) factor $Z(\gamma)$ depends on the cusp angle (or angles) γ (or γ 's) and $W_{\text{ren}}(C)$ is finite when expressed via the renormalized charge g_R .

8.2 Schwinger–Dyson equations for Wilson loop

Dynamics of (quantum) Yang–Mills theory is described by the quantum equation of motion

$$\nabla_\mu^{ab} F_{\mu\nu}^b(x) \stackrel{\text{w.s.}}{=} \hbar \frac{\delta}{\delta A_\nu^a(x)} \quad (8.11)$$

which is understood in the weak sense, i.e. for the averages

$$\langle \nabla_\mu^{ab} F_{\mu\nu}^b(x) Q[A] \rangle = \hbar \left\langle \frac{\delta}{\delta A_\nu^a(x)} Q[A] \right\rangle. \quad (8.12)$$

The standard set of Schwinger–Dyson equations of Yang–Mills theory emerges when the functional $Q[A]$ is chosen in the form of the product of A 's as in Eq. (7.25).

Strictly speaking, the last statement is incorrect, since we have not added, in Eqs. (8.11) and (8.12), contributions coming from the variation of gauge-fixing and ghost terms in the Yang–Mills action. However, these two contributions are mutually cancelled for gauge-invariant functionals $Q[A]$. We shall deal below only with such gauge-invariant functionals (the Wilson loops). This is why we have not considered the contribution of the gauge-fixing and ghost terms.

It is also convenient to use the matrix notation (2.24), when Eq. (8.11) for the Wilson loop takes on the form

$$\left\langle \frac{1}{N_c} \text{tr} \mathbf{P} \nabla_\mu \mathcal{F}_{\mu\nu}(x) e^{\oint_C d\xi^\mu \mathcal{A}_\mu} \right\rangle = \left\langle \frac{g^2}{2N_c} \text{tr} \frac{\delta}{\delta \mathcal{A}_\nu(x)} \mathbf{P} e^{\oint_C d\xi^\mu \mathcal{A}_\mu} \right\rangle, \quad (8.13)$$

where we have restored the units with $\hbar = 1$.

The variational derivative on the RHS can be calculated by virtue of the formula

$$\frac{\delta A_\mu^{ij}(y)}{\delta A_\nu^{kl}(x)} = \delta_{\mu\nu} \delta^{(d)}(x-y) \left(\delta^{ij} \delta^{kl} - \frac{1}{N} \delta^{ij} \delta^{kl} \right) \quad (8.14)$$

which is a consequence of

$$\frac{\delta A_\mu^a(y)}{\delta A_\nu^b(x)} = \delta_{\mu\nu} \delta^{(d)}(x-y) \delta^{ab}. \quad (8.15)$$

The second term in the parentheses in Eq. (8.14) — same as in Eq. (7.6) — is because $\mathcal{A}_\mu$ is a matrix from the adjoint representation of $\text{SU}(N_c)$.

By using Eq. (8.14), we get for the variational derivative on RHS of Eq. (8.13):

$$\begin{aligned} \text{tr} \frac{\delta}{\delta \mathcal{A}_\nu(x)} \mathbf{P} e^{\oint_C d\xi^\mu \mathcal{A}_\mu} &= \oint_C dy_\nu \delta^{(d)}(x-y) \\ &\times \left[\frac{1}{N_c} \text{tr} \mathbf{P} e^{\int_{C_{yx}} d\xi^\mu \mathcal{A}_\mu} \frac{1}{N_c} \text{tr} \mathbf{P} e^{\int_{C_{xy}} d\xi^\mu \mathcal{A}_\mu} - \frac{1}{N_c^2} \text{tr} \mathbf{P} e^{\oint_C d\xi^\mu \mathcal{A}_\mu} \right]. \end{aligned} \quad (8.16)$$

The contours C_{yx} and C_{xy} , which are depicted in Fig. 46, are the parts of the loop C : from x to y and from y to x , respectively. They are always closed due to the presence of the delta-function. It implies that x and y should be the same points of *space* but not necessarily of the *contour* (i.e. they may be associated with different values of the parameter σ).

We finally rewrite Eq. (8.13) as

$$\begin{aligned} \left\langle \frac{1}{N_c} \text{tr} \mathbf{P} \nabla_\mu \mathcal{F}_{\mu\nu}(x) e^{\oint_C d\xi^\mu \mathcal{A}_\mu} \right\rangle \\ = \lambda \oint_C dy_\nu \delta^{(d)}(x-y) \left[\langle \Phi(C_{yx}) \Phi(C_{xy}) \rangle - \frac{1}{N^2} \langle \Phi(C) \rangle \right] \end{aligned} \quad (8.17)$$

where we have introduced

$$\lambda = \frac{g^2 N_c}{2}. \quad (8.18)$$

Notice that the RHS of Eq. (8.17) is completely represented via the (closed) Wilson loops.

Fig. 46: Contours C_{yx} and C_{xy} which enter the RHS's of Eqs. (8.13) and (8.17).

8.3 Path and area derivatives

As we already mentioned, the RHS of Eq. (8.17) is completely represented via the (closed) Wilson loops. It is crucial for the loop-space formulation of QCD that the LHS of Eq. (8.17) can also be represented in loop space as some operator applied to the Wilson loop. To do this we need to develop a differential calculus in loop space.

Loop space consists of arbitrary continuous closed loops, C . They can be described in a parametric form by the functions $x_\mu(\sigma) \in L_2$,²⁴ where $\sigma_0 \leq \sigma \leq \sigma_f$ and $\mu = 1, \dots, d$, which take on values in a d -dimensional Euclidean space. The functions $x_\mu(\sigma)$ can be discontinuous, generally speaking, for an arbitrary choice of the parameter σ . The continuity of the loop C implies a continuous dependence on parameters of the type of proper length.

The functions $x_\mu(\sigma) \in L_2$ which are associated with the elements of loop space obey the following restrictions:

- i) The points $\sigma = \sigma_0$ and $\sigma = \sigma_f$ are identified: $x_\mu(\sigma_0) = x_\mu(\sigma_f)$ — the loops are closed.
- ii) The functions $x_\mu(\sigma)$ and $\Lambda_{\mu\nu} x_\nu(\sigma) + \alpha_\mu$, with $\Lambda_{\mu\nu}$ and α_μ independent of σ , represent the same element of the loop space — rotational and translational invariance.
- iii) The functions $x_\mu(\sigma)$ and $x_\mu(\sigma')$ with $\sigma' = f(\sigma)$, $f'(\sigma) \geq 0$ describe the same loop — reparametrization invariance.

An example of functionals which are defined on the elements of loop space is the Wilson loop average (8.2) or, more generally, the n -loop average (8.1).

The differential calculus in loop space is built out of the path and area derivatives.

The *area derivative* of a functional $\mathcal{F}(C)$ is defined by the difference

$$\frac{\delta \mathcal{F}(C)}{\delta \sigma_{\mu\nu}(x)} \equiv \frac{1}{|\delta \sigma_{\mu\nu}|} \left[\mathcal{F} \left(\text{loop with } \delta C_{\mu\nu}(x) \right) - \mathcal{F} \left(\text{loop } C \right) \right] \quad (8.19)$$

where an infinitesimal loop $\delta C_{\mu\nu}(x)$ is attached to a given loop at the point x in the $\mu\nu$ -plane and $|\delta \sigma_{\mu\nu}|$ stands for the area enclosed by the $\delta C_{\mu\nu}(x)$.

Analogously, the *path derivative* is defined by

²⁴Let us remind that L_2 stands for the Hilbert space of functions $x_\mu(\sigma)$ whose square is integrable over the Lebesgue measure: $\int_{\sigma_0}^{\sigma_f} d\sigma x_\mu^2(\sigma) < \infty$.

$$\partial_\mu^x \mathcal{F}(C_{xx}) \equiv \frac{1}{|\delta x_\mu|} \left[\mathcal{F} \left(\text{loop with } \delta x_\mu \right) - \mathcal{F} \left(\text{loop } C \right) \right] \quad (8.20)$$

where δx_μ is an infinitesimal path along which the point x is shifted from the loop and $|\delta x_\mu|$ stands for the length of the δx_μ .

These two differential operations are well-defined for so-called functionals of the Stokes type which satisfy the backtracking condition — they do not change when an appendix passing back and forth is added to the loop at some point x :

$$\mathcal{F} \left(\text{loop with backtracking} \right) = \mathcal{F} \left(\text{loop } C \right). \quad (8.21)$$

This condition is equivalent to the Bianchi identity of Yang-Mills theory and is obviously satisfied by the Wilson loop (8.2) due to the properties of the non-Abelian phase factor (see Eq. (2.55)). Such functionals are known in mathematics as Chen integrals.

A simple example of the Stokes functional is the area of the minimal surface, $A_{\min}(C)$. It obviously satisfies Eq. (8.21). Otherwise, the length $L(C)$ of the loop C is not a Stokes functional, since the lengths of contours on the LHS and RHS of Eq. (8.21) are different.

For the Stokes functionals, the variation on the RHS of Eq. (8.19) is proportional to the area enclosed by the infinitesimally small loop $\delta C_{\mu\nu}(x)$ and does not depend on its shape. Analogously, the variation on the RHS of Eq. (8.20) is proportional to the length of the infinitesimal path δx_μ and does not depend on its shape.

If x is a regular point (like any point of the contour for the functional (8.2)), the RHS of Eq. (8.20) vanishes due to the backtracking condition (8.21). In order for the result to be nonvanishing, the point x should be a *marked* (or irregular) point. A simple example of the functional with a marked point x is

$$\Phi^a[C_{xx}] \equiv \frac{\text{tr}}{N} \left(t^a \mathbf{P} e^{\int_{C_{xx}} d\xi^\mu \mathcal{A}_\mu(\xi)} \right) \quad (8.22)$$

with the $\text{SU}(N_c)$ generator t^a inserted in the path-ordered product at the point x .

The area derivative of the Wilson loop is given by the Mandelstam formula

$$\frac{\delta}{\delta \sigma_{\mu\nu}(x)} \frac{1}{N_c} \text{tr} \mathbf{P} e^{\int_C d\xi^\mu \mathcal{A}_\mu} = \frac{1}{N_c} \text{tr} \mathbf{P} \mathcal{F}_{\mu\nu}(x) e^{\int_C d\xi^\mu \mathcal{A}_\mu}. \quad (8.23)$$

In order to prove it, it is convenient to choose $\delta C_{\mu\nu}(x)$ to be a rectangle in the $\mu\nu$ -plane, as is depicted in Fig. 2, and straightforwardly use the definition (8.19). The sense of Eq. (8.23) is very simple: $\mathcal{F}_{\mu\nu}$ is a curvature associated with the connection $\mathcal{A}_\mu$, as we already mentioned.

The functional on the RHS of Eq. (8.23) has a marked point x , and is of the type in Eq. (8.22). When the path derivative acts on such a functional according to the definition (8.20), the result reads

$$\partial_\mu^x \frac{1}{N_c} \text{tr} \mathbf{P} B(x) e^{\int_C d\xi^\mu \mathcal{A}_\mu} = \frac{1}{N_c} \text{tr} \mathbf{P} \nabla_\mu B(x) e^{\int_C d\xi^\mu \mathcal{A}_\mu}, \quad (8.24)$$

Ordinary space		Loop space	
$\Phi[A]$	phase factor	$\Phi(C)$	loop functional
$F_{\mu\nu}(x)$	field strength	$\frac{\delta}{\delta\sigma_{\mu\nu}(x)}$	area derivative
∇_μ^x	covariant derivative	∂_μ^x	path derivative
$\nabla \wedge F = 0$	Bianchi identity		Stokes functionals
$\nabla_\mu F_{\mu\nu} = \delta/\delta A_\nu$	Schwinger-Dyson equations		Loop equations

Table 2: Vocabulary for translation of Yang-Mills theory from ordinary space in loop space.

where

$$\nabla_\mu B = \partial_\mu B - [A_\mu, B] \quad (8.25)$$

is the covariant derivative (2.30) in the adjoint representation (see also Eq. (2.62)).

Combining Eqs. (8.23) and (8.24), we finally represent the expression on the LHS of Eq. (8.13) (or Eq. (8.17)) as

$$\frac{1}{N_c} \text{tr} \mathbf{P} \nabla_\mu \mathcal{F}_{\mu\nu}(x) e^{\oint_C d\xi^\mu A_\mu} = \partial_\mu^x \frac{\delta}{\delta\sigma_{\mu\nu}(x)} \frac{1}{N_c} \text{tr} \mathbf{P} e^{\oint_C d\xi^\mu A_\mu}, \quad (8.26)$$

i.e. via the action of the path and area derivatives on the Wilson loop. It is therefore rewritten in loop space.

A résumé of the results of this Subsection is presented in Table 2 as a vocabulary for translation of Yang-Mills theory from the language of ordinary space in the language of loop space.

Remark on Bianchi identity for Stokes functionals

The backtracking relation (8.21) can be equivalently represented as

$$\epsilon_{\mu\nu\lambda\rho} \partial_\mu^x \frac{\delta}{\delta\sigma_{\nu\lambda}(x)} \Phi(C) = 0, \quad (8.27)$$

by choosing the appendix in Eq. (8.21) to be an infinitesimal straight line in the ρ -direction and geometrically applying the Stokes theorem. Using Eqs. (8.23) and (8.24), Eq. (8.27) can in turn be rewritten as

$$\epsilon_{\mu\nu\lambda\rho} \frac{1}{N_c} \text{tr} \mathbf{P} \nabla_\mu \mathcal{F}_{\nu\lambda}(x) e^{\oint_C d\xi^\mu A_\mu} = 0. \quad (8.28)$$

Therefore, Eq. (8.27) represents the Bianchi identity (2.38) in loop space.

Remark on relation to variational derivative

The standard variational derivative, $\delta/\delta x_\mu(\sigma)$, can be expressed via the path and area derivatives by the formula

$$\frac{\delta}{\delta x_\mu(\sigma)} = \dot{x}_\nu(\sigma) \frac{\delta}{\delta\sigma_{\mu\nu}(x(\sigma))} + \sum_{i=1}^m \partial_\mu^x \delta(\sigma - \sigma_i), \quad (8.29)$$

where the sum on the RHS is present for the case of a functional having m marked (irregular) points $x_i \equiv x(\sigma_i)$. A simplest example of the functional with m marked points is just a function of m variables $x_1, \dots, x_m$.

By using Eq. (8.29), the path derivative can be calculated as the limiting procedure

$$\partial_\mu^{x(\sigma)} = \int_{\sigma-0}^{\sigma+0} d\sigma' \frac{\delta}{\delta x_\mu(\sigma')}. \quad (8.30)$$

The result is obviously nonvanishing only when ∂_μ^x is applied to a functional with $x(\sigma)$ being a marked point.

It is nontrivial that the area derivative can also be expressed via the variational derivative [74]:

$$\frac{\delta}{\delta\sigma_{\mu\nu}(x(\sigma))} = \int_{\sigma-0}^{\sigma+0} d\sigma' (\sigma' - \sigma) \frac{\delta}{\delta x_\mu(\sigma')} \frac{\delta}{\delta x_\nu(\sigma)}. \quad (8.31)$$

The point is that the six-component quantity, $\delta/\delta\sigma_{\mu\nu}(x(\sigma))$, is expressed via the four-component one, $\delta/\delta x_\mu(\sigma)$, which is possible because the components of $\delta/\delta\sigma_{\mu\nu}(x(\sigma))$ are dependent due to the loop-space Bianchi identity (8.27).

8.4 Loop equations

By virtue of Eq. (8.26), Eq. (8.17) can be represented completely in loop space:

$$\partial_\mu^x \frac{\delta}{\delta\sigma_{\mu\nu}(x)} \langle \Phi(C) \rangle = \lambda \oint_C dy_\nu \delta^{(d)}(x - y) \left\langle \left[\Phi(C_{yx}) \Phi(C_{xy}) - \frac{1}{N_c^2} \Phi(C) \right] \right\rangle. \quad (8.32)$$

or, using the definitions (8.1) and (8.2) of the loop averages, as

$$\partial_\mu^x \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W(C) = \lambda \oint_C dy_\nu \delta^{(d)}(x - y) \left[W_2(C_{yx}, C_{xy}) - \frac{1}{N_c^2} W(C) \right]. \quad (8.33)$$

This equation is not closed. Having started from $W(C)$, we obtain another quantity, $W_2(C_1, C_2)$, so that Eq. (8.33) connects the one-loop average with a two-loop one. This is similar to the case of the (quantum) φ^3 -theory, whose Schwinger–Dyson equations (7.87) connect the n -point Green functions with different n . We shall derive this complete set of equations for the n -loop averages in this Subsection later on.

However, the two-loop average factorizes in the large- N_c limit:

$$W_2(C_1, C_2) = W(C_1)W(C_2) + \mathcal{O}\left(\frac{1}{N_c^2}\right), \quad (8.34)$$

as was discussed in Subsect. 7.6. Keeping the constant λ (defined by Eq. (8.18)) fixed in the large- N_c limit as is prescribed by Eq. (7.13), we get [70]

$$\partial_\mu^\epsilon \frac{\delta}{\delta \sigma_{\mu\nu}(x)} W(C) = \lambda \oint_C dy_\nu \delta^{(d)}(x-y) W(C_{yx}) W(C_{xy}) \quad (8.35)$$

as $N_c \rightarrow \infty$.

Equation (8.35) is a closed equation for the Wilson loop average in the large- N_c limit. It is referred to as the *loop equation* or the *Makeenko–Migdal equation*.

To find $W(C)$, Eq. (8.35) should be solved in the class of Stokes functionals with the initial condition

$$W(0) = 1 \quad (8.36)$$

for loops which are shrunk to points. This is a consequence of the obvious property of the Wilson loop

$$e^{\oint_C d\xi^\mu A_\mu} = 1 \quad (8.37)$$

and the normalization $\langle 1 \rangle = 1$ of the averages.

The factorization (8.34) can itself be derived from the chain of loop equations. Proceeding as before, we get

$$\begin{aligned} & \frac{1}{\lambda} \partial_\mu^\epsilon \frac{\delta}{\delta \sigma_{\mu\nu}(x)} W_n(C_1, \dots, C_n) \\ &= \oint_{C_1} dy_\nu \delta^{(d)}(x-y) \left[W_{n+1}(C_{xy}, C_{yx}, \dots, C_n) - \frac{1}{N_c^2} W_n(C_1, \dots, C_n) \right] \\ &+ \sum_{j \geq 2} \frac{1}{N_c^2} \oint_{C_j} dy_\nu \delta^{(d)}(x-y) \left[W_{n-1}(C_1 C_j, \dots, \underline{C_j}, \dots, C_n) \right. \\ &\quad \left. - W_n(C_1, \dots, C_n) \right]. \end{aligned} \quad (8.38)$$

Here x belongs to C_1 ; $C_1 C_j$ stands for the joining of C_1 and C_j ; $\underline{C_j}$ means that C_j is omitted.

Equation (8.38) looks like Eq. (7.87) for the φ^3 -theory. Moreover, the number of colors N_c enters Eq. (8.38) simply as a scalar factor N_c^{-2} , likewise Plank's constant $\hbar$ enters Eq. (7.87). It is the major advantage of the use of loop space. What is said in Subsect. 7.8 about the

“semiclassical” nature of the $1/N_c$ -expansion of QCD is explicitly realized in Eq. (8.38). Its expansion in $1/N_c$ is straightforward.

At $N_c = \infty$, Eq. (8.38) is simplified to

$$\partial_\mu^\epsilon \frac{\delta}{\delta \sigma_{\mu\nu}(x)} W_n(C_1, \dots, C_n) = \lambda \oint_{C_1} dy_\nu \delta^{(d)}(x-y) W_{n+1}(C_{yx}, C_{xy}, \dots, C_n). \quad (8.39)$$

This equation possesses a factorized solution

$$\begin{aligned} W_n(C_1, \dots, C_n) &= \langle \Phi(C_1) \rangle \dots \langle \Phi(C_n) \rangle + \mathcal{O}\left(\frac{1}{N_c^2}\right) \\ &\equiv W(C_1) \dots W(C_n) + \mathcal{O}\left(\frac{1}{N_c^2}\right) \end{aligned} \quad (8.40)$$

provided $W(C)$ obeys Eq. (8.35) which plays the role of a “classical” equation in the large- N_c limit. Thus, we have given a non-perturbative proof of the large- N_c factorization of the Wilson loops.

8.5 Relation to planar diagrams

The perturbation-theory expansion of the Wilson loop average can be calculated from Eq. (7.26) which we represent in the form

$$W(C) = 1 + \sum_{n=2}^{\infty} \oint_C dx_1^{\mu_1} \oint_C dx_2^{\mu_2} \dots \oint_C dx_n^{\mu_n} \theta_c(1, 2, \dots, n) G_{\mu_1 \mu_2 \dots \mu_n}^{(n)}(x_1, x_2, \dots, x_n), \quad (8.41)$$

where $\theta_c(1, 2, \dots, n)$ orders the points $x_1, \dots, x_n$ along contour in the cyclic order and $G_{\mu_1 \dots \mu_n}^{(n)}$ is given by Eq. (7.51). This θ -function has the meaning of the propagator of a test heavy particle which lives in the contour C .

We assume, for definitiveness, the dimensional regularization throughout this Subsection to make all the integrals well-defined.

Each term on the RHS of Eq. (8.41) can be conveniently represented by the diagram in Fig. 47, where the integration over the contour C is associated with each point x_i lying in the contour C .

These diagrams are analogous to those discussed in Subsect. 7.3 with one external boundary — the Wilson loop in the given case. This was already mentioned in the Remark to Subsect. 7.3. In the large- N_c limit, only planar diagrams survive. Some of them, which are of the lowest order in λ , are depicted in Fig. 48.

The large- N_c loop equation (8.35) describes the sum of the planar diagrams. Its iterative solution in λ reproduces the set of planar diagrams for $W(C)$ provided the initial condition (8.36) and some boundary conditions for asymptotically large contours are imposed.

Equation (8.41) can be viewed as an ansatz for $W(C)$ with some unknown functions $G_{\mu_1 \dots \mu_n}^{(n)}(x_1, \dots, x_n)$ to be determined by the substitution into the loop equation. To preserve symmetry properties of $W(C)$, the functions $G^{(n)}$ must be symmetric under a cyclic permutation of the points $1, \dots, n$ and depend only on $x_i - x_j$ (translational invariance). A main

Fig. 47: Graphic representation of the terms on the RHS of Eq. (8.41).

Fig. 48: Planar diagrams for $W(C)$: a) of order λ with gluon propagator, and of order λ^2 b) with two noninteracting gluons and c) with the three-gluon vertex. Diagrams of order λ^2 with one-loop insertions to gluon propagator are not drawn.

advantage of this ansatz is that it automatically corresponds to a Stokes functional, due to the properties of vector integrals, and the initial condition (8.36) is satisfied.

The action of the area and path derivatives on the ansatz (8.41) is easily calculable. For instance, the area derivative reads

$$\begin{aligned} \frac{\delta W(C)}{\delta \sigma_{\mu\nu}(z)} &= \sum_{n=1}^{\infty} \oint_C dx_1^{\mu_1} \dots \oint_C dx_n^{\mu_n} \theta_c(1, 2, \dots, n) \\ &\times \left[\left(\partial_\mu^z \delta_{\nu\alpha} - \partial_\nu^z \delta_{\mu\alpha} \right) G_{\alpha\mu_1 \dots \mu_n}^{(n+1)}(z, x_1, \dots, x_n) \right. \\ &\left. + (\delta_{\mu\beta} \delta_{\nu\alpha} - \delta_{\mu\alpha} \delta_{\nu\beta}) G_{\alpha\beta\mu_1 \dots \mu_n}^{(n+2)}(z, z, x_1, \dots, x_n) \right]. \end{aligned} \quad (8.42)$$

The analogy with the Mandelstam formula (8.23) is obvious.

More about solving the loop equation by the ansatz (8.41) can be found in Refs. [71, 77, 78].

8.6 Loop-space Laplacian and regularization

The loop equation (8.35) is *not* yet entirely formulated in loop space. It is a d -vector equation whose both sides depend explicitly on the point x which does not belong to loop

space. The fact that we have a d -vector equation for a scalar quantity means, in particular, that Eq. (8.35) is overspecified.

A practical difficulty in solving Eq. (8.35) is that the area and path derivatives, $\delta/\delta\sigma_{\mu\nu}(x)$ and ∂_μ^x , which enter the LHS are complicated, generally speaking, non-commutative operators. They are intimately related to the Yang-Mills perturbation theory where they correspond to the non-Abelian field strength $F_{\mu\nu}$ and the covariant derivative ∇_μ . However, it is not easy to apply these operators to a generic functional $W(C)$ which is defined on elements of loop space.

A much more convenient form of the loop equation can be obtained by integrating both sides of Eq. (8.35) over dx_ν along the same contour C , which yields

$$\oint_C dx_\nu \partial_\mu^x \frac{\delta}{\delta \sigma_{\mu\nu}(x)} W(C) = \lambda \oint_C dx_\mu \oint_C dy_\mu \delta^{(d)}(x-y) W(C_{xy}) W'(C_{xy}). \quad (8.43)$$

Now both the operator on the LHS and the functional on the RHS are scalars without labeled points and are well-defined in loop space. The operator on the LHS of Eq. (8.43) can be interpreted as an infinitesimal variation of elements of loop space.

Equations (8.35) and (8.43) are completely equivalent. A proof of equivalence of scalar Eq. (8.43) and original d -vector Eq. (8.35) is based on the important property of Eq. (8.35), both sides of which are annihilated identically by the operator ∂_ν^x . It is a consequence of the identity (see Subsect. 2.2)

$$\nabla_\mu \nabla_\nu \mathcal{F}_{\mu\nu} = -\frac{1}{2} [\mathcal{F}_{\mu\nu}, \mathcal{F}_{\mu\nu}] = 0 \quad (8.44)$$

in the ordinary space. Due to this property, the vanishing of the contour integral of some vector is equivalent to vanishing of the vector itself, so that Eq. (8.35) can in turn be deduced from Eq. (8.43).

Equation (8.43) is associated with the so-called second-order Schwinger-Dyson equation

$$\int d^d x \nabla_\mu F_{\mu\nu}^a(x) \frac{\delta}{\delta A_\nu^a(x)} \stackrel{w.s.}{=} \hbar \int d^d x d^d y \delta^{(d)}(x-y) \frac{\delta}{\delta A_\nu^a(y)} \frac{\delta}{\delta A_\mu^a(x)} \quad (8.45)$$

in the same sense as Eq. (8.35) is associated with Eq. (8.11). It is called “second order” since the RHS involves two variational derivatives with respect to A_μ .

The operator on the LHS of Eq. (8.43) is a well-defined object in loop space. When applied to regular functionals which do not have marked points, it can be represented, using Eqs. (8.30) and (8.31), in an equivalent form

$$\Delta \equiv \oint_C dx_\nu \partial_\mu^x \frac{\delta}{\delta \sigma_{\mu\nu}(x)} = \int_{\sigma_0}^{\sigma_f} d\sigma \int_{\sigma=0}^{\sigma+0} d\sigma' \frac{\delta}{\delta x_\mu(\sigma')} \frac{\delta}{\delta x_\mu(\sigma)}. \quad (8.46)$$

As was first pointed out by Gervais and Neveu [79], this operator is nothing but a functional extension of the Laplace operator, which is known in mathematics as the Levy operator [80].

Equation (8.43) can be represented in turn as an (inhomogeneous) functional Laplace equation

$$\Delta W(C) = \lambda \oint_C dx_\mu \oint_C dy_\mu \delta^{(d)}(x-y) W(C_{yx}) W(C_{xy}). \quad (8.47)$$

We shall refer to this equation as the loop-space Laplace equation.

The form (8.47) of the loop equation is convenient for a non-perturbative ultraviolet regularization.

The idea is to start from the regularized version of Eq. (8.45), replacing the delta-function on the RHS by the kernel of the regularizing operator:

$$\delta^{ab} \delta^{(d)}(x-y) \xrightarrow{\text{Reg}} \langle y | \mathbf{R}^{ab} | x \rangle = \mathbf{R}^{ab} \delta^{(d)}(x-y) \quad (8.48)$$

with

$$\mathbf{R}^{ab} = \left(e^{a^2 \nabla^2 / 2} \right)^{ab}, \quad (8.49)$$

where ∇_μ is the covariant derivative in the adjoint representation. The regularized version of Eq. (8.45) is

$$\int d^d x \nabla_\mu F_{\mu\nu}^a(x) \frac{\delta}{\delta A_\nu^a(x)} \stackrel{\text{w.s.}}{=} \hbar \int d^d x d^d y \langle y | \mathbf{R}^{ab} | x \rangle \frac{\delta}{\delta A_\nu^b(y)} \frac{\delta}{\delta A_\mu^a(x)}. \quad (8.50)$$

To translate Eq. (8.50) in loop space, we use the path-integral representation (2.53) which gives

$$\langle y | \mathbf{R}^{ab} | x \rangle = \int_{\substack{r(0)=x \\ r(a^2)=y}} Dr(t) e^{-\frac{1}{2} \int_0^a dt \dot{r}^2(t)} 2 \text{tr} [t^a U(r_{yx}) t^b U(r_{xy})] \quad (8.51)$$

with

$$U(r_{yx}) = \mathbf{P} e^{\int_x^y dr_\mu \mathcal{A}_\mu(r)}, \quad (8.52)$$

where the integration is over regulator paths $r_\mu(t)$ from x to y whose typical length is $\sim a$. The conventional measure is implied in (8.51) so that

$$\begin{aligned} & \int_{\substack{r(0)=x \\ r(a^2)=y}} Dr(t) e^{-\frac{1}{2} \int_0^a dt \dot{r}^2(t)} 2 \text{tr} [t^a t^b] \\ &= \delta^{ab} e^{a^2 \partial^2 / 2} \delta^{(d)}(x-y) = \delta^{ab} \frac{1}{(2\pi a^2)^{d/2}} e^{-\frac{(x-y)^2}{2a^2}}. \end{aligned} \quad (8.53)$$

Calculating the variational derivatives on the RHS of Eq. (8.50), using Eq. (8.51) and the completeness condition (7.6), we get as $N \rightarrow \infty$:

$$\begin{aligned} & \int d^d x d^d y \langle y | \mathbf{R}^{ab} | x \rangle \frac{\delta}{\delta A_\nu^b(y)} \frac{\delta}{\delta A_\mu^a(x)} \Phi(C) \\ &= \lambda \oint_C dx_\mu \oint_C dy_\mu \int_{\substack{r(0)=x \\ r(a^2)=y}} Dr(t) e^{-\frac{1}{2} \int_0^a dt \dot{r}^2(t)} \Phi(C_{yx} r_{xy}) \Phi(C_{xy} r_{yx}), \end{aligned} \quad (8.54)$$

Fig. 49: Contours $C_{yx} r_{xy}$ and $C_{xy} r_{yx}$ which enter the LHS's of Eqs. (8.54) and (8.55).

where the contours $C_{yx} r_{xy}$ and $C_{xy} r_{yx}$ are depicted in Fig. 49. Averaging over the gauge field and using the large- N_c factorization, we arrive at the regularized loop-space Laplace equation [81]

$$\Delta W(C) = \lambda \oint_C dx_\mu \oint_C dy_\mu \int_{\substack{r(0)=x \\ r(a^2)=y}} Dr(t) e^{-\frac{1}{2} \int_0^a dt \dot{r}^2(t)} W(C_{yx} r_{xy}) W(C_{xy} r_{yx}) \quad (8.55)$$

which manifestly recovers Eq. (8.47) when $a \rightarrow 0$.

The constructed regularization is non-perturbative while perturbatively reproduces regularized Feynman diagrams. An advantage of this regularization of the loop equation is that the contours $C_{yx} r_{xy}$ and $C_{xy} r_{yx}$ on the RHS of Eq. (8.55) both are closed and do not have marked points if C does not have. Therefore, Eq. (8.55) is written entirely in loop space.

Remark on functional Laplacian

It is worth noting that the representation of the functional Laplacian on the RHS of Eq. (8.46) is defined for a wider class of functionals than Stokes functionals. The point is that the standard definition of the functional Laplacian from the book by Levy [80] uses solely the concept of the second variation of a functional $U[x]$, namely the term in the second variation which is proportional to $(\delta x_\mu(\sigma))^2$:

$$\delta^2 U[x] = \frac{1}{2} \int_{\sigma_0}^{\sigma_f} d\sigma U''_{xx}[x] (\delta x_\mu(\sigma))^2 + \dots \quad (8.56)$$

The functional Laplacian Δ is then defined by the formula

$$\Delta U[x] = \int_{\sigma_0}^{\sigma_f} d\sigma U''_{xx}[x]. \quad (8.57)$$

Here $U[x]$ can be an arbitrary, not necessarily parametric invariant, functional. To emphasize this obstacle, we use the notation $U[x]$ for generic functionals which are defined on L_2 space in comparison to $U(C)$ for the functionals which are defined on elements of loop space. It is easier to deal with the whole operator Δ , rather than separately with the area and path derivatives.

The functional Laplacian is parametric invariant and possesses a number of remarkable properties. While a finite-dimensional Laplacian is an operator of the second order, the functional Laplacian is that of the first order and satisfies the Leibnitz rule

$$\Delta(UV) = \Delta(U)V + U\Delta(V). \quad (8.58)$$

The functional Laplacian can be approximated [82] in loop space by a (second-order) partial differential operator in such a way to preserve these properties in the continuum limit. This loop-space Laplacian can be inverted to determine a Green function $G(C, C')$ in the form of a sum over surfaces $S_{C, C'}$ connecting two loops:

$$G(C, C') = \sum_{S_{C, C'}} \dots, \quad (8.59)$$

which is analogous to the sum-over-path representation of the Green function of the ordinary Laplacian. The standard perturbation theory can then be recovered by iterating Eq. (8.47) (or its regularized version (8.55)) in λ with the Green function (8.59).

8.7 Survey of non-perturbative solutions

While the loop equations were proposed long ago, not much is known about their non-perturbative solutions. We briefly list some of the results.

It was shown in Ref. [83] that area law

$$W(C) \equiv \langle \Phi(C) \rangle \propto e^{-K \cdot A_{\min}(C)} \quad (8.60)$$

satisfies the large- N_c loop equation for asymptotically large C . However, a self-consistency equation for K , which should relate it to the bare charge and the cutoff, was not investigated. In order to do this, one needs more detailed information about the behavior of $W(C)$ for intermediate loops.

The *free* bosonic Nambu-Goto string which is defined as a sum over surfaces spanned by C

$$W(C) = \sum_{S: \partial S = C} e^{-K \cdot A(S)} \quad (8.61)$$

with the action being the area $A(S)$ of the surface S , is *not* a solution for intermediate loops. Consequently, QCD does not reduce to this kind of string, as was originally expected in Refs. [84, 85, 86]. Roughly speaking, the ansatz (8.61) is not consistent with the factorized structure on the RHS of Eq. (8.35).

Nevertheless, it was shown that if a free string satisfies Eq. (8.35), then the same interacting string satisfies the loop equations for finite N_c . Here “free string” means, as usual in string theory, that only surfaces of genus zero are present in the sum over surfaces, while surfaces of higher genera are associated with a string interaction. The coupling constant of this interaction is $\mathcal{O}(N_c^{-2})$.

A formal solution of Eq. (8.35) for all loops was found by Migdal [87] in the form of a fermionic string

$$W(C) = \sum_{S: \partial S = C} \int D\psi e^{-\int d^2\xi [\bar{\psi} \sigma_\mu \partial_\mu \psi + \bar{\psi} \psi m^4 \sqrt{g}]}, \quad (8.62)$$

where the world sheet of the string is parametrized by the coordinates ξ_1 and ξ_2 for which the 2-dimensional metric is conformal, *i.e.* diagonal. The field $\psi(\xi)$ describes 2-dimensional elementary fermions (elves) living in the surface S , and m stands for their mass. Elves were introduced to provide factorization which now holds due to some remarkable properties of 2-dimensional fermions. For large loops, the internal fermionic structure becomes frozen, so that the empty string behavior (8.60) is recovered. For small loops, the elves are necessary for asymptotic freedom. However, it is unclear whether or not the string solution (8.62) is practically useful for study of multicolor QCD, since the methods of dealing with the string theory in four dimensions are not yet developed.

A very interesting solution of the large- N_c loop equation on a lattice was found by Eguchi and Kawai [88]. They showed that the $SU(N_c)$ gauge theory on an infinite lattice reduces at $N_c = \infty$ to the model on a hypercube, which is described by the partition function

$$Z_{\text{EK}}(\beta) = \int \prod_{\rho=1}^d dU_\rho e^{\beta \sum_{\mu > \nu} \text{Re} \frac{1}{N_c} \text{tr}(U_\mu U_\nu U_\mu^\dagger U_\nu^\dagger)}. \quad (8.63)$$

Here the matrices U_ρ ($\rho = 1, \dots, d$) depend only on the direction ρ , rather than on a space-time point x . The equivalence is possible only at $N_c = \infty$, when the space-time dependence is absorbed by the internal symmetry group. More about this large- N_c reduction can be found in the review [89].

References

- [1] K.G. Wilson, *Confinement of quarks*, Phys. Rev. **D10** (1974) 2445.
- [2] G. 't Hooft, *A planar diagram theory for strong interactions*, Nucl. Phys. **B72** (1974) 461.
- [3] C.N. Yang and R.L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954) 191.
- [4] R.P. Feynman, *Mathematical formulation of the quantum theory of electromagnetic interaction*, Phys. Rev. **80** (1950) 440.
- [5] H. Weyl, *Eine neue Erweiterung der Relativitätstheorie*, Ann. der Phys. **59** (1919) v.10, pp.101-133.
- [6] F. London, *Quantenmechanische Deutung der Theorie von Weyl*, Z. Phys. **42** (1927) 375.
- [7] D.H. Kobe, *Aharonov-Bohm effect revisited*, Ann. Phys. **123** (1979) 381.
- [8] F.J. Wegner, *Duality in generalized Ising models and phase transitions without local order parameter*, J. Math. Phys. **12** (1971) 2259.
- [9] J. Kogut and L. Susskind, *Hamiltonian formulation of Wilson's lattice gauge theories*, Phys. Rev. **D11** (1975) 395.
- [10] G. Bhanot and M. Creutz, *Variant actions and phase structure in lattice gauge theory*, Phys. Rev. **D24** (1981) 3212.
- [11] S.B. Khokhlachev and Yu.M. Makeenko, *Phase transition over the gauge group center and quark confinement in QCD*, Phys. Lett. **101B** (1981) 403.
- [12] V.N. Gribov, *Quantization of non-Abelian gauge theories*, Nucl. Phys. **B139** (1978) 1.
- [13] K.G. Wilson, *Quarks and strings on a lattice*, in New phenomena in a subnuclear physics (Erice 1975), ed. A. Zichichi, Plenum, N.Y., 1977.
- [14] M. Creutz, *Quarks gluons and lattices*, Cambridge Univ. Press, 1983 (Russian transl.: Mir., M., 1987).
- [15] J.-M. Drouffe and J.-B. Zuber, *Strong coupling and mean field methods in lattice gauge theories*, Phys. Rep. **102** (1983) 1.
- [16] M. Creutz, *Asymptotic-freedom scales*, Phys. Rev. Lett. **45** (1980) 313.
- [17] R. Balian, J.M. Drouffe and C. Itzykson, *Gauge fields on a lattice. I. General outlook*, Phys. Rev. **D10** (1974) 3376.
- [18] A.A. Migdal, *Recursion equations in gauge theories*, ZhETF **69** (1975) 810 (English transl.: Sov. Phys. JETP **42** (1975) 413).
- [19] L.P. Kadanoff, *Recursion relations in statistical physics and field theory*, Ann. Phys. **100** (1976) 359.
- [20] K.G. Wilson, *Quantum chromodynamics on a lattice*, in New developments in quantum field theory and statistical mechanics, eds. M. Levy and P. Mitter, Plenum, N.Y., 1977.
- [21] M. Creutz, L. Jacobs and C. Rebbi, *Experiments with a gauge-invariant Ising system*, Phys. Rev. Lett. **42** (1979) 1390; *Monte Carlo study of Abelian lattice gauge theories*, Phys. Rev. **D20** (1979) 1915.
- [22] M. Creutz, *Confinement and critical dimensionality of space-time*, Phys. Rev. Lett. **43** (1979) 553; *Monte Carlo study of quantized SU(2) gauge theory*, Phys. Rev. **D21** (1980) 2308.

- [23] Yu.M. Makeenko, *The Monte Carlo method in lattice gauge theories*, Usp. Fiz. Nauk **143** (1984) 161 (English translation: Sov. Phys. Usp. **27** (1984) 401).
- [24] B. Lautrup and M. Nauenberg, *Phase transition in four-dimensional compact QED*, Phys. Lett. **95B** (1980) 63.
- [25] Lattice 94, Proc. of the XIth Int. Symp. on lattice field theory, ed. F. Karsch *et al.*, Nucl. Phys. B (Proc. Suppl.) **42**, 1995.
- [26] K.G. Wilson and J. Kogut, *The renormalization group and the ϵ -expansion*, Phys. Rep. **12C** (1974) 75 (Russian transl.: Mir., M., 1975).
- [27] J. Greensite and B. Lautrup, *Phase transition and mean-field methods in lattice gauge theories*, Phys. Lett. **104B** (1981) 41.
- [28] P. Cvitanovich, J. Greensite and B. Lautrup, *The crossover point in lattice gauge theories with continuous gauge groups*, Phys. Lett. **105B** (1981) 197.
- [29] S. Elitzur, *Impossibility of spontaneously breaking local symmetries*, Phys. Rev. **D12** (1975) 3978.
- [30] R.P. Feynman, *Statistical mechanics*, Benjamin, Mass., 1972 (Russian transl.: Mir., M., 1975).
- [31] J.M. Drouffe, *The mean-field approximation in the SU(2) pure lattice gauge theory*, Phys. Lett. **105B** (1981) 46.
- [32] H. Flyvbjerg, B. Lautrup and J.B. Zuber, *Mean field with corrections in lattice gauge theory*, Phys. Lett. **110B** (1982) 279.
- [33] K. Akemi *et al.*, *Scaling study of pure gauge lattice QCD by Monte Carlo renormalization group method*, Phys. Rev. Lett. **71** (1993) 3063.
- [34] M. Creutz, L. Jacobs and C. Rebbi, *Monte Carlo computations in lattice gauge theories*, Phys. Rep. **95** (1983) 201.
- [35] C.B. Lang and C. Rebbi, *Potential and restoration of rotational symmetry in SU(2) lattice gauge theory*, Phys. Lett. **115B** (1982) 137.
- [36] L.H. Karsten and J. Smit, *Lattice fermions: species doubling, chiral invariance and the triangle anomaly*, Nucl. Phys. **B183** (1981) 103.
- [37] H.B. Nielsen and M. Ninomiya, *Absence of neutrinos on a lattice*, 1. *Proof by homotopy theory*, Nucl. Phys. **B185** (1981) 20; 2. *Intuitive topological proof*, Nucl. Phys. **B193** (1981) 173.
- [38] L. Susskind, *Lattice fermions*, Phys. Rev. **D16** (1977) 3031.
- [39] H.S. Sharatchandra, H.J. Thun and P. Weisz, *Susskind fermions on a Euclidean lattice*, Nucl. Phys. **B192** (1981) 205.
- [40] N. Kawamoto and J. Smit, *Effective Lagrangian and dynamical symmetry breaking in strongly coupled lattice QCD*, Nucl. Phys. **B192** (1981) 100.
- [41] H. Kluberg-Stern, A. Morel, O. Napoly and B. Petersson, *Flavors of Lagrangian Susskind fermions*, Nucl. Phys. **B220** (1983) 447.
- [42] M. Bander, *Equivalence of lattice gauge and spin theories*, Phys. Lett. **126B** (1983) 463.
- [43] H.W. Hamber, *Lattice gauge theories at large N_F* , Phys. Lett. **126B** (1983) 471.
- [44] H. Hamber and G. Parisi, *Numerical estimates of hadron masses in a pure SU(3) gauge theory*, Phys. Rev. Lett. **47** (1981) 1792.

- [45] H.E. Stanley, *Spherical model as the limit of infinite spin dimensionality*, Phys. Rev. **176** (1968) 718.
- [46] K.G. Wilson, *Quantum field-theory models in less than 4 dimensions*, Phys. Rev. **D7** (1973) 2911.
- [47] D.J. Gross and A. Neveu, *Dynamical symmetry breaking in asymptotically free field theories*, Phys. Rev. **D10** (1974) 3235.
- [48] A.A. Anselm, *A model of a field theory with nonvanishing renormalized charge*, Zh. Exp. Teor. Fiz. **36** (1959) 863 (English transl.: Sov. Phys. JETP **9** (1959) 608).
- [49] K. Johnson, *Solution of the equations for the Green's functions of a two dimensional relativistic field theory*, Nuovo Cim. **20** (1961) 773.
- [50] G. Parisi, *The theory of non-renormalizable interaction*, Nucl. Phys. **B100** (1975) 368.
- [51] G. Mack and A. Salam, *Finite-component field representation of the conformal group*, Ann. Phys. **53** (1969) 144.
- [52] D. Gross and J. Wess, *Scale invariance, conformal invariance, and the high-energy behavior of scattering amplitudes*, Phys. Rev. **D2** (1970) 753.
- [53] R. Jackiw, *Field theoretic investigations in current algebra*, in *Lectures on Current Algebra and its Applications* by S.B. Treiman, R. Jackiw, and D.J. Gross, Princeton Univ. Press, 1972, p. 97; also in *Current Algebra and Anomalies* by S.B. Treiman, R. Jackiw, B. Zumino, and E. Witten, World Sci., 1985, p. 81.
- [54] A.M. Polyakov, *Conformal symmetry of critical fluctuations*, Pis'ma v ZhETF **12** (1970) 381 (English transl.: JETP Lett. **12** (1970) 381).
- [55] A.A. Migdal, *On hadronic interactions at small distances*, Phys. Lett. **37B** (1971) 98.
- [56] A.M. Polyakov, *Interaction of Goldstone particles in two dimensions. Applications to ferromagnets and massive Yang-Mills fields*, Phys. Lett. **59B** (1975) 79.
- [57] A.A. Belavin and A.M. Polyakov, *Metastable states of two-dimensional isotropic ferromagnet*, Pis'ma v ZhETF **22** (1975) 503 (English transl.: JETP Lett. **22** (1975) 245).
- [58] G. Veneziano, *Some aspects of a unified approach to gauge, dual and Gribov theories*, Nucl. Phys. **B117** (1976) 519.
- [59] W.T. Tutte, *A census of planar triangulations*, Can. J. Math. **14** (1962) 21.
- [60] J. Koplik, A. Neveu and S. Nussinov, *Some aspects of the planar perturbation series*, Nucl. Phys. **B123** (1977) 109.
- [61] P. Cvitanović, *Planar perturbation expansion*, Phys. Lett. **99B** (1981) 49.
- [62] P. Cvitanović, P.G. Lauwers, and P.N. Scharbach, *The planar sector of field theories*, Nucl. Phys. **B203** (1982) 385.
- [63] E. Witten, *The $1/N$ expansion in atomic and particle physics*, in *Recent developments in gauge theories*, eds. G. 't Hooft et al., Plenum, 1980, p.403.
- [64] S. Coleman, *$1/N$* , in Proc. of Erice Int. School of Subnuclear Physics 1979, Plenum, N.Y., 1982, p. 805 (Reprinted in S. Coleman, *Aspects of symmetry*, Cambridge Univ. Press, 1985, pp.351-402).
- [65] O. Haan, *Large N as a thermodynamic limit*, Phys. Lett. **106B** (1981) 207.

- [66] D.V. Voiculescu, K.J. Dykema and A. Nica, *Free Random Variables*, AMS, Providence 1992.
- [67] M.R. Douglas, *Stochastic master fields*, Phys. Lett. **B344** (1995) 117.
- [68] R. Gopakumar and G.J. Gross, *Mastering the master field*, Nucl. Phys. **B451** (1995) 379.
- [69] S. Mandelstam, *Charge-monopole duality and the phases of non-Abelian gauge theories*, Phys. Rev. **D19** (1979) 2391.
- [70] Yu.M. Makeenko and A.A. Migdal, *Exact equation for the loop average in multicolor QCD*, Phys. Lett. **88B** (1979) 135.
- [71] Yu.M. Makeenko and A.A. Migdal, *Quantum chromodynamics as dynamics of loops*, Nucl. Phys. **B188** (1981) 269 (Russian version: Yad. Fiz. **32** (1980) 838).
- [72] R.A. Brandt, F. Neri and D. Zwanziger, *Lorentz invariance from classical particle paths in quantum field theory of electric and magnetic charge*, Phys. Rev. **D19** (1979) 1153.
- [73] J.L. Gervais and A. Neveu, *The slope of the leading Regge trajectory in quantum chromodynamics*, Nucl. Phys. **B163** (1980) 189.
- [74] A.M. Polyakov, *Gauge fields as rings of glue*, Nucl. Phys. **B164** (1980) 171.
- [75] V.S. Dotsenko and S.N. Vergeles, *Renormalizability of phase factors in non-Abelian gauge theory*, Nucl. Phys. **B169** (1980) 527.
- [76] R.A. Brandt, F. Neri and M. Sato, *Renormalization of loop functions for all loops*, Phys. Rev. **D24** (1981) 879.
- [77] R.A. Brandt, A. Gocksch, M. Sato, and F. Neri, *Loop space*, Phys. Rev. **D26** (1982) 3611.
- [78] A.A. Migdal, *Loop equations and $1/N$ expansion*, Phys. Rep. **102** (1983) 199.
- [79] J.L. Gervais and A. Neveu, *Local harmonicity of the Wilson loop integral in classical Yang Mills theory*, Nucl. Phys. **B153** (1979) 445.
- [80] P. Lévy, *Problèmes concrets d'analyse fonctionnelle*, Paris, 1951 (Russian transl.: Nauka, M., 1967).
- [81] M.B. Halpern and Yu.M. Makeenko, *Continuum-regularized loop-space equation*, Phys. Lett. **218B** (1989) 230.
- [82] Yu.M. Makeenko, *Polygon discretization of the loop-space equation*, Phys. Lett. **212B** (1988) 221.
- [83] Yu.M. Makeenko and A.A. Migdal, *Self-consistent area law QCD*, Phys. Lett. **97B** (1980) 235.
- [84] J.L. Gervais and A. Neveu, *The quantum dual string wave functional in Yang Mills theories*, Phys. Lett. **80B** (1979) 255.
- [85] Y. Nambu, *QCD and the string model*, Phys. Lett. **80B** (1979) 372.
- [86] A.M. Polyakov, *String representations and hidden symmetries for gauge fields*, Phys. Lett. **82B** (1979) 247.
- [87] A.A. Migdal, *QCD = Fermi string theory*, Nucl. Phys. **189** (1981) 253.
- [88] T. Eguchi and H. Kawai, *Reduction of dynamical degrees of freedom in the large- N gauge theory*, Phys. Rev. Lett. **48** (1982) 1063.
- [89] S.R. Das, *Some aspects of large- N theories*, Rev. Mod. Phys. **59** (1987) 235.