Non-perturbative Methods in Gauge Theory

A Set of Lectures

by

Y. M. Makeenko[*]

Samizdat, 1995

[*]Institute of Theoretical and Experimental Physics, Moscow, Russia and The Niels Bohr Institute, Copenhagen, Denmark

He reached a much higher plane of creativity when he blacked out everything but *a*, *an* and *the*. That erected more dynamic intralinear tensions.

J. Heller, *Catch-22*

Preface

These lecture notes are based on courses given at:

1) Autonoma University of Madrid, Winter Semester of 1993;
2) Leipzig University, Winter Semester of 1995;
3) Moscow Physical and Technical Institute, Spring Semester of 1995.

My intention was to introduce graduate and Ph.D. students to the methods of contemporary quantum field theory. The term "non-perturbative" in the title means literally "beyond the scope of perturbation theory". Therefore, it is assumed that the reader is familiar with quantum mechanics as well as with the standard methods of perturbative expansion in quantum field theory and, in particular, with the theory of renormalization.

The second purpose was to make the course useful for senior people (including those working in condensed-matter theory), as a survey of ideas, terminology and methods, which were developed in quantum field theory in the seventies and the beginning of the eighties. For this reason, these notes do not go very deeply into details, so the presentation is sometimes a bit schematic. Correspondingly, the subjects which are usually covered by modern courses in string theory, such as the two-dimensional conformal field theories, are not touched. It is assumed that such a course will follow this one.

The main body of the lecture notes deals with lattice gauge theories and large-N methods. These two Chapters are preceded by Chapter 1 which is devoted to the method of path integrals. The path-integral approach is loosely used in quantum field theory and statistical mechanics. In Chapter 1, I shall pay most attention to aspects of the path integrals, which are then used in the next two Chapters.

In each Chapter, I was going to be as closed to the original papers, where the involved methods were proposed, as possible. The list of these papers respectively includes:

1. R.P. Feynman, *An operator calculus having applications in quantum electrodynamics*, Phys. Rev. **84** (1951) 108.
2. K.G. Wilson, *Confinement of quarks*, Phys. Rev. **D10** (1974) 2445.
3. G.'t Hooft, *A planar diagram theory for strong interactions*, Nucl. Phys. **B72** (1974) 461.

The lectures were followed by seminars where some problems for deeper studies had been solved on a blackboard. They are inserted in the text as the problems, which can be omitted at first reading. Some more information is also added as remarks after the main text. Both of them contain some relevant references.

The references, which are collected at the end of each Chapter, are usually given only to either a first paper (or papers) in a series or those containing a pedagogical presentation of the material. With the modern electronic database at `qspires` (SLAC), a list of subsequent papers can, in most cases, be retrieved by downloading citations of the first paper.

I would like to thank the students for their attention, patience, and questions. I am indebted to Martin Gürtler for his help in preparing these lecture notes.

1995–1996 *Y. M.*

Contents

List of Figures

"Yossarian? What kind of a name is Yossarian?"
He had the facts at his finger tips. "It's Yossarian's name," he explained.

J. Heller, *Catch-22*

Chapter 1

Path Integrals

The path integral is a method of quantization which is equivalent to the operator formalism. It recovers the operator formalism in quantum mechanics and perturbation theory in quantum field theory (QFT).

The approach based on path integrals has several advantages over the operator formalism. It provides a useful tool for non-perturbative studies including:

- instantons,
- analogy with statistical mechanics,
- numerical methods.

A standard way of deriving the path integral is from the operator formalism:

$$\boxed{\text{operator formalism}} \quad \Longleftrightarrow \quad \boxed{\text{path integral}} \ .$$

We shall proceed in the opposite direction, following the original paper by Feynman [Fey51].

1.1 Operator calculus

The operator calculus developed by Feynman [Fey51] makes it possible to represent functions of (non-commuting) operators as path integrals, with the integrand being the path-ordered exponential of operators whose order is controlled by a parameter which varies along the trajectory. This procedure is termed as the Feynman disentangling. It is applicable also to functions of matrices (say, γ-matrices which are associated with a spinor particle). When applied to the evolution operator, this procedure results in the standard path-integral representation of quantum mechanics.

We first demonstrate in this Section the general technique using the simplest example: a free propagator in Euclidean space, and then consider the path-integral representation of quantum mechanics, as well as propagators in an external electromagnetic field.

1.1.1 Free propagator

Let us first consider the simplest propagator of a free scalar field which is given in the operator formalism by the vacuum expectation value of the T–product[1]

$$G(x-y) \;=\; \langle 0|\mathbf{T}\,\varphi(x)\,\varphi(y)\,|0\rangle \tag{1.1.1}$$

with φ being the field-operator.

The T–product (1.1.1) obeys the equation

$$\left(-\partial^2 - m^2\right) G(x-y) \;=\; i\,\delta^{(d)}(x-y)\,, \tag{1.1.2}$$

where $d = 4$ is the dimension of space-time, however the formulas are applicable at any value of d. In the operator formalism, Eq. (1.1.2) is a consequence of the free equations

$$\begin{aligned} \left(-\partial^2 - m^2\right)\varphi(x)\big|\,0\big\rangle &= 0\,, \\ \big\langle 0\,\big|\left(-\partial^2 - m^2\right)\varphi(x) &= 0 \end{aligned} \tag{1.1.3}$$

and canonical equal-time commutators

$$\begin{aligned} \left[\varphi(t,\vec{x})\,,\dot{\varphi}(t,\vec{y})\right] &= i\,\delta^{(d-1)}(\vec{x}-\vec{y})\,, \\ \left[\varphi(t,\vec{x})\,,\varphi(t,\vec{y})\right] &= 0. \end{aligned} \tag{1.1.4}$$

[1]The ordered products of operators were introduced by Dyson [Dys49]. This paper and other classical papers on quantum electrodynamics are collected in the book edited by Schwinger [Sch58].

The delta-function $\delta^{(1)}(x_0 - y_0)$ emerges when $(\partial/\partial t)^2$ is applied to the operator of the T–product in (1.1.1).

Problem 1.1 Derive Eq. (1.1.2) in the operator formalism.
Solution Let us apply the operator on the left-hand side (LHS) of Eq. (1.1.2) to the T–product which is defined by

$$\mathbf{T}\,\varphi(x)\,\varphi(y) \;=\; \theta(x_0-y_0)\varphi(x)\,\varphi(y) + \theta(y_0-x_0)\varphi(y)\,\varphi(x) \tag{1.1.5}$$

with

$$\theta(x_0-y_0) \;=\; \begin{cases} 1 \text{ for } x_0 \geq y_0 \\ 0 \text{ for } x_0 < y_0 \end{cases}. \tag{1.1.6}$$

Eq. (1.1.3) implies a nonvanishing result to emerge only when $(\partial/\partial x_0)^2$ is applied to the operator of the T–product. One gets

$$\begin{aligned} \left(-\partial^2 - m^2\right)\langle 0|\mathbf{T}\,\varphi(x)\,\varphi(y)\,|0\rangle &= -\frac{\partial}{\partial x_0}\langle 0|\mathbf{T}\,\dot{\varphi}(x)\,\varphi(y)\,|0\rangle \\ &= \delta^{(1)}(x_0-y_0)\,\langle 0|\left[\varphi(y)\,,\dot{\varphi}(x)\right]|0\rangle = i\,\delta^{(d)}(x-y)\,, \end{aligned} \tag{1.1.7}$$

where the canonical commutation relations (1.1.4) are used.

The explicit solution to Eq. (1.1.2) for the free propagator is well-known and is most simply given by the Fourier transform:

$$G(x-y) \;=\; \int \frac{d^dp}{(2\pi)^d}\,\mathrm{e}^{\,ip(x-y)}\,\frac{i}{p^2-m^2+i\varepsilon}. \tag{1.1.8}$$

An extra $i\varepsilon$ (with $\varepsilon \to +0$) in the denominator is due to the T–product in the definition (1.1.1) and unambiguously determines the integral over p_0. The propagator (1.1.8) is known as the Feynman propagator which respects causality.

Problem 1.2 Perform the Fourier transformation of the free momentum-space propagator in the energy p_0:

$$G_\omega(t-t') \;=\; \int\limits_{-\infty}^{+\infty} \frac{dp_0}{2\pi}\,\mathrm{e}^{\,ip_0(t-t')}\,\frac{i}{p_0^2-\omega^2+i\varepsilon}, \qquad \omega \;=\; \sqrt{\vec{p}^{\,2}+m^2}. \tag{1.1.9}$$

Solution The poles of the momentum-space propagator are at

$$p_0 \;=\; \pm\,\omega \mp i\varepsilon\,. \tag{1.1.10}$$

For $t > t'$ $(t < t')$, the contour of integration can be closed in the upper (lower) half-plane which gives

$$G_\omega(t-t') = \theta(t-t')\frac{e^{-i\omega(t-t')}}{2\omega} + \theta(t'-t)\frac{e^{i\omega(t-t')}}{2\omega} = \frac{e^{-i\omega|t-t'|}}{2\omega}. \tag{1.1.11}$$

The Green function (1.1.11) obeys the equation

$$\left(-\frac{\partial^2}{\partial t^2} - \omega^2\right) G_\omega\left(t-t'\right) = i\,\delta^{(1)}\left(t-t'\right) \tag{1.1.12}$$

and therefore coincides with the causal Green function for a harmonic oscillator with frequency ω.

Remark on operator notations

In mathematical language, the Green function $G(x-y)$ is termed as the *resolvent* of the operator on the LHS of Eq. (1.1.2), and is often denoted as the matrix element of the inverse operator

$$G(x-y) = \left\langle y \left| \frac{i}{-\partial^2 - m^2} \right| x \right\rangle. \tag{1.1.13}$$

The operators act in an infinite-dimensional Hilbert space whose elements in Dirac's notation [Dir58] are the *bra* and *ket* vectors: $\langle g|$ and $|f\rangle$, respectively. The coordinate representation emerges when these vectors are chosen to be the eigenstates of the position operator $\boldsymbol{x}_\mu$:

$$\boldsymbol{x}_\mu |x\rangle = x_\mu |x\rangle. \tag{1.1.14}$$

These basis vectors obey the completeness relation

$$\int d^d x\, |x\rangle\langle x| = 1, \tag{1.1.15}$$

while the wave functions, associated with $\langle g|$ and $|f\rangle$, are given by

$$\langle g \mid x\rangle = g(x), \qquad \langle x \mid f\rangle = f(x). \tag{1.1.16}$$

These wave functions appear in the expansions

$$|f\rangle = \int d^d x\, f(x)|x\rangle, \qquad \langle g| = \int d^d y\, g(y)\langle y|. \tag{1.1.17}$$

The action of a linear operator $\boldsymbol{O}$ on the bra and ket vectors in the Hilbert space is determined by its matrix element $\langle y|\boldsymbol{O}|x\rangle$, which is also known as the *kernel* of the operator $\boldsymbol{O}$ and is denoted by

$$\langle y|\boldsymbol{O}|x\rangle = O(y,x). \tag{1.1.18}$$

Using the expansion (1.1.17), one gets

$$\langle g|\boldsymbol{O}|f\rangle = \int d^d x \int d^d y\, g(y)\, O(y,x)\, f(x). \tag{1.1.19}$$

Since the kernel of the unit operator is the delta-function,

$$\langle y|\mathbf{1}|x\rangle = \langle y|x\rangle = \delta^{(d)}(x-y), \tag{1.1.20}$$

the formula

$$\langle y|\boldsymbol{O}|x\rangle = \boldsymbol{O}\,\delta^{(d)}(x-y) \tag{1.1.21}$$

can also be written down as a direct consequence of Eq. (1.1.20), where the operator $\boldsymbol{O}$ on the right-hand side (RHS) acts on the variable x.

Therefore, when the operator acts on a function $f(x)$, the result is expressed via the kernel by the standard formula

$$(\boldsymbol{O}f)(y) \equiv \langle y|\boldsymbol{O}|f\rangle = \int d^d x\, O(y,x)\, f(x). \tag{1.1.22}$$

Eq. (1.1.21) is obviously reproduced when f is substituted by a delta-function, while Eq. (1.1.19) takes the form

$$\langle g|\boldsymbol{O}|f\rangle = \int d^d x\, g(x)\, (\boldsymbol{O}f)(x). \tag{1.1.23}$$

If space-time is approximated by a discrete set of points, then the operator $\boldsymbol{O}$ is approximated by a matrix with $\langle y|\boldsymbol{O}|x\rangle$ being its elements.

1.1.2 Euclidean formulation

Eq. (1.1.8) can be alternatively obtained by inverting the operator on the LHS of Eq. (1.1.2). Before doing that, it is convenient to make an analytic continuation in the time-variable t, and to pass to the Euclidean formulation of QFT where one substitutes

$$t = -i x_4. \tag{1.1.24}$$

The four-momentum operator in Minkowski space reads

$$\boldsymbol{p}_{\rm M}^{\mu} \;=\; i\,\partial_{\rm M}^{\mu} \;\equiv\; \left(i\frac{\partial}{\partial t},-i\frac{\partial}{\partial \vec{x}}\right) \qquad \boxed{\text{Minkowski space}}\,, \tag{1.1.25}$$

while its Euclidean counterpart is given by

$$\boldsymbol{p}_{\rm E}^{\mu} \;=\; -\,i\,\partial_{\rm E}^{\mu} \;\equiv\; \left(-i\frac{\partial}{\partial \vec{x}},-i\frac{\partial}{\partial x_4}\right) \qquad \boxed{\text{Euclidean space}}\,. \tag{1.1.26}$$

These two formulas together with Eq. (1.1.24) yield

$$E \;\equiv\; p_0 \;=\; -\,i\,p_4 \tag{1.1.27}$$

for the relation between energy and the fourth component of the Euclidean four-momentum.

The passage to Euclidean space results in changing the Minkowski signature of the metric $g_{\mu\nu}$ to the Euclidean one:[2]

$$\begin{array}{ccc} (+\,-\,-\,-) & \longrightarrow & (+\,+\,+\,+) \\ \boxed{\text{Minkowski signature}} & \longrightarrow & \boxed{\text{Euclidean signature}} \end{array}\,. \tag{1.1.28}$$

As such, one gets

$$p_{\rm M}^2 = p_0^2 - \vec{p}^{\,2} \quad\longrightarrow\quad -p_{\rm E}^2 = -\vec{p}^{\,2} - p_4^2\,. \tag{1.1.29}$$

The exponent in the Fourier transformation changes analogously:

$$-p_\mu x^\mu = -Et + \vec{p}\vec{x} \;\longrightarrow\; p_{\rm E}^\mu x_{\rm E}^\mu = \vec{p}\vec{x} + p_4 x_4\,. \tag{1.1.30}$$

This reproduces the standard Fourier transformation in Euclidean space

$$\begin{aligned} f(p) \;&=\; \int d^d x\,{\rm e}^{\,-ipx} f(x)\,, \\ f\,(x) \;&=\; \int \frac{d^d p}{(2\pi)^d}\,{\rm e}^{\,ipx} f(p)\,. \end{aligned} \tag{1.1.31}$$

[2]An older generation is familiar with Euclidean notations which are used throughout the book by Akhiezer and Berestetskii [AB69]. On the contrary, the two other canonical books on quantum field theory by Bogoliubov and Shirkov [BS76] and by Bjorken and Drell [BD65] use Minkowskian notations which are due to Feynman.

We shall use the same notation v^μ for a four-vector in Minkowski and Euclidean spaces:

$$\begin{aligned} v_{\rm M}^{\mu} \;&=\; (v_0,\vec{v}) \qquad \boxed{\text{Minkowski space}}\,, \\ v_{\rm E}^{\mu} \;&=\; (\vec{v},v_4) \qquad \boxed{\text{Euclidean space}}\,, \end{aligned} \tag{1.1.32}$$

with

$$v_0 \;=\; -\,iv_4\,. \tag{1.1.33}$$

The only difference resides in the metric. We do not distinguish between upper and lower indices in Euclidean space.

Using Eqs. (1.1.24) and (1.1.26), we see that in Euclidean space Eq. (1.1.2) takes the form

$$\left(-\partial^2 + m^2\right) G\,(x-y) \;=\; \delta^{(d)}\,(x-y) \tag{1.1.34}$$

with the positive sign in front of m^2.

The passage to the Euclidean formulation is justified in perturbation theory where it is associated with the Wick rotation. The direction in which the rotation is performed is unambiguously prescribed by the $+i\varepsilon$ term in Eq. (1.1.8), and is depicted in Fig. 1.1. The variable $t = x_0$ rotates through $-\pi/2$ while $E = p_0$ rotates through $\pi/2$.

Figure 1.1a explains the sign in Eq. (1.1.24). Figure 1.1b and Eq. (1.1.27) implies that the integration over p_4 goes in the opposite direction, so that

$$\int\limits_{-\infty}^{+\infty} \frac{dp_0}{2\pi}\cdots \;=\; i\int\limits_{-\infty}^{+\infty}\frac{dp_4}{2\pi}\cdots\,. \tag{1.1.35}$$

Thus when passing in to Euclidean variables, Eq. (1.1.8) becomes

$$G\,(x-y) \;=\; \int \frac{d^d p}{(2\pi)^d}\,{\rm e}^{\,ip(y-x)}\frac{1}{p^2+m^2}\,. \tag{1.1.36}$$

Note that the RHS of Eq. (1.1.36) is nothing but the Fourier transform of the free momentum-space Euclidean propagator, and there is no need to retain an $i\varepsilon$ in the denominator since the integration prescription is now unambiguous.

It is now clear why we keep the same notation for the coordinate-space Green functions: the Feynman propagator in Minkowski space and the Euclidean propagator. They are the same analytic function of the time-variable.

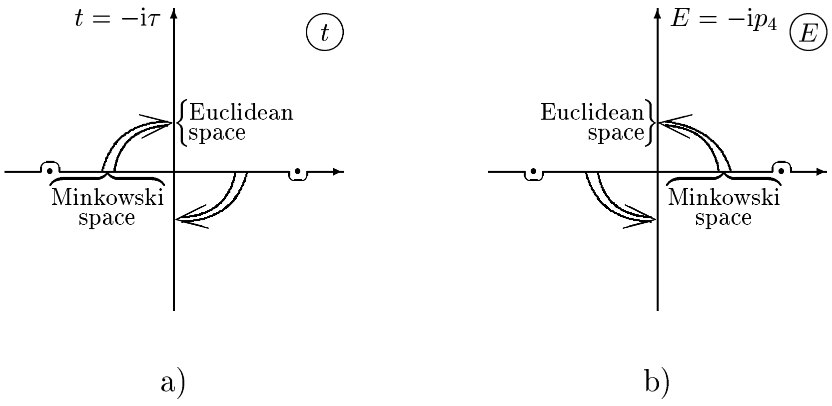


Fig. 1.1: Direction of the Wick rotation from Minkowski to Euclidean space (indicated by the arrows) for a) time and b) energy. The dots represent singularities of a free propagator in a) coordinate and b) momentum spaces. The contours of integration in Minkowski space are associated with causal Green functions. They can obviously be deformed in the directions of the arrows.

Problem 1.3 Repeat the calculation of Problem 1.2 in Euclidean space.
Solution According to Eq. (1.1.36) we need to calculate

$$G_\omega(\tau-\tau') = \int_{-\infty}^{+\infty} \frac{dp_4}{2\pi}\, \mathrm{e}^{\,ip_4(\tau'-\tau)} \frac{1}{p_4^2+\omega^2}\,. \tag{1.1.37}$$

The integral on the RHS can be calculated for $\tau > \tau'$ ($\tau < \tau'$) by closing the contour in the lower (upper) half-plane, and taking the residues at $p_4 = -i\omega$ ($p_4 = i\omega$), respectively. This yields

$$\begin{aligned} G_\omega(\tau-\tau') &= \theta(\tau-\tau')\frac{\mathrm{e}^{\,\omega(\tau'-\tau)}}{2\omega} + \theta(\tau'-\tau)\frac{\mathrm{e}^{\,\omega(\tau-\tau')}}{2\omega} \\ &= \frac{\mathrm{e}^{\,-\omega|\tau-\tau'|}}{2\omega}\,. \end{aligned} \tag{1.1.38}$$

The Euclidean Green function (1.1.38) can obviously be obtained from the Minkowskian one, Eq. (1.1.11), by the substitution

$$\tau = it\,, \qquad \tau' = it' \tag{1.1.39}$$

and vise versa. $G_\omega(\tau-\tau')$ obeys the equation

$$\left(-\frac{\partial^2}{\partial\tau^2}+\omega^2\right) G_\omega\left(\tau-\tau'\right) = \delta^{(1)}\left(\tau-\tau'\right) \tag{1.1.40}$$

and, therefore, is the Green function for a Euclidean harmonic oscillator with frequency ω.

As we shall see in a moment, the Euclidean formulation makes path integrals well-defined, and allows non-perturbative investigations analogous to statistical mechanics to be carried out. There are no reasons, however, why Minkowski and Euclidean formulations should always be equivalent non-perturbatively.

Remark on Euclidean γ-matrices

The γ-matrices in Minkowski space satisfy

$$\{\gamma_{\mathrm{M}}^\mu, \gamma_{\mathrm{M}}^\nu\} = 2\, g^{\mu\nu}\, \mathrm{I}\,. \tag{1.1.41}$$

Therefore, γ_0 is Hermitean while the Minkowskian spatial γ-matrices are anti-Hermitean.

Analogously, the Euclidean γ-matrices satisfy

$$\{\gamma_\mu, \gamma_\nu\} = 2\, \delta_{\mu\nu}\, \mathrm{I}\,, \tag{1.1.42}$$

so that all of them are Hermitean. We compose them from 2×2 matrices as

$$\gamma_4 = \gamma_0 = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} \tag{1.1.43}$$

and

$$\vec{\gamma} = \begin{pmatrix} 0 & -i\vec{\sigma} \\ i\vec{\sigma} & 0 \end{pmatrix}, \tag{1.1.44}$$

where $\vec{\sigma}$ are the usual Pauli matrices. Notice that the Euclidean spatial γ-matrices differ from the Minkowskian ones by a factor of i.

The free Dirac equation in Euclidean space reads

$$\left(\hat{\partial}+m\right)\psi = 0\,, \qquad \hat{\partial} = \gamma_\mu\partial_\mu \tag{1.1.45}$$

or

$$(i\widehat{\boldsymbol{p}}+m)\,\psi = 0 \tag{1.1.46}$$

with $\boldsymbol{p}$ given by Eq. (1.1.26).

1.1.3 Path-ordering of operators

There are no problems in defining a function of an operator A, say via the Taylor series. For instance,

$$e^{A} = \sum_{n=0}^{\infty} \frac{1}{n!} A^n \,. \tag{1.1.47}$$

However, it is more complicated to define a function of several non-commuting operators (or matrices), *e.g.* A and B having

$$[A, B] \neq 0\,, \tag{1.1.48}$$

since the order of operators is now essential. In particular, one has

$$e^{A+B} \neq e^{A} e^{B}\,, \tag{1.1.49}$$

so that the law of addition of exponents fail. Certainly, the exponential on the LHS is a well-defined function of $A+B$, but since A and B are intermixed in the Taylor expansion, this expansion is of little use in practice. We would like to have an expression where all B's are written, say, to the right of all A's.

This can be achieved by the following formal trick [Fey51].

Let us write

$$e^{A+B} = \lim_{M\to\infty} \left[1 + \frac{1}{M}(A+B)\right]^M = \lim_{M\to\infty} \underbrace{\left[1 + \frac{1}{M}(A+B)\right]\left[1 + \frac{1}{M}(A+B)\right]\cdots\left[1 + \frac{1}{M}(A+B)\right]}_{M \text{ times}}. \tag{1.1.50}$$

The structure of the product on the RHS prompts us to introduce an index i running from 1 to M and replace $(A+B)$ in each multiplier by $(A_i + B_i)$. Therefore, one writes

$$\begin{aligned} e^{A+B} &= \lim_{M\to\infty} \prod_{i=1}^{M} \left[1 + \frac{1}{M}(A_i + B_i)\right] \\ &= \lim_{M\to\infty} \left[1 + \frac{1}{M}(A_M + B_M)\right]\cdots\left[1 + \frac{1}{M}(A_1 + B_1)\right], \end{aligned} \tag{1.1.51}$$

where the index i controls the order of the operators which are all treated *different*. The ordering is such that the larger i is, the later the operator with the index i acts. This order of operators is prescribed by quantum mechanics, where initial and final states are represented by ket and bra vectors, respectively.

The RHS of Eq. (1.1.51) can be rewritten as

$$e^{A+B} = \mathbf{P} \lim_{M\to\infty} e^{\frac{1}{M}\sum_{i=1}^{M}(A_i+B_i)} \tag{1.1.52}$$

where the symbol $\mathbf{P}$ stands for the ordering operation. There is no ambiguity on the RHS of Eq. (1.1.52) concerning ordering A_i and B_i with the same index i, since such terms are $\mathcal{O}(1/M^2)$ and are negligible as $M \to \infty$.

To describe the continuum limit as $M \to \infty$, one introduces the continuum variable $\sigma = i/M$ which belongs to the interval $[0, 1]$. The continuum limit of Eq. (1.1.52) reads

$$e^{A+B} = \mathbf{P}\, e^{\int_0^1 d\sigma (A(\sigma)+B(\sigma))}, \tag{1.1.53}$$

where $A(i/M) = A_i$ and $B(i/M) = B_i$ while the operator $A(\sigma) + B(\sigma)$ acts at order σ.

Eq. (1.1.53) is in fact obvious since it only involves the operator $A + B$ which commutes with itself. For commuting operators there is no need for ordering so that $A(\sigma) + B(\sigma)$ does not depend on σ in this case. The integral in the exponent on the RHS of Eq. (1.1.53) can then be done, and reproduces the LHS.

Eq. (1.1.53) can however be manipulated as though $A(\sigma)$ and $B(\sigma)$ were just functions rather than operators since the order would be automatically specified by the path-ordering operation. This is analogous to the well-known fact that operators can be written in an arbitrary order under the T–product. Therefore, we can rewrite Eq. (1.1.53) as

$$e^{A+B} = \mathbf{P}\, e^{\int_0^1 d\sigma' A(\sigma')}\, e^{\int_0^1 d\sigma B(\sigma)}. \tag{1.1.54}$$

This is the operator analog of the law of addition of exponents.

Problem 1.4 Calculate explicitly the first term of the expansion of $\exp(A+B)$ in B.

Solution Expanding the RHS of Eq. (1.1.54) in B, one gets

$$e^{A+B} = e^{A} + \int_0^1 d\sigma\, e^{\int_\sigma^1 d\sigma' A(\sigma')} B(\sigma)\, e^{\int_0^\sigma d\sigma' A(\sigma')} + \ldots \tag{1.1.55}$$

There is no need for a path-ordering sign in this formula, since the order of the operators A and B is written explicitly. There is also no ambiguity in defining the exponentials of the operator A as already explained.

Since the order is explicit, one drops the formal dependence of A and B on the ordering parameter which gives

$$\mathrm{e}^{A+B} = \mathrm{e}^{A} + \int_0^1 d\sigma\, \mathrm{e}^{(1-\sigma)A} B\, \mathrm{e}^{\sigma A} + \ldots \tag{1.1.56}$$

The formulas (1.1.55) and (1.1.56) are known from time-dependent perturbation theory in quantum mechanics.

1.1.4 Feynman disentangling

The operator on the LHS of Eq. (1.1.34) can be inverted as follows:

$$\begin{aligned} G(x-y) &= \frac{1}{-\partial^2+m^2}\delta^{(d)}(x-y) \\ &= \frac{1}{2}\int_0^\infty d\tau\ \mathrm{e}^{\frac{1}{2}\tau\left(\partial^2-m^2\right)}\delta^{(d)}(x-y) \\ &= \frac{1}{2}\int_0^\infty d\tau\, \mathrm{e}^{-\frac{1}{2}m^2\tau}\mathbf{P}\,\mathrm{e}^{\frac{1}{2}\int_0^\tau dt\partial^2(t)}\delta^{(d)}(x-y)\,, \end{aligned} \tag{1.1.57}$$

where we have formally labelled the derivatives by an ordering parameter $t \in [0,\tau]$, which is an analog of σ from the previous subsection. This is the general procedure which the Feynman disentangling is built on.

Since the operators ∂_μ and ∂_ν commute in the free case, we could manage without introducing the t-dependence, however the operators do not commute in general. The simple example of the non-relativistic Hamiltonian and the propagator in an external electromagnetic field are considered later on in this Section. Other cases where the disentangling is needed are related to inverting an operator which is in addition a matrix in some symmetry space.

Continuing with the disentangling, the RHS of Eq. (1.1.57) can be rewritten as

$$\begin{aligned} G(x-y) &= \frac{1}{2}\int_0^\infty d\tau\, \mathrm{e}^{-\frac{1}{2}m^2\tau}\int_{z_\mu(0)=x_\mu} Dz_\mu(t)\ \mathrm{e}^{-\frac{1}{2}\int_0^\tau dt\,\dot{z}_\mu^2(t)} \\ &\quad \cdot \mathbf{P}\,\mathrm{e}^{\int_0^\tau dt\,\dot{z}_\mu(t)\partial_\mu(t)}\ \delta^{(d)}(x-y)\,, \end{aligned} \tag{1.1.58}$$

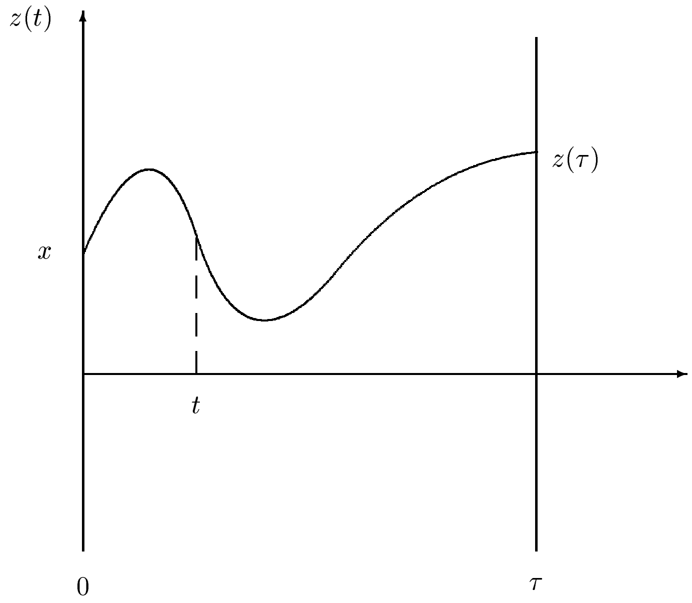


Fig. 1.2: The trajectory $z_\mu(t)$. The operator $\partial_\mu(t)$ acts at the order t.

where the integration runs over all trajectories $z_\mu(t)$ which begin at the point x, as depicted in Fig. 1.2.

Since the operator $\partial_\mu(t)$ acts at the order t, these operators are ordered along the trajectory $z_\mu(t)$ with $\mathbf{P}$, in Eq. (1.1.58), standing for the path-ordering operator. Note, that $\dot{z}_\nu(t)$ and $\partial_\mu(t)$ commute since

$$\partial_\mu(t)\dot{z}_\nu(t) = \frac{d}{dt}\delta_{\mu\nu} = 0 \tag{1.1.59}$$

so that their order is inessential in Eq. (1.1.58). With these rules of manipulations, Eq. (1.1.58) can be proven by the "translation"

$$z_\mu(t) \rightarrow z'_\mu(t) = z_\mu(t) + \int_0^t dt'\,\partial_\mu(t') \tag{1.1.60}$$

of the integration variable $z_\mu(t)$ in the Gaussian integral.

The integral over the functions $z_\mu(t)$ in Eq. (1.1.58) is called a *path integral* or a *functional integral.*

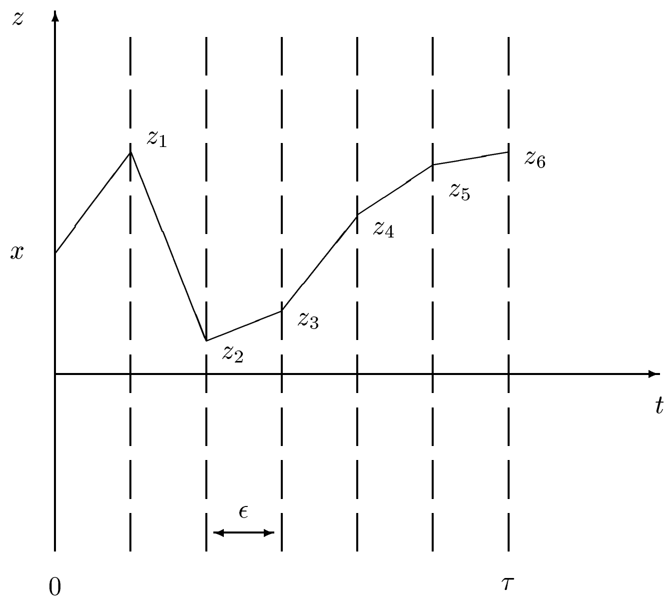


Fig. 1.3: Discretization of the trajectory $z_\mu(t)$ (depicted for $M = 6$).

The continual path integral can be approximated by a finite one. To this aim, let us choose M points $t_i = i\epsilon$, where ϵ is the step of discretization, and $M = \tau/\epsilon$. We then connect the points

$$z_0 \;=\; x\,, \quad z_i \;=\; z(i\epsilon) \quad i = 1, 2, \ldots, M \tag{1.1.61}$$

by straight lines. Such a discretization of the trajectory $z_\mu(t)$ is depicted in Fig. 1.3. The measure in Eq. (1.1.58) can then be discretized by

$$\int Dz_\mu\,(t) \ldots \;=\; \prod_{i=1}^{M} \int \frac{d^d z_i}{(2\pi\epsilon)^{d/2}} \cdots\,. \tag{1.1.62}$$

The explicit form of the operator ∂_μ in Eq. (1.1.34) was not essential in deriving Eq. (1.1.58). If ∂_μ in Eq. (1.1.34) is replaced by an arbitrary operator D_μ with non-commuting components, then Eq. (1.1.58) holds with $\partial_\mu(t)$ substituted by $D_\mu(t)$. The discretized path-ordered exponential of a general operator $D_\mu(t)$ is given by

$$\mathbf{P}\,\mathrm{e}^{\int_0^\tau dt\, \dot z^\mu(t) D_\mu(t)} \;=\; \lim_{\epsilon\to 0} \prod_{i=1}^{M} \left[1 + (z_i - z_{i-1})^\mu\, D_\mu(i\epsilon)\right]. \tag{1.1.63}$$

The order of multiplication here is the same as in Eq. (1.1.51).

The explicit form of the operator ∂_μ is essential when we calculate how it acts on the delta-function as is prescribed by the RHS of Eq. (1.1.58). For the free case, when the t-dependence of $\partial_\mu(t)$ is inessential, one simply gets

$$\mathbf{P}\,\mathrm{e}^{\int_0^\tau dt\, \dot z^\mu(t) \partial_\mu(t)} \;=\; \exp\left[\left(z^\mu(\tau) - x^\mu\right) \frac{\partial}{\partial x^\mu}\right], \tag{1.1.64}$$

which is nothing but the shift operator. Applying the operator on the RHS of Eq. (1.1.64) to the delta-function, one gets

$$\mathbf{P}\,\mathrm{e}^{\int_0^\tau dt\, \dot z^\mu(t) \partial_\mu(t)}\, \delta^{(d)}\,(x-y) \;=\; \delta^{(d)}\,(z\,(\tau) - y)\,. \tag{1.1.65}$$

Therefore, $z_\mu(\tau)$ has to coincide with y_μ due to the delta-function, which disappears after the integration over $z_\mu(\tau)$ has been performed. Thus the final answer is

$$G\,(x-y) \;=\; \frac{1}{2} \int_0^\infty d\tau\; \mathrm{e}^{-\frac{1}{2}\tau m^2} \int\limits_{\substack{z_\mu(0)=x_\mu \\ z_\mu(\tau)=y_\mu}} Dz_\mu\,(t)\;\; \mathrm{e}^{-\frac{1}{2}\int_0^\tau dt\, \dot z_\mu^2(t)}. \tag{1.1.66}$$

This path integral goes over all trajectories $z_\mu(t)$ which connect the initial point x_μ and the final point y_μ.

Problem 1.5 Derive Eqs. (1.1.58) and (1.1.66) by introducing a path integral over velocity $v_\mu(t) = \dot z_\mu(t)$.
Solution The operator on the RHS of Eq. (1.1.57) can be disentangled by the following Gaussian path integral

$$\mathbf{P}\;\mathrm{e}^{\frac{1}{2}\int_0^\tau dt D_\mu^2(t)} = \int Dv_\mu(t)\,\mathrm{e}^{-\frac{1}{2}\int_0^\tau dt v_\mu^2(t)}\;\mathbf{P}\;\mathrm{e}^{\int_0^\tau dt v^\mu(t) D_\mu(t)}. \tag{1.1.67}$$

This formula holds for an arbitrary operator D_μ and can be proven by calculating the Gaussian integral after shifting $v_\mu(t)$.

Substituting $D_\mu(t) = \partial_\mu(t)$ and calculating the action of the path-ordered exponential on $\delta^{(d)}(x-y)$, we get

$$G(x-y) = \frac{1}{2}\int_0^\infty d\tau\, \mathrm{e}^{-\frac{1}{2}\tau m^2} \int Dv_\mu(t)\, \mathrm{e}^{-\frac{1}{2}\int_0^\tau dt v_\mu^2(t)} \cdot \delta^{(d)}(x + \int_0^\tau dt\, v(t) - y). \tag{1.1.68}$$

The integration over $Dv_\mu(t)$ in this formula has no restrictions.

To derive Eq. (1.1.66) from Eq. (1.1.68), let us note that the discretized velocities read

$$v_i^\mu = \frac{z_i^\mu - z_{i-1}^\mu}{\epsilon}. \tag{1.1.69}$$

Since

$$\int_0^\tau dt\, v^2(t) \to \epsilon \sum_{i=1}^M v_i^2\,, \tag{1.1.70}$$

the measure

$$\int Dv_\mu(t)\ldots = \prod_{i=1}^M \int \frac{d^d v_i}{(2\pi/\epsilon)^{d/2}}\cdots \tag{1.1.71}$$

obviously recovers Eq. (1.1.62) after calculating the Jacobian from the variables v_i to the variables z_i. Therefore, Eq. (1.1.68) reproduces Eq. (1.1.66) provided

$$z_\mu(t) = x_\mu + \int_0^t dt' v_\mu(t')\,. \tag{1.1.72}$$

Remark on definition of the measure

The discretized trajectory in Fig. 1.3 can be analytically written as the expansion

$$z^\mu(t) = \sum_{i=1}^M z_i^\mu f_i(t) + x^\mu\,(1 - t/\epsilon)\,\theta(\epsilon - t)\,, \tag{1.1.73}$$

where the basis functions

$$f_i(t) = \begin{cases} 1 + (t/\epsilon - i) & \text{for } t \in [(i-1)\epsilon, i\epsilon]\,, \\ 1 - (t/\epsilon - i) & \text{for } t \in [i\epsilon, (i+1)\epsilon]\,, \\ 0 & \text{otherwise} \end{cases} \tag{1.1.74}$$

are nonvanishing only for the i-th and $(i+1)$-th intervals. The measure (1.1.62) is defined, therefore, via the coefficients z_i as a multiple product of dz_i's.

While the basis functions $f_i(t)$ are not orthogonal:

$$\frac{1}{\epsilon}\int_0^\tau dt\, f_i(t) f_j(t) = \frac{2}{3}\delta_{ij} + \frac{1}{6}\delta_{i(j+1)} + \frac{1}{6}\delta_{i(j-1)}\,, \tag{1.1.75}$$

the orthogonal set appears in the expansion of the velocity

$$\dot{z}^\mu(t) = \sum_{i=1}^M (z_i^\mu - z_{i-1}^\mu)\phi_i(t)\,, \tag{1.1.76}$$

where

$$\phi_i(t) = \begin{cases} 1/\epsilon & \text{for } t \in [(i-1)\epsilon, i\epsilon]\,, \\ 0 & \text{otherwise} \end{cases}\,. \tag{1.1.77}$$

This shows why the discretized velocities from Problem 1.5 are natural variables.

One can choose, instead, another set of (orthogonal) basis functions and expand

$$z^\mu(t) = \sum_{n=1}^M c_n^\mu \phi_n(t) \tag{1.1.78}$$

with some coefficients c_n^μ. Then the measure (1.1.62) takes the form

$$Dz^\mu(t)\ldots \propto \prod_{n=1}^M d^d c_n \ldots \tag{1.1.79}$$

modulo a c-independent Jacobian. Mathematically, this implies that one approximates the functional space by M-dimensional spaces.

1.1.5 Calculation of the Gaussian path integral

The Gaussian path integral (1.1.66) can be easily calculated by the following trick (see, *e.g.*, the book by Feynman [Fey72], Chapter 3). Let us substitute the variable $z(t)$ by a new variable $\xi(t)$ which are related by the formula

$$z\,(t) = \frac{y - x}{\tau}\, t + \xi\,(t) + x. \tag{1.1.80}$$

The boundary conditions for the variable $\xi(t)$ are determined by Eq. (1.1.80) to be

$$\xi(0) = \xi(\tau) = 0. \tag{1.1.81}$$

Substituting Eq. (1.1.80) into the exponent in Eq. (1.1.66), one gets

$$\int_0^\tau dt\, \dot z^2(t) = \frac{(y-x)^2}{\tau} + 2\frac{(y-x)}{\tau}\left(\xi(\tau)-\xi(0)\right) + \int_0^\tau dt\, \dot\xi^2(t). \tag{1.1.82}$$

The second term on the RHS vanishes due to the boundary conditions (1.1.81) so that the propagator becomes

$$G(x-y) = \frac{1}{2}\int_0^\infty d\tau\, \mathrm{e}^{-\frac{1}{2}\tau m^2}\, \mathrm{e}^{-\frac{(y-x)^2}{2\tau}} \int\limits_{\xi(0)=\xi(\tau)=0} D\xi\, \mathrm{e}^{-\frac{1}{2}\int_0^\tau dt\, \dot\xi^2(t)}. \tag{1.1.83}$$

The path integral over ξ on the RHS of Eq. (1.1.83) is a function solely of τ:

$$\int\limits_{\xi(0)=\xi(\tau)=0} D\xi\, \mathrm{e}^{-\frac{1}{2}\int_0^\tau dt\, \dot\xi^2(t)} = \mathcal{F}(\tau). \tag{1.1.84}$$

This expression is to be compared with the proper-time representation of the Euclidean free propagator which reads

$$\begin{aligned} G(x-y) &= \int \frac{d^d p}{(2\pi)^d}\, \mathrm{e}^{ip(x-y)} \frac{1}{2}\int_0^\infty d\tau\, \mathrm{e}^{-\frac{\tau}{2}(p^2+m^2)} \\ &= \frac{1}{2}\int_0^\infty d\tau\, \mathrm{e}^{-\frac{1}{2}\tau m^2}\, \mathrm{e}^{-\frac{(x-y)^2}{2\tau}} \frac{1}{(2\pi\tau)^{d/2}}. \end{aligned} \tag{1.1.85}$$

These two expressions coincide provided that

$$\mathcal{F}(\tau) = \frac{1}{(2\pi\tau)^{d/2}}. \tag{1.1.86}$$

Problem 1.6 Calculate $\mathcal{F}(\tau)$ from the discretized path integral.
Solution The discretized version of Eq. (1.1.66) is

$$\begin{aligned} &\int\limits_{\substack{z_\mu(0)=x_\mu \\ z_\mu(\tau)=y_\mu}} Dz_\mu(t)\, \mathrm{e}^{-\frac{1}{2}\int_0^\tau dt\, \dot z_\mu^2(t)} \\ &= \frac{1}{(2\pi\epsilon)^{d/2}} \int \prod_{i=1}^{M-1} \frac{d^d z_i}{(2\pi\epsilon)^{d/2}}\, \mathrm{e}^{-\frac{1}{2}\sum_{i=1}^M (z_i - z_{i-1})^2/\epsilon} \end{aligned} \tag{1.1.87}$$

where $z_0 = x$ and $z_M = y$. The integral can be calculated by well-known formula for the Gaussian integral

$$\int \frac{d^d z}{(2\pi)^{d/2}}\, \mathrm{e}^{-\frac{(x-z)^2}{2\tau_1} - \frac{(z-y)^2}{2\tau_2}} = \left(\frac{\tau_1\tau_2}{\tau_1+\tau_2}\right)^{d/2} \mathrm{e}^{-\frac{(x-y)^2}{2(\tau_1+\tau_2)}}. \tag{1.1.88}$$

After applying this formula $M-1$ times, one arrives at Eq. (1.1.86). Notice that ϵ cancels in the final answer.

Problem 1.7 What trajectories are essential in the path integral?
Solution It is seen from the discretization on the RHS of Eq. (1.1.87) that only trajectories with

$$|z_i - z_{i-1}| \sim \sqrt{\epsilon} \tag{1.1.89}$$

are essential as $\epsilon \to 0$. Such trajectories are typical *Brownian* trajectories. They are continuous as $\epsilon \to 0$ but not smooth ($|z_i - z_{i-1}| \sim \epsilon$ for smooth trajectories). In mathematical language, these functions are said to belong to the Lipshitz class 1/2.

Remark on mathematical structure

The measure (1.1.62) for integration over functions is sometimes called the Lebesgue measure. It was introduced in mathematics by Wiener [Wie23] in connection with the problem of Brownian motion. With the Gaussian factor incorporated, it is also known as the Wiener measure while the proper path integral is known as the Wiener integral[3]. The measure (1.1.62) is defined on the space L_2 (*i.e.* the space of functions whose square is integrable, in the sense of the Lebesgue integral, $\int dt z^2(t) < \infty$). The integration on L_2 goes over trajectories $z_\mu(t)$, which are generically discontinuous. However, the extra weight factor $\exp\{-\frac{1}{2}\int_0^\tau \dot z^2(t)\}$ restricts the trajectories in the above path integrals to be continuous.

[3]See, *e.g.*, the books [Kac59, Sch81, Wie86] for a description of the path-integral approach to Brownian motion.

1.1.6 Transition amplitudes

As is well-known in quantum mechanics, $G(x-y)$ is the probability for a (scalar) particle to propagate from x to y. A convenient notation for a trajectory $z_\mu(t)$ that connects x_μ and y_μ is

$$\Gamma_{yx} \equiv \{z_\mu(t);\ 0 \le t \le \tau, \quad z_\mu(0)=x_\mu,\, z_\mu(\tau)=y_\mu\}\,. \tag{1.1.90}$$

Notice that 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*

$$t \to \sigma(t)\,, \qquad \frac{d\sigma}{dt} \ge 0\,, \tag{1.1.91}$$

with σ being a new parameter.

A convenient parametrization is via the proper length of Γ_{yx} which is given by

$$s = \int_{\Gamma_{yx}} ds \tag{1.1.92}$$

where

$$ds = \sqrt{\dot z^2(\sigma)}d\sigma \tag{1.1.93}$$

and $\sigma \in [\sigma_i, \sigma_f]$ is some parametrization. For the obvious reasons the parametrization

$$t = \frac{1}{m}s \tag{1.1.94}$$

with s given by Eq. (1.1.92) is called the *proper time* parametrization. Notice that the dimension of t is $[\text{length}]^2$ according to Eq. (1.1.94).

Let us denote[4]

$$S[\Gamma_{yx}] \equiv \frac{m^2\tau}{2} + \frac{1}{2}\int_0^\tau dt\, \dot z^2(t)\,. \tag{1.1.95}$$

[4]The notation $S[\Gamma]$ with square brackets means that S is a functional of Γ while $f(x)$ with parentheses stands for functions.

The sense of this notation is that the RHS coincides with the classical action of a relativistic free (scalar) particle in the proper-time parametrization (1.1.94) when

$$\int_0^\tau dt\dot z^2(t) = m\int_0^\tau ds = m\,\text{Length}[\Gamma] \tag{1.1.96}$$

since

$$\left(\frac{dz_\mu(s)}{ds}\right)^2 = 1 \tag{1.1.97}$$

and $m\tau = \text{Length}[\Gamma]$ by the definition of the proper time.

Therefore, the path-integral representation (1.1.66) is nothing but the sum over trajectories with the weight being an exponential of (minus) classical action:

$$G(x-y) = \sum_{\Gamma_{yx}} \mathrm{e}^{-S[\Gamma_{yx}]}. \tag{1.1.98}$$

This sum is split in Eq. (1.1.66) into the trajectories along which the particle propagates during the proper time τ and the integral over τ.

Equation (1.1.98) implies that the transition amplitude in quantum mechanics is a sum over all paths which connects x and y. In other words, a particle propagates from x to y along all paths Γ_{yx} including the ones which are forbidden by the free classical equation of motion

$$\ddot z_\mu(t) = 0. \tag{1.1.99}$$

Only the classical trajectory (1.1.99) survives the path integral in the classical limit $\hbar \to 0$. The reason for this is that if the dependence on the Planck's constant is restored, it appears in the exponent:

$$G(x-y) = \sum_{\Gamma_{yx}} \mathrm{e}^{-S[\Gamma_{yx}]/\hbar}. \tag{1.1.100}$$

As $\hbar \to 0$ the path integral is dominated by a saddle point which is given in the free case by the classical equation of motion (1.1.99).

It is worth noting that the sum-over-path representation (1.1.98) is written entirely in terms of trajectories as geometric objects and does not refer to a concrete parametrization. For the free theory $S[\Gamma]$ is proportional to the length of the trajectory Γ:

$$S_{\text{free}}[\Gamma] \;=\; m\,\text{Length}[\Gamma], \tag{1.1.101}$$

where the length is given for some parametrization σ of the trajectory Γ by

$$\text{Length}[\Gamma] \;=\; \int_{\sigma_i}^{\sigma_f} d\sigma \sqrt{\dot{z}^2(\sigma)}. \tag{1.1.102}$$

The sum-over-path representation (1.1.98) with $S[\Gamma]$ given by the classical action (Eq. (1.1.101) in the free case) is often considered as a first principle of constructing quantum mechanics given the classical action $S[\Gamma]$.

Problem 1.8 Represent the matrix element of the (Euclidean) evolution operator $\langle y\,|\exp\{-\boldsymbol{H}\tau\}|\,x\rangle$ for the non-relativistic Hamiltonian

$$\boldsymbol{H} = -\frac{\partial^2}{2m} + V(x) \tag{1.1.103}$$

as a path integral.

Solution The calculation is similar to the one which is already done in Subsect. 1.1.4. It is most convenient to use a path-integral over velocity which was considered in Problem 1.5. The proper disentangling formula reads

$$\left\langle y \left| \mathrm{e}^{-\boldsymbol{H}\tau} \right| x \right\rangle = \int Dv_\mu(t)\, \mathrm{e}^{-\frac{m}{2}\int_0^\tau dt\, v_\mu^2(t)}\, \mathbf{P}\,\mathrm{e}^{-\int_0^\tau dt v^\mu(t)\partial_\mu(t) - \int_0^\tau dt V(x;t)}\delta^{(d)}(x-y). \tag{1.1.104}$$

Here the argument t in $V(x;t)$ is just the ordering parameter, while the same formula holds when the potential is explicitly time dependent.

In contrast to Eq. (1.1.67), we have put the minus sign in front of the linear-in-v term in the exponent in Eq. (1.1.104), so that it agrees with Appendix B of Feynman's paper [Fey51]. In fact it does not matter what sign is used since the integral over $v(t)$ is Gaussian so that only even powers of v survive after the integration.

The path-ordered exponential in Eq. (1.1.104) reads explicitly

$$\mathbf{P}\,\mathrm{e}^{-\int_0^\tau dt\, v^\mu(t)\partial_\mu(t) - \int_0^\tau dt V(x;t)} \;=\; \lim_{\epsilon\to 0}\prod_{i=1}^{M}\left[1 - \epsilon v_i^\mu \frac{\partial}{\partial x^\mu} - \epsilon V(x;i\epsilon)\right], \tag{1.1.105}$$

which can be rewritten as

$$\mathbf{P}\,\mathrm{e}^{-\int_0^\tau dt v^\mu(t)\partial_\mu(t) - \int_0^\tau dt V(x;t)} \;=\; \lim_{\epsilon\to 0}\prod_{i=1}^{M}\left[1 - \epsilon v_i^\mu \frac{\partial}{\partial x^\mu}\right]\left[1 - \epsilon V(x;i\epsilon)\right], \tag{1.1.106}$$

if terms which vanish as $\epsilon \to 0$ are neglected, or equivalently as

$$\mathbf{P}\,\mathrm{e}^{-\int_0^\tau dt v^\mu(t)\partial_\mu(t) - \int_0^\tau dt V(x;t)} \;=\; \prod_{t=0}^{\tau}\left[1 - dt v^\mu(t) \frac{\partial}{\partial x^\mu}\right]\left[1 - dt\, V(x;t)\right]. \tag{1.1.107}$$

There is no need to write down the t-dependence of $\partial_\mu(t)$ in these formulas since the order of the operators is explicit.

To disentangle the operator expression (1.1.107), let us note that

$$[1 - dtv^\mu(t)\partial_\mu] \;=\; U^{-1}(t+dt)U(t) \tag{1.1.108}$$

with

$$U(t) \;=\; \exp \int_0^t dt' v^\mu(t')\partial_\mu \tag{1.1.109}$$

being the shift operator. It obviously obeys the differential equation

$$\frac{d}{dt}U(t) \;= v^\mu(t)\partial_\mu\, U(t)\,. \tag{1.1.110}$$

Now since

$$U(t)\left[1 - dt\, V(x;t)\right]U^{-1}(t) \;=\; \left[1 - dt\, V\left(x + \int_0^t dt' v^\mu(t');t\right)\right], \tag{1.1.111}$$

the RHS of Eq. (1.1.107) can be written in the form

$$\begin{aligned} &\prod_{t=0}^{\tau}\left[1 - dt\, v^\mu(t)\frac{\partial}{\partial x^\mu}\right]\left[1 - dt\, V(x;t)\right] \\ &= \; U^{-1}(\tau)\prod_{t=0}^{\tau}\left[1 - dt\, V\left(x + \int_0^t dt' v^\mu(t');t\right)\right] \\ &= \; U^{-1}(\tau)\exp\left[-\int_0^\tau dt\, V\left(x + \int_0^t dt' v^\mu(t');t\right)\right], \end{aligned} \tag{1.1.112}$$

which is completely disentangled.

The operator $U^{-1}(\tau)$ is now in the proper order to be applied to the variable y in the argument of the delta-function, which results in the shift

$$\delta^{(d)}(x-y) \implies \delta^{(d)}\left(x+\int_0^\tau dt\, v^\mu(t)-y\right). \tag{1.1.113}$$

This will be explained in more detail in the next paragraphs.

Passing to the variable (1.1.72), we get finally

$$\left\langle y \left| \mathrm{e}^{-\boldsymbol{H}\tau} \right| x \right\rangle = \int_{\substack{z_\mu(0)=x_\mu \\ z_\mu(\tau)=y_\mu}} Dz_\mu(t)\; \mathrm{e}^{-\int_0^\tau dt\, \mathcal{L}(t)}, \tag{1.1.114}$$

where

$$\mathcal{L}(t) = \frac{m}{2}\dot{z}_\mu^2(t)+V(z(t)) \tag{1.1.115}$$

is the Lagrangian associated with the Hamiltonian $\boldsymbol{H}$. The unusual plus-sign in this formula is due to Euclidean-space formalism. It is clear from the derivation that Eq. (1.1.114) holds for time-dependent potentials as well.

Notice that the path integral in Eq. (1.1.114) is now over trajectories which the particle propagates in the fixed proper time τ with no integration over τ.

A special comment about the operator $U^{-1}(\tau)$ in Eq. (1.1.112) is needed. In the Schrödinger representation of quantum mechanics, one is interested in the matrix elements on the evolution operator between some vectors $\langle g|$ and $|f\rangle$ in the Hilbert space. According to Eq. (1.1.23) one has in the coordinate representation

$$\left\langle g \left| \mathrm{e}^{-\boldsymbol{H}\tau} \right| f \right\rangle = \int d^d x\, g(x)\, \mathrm{e}^{-\boldsymbol{H}\tau} f(x)\,. \tag{1.1.116}$$

Integrating by parts, the operator $U^{-1}(\tau)$ can then be applied to $g(x)$ which results in the shift

$$g(x) \implies U(\tau)\,g(x)\,U^{-1}(\tau)=g\left(x+\int_0^\tau dt\, v^\mu(t)\right). \tag{1.1.117}$$

Passing to the variable (1.1.72), Eq. (1.1.116) becomes

$$\left\langle g \left| \mathrm{e}^{-\boldsymbol{H}\tau} \right| f \right\rangle = \int Dz_\mu(t)\; \mathrm{e}^{-\int_0^\tau dt\, \mathcal{L}(t)} g(z(\tau))f(z(0))\,. \tag{1.1.118}$$

There are no restrictions on the initial and final points of the trajectories $z_\mu(t)$ in this formula.

Problem 1.9 Calculate the diagonal resolvent of the Schrödinger operator in the potential $V(x)$:

$$R_\omega(x,x;V) = \left\langle x \left| \frac{1}{-\mathcal{G}\partial^2+\omega^2+V} \right| x \right\rangle, \tag{1.1.119}$$

in the limit $\mathcal{G}\to 0$ for $d=1$.

Solution Using the formula of the type (1.1.114), we represent $R_\omega(x,x;V)$ as the path integral

$$R_\omega(x,x;V) = \frac{1}{2}\int_0^\infty d\tau\; \mathrm{e}^{-\frac{1}{2}\tau\omega^2} \int_{\substack{z_\mu(0)=x_\mu \\ z_\mu(\tau)=x_\mu}} Dz_\mu(t)\; \mathrm{e}^{-\frac{1}{2\mathcal{G}}\int_0^\tau dt\, \dot{z}_\mu^2(t)-\int_0^\tau dt V(z(t))}. \tag{1.1.120}$$

As $\mathcal{G}\to 0$ this path integral is dominated by the t-independent saddle-point trajectory

$$z(t) = x\,, \tag{1.1.121}$$

which is associated with a particle standing at the point x. Substituting V at this saddle-point, *i.e.* replacing $V(z(t))$ by $V(x)$, and calculating the Gaussian integral over quantum fluctuations around the trajectory (1.1.121) by the use of Eqs. (1.1.84) and (1.1.86), one gets

$$R_\omega(x,x;V) = \frac{1}{2\sqrt{\omega^2+V(x)}} \tag{1.1.122}$$

in $d=1$.

Equation (1.1.122) can be alternatively derived by applying the Gelfand–Dickey technique [GD75] which says that $R_\omega(x,x;V)$ obeys the third order linear differential equation

$$\tfrac{1}{2}\left(\tfrac{1}{2}\mathcal{G}\partial^3-\partial V(x)-V(x)\partial\right)R_\omega(x,x;V) = \omega^2\partial R_\omega(x,x;V)\,. \tag{1.1.123}$$

$R_\omega(x,x;V)$ given by Eq. (1.1.122) obviously satisfies this equation as $\mathcal{G}\to 0$.

One more way to derive Eq. (1.1.122) is to perform a semiclassical WKB expansion of $R_\omega(x,y;V)$ in the parameter $\mathcal{G}$. This is explained in Chapter 7 of the book [LL74].

Problem 1.10 Derive Eq. (1.1.123).

Solution The resolvent

$$R_\omega(x,y;V) = \left\langle y \left| \frac{1}{-\mathcal{G}\partial^2+\omega^2+V} \right| x \right\rangle \tag{1.1.124}$$

obeys the equations

$$\begin{aligned}\left(-\mathcal{G}\frac{\partial^2}{\partial x^2}+\omega^2+V(x)\right)R_\omega(x,y;V) &= \delta^{(1)}(x-y)\,,\\ \left(-\mathcal{G}\frac{\partial^2}{\partial y^2}+\omega^2+V(y)\right)R_\omega(x,y;V) &= \delta^{(1)}(x-y)\,.\end{aligned} \tag{1.1.125}$$

It can be expressed via the two solutions $f_\pm(x)$ of the homogeneous equation

$$\left(-\mathcal{G}\frac{\partial^2}{\partial x^2}+\omega^2+V(x)\right)f_\pm(x) = 0\,, \tag{1.1.126}$$

where f_+ or f_- are regular at $+\infty$ or $-\infty$, respectively. Then the full solution is

$$R_\omega(x,y;V) = \frac{f_+(x)f_-(y)\theta(x-y)+f_-(x)f_+(y)\theta(y-x)}{\mathcal{G}\,W_\omega} \tag{1.1.127}$$

with

$$W_\omega = f_+(x)f_-'(x)-f_+'(x)f_-(x) \tag{1.1.128}$$

being the Wronskian of these solutions. Applying $\partial/\partial x$ to Eq. (1.1.128), it is easy to show that W_ω is an x-independent function of ω.

The simplest way to prove Eq. (1.1.123) is to differentiate

$$R_\omega(x,x;V) = \frac{f_+(x)f_-(x)}{\mathcal{G}\,W_\omega} \tag{1.1.129}$$

using Eq. (1.1.126), in order to verify that it satisfies the nonlinear differential equation

$$-2\mathcal{G}R_\omega R_\omega''+\mathcal{G}\left(R_\omega'\right)^2+4\left[\omega^2+V\right]R_\omega^2 = 1\,. \tag{1.1.130}$$

One more differentiation of Eq. (1.1.130) with respect to x results in Eq. (1.1.123).

It is worth noting that Eq. (1.1.130) is very convenient for calculating the semiclassical expansion of $R_\omega(x,x;V)$ in $\mathcal{G}$. In particular, the leading order (1.1.122) is obvious.

Remark on parametric invariant representation

The Green function $G(x-y)$ can be alternatively calculated from the parametric invariant representation

$$G(x-y) \propto \int\limits_{\substack{z_\mu(\sigma_i)=x_\mu\\ z_\mu(\sigma_f)=y_\mu}} Dz(\sigma)\,\mathrm{e}^{-m_0\int_{\sigma_i}^{\sigma_f}d\sigma\sqrt{\dot z^2(\sigma)}} \tag{1.1.131}$$

as is prescribed by Eqs. (1.1.101) and (1.1.102). In contrast to (1.1.66), this path integral is not easy to calculate. The integration over $Dz(\sigma)$ in Eq. (1.1.131) involves integration over the reparametrization group which gives the proper group-volume factor since the exponent is parametric invariant. Eq. (1.1.66) is recovered after fixing parametrization to be proper time. How to perform this calculation is explained in Chapter 9 of the book by Polyakov [Pol87].

If one makes a naive discretization of the parameter σ by equidistant intervals, the exponent in Eq. (1.1.131) is highly nonlinear in the variables z_i, leading to complicated integrals. On the contrary the discretization (1.1.87) of the path integral (1.1.66), where the parametric invariance is fixed, results in a Gaussian integral which is easily calculable.

Problem 1.11 Calculate the path integral in Eq. (1.1.131), discretizing the measure by

$$Dz_\mu \;\to\; \sum_{M=1}^{\infty}\prod_{i=1}^{M}\frac{d^dz_i}{(2\pi\epsilon)^{d/2}} \tag{1.1.132}$$

and applying the central limit theorem as $M\to\infty$.

Solution By making the discretization, we represent the RHS of Eq. (1.1.131) as the probability integral

$$\begin{aligned}G_\epsilon(x-y) &= \frac{1}{(2\pi\epsilon)^{d/2}}\sum_{M=1}^{\infty}\int\prod_{i=1}^{M-1}\frac{d^dz_i}{(2\pi\epsilon)^{d/2}}\\ &\quad\times\rho(x\to z_1)\,\rho(z_1\to z_2)\cdots\rho(z_{M-1}\to y)\end{aligned} \tag{1.1.133}$$

with

$$\rho(z_{i-1}\to z_i) = \mathrm{e}^{-m_0|z_i-z_{i-1}|} \tag{1.1.134}$$

being an (unnormalized) probability function and ϵ is a parameter of the dimension of $[\text{length}]^2$. The probability interpretation of each term in the sum is standard for random walk models, and means, as usual, that a particle propagates via independent intermediate steps. The discretization of the measure given by Eq. (1.1.133) looks like that in Eq. (1.1.62), but the summation over M is now added.

Since the integral in Eq. (1.1.133) is a convolution, the central limit theorem states that

$$\begin{aligned}G_\epsilon(x-y) &= \frac{1}{(2\pi\epsilon)^{d/2}}\sum_M\left[\frac{c_0}{(2\pi\epsilon m_0^2)^{d/2}}\right]^M\\ &\quad\times\frac{1}{(2\pi\sigma^2M)^{d/2}}\,\mathrm{e}^{-m_0^2(x-y)^2/(2\sigma^2M)+\mathcal{O}(M^{-2})}\end{aligned} \tag{1.1.135}$$

at large M, where c_0 and σ^2 are the zeroth and (normalized) second moments of ρ:

$$\begin{aligned} c_0 &= \int d^d x \,\mathrm{e}^{-|x|} = 2\pi^{d/2} \frac{\Gamma(d)}{\Gamma(d/2)}, \\ \sigma^2 &= \frac{1}{c_0} \int d^d x \, x^2 \,\mathrm{e}^{-|x|} = d(d+1). \end{aligned} \tag{1.1.136}$$

The sum over M in Eq. (1.1.135) is convergent for

$$m_0 > m_c = \frac{c_0^{1/d}}{\sqrt{2\pi\epsilon}} \tag{1.1.137}$$

and is divergent for $m_0 < m_c$. Choosing $m_0 > m_c$, but $m_0^2 - m_c^2 \sim 1$ in the limit $\varepsilon \to 0$, the sum over M will be convergent, while dominated by terms with large

$$M \sim m_c^2 \sim \frac{1}{\epsilon}. \tag{1.1.138}$$

This is easily seen by rewriting Eq. (1.1.135) as

$$\begin{aligned} G_\epsilon(x-y) &= \sum_M \left(\frac{m_c^2}{2\pi\sigma^2 M} \right)^{d/2} \\ &\quad \times \mathrm{e}^{-dM \ln(m_o/m_c) - m_0^2 (x-y)^2/(2\sigma^2 M) + \mathcal{O}(M^{-2})}. \end{aligned} \tag{1.1.139}$$

Each term with $M \sim m_c^2$ contributes $\mathcal{O}(1)$ to the sum, so that

$$G_\epsilon(x-y) \sim m_c^2. \tag{1.1.140}$$

This justifies the using of the central limit theorem in this case. The typical distances between the z_i's, which are essential in the integral on the RHS of Eq. (1.1.133), are

$$|z_i - z_{i-1}| \sim \frac{1}{m_0} \sim \sqrt{\epsilon} \tag{1.1.141}$$

like in Eq. (1.1.89). The relation (1.1.138) between the essential values of M and ϵ is also similar to what we had in Subsect. 1.1.4.

The sum over M in Eq. (1.1.139) can be replaced by a continuous integral over the variable

$$\tau = \frac{\sigma^2 M}{m_c^2}, \tag{1.1.142}$$

which is $\mathcal{O}(1)$ for $M \sim m_c^2$. Introducing also the variable m by

$$m^2 \equiv \frac{d}{\sigma^2} \left(m_0^2 - m_c^2 \right) > 0, \tag{1.1.143}$$

we rewrite Eq. (1.1.139) as

$$G_\epsilon(x-y) \overset{\epsilon \to 0}{\to} \frac{m_c^2}{\sigma^2} \int_0^\infty d\tau \frac{1}{(2\pi\tau)^{d/2}} \mathrm{e}^{-\frac{1}{2} m^2 \tau - \frac{(x-y)^2}{2\tau}}, \tag{1.1.144}$$

whose RHS is proportional to that in Eq. (1.1.85) for the Euclidean propagator.

Remark on discretized path-ordered exponential

As is discussed in Subsect. 1.1.3, the order of operators A_i and B_i with the same index i is not essential in the path-ordered exponential (1.1.52) as $M \to \infty$. If Eq. (1.1.52) is promoted to be valid at finite M (or at least to the order of $\mathcal{O}(1/M)$), this specifies the commutator of A_i and B_i. Analogously, a discretization of Eq. (1.1.114) specifies in which order the product of $x_i p_i$ in the classical theory should be substituted by the operators $\boldsymbol{x}_i$ and $\boldsymbol{p}_i$ in the operator formalism. See details in the books by Berezin [Ber86] (Chapter 1 of the Part II) and Sakita [Sak85] (Chapter 6).

1.1.7 Propagators in external field

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

$$\boldsymbol{p}^\mu \longrightarrow \boldsymbol{p}^\mu - eA^\mu(x). \tag{1.1.145}$$

Recalling the definition (1.1.26) of the Euclidean four-momentum, ∂_μ needs to be replaced by the covariant derivative

$$\partial_\mu \longrightarrow \nabla_\mu = \partial_\mu - ieA_\mu(x). \tag{1.1.146}$$

Inverting the operator ∇_μ^2 by the use of the disentangling procedure, one gets

$$\begin{aligned} G(x,y;A) &\equiv \left\langle y \left| \frac{1}{-\nabla_\mu^2 + m^2} \right| x \right\rangle \\ &= \frac{1}{2} \int_0^\infty d\tau \,\mathrm{e}^{-\frac{1}{2}\tau m^2} \int_{\substack{z(0)=x \\ z(\tau)=y}} Dz(t) \;\mathrm{e}^{-\frac{1}{2}\int_0^\tau dt\, \dot z^2(t) + ie\int_0^\tau dt\, \dot z^\mu(t) A_\mu(z(t))}. \end{aligned} \tag{1.1.147}$$

Notice that the exponent is just the classical (Euclidean) action of a particle in an external electromagnetic field. Therefore, this expression is again of the type in Eq. (1.1.98).

The path-integral representation (1.1.147) for the propagator of a scalar particle in an external electromagnetic field is due to Feynman [Fey50] (Appendix A).

Problem 1.12 Derive Eq. (1.1.147) using Eq. (1.1.67) with $D_\mu = -\nabla_\mu$.
Solution The calculation is analogous to that of Problem 1.8. We have

$$D_\mu(t) \equiv -\nabla_\mu(t) = -\partial_\mu(t) + ieA_\mu(x;t) \tag{1.1.148}$$

so that explicitly

$$\begin{aligned} \mathbf{P}\,\mathrm{e}^{-\int_0^\tau dt\, v^\mu(t)\nabla_\mu(t)} &= \prod_{t=0}^{\tau}\left[1 - dt\, v^\mu(t)\frac{\partial}{\partial x^\mu} + ie\, dt\, v^\mu(t)A_\mu(x;t)\right] \\ &= \prod_{t=0}^{\tau}\left[1 - dtv^\mu(t)\frac{\partial}{\partial x^\mu}\right][1 + ie\, dt\, v^\mu(t)A_\mu(x;t)]\,. \end{aligned} \tag{1.1.149}$$

This looks exactly like the expression (1.1.107) with

$$V(x;t) = -ie\, v^\mu(t)A_\mu(x;t)\,. \tag{1.1.150}$$

Substituting this potential in Eq. (1.1.114) and remembering the additional integration over τ, we obtain the path-integral representation (1.1.147).

We can alternatively rewrite Eq. (1.1.147) as

$$G\,(x,y;A) = {\sum_{\Gamma_{yx}}}' \mathrm{e}^{\,ie\int_{\Gamma_{yx}} dz^\mu\, A_\mu(z)}, \tag{1.1.151}$$

where, proceeding as in Subsect. 1.1.6, we have included the free action in the definition of the sum over trajectories:

$${\sum_{\Gamma_{yx}}}' = \sum_{\Gamma_{yx}} \mathrm{e}^{-S_{\rm free}[\Gamma_{yx}]}, \tag{1.1.152}$$

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

$$dz^\mu = dt\ \dot z^\mu(t) \tag{1.1.153}$$

along the trajectory Γ_{yx}.

The meaning of Eq. (1.1.151) is that the transition amplitude of a quantum particle in a classical electromagnetic field is the sum over paths of the Abelian *phase factor*

$$U[\Gamma_{yx}] = \mathrm{e}^{\,ie\int_{\Gamma_{yx}} dz^\mu A_\mu(z)}. \tag{1.1.154}$$

Under the gauge transformation

$$A_\mu\,(z) \stackrel{\rm g.t.}{\longrightarrow} A_\mu\,(z) + \frac{1}{e}\partial_\mu\alpha\,(z)\,, \tag{1.1.155}$$

the Abelian phase factor transforms as

$$U[\Gamma_{yx}] \stackrel{\rm g.t.}{\longrightarrow} \mathrm{e}^{\,i\alpha(y)}\, U[\Gamma_{yx}]\ \mathrm{e}^{\,-i\alpha(x)}\,. \tag{1.1.156}$$

Noting that a wave function at the point x is transformed under the gauge transformation (1.1.155) as

$$\varphi(x) \stackrel{\rm g.t.}{\longrightarrow} \mathrm{e}^{\,i\alpha(x)}\,\varphi(x), \tag{1.1.157}$$

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

$$U[\Gamma_{yx}] \stackrel{\rm g.t.}{\sim} \text{“}\varphi(y)\varphi^\dagger(x)\text{”}\,, \tag{1.1.158}$$

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

As a consequence of Eqs. (1.1.156) and (1.1.157), a wave function at the point x transforms like one at the point y after multiplication by the phase factor:

$$U[\Gamma_{yx}]\,\varphi(x) \stackrel{\rm g.t.}{\sim} \text{“}\varphi(y)\text{”}\,, \tag{1.1.159}$$

and analogously

$$\varphi^\dagger(y)\,U[\Gamma_{yx}] \stackrel{\rm g.t.}{\sim} \text{“}\varphi^\dagger(x)\text{”}\,. \tag{1.1.160}$$

Equations (1.1.159) and (1.1.160) 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 factor[5].

Below we shall deal with determinants of various operators. Analogous to Eq. (1.1.147), one gets

$$\begin{aligned}\ln\det\nabla^2_\mu &= \frac{1}{2}\int\limits_0^\infty \frac{d\tau}{\tau}\,\mathrm{Sp}\;\mathrm{e}^{\,\frac{1}{2}\tau\nabla^2_\mu} \\ &= \frac{1}{2}\int_0^\infty \frac{d\tau}{\tau}\int\limits_{z(0)=z(\tau)} Dz\,(t)\;\mathrm{e}^{-\frac{1}{2}\int_0^\tau dt\,\dot z^2(t)+ie\oint_\Gamma dz^\mu A_\mu(z)},\end{aligned} \tag{1.1.161}$$

where the path integral goes over trajectories which are closed due to the boundary condition $z_\mu(0)=z_\mu(\tau)$. To derive Eq. (1.1.161), we have used the formula

$$\ln\det\boldsymbol{D} \;=\; \mathrm{Sp}\ln\boldsymbol{D}\,, \tag{1.1.162}$$

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

Problem 1.13 Prove Eq. (1.1.162).
Solution Let us first reduce $\boldsymbol{D}$ to a diagonal form by a unitary transformation and denote eigenvalues as D_i. Then Eq. (1.1.162) reads

$$\ln\prod_i D_i \;=\; \sum_i \ln D_i \tag{1.1.163}$$

which is obviously true.

The phase factor for a closed contour Γ enters Eq. (1.1.161). It describes a parallel transportation along a closed loop, and is gauge invariant as a consequence of Eq. (1.1.156):

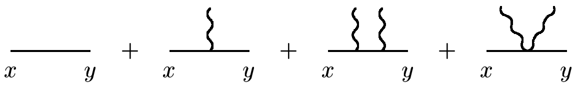
$$\mathrm{e}^{\,ie\oint_\Gamma dz^\mu A_\mu(z)} \;\xrightarrow{\text{g.t.}}\; \mathrm{e}^{\,ie\oint_\Gamma dz^\mu A_\mu(z)}. \tag{1.1.164}$$

This quantity, which plays a crucial role in modern formulations of gauge theories, will be discussed in more detail in Sect. 2.1.

[5]Strictly 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. 2.1.4.

Problem 1.14 Show how the path-integral representation (1.1.147) recovers for $G(x,y;A)$ the diagrammatic expansion of propagator in an external field A_μ:

$$G(x,y;A) \;=$$



$$+\;\cdots \tag{1.1.165}$$

Solution Use the formula

$$\sum_{\Gamma_{yx}}{}' \int\limits_{\Gamma_{yx}} d\xi_\mu\,\delta^{(d)}(\xi-z)\ldots \;=\; \sum_{\Gamma_{yz}}{}' \frac{\overleftrightarrow{\partial}}{\partial z_\mu}\sum_{\Gamma_{zx}}{}'\ldots\,, \tag{1.1.166}$$

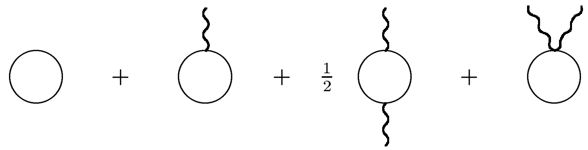
where

$$\overleftrightarrow{\partial}_\mu \;=\; -\overleftarrow{\partial}_\mu + \overrightarrow{\partial}_\mu\,. \tag{1.1.167}$$

The proof is given in Appendix A of the paper [MM81].

Problem 1.15 Establish the equivalence of the path-integral representation (1.1.161) of $\ln\det\nabla^2_\mu$ and the sum of one-loop diagrams in an external field A_μ:

$$\ln\det\nabla^2_\mu \;=$$



$$+\;\ldots\,, \tag{1.1.168}$$

with the combinatoric factor 1/2 in the third diagram being associated with a symmetry factor.
Solution The same as in the previous problem.

Remark on analogy with statistical mechanics

A formula of the type (1.1.161), which represents the trace of an operator via path integral over closed trajectories, is known as the Feynman–Kac formula. The terminology comes from statistical mechanics where the partition function (or equivalently the statistical sum) is given by the Boltzmann formula

$$Z \;=\; \mathrm{Sp}\;\mathrm{e}^{-\beta\boldsymbol{H}} \tag{1.1.169}$$

(with β being the inverse temperature) whose path-integral representation is of the type in Eq. (1.1.161). The expression which is integrated on the RHS of Eq. (1.1.161) 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. We shall explain this analogy in more detail in Chapter 2 (Sect. 2.5) when discussing quantum field theory at finite temperature.

1.2 Second quantization

In the previous Section we have considered first quantization of particles, where the operators of coordinate and momentum, $\boldsymbol{x}$ and $\boldsymbol{p}$ respectively, are represented in the coordinate space by

$$\boldsymbol{x} = x\,, \qquad \boldsymbol{p} = -i\frac{\partial}{\partial x} \qquad \boxed{= \text{ first quantization}}\,. \tag{1.2.1}$$

In the language of path integrals, first quantization is associated with integrals over trajectories in the coordinate space.

While propagators can be easily represented as path integrals, it is very difficult to describe, in this language, a (non-geometric) self-interaction of a particle, since this would correspond to extra weights for self-intersecting paths. For the free case (or a particle in an external gauge field) there are no such extra weights, and the transition amplitude is completely described by the classical action of the particle.

In the operator formalism, self-interactions of a particle is described by the use of second quantization — this is where the transition from quantum mechanics to quantum field theory begins. Second quantization is a quantization of fields, and is associated with path integrals over fields which is the subject of this Section. We demonstrate how perturbation theory and the Schwinger–Dyson equations can be derived by using path integrals.

1.2.1 Integration over fields

Let us define the following (Euclidean) *partition function*[6]

$$Z \;=\; \int D\varphi\,(x)\;\, \mathrm{e}^{\,-S} \tag{1.2.2}$$

where the action in the exponent is given for the free case by

$$S_{\text{free}}[\varphi] \;=\; \frac{1}{2}\int d^dx\, \Big((\partial_\mu\varphi)^2 + m^2\varphi^2\Big)\,. \tag{1.2.3}$$

The measure $D\varphi(x)$ is defined analogously to Eq. (1.1.62):

$$\int D\varphi\,(x)\ldots \;=\; \prod_x \int_{-\infty}^{+\infty} d\varphi\,(x)\ldots \tag{1.2.4}$$

[6] It is also called the statistical sum in an analogy with statistical mechanics (see Subsect. 1.1.7).

where the product runs over all space points x and the integral over $d\varphi$ is the Lebesgue one.

The propagator is given by the average

$$G(x,y) \;=\; \langle\, \varphi(x)\, \varphi(y)\, \rangle \tag{1.2.5}$$

where a generic average is defined by the formula

$$\langle\, F[\varphi]\, \rangle \;=\; Z^{-1} \int D\varphi(x)\, \mathrm{e}^{-S[\varphi]} F[\varphi]\,. \tag{1.2.6}$$

The notation is obvious since we average on the RHS of Eq. (1.2.6) over all field configurations with the same weight as in the partition function (1.2.2). The normalization factor Z^{-1} provides the necessary property of an average

$$\langle\, 1\, \rangle = 1\,. \tag{1.2.7}$$

Since the free action (1.2.3) is Gaussian, the average (1.2.5) equals

$$G(x-y) \;=\; \left\langle y \left| \frac{1}{-\partial^2 + m^2} \right| x \right\rangle, \tag{1.2.8}$$

which coincides with (1.1.57). Therefore, we have obtained the same propagator (1.1.85) as in the previous Section.

Problem 1.16 By discretizing the (Euclidean) space, derive

$$Z^{-1} \int \prod_x d\varphi_x \exp\left\{ -\tfrac{1}{2} \sum_{x,y} \varphi_x D_{xy} \varphi_y \right\} \varphi_x \varphi_y \;=\; D^{-1}_{xy}\,. \tag{1.2.9}$$

Solution Calculate the Gaussian integral by the change of variable

$$\varphi_x \;\to\; \varphi'_x \;=\; \sum_y \left(D^{-1/2}\right)_{xy} \varphi_y\,. \tag{1.2.10}$$

Note that the integrals over $\varphi(x)$ are convergent in Euclidean space. If a discretization of space is introduced, the path integrals in Eqs. (1.2.2) or (1.2.6) are rigorously defined.

Remark on Minkowski-space formulation

In Minkowski space, perturbation theory is well-defined since the Gaussian path integral, which determines the propagator

$$\int D\varphi\, \mathrm{e}^{\,iS} \;=\; \int D\varphi\, \mathrm{e}^{\,i \int d^dx\, \varphi \boldsymbol{D} \varphi}, \tag{1.2.11}$$

equals

$$\langle\, \varphi_x \varphi_y\, \rangle \;=\; \left\langle y \left| \frac{1}{i\boldsymbol{D}} \right| x \right\rangle. \tag{1.2.12}$$

Here $\boldsymbol{D}$ is a proper operator in Minkowski space.

It can not be said *a priory* whether a non-perturbative formulation of a given (interacting) theory via the path integral in Minkowski space exists since the weight factor is complex and the integral may be divergent.

1.2.2 Grassmann variables

Path integrals over anticommuting Grassmann variables are used to describe fermionic systems.

The Grassmann variables ψ_x and $\bar\psi_y$ obey the anticommutation relations

$$\begin{aligned} \{\psi_y, \psi_x\} &= 0\,, \\ \{\bar\psi_y, \bar\psi_x\} &= 0\,, \\ \{\bar\psi_y, \psi_x\} &= 0\,. \end{aligned} \tag{1.2.13}$$

As a consequence, the square of a Grassmann variable vanishes

$$\psi_x^2 \;=\; \bar\psi_x^2 \;=\; 0\,. \tag{1.2.14}$$

The path integral over the Fermi fields equals

$$\int D\bar\psi\, D\psi\, \mathrm{e}^{-\int d^dx\, \bar\psi \boldsymbol{D} \psi} \;=\; \det \boldsymbol{D} \tag{1.2.15}$$

while an analogous integral over the Bose fields is

$$\int D\varphi^\dagger\, D\varphi\, \mathrm{e}^{-\int d^dx\, \varphi^\dagger \boldsymbol{D} \varphi} \;=\; (\det \boldsymbol{D})^{-1}\,. \tag{1.2.16}$$

Problem 1.17 Define integrals over Grassmann variables.
Solution Assuming that ψ and $\bar{\psi}$ belong to the same Grassmann algebra, the Berezin integrals are defined by

$$\begin{aligned} &\int d\psi_x = \int d\bar{\psi}_x = 0 \,, \\ &\int d\psi_x \psi_y = \int d\bar{\psi}_x \bar{\psi}_y = \delta_{xy} \,, \\ &\int d\psi_x \bar{\psi}_y = \int d\bar{\psi}_x \psi_y = 0 \,. \end{aligned} \tag{1.2.17}$$

The simplest interesting integral is

$$\int d\bar{\psi}_x \, d\psi_x \, \mathrm{e}^{-\bar{\psi}_x \psi_x} = 1 \,. \tag{1.2.18}$$

The formula (1.2.15) can now easily be derived: representing $\langle y \,|\boldsymbol{D}|\, x\rangle$ in the diagonal form, expanding the exponential up to a term which is linear in all the Grassmann variables and calculating the integrals of this term according to Eq. (1.2.17). See more details in the book by Berezin [Ber86] (§3 of Part I).

The average over the Fermi fields, defined with the same weight as in Eq. (1.2.15), equals

$$\langle\, \psi\,(x)\; \bar{\psi}\,(y)\,\rangle = \langle y \,\big|\boldsymbol{D}^{-1}\big|\, x\rangle \tag{1.2.19}$$

which is the same as for bosons.

Notice that the fermionic partition function (1.2.15) can be rewritten according to Eq. (1.1.162) as

$$\det \boldsymbol{D} = \mathrm{e}^{\,\mathrm{Sp}\ln \boldsymbol{D}}. \tag{1.2.20}$$

The analogous formula for bosons (1.2.16) is rewritten as

$$(\det \boldsymbol{D})^{-1} = \mathrm{e}^{\,-\mathrm{Sp}\ln \boldsymbol{D}}. \tag{1.2.21}$$

The relative minus sign in the exponents on the RHS's of Eqs. (1.2.20) and (1.2.21) is a famous minus sign which emerges for closed fermionic loops which contribute to the (logarithm of the) partition function.

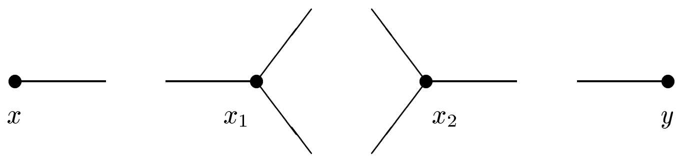


Fig. 1.4: Diagrammatic representation of the last term on the RHS of Eq. (1.2.23).

1.2.3 Perturbation theory

The cubic self-interaction of the scalar field is described by the action

$$S[\varphi] = \int d^d x \left(\frac{1}{2}\,(\partial_\mu \varphi)^2 + \frac{1}{2} m^2 \varphi^2 + \frac{\lambda}{3!}\varphi^3 \right), \tag{1.2.22}$$

where λ is the coupling constant.

To construct perturbation theory, we expand the exponential in λ and calculate the Gaussian averages with the free action (1.2.3).

To order λ^2, the expansion is

$$\begin{aligned} \langle\, \varphi\,(x)\, \varphi\,(y)\,\rangle &= \langle\, \varphi\,(x)\, \varphi\,(y)\,\rangle_{\text{free}} \\ &\cdot \left(1 - \left\langle \frac{\lambda}{3!} \int d^d x_1\, \varphi^3\,(x_1)\, \frac{\lambda}{3!} \int d^d x_2\, \varphi^3\,(x_2) \right\rangle_{\text{free}} \right) \\ &+ \left\langle \varphi\,(x)\, \frac{\lambda}{3!} \int d^d x_1\, \varphi^3\,(x_1)\, \frac{\lambda}{3!} \int d^d x_2\, \varphi^3\,(x_2)\, \varphi\,(y) \right\rangle_{\text{free}} + \ldots \end{aligned} \tag{1.2.23}$$

The term which is linear in λ (as well as the Gaussian average of any odd power of φ) vanishes due to the reflection symmetry $\varphi \to -\varphi$ of the Gaussian action. The last term on the RHS is depicted graphically in Fig. 1.4.

Further calculation of the RHS of Eq. (1.2.23) is based on the free average

$$\langle\, \varphi\,(x_i)\, \varphi\,(x_j)\,\rangle_{\text{free}} = G\,(x_i - x_j) \tag{1.2.24}$$

and rules of the Wick pairing of Gaussian averages, which allow us to represent the average of a product as the sum of all possible products of pair averages. Some of the diagrams which emerge after the Wick contraction are depicted in Fig. 1.5. These diagrams are called Feynman diagrams.

The diagram of Fig. 1.5b is disconnected. Its disconnected part with the two loops cancels with the same contribution from Z^{-1} (which yields the

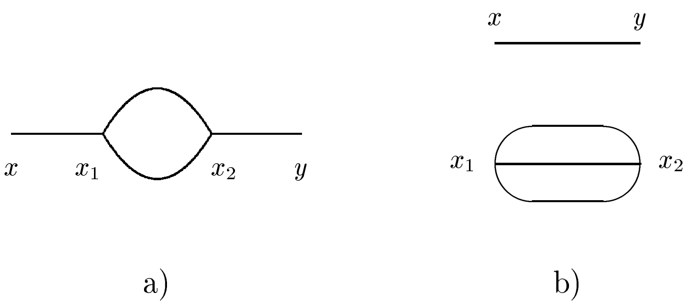


Fig. 1.5: Some of the Feynman diagrams which appear from (1.2.23) after the Wick pairing.

factor in the first term on the RHS of Eq. (1.2.23)). This is a general property that only connected diagrams are left in $\langle\, \varphi\,(x_i)\, \varphi\,(x_j)\,\rangle$.

Let us note finally that the combinatoric factor $1/2$ is correctly reproduced for the diagram of Fig. 1.5a. For an arbitrary diagram, this procedure of pairing reproduces the usual combinatoric factor which is equal to the number of automorphisms of the diagram (*i.e.* the symmetries of a given graph).

1.2.4 Schwinger–Dyson equations

Feynman diagrams can be alternatively derived by iterations in the coupling constant of the set of the Schwinger–Dyson equations which is a quantum analogue of the classical equation of motion.

To derive the Schwinger–Dyson equations, let us utilize the fact that the measure (1.2.4) is invariant under an arbitrary shift of the field

$$\varphi\,(x) \quad\to\quad \varphi\,(x) + \delta\varphi\,(x)\,. \tag{1.2.25}$$

This invariance is obvious since the functional integration goes over all the fields while the shift (1.2.25) is just a transformation from one field configuration to another.

Since the measure is invariant, the path integral in the average (1.2.6) does not change under the shift (1.2.25):

$$\int d^d x\, \delta\varphi\,(x) \int D\varphi\; \mathrm{e}^{\,-S[\varphi]} \left[-\frac{\delta S\,[\varphi]}{\delta\varphi(x)} F[\varphi] + \frac{\delta F\,[\varphi]}{\delta\varphi(x)} \right] \;=\; 0. \tag{1.2.26}$$

Since $\delta\varphi(x)$ is arbitrary, Eq. (1.2.26) results in the following *quantum equation of motion*

$$\frac{\delta S\,[\varphi]}{\delta\varphi\,(x)} \;\overset{\text{w.s.}}{=}\; \hbar \frac{\delta}{\delta\varphi\,(x)}, \tag{1.2.27}$$

where we have explicitly written the dependence on the Planck's constant $\hbar$. It appears this way since the action S is divided by $\hbar$ in Eq. (1.2.2) when $\hbar$ is restored.

Equation (1.2.27) is to be understood in the weak sense, *i.e.* it is valid under the sign of averaging when applied to a functional $F[\varphi]$. In other words the variation of the action on the LHS of Eq. (1.2.27) can always be substituted by the variational derivative on the RHS when integrated over fields with the same weight as in Eq. (1.2.6). Therefore, one arrives at the following functional equation

$$\left\langle \frac{\delta S\,[\varphi]}{\delta\varphi\,(x)} F\,[\varphi] \right\rangle \;=\; \hbar \left\langle \frac{\delta F\,[\varphi]}{\delta\varphi\,(x)} \right\rangle. \tag{1.2.28}$$

This equation is quite similar to that which Schwinger considered in the framework of his variational technique.

1.2.5 Commutator terms

In order to show how Eq. (1.2.28) reproduces Eq. (1.1.34) for the free propagator, let us choose

$$F\,[\varphi] \;=\; \varphi\,(y)\,. \tag{1.2.29}$$

Substituting into Eq. (1.2.28) and calculating the variational derivative, one gets

$$\left(-\partial^2 + m^2\right) \langle \varphi\,(x)\, \varphi\,(y) \rangle \;=\; \hbar \left\langle \frac{\delta\varphi\,(y)}{\delta\varphi\,(x)} \right\rangle \;=\; \hbar\, \delta^{(d)}\,(x-y) \tag{1.2.30}$$

which coincides with Eq. (1.1.34).

The LHS of Eq. (1.2.30) emerges from the variation of the free classical action (1.2.3)

$$\frac{\delta S_{\text{free}}}{\delta\varphi\,(x)} \;=\; \left(-\partial^2 + m^2\right) \varphi\,(x) \tag{1.2.31}$$

while the RHS, which results from the variational derivative, emerges in the operator formalism from the canonical commutation relations

$$\delta\left(x_0-y_0\right)\left[\boldsymbol{\varphi}\left(x_0,\vec{x}\right), \dot{\boldsymbol{\varphi}}\left(y_0,\vec{y}\right)\right] \;=\; i\,\delta^{(d)}\left(x-y\right) \tag{1.2.32}$$

as is explained in Subsect. 1.1.1.

For this reason, the RHS of Eq. (1.2.30) and, more generally, the RHS of Eq. (1.2.28) is called the commutator term. The variational derivative on the RHS of Eq. (1.2.27) plays the role of the conjugate momentum in the operator formalism. The calculation of this variational derivative in Euclidean space is equivalent to differentiating the T–product and using canonical commutation relations in Minkowski space.

When the Planck's constant vanishes, $\hbar \to 0$, the RHS of Eq. (1.2.27) (or Eq. (1.2.28)) vanishes. Therefore it reduces to the classical equation of motion for the field φ:

$$\frac{\delta S\left[\varphi\right]}{\delta\varphi\left(x\right)} \;=\; 0. \tag{1.2.33}$$

This implies that the path integral over fields has a saddle point as $\hbar \to 0$ which is given by Eq. (1.2.33).

Another lesson we have learned is that the average (1.2.5), which is defined via the Euclidean path integral, is associated with the Wick rotated T–product. We have already seen this property in the previous Section in the language of first quantization. More generally, the Euclidean average (1.2.6) is associated with the vacuum expectation value of $\langle 0\,|\,\mathbf{T}F[\boldsymbol{\varphi}]\,|\,0\rangle$ in Minkowski space.

1.2.6 Schwinger–Dyson equations (continued)

The set of the Schwinger–Dyson equations for an interacting theory can be derived analogously to the free case.

Let us consider the cubic interaction which is described by the action (1.2.22). Choosing again $F[\varphi]$ to be given by Eq. (1.2.29) and calculating the variation of the action (1.2.22), one gets

$$\left(-\partial^2+m^2\right)\left\langle\,\varphi\left(x\right)\varphi\left(y\right)\right\rangle + \frac{\lambda}{2}\left\langle\,\varphi^2\left(x\right)\varphi\left(y\right)\right\rangle \;=\; \delta^{(d)}\left(x-y\right). \tag{1.2.34}$$

Problem 1.18 Rederive Eq. (1.2.34) by analyzing Feynman diagrams.
Solution Let us introduce the Fourier-transformed 2- and 3-point Green functions

$$G(p) \;=\; \int d^d x\, \mathrm{e}^{-ipx}\left\langle\,\varphi(x)\varphi(0)\,\right\rangle, \tag{1.2.35}$$

and

$$G_3(p,q,-p-q) \;=\; \int d^d x d^d y\, \mathrm{e}^{-ipx-iqy}\left\langle\,\varphi(x)\varphi(y)\varphi(0)\,\right\rangle. \tag{1.2.36}$$

Let us also denote the free momentum-space propagator as

$$G_0(p) \;=\; \frac{1}{p^2+m^2}\,. \tag{1.2.37}$$

The perturbative expansion of G_3 starts from

$$G_3(p,q,-p-q) \;=\; -\,\lambda G_0(p)G_0(q)G_0(p+q)+\ldots\,. \tag{1.2.38}$$

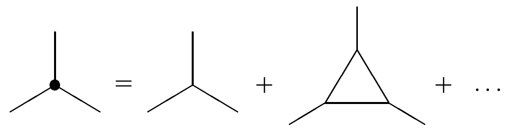
It is standard to truncate three external legs, introducing the vertex function

$$\Gamma(p,q,-p-q) \;=\; G_3(p,q,-p-q)G^{-1}(p)G^{-1}(q)G^{-1}(p+q) \tag{1.2.39}$$

whose perturbative expansion starts from λ:

$$\Gamma(p,q,-p-q) \;=\; -\,\lambda-\lambda^3\int\frac{d^d k}{(2\pi)^d}G_0(k-p)G_0(k)G_0(k+q)+\ldots\,. \tag{1.2.40}$$

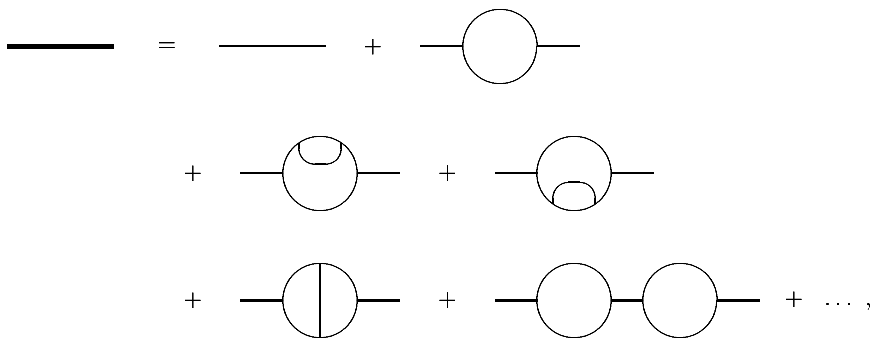
This expansion can be represented diagrammatically as



(1.2.41)

where the filled circle on the LHS represents the exact vertex and the thin lines are associated with the bare propagator (1.2.37).

An analogous expansion of the propagator is

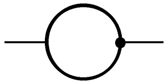


(1.2.42)

where the bold line represents the exact propagator. It is commonly rewritten as an equation for the self-energy $G_0^{-1}(p) - G^{-1}(p)$ which involves only the 1-particle irreducible (1PI) diagrams. The last diagram on the RHS of Eq. (1.2.42) is not 1PI.

Resumming the diagrams according to the definition (1.2.41) of the exact vertex Γ and the exact propagator G, the propagator equation can be represented graphically as

$$G_0(p) - G_0(p)G^{-1}(p)G_0(p) \;=\; \tag{1.2.43}$$



where the bold lines represent the exact propagator G while the external (thin) ones are associated with the bare propagator G_0. One vertex on the RHS of Eq. (1.2.43) is exact and the other is bare. It does not matter which one is exact and which one is bare since we can collect the diagrams of Eq. (1.2.42) into the exact vertex either on the left or on the right. Eq. (1.2.43) can be written analytically as

$$G_0^{-1}(p) - G^{-1}(p) \;=\; -\frac{\lambda}{2}\int \frac{d^d q}{(2\pi)^d} G(q)\Gamma(-q,p,q-p)G(p-q)\,. \tag{1.2.44}$$

Multiplying Eq. (1.2.44) by $G(p)$ and using the definition (1.2.36), we obtain exactly the Fourier transform of Eq. (1.2.34).

Notice that Eq. (1.2.34) is not closed. It relates the 2-point average (propagator) to the 3-point average (which is associated with a vertex). The closed set of the Schwinger–Dyson equations can be obtained for the n-point averages

$$G_n(x_1,x_2,\ldots,x_n) \;=\; \langle\,\varphi(x_1)\,\varphi(x_2)\ldots\varphi(x_n)\,\rangle\,. \tag{1.2.45}$$

They are also called the correlators, in analogy with statistical mechanics, or the n-point Green functions, in analogy with the Green functions in Minkowski space.

Choosing

$$F[\varphi] \;=\; \varphi(x_2)\ldots\varphi(x_n) \tag{1.2.46}$$

and calculating the variational derivative, one gets from Eq. (1.2.28) the following chain of equations

$$\begin{aligned}
&\left(-\partial^2+m^2\right)\langle\,\varphi(x)\,\varphi(x_2)\ldots\varphi(x_n)\,\rangle \\
&\quad+\frac{\lambda}{2}\langle\,\varphi^2(x)\,\varphi(x_2)\ldots\varphi(x_n)\,\rangle \\
&= \sum_{j=2}^{n}\delta^{(d)}(x-x_j)\left\langle\varphi(x_2)\ldots\underline{\varphi(x_j)}\ldots\varphi(x_n)\right\rangle
\end{aligned} \tag{1.2.47}$$

where $\underline{\varphi(x_j)}$ means that the corresponding term $\varphi(x_j)$ is missing in the product. Using the notation (1.2.45), Eq. (1.2.47) can be rewritten as

$$\begin{aligned}
&\left(-\partial^2+m^2\right)G_n(x,x_2,\ldots,x_n)+\frac{\lambda}{2}G_{n+1}(x,x,x_2,\ldots,x_n) \\
&= \sum_{j=2}^{n}\delta^{(d)}(x-x_j)\,G_{n-2}(x_2,\ldots,\underline{x_j},\ldots,x_n)
\end{aligned} \tag{1.2.48}$$

with $\underline{x_j}$ again standing for the missing argument.

Remark on connected correlators

The n-point correlators (1.2.45) include both connected and disconnected parts. The presence of disconnected parts is most easy to see in the free case when all connected parts disappear while G_n for an even n is given by the Wick pairing as is discussed in Subsect. 1.2.3. .

The correlators can be alternatively defined by introducing the partition function which is a functional of an external source $J(x)$:

$$Z[J] \;=\; \int D\varphi(x)\; \mathrm{e}^{-S+\int d^d x J(x)\varphi(x)}, \tag{1.2.49}$$

and varying with respect to the source. The n-point functions (1.2.45) are then given by

$$G_n(x_1,\ldots,x_n) \;=\; \frac{1}{Z[J]}\frac{\delta}{\delta J(x_1)}\cdots\frac{\delta}{\delta J(x_n)}Z[J] \tag{1.2.50}$$

while the connected parts are given by

$$\langle\,\varphi(x_1)\ldots\varphi(x_n)\,\rangle_{\rm conn} \;=\; \frac{\delta}{\delta J(x_1)}\cdots\frac{\delta}{\delta J(x_n)}\ln Z[J]. \tag{1.2.51}$$

The point is that

$$W[J] \;=\; \ln Z[J] \tag{1.2.52}$$

involves only a set of connected diagrams while disconnected ones emerge in $Z[J]$ after the exponentiation. We already touched this property in Subsect. 1.2.3 to order λ^2. The functional $W[J]$ is called, for this reason, the generating functional for connected diagrams.

Remark on the LSZ reduction formula

The (analytically continued to Minkowski space) correlators $G_n(x_1, \ldots, x_n)$ determine an amplitude of the process when n, generally speaking, virtual particles produces k on-mass-shell particles. Let us denote this amplitude as $A_{n\to k}(q_1, \ldots, q_n; p_1, \ldots, p_k)$ where $q_1, \ldots, q_n$ and $p_1, \ldots, p_k$ are the four-momenta of the incoming and outgoing particles, respectively. The four-momentum conservation says

$$q_1 + \ldots + q_n = p_1 + \ldots + p_k \,. \tag{1.2.53}$$

The Lehman–Symanzik–Zimmerman (LSZ) reduction formula reads

$$\begin{aligned} & A_{n\to k}(q_1, \ldots, q_n; p_1, \ldots, p_k) \\ & = \prod_{j=1}^{n} (q_j^2 + m^2) \prod_{i=1}^{k} \lim_{p_i^2 \to -m^2} (p_i^2 + m^2) \int \prod_{i=1}^{k} \frac{d^d p_i}{(2\pi)^d} \int \prod_{j=1}^{n-1} \frac{d^d q_j}{(2\pi)^d} \\ & \quad \cdot \mathrm{e}^{-i\sum_{i=1}^{k} p_i x_i + i\sum_{j=1}^{n-1} q_j x_{k+j}} G_n(x_1, \ldots, x_{n+k-1}, 0) \,. \end{aligned} \tag{1.2.54}$$

The unusual sign of the square of the particle mass m is due to the Euclidean metric.

Equation (1.2.54) makes sense for timelike p_i, when $p_i^2 < 0$, while q_j^2 is arbitrary. The amplitude for the case of on-mass-shell incoming particles is given by Eq. (1.2.54) with $q_j^2 \to -m^2$.

1.2.7 Regularization

The ultraviolet divergencies (*i.e.* those at small distances or large momenta) are an intrinsic property of QFT which makes it different from the quantum mechanics of a finite number of degrees of freedom. The divergencies emerge, roughly speaking, because of the delta-function in the canonical commutation relations.

The idea of regularization is to somehow smooth the effect of the delta-function. The usual procedures of regularization are to

1) smear the delta-function by

- point splitting,
- Schwinger proper-time regularization,
- latticizing;

2) add negative-norm regulating term to the action

- Pauli–Villars regularization;

3) introduce higher derivatives in the kinetic term[7];
4) change the dimension $4 \to 4-\epsilon$;
5) regularize the measure in the path integral.

As an example of point splitting, let us consider the regularization when the delta-function in the commutator term is replaced by

$$\delta^{(d)}(x-y) \stackrel{\text{Reg.}}{\Longrightarrow} \boldsymbol{R}\,\delta^{(d)}(x-y) = R(x,y)\,. \tag{1.2.55}$$

The regularizing operator $\boldsymbol{R}$ is for instance

$$\boldsymbol{R} = \mathrm{e}^{a^2(\partial^2 - m^2)}\,, \tag{1.2.56}$$

where the parameter a with the dimension of length plays the role of an ultraviolet cutoff. The cutoff disappears as $a \to 0$ when

$$\begin{aligned} \boldsymbol{R} &\to \boldsymbol{1}\,, \\ R(x,y) &\to \delta^{(d)}(x-y)\,. \end{aligned} \tag{1.2.57}$$

It is easy to calculate how the regularization (1.2.55) modifies the propagator. The result reads

$$\begin{aligned} G_R(x-y) &= \frac{1}{-\partial^2 + m^2} \boldsymbol{R}\,\delta^{(d)}(x-y) \\ &= \frac{1}{2} \int_{a^2}^{\infty} d\tau\, \mathrm{e}^{-\frac{1}{2}\tau m^2}\, \mathrm{e}^{-\frac{(x-y)^2}{2\tau}} \frac{1}{(2\pi\tau)^{d/2}}. \end{aligned} \tag{1.2.58}$$

The lower limit in the integral over the proper time τ is now a^2 rather than 0 as in the non-regularized expression (1.1.85). This particular way of point splitting coincides with the Schwinger proper-time regularization.

A regularization via point splitting can be done non-perturbatively while the dimensional regularization, which is listed in item 4, is defined only in the framework of perturbation theory. The regularization of the measure, which is listed in item 5, will be considered in the next Section.

When a regularization is introduced, some of the first principles (called the axioms), on which quantum field theory is built on, are violated. For instance, the regularization via the point splitting (1.2.55) violates locality.

[7]That is in the quadratic-in-fields part of the action.

1.3 Quantum anomalies from path integral

As is well-known, the Lagrangian approach in classical field theory is very useful for constructing conserved currents associated with symmetries of the Lagrangian. Noether's theorems[8] say how to construct corresponding currents and when they are conserved.

An analogous approach in quantum field theory is based on path integrals over fields. It naturally incorporates the classical results since the weight in the path integral is given by the classical action.

However, anomalous (*i.e.* additional to classical) terms in the divergencies of currents can appear in the quantum case due to a contribution from regulators which make the theory finite in the ultraviolet limit. They are called quantum anomalies.

In this Section we first consider the chiral anomaly, *i.e.* the quantum anomaly in the divergence of the axial current, which appears in the path-integral approach as a result of the non-invariance of the regularized measure. Then we briefly repeat the analysis for the scale anomaly, *i.e.* the quantum anomaly in the divergence of the dilatation current.

1.3.1 QED via path integral

Let us restrict ourselves to the case of *quantum electrodynamics* (QED), though most of the formulas will be valid for a non-Abelian Yang–Mills theory as well.

QED is described by the following partition function

$$Z = \int DA_\mu \int D\bar\psi D\psi \, \mathrm{e}^{-S[A,\psi,\bar\psi]} \tag{1.3.1}$$

where A_μ is the vector-potential of the electromagnetic field, ψ_i and $\bar\psi_i$ are the Grassmann variables which describe the electron-positron field with i being the spinor index. They are independent but are interchangeable under involution

$$\psi \overset{\text{inv.}}{\longleftrightarrow} \bar\psi \,, \tag{1.3.2}$$

which is defined such that[9]

$$\psi_1\psi_2 \xrightarrow{\text{inv.}} \bar\psi_2\bar\psi_1 \,. \tag{1.3.3}$$

[8]See, *e.g.*, §2 of the book [BS76].

[9]See the book by Berezin [Ber86] (§3.5 of Part I). Sometimes involution is defined with an opposite sign (*i.e.* $\bar\psi$ is substituted by $i\bar\psi$) which results in a multiplication of the fermionic part of the action (1.3.4) by an extra factor of i.

In particular, $\bar\psi\psi$ is invariant under involution. Therefore, $\bar\psi$ is an analogue of $i\bar{\boldsymbol{\psi}} = i\boldsymbol{\psi}^\dagger\gamma_0$ in the operator formalism while involution is analogous to Hermitean conjugation.

The Euclidean QED action in Eq. (1.3.1) reads

$$S[A,\psi,\bar\psi] = \int d^dx \left(\bar\psi\gamma_\mu\nabla_\mu\psi + m\bar\psi\psi + \frac{1}{4}F^2_{\mu\nu}\right), \tag{1.3.4}$$

where

$$\nabla_\mu = \partial_\mu - ieA_\mu(x) \tag{1.3.5}$$

is the covariant derivative as before,

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu \tag{1.3.6}$$

is the field strength, and γ_μ are the Euclidean γ-matrices which are discussed in Subsect. 1.1.2.

1.3.2 Chiral Ward identity

Let us perform the local *chiral transformation*

$$\begin{aligned} \psi(x) &\xrightarrow{\text{c.t.}} \psi'(x) = \mathrm{e}^{\,i\alpha(x)\gamma_5}\psi(x)\,, \\ \bar\psi(x) &\xrightarrow{\text{c.t.}} \bar\psi'(x) = \bar\psi(x)\,\mathrm{e}^{\,i\alpha(x)\gamma_5}. \end{aligned} \tag{1.3.7}$$

Here the parameter of the transformation $\alpha(x)$ is a function of x and γ_5 is the Hermitean matrix

$$\gamma_5 = \gamma_1\gamma_2\gamma_3\gamma_4 \tag{1.3.8}$$

in $d = 4$. Note that both ψ and $\bar\psi$ have the same transformation law since in Minkowski space

$$\bar\psi = \psi^\dagger\gamma_0\,, \qquad \gamma_5^\dagger = \gamma_5\,, \quad \gamma_0^\dagger = \gamma_0 \tag{1.3.9}$$

while γ_5 and γ_0 anticommute.

The variation of the classical action (1.3.4) under the chiral transformation (1.3.7) reads

$$\delta S = \int d^dx \left(\partial_\mu\alpha(x)\, J^5_\mu(x) + 2im\alpha(x)\bar\psi(x)\gamma_5\psi(x)\right), \tag{1.3.10}$$

where the axial current

$$J^5_\mu \;=\; i\bar\psi\gamma_\mu\gamma_5\psi \tag{1.3.11}$$

is the Noether current associated with the chiral transformation.

It follows from Eq. (1.3.10) that the divergence of the axial current (1.3.11) reads

$$\partial_\mu J^5_\mu \;=\; 2im\bar\psi\gamma_5\psi, \tag{1.3.12}$$

so that it is conserved in the massless case ($m=0$) at the classical level:

$$\partial_\mu J^5_\mu \;\stackrel{m=0}{=}\; 0. \tag{1.3.13}$$

Problem 1.19 Verify Eq. (1.3.12) using the classical Dirac equation

$$\left(\hat\nabla + m\right)\psi(x) \;=\; 0\,, \qquad \hat\nabla \;=\gamma_\mu\nabla_\mu\,. \tag{1.3.14}$$

Solution Calculate straightforwardly the divergence of the axial current (1.3.11) using Eq. (1.3.14) and the conjugate one

$$\bar\psi(x)\left(\overleftarrow{\hat\nabla}\; - m\right) \;=\; 0 \tag{1.3.15}$$

with

$$\overleftarrow{\nabla}_\mu \;=\; \overleftarrow{\partial}_\mu + ieA_\mu(x)\,. \tag{1.3.16}$$

Let us now discuss how the measure in the path integral changes under the chiral transformation (1.3.7). The old and new measures are related by

$$D\bar\psi D\psi \;=\; D\bar\psi' D\psi' \,\det\left[\mathrm{e}^{\,2i\alpha(x)\gamma_5}\delta^{(d)}\left(x-y\right)\right], \tag{1.3.17}$$

where the determinant is over the space indices x and y as well as over the γ-matrix indices i and j. Notice that the determinant, which is nothing but the Jacobian of the transformation (1.3.7), emerges for the Grassmann variables to the positive rather than negative power as for commuting variables. This is a known property of the integrals (1.2.17) over Grassmann variables [Ber86] which look more like derivatives.

The logarithm of the Jacobian in Eq. (1.3.17) can be calculated as

$$\begin{aligned}
\ln\det\left[\mathrm{e}^{\,2i\alpha(x)\gamma_5}\delta^{(d)}\left(x-y\right)\right] \;&=\; \mathrm{Sp}\ln\left[\mathrm{e}^{\,2i\boldsymbol{\alpha}\gamma_5}\right] \\
=\;\; \mathrm{Sp}\left[2i\,\boldsymbol{\alpha}\,\gamma_5\right] \;&=\; 2i\int d^dx\,\alpha\left(x\right)\,\delta^{(d)}\left(0\right)\,\mathrm{sp}\,\gamma_5 \;=\; 0\,,
\end{aligned} \tag{1.3.18}$$

where sp is the trace only over the γ-matrix indices i and j. The RHS vanishes naively since the trace vanishes. A subtlety with the appearance of the infinite factor $\delta^{(d)}(0)$ will be discussed in the next Subsection.

Note that the infinitesimal version the transformation (1.3.7) is a particular case of the more general one

$$\begin{aligned}
\psi\left(x\right) \quad&\longrightarrow\quad \psi'\left(x\right) = \psi\left(x\right) + \delta\psi\left(x\right), \\
\bar\psi\left(x\right) \quad&\longrightarrow\quad \bar\psi'\left(x\right) = \bar\psi\left(x\right) + \delta\bar\psi\left(x\right),
\end{aligned} \tag{1.3.19}$$

which is an analog of the transformation (1.2.25) and leaves the measure invariant. The calculation given in Eq. (1.3.18) is an explicit illustration of this fact.

The general transformation (1.3.19) leads, when applied to the path integral in Eq. (1.3.1), to the Schwinger–Dyson equations

$$\begin{aligned}
\left(\hat\nabla + m\right)\psi(x) \quad&\stackrel{\text{w.s.}}{=}\quad \frac{\delta}{\delta\bar\psi(x)}\,, \\
\bar\psi(x)\left(\overleftarrow{\hat\nabla}\; - m\right) \quad&\stackrel{\text{w.s.}}{=}\quad \frac{\delta}{\delta\psi(x)}\,,
\end{aligned} \tag{1.3.20}$$

which hold in the weak sense, *i.e.* under the sign of averaging over $\bar\psi$ and ψ.

More restrictive transformations of the type of (1.3.7), which are associated with symmetries of the classical action and result in conserved currents, lead to some (less restrictive) relations between correlators which are called *Ward identities.* This terminology goes back to the fifties when a proper relation between the 2- and 3-point Green functions was first derived for the gauge symmetry in QED.

The simplest Ward identity, which is associated with the chiral transformation (1.3.7), reads

$$\begin{aligned}
&\left\langle \partial_\mu J^5_\mu(0)\psi_i(x)\bar\psi_j(y)\right\rangle \\
&\quad\stackrel{m=0}{=}\; i\,\delta^{(d)}(x)\left\langle (\gamma_5\psi)_i\,(0)\bar\psi_j(y)\right\rangle - i\,\delta^{(d)}(y)\left\langle \psi_i(x)\left(\bar\psi\gamma_5\right)_j(0)\right\rangle.
\end{aligned} \tag{1.3.21}$$

It is clear from the way Eq. (1.3.21) was derived, that it is always satisfied as a consequence of the quantum equations of motion (1.3.20).

Problem 1.20 Derive Eq. (1.3.21) in the operator formalism when the averages are substituted by the vacuum expectation values of the T–products.

Solution Use Eq. (1.3.14) (which acquires an extra $-i$ in Minkowski space, where the spatial γ-matrices are anti-Hermitean rather than Hermitean as in Euclidean space, and holds in the quantum case in the weak sense, *i.e.* when applied to a state) and canonical equal-time anticommutation relations for ψ and $\bar{\psi}$ with the only non-vanishing anticommutator being

$$\delta(x_0 - y_0)\left\{\bar{\boldsymbol{\psi}}_i(y), \boldsymbol{\psi}_j(x)\right\} \;=\; \delta_{ij}\delta^{(d)}(x-y)\,. \tag{1.3.22}$$

Remark on γ_5 in d dimensions

Let us recall that

$$\gamma_5 \;=\; \gamma_1\gamma_2\cdots\gamma_d \tag{1.3.23}$$

exists only for an even d when the size of the γ-matrices is $2^{d/2}\times 2^{d/2}$. For this reason the dimensional regularization is not applicable in calculations of the chiral anomaly.

Remark on gauge-fixing

Note that we did not add a gauge-fixing term to the action (1.3.4). It is harmless to do that since the gauge-fixing term does not contribute to the variation of the action under the chiral transformation. Moreover, all gauge invariant quantities do not depend on the gauge fixing. How to quantize a gauge theory without adding a gauge-fixing term will be explained in the next Chapter.

1.3.3 Chiral anomaly

As is already mentioned, Eq. (1.3.18) involves the uncertainty

$$\delta^{(d)}(0)\cdot \mathrm{sp}\,\gamma_5 \;=\; \infty\cdot 0\,. \tag{1.3.24}$$

To regularize $\delta^{(d)}(0)$, one needs [Ver78, Fuj79] to regularize the measure in the path integral over ψ and $\bar{\psi}$, since this term comes from the change of the measure under the chiral transformation.

Let us expand the fields ψ and $\bar{\psi}$ over some set of the orthogonal basis functions, similarly to how it is done in Eq. (1.1.78):

$$\psi^i(x) \;=\; \sum_n c_n^i \phi_n^i(x)\,, \qquad \bar{\psi}^i(x) \;=\; \sum_n \bar{c}_n^i \phi_n^{i\,\dagger}(x)\,, \tag{1.3.25}$$

where there is no summation over the spinor index i. Here c_n^i and $\bar{c}_n^i$ are Grassmann variables. The measure is then similar to that of Eq. (1.1.79) and

reads explicitly

$$D\bar{\psi}D\psi \;=\; \prod_{n=1}^{\infty}\prod_i d\bar{c}_n^i \prod_{m=1}^{\infty}\prod_j dc_m^j\,. \tag{1.3.26}$$

The idea of regularizing the measure is to restrict ourselves to a large but finite number of the basis functions. This is analogous to the discretization of Subsect. 1.1.4. We therefore define the regularized measure as

$$(D\bar{\psi})_R(D\psi)_R \;=\; \prod_{n=1}^{M}\prod_i d\bar{c}_n^i \prod_{m=1}^{M}\prod_j dc_m^j\,. \tag{1.3.27}$$

The change of the measure under the chiral transformation reads

$$\begin{aligned}&(D\bar{\psi})_R(D\psi)_R\\ &\;=\; (D\bar{\psi}')_R(D\psi')_R \det\left[\int d^dx\,\phi_n^{k\,\dagger}(x)\,\mathrm{e}^{\,2i\alpha(x)\gamma_5^{kj}}\,\phi_m^j(x)\right],\end{aligned} \tag{1.3.28}$$

where the determinant is both over the n and m indices and over the spinor indices k and j. This is the regularized analog of the nonregularized expression (1.3.17).

Using the orthogonality of the basis functions:

$$\int d^dx\,\phi_n^{j\,\dagger}(x)\phi_m^i(x) \;=\; \delta_{nm}\delta^{ij}\,, \tag{1.3.29}$$

and Eq. (1.2.20), we rewrite the determinant on the RHS of Eq. (1.3.28) for an infinitesimal parameter α as

$$\begin{aligned}&\det\left[\int d^dx\,\phi_n^{k\,\dagger}(x)\,\mathrm{e}^{\,2i\alpha(x)\gamma_5^{kj}}\,\phi_m^j(x)\right]\\ &\;=\; 1+2i\sum_{n=1}^{M}\int d^dx\,\phi_n^\dagger(x)\alpha\,(x)\,\gamma_5\phi_n(x)\,,\end{aligned} \tag{1.3.30}$$

where the spinor indices are contracted in the usual way.

It is easy to see how this formula recovers Eq. (1.3.18) since

$$\sum_{n=1}^{\infty}\phi_n^i(x)\phi_n^{j\,\dagger}(y) \;=\; \delta^{(d)}(x-y)\,\delta^{ij} \tag{1.3.31}$$

in the nonregularized case due to the completeness of the basis functions.

In the regularized case, the sum over n on the LHS of Eq. (1.3.31) is restricted by M from above so that the RHS is no longer equal to the delta-function. We substitute

$$\sum_{n=1}^{M} \phi_n^i(x)\phi_n^{j\,\dagger}(y) \;=\; R^{ij}(x,y)\,, \tag{1.3.32}$$

with the RHS being the matrix element of some regularizing operator $\boldsymbol{R}$.

It can be chosen in many ways. We shall work with several forms:

$$\boldsymbol{R} \;=\; \mathrm{e}^{\,a^2\widehat{\nabla}^2}\,, \qquad \text{or} \tag{1.3.33}$$

$$\boldsymbol{R} \;=\; \frac{1}{1-a^2\widehat{\nabla}^2}\,, \qquad \text{or} \tag{1.3.34}$$

$$\boldsymbol{R} \;=\; \frac{1}{1+a\widehat{\nabla}}\,, \qquad \ldots\,, \tag{1.3.35}$$

where again $\widehat{\nabla} = \gamma_\mu \nabla_\mu$. The parameter a is the ultraviolet cutoff. The cutoff disappears as $a \to 0$ when Eq. (1.2.57) holds.

These regularizations (1.3.33) – (1.3.35) are non-perturbative, and preserve gauge invariance since they are constructed from the covariant derivative ∇_μ. A consistent regularization occurs when $\boldsymbol{R}$ commutes with $\widehat{\nabla}$, which is obviously true for the regularizations (1.3.33) – (1.3.35)[10].

Therefore, we get

$$\begin{aligned}\int d^dx\,\alpha\,(x)\,\partial_\mu J_\mu^5 \;&=\; 2i\,\mathrm{Sp}\;\boldsymbol{\alpha}\gamma_5\boldsymbol{R} \\ &=\; 2i\int d^dx\,\alpha\,(x)\,\mathrm{sp}\,\gamma_5 R\,(x,x)\,. \end{aligned} \tag{1.3.36}$$

It is worth noting that an extra R in Eq. (1.3.36) is a consequence of a more general formula

$$\mathrm{Sp}\,\boldsymbol{O} \;\longrightarrow\; \mathrm{Sp}\,\boldsymbol{OR}\,, \tag{1.3.37}$$

which says how to regularize traces of operators.

[10] This can be shown by choosing the basis functions to be eigenfunctions of the Hermitean operator $i\widehat{\nabla}$ ($i\widehat{\nabla}\phi_n = E_n\phi_n$) and applying $\widehat{\nabla}^{ki}(x)(\widehat{\nabla}^{-1}(y))^{jl}$ to Eq. (1.3.32). The LHS then does not change (because $E_nE_n^{-1}=1$) while $\langle x|\widehat{\nabla}\boldsymbol{R}\widehat{\nabla}^{-1}|y\rangle$ appears on the RHS. It coincide with the RHS of Eq. (1.3.32) when $\widehat{\nabla}$ and $\boldsymbol{R}$ commute.

Remark on regularization of the measure

The regularization of the measure in the path integral by Eq. (1.3.27) is equivalent to the point-splitting procedure where the delta-function in the commutator term is smeared according to Eq. (1.2.55).

To show this, let us note that the variational derivative can be approximated for the finite number of basis functions by

$$\frac{\delta}{\delta\psi_R^j(y)} \;=\; \sum_{n=1}^{M}\phi_n^{j\,\dagger}(y)\sum_k\frac{\partial}{\partial c_n^k}\,. \tag{1.3.38}$$

This definition extends the standard mathematical one[11] to the case of spinor indices. The sum over k is included in order for the regularized variational derivative to feel variations of all the spinor components of c_n when the variation is not diagonal in the spinor indices.

When applied to

$$\psi_R^i(x) \;=\; \sum_{n=1}^{M} c_n^i\phi_n^i(x)\,, \tag{1.3.39}$$

it yields

$$\frac{\delta\psi_R^i(x)}{\delta\psi_R^j(y)} \;=\; \sum_{n=1}^{M}\phi_n^i(x)\phi_n^{j\,\dagger}(y) \;=\; R^{ij}(x,y)\,, \tag{1.3.40}$$

or, equivalently,

$$\delta^{ij}\delta^{(d)}(x-y) \;\overset{\text{Reg.}}{\Longrightarrow}\; R^{ij}(x,y)\,. \tag{1.3.41}$$

which is a fermionic analog of Eq. (1.2.55).

Thus, we conclude that the regularization of the measure in the path integral is equivalent to smearing the delta-function in commutator terms.

Remark on regularized Schwinger–Dyson equations

According to Subsect. 1.2.5, the procedure from the previous paragraph results in the following regularized Schwinger–Dyson equations

$$\begin{aligned}\left(\widehat{\nabla}+m\right)\psi(x) \;&\overset{\text{w.s.}}{=}\; \int d^dyR(x,y)\frac{\delta}{\delta\bar{\psi}(y)}\,, \\ \bar{\psi}(x)\left(\overleftarrow{\widehat{\nabla}}-m\right) \;&\overset{\text{w.s.}}{=}\; \int d^dyR(x,y)\frac{\delta}{\delta\psi(y)}\,. \end{aligned} \tag{1.3.42}$$

[11] See, *e.g*, the book by Lévy [Lev51].

These equations are understood again in the weak sense, *i.e.* under the sign of averaging over $\bar\psi$ and ψ and obviously reproduce Eq. (1.3.20) as $a \to 0$.

Problem 1.21 Derive Eq. (1.3.36) using the regularized Schwinger–Dyson equation (1.3.42).
Solution The calculation is similar to that of Problem 1.19 except for the additional terms which are due to the RHS of Eq. (1.3.42). One gets for $m = 0$

$$\begin{aligned}\partial_\mu J^5_\mu &\overset{\text{w.s.}}{=} i \int d^d y \frac{\delta}{\delta\psi(y)} R(x,y)\gamma_5\psi(x) - i\bar\psi(x)\gamma_5 \int d^d y R(x,y) \frac{\delta}{\delta\bar\psi(y)} \\ &= 2i \,\mathrm{sp}\, \gamma_5 R\,(x,x)\,, \end{aligned} \tag{1.3.43}$$

which is equivalent to Eq. (1.3.36) since the $\alpha(x)$ there is an arbitrary function.

1.3.4 Chiral anomaly (calculation)

In order to derive an explicit expression for the chiral anomaly, we should calculate the RHS of Eq. (1.3.36) for some choice of the regularizing operator $\boldsymbol{R}$. Let us choose $\boldsymbol{R}$ given by Eq. (1.3.34). The operator $\widehat\nabla^2$ in the denominator can be transformed as

$$\begin{aligned}\widehat\nabla^2 &= \nabla^2 + \frac{1}{2}\left[\gamma_\mu,\gamma_\nu\right]\nabla_\mu\nabla_\nu \\ &= \nabla^2 - \frac{ie}{4}\left[\gamma_\mu,\gamma_\nu\right] F_{\mu\nu} \\ &= \nabla^2 + \frac{e}{2}\Sigma_{\mu\nu}F_{\mu\nu}\,,\end{aligned} \tag{1.3.44}$$

where the trace of the spin matrices

$$\Sigma_{\mu\nu} = \frac{1}{2i}\left[\gamma_\mu,\gamma_\nu\right] \tag{1.3.45}$$

is given by

$$\mathrm{sp}\ \Sigma_{\mu\nu}\Sigma_{\lambda\rho}\gamma_5 = -4\epsilon_{\mu\nu\lambda\rho}\,. \tag{1.3.46}$$

Expanding in e:

$$\boldsymbol{R} = \boldsymbol{R}_0 + \boldsymbol{R}_0\,(\ldots)\,\boldsymbol{R}_0 + \ldots \tag{1.3.47}$$

with

$$\boldsymbol{R}_0 = \frac{1}{1 - a^2\partial^2}\,, \tag{1.3.48}$$

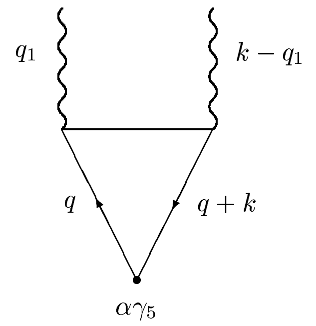


Fig. 1.6: Triangular diagram associated with chiral anomaly in $d = 4$. The solid lines correspond to R_0 given by Eq. (1.3.51). The wavy lines correspond to the field strength.

we get schematically

$$\begin{aligned}\mathrm{Sp}\,\boldsymbol{\alpha}\gamma_5\boldsymbol{R} &= a^4\mathrm{Sp}\,\boldsymbol{\alpha}\gamma_5\boldsymbol{R}_0\left(\frac{e\Sigma F}{2}\right)\boldsymbol{R}_0\left(\frac{e\Sigma F}{2}\right)\boldsymbol{R}_0 \\ &= -\int d^4x\,\alpha\,(x)\,\frac{e^2}{16\pi^2}F_{\mu\nu}\tilde F_{\mu\nu}\,,\end{aligned} \tag{1.3.49}$$

where

$$\tilde F_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\lambda\rho}F_{\lambda\rho} \tag{1.3.50}$$

is the dual field strength.

The calculation described in Eq. (1.3.49) is most easy to do in momentum space where it is associated with one-loop diagrams. The analytic expression to be calculated can be represented in $d = 4$ graphically by the triangular diagram in Fig. 1.6. The solid lines are associated with R_0 given by Eq. (1.3.48), which reads in momentum space as

$$R_0(p) = \frac{1}{1 + a^2p^2}\,. \tag{1.3.51}$$

The wavy lines correspond to the field strength. The lower vertex is associated with $\alpha\,\gamma_5$.

The integral over the four-momentum q, which circulates along the triangular loop, can be easily calculated by introducing $\omega = aq$ and transforming the integral as

$$\int d^4q\, f(q) \quad\to\quad \frac{1}{a^4}\int d^4\omega\, f\left(\frac{\omega}{a}\right). \tag{1.3.52}$$

Notice that the integral involves a^{-4} which cancels a^4 coming from the expansion in e whose proper term is given by the intermediate expression in Eq. (1.3.49). Therefore, the result is non-vanishing and a independent as $a \to 0$. Higher terms of the expansion in e are proportional to higher powers in a and vanish as $a \to 0$.

From Eqs. (1.3.36) and (1.3.49), we get finally

$$\partial_\mu J^5_\mu \;=\; -\frac{ie^2}{8\pi^2} F_{\mu\nu}\tilde{F}_{\mu\nu}\,. \tag{1.3.53}$$

The anomaly on the RHS is known as the Adler–Bell–Jackiw anomaly. Its appearance is usually related to the fact that any regularization can not be simultaneously gauge and chiral invariant.

Problem 1.22 Calculate the coefficient in Eq. (1.3.49) and show that it is regulator independent.
Solution The contribution of the triangular diagram of Fig. 1.6, which represents the intermediate expression in Eq. (1.3.49), reads explicitly

$$\begin{aligned} 2\,\mathrm{Sp}\,\alpha\gamma_5 R \;=\; & -4e^2a^4\int d^4x\int d^4y\int d^4z \\ & \cdot\,\alpha(x)R_0(x,y)F_{\mu\nu}(y)R_0(y,z)\tilde{F}_{\mu\nu}(z)R_0(z,x)\,. \end{aligned} \tag{1.3.54}$$

In momentum space, it becomes

$$\begin{aligned} & -2e^2a^4\int d^4x\,\alpha(x)\int\frac{d^4k}{(2\pi)^4}\,\mathrm{e}^{ikx}\int\frac{d^4q_1}{(2\pi)^4}F_{\mu\nu}(q_1)\tilde{F}_{\mu\nu}(k-q_1) \\ & \cdot\int\frac{d^4q}{(2\pi)^4}\,\frac{1}{(1+a^2q^2)\,(1+a^2(q+q_1)^2)\,(1+a^2(q+k)^2)} \overset{(1.3.52)}{=} \\ & -\frac{2e^2}{16\pi^2}\int d^4x\,\alpha(x)\int\frac{d^4k}{(2\pi)^4}\,\mathrm{e}^{ikx}\int\frac{d^4q_1}{(2\pi)^4}F_{\mu\nu}(q_1)\tilde{F}_{\mu\nu}(k-q_1)\int_0^\infty\frac{\omega^2d\omega^2}{(1+\omega^2)^3} \end{aligned} \tag{1.3.55}$$

which recovers the RHS of Eq. (1.3.49).

An analogous calculation can be repeated for other regulators (1.3.33) and (1.3.35). Let us denote

$$r(a^2p^2) \;\equiv\; R_0(p)\,. \tag{1.3.56}$$

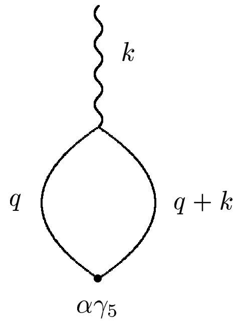


Fig. 1.7: The diagram associated with chiral anomaly in $d = 2$. The solid lines correspond to R_0 given by Eq. (1.3.51). The wavy line corresponds to the field strength.

Then the only difference with Eq. (1.3.55) is that the last integral over ω^2 is replaced by

$$\int_0^\infty \omega^2 d\omega^2 r''(\omega^2) \;=\; r(0) \;=\; 1 \tag{1.3.57}$$

for reasonable functions r which look like those given by Eqs. (1.3.33) – (1.3.35).

An anomaly which is analogous to the Adler–Bell–Jackiw anomaly (1.3.53) exists in $d = 2$ where

$$\partial_\mu J^5_\mu \;=\; -\frac{e}{2\pi}\epsilon_{\mu\nu}F_{\mu\nu}\,. \tag{1.3.58}$$

This anomaly is given by the diagram depicted in Fig. 1.7. It involves only two lines with the regulator $R_0(p)$ since in $d = 2$

$$\int d^2q\, f(q) \quad\to\quad \frac{1}{a^2}\int d^2\omega\, f\left(\frac{\omega}{a}\right) \tag{1.3.59}$$

so that all terms with more lines vanish as $a \to 0$.

Problem 1.23 Calculate $2\,\mathrm{Sp}\,\alpha\gamma_5 R$ in $d = 2$.
Solution Proceeding as before, we see that only the diagram of Fig. 1.7 is essential in $d = 2$ which yields

$$2iea^2\int d^2x\int d^2y\,\alpha(x)R_0(x,y)F_{\mu\nu}(y)\epsilon_{\mu\nu}R_0(y,x) \;=$$

$$2iea^2\int d^2x\alpha(x)\int\frac{d^2k}{(2\pi)^2}\,\mathrm{e}^{\,ikx}F_{\mu\nu}(k)\epsilon_{\mu\nu}\int\frac{d^2q}{(2\pi)^2}\frac{1}{(1+a^2q^2)\,(1+a^2(q+k)^2)}$$

$$\stackrel{(1.3.59)}{=}2\frac{ie}{4\pi}\int d^2x\alpha(x)\int\frac{d^2k}{(2\pi)^2}\,\mathrm{e}^{\,ikx}F_{\mu\nu}(k)\epsilon_{\mu\nu}\int_0^\infty\frac{d\omega^2}{(1+\omega^2)^2}$$

$$=\int d^2x\alpha(x)\frac{ieF_{\mu\nu}(x)\epsilon_{\mu\nu}}{2\pi}. \qquad (1.3.60)$$

The linear-in-$F_{\mu\nu}$ term is nonvanishing since

$$\mathrm{sp}\,\Sigma_{\mu\nu}\gamma_5 \;=\; 2i\epsilon_{\mu\nu} \qquad (1.3.61)$$

in $d=2$.

The result is again regulator independent since the integral over ω is replaced for an arbitrary $R_0(p)$ by

$$-\int_0^\infty d\omega^2 r'(\omega^2) \;=\; r(0) \;=\; 1 \qquad (1.3.62)$$

where Eq. (1.3.56) has been used.

Remark on non-Abelian chiral anomaly

Equation (1.3.53) holds also in the case of a non-Abelian gauge group with $F^a_{\mu\nu}$ being the non-Abelian field strength

$$F^a_{\mu\nu}(x)=\partial_\mu A^a_\nu(x)-\partial_\nu A^a_\mu(x)+gf^{abc}A^b_\mu(x)A^c_\nu(x)\,. \qquad (1.3.63)$$

Here f^{abc} are the structure constants of the gauge group and g is the coupling constant. The non-Abelian analog of Eq. (1.3.53) for the axial current, which is a singlet w.r.t. the gauge group, reads[12]

$$\partial_\mu J^5_\mu \;=\; -\frac{ig^2}{16\pi^2}\sum_a F^a_{\mu\nu}\tilde F^a_{\mu\nu}\,. \qquad (1.3.64)$$

The coefficient in this formula differs by 1/2 from that in Eq. (1.3.53) which is due to different normalizations in the Abelian and non-Abelian cases. This will be explained in more detail in Subsect. 2.1.1 below.

The $d=2$ anomaly (1.3.58) exists for the singlet axial current only in the Abelian case.

[12] See Ref. [Mor86] for a review of anomalies in gauge theories.

1.3.5 Scale anomaly

The scale transformation is defined by

$$x_\mu \;\longrightarrow\; x'_\mu=\rho\, x_\mu\,, \qquad (1.3.65)$$

$$\varphi\,(x) \;\longrightarrow\; \varphi'\,(x')=\rho^{l_\varphi}\varphi\,(x')\,. \qquad (1.3.66)$$

The index l_φ is called the *scale dimension* of the field φ. The value of l_φ in a free theory is called the canonical dimension, which equals $(d-2)/2$ for bosons (scalar or vector fields) and $(d-1)/2$ for the spinor Dirac field, *i.e.* 1 and 3/2 in $d=4$, respectively. Sometimes l_φ is called, for historical reasons, the anomalous dimension. More often the term "anomalous dimension" is used for the difference between l_φ and the canonical value.

The proper Noether current, which is called the *dilatation current*, is expressed via the energy-momentum tensor $\theta_{\mu\nu}$ as

$$D_\mu \;=\; x_\nu\theta_{\mu\nu} \qquad (1.3.67)$$

so that its divergence equals the trace of the energy-momentum tensor over the spatial indices:

$$\partial_\mu D_\mu \;=\; \theta_{\mu\mu}\,, \qquad (1.3.68)$$

since the energy-momentum tensor is conserved. For the action (1.3.4) one gets

$$\theta_{\mu\mu} \;=\; -m\bar\psi\psi \qquad (1.3.69)$$

at the classical level.

The above formulas can be obtained from the Noether theorems which state

$$\delta S \;=\; \int d^dx\,\rho\,(x)\,\partial_\mu D_\mu(x) \qquad (1.3.70)$$

or

$$\partial_\mu D_\mu(x) \;=\; \frac{\delta S}{\delta\rho\,(x)}\,. \qquad (1.3.71)$$

In the massless case, $m=0$, the RHS of Eq. (1.3.69) vanishes and the dilatation current is conserved. This is a well-known property of electrodynamics with massless electron which is scale invariant at the classical level. A generic scale-invariant theory does not depend on parameters of the dimension of mass or length. This usual dimension is to be distinguished from the scale dimension which is defined by Eq. (1.3.66). The dimensional parameters do not change under the scale transformation (1.3.65).

In the quantum case, the scale invariance is broken by the (dimensional) cutoff a. The energy-momentum tensor is no longer traceless due to loop effects. The relation (1.3.68) holds in the quantum case in the weak sense, *i.e.* for the averages

$$\left\langle \partial_\mu D_\mu F[A,\psi,\bar\psi]\right\rangle \;=\; \left\langle \theta_{\mu\mu} F[A,\psi,\bar\psi]\right\rangle , \tag{1.3.72}$$

where $F[A,\psi,\bar\psi]$ is a gauge-invariant functional of A, ψ and $\bar\psi$.

For a renormalizable theory like QED, the RHS of Eq. (1.3.72) is proportional to the Gell-Mann—Low function $\mathcal{B}(e^2)$ which is defined by

$$-\frac{a\,de^2}{da} \;=\; \mathcal{B}\left(e^2\right). \tag{1.3.73}$$

A non-trivial property of a renormalizable theory is that the RHS in this formula is a function solely of e^2 — the bare charge.

The meaning of the *renormalizability* is very simple: physical quantities do not depend on the cutoff a, provided the bare charge e is chosen to be cutoff dependent according to Eq. (1.3.73). This dependence of e on a effectively accounts for the distances smaller than a, which are excluded from the theory.

The precise relation between the trace of the energy-momentum tensor and the Gell-Mann–Low function reads

$$\theta_{\mu\mu} \;\stackrel{\text{w.s.}}{=}\; \frac{\mathcal{B}\left(e^2\right)}{4e^2} F^2_{\mu\nu}\,, \tag{1.3.74}$$

where the equality is understood again in the weak sense. This formula was first obtained in Refs. [Cre72, CE72] to leading order in e^2 and proven in Ref. [ACD77] to all orders in e^2.

Note that this formula holds in the operator formalism only when applied to a gauge invariant state. The reason is that otherwise a contribution from a gauge-fixing term in the action would be essential. It does not contribute, however, to gauge-invariant averages which can be formally proven using the gauge Ward identity.

Problem 1.24 Prove the relation (1.3.74).
Solution Let us absorb the coupling e into A_μ introducing

$$\begin{aligned} \mathcal{A}_\mu &= ieA_\mu\,, \\ \mathcal{F}_{\mu\nu} &= ieF_{\mu\nu}\,. \end{aligned} \tag{1.3.75}$$

The Lagrangian density of massless QED then reads

$$\mathcal{L} \;=\; \bar\psi\left(\hat\partial-\hat{\mathcal{A}}\right)\psi - \frac{1}{4e^2}\mathcal{F}^2\,. \tag{1.3.76}$$

To prove Eq. (1.3.74), let us use the chain of Eqs. (1.3.68) and (1.3.71). It is crucial that in the absence of other dimensional parameters the derivative $\partial/\partial\rho$ can be replaced by $\partial/\partial a$ since all dimensionless quantities depend in a theory with a cutoff only on the ratios of the type x/a.[13] Since the dependence on the cutoff a enters in Eq. (1.3.76) formally only via e^{-2} in front of $\mathcal{F}^2_{\mu\nu}$, Eq. (1.3.74) can be heuristically proven by first differentiating with respect to a and then expressing the result via $F_{\mu\nu}$ again. We have used herein the fact that $\mathcal{F}_{\mu\nu}$ is invariant under the renormalization-group transformation and, therefore, does not depend on a.

In the path-integral approach, a contribution to the scale anomaly comes from the regularized quantum measure. Proceeding as in Subsect. 1.3.3, we get

$$\partial_\mu D_\mu\left(x\right) \;=\; -\mathrm{sp}\,R(x,x)\,. \tag{1.3.77}$$

Problem 1.25 Derive Eq. (1.3.77) using the regularized Schwinger–Dyson equations (1.3.42).
Solution The energy-momentum tensor of QED reads

$$\theta_{\mu\nu} \;=\; F_{\mu\lambda}F_{\nu\lambda} - \frac14\delta_{\mu\nu}F^2_{\rho\lambda} + \frac14\left(\bar\psi\gamma_\mu \overleftrightarrow{\nabla}_\nu \psi + \bar\psi\gamma_\nu \overleftrightarrow{\nabla}_\mu \psi\right). \tag{1.3.78}$$

Taking the trace, one gets

$$\theta_{\mu\mu} \;=\; \frac12\bar\psi\gamma_\mu \overleftrightarrow{\nabla}_\mu \psi\,. \tag{1.3.79}$$

Using Eq. (1.3.42), it can be transformed as

$$\begin{aligned} \theta_{\mu\mu} &= -m\bar\psi\psi + \frac12\left(\int d^dy R(x,y)\frac{\delta}{\delta\psi(y)}\psi(x) - \bar\psi(x)\int d^dy R(x,y)\frac{\delta}{\delta\bar\psi(y)}\right) \\ &= -m\bar\psi\psi - \mathrm{sp}\,R\left(x,x\right), \end{aligned} \tag{1.3.80}$$

which reproduces Eq. (1.3.77) as $m\to 0$.

[13] This is the reason why the Callan–Symanzik equations, which are nothing but the dilatation Ward identities, coincide with the renormalization-group equations.

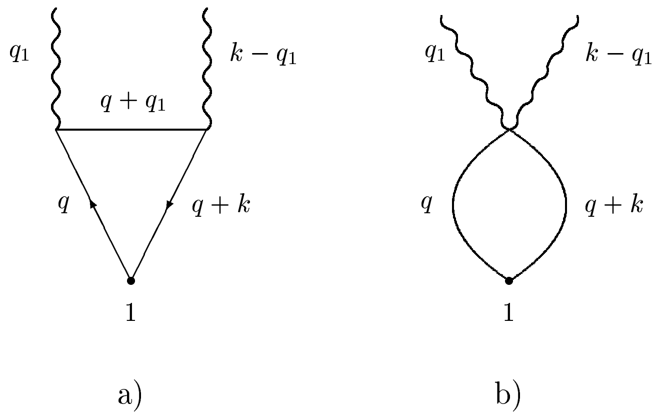


Fig. 1.8: The diagrams which contribute to the scale anomaly in $d=4$. The wavy line corresponds to the field strength.

To calculate the scale anomaly we should therefore perform a one-loop calculation of

$$\begin{aligned} \operatorname{sp} R(x,x) &= \operatorname{sp}\left\langle x \left| \frac{1}{1+a^2\left(i\hat{\nabla}^2\right)} \right| x \right\rangle \\ &= \operatorname{sp}\left\langle x \left| \frac{1}{1+a^2\left(i\partial_\mu+eA_\mu\right)^2-\frac{1}{2}a^2e\Sigma_{\mu\nu}F_{\mu\nu}} \right| x \right\rangle \end{aligned} \quad (1.3.81)$$

which is again most convenient to do in momentum space. The propagator is given by Eq. (1.3.48) while the vertices, which emerges in the corresponding Feynman rules for the expansion in e, come from the operators

$$-2iea^2A_\mu\partial_\mu\,, \qquad -e^2a^2A_\mu^2\,, \qquad \tfrac{1}{2}ea^2\Sigma_{\mu\nu}F_{\mu\nu}\,.$$

The only diagrams which survive as $a\to 0$ in $d=4$ are depicted in Fig. 1.8. The calculation of the diagram of Fig. 1.8a is the same as in Subsect. 1.3.3 while the diagram of Fig. 1.8b gives a total derivative which does not contribute to the scale anomaly.

The calculation of the diagram of Fig. 1.8a yields

$$\operatorname{sp} R(x,x) = -\frac{e^2F_{\mu\nu}^2(x)}{48\pi^2}\,. \quad (1.3.82)$$

The one-loop Gell-Mann–Low function can now be calculated using Eqs. (1.3.77) and (1.3.74), which reproduces the known result for QED. The higher-order corrections in e do not vanish for the scale anomaly.

Remark on non-Abelian scale anomaly

Equation (1.3.74) holds in the non-Abelian Yang–Mills theory as well if $F_{\mu\nu}$ is substituted by the non-Abelian field strength $F_{\mu\nu}^a$ given by Eq. (1.3.63). The corresponding formula, reads

$$\theta_{\mu\mu} \overset{\text{w.s.}}{=} \frac{\mathcal{B}\left(g^2\right)}{4g^2}\sum_a F_{\mu\nu}^a\tilde{F}_{\mu\nu}^a\,. \quad (1.3.83)$$

A heuristic proof, presented in Problem 1.24 for the Abelian case, can be repeated. The equality is again understood in the weak sense when averaged between gauge-invariant states. The contribution of gauge-fixing and ghost terms are then cancelled due to the gauge Ward identity which is called in this case the Slavnov–Taylor identity. The proof of Eq. (1.3.83) was given in Refs. [CDJ77, Nie77].

1.4 Instantons in quantum mechanics

Instantons are solutions of the classical equations of motion with a finite Euclidean action. Such field configurations are not taken into account in perturbation theory. Instantons are characterized by a topological charge which may result in a conserved quantum number and never shows up in perturbation theory. In Minkowski space, instantons are associated with tunneling processes between vacua labelled by a distinct topological charge.

Instantons first appear in Yang–Mills theory [BPST75] although this kind of classical solution was used long before in statistical physics [Lan67]. In this Section we consider instantons in quantum mechanics as an illustration of path-integral calculations. The consideration follows the original paper by Polyakov [Pol77] except for technical details.

1.4.1 Double-well potential

Let us consider a one-dimensional quantum-mechanical system with the double-well potential

$$V(x) = \frac{\lambda}{4}\left(x^2-\frac{\mu^2}{\lambda}\right)^2 = -\frac{1}{2}\mu^2x^2+\frac{1}{4}\lambda x^4+\frac{\mu^4}{4\lambda}. \tag{1.4.1}$$

This is nothing but an anharmonic oscillator with a wrong sign for the coefficient of the quadratic term,[14] which usually appears with a positive coefficient $\omega^2/2$. We have introduced

$$\mu^2 = -\omega^2 \tag{1.4.2}$$

in order to work with real numbered values. The constant term is added for later convenience. The potential (1.4.1) as a function x is depicted in Fig. 1.9.

The (Euclidean) action is defined by

$$S[x] = \int d\tau \left(\frac{1}{2}\dot{x}^2(\tau)+V(x(\tau))\right) \tag{1.4.3}$$

[14]It is often called the mass term. This terminology comes from quantum field theory, where the potential (1.4.1) is considered in the context of a spontaneous breaking of the reflection symmetry $x \to -x$. In our quantum-mechanical problem, defined by the Euclidean action (1.4.3), the mass of the non-relativistic particle is absorbed in τ which has, therefore, the dimension of [length]^2. This is explained already in Subsect. 1.1.6.

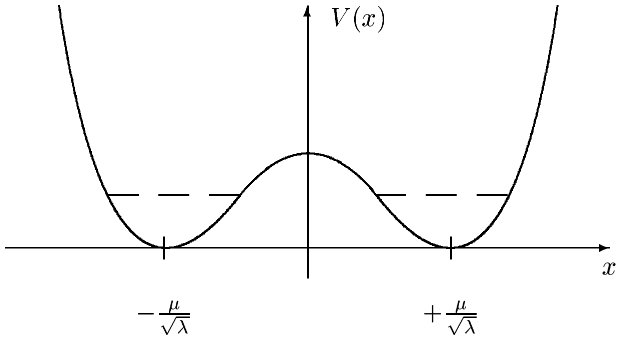


Fig. 1.9: The double-well potential (1.4.1). The short vertical lines represent the position of the minima (1.4.4). The dashed lines correspond to the energy E_0 of the lowest state in a single well, *i.e.* to that in the limit $\lambda \to 0$.

with $V(x)$ given by Eq. (1.4.1). The plus-sign between the kinetic and potential energies is because we are in Euclidean space.

It follows from Eqs. (1.4.1) and (1.4.3) that the parameter μ has the dimension of [length]$^{-2}$ or, in other words, the dimensions of x and τ are $[\mu]^{-1/2}$ and $[\mu]^{-1}$, respectively. Analogously, the dimension of the constant λ is $[\mu]^3$.

For $\lambda \ll \mu^3$, the potential (1.4.1) has superficially two degenerate vacua

$$x_0^{\pm} = \pm\frac{\mu}{\sqrt{\lambda}}, \tag{1.4.4}$$

whose positions coincide with the minima of the potential in Fig. 1.9.

The degeneracy between the two minima is preserved at all orders of perturbation theory, where an expansion near one of the minima of the potential (either the left or right one) is carried out:

$$x(\tau) = \pm\frac{\mu}{\sqrt{\lambda}}+\chi(\tau) \tag{1.4.5}$$

with $\chi(\tau) \ll \mu/\sqrt{\lambda}$. The correlator at asymptotically large τ is

$$\langle x(0)\,x(\tau)\rangle \to \frac{\mu^2}{\lambda}+\dots. \tag{1.4.6}$$

Its nonvanishing value means that a particle is localized at one of the two vacua.

The next terms of the perturbative expansion in λ do not spoil this result since the potential (1.4.1) becomes

$$V = \mu^2\chi^2 \mp \sqrt{\lambda}\mu\chi^3 + \frac{\lambda}{4}\chi^4 \tag{1.4.7}$$

after the shift (1.4.5), and has the positive sign of the quadratic term. Therefore, perturbation theory around the vacuum $x_0^\pm$ is a usual one, and the particle lives perturbatively in one of the two vacua.

However, we know from quantum mechanics that (non-perturbatively)

$$\langle x(0)\, x(\tau)\rangle = \sum_n |x_{n0}|^2\, \mathrm{e}^{-(E_n-E_0)\tau} \tag{1.4.8}$$

at imaginary time $\tau = it$ with E_n being the energy of the n-th eigenstate of the Hamiltonian, and x_{n0} being the proper matrix element. Therefore,

$$\langle x(0)\, x(\tau)\rangle \sim \mathrm{e}^{-\Delta E\,\tau} \tag{1.4.9}$$

for large τ, where

$$\Delta E = \mu\sqrt{\frac{48}{\pi}}\sqrt{\frac{2\sqrt{2}\,\mu^3}{3\lambda}}\exp\left(-\frac{2\sqrt{2}\,\mu^3}{3\lambda}\right) \tag{1.4.10}$$

is the energy splitting between the two lowest states (symmetric and anti-symmetric) for $\lambda \ll \mu^3$ which exponentially vanishes as $\lambda \to 0$.

The appearance of imaginary time in Eq. (1.4.8) is because under a barrier particles live in imaginary time. We may say that imaginary time is an adequate language to describe a tunneling through a barrier.

Since the RHS of Eq. (1.4.9) vanishes as $\tau \to \infty$, the reflection symmetry $x \to -x$, which is broken in perturbation theory, is restored non-perturbatively as $\tau \to \infty$.

Problem 1.26 Derive Eq. (1.4.10) modulo a constant factor within standard quantum mechanics.

Solution Let us use the semiclassical formula [LL74] (Problem 3 in §50)

$$\Delta E = \frac{\sqrt{2}\,\mu}{\pi}\,\mathrm{e}^{-\int_{-a}^{+a} dx\sqrt{2(V(x)-E_0)}}, \tag{1.4.11}$$

where $\pm a$ are the classical turning points, which are determined by

$$V(\pm a) = E_0\,, \tag{1.4.12}$$

and

$$E_0 = \sqrt{2}\,\mu \tag{1.4.13}$$

is the lowest energy for the oscillator potential (1.4.7) as $\lambda \to 0$. Denoting

$$h = \sqrt{\frac{\lambda}{\sqrt{2}\,\mu^3}}, \qquad z = \frac{\sqrt{\lambda}}{\mu}\,x\,, \tag{1.4.14}$$

the integral in the exponent on the RHS of Eq. (1.4.11) can be calculated by an expansion in h which gives

$$\frac{1}{2h^2}\int_{-1+h}^{1-h} dz\sqrt{(1-z^2)^2-4h^2} = \frac{2}{3h^2} + \ln h + \mathcal{O}(1)\,. \tag{1.4.15}$$

Substituting into Eq. (1.4.11), one recovers Eq. (1.4.10) modulo a constant factor.

1.4.2 The instanton solution

In the path-integral approach, the correlator (1.4.8) is given by

$$\langle\, x(0)\, x(\tau)\,\rangle = \frac{\int Dx\ \mathrm{e}^{-S[x]}\, x(0)\, x(\tau)}{\int Dx\ \mathrm{e}^{-S[x]}} \tag{1.4.16}$$

with no restrictions on the integration over x. This is a quantum mechanical analogue of the path integrals defined in Subsect. 1.2.1.

At small λ, the path integral (1.4.16) can be evaluated by the saddle-point method. The reason is that for x given by Eq. (1.4.4), which are the minima of the action (1.4.3), the Gaussian fluctuations around (1.4.4) are not essential as $\lambda \to 0$. This is most easily seen by making the shift (1.4.5) and noting that $\chi(\tau)$ is $\mathcal{O}(1)$ at the saddle points according to Eq. (1.4.7) whose RHS is quadratic in $\chi(\tau)$ as $\lambda \to 0$.

Performing the saddle-point evaluation of the path integral (1.4.16), one gets

$$\langle x(0)\, x(\tau)\rangle = \frac{\mu^2}{\lambda} + \ldots\,. \tag{1.4.17}$$

Notice that $x(0)$ and $x(\tau)$ in the integrand can be substituted by the saddle-point values after which the integral over Gaussian fluctuations cancels by the same one in the denominator. In other words we have reproduced that each of the trivial minima (1.4.4) results in Eq. (1.4.6).

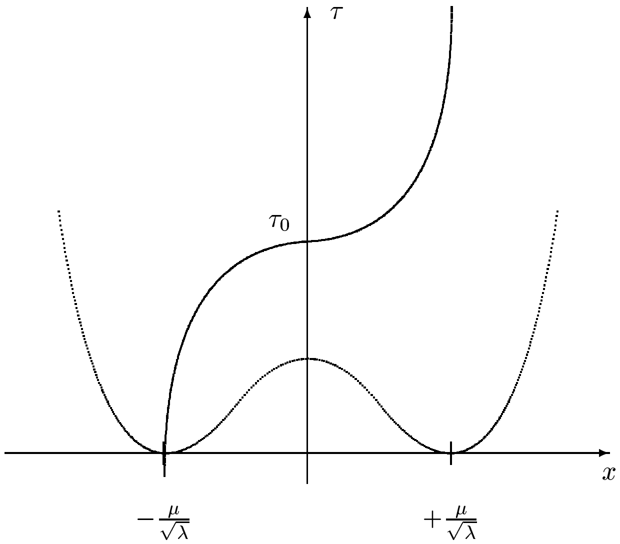


Fig. 1.10: Graphic representation of the one-kink solution (1.4.19) as a function of τ.

Minima of the action (1.4.3) can be alternatively obtained from the classical equation of motion

$$-\ddot{x} - \mu^2 x + \lambda x^3 \;=\; 0\,. \tag{1.4.18}$$

The trivial minima (1.4.4) obviously satisfy this equation.

However, another solution of the classical equation of motion (1.4.18) exists:

$$x_{\rm inst}\,(\tau-\tau_0) \;=\; \frac{\mu}{\sqrt{\lambda}}\tanh\frac{\mu\,(\tau-\tau_0)}{\sqrt{2}}\,, \tag{1.4.19}$$

which is associated with another (local) minimum of the classical action. This solution is called an *instanton* or a pseudoparticle. The arbitrary constant τ_0 in Eq. (1.4.19) is the position of the center of instanton.

The solution (1.4.19) is also known as a *kink* in this quantum mechanical problem. It interpolates between the two minima (1.4.4) when τ changes from $-\infty$ to $+\infty$ as is depicted in Fig. 1.10. On the same figure we kept the shape of $V(x)$ from Fig. 1.9.

An analogous solution which interpolates between $\mu/\sqrt{\lambda}$ at $\tau=-\infty$ and $-\mu/\sqrt{\lambda}$ at $\tau=+\infty$ is called an *anti-instanton.* It differs from Eq. (1.4.19) by an overall minus sign:

$$x_{\rm ainst}\,(\tau-\tau_0) \;=\; -\frac{\mu}{\sqrt{\lambda}}\tanh\frac{\mu\,(\tau-\tau_0)}{\sqrt{2}}, \tag{1.4.20}$$

and is obviously also a solution of the classical equation (1.4.18).

Problem 1.27 Find all solutions of Eq. (1.4.18) with the boundary conditions $x(-\infty)=-\mu/\sqrt{\lambda}$ and $x(+\infty)=\mu/\sqrt{\lambda}$.

Solution Equation (1.4.18) looks like the Newton equation for a classical particle, whose mass is equal to unity, in the upside-down potential $-V(x)$ (its shape can be obtained from the one depicted in Fig. 1.9 by reflecting with respect to the horizontal axis $V=0$). The first integral of motion is the energy

$$\mathcal{E} \;=\; \frac{1}{2}\dot{x}^2 - \frac{\lambda}{4}\left(x^2-\frac{\mu^2}{\lambda}\right)^2 \tag{1.4.21}$$

which is obviously conserved due to Eq. (1.4.18).

Equation (1.4.21) can easily be solved for the velocity

$$\dot{x} \;=\; \sqrt{2(\mathcal{E}+V(x))}, \tag{1.4.22}$$

where we have chosen the positive sign according to the boundary condition. It says also that $\mathcal{E}=0$ in order for the particle to stay at $x=\mu/\sqrt{\lambda}$ for $\tau\to\infty$ since this point is associated with the maximum of $-V(x)$. Therefore, we get

$$\dot{x} \;=\; \sqrt{\frac{\lambda}{2}}\left(\frac{\mu^2}{\lambda}-x^2\right), \tag{1.4.23}$$

which results after integration in Eq. (1.4.19) with τ_0 being the integration constant. It is evident that the solution is unique.

For the instanton (or anti-instanton) minimum, one gets, substituting in Eq. (1.4.3),

$$S\,[x_{\rm inst}] \;=\; \frac{2\sqrt{2}\,\mu^3}{3\lambda}\,, \tag{1.4.24}$$

which coincides with the (minus) exponent in Eq. (1.4.10) for the energy splitting ΔE.

1.4.3 Instanton contribution to path integral

The contribution of the instanton configuration looks as if it is suppressed in the path integral by a factor of $\exp\left(-S\left[x_{\rm inst}\right]\right)$, but in fact this exponential is multiplied by τ since the instanton has a zero mode. This factor of τ appears after an integration over the collective coordinate τ_0 — the instanton center. The explicit result for the one-kink contribution to the correlator (1.4.16) reads [Pol77]

$$\langle\, x\,(0)\, x\,(\tau)\,\rangle \;=\; \frac{\mu^2}{\lambda}\left[1 - C\tau\sqrt{\frac{2\sqrt{2}\,\mu^3}{3\lambda}}\exp\left(-\frac{2\sqrt{2}\,\mu^3}{3\lambda}\right)\right], \tag{1.4.25}$$

where C is a (dimensional) constant.

Problem 1.28 Derive Eq. (1.4.25) using the Faddeev–Popov trick to deal with the collective coordinate τ_0.
Solution Let us approximate the path integrals in the numerator and denominator of Eq. (1.4.16) for small λ by the sum of the contributions from the trivial minima (1.4.4) and the one-kink minima (1.4.19) and (1.4.20). Since the one-kink contribution is suppressed by $\exp\left(-S\left[x_{\rm inst}\right]\right)$, we can expand the denominator to give

$$\begin{aligned}\langle\, x\,(0)\, x\,(\tau)\,\rangle \;=\;& \frac{\mu^2}{\lambda} \\ &+ {\rm e}^{-S[x_{\rm inst}]}\,\frac{\int Dx(\tau)\left(x(0)x(\tau)-\frac{\mu^2}{\lambda}\right)\,{\rm e}^{-(S[x]-S[x_{\rm inst}])}}{\int D\chi(\tau)\,{\rm e}^{-\int d\tau\left(\frac12\dot\chi^2+\mu^2\chi^2\right)}}\,,\end{aligned} \tag{1.4.26}$$

where the path integral in the numerator is over fluctuations around the instanton solution (1.4.19). The normalizing factor in the denominator is associated with averaging over the Gaussian fluctuations around the trivial minima (1.4.4) whose potential energy is described by the quadratic term in Eq. (1.4.7). There are two such trivial minima (x_+ and x_-) and two one-kink minima (instanton and anti-instanton) so these factors of 2 cancel.

Keeping the quadratic term in the expansion around the instanton:

$$x(\tau) \;=\; x_{\rm inst}(\tau-\tau_0) + \chi(\tau-\tau_0)\,, \tag{1.4.27}$$

one gets

$$S[x] - S[x_{\rm inst}] \;=\; \frac12\int d\tau\left(\dot\chi^2 - \mu^2\chi^2 + 3\lambda x_{\rm inst}^2\chi^2\right)\,. \tag{1.4.28}$$

The fluctuations around the instanton are Gaussian except for one mode, which is associated with a translation of the instanton center, τ_0. This zero mode is given by

$$\chi_0(\tau) \;\propto\; \dot x_{\rm inst}(\tau)\,. \tag{1.4.29}$$

This is obvious because

$$\left(-\frac{d^2}{d\tau^2} - \mu^2 + 3\lambda x_{\rm inst}^2\right)\dot x_{\rm inst} \;=\; 0 \tag{1.4.30}$$

as a result of differentiating Eq. (1.4.18) w.r.t. τ_0.

To deal with the zero mode, let us insert the unity

$$1 \;=\; \int\limits_{-\infty}^{+\infty} d\tau\,\delta\left(u[x]-\tau\right) \tag{1.4.31}$$

in the path integral in the numerator on the RHS of Eq. (1.4.26). Here $u[x]$ is determined by the equation

$$\int\limits_{-\infty}^{+\infty} d\tau\, y\left(\tau - u[x]\right) x(\tau) \;=\; 0 \tag{1.4.32}$$

with

$$y(\tau) \;=\; \frac{\dot x(\tau)}{\left[\int_{-\infty}^{+\infty} dt\,\dot x^2(t)\right]^{1/2}} \tag{1.4.33}$$

being the normalized derivative of $x(\tau)$.

Under the translation,

$$\tau \;\to\; \tau' = \tau - \tau_0\,, \tag{1.4.34}$$

one gets

$$x(\tau) \;\to\; x(\tau') = x(\tau-\tau_0)\,. \tag{1.4.35}$$

This leaves the measure and the action in the path integral (1.4.26) invariant, while

$$u[x] \;\to\; u[x] + \tau_0\,. \tag{1.4.36}$$

Therefore, the integration over the instanton center, τ_0, in the numerator of Eq. (1.4.26) factorizes and we get

$$\int Dx(\tau)\left(x(0)x(\tau)-\frac{\mu^2}{\lambda}\right)\,\mathrm{e}^{-(S[x]-S[x_{\rm inst}])} = \int_{-\infty}^{+\infty} d\tau_0\left(x_{\rm inst}(-\tau_0)x_{\rm inst}(\tau-\tau_0)-\frac{\mu^2}{\lambda}\right) \cdot\int D\chi(\tau)\,\delta(u[x_{\rm inst}(\tau)+\chi(\tau)])\,\mathrm{e}^{-\frac12\int d\tau\left(\dot\chi^2-\mu^2\chi^2+3\lambda x^2_{\rm inst}\chi^2\right)}. \tag{1.4.37}$$

We have substituted the integration over the zero mode χ_0 by the integration over the collective coordinate τ_0. The remaining path integral is finite since the integration runs over the directions which are orthogonal to the zero mode.

The integral over τ_0 is equal to

$$\int_{-\infty}^{+\infty} d\tau_0\left(x_{\rm inst}(-\tau_0)x_{\rm inst}(\tau-\tau_0)-\frac{\mu^2}{\lambda}\right) = -\frac{2\mu^2}{\lambda}\tau \tag{1.4.38}$$

as $\lambda\to0$. This is because

$$x_{\rm inst}(\tau-\tau_0) = \frac{\mu}{\sqrt\lambda}\,\mathrm{sign}\,(\tau-\tau_0) \tag{1.4.39}$$

as $\lambda\to0$.

Expanding the delta-function in χ:

$$\delta(u[x]) = \left|\int_{-\infty}^{+\infty} d\tau\,\dot y_{\rm inst}(\tau)x_{\rm inst}(\tau)\right|\delta\left(\int_{-\infty}^{+\infty} d\tau\,y_{\rm inst}(\tau)\chi(\tau)\right), \tag{1.4.40}$$

and noting that

$$\int_{-\infty}^{+\infty} d\tau\,\dot x^2_{\rm inst}(\tau) = \frac{2\sqrt2\,\mu^3}{3\lambda}, \tag{1.4.41}$$

we get

$$\int D\chi(\tau)\,\delta(u[x_{\rm inst}(\tau)+\chi(\tau)])\,\mathrm{e}^{-\frac12\int d\tau\left(\dot\chi^2-\mu^2\chi^2+3\lambda x^2_{\rm inst}\chi^2\right)} = \sqrt{\frac{2\sqrt2\,\mu^3}{3\lambda}}\int D\chi(\tau)\,\delta\left(\int_{-\infty}^{+\infty} d\tau\,y_{\rm inst}(\tau)\chi(\tau)\right)\,\mathrm{e}^{-\frac12\int d\tau\left(\dot\chi^2-\mu^2\chi^2+3\lambda x^2_{\rm inst}\chi^2\right)}. \tag{1.4.42}$$

Notice the appearance of the factor $\sqrt{S[x_{\rm inst}]}$.

We have thus obtained Eq. (1.4.25) with

$$C = 2\frac{\int D\chi(\tau)\,\delta\left(\int_{-\infty}^{+\infty} d\tau\,y_{\rm inst}(\tau)\chi(\tau)\right)\,\mathrm{e}^{-\frac12\int d\tau\left(\dot\chi^2-\mu^2\chi^2+3\lambda x^2_{\rm inst}\chi^2\right)}}{\int D\chi(\tau)\,\mathrm{e}^{-\int d\tau\left(\frac12\dot\chi^2+\mu^2\chi^2\right)}}. \tag{1.4.43}$$

Problem 1.29 Calculate the ratio of determinants in Eq. (1.4.43).
Solution Let us introduce the notation

$$z = \frac{\mu\tau}{\sqrt2}, \qquad D = \frac{d}{dz}. \tag{1.4.44}$$

Noting that

$$\lambda x^2_{\rm inst}(\tau) = \mu^2\left(1-\frac{1}{\cosh^2 z}\right), \tag{1.4.45}$$

we can rewrite the ratio of determinants as

$$B^{-2} = \frac{4\pi}{\mu^2}\frac{\det'\left[-D^2+4-6/\cosh^2 z\right]}{\det\left[-D^2+4\right]}. \tag{1.4.46}$$

The notation $\det'$ means that the zero eigenvalue is excluded. An extra factor of 2π comes from the normalization of the Gaussian integral in the denominator which involves one integral more.

The RHS of Eq. (1.4.46) can be calculated by the limiting procedure

$$\frac{\det'\left[-D^2+4-6/\cosh^2 z\right]}{\det\left[-D^2+4\right]} = \lim_{\omega\to2}\frac{\det\left[-D^2+\omega^2-6/\cosh^2 z\right]}{(\omega^2-4)\det\left[-D^2+\omega^2\right]}. \tag{1.4.47}$$

To compute the ratio of the Fredholm determinants

$$\mathcal{R}_\omega[v] \equiv \frac{\det\left[-D^2+\omega^2+v(z)\right]}{\det\left[-D^2+\omega^2\right]} \tag{1.4.48}$$

for the potential

$$v(z) = -\frac{6}{\cosh^2 z}, \tag{1.4.49}$$

let us note that

$$\frac{\partial}{\partial\omega^2}\ln\mathcal{R}_\omega[v] = \mathrm{Sp}\left[\frac{1}{-D^2+\omega^2+v(z)}\right]-\mathrm{Sp}\left[\frac{1}{-D^2+\omega^2}\right] = \int_{-\infty}^{+\infty} dz\left(R_\omega(z,z;v)-\frac{1}{2\omega}\right), \tag{1.4.50}$$

where the diagonal resolvent $R_\omega(z,z;v)$ is defined by Eq. (1.1.119) with $G=1$ and $V\equiv v$. The term $1/2\omega$ on the RHS, which equals the diagonal resolvent in the free case when $v=0$ (see Eq. (1.1.38)), comes from the free determinant in the denominator on the RHS of Eq. (1.4.48).

A crucial observation is that the diagonal resolvent for the potential (1.4.49) is given by the simple formula

$$R_\omega(z,z;v) \;=\; \frac{1}{2\omega} - \frac{v(z)}{4\omega(\omega^2-1)} + \frac{v^2(z)}{8\omega(\omega^2-1)(\omega^2-4)}\,, \tag{1.4.51}$$

which can easily be verified by substituting into the Gelfand–Dickey equation (1.1.123) with $\mathcal{G}=1$. The reason is that the potential (1.4.49) is integrable and possesses two bound states (see, *e.g.*, §23 of Ref. [LL74]).

Calculating the integral over z on the RHS of Eq. (1.4.50), by the formulas

$$\int\limits_{-\infty}^{+\infty} \frac{dz}{\cosh^2 z} \;=\; 2\,, \qquad \int\limits_{-\infty}^{+\infty} \frac{dz}{\cosh^4 z} \;=\; \frac{4}{3}\,, \tag{1.4.52}$$

we obtain

$$\frac{\partial}{\partial\omega^2} \ln \mathcal{R}_\omega[v] \;=\; \frac{1}{\omega}\left(\frac{1}{\omega^2-1} + \frac{2}{\omega^2-4}\right)\,, \tag{1.4.53}$$

which is easily integrated over ω to give

$$\frac{\det\left[-D^2+\omega^2-6/\cosh^2 z\right]}{\det\left[-D^2+\omega^2\right]} \;=\; \frac{(\omega-2)(\omega-1)}{(\omega+2)(\omega+1)}\,. \tag{1.4.54}$$

The integration constant has been determined requiring that

$$\lim_{\omega\to\infty} \mathcal{R}_\omega[v] \;=\; 1\,. \tag{1.4.55}$$

Substituting into Eq. (1.4.47), we get

$$C \;=\; 2B \;=\; \sqrt{\frac{48}{\pi}}\mu \tag{1.4.56}$$

which coincides with the constant in Eq. (1.4.10).

For other methods of calculating the ratio of determinants in the one-instanton contribution, see the original papers [Lan67, Pol77], the reviews [Col77, VZNS82] or Chapter 4 of the book by Polyakov [Pol87].

1.4.4 Symmetry restoration by instantons

At $\tau \sim 1/\Delta E$, many kinks become essential. A many-kink "solution" can be approximately constructed from several single kinks and anti-kinks, which are separated along the τ-axis by the some distance $R \gg 1/\mu$, since the interaction between kinks would be $\sim \exp\{-\mu R\}$. Such a configuration is

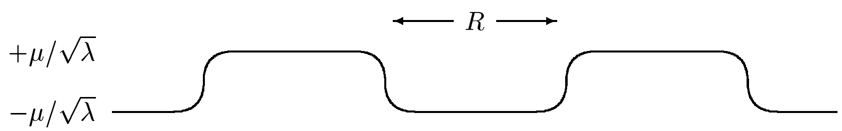


Fig. 1.11: The many-kink configuration $x_{M\text{-kink}}(\tau)$ which is combined from the solution (1.4.19).

depicted in Fig. 1.11 for the case when the number of kinks is equal to the number of anti-kinks. An analogous configuration with the number of kinks one more than the number of anti-kinks connects the $-\mu/\sqrt{\lambda}$ and $\mu/\sqrt{\lambda}$ vacua.

It is not an exact solution of Eq. (1.4.18) since the kink and the anti-kink attract and have a tendency to annihilate. However, it is an approximate solution as $\lambda \to 0$.

Analytically, the M-kink configuration can be represented as

$$x_{M\text{-kink}}(\tau) \;=\; \frac{\mu}{\sqrt{\lambda}} \prod_{i=1}^{M} \operatorname{sign}(\tau-\tau_i)\,, \tag{1.4.57}$$

where τ_i are the centers of the instantons (or anti-instantons), of which the M-kink configuration is built out, and

$$\tau_1 \le \tau_2 \le \ldots \le \tau_M\,. \tag{1.4.58}$$

Equation (1.4.57) assumes that the kinks do not interact and are infinitely thin as $\lambda \to 0$. The action of the configuration (1.4.57) is therefore given by

$$S[x_{M\text{-kink}}] \;=\; \frac{2\sqrt{2}\mu^3}{3\lambda}M\,, \tag{1.4.59}$$

i.e. it equals M times the action for the one kink.

Summing over many-kink configurations, one gets [Pol77]

$$\langle\, x(0)\, x(\tau)\,\rangle \;=\; \frac{\mu^2}{\lambda}\,\mathrm{e}^{-\tau\Delta E}\,, \tag{1.4.60}$$

where ΔE is given by Eq. (1.4.10). The $x \to -x$ symmetry is now restored as $\tau \to \infty$. This restoration is produced by instantons = classical trajectories with a finite (Euclidean) action.

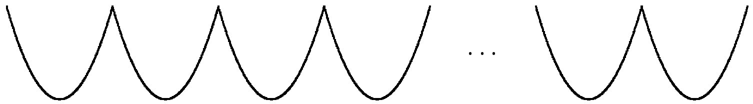


Fig. 1.12: Graphic representation of a periodic potential.

Problem 1.30 Obtain the exponentiation of the one-kink contribution (1.4.25) after summing over the M-kink configurations (1.4.57) in the dilute gas approximation when the interaction between kinks is disregarded.
Solution The calculation of the contribution of the M-kink configuration (1.4.57) to the path integral is quite analogous to that for the one kink which is described in Problem 1.28. One gets

$$\langle x(0)x(\tau)\rangle = \frac{\mu^2}{\lambda}\sum_{M=0}^{\infty}(-\Delta E)^M \int_0^\tau d\tau_1 \int_0^{\tau_1} d\tau_2 \ldots \int_0^{\tau_{M-1}} d\tau_M\,, \qquad (1.4.61)$$

which reproduces Eq. (1.4.60) by noting that the ordered integral is equal to

$$\int_0^\tau d\tau_1 \int_0^{\tau_1} d\tau_2 \ldots \int_0^{\tau_{M-1}} d\tau_M = \frac{\tau^M}{M!}\,. \qquad (1.4.62)$$

This calculation is very similar to the one in statistical mechanics for the exponentiation of a single-particle contribution to the partition function in the case of an ideal gas.

1.4.5 Topological charge and θ-vacua

Let us consider a periodic potential whose period equals 1, which is depicted in Fig. 1.12. It can be viewed as being defined on a circle S^1 of a unit length. The boundary conditions are

$$\begin{aligned} x(1) &= x(0) && \boxed{\text{in perturbation theory}}\,, \\ x(1) &= x(0)+n && \boxed{\text{for } n\text{-instanton solution}}\,. \end{aligned} \qquad (1.4.63)$$

The multi-instanton solution always exists because of the topological formula[15]

$$\pi_1(S^1) = \boldsymbol{Z}\,, \qquad (1.4.64)$$

[15]See, *e.g.*, the book [DNF86] (§17.5 of Part II).

where $\pi_k(M)$ is the k-th homotopy group whose elements are classes of continuous maps of the k-sphere S^k onto M. Eq. (1.4.64) says that an (integer) winding number $n \in \boldsymbol{Z}$ is associated with the mapping $S^1 \to S^1$, which counts how many times the target is covered.

We see the difference between the M-kink configuration for the double-well potential and the multi-instanton solution for the periodic potential. The former was not an exact solution of the classical field equation (1.4.18). Only a single instanton or anti-instanton was a solution that connects the two vacua. This is why we need a periodic potential for the multi-instanton solution to exist due to the topological argument.

The value of n in the boundary condition (1.4.63) is called the *topological charge* of the instantons, while $n<0$ is associated with anti-instantons. The vacuum states are labelled by n: $|n\rangle$. The n-instanton configuration connects the $|m\rangle$ and $\langle m+n|$ states. Therefore, instantons are associated in Minkowski space with the process of tunneling between topologically distinct vacua[16] rather than with real particles. For this reason, they are sometimes called pseudoparticles in Euclidean space.

It is convenient to consider another representation of vacuum states

$$|\,\theta\,\rangle = \sum_{n=-\infty}^{\infty} \mathrm{e}^{\,i\theta n}\,|\,n\rangle\,, \qquad (1.4.65)$$

which are called the θ-vacua. The θ-vacua are orthogonal

$$\langle\theta\mid\theta'\rangle = \sum_{m=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} \mathrm{e}^{\,i(\theta n-\theta' m)}\langle m\mid n\rangle = \delta_{2\pi}(\theta-\theta')\,, \qquad (1.4.66)$$

where $\delta_{2\pi}$ is a periodic delta-function with period 2π. Here we used the orthogonality of the n-states:

$$\langle m\mid n\rangle = \delta_{mn}\,. \qquad (1.4.67)$$

[16]The Minkowski-space interpretation of instantons is attributed to V.N. Gribov (unpublished). It is based on the fact that when the particle is localized in one of the two wells its momentum is indefinite and can sometimes be very large so that the proper energy is above the barrier between the two wells. Such a particle jumps from the given well to the other one. The characteristic time of this process is small in the typical units given by μ. In other words this process is instantaneous that explains the term "instanton" which was introduced by 't Hooft. The exponential suppression with λ of the one-instanton contribution (1.4.25) represents quantitatively the fact that the probability of having large momentum is small.

The θ-vacuum partition function reads

$$Z_\theta \;=\; \int Dx \; \mathrm{e}^{-S[x]+i\theta \int_0^1 d\tau\, \dot{x}(\tau)}. \tag{1.4.68}$$

Here θ is multiplied in the exponent by the topological charge

$$\int_0^1 d\tau\, \dot{x}\,(\tau) \;=\; x\,(1) - x\,(0) \tag{1.4.69}$$

which never shows up in perturbation theory. Therefore, the partition function (1.4.68) can be alternatively represented as

$$Z_\theta \;=\; \sum_n \int_{x(1)=x(0)+n} Dx \; \mathrm{e}^{-S[x]+i\theta n}. \tag{1.4.70}$$

The second term in the exponent in Eq. (1.4.68) is known as the θ*-term*. The parameter θ plays a role of a new fundamental constant which does not show up in perturbation theory. Amplitudes of physical processes generated by instantons may depend on the θ.

Remark on description of instantons

A description of instantons in the first-quantized language can be done only in quantum mechanics (where the first and second quantizations do not differ essentially). The path-integral representation (1.4.16) is more in the spirit of second quantization, which is discussed in Sect. 1.2, where $x(\tau)$ plays the role of a field which depends on the one-dimensional coordinate τ.

Remark on instantons in Yang–Mills theory

In the Yang–Mills theory, instantons are conveniently described by a (Euclidean) path integral over fields. The saddle-point equation, which describes instantons in the SU(2) Yang–Mills theory, reads [BPST75]

$$F^a_{\mu\nu}(x) \;=\; \tilde{F}^a_{\mu\nu}(x), \tag{1.4.71}$$

whose non-trivial solutions exist due to the fact that the mapping of the asymptotics S^3 of four-dimensional Euclidean space onto SU(2) is nontrivial:

$$\pi_3\,(\mathrm{SU}(2)) \;=\; \boldsymbol{Z}\,. \tag{1.4.72}$$

Correspondingly, the topological charge is given by

$$n \;=\; \frac{g^2}{32\pi^2} \int d^4x \sum_{a=1}^{3} F^a_{\mu\nu}(x) \tilde{F}^a_{\mu\nu}(x)\,, \tag{1.4.73}$$

which equals one half of the nonconservation of the axial charge given by the Minkowski-space integral of the chiral anomaly (1.3.64). This expression is also known in topology as the Pontryagin index or the second Chern class. See, *e.g.*, Refs. [Col77, VZNS82] for an introduction to instantons in the Yang–Mills theory.

1.5 Reference guide

The operator formalism in quantum field theory is described in the canonical books [AB69,BS76,BD65] which were written in the fifties or the beginning of the sixties.

The Feynman disentangling is contained in the original paper [Fey51] whose appendices are especially relevant. A classical book on path integrals in quantum mechanics is the one by Feynman and Hibbs [FH65]. The path-integral approach to a very closely related problem of Brownian motion is discussed in the books [Kac59, Sch81, Wie86]. Many useful information on path integrals can be found in the book by Kleinert [Kle95].

An introduction to path integrals in quantum mechanics and quantum field theory can be found in many books. I shall list some which I have on my bookshelf: [Ber86, Pop76, SF88, IZ80, Ram81, Sak85]. The ordering is according to the appearance of the first edition. The book by Berezin [Ber86], which pretends to be mathematically more rigorous, contains an excellent description of operations with Grassmann variables.

An introduction to path integrals in statistical mechanics can be found in Refs. [Kac59, Fey72, Pop76, Wie86]. The well-written book by Parisi [Par88] describes a modern view on the relation between statistical mechanics and quantum field theory. A very good, while a bit more advanced, book where contemporary problems of quantum field theory and statistical mechanics are discussed in an unified language of Euclidean path integrals is the one by Polyakov [Pol87].

The derivation of quantum anomalies from a non-invariance of the measure in path integral is contained in the original papers [Ver78, Fuj79] (see also the review [Mor86]). It can also be found in Sect. 8.9 of the Second edition of the book by Ramond [Ram81].

Instantons in the Yang–Mills theory were discovered by Belavin, Polyakov, Schwartz and Tyupkin [BPST75]. The role of instantons in quantum mechanics is clarified in the original paper by Polyakov [Pol77]. Their description is given in the books [Sak85, Pol87]. The review articles [Col77, VZNS82] are also useful for an introduction to the subject.

References to Chapter 1

[AB69] A.I. Akhiezer and V.B. Berestetskii, *Quantum electrodynamics*, Nauka, M., 1969 (English transl.: Oldbourne Press, L., 1962).

[ACD77] S.L. Adler, J.C. Collins and A. Duncan, *Energy-momentum-tensor trace anomaly in spin-1/2 quantum electrodynamics*, Phys. Rev. **D15** (1977) 1712.

[BD65] J.D. Bjorken and S.D. Drell, *Relativistic quantum fields*, McGraw-Hill, N.Y., 1965 (Russian transl.: Nauka, M., 1978).

[Ber86] F.A. Berezin, *The method of second quantization*, Nauka, M., 1986 (English transl.: Acad. Press, N.Y., 1966).

[BPST75] A.A. Belavin, A.M. Polyakov, A.S. Schwartz and Yu.S. Tyupkin, *Pseudoparticle solutions of the Yang–Mills equations*, Phys. Lett. **B59** (1975) 85.

[BS76] N.N. Bogoliubov and D.V. Shirkov, *Introduction to the theory of quantized fields*, Nauka, M., 1976 (English transl.: Interscience, N.Y., 1959).

[CDJ77] J.C. Collins, A. Duncan and S.D. Joglekar, *Trace and dilatation anomalies in gauge theories*, Phys. Rev. **D16** (1977) 438.

[CE72] M.S. Chanowitz and J. Ellis, *Canonical anomalies and broken scale invariance*, Phys. Lett. **40B** (1972) 397.

[Col77] S. Coleman, *The uses of instantons*, in Proc. of Erice Int. School of Subnuclear Physics 1977, Plenum, N.Y., 1979, p. 805 (Reprinted in S. Coleman, *Aspects of symmetry*, Cambridge Univ. Press, 1985, pp.265–350).

[Cre72] R. Crewther, *Nonperturbative evaluation of the anomalies in low-energy theorems*, Phys. Rev. Lett. **28** (1972) 1421.

[Dir58] P.A.M. Dirac, *The principles of quantum mechanics*, Oxford Univ. Press, 1958 (Russian transl.: Nauka, M., 1979).

[DNF86] B.A. Dubrovin, C.P. Novikov and A.T. Fomenko, *Modern geometry. Methods and applications*, Nauka, M., 1986 (English transl.: Springer, N.Y., Part I 1984, Part II 1985).

[Dys49] F.J. Dyson, *The radiation theories of Tomonaga, Schwinger, and Feynman*, Phys. Rev. **75** (1949) 486.

[Fey50] R.P. Feynman, *Mathematical formulation of the quantum theory of electromagnetic interaction*, Phys. Rev. **80** (1950) 440.

[Fey51] R.P. Feynman, *An operator calculus having applications in quantum electrodynamics*, Phys. Rev. **84** (1951) 108.

[Fey72] R.P. Feynman, *Statistical mechanics*, Benjamin, Mass., 1972 (Russian transl.: Mir, M., 1975).

[FH65] R.P. Feynman and A.R. Hibbs, *Quantum mechanics and path integrals*, McGraw-Hill, N.Y., 1965 (Russian transl.: Mir, M, 1968).

[Fuj79] K. Fujikawa, *Path integral measure for gauge invariant fermion theories*, Phys. Rev. Lett. **42** (1979) 1195.

[GD75] I.M. Gelfand and L.A. Dickey, *Asymptotics of development of the Sturm–Liouville equations and the Korteveg–de Vries algebra*, Usp. Mat. Nauk **30** (1975) v.5, pp.67-100 (English transl.: Russ. Math. Surv. **30** (1975) v.5, 77).

[IZ80] C. Itzykson and J.-B. Zuber, *Quantum field theory*, McGraw-Hill, N.Y., 1980 (Russian transl.: Mir, M., 1984).

[Kac59] M. Kac, *Probability and related topics in the physical science*, Interscience, N.Y., 1959 (Russian transl.: Mir, M., 1972).

[Kle95] H. Kleinert, *Path integrals in quantum mechanics, statistics, and polymer physics*, 2nd Ed., World Sci., 1995.

[Lan67] J.S. Langer, *Theory of the condensation point*, Ann. Phys. **41** (1967) 108.

[Lev51] P. Lévy, *Probleèms concrets d'analyse fonctionnelle*, Paris, 1951 (Russian transl.: Nauka, M., 1967).

[LL74] L.D. Landau and E.M. Lifshits, *Quantum mechanics. Nonrelativistic theory*, Nauka, M., 1974 (English transl.: Pergamon, N.Y., 1977).

[MM81] 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).

[Mor86] A.Yu. Morozov, *Anomalies in gauge theories*, Usp. Fiz. Nauk **150** (1986) 337 (English transl.: Sov. Phys. Usp. **29** (1986) 993).

[Nie77] N.K. Nielsen, *Energy-momentum tensor in a non-Abelian quark gluon theory*, Nucl. Phys. **B120** (1977) 212.

[Par88] G. Parisi, *Statistical field theory*, Addison-Wesley, 1988.

[Pol77] A.M. Polyakov, *Quark confinement and topology of gauge groups*, Nucl. Phys. **B120** (1977) 429.

[Pol87] A.M. Polyakov, *Gauge fields and strings*, Harwood Acad. Pub., 1987.

[Pop76] V.N. Popov, *Functional integrals in quantum field theory and statistical physics*, Atomizdat, M., 1976 (English transl.: Reidel, Dordrecht, 1983).

[Ram81] P. Ramond, *Field theory. A modern primer*, Benjamin/Cummings, 1981 (Russian transl.: Mir, M., 1984); 2nd ed. Addison-Wesley, 1989.

[Sak85] B. Sakita, *Quantum theory of many-variable systems and fields*, World Sci., 1985.

[Sch58] J. Schwinger, *Selected papers on quantum electrodynamics*, Dover, N.Y., 1958.

[Sch81] L.S. Schulman, *Techniques and applications of path integration*, Wiley, N.Y., 1981.

[SF88] A.A. Slavnov, L.D. Faddeev, *Introduction to quantum theory of gauge fields*, Nauka, M., 1988 (English transl.: Benjamin, Mass., 1980).

[Ver78] S.N. Vergeles, unpublished, as quoted in A.A. Migdal, *Effective low-energy Lagrangian for QCD*, Phys. Lett. **81B** (1979) 37.

[VZNS82] A.I. Vainshtein, V.I. Zakharov, V.A. Novikov and M.A. Shifman, *ABC of instantons*, Usp. Fiz. Nauk **136** (1982) 553 (English transl.: Sov. Phys. Usp. **25** (1982) 195).

[Wie23] N. Wiener, *Differential space*, J. Math. Phys. **2** (1923) v.3, pp.131–174.

[Wie86] F.W. Wiegel, *Introduction to path-integral methods in physics and polymer science*, World Sci., 1986.

"I never said it."
"Now you are telling us when you did say it. I'm asking you to tell us when you didn't say it."
J. Heller, *Catch-22*

Chapter 2

Lattice Gauge Theories

Lattice gauge theories in their modern form were proposed in 1974 by Wilson [Wil74] 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 non-trivial 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 shall start this Chapter by a description of the continuum formulation of non-Abelian gauge theories, and will return to it 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 [Wil74] on a lattice, are applicable for the continuum theory as well.

2.1 Observables in gauge theories

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 [YM54].

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 the matrix notation for the non-Abelian gauge fields and introduce proper non-Abelian phase factors. Then we discuss the relation between observables in classical and quantum theories.

2.1.1 Gauge invariance

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

$$\psi\,(x) \quad \overset{\text{g.t.}}{\longrightarrow} \quad \psi'\,(x) = \Omega\,(x)\,\psi\,(x)\ . \tag{2.1.1}$$

Here $\Omega\,(x) \in G$ with G being a semisimple Lie group which is called the gauge group ($G = \mathrm{SU}(3)$ for QCD). Equation (2.1.1) means that ψ belongs to the fundamental representation of G.

We restrict ourselves to a unitary gauge group when

$$\Omega^{-1}(x) \;=\; \Omega^{\dagger}(x)\,, \tag{2.1.2}$$

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

$$\psi^{\dagger}\,(x) \quad \overset{\text{g.t.}}{\longrightarrow} \quad \psi'^{\,\dagger}\,(x) = \psi^{\dagger}\,(x)\,\Omega^{\dagger}\,(x)\,. \tag{2.1.3}$$

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

The gauge transformation (2.1.1) 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:

$$\mathcal{A}_\mu\,(x) \quad \overset{\text{g.t.}}{\longrightarrow} \quad \mathcal{A}'_\mu\,(x) = \Omega\,(x)\,\mathcal{A}_\mu\,(x)\,\Omega^{\dagger}\,(x) - \Omega\,(x)\,\partial_\mu\Omega^{\dagger}\,(x)\,. \tag{2.1.4}$$

We have introduced in Eq. (2.1.4) the anti-Hermitean matrix $\mathcal{A}_\mu\,(x)$ with the elements

$$[\mathcal{A}_\mu\,(x)]^{ij} \;=\; ig\sum_a A^a_\mu\,(x)\,[t^a]^{ij}\ . \tag{2.1.5}$$

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

$$\mathrm{tr}\,t^a t^b \;=\; \frac{1}{2}\delta^{ab}, \tag{2.1.6}$$

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

$$t^a \;=\; \frac{\sigma^a}{2} \tag{2.1.7}$$

for the SU(2) group, with σ^a being the Pauli matrices.

Equation (2.1.5) can be inverted to give

$$A^a_\mu\,(x) \;=\; \frac{2}{ig}\,\mathrm{tr}\,\mathcal{A}_\mu\,(x)\,t^a\ . \tag{2.1.8}$$

Substituting

$$\Omega(x) \;=\; \mathrm{e}^{\,i\alpha(x)}, \tag{2.1.9}$$

we get for an infinitesimal α:

$$\delta\,\mathcal{A}_\mu(x) \;\overset{\text{g.t.}}{=}\; i\,\nabla^{\mathrm{adj}}_\mu\,\alpha(x). \tag{2.1.10}$$

Here

$$\nabla^{\mathrm{adj}}_\mu\,\alpha \;\equiv\; \partial_\mu\alpha - [\mathcal{A}_\mu,\alpha] \tag{2.1.11}$$

is the covariant derivative in the adjoint representation of G while

$$\nabla^{\mathrm{fun}}_\mu\,\psi \;\equiv\; \partial_\mu\psi - \mathcal{A}_\mu\psi \tag{2.1.12}$$

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

$$\nabla^{\mathrm{adj}}_\mu\,B(x) \;=\; [\nabla^{\mathrm{fun}}_\mu, B(x)] \tag{2.1.13}$$

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

The QCD action reads in the matrix notation as

$$S[\mathcal{A},\psi,\bar{\psi}] \;=\; \int d^4x \left[\bar{\psi}\gamma_\mu\left(\partial_\mu - \mathcal{A}_\mu\right)\psi + m\bar{\psi}\psi - \frac{1}{2g^2}\,\mathrm{tr}\,\mathcal{F}^2_{\mu\nu}\right], \tag{2.1.14}$$

where

$$\mathcal{F}_{\mu\nu} \;=\; \partial_\mu\mathcal{A}_\nu - \partial_\nu\mathcal{A}_\mu - [\mathcal{A}_\mu,\mathcal{A}_\nu] \tag{2.1.15}$$

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

The action (2.1.14) is manifestly invariant under the local gauge transformation (2.1.1), (2.1.4) since

$$\mathcal{F}_{\mu\nu}(x) \;\stackrel{\text{g.t.}}{\longrightarrow}\; \Omega(x)\,\mathcal{F}_{\mu\nu}(x)\,\Omega^\dagger(x) \tag{2.1.16}$$

or

$$\delta\mathcal{F}_{\mu\nu}(x) \;\stackrel{\text{g.t.}}{=}\; -i\left[\mathcal{F}_{\mu\nu}(x),\alpha(x)\right] \tag{2.1.17}$$

for the infinitesimal gauge transformation.

For the Abelian group $G = \mathrm{U}(1)$, the above formulas recover those of the previous Chapter for QED where we have already used the calligraphic notations in Problem 1.24.

Problem 2.1 Rewrite classical equations of motion in the matrix notations.
Solution The non-Abelian Maxwell equation and the Bianchi identity read, respectively, as

$$\nabla^{\text{adj}}_\mu\,\mathcal{F}_{\mu\nu} \;=\; 0 \tag{2.1.18}$$

and

$$\nabla^{\text{adj}}_\mu\,\tilde{\mathcal{F}}_{\mu\nu} \;=\; 0\,, \tag{2.1.19}$$

where the dual tensor is defined by Eq. (1.3.50). Rewriting Eq. (2.1.15) as

$$\mathcal{F}_{\mu\nu} \;=\; -\left[\nabla^{\text{fun}}_\mu,\nabla^{\text{fun}}_\nu\right] \tag{2.1.20}$$

and using Eq. (2.1.13), we represent the Bianchi identity as

$$\epsilon_{\mu\nu\lambda\rho}\left[\nabla^{\text{fun}}_\mu,\left[\nabla^{\text{fun}}_\nu,\nabla^{\text{fun}}_\lambda\right]\right] \;=\; 0 \tag{2.1.21}$$

which is obviously satisfied due to the Jacobi identity.

We have thus proven 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. (2.1.15). On the contrary, $\mathcal{A}_\mu$ and $\mathcal{F}_{\mu\nu}$ are considered as independent variables in the first-order formalism, where both equations (2.1.18) and (2.1.19) are essential. The concept of the first- and second-order formalisms comes from the theory of gravity.

2.1.2 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. 1.1.7. The proper extension of the Abelian formula (1.1.154) is written as

$$U[\Gamma_{yx}] \;=\; \mathbf{P}\,\mathrm{e}^{\int_{\Gamma_{yx}} dz^\mu \mathcal{A}_\mu(z)}. \tag{2.1.22}$$

While the matrices $\mathcal{A}_\mu(z)$ do not commute, the path-ordered exponential on the RHS of Eq. (2.1.22) is unambiguously defined by the general method of Subsect. 1.1.3. This is obvious after rewriting the phase factor in an equivalent form

$$\mathbf{P}\,\mathrm{e}^{\int_{\Gamma_{yx}} dz^\mu \mathcal{A}_\mu(z)} \;=\; \mathbf{P}\,\mathrm{e}^{\int_0^1 d\sigma\, \dot{z}^\mu(\sigma)\mathcal{A}_\mu(z(\sigma))}. \tag{2.1.23}$$

Therefore, the path-ordered exponential in Eq. (2.1.22) is to be understood as[1]

$$U[\Gamma_{yx}] \;=\; \prod_{t=0}^{\tau}\left[1 + dt\,\dot{z}^\mu(t)\mathcal{A}_\mu(z(t))\right]. \tag{2.1.24}$$

We already used this notation for the product on the RHS in Problem 1.8. Using Eq. (1.1.153), Eq. (2.1.24) can also be written as

$$U[\Gamma_{yx}] \;=\; \prod_{x}^{y}\left[1 + dz^\mu\mathcal{A}_\mu(z)\right]. \tag{2.1.25}$$

If the contour Γ_{yx} is discretized as is shown in Fig. 1.3, then the non-Abelian phase factor is approximated by

$$U[\Gamma_{yx}] \;=\; \lim_{M\to\infty}\prod_{i=1}^{M}\left[1 + (z_i - z_{i-1})^\mu\mathcal{A}_\mu\left(\frac{z_i + z_{i-1}}{2}\right)\right], \tag{2.1.26}$$

which obviously reproduces (2.1.25) in the limit $\epsilon \to 0$.

Notice that the non-Abelian phase factor (2.1.22) is, by construction, an element of the gauge group G itself, while $\mathcal{A}_\mu$ belongs to the Lie algebra of G.

[1] Sometimes the phase factor is defined by a similar formula but with the inverse order of multipliers. Our definition by Eq. (2.1.24) is exactly equivalent to the Dyson's definition of the P–product (see the footnote on p. 2) which can be seen by choosing the contour Γ_{yx} to coincide with the temporal axis.

Problem 2.2 Write down an explicit expansion of the non-Abelian phase factor (2.1.22) in $\mathcal{A}_\mu$.
Solution Let us use the notation

$$\int_x^y dz^\mu \ldots \equiv \int_{\Gamma_{yx}} dz^\mu \ldots \tag{2.1.27}$$

for the integral along the contour Γ_{yx}. Then we get

$$\begin{aligned}
&\mathbf{P}\,\mathrm{e}^{\int_x^y dz^\mu \mathcal{A}_\mu(z)} \\
&= \sum_{k=0}^{\infty} \int_x^y dz_1^{\mu_1} \int_{z_1}^y dz_2^{\mu_2} \ldots \int_{z_{k-1}}^y dz_k^{\mu_k} \mathcal{A}_{\mu_k}(z_k)\cdots\mathcal{A}_{\mu_2}(z_2)\mathcal{A}_{\mu_1}(z_1)\,.
\end{aligned} \tag{2.1.28}$$

The ordered integral in this formula can be rewritten in a more symmetric form as

$$\begin{aligned}
&\int_0^\tau dt_1 \int_{t_1}^\tau dt_2 \ldots \int_{t_{k-1}}^\tau dt_k\, \dot z^{\mu_1}(t_1)\dot z^{\mu_2}(t_2)\cdots \dot z^{\mu_k}(t_k) \\
&\qquad \cdot\mathcal{A}_{\mu_k}(z(t_k))\cdots\mathcal{A}_{\mu_2}(z(t_2))\mathcal{A}_{\mu_1}(z(t_1)) \\
&= \int_0^\tau dt_1 \int_0^\tau dt_2 \ldots \int_0^\tau dt_k\, \theta(t_k,t_{k-1},\ldots,t_2,t_1)\, \dot z^{\mu_1}(t_1)\dot z^{\mu_2}(t_2)\cdots \dot z^{\mu_k}(t_k) \\
&\qquad \cdot\mathcal{A}_{\mu_k}(z(t_k))\cdots\mathcal{A}_{\mu_2}(z(t_2))\mathcal{A}_{\mu_1}(z(t_1))\,,
\end{aligned} \tag{2.1.29}$$

where

$$\theta(t_k,t_{k-1},t_{k-2},\ldots,t_2,t_1) = \theta(t_k-t_{k-1})\,\theta(t_{k-1}-t_{k-2})\cdots\theta(t_2-t_1) \tag{2.1.30}$$

orders the points along the contour. We shall also denote this theta in a parametrization independent form as

$$\theta(k,k-1,k-2,\ldots,2,1) \equiv \theta(t_k,t_{k-1},t_{k-2},\ldots,t_2,t_1)\,. \tag{2.1.31}$$

It satisfies the obvious identity

$$\begin{aligned}
&\theta(k,k-1,k-2,\ldots,2,1)+\theta(k-1,k,k-2,\ldots,2,1) \\
&+ \text{ (other permutations of } k,\ldots,1) = 1.
\end{aligned} \tag{2.1.32}$$

For the Abelian case, when $\mathcal{A}(z_i)$ commute, Eq. (2.1.32) results in

$$\begin{aligned}
&\int_x^y dz_1^{\mu_1} \int_{z_1}^y dz_2^{\mu_2} \ldots \int_{z_{k-1}}^y dz_k^{\mu_k} \mathcal{A}_{\mu_k}(z_k)\cdots\mathcal{A}_{\mu_2}(z_2)\mathcal{A}_{\mu_1}(z_1) \\
&= \frac{1}{k!}\left(\int_x^y dz^\mu \mathcal{A}_\mu(z)\right)^k
\end{aligned} \tag{2.1.33}$$

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

Problem 2.3 Disentangle the non-Abelian phase factor using a path integral over Grassmann variables living in a contour.
Solution Let us define the average

$$\left\langle F[\psi,\bar\psi]\right\rangle_\psi = \frac{\int D\bar\psi(t)D\psi(t)\,\mathrm{e}^{-\int_0^\tau dt\,\bar\psi(t)\dot\psi(t)-\bar\psi(0)\psi(0)}\,F[\psi,\bar\psi]}{\int D\bar\psi(t)D\psi(t)\,\mathrm{e}^{-\int_0^\tau dt\,\bar\psi(t)\dot\psi(t)-\bar\psi(0)\psi(0)}}\,. \tag{2.1.34}$$

The path integral in this formula looks like those of Sect. 1.2 with $\bar\psi_i(t)$ and $\psi_j(t)$ being Grassmann variables which depend on the one-dimensional variable $t\in[0,\tau]$ that parametrizes a contour, and i and j are the color indices.

The simplest average, which describes a propagation of the color indices along the contour, is

$$\left\langle \psi_i(t_2)\bar\psi_j(t_1)\right\rangle_\psi = \delta_{ij}\,\theta(t_2-t_1)\,, \qquad 0\le t_1,t_2\le\tau\,. \tag{2.1.35}$$

This can be easily checked, say, by deriving the Schwinger–Dyson equation

$$\frac{\partial}{\partial t_2}\left\langle \psi_i(t_2)\bar\psi_j(t_1)\right\rangle_\psi = \delta_{ij}\,\delta^{(1)}(t_2-t_1)\,, \qquad 0< t_1,t_2<\tau\,. \tag{2.1.36}$$

as was done in Sect. 1.3. We see now that we need the Grassmann variables because the operator in the action in Eq. (2.1.34) is $\partial/\partial t$.

A special comment is needed concerning the term $\bar\psi(0)\psi(0)$ in the exponent whose appearance in the disentangling procedure is clarified in Ref. [HJS77]. The need for this term can be seen from the discretized version of the exponent:

$$\int_0^\tau dt\,\bar\psi(t)\dot\psi(t)+\bar\psi(0)\psi(0) \;\to\; \sum_{n=1}^{M}\bar\psi(n\epsilon)\left[\psi(n\epsilon)-\psi(n\epsilon-\epsilon)\right]+\bar\psi(0)\psi(0). \tag{2.1.37}$$

For this discretization we immediately get

$$\left\langle\psi_i(n\epsilon)\bar\psi_j(m\epsilon)\right\rangle_\psi = \begin{cases}\delta_{ij} & \text{for } n\ge m\,,\\ 0 & \text{for } n<m\,.\end{cases} \tag{2.1.38}$$

The term $\bar\psi(0)\psi(0)$ is needed to provide nonvanishing integrals over $\bar\psi(0)$ and $\psi(0)$. It is also seen from the discretized version that the path integral in the denominator on the RHS of Eq. (2.1.34) is equal to unity.

The fermionic path-integral representation for the non-Abelian phase factor reads (see, *e.g.*, Ref. [GN80]) as

$$\left[\mathbf{P}\,\mathrm{e}^{\int_0^\tau dt\,\dot z^\mu(t)\mathcal{A}_\mu(z(t))}\right]_{ij} \;=\; \left\langle \mathrm{e}^{\int_0^\tau dt\,\dot z^\mu(t)\bar\psi(t)\mathcal{A}_\mu(z(t))\psi(t)}\psi_i(\tau)\bar\psi_j(0)\right\rangle_\psi . \tag{2.1.39}$$

There is no path-ordering sign on the RHS since the matrix indices of $\mathcal{A}_\mu$ are contacted by ψ and $\bar\psi$.

In order to prove Eq. (2.1.39), one expands the exponential in $\mathcal{A}_\mu$ and calculates the average using Eq. (2.1.35) and the rules of Wick's pairing which yields

$$\begin{aligned}
&\frac{1}{k!}\left\langle \psi_i(\tau)\left(\int_0^\tau dt\,\dot z^\mu(t)\,\bar\psi(t)\mathcal{A}_\mu(z(t))\psi(t)\right)^k \bar\psi_j(0)\right\rangle_\psi \\
&= \int_0^\tau dt_1\int_0^\tau dt_2\ldots\int_0^\tau dt_k\,\theta(\tau,t_k,\ldots,t_2,t_1,0)\,\dot z^{\mu_1}(t_1)\dot z^{\mu_2}(t_2)\cdots\dot z^{\mu_k}(t_k) \\
&\qquad\cdot\left[\mathcal{A}_{\mu_k}(z(t_k))\cdots\mathcal{A}_{\mu_2}(z(t_2))\mathcal{A}_{\mu_1}(z(t_1))\right]_{ij} ,
\end{aligned} \tag{2.1.40}$$

where $\theta(\tau,t_k,\ldots,t_2,t_1,0)$ is given by Eq. (2.1.30). It is crucial in the derivation of this formula that only connected terms contribute to the average (2.1.34). Equation (2.1.40) reproduces Eq. (2.1.28) from the previous Problem, which completes the proof of Eq. (2.1.39). Moreover, we can say that the path integral (2.1.34) is nothing but a nice representation of the thetas (2.1.30).

Problem 2.4 Invert $(-\nabla^2+m^2)$ when ∇_μ is in the fundamental representation.
Solution The calculation is quite analogous to that of the Problem 1.12. We first use the path-integral representation of the inverse operator:

$$\begin{aligned}
G(x,y;\mathcal{A}) \;&\equiv\; \left\langle y\left|\frac{1}{-\nabla_\mu^{\rm fun}\nabla_\mu^{\rm fun}+m^2}\right|x\right\rangle \\
&= \frac12\int_0^\infty d\tau\;\mathrm{e}^{-\frac12\tau m^2}\int_{z_\mu(0)=x_\mu} Dz_\mu(t)\;\mathrm{e}^{-\frac12\int_0^\tau dt\,\dot z_\mu^2(t)}\left\langle y\left|\mathbf{P}\,\mathrm{e}^{-\int_x^{z(\tau)}dz^\mu\nabla_\mu^{\rm fun}}\right|x\right\rangle .
\end{aligned} \tag{2.1.41}$$

The integral over $z(\tau)$ — the final point of the trajectory — of the matrix element on the RHS equals

$$\int d^d z(\tau)\left\langle y\left|\mathbf{P}\,\mathrm{e}^{-\int_x^{z(\tau)}dz^\mu\nabla_\mu^{\rm fun}}\right|x\right\rangle \;=\; \mathbf{P}\,\mathrm{e}^{\int_x^y dz^\mu\mathcal{A}_\mu(z)} . \tag{2.1.42}$$

Therefore, the result reads

$$G(x,y;\mathcal{A}) \;=\; \sum_{\Gamma_{yx}}{}'\,\mathbf{P}\,\mathrm{e}^{\int_{\Gamma_{yx}}dz^\mu\mathcal{A}_\mu(z)} , \tag{2.1.43}$$

where $\sum'$ is defined by Eq. (1.1.152).

Problem 2.5 Invert $(-\nabla^2+m^2)$ when ∇_μ is in the adjoint representation.
Solution Let us introduce

$$\nabla_\mu^{ab} \;=\; \partial_\mu\delta^{ab}-gf^{abc}A_\mu^c \tag{2.1.44}$$

and the Green function $G^{ab}(x,y;\mathcal{A})$ which obeys

$$\left(-\nabla_\mu^{ac}\nabla_\mu^{cb}+m^2\delta^{ab}\right)G^{bd}(x,y;\mathcal{A}) \;=\; \delta^{ad}\,\delta^{(d)}(x-y) . \tag{2.1.45}$$

Then we get

$$G^{ab}(x,y;\mathcal{A}) \;=\; \sum_{\Gamma_{yx}}{}'\,2\,\mathrm{tr}\,t^b\,U[\Gamma_{yx}]\,t^a\,U^\dagger[\Gamma_{yx}] , \tag{2.1.46}$$

where $U[\Gamma_{yx}]$ is given by Eq. (2.1.22).

Since matrices are rearranged in an inverse order under Hermitean conjugation, one has[2]

$$U^\dagger[\Gamma_{yx}] \;=\; U[\Gamma_{xy}] . \tag{2.1.47}$$

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

$$U[\Gamma_{yx}]\,U[\Gamma_{xy}] \;=\; 1 . \tag{2.1.48}$$

We have chosen $\mathcal{A}_\mu$ in the discretized phase factor (2.1.26) at the center on the i-th interval in order to satisfy Eq. (2.1.48) at finite ϵ.

Problem 2.6 Establish the relation between non-Abelian phase factors and the group of paths.
Solution The group of paths (or loops) is defined as follows. The elements of the group are the paths Γ_{yx}. The product of two elements Γ_{zx} and Γ_{yz} is the path Γ_{yx} which is a composition of Γ_{zx} and Γ_{yz}. In other words one passes first the path Γ_{zx} and then the path Γ_{yz}. The product is denoted as

$$\Gamma_{yz}\,\Gamma_{zx} \;=\; \Gamma_{yx} . \tag{2.1.49}$$

[2] The 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 (2.1.25), these two contours result in the opposite orders of multiplying the matrices.

The multiplication of paths is obviously associative but noncommutative. The inverse element is defined as

$$\Gamma_{yx}^{-1} = \Gamma_{xy}\,, \tag{2.1.50}$$

i.e. the path with opposite orientation.

It follows from definition (2.1.25) that

$$U[\Gamma_{yz}]\,U[\Gamma_{zx}] = U[\Gamma_{yz}\,\Gamma_{zx}]\,. \tag{2.1.51}$$

The backtracking relation (2.1.48) then reads

$$U[\Gamma_{yx}\,\Gamma_{xy}] = 1\,. \tag{2.1.52}$$

In other words the paths of opposite orientations cancel each other in the phase factors.

2.1.3 Phase factors (properties)

Under the gauge transformation (2.1.4) the non-Abelian phase factor (2.1.22) transforms as

$$U[\Gamma_{yx}] \stackrel{\text{g.t.}}{\longrightarrow} \Omega\,(y)\;U[\Gamma_{yx}]\;\Omega^\dagger\,(x)\;. \tag{2.1.53}$$

This formula stems from the fact that

$$\begin{aligned}[1+dz^\mu\mathcal{A}_\mu(z)] &\stackrel{\text{g.t.}}{\longrightarrow} [1+dz^\mu\mathcal{A}'_\mu(z)]\\ &= \Omega(z+dz)\,[1+dz^\mu\mathcal{A}_\mu(z)]\,\Omega^\dagger(z)\end{aligned} \tag{2.1.54}$$

under the gauge transformation, which can be proven by substituting Eq. (2.1.4), so that $\Omega^\dagger(z)$ and $\Omega(z)$ cancel in the definition (2.1.25) at the intermediate point z.

One of the consequences of Eq. (2.1.53) is that $\psi(x)$, transported by the matrix $U[\Gamma_{yx}]$ to the point y, transforms under the gauge transformation as $\psi(y)$:

$$U[\Gamma_{yx}]\;\psi(x) \stackrel{\text{g.t.}}{\sim} \text{“}\psi(y)\text{”}\,, \tag{2.1.55}$$

and, analogously,

$$\bar\psi(y)\;U[\Gamma_{yx}] \stackrel{\text{g.t.}}{\sim} \text{“}\bar\psi(x)\text{”}\,. \tag{2.1.56}$$

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:

$$\bar\psi(y)\,U[\Gamma_{yx}]\,\psi(x) \stackrel{\text{g.t.}}{\longrightarrow} \bar\psi(y)\,U[\Gamma_{yx}]\,\psi(x)\,. \tag{2.1.57}$$

Another consequence of Eq. (2.1.53) is that the trace of the phase factor for a closed contour Γ is gauge invariant:

$$\operatorname{tr}\mathbf{P}\,\mathrm{e}^{\oint_\Gamma dz^\mu\mathcal{A}_\mu(z)} \stackrel{\text{g.t.}}{\longrightarrow} \operatorname{tr}\mathbf{P}\,\mathrm{e}^{\oint_\Gamma dz^\mu\mathcal{A}_\mu(z)}\,. \tag{2.1.58}$$

These properties of the non-Abelian phase factor are quite similar to those of the Abelian one which was considered in Subsect. 1.1.7.

Problem 2.7 Calculate $\partial U[\Gamma_{yx}]/\partial x_\mu$ and $\partial U[\Gamma_{yx}]/\partial y_\mu$.

Solution It is convenient to start from Eq. (2.1.26). 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 $\epsilon\to 0$, we get

$$\begin{aligned}\frac{\partial}{\partial x_\mu}\mathbf{P}\,\mathrm{e}^{\int_x^y dz^\mu\mathcal{A}_\mu(z)} &= -\mathbf{P}\,\mathrm{e}^{\int_x^y dz^\mu\mathcal{A}_\mu(z)}\,\mathcal{A}_\mu\,(x)\,,\\ \frac{\partial}{\partial y_\mu}\mathbf{P}\,\mathrm{e}^{\int_x^y dz^\mu\mathcal{A}_\mu(z)} &= \mathcal{A}_\mu\,(y)\;\mathbf{P}\,\mathrm{e}^{\int_x^y dz^\mu\mathcal{A}_\mu(z)}\,.\end{aligned} \tag{2.1.59}$$

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.

One can rewrite Eq. (2.1.59) via the covariant derivatives as

$$\begin{aligned}\nabla_\mu^{\text{fun}}(y)\,U[\Gamma_{yx}] &= 0\,,\\ U[\Gamma_{yx}]\;\overleftarrow{\nabla}_\mu^{\text{fun}}(x) &= 0\,.\end{aligned} \tag{2.1.60}$$

It is the property of the parallel transporter which is annihilated by the covariant derivative.

Problem 2.8 Prove that the sufficient and necessary condition for the phase factor to be independent on a local variation of the path is the vanishing of $\mathcal{F}_{\mu\nu}$.

Solution Let us add to Γ_{yx} at the point $z\in\Gamma_{yx}$ an infinitesimal loop δC_{zz} which lies in the μ,ν-plane and encloses the area $\delta\sigma_{\mu\nu}(z)$. Then the variation of the phase factor is

$$\delta U[\Gamma_{yx}] \equiv U[\Gamma_{yz}\,\delta C_{zz}\,\Gamma_{zx}]-U[\Gamma_{yx}] = U[\Gamma_{yz}]\,\mathcal{F}_{\mu\nu}(z)\,U[\Gamma_{zx}]\,\delta\sigma_{\mu\nu}(z)\,. \tag{2.1.61}$$

We can rewrite Eq. (2.1.61) as

$$\delta U[\Gamma_{yx}] = \mathbf{P}\,U[\Gamma_{yx}]\;\mathcal{F}_{\mu\nu}(z)\,\delta\sigma_{\mu\nu}(z) \tag{2.1.62}$$

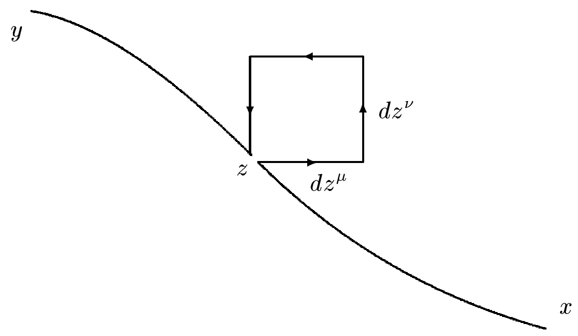


Fig. 2.1: The rectangular loop δC_{zz}, which is added to the contour Γ_{yx} at the intermediate point z in the μ,ν-plane.

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

A convenient way to prove Eq. (2.1.61) is to choose δC_{zz} to be a rectangle which is built on the vectors dz^μ and dz^ν, as is depicted in Fig. 2.1. Using the representation (2.1.42), we see that the phase factor acquires the extra factor

$$[1 + dz^\nu \nabla_\nu]\,[1 + dz^\mu \nabla_\mu]\,[1 - dz^\nu \nabla_\nu]\,[1 - dz^\mu \nabla_\mu] \;=\; 1 - dz^\mu dz^\nu\,[\nabla_\mu, \nabla_\nu] \quad (2.1.63)$$

at the proper order in the path-ordered product. Then Eq. (2.1.20) results in Eq. (2.1.62). Alternatively, one can prove Eq. (2.1.62) using the discretized formula (2.1.26).

Problem 2.9 Derive a non-Abelian version of the Stokes theorem.
Solution The ordered contour integral can be represented as the double-ordered surface integral [Are80, Bra80]

$$\mathbf{P}\exp\oint\limits_{C_{xx}} dz^\mu\,\mathcal{A}_\mu(z) \;=\; \mathbf{P}_\sigma\,\mathbf{P}_\tau\,\exp\int\limits_S d\sigma^{\mu\nu}\;\text{“}\mathcal{F}_{\mu\nu}(x)\text{”} \quad (2.1.64)$$

where τ and σ parametrize the surface S (spanned by C but arbitrary otherwise) whose element is given by

$$d\sigma^{\mu\nu} \;=\; d\tau d\sigma\left(\frac{\partial z_\mu}{\partial\tau}\frac{\partial z_\nu}{\partial\sigma} - \frac{\partial z_\mu}{\partial\sigma}\frac{\partial z_\nu}{\partial\tau}\right). \quad (2.1.65)$$

“$\mathcal{F}_{\mu\nu}(x)$” in Eq. (2.1.64) means that $\mathcal{F}_{\mu\nu}(z(\tau,\sigma))$ is parallel-transported, say, to the initial point x.

Remark on analogy with differential geometry

The formulas of the type of Eq. (2.1.61) 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 [Wey19] 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, *i.e.* the vector-potential was not multiplied by i as in Eq. (1.1.154). This explains the term “gauge invariance” — gauging literally means fixing a scale. The factor of i was inserted by London [Lon27] 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.

2.1.4 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. 1.1.7, to its value at some reference point x by the parallel transport which is given by Eq. (1.1.159). 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. 2.2.

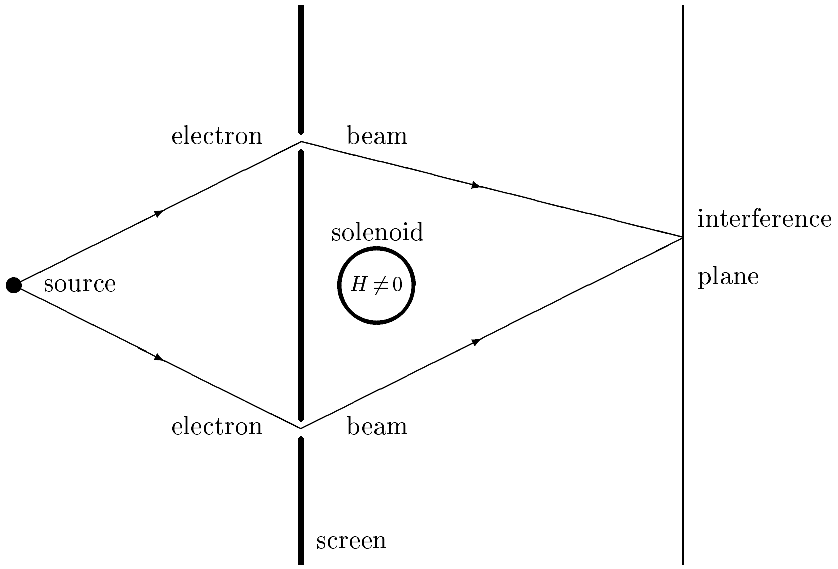


Fig. 2.2: Principal scheme of the experiment which demonstrates the Aharonov–Bohm effect. Electrons do not pass inside the solenoid where the magnetic field is concentrated. Nevertheless, a phase difference arises between the electron beams passing through the two slits. The interference picture changes with the value of electric current.

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. (1.1.151):

$$\Psi(x,y) \;=\; \sum_{\Gamma^+_{yx}}{}' \, \mathrm{e}^{\,ie\int_{\Gamma^+_{yx}} dz^\mu \, A_\mu(z)} + \sum_{\Gamma^-_{yx}}{}' \, \mathrm{e}^{\,ie\int_{\Gamma^-_{yx}} dz^\mu \, A_\mu(z)} \,, \qquad (2.1.66)$$

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

$$\mathrm{e}^{\,ie\int_{\Gamma^+_{yx}} dz^\mu \, A_\mu(z)} \, \mathrm{e}^{\,-ie\int_{\Gamma^-_{yx}} dz^\mu \, A_\mu(z)} \;=\; \mathrm{e}^{\,ie\oint_{\Gamma} dz^\mu \, A_\mu(z)} \,, \qquad (2.1.67)$$

where the closed contour Γ is composed from Γ^+_{yx} and Γ^-_{xy}. 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

$$\mathrm{e}^{\,ie\oint_{\Gamma} d\xi^\mu \, A_\mu} \;=\; e^{ie\int d\sigma^{\mu\nu} \, F_{\mu\nu}} \;=\; e^{ieHS} \,, \qquad (2.1.68)$$

where HS is the magnetic flow through the solenoid. Therefore, the interference picture changes when H changes[3].

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 $G = \mathrm{SU}(N_c)$, a quark can alter its color under the parallel transport so the non-Abelian phase factor (2.1.22) 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. (2.1.58).

It looks suggestive to reformulate gauge theories entirely in terms of these observable quantities. How to do that will be explained in Chapter 3.

[3] A detailed computation of the interference picture for the Aharonov–Bohm experiment is contained, *e.g.*, in the review by Kobe [Kob79].

2.2 Gauge fields on a lattice

The modern formulation of non-Abelian lattice gauge theories is due to Wilson [Wil74]. Independently, gauge theories were discussed on a lattice by Wegner [Weg71] 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.

2.2.1 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 approach[4]. We shall discuss only the Euclidean formulation of lattice gauge theories.

[4]A Hamiltonian formulation of lattice gauge theories was developed by Kogut and Susskind [KS75].

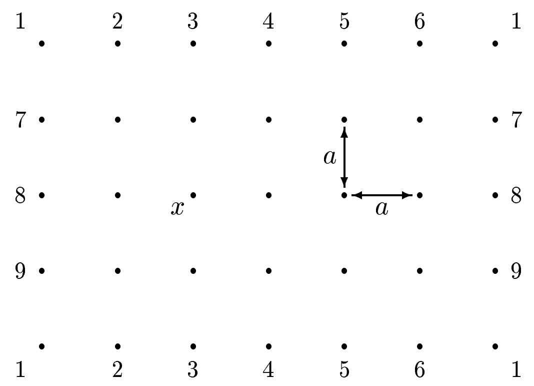


Fig. 2.3: Two-dimensional lattice with periodic boundary conditions. The sites labelled by the same numbers are identified. The lattice spacing equals a while the spatial size of the lattice corresponds to $L_1 = 6$ and $L_2 = 4$.

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

$$x_\mu \;=\; n_\mu a\,, \tag{2.2.1}$$

where the components of the vector

$$n_\mu \;=\; (n_1, n_2, \ldots, n_d) \tag{2.2.2}$$

are integer numbers. The points (2.2.1) are called the lattice *sites*.

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. 2.3. A four-dimensional lattice for which the distances between sites are the same in all directions (as for the lattice in Fig. 2.3) 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 $\mu = 1, \ldots, d$:

$$l \;=\; \{x, \mu\}\,. \tag{2.2.3}$$

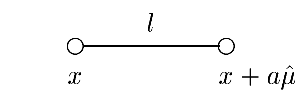


Fig. 2.4: A link of a lattice. The link connects the sites x and $x+a\hat{\mu}$.

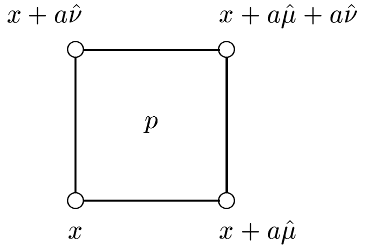


Fig. 2.5: A plaquette of a lattice. The plaquette boundary is made of four links.

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. 2.4. 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:

$$p = \{x;\mu,\nu\}\,. \tag{2.2.4}$$

A plaquette is depicted in Fig. 2.5. The set of four links which bound the plaquette p is 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. 2.3).

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 $(0,n_2,\ldots,n_d)$ and $(L_1,n_2,\ldots,n_d)$ are identified and similarly along other axes.

Problem 2.10 Calculate the numbers of sites, links and plaquettes for a symmetric hypercubic lattice with periodic boundary conditions.
Solution Let us denote $L_1 = L_2 = \ldots = L_d = L$. Then

$$N_s = L^d\,, \qquad N_l = dL^d\,, \qquad N_p = \frac{d(d-1)}{2}L^d\,. \tag{2.2.5}$$

Problem 2.11 Label the lattice links by a natural number $l \in [1, N_l]$.
Solution One of the choices is as follows

$$l = \mu + n_1 d + n_2 dL + \ldots + n_d dL^{d-1}\,, \tag{2.2.6}$$

where $n_\nu = x_\nu/a$ and μ is the direction of the link $\{x,\mu\}$.

2.2.2 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 sites[5]:

$$\varphi(x) \Longrightarrow \varphi_x\,. \tag{2.2.7}$$

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. 2.6.

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

$$\mathcal{A}_\mu(x) \Longrightarrow U_{x,\mu}\,. \tag{2.2.8}$$

It looks natural since a link is characterized by a coordinate and a direction (see Eq. (2.2.3)) — the same as $\mathcal{A}_\mu(x)$. Sometimes the notation $U_\mu(x)$ is alternatively used for $U_{x,\mu}$.

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

$$U_{x,\mu} = \mathbf{P}\,\mathrm{e}^{\int_x^{x+a\hat{\mu}} dz^\mu\,\mathcal{A}_\mu(z)}\,, \tag{2.2.9}$$

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

$$U_{x,\mu} \to \mathrm{e}^{\,a\mathcal{A}_\mu(x)} \tag{2.2.10}$$

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

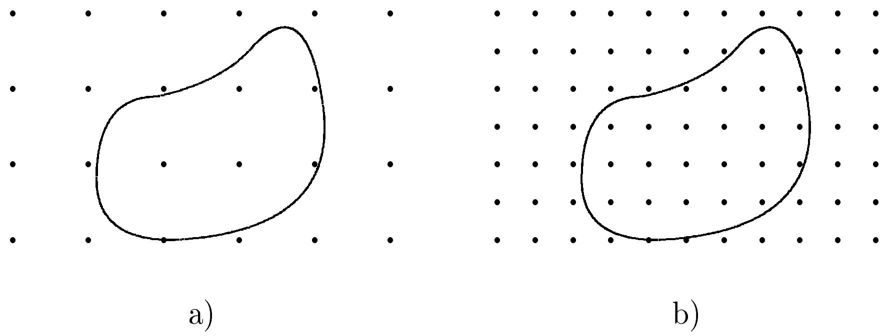


Fig. 2.6: Description of continuum field configurations by lattices a) "coarse" and b) "fine". The lattice a) can represent the given continuum field configuration very roughly, while the lattice b) has the spacing which is small enough.

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. (2.1.26).

Since the path-ordered integral in Eq. (2.2.9) 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

$$U_{x+a\hat{\mu},-\mu} \;=\; U^{\dagger}_{x,\mu}\,. \tag{2.2.11}$$

This is a one-link analog of Eq. (2.1.47).

It is clear from the relation (2.2.9) 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. 2.7. The plaquette variable is composed from the link variables (2.2.9) as

$$U_{\partial p} \;=\; U^{\dagger}_{x,\nu}U^{\dagger}_{x+a\hat{\nu},\mu}U_{x+a\hat{\mu},\nu}U_{x,\mu}\,. \tag{2.2.12}$$

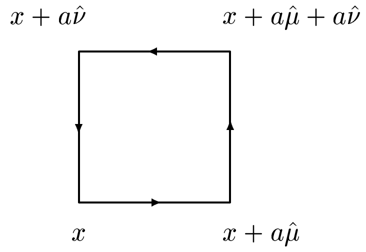


Fig. 2.7: A contour in the form of an oriented boundary of a plaquette.

The link variable transforms under the gauge transformation, according to Eq. (2.1.53), as

$$U_{x,\mu} \;\stackrel{\text{g.t.}}{\longrightarrow}\; \Omega_{x+a\hat{\mu}}U_{x,\mu}\Omega^{\dagger}_{x}\,, \tag{2.2.13}$$

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

$$U_{\partial p} \;\stackrel{\text{g.t.}}{\longrightarrow}\; \Omega_x U_{\partial p}\Omega^{\dagger}_{x}\,. \tag{2.2.14}$$

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

$$\operatorname{tr} U_{\partial p} \;\stackrel{\text{g.t.}}{\longrightarrow}\; \operatorname{tr} U_{\partial p}\,. \tag{2.2.15}$$

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

$$S_{\text{Lat}}[U] \;=\; \sum_p \left(1 - \frac{1}{N_c}\operatorname{Re}\operatorname{tr} U_{\partial p}\right)\,. \tag{2.2.16}$$

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. (2.1.47), in complex conjugation:

$$\operatorname{tr} U_{\partial p} \;\stackrel{\text{reorient}}{\longrightarrow}\; \operatorname{tr} U^{\dagger}_{\partial p} = \left(\operatorname{tr} U_{\partial p}\right)^{*}\,, \tag{2.2.17}$$

one can rewrite the action (2.2.16) in the equivalent form

$$S_{\mathrm{Lat}}[U] \;=\; \frac{1}{2}\sum_{\text{oriented } p}\left(1-\frac{1}{N_c}\mathrm{tr}\,U_{\partial p}\right), \tag{2.2.18}$$

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

In the limit $a \to 0$, the lattice action (2.2.16) becomes in $d=4$ the action of a continuum gauge theory. In order to show this, let us first note that

$$U_{\partial p} \;\to\; \exp\left[a^2\mathcal{F}_{\mu\nu}(x)+\mathcal{O}(a^3)\right], \tag{2.2.19}$$

where $\mathcal{F}_{\mu\nu}(x)$ is defined by Eq. (2.1.15).

In the Abelian theory, the expansion (2.2.19) 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. Equation (2.2.19) was in fact already derived in Problem 2.8.

The transition to the continuum limit is performed by virtue of

$$a^4\sum_p \;\stackrel{a\to 0}{\longrightarrow}\; \frac{1}{2}\int d^4x\sum_{\mu,\nu}. \tag{2.2.20}$$

Expanding the exponential on the RHS of Eq. (2.2.19) in a, we get

$$S_{\mathrm{Lat}} \;\stackrel{a\to 0}{\longrightarrow}\; -\frac{1}{4N_c}\int d^4x\sum_{\mu,\nu}\mathrm{tr}\,\mathcal{F}^2_{\mu\nu}(x), \tag{2.2.21}$$

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

Problem 2.12 Derive the lattice version of the non-Abelian Maxwell equation (2.1.18).
Solution : Let us perform the variation of the link variable

$$U_{x,\mu}\to U_{x,\mu}\left(1-i\epsilon_{x,\mu}\right), \qquad U^\dagger_{x,\mu}\to\left(1+i\epsilon_{x,\mu}\right)U^\dagger_{x,\mu}, \tag{2.2.22}$$

where $\epsilon_{x,\mu}$ is an infinitesimal traceless Hermitean matrix.

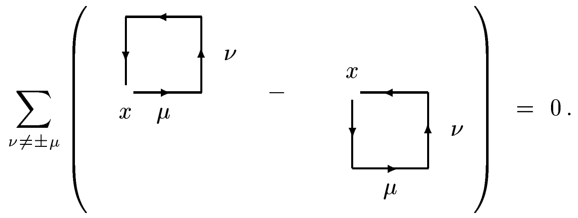
A given link $\{x,\mu\}$ enters $4(d-1)$ plaquettes $p=\{x;\mu,\nu\}$ in the action (2.2.18). One half of them has boundary with positive orientation and the other half with negative one. The variation of the action (2.2.18) under the shift (2.2.22) is

$$\delta S[U] \;=\; \frac{i}{2N_c}\sum_{\nu\neq\pm\mu}\left(\mathrm{tr}\,U_{\partial p}\epsilon_{x,\mu}-\mathrm{tr}\,\epsilon_{x,\mu}U^\dagger_{\partial p}\right). \tag{2.2.23}$$

Since $\epsilon_{x,\mu}$ is arbitrary, we get

$$\sum_{\nu\neq\pm\mu}\left(U_{\partial\{x;\mu,\nu\}}-U^\dagger_{\partial\{x;\mu,\nu\}}\right) \;=\; 0, \tag{2.2.24}$$

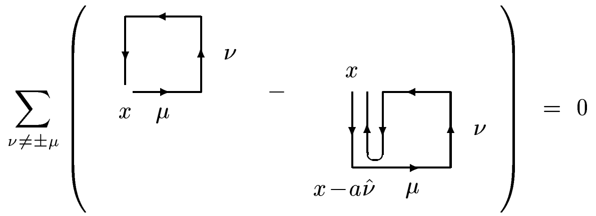
or, graphically,



(2.2.25)

In the last equation we have depicted only plaquettes with the positive orientation, while those with the negative one are recovered by the sum over ν for $\nu<0$. Equation (2.2.24) (or (2.2.25)) is the lattice analog of the non-Abelian Maxwell equation.

In order to show how this equation reproduces the continuum one (2.1.18) as $a\to 0$, let us rewrite the second term on the LHS of Eq. (2.2.25) using (2.2.11):



(2.2.26)

or, analytically,

$$\sum_{\nu\neq\pm\mu}\left(U_{\partial\{x;\mu,\nu\}}-U_{x-a\hat{\nu},\nu}U_{\partial\{x-a\hat{\nu};\mu,\nu\}}U^\dagger_{x-a\hat{\nu},\nu}\right) \;=\; 0. \tag{2.2.27}$$

It is now clear that the plaquette boundary in the second term on the LHS, which is the same as the first one but transported by one lattice spacing in the ν-direction, is associated with $\mathcal{F}_{\mu\nu}(x-a\hat{\nu})$. Using Eqs. (2.2.10) and (2.2.19), we recover the continuum Maxwell equation (2.1.18).

Remark on the naive continuum limit

The limit $a\to 0$, when Eqs. (2.2.10) and (2.2.19) hold reproducing the continuum action (2.2.21), 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 (2.2.16), say the term $\propto a^2 \mathrm{tr}\, \mathcal{F}^3$, 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. 2.2.7 below.

Remark on ambiguities of the lattice action

The Wilson action (2.2.16) 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 $\mathrm{SU}(N_c)$, *e.g.* in the adjoint representation

$$\chi_{\mathrm{adj}}(U) \;=\; |\mathrm{tr}\, U|^2 - 1\,, \tag{2.2.28}$$

to construct the lattice action.

The adjoint-representation lattice action reads

$$S_{\mathrm{adj}}[U] \;=\; \sum_p \left(1 - \frac{1}{N_c^2}\,|\mathrm{tr}\, U_{\partial p}|^2\right)\,. \tag{2.2.29}$$

The naive continuum limit will be the same as for the Wilson action (2.2.16).

Moreover, one can define the lattice action as a mixture of the fundamental and adjoint representations [BC81, KM81]:

$$S_{\mathrm{mixed}}[U] = \sum_p \left(1 - \frac{1}{N_c}\mathrm{Re}\,\mathrm{tr}\, U_{\partial p}\right) + \frac{\beta_A}{2\beta}\sum_p \left(1 - \frac{1}{N_c^2}\,|\mathrm{tr}\, U_{\partial p}|^2\right)\,. \tag{2.2.30}$$

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 (2.2.29) for $N_c = 2$ is associated with the action of the SO(3) lattice gauge theory. Since algebras of the SU(2) and 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 to construct the action. 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 in called the *universality.* We shall say more about this in the next Section when discussing renormalization group on the lattice.

2.2.3 The Haar measure

The partition function of a pure[6] lattice gauge theory is defined by

$$Z(\beta) \;=\; \int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{e}^{\,-\beta S[U]}\,, \tag{2.2.31}$$

where the action is given by Eq. (2.2.16).

This is an analog of a partition function in statistical mechanics at an inverse temperature β given by[7]

$$\beta = \frac{2N_c}{g^2}\,. \tag{2.2.32}$$

This formula results from Eq. (2.2.21).

A subtle question is what is the measure $dU_{x,\mu}$ in Eq. (2.2.31). 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:

$$dU = d(\Omega U) = d(U\Omega')\,, \tag{2.2.33}$$

guarantees the gauge invariance of the partition function (2.2.31).

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

[6] Here "pure" means without matter fields.

[7] One has instead $\beta = 2N_c/g^2a^{4-d}$ on a d-dimensional lattice since the Yang–Mills coupling g is dimensionful for $d \neq 4$.

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

$$U = a_4 \mathrm{I} + i\vec{a}\vec{\sigma} \tag{2.2.34}$$

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

$$dU = \frac{1}{\pi^2} \prod_{\mu=1}^{4} da_\mu \, \delta^{(1)}\left(a_\mu^2 - 1\right), \tag{2.2.35}$$

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

Problem 2.13 Rewrite the Haar measure on SU(2) via a unit three-vector $\vec{n}$ ($\vec{n}^2 = 1$) and an angle φ ($\varphi \in [0, 2\pi]$).
Solution An element of SU(2) reads in this parametrization as

$$U = \mathrm{e}^{\,i\varphi\vec{n}\vec{\sigma}/2} = \cos\frac{\varphi}{2} + i\vec{n}\vec{\sigma}\sin\frac{\varphi}{2}. \tag{2.2.36}$$

The geometric meaning of this parametrization is simple: the element (2.2.36) is associated with a rotation through the angle φ around the $\vec{n}$-axis. The Haar measure for the SU(2) group then is

$$dU = \frac{d^2\vec{n}}{4\pi}\frac{d\varphi}{\pi}\sin^2\frac{\varphi}{2}. \tag{2.2.37}$$

This formula can be obtained from Eq. (2.2.35) by integrating over $|\vec{a}|$.

Problem 2.14 For the U(N_c) group represent the Haar measure as a multiple integral over the matrix elements of U.
Solution Elements of a unitary matrix U are complex numbers $U_{ij} = U_{ji}^*$. The Haar measure can be represented as

$$\int dU \ldots = \int_{-\infty}^{\infty} \prod_{i,j} d\,\mathrm{Re}\,U_{ij}\, d\,\mathrm{Im}\,U_{ij}\, \delta^{\left(N_c^2\right)}\left(UU^\dagger - 1\right)\ldots \tag{2.2.38}$$

The integral in this formula goes over unrestricted U_{ij} as if U were a general complex matrix while the delta-function restricts U to be unitary.

The partition function (2.2.31) characterizes vacuum effects in the quantum theory. Physical quantities are given by the averages of the type of Eq. (1.2.6):

$$\langle F[U] \rangle = Z^{-1}(\beta) \int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{e}^{-\beta S[U]} F[U], \tag{2.2.39}$$

where $F[U]$ is a gauge invariant functional of the link variable $U_{x,\mu}$. The averages (2.2.39) become the corresponding expectation values in the continuum theory as $a \to 0$ and β is related to g^2 by Eq. (2.2.32).

Remark on the lattice quantization

On a lattice of finite size, the integral over the gauge group in Eq. (2.2.39) 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 (2.2.39) 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. (2.2.39). 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 [Gri78]. In contrast, the lattice path integral (2.2.39) 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.

2.2.4 Wilson loops

As is already mentioned in Subsect. 2.2.2, 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:

$$C = \{x; \mu_1, \ldots, \mu_n\}. \tag{2.2.40}$$

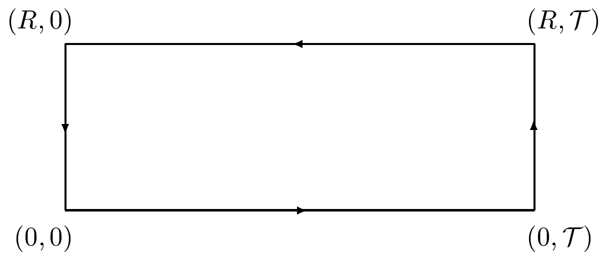


Fig. 2.8: Rectangular loop of the size $R \times \mathcal{T}$.

Then the lattice phase factor U_C reads as

$$U_C \;=\; U_{x+a\hat{\mu}_1+\ldots+a\hat{\mu}_{n-1},\mu_n} \cdots U_{x+a\hat{\mu}_1,\mu_2} U_{x,\mu_1} \,. \tag{2.2.41}$$

For the links with a negative direction it is again convenient to use Eq. (2.2.11).

A closed contour has $\hat{\mu}_1+\ldots+\hat{\mu}_n=0$. 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 (2.2.39) to be

$$\begin{aligned} W_C \;&\equiv\; \left\langle \frac{1}{N_c} \operatorname{tr} U_C \right\rangle \\ &=\; Z^{-1}(\beta) \int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{e}^{-\beta S[U]} \frac{1}{N_c} \operatorname{tr} U_C \,. \end{aligned} \tag{2.2.42}$$

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. 2.8.

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

$$W_{R\times\mathcal{T}} \;\stackrel{\mathcal{T}\gg R}{=}\; \mathrm{e}^{-E_0(R)\cdot\mathcal{T}} \,. \tag{2.2.43}$$

Problem 2.15 Derive Eq. (2.2.43) by fixing the gauge $\mathcal{A}_4=0$.
Solution In the axial gauge $\mathcal{A}_4=0$, we have $U_{x,4}=1$ so that only vertical

segments of the rectangle in Fig. 2.8 contribute to $U_{R\times\mathcal{T}}$. Denoting

$$\Psi_{ij}(t) \equiv \left[\mathbf{P}\, \mathrm{e}^{\int_0^R dz_1\, A_1(z_1,\ldots,t)} \right]_{ij} , \tag{2.2.44}$$

we then have

$$W_{R\times\mathcal{T}} \;=\; \left\langle \frac{1}{N_c} \operatorname{tr} \Psi_{ij}(0)\, \Psi^\dagger_{ij}(\mathcal{T}) \right\rangle . \tag{2.2.45}$$

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

$$\sum_n |n\rangle\langle n| \;=\; 1 \,, \tag{2.2.46}$$

we get

$$\begin{aligned} W_{R\times\mathcal{T}} \;&=\; \sum_n \frac{1}{N_c} \operatorname{tr} \left\langle \Psi_{ij}(0) \right| n \rangle \left\langle n \left| \Psi^\dagger_{ji}(\mathcal{T}) \right\rangle \right. \\ &=\; \sum_n \frac{1}{N_c} \left| \left\langle \Psi_{ij}(0) \right| n \rangle \right|^2 \mathrm{e}^{-E_n \mathcal{T}} \end{aligned} \tag{2.2.47}$$

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

$$W_{R\times\mathcal{T}} \;\stackrel{\text{large } \mathcal{T}}{\longrightarrow}\; \mathrm{e}^{-E_0\mathcal{T}} , \tag{2.2.48}$$

which results in Eq. (2.2.43).

Notice that nothing relies in this derivation on the lattice. Therefore, Eq. (2.2.43) holds for a rectangular loop in a continuum theory as well.

Equation (2.2.43) 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

$$e \oint_C dz^\mu A_\mu(z) \;=\; \int d^d x J^\mu(x) A_\mu(x) \,, \tag{2.2.49}$$

where

$$J^\mu(x) \;=\; e \oint_C dz^\mu \delta^{(d)}(x-z) \tag{2.2.50}$$

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

$$-\ln W(C) \;=\; -\ln \left\langle \mathrm{e}^{\,i\int d^4x J^\mu(x) A_\mu(x)} \right\rangle \tag{2.2.51}$$

determines a change of the action of the classical particle due to electromagnetic interaction in accordance with Eq. (2.2.43).

A similar interpretation of Eq. (2.2.43) 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 $\mathcal{A}_\mu(x)$, 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. (2.2.43) is associated with the interaction energy averaged over color in this way.

Problem 2.16 Derive a non-Abelian analog of Eq. (2.2.50).
Solution The proper non-Abelian extension of Eq. (2.2.50) is [Won70]

$$\mathcal{J}^a_\mu(x) \;=\; g\int_0^\tau dt\, \dot z_\mu(t)\, \delta^{(d)}(x-z(t))\, I^a(t) \tag{2.2.52}$$

where $I^a(t)$, which describes the color state of a classical particle moving along the trajectory $z^\mu(t)$ in an external Yang–Mills field $A_\mu(z)$, is a solution of the equation

$$\dot I^a(t) + g f^{abc} \dot z^\mu(t)\, A^b_\mu(z(t))\, I^c(t) \;=\; 0\,. \tag{2.2.53}$$

It is convenient to use again Grassmann variables to describe color degrees of freedom as in Problem 2.3. Then [BCL77, BSSW77]

$$\mathcal{J}^a_\mu(x) \;=\; \bar\psi(t) t^a \psi(t) \tag{2.2.54}$$

and $\psi(t)$ is a solution of

$$\dot\psi(t) - z^\mu(t)\, \mathcal{A}_\mu(z(t))\, \psi(t) \;=\; 0. \tag{2.2.55}$$

Remark on mass renormalization

By definition, $E_0(R)$ in Eq. (2.2.43) 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

$$\Delta E_{\mathrm{mass}} \;=\; \frac{g^2}{4\pi a}\frac{N_c^2-1}{2N_c} \tag{2.2.56}$$

as $a \to 0$.[8]

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

$$E(R) \;=\; E_0(R) - \Delta E_{\mathrm{mass}}\,. \tag{2.2.57}$$

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

2.2.5 Strong coupling expansion

We already mentioned in Subsect. 2.2.3 that the path integral (2.2.39) can be calculated by the lattice perturbation theory in g^2. As was pointed out by Wilson [Wil74], 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. (2.2.32). 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. (2.2.42), in β. Then the problem is to calculate the integrals over the unitary group of the form

$$I^{\,i_1\cdots i_m, k_1\cdots k_n}_{\,j_1\cdots j_m, l_1\cdots l_n} \;=\; \int dU\, U^{i_1}_{j_1}\ldots U^{i_m}_{j_m} U^{\dagger k_1}_{\;l_1}\ldots U^{\dagger k_n}_{\;l_n}\,, \tag{2.2.58}$$

where the Haar measure (given for SU(2) by Eq. (2.2.35)) is normalized by

$$\int dU \;=\; 1\,. \tag{2.2.59}$$

[8]The calculation is presented below in Problem 3.20.

It is clear from general arguments that the integral (2.2.58) 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:

$$\int dU\, U^i_j\, {U^\dagger}^k_l \;=\; \frac{1}{N_c}\delta^i_l\delta^k_j\,. \tag{2.2.60}$$

Problem 2.17 Prove Eq. (2.2.60) for the U(N_c) group.
Solution From the general arguments we get

$$\int dU\, U^i_j\, {U^\dagger}^k_l \;=\; A\,\delta^i_l\delta^k_j + B\,\delta^i_j\delta^k_l\,. \tag{2.2.61}$$

Contracting by δ^l_i, using the unitarity of U, and Eq. (2.2.59), we have

$$AN_c + B \;=\; 1\,. \tag{2.2.62}$$

One more relation between A and B comes from the fact that the character in the adjoint representation is given by Eq. (2.2.28). Contracting Eq. (2.2.61) by δ^j_i and δ^l_k, and using the orthogonality of characters which says

$$\int dU\,\left(|\mathrm{tr}\,U|^2 - 1\right) \;=\; 1\,, \tag{2.2.63}$$

we get

$$AN_c + BN_c^2 \;=\; 1\,. \tag{2.2.64}$$

Therefore, $A = 1/N_c$ and $B = 0$ which proves Eq. (2.2.60).

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. 2.7). It is called the plaquette average and is denoted by

$$W_{\partial p} \;=\; \left\langle \frac{1}{N_c}\mathrm{tr}\,U_{\partial p}\right\rangle\,. \tag{2.2.65}$$

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. (2.2.42):

$$W_{\partial p} \;=\; \frac{\int\prod\limits_{x,\mu} dU_{x,\mu}\left(1+\beta\sum\limits_{p'}\frac{1}{N_c}\mathrm{Re\,tr}\,U_{\partial p'}\right)\frac{1}{N_c}\mathrm{tr}\,U_{\partial p}}{\int\prod\limits_{x,\mu} dU_{x,\mu}\left(1+\beta\sum\limits_{p'}\frac{1}{N_c}\mathrm{Re\,tr}\,U_{\partial p'}\right)} + \mathcal{O}\left(\beta^2\right)\,. \tag{2.2.66}$$

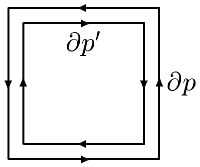


Fig. 2.9: Boundaries of the plaquettes p and p' with the opposite orientations: ∂p and $\partial p'$, respectively.

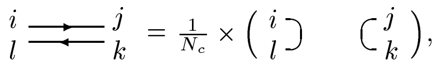
The group integration can then be performed remembering that

$$\int dU_{x,\mu}\,\left[U_{x,\mu}\right]^i_j\,\left[U^\dagger_{y,\nu}\right]^k_l \;=\; \frac{1}{N_c}\delta_{xy}\,\delta_{\mu\nu}\,\delta^i_l\,\delta^k_j \tag{2.2.67}$$

at different links.

Using this property of the group integral in Eq. (2.2.66), 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. 2.9.

It is convenient to use the graphic notations[9] for Eq. (2.2.60) at each link of ∂p:



(2.2.68)

where the semicircles are associated with the Kronecker deltas:

$${}^{i}_{l}\!\supset \;=\; \delta^i_l\,. \tag{2.2.69}$$

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,

$$\bigcirc \;=\; \delta^i_i \;=\; N_c\,. \tag{2.2.70}$$

[9]A calculation of more complicated group integrals (2.2.58) in the graphic notations is discussed in the lectures by Wilson [Wil75] and in the Chapter 8 of the book by Creutz [Cre83]. An alternative way of calculating the group integrals by the character expansion is described in the review by Drouffe and Zuber [DZ83].

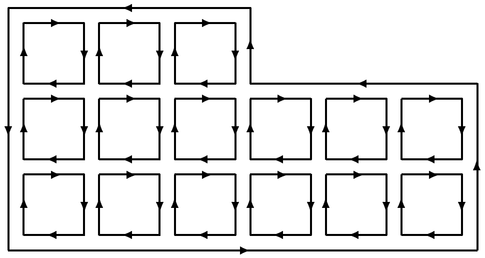
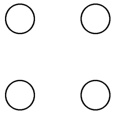


Fig. 2.10: Filling of a loop with elementary plaquettes.

Using the graphic representation (2.2.68) for each of the four links depicted in Fig. 2.9, we get

$$\int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{tr}\, U_{\partial p}\, \mathrm{tr}\, U^{\dagger}_{\partial p'} \;=\; \frac{1}{N_c^4}\times$$



$$=\; 1\,, \tag{2.2.71}$$

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

The final answer for the plaquette average is

$$\begin{aligned} W_{\partial p} &= \frac{\beta}{2N_c^2} \qquad \text{for SU}(N_c) \text{ with } N_c \geq 3\,, \\ W_{\partial p} &= \frac{\beta}{4} \qquad \text{for SU}(2)\,. \end{aligned} \tag{2.2.72}$$

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

The graphic representation (2.2.68) is useful for evaluating the leading order of the strong coupling expansion for more complicated loops. According to Eq. (2.2.67), a nonvanishing result emerges only when plaquettes, coming from the expansion of the exponentials of Eq. (2.2.42) in β, completely cover a surface enclosed by the given loop C as is depicted in Fig. 2.10. 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

$$W_C \;=\; (W_{\partial p})^{A_{\min}(C)}\,, \tag{2.2.73}$$

where $W_{\partial p}$ is given by Eq. (2.2.72) 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. 2.8, the minimal surface is just a piece of the plane bounded by the rectangle. Therefore, we get

$$W_{R\times\mathcal{T}} \;=\; (W_{\partial p})^{R\,\mathcal{T}} \tag{2.2.74}$$

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 β^{Area}. They are suppressed at small β since their areas are larger than $A_{\min}$.

2.2.6 Area law and confinement

The exponential dependence of the Wilson loop average on the area of the minimal surface (as in Eq. (2.2.73)) 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 SU(3) gauge theory) then quarks are confined. In other words, there are no physical $|in\rangle$ or $\langle out|$ 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 [Wil74], where this point is clarified.

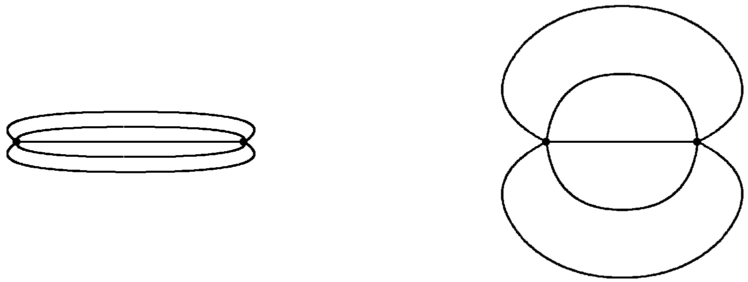
Another, somewhat oversimplified, justification for the Wilson criterion is based on the relationship (2.2.43) between the Wilson loop average and the potential energy of interaction between static quarks. When the area law

$$W(C) \;\stackrel{\text{large } C}{\longrightarrow}\; \mathrm{e}^{-K A_{\min}(C)} \tag{2.2.75}$$

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

$$E\,(R) \;=\; K R\,. \tag{2.2.76}$$

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. 2.11a. The value of K



a) b)

Fig. 2.11: 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.

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 (2.2.74) gives

$$K = \frac{1}{a^2} \ln \frac{2N_c^2}{\beta} = \frac{1}{a^2} \ln \left(N_c g^2\right) \tag{2.2.77}$$

for the string tension to the leading order of the strong coupling expansion. Next orders of the strong coupling expansion result 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

$$E(R) = -\frac{g^2}{4\pi R} \frac{N_c^2 - 1}{2N_c}, \tag{2.2.78}$$

the gauge field between quarks would be distributed over the whole space as is depicted in Fig. 2.11b. The Wilson loop average would have the behavior

$$W(C) \stackrel{\text{large } C}{\longrightarrow} \mathrm{e}^{-\mathrm{const}\cdot L(C)}, \tag{2.2.79}$$

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 (2.2.79), rather than the area law (2.2.75), 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 [Cre80] proposed to consider the ratio

$$\chi(I,J) = -\ln \frac{W_{I\times J} W_{(I-1)\times(J-1)}}{W_{(I-1)\times J} W_{I\times(J-1)}}, \tag{2.2.80}$$

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

$$L(I\times J) = 2I + 2J, \tag{2.2.81}$$

cancel out in the ratio (2.2.80). In particular, the mass renormalization (2.2.56) cancels out, which is essential for the continuum limit.

The Creutz ratio (2.2.80) 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. 2.8). If the area law (2.2.75) holds for asymptotically large I and J, then

$$\chi(I,J) \stackrel{\text{large } I,J}{\longrightarrow} a^2 K, \tag{2.2.82}$$

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.

2.2.7 Asymptotic scaling

Equation (2.2.77) 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 value[10]

$$K = (400 \text{ MeV})^2 \approx 1 \text{ Gev/fm}. \tag{2.2.83}$$

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. (2.2.77), this procedure calls for $a \to \infty$ as $g^2 \to \infty$. 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. 2.6a. 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. 2.6b. Equation (2.2.77) 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. (2.2.77) is replaced at small g^2 by

$$K = \text{const} \cdot \frac{1}{a^2} \left(\frac{16\pi^2}{11g^2} \right)^{\frac{102}{121}} \mathrm{e}^{-\frac{16\pi^2}{11g^2}}. \tag{2.2.84}$$

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. 2.12. 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^2K 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.

[10]This 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 $\alpha' = 1$ 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.

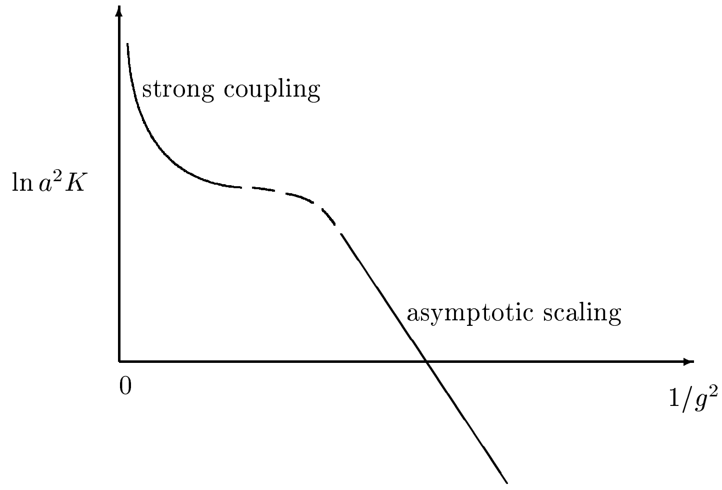


Fig. 2.12: Dependence of the string tension on $1/g^2$. The strong coupling formula (2.2.77) holds for small $1/g^2$. The asymptotic scaling formula (2.2.84) sets in for large $1/g^2$. Both formulas are not applicable in the intermediate region of $1/g^2 \sim 1$ which is depicted by the dashed line.

Remark on dimensional transmutation

The QCD action (2.1.14) 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, these 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 (1.3.73). It can be integrated to give the integration constant

$$\Lambda_{\text{QCD}} = \frac{1}{a} \exp \left[- \int \frac{dg^2}{\mathcal{B}(g^2)} \right]. \tag{2.2.85}$$

Up to now there was no difference between QCD and QED. The difference stems from the fact that the Gell-Mann–Low function $\mathcal{B}(g^2)$ 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 ($= 1/137$) is measurable in experiments while in QCD the constant $\Lambda_{\rm QCD}$ 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 $\Lambda_{\rm QCD}$. 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. (2.2.85):

$$\xi \sim \Lambda^{-1}_{\rm QCD} = a \exp \left[\int \frac{dg^2}{\mathcal{B}(g^2)} \right] . \tag{2.2.86}$$

The only chance for the RHS to diverge is to have a zero of the Gell-Mann–Low function $\mathcal{B}(g^2)$ 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, $\mathcal{B}(0) = 0$ 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 $g^2 \to 0$ as is said above.

2.3 Lattice methods

Analytic calculations of observables in the non-Abelian lattice gauge theories are available only in the strong coupling regime $g^2 \to \infty$, while one needs $g^2 \to 0$ 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 [Wil74, BDI74]. 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 [Mig75, Kad76] 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 [Wil77] while the practical implementation was done by Creutz, Jacobs and Rebbi [CJR79] for Abelian gauge groups and by Creutz [Cre79, Cre80] 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.

2.3.1 Phase transitions

As is pointed out in the previous Section, analytic calculations of the string tension are available only in the strong coupling regime $g^2 \to \infty$, while one needs $g^2 \to 0$ 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. 2.12?

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

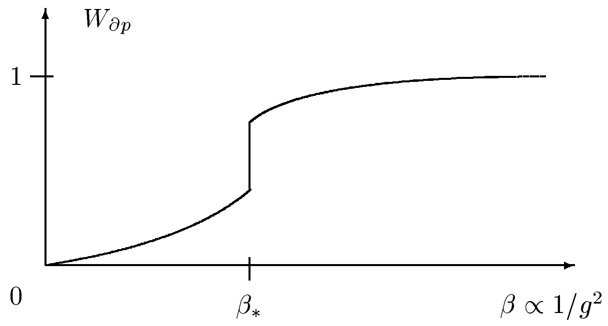


Fig. 2.13: Typical β-dependence of the plaquette average for a first-order phase transition which occurs at $\beta = \beta_*$.

the plaquette average (2.2.65) as is depicted in Fig. 2.13. The form of $W_{\partial p}(\beta)$ at small β is given to the leading order of the strong coupling expansion by Eq. (2.2.72), while that at large β is prescribed by the lattice perturbation theory[11] to be

$$W_{\partial p}(\beta) \;=\; 1 - \frac{d_G}{\beta\, d} + \mathcal{O}(\beta^{-2})\,, \tag{2.3.1}$$

where d_G is the dimensionality of the gauge group ($d_G = N_c^2 - 1$ for $\mathrm{SU}(N_c)$, $d_G = N_c^2$ for $\mathrm{U}(N_c)$) 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. (2.2.65) as

$$W_{\partial p}(\beta) \;=\; 1 + \frac{1}{N_p}\frac{\partial}{\partial \beta} \ln Z(\beta)\,, \tag{2.3.2}$$

where the partition function is given by Eq. (2.2.31) and the number of plaquettes N_p is an analog of the volume of a statistical system.

Problem 2.18 Derive Eq. (2.3.1) for the $\mathrm{SU}(N_c)$ gauge group.
Solution The partition function (2.2.31) can be calculated at large β by the saddle-point method. The saddle-point configurations are given by solutions of the classical

[11] It is often called, for obvious reasons, the weak coupling expansion.

equation (2.2.24). The proper solution reads

$$U^{\mathrm{SP}}_{x,\mu} \;=\; Z_\mu\,, \tag{2.3.3}$$

where Z_μ is an element of the $\mathrm{Z}(N_c)$ group, the center of the $\mathrm{SU}(N_c)$,

$$Z_\mu \;=\; \mathrm{I}\cdot \mathrm{e}^{\,2\pi i n_\mu/N_c}\,, \qquad n_\mu = 1, \ldots, N_c\,. \tag{2.3.4}$$

It is evident that this is a solution because elements of the center commute so that Z_μ and $Z_{-\mu}$ cancel each other in $U_{x;\mu,\nu}$.

In order to take into account fluctuations around the saddle-point solution (2.3.3), let us expand

$$U_{x,\mu} \;=\; U^{\mathrm{SP}}_{x,\mu}\, \mathrm{e}^{\,i\,t^a \epsilon^a_{x,\mu}}\,, \tag{2.3.5}$$

where the order of multiplication is inessential since Z_μ commute with the generators t^a. The expansion of $\mathrm{tr}\, U_{\partial p}$ to the quadratic order in ϵ^a reads

$$\frac{1}{N_c}\mathrm{tr}\, U_{x;\mu,\nu} \;=\; 1 - \frac{1}{4N_c}\mathcal{E}^2_{x;\mu,\nu}\,, \tag{2.3.6}$$

where

$$\mathcal{E}^a_{x;\mu,\nu} \;=\; \epsilon^a_{x,\mu} + \epsilon^a_{x+a\hat{\mu},\nu} - \epsilon^a_{x+a\hat{\nu},\mu} - \epsilon^a_{x,\nu}\,. \tag{2.3.7}$$

Due to the local gauge invariance, we can always choose, say, $\epsilon_{x,d} = 0$ so that there are only $N_l - N_s$ independent ϵ's.

Substituting into Eq. (2.2.31) and expanding the Haar measure, we get

$$Z(\beta) \;\propto\; \prod_{\nu=1}^{d} \sum_{n_\nu=1}^{N_c} \prod_{x,\mu<d,a} \int_{-\infty}^{+\infty} d\epsilon^a_{x,\mu}\, \mathrm{e}^{\,-\beta\, \mathcal{E}^2_{x;\mu,\nu}/4N_c}\,. \tag{2.3.8}$$

The sum over n_ν, which is due to the degenerate saddle points, is just an irrelevant constant.

We see from Eq. (2.3.8) that only

$$\epsilon^a_{x,\mu} \;\sim\; \frac{1}{\sqrt{\beta}} \tag{2.3.9}$$

are essential which justifies the expansion in ϵ. Rescaling the integration variables in Eq. (2.3.8), we get, therefore,

$$Z(\beta) \;\propto\; \beta^{-(N_l - N_s)\, d_G/2}\,. \tag{2.3.10}$$

Substituting into Eq. (2.3.2) and remembering that $(N_l - N_s)/N_p = 2/d$ (see Eq. (2.2.5)), we obtain Eq. (2.3.1).

Problem 2.19 Repeat the derivation of the previous Problem for the adjoint action (2.2.29).
Solution The only difference from the Wilson action (2.2.16) is that the saddle-point solution (2.3.3) is now modified as

$$U^{\rm SP}_{x,\mu} \;=\; Z_{x,\mu}\,, \tag{2.3.11}$$

i.e. may take on different values at different links. It is evident that this is a minimum of the action (2.2.29).

The only modification of Eq. (2.3.8) is

$$\prod_{\nu}^{d}\sum_{n_\nu=1}^{N_c} \;\Longrightarrow\; \prod_{x,\nu}^{d}\sum_{n_{x,\nu}=1}^{N_c}\,, \tag{2.3.12}$$

which only changes an irrelevant overall constant. Therefore, Eq. (2.3.1) remains unchanged providing the plaquette average is also taken in the adjoint representation. This supports the expectation that the continuum limits for the both actions coincide.

The first-order phase transitions of the type in Fig. 2.13 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 [Mak84].

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. 2.14. Given Eq. (2.3.2), 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. (2.2.72) and (2.3.1), respectively.

Differentiating Eq. (2.2.65) with respect to β, the derivative $\partial W_{\partial p}/\partial\beta$ can be expressed via the sum of the connected correlators:

$$\frac{\partial W_{\partial p}}{\partial \beta} \;=\; \frac{1}{2}\sum_{\text{oriented }p'}\left\langle \frac{1}{N_c}{\rm tr}\,U_{\partial p}\,\frac{1}{N_c}{\rm tr}\,U_{\partial p'}\right\rangle_{\rm conn}. \tag{2.3.13}$$

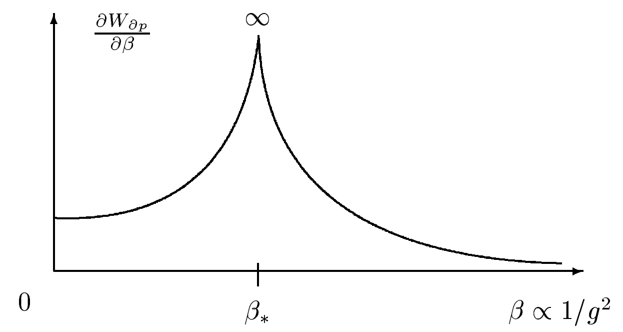


Fig. 2.14: Typical β-dependence of $\partial W_{\partial p}/\partial\beta$ for a second-order phase transition which occurs at $\beta = \beta_*$.

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

$$\begin{aligned}
&\frac{1}{2}\sum_{\text{oriented }p'}\left\langle \frac{1}{N_c}{\rm tr}\,U_{\partial p}\,\frac{1}{N_c}{\rm tr}\,U_{\partial p'}\right\rangle_{\rm conn}\\
&=\; \frac{1}{4N_p}\left\langle\left(\sum_{\text{oriented }p}\frac{1}{N_c}{\rm tr}\,U_{\partial p}\right)^2\right\rangle - \frac{1}{4N_p}\left(\left\langle\sum_{\text{oriented }p}\frac{1}{N_c}{\rm tr}\,U_{\partial p}\right\rangle\right)^2\\
&\geq\; 0\,,
\end{aligned} \tag{2.3.14}$$

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. (2.3.13) 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. 2.11b for $e^2 < e_*^2$ and like that of Fig. 2.11a in the confinement region $e^2 > e_*^2$. [12] 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. 2.2.7.

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 [LN80] 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 [Lat94], 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. 2.12. The second-order phase transition occurs in four dimensions at $\beta = \infty$ (or $g^2 = 0$) according to the general arguments of Subsect. 2.2.7, 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 [WK74]. 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 $d = 4$. The deconfining phase is associated with $g < g_*$ while the confining phase is associated with $g > g_*$.

2.3.2 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 [Wil74] and first implemented for Abelian theories by Balian, Drouffe and Itzykson [BDI74]. A mean field usually works well when there are many neighbor degrees of freedom, interacting with a given one.

[12] The latter statement is not quite correct for the reasons which are discussed in Subsect. 2.5.4.

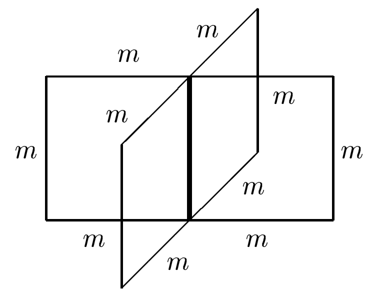


Fig. 2.15: Graphic representation of the self-consistency condition (2.3.17). The link variables are replaced by $m \cdot \mathrm{I}$ at all links except for a given one denoted by the bold line.

In the simplest version of the mean-field method, the link variable $U_{x,\mu}$ is replaced by the mean-field value $m \cdot \mathrm{I}$ everywhere but at a given link (see Fig. 2.15) at which the self-consistency condition

$$\left\langle [U_{x,\mu}]^{ij} \right\rangle_0 \;=\; m\,\delta^{ij} \tag{2.3.15}$$

is imposed.

The average on the LHS of Eq. (2.3.15) is calculated with the action which is obtained from (2.2.16) by the substitution of $m \cdot \mathrm{I}$ 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. (2.3.15) is to be calculated with the action

$$S_0\,[U] \;=\; 2\,(d-1)\,m^3\,\mathrm{Re\,tr}\,U_{x,\mu} + \mathrm{const}\,. \tag{2.3.16}$$

Therefore, the self-consistency condition (2.3.15) can be written by the substitution of the mean-field ansatz into the lattice partition function (2.2.31) as

$$\frac{\int dU\, e^{\bar\beta N_c \,\mathrm{Re\,tr\,U}}\, \frac{1}{N_c}\mathrm{tr}\, U}{\int dU\, \mathrm{e}^{\,\bar\beta N_c\,\mathrm{Re\,tr}\,U}} \;=\; m \tag{2.3.17}$$

with

$$\bar\beta \;=\; 2\,(d-1)\,m^3\,\frac{\beta}{N_c^2}\,. \tag{2.3.18}$$

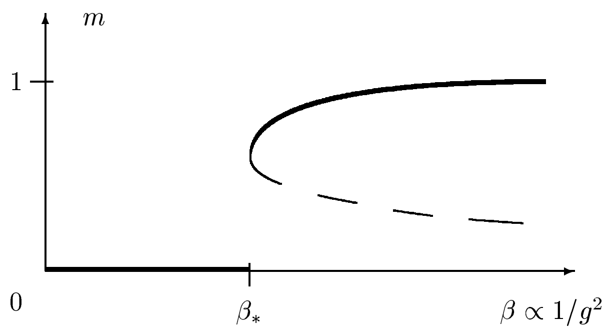


Fig. 2.16: Typical behavior of the solutions of the self-consistency equation (2.3.17). The only solution with $m = 0$ exists for $\beta < \beta_*$. Two more solutions appear for $\beta > \beta_*$. The one depicted by the dashed line is unstable. The actual value of m versus β is depicted by the bold lines. A first-order phase transition is associated with $\beta = \beta_*$.

The meaning of Eq. (2.3.17) 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 (2.3.17) 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. 2.16. 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

$$W_{\partial p} \;=\; m^4 \tag{2.3.19}$$

in the mean-field approximation which follows from the substitution of the link-variables in the definition (2.2.65) 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. [GL81, CGL81]. 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.

Problem 2.20 Calculate β_* for the SU(∞) lattice gauge theory, when the group integral on the LHS of Eq. (2.3.17) equals: $\bar\beta/2$ for $\bar\beta \leq 1$ (a strong coupling phase) and $1 - 1/2\bar\beta$ for $\bar\beta \geq 1$ (a weak coupling phase).
Solution For the strong coupling phase, the self-consistency equation

$$(d-1)\, m^3 \frac{\beta}{N_c^2} \;=\; m \tag{2.3.20}$$

has the only solution $m = 0$. The other solutions are unacceptable due to stability consideration.

The nontrivial solutions of the self-consistency equation appear in the weak coupling phase when $dm/d\beta = \infty$ or $d\beta/dm = 0$. Differentiating, we get in $d = 4$

$$\frac{\partial \beta^{-1} N_c^2}{\partial m} \;=\; 12\left(3m^2 - 4m^3\right)\,, \tag{2.3.21}$$

which yields

$$m_* \;=\; \frac{3}{4}\,, \qquad \frac{\beta_*}{N_c^2} \;=\; \frac{4^3}{3^4} \;=\; 0,79\,. \tag{2.3.22}$$

It is still left to verify that the proper $\bar\beta$ is indeed associated with the weak coupling phase. From Eq. (2.3.18), we get $\bar\beta_* = 2$ and this is the case.

2.3.3 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. (2.2.13) while the action and the measure are gauge invariant). This is in accordance with Elitzur's theorem [Eli75] 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 [BDI74] which is similar to that proposed by R. Peierls in thirties. It is based on Jensen's inequality[13]

$$\left\langle \mathrm{e}^{\,F} \right\rangle_0 \;\geq\; \mathrm{e}^{\,\langle F \rangle_0} \tag{2.3.23}$$

[13] More detail can be found, *e.g.*, in the books [Fey72, Sak85] cited in the reference list to Chapter 1.

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

Let us choose the trial partition function

$$Z_0 \;=\; \int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{e}^{\,\bar\beta N_c \sum_{x,\mu} \mathrm{Re\,tr}\, U_{x,\mu}} \tag{2.3.24}$$

as a product of one-link integrals. Adding and subtracting the trial action, we write down the following bound on the partition function (2.2.31):

$$Z \;\geq\; Z_0\, \mathrm{e}^{\left\langle \frac{\beta}{N_c} \sum_p \mathrm{Re\,tr}\, U_{\partial p} - \bar\beta N_c \sum_{x,\mu} \mathrm{Re\,tr}\, U_{x,\mu} \right\rangle_0}\,. \tag{2.3.25}$$

Here $\langle \ldots \rangle_0$ means averaging with respect to the same action as in Eq. (2.3.24).

Since the expression which is averaged in the exponent in Eq. (2.3.25) is linear in each of the link variables, it can be calculated via the one-matrix integral given by the LHS of Eq. (2.3.17). Therefore, we get

$$\begin{aligned} &\left\langle \frac{\beta}{N_c} \sum_p \mathrm{Re\,tr}\, U_{\partial p} - \bar\beta N_c \sum_{x,\mu} \mathrm{Re\,tr}\, U_{x,\mu} \right\rangle_0 \\ &\qquad = \;\; \beta N_p m^4 - \bar\beta N_c^2 N_l m\,, \end{aligned} \tag{2.3.26}$$

where Eq. (2.3.19) has been used.

The idea of the variational mean-field method is to fix $\bar\beta$ from the condition for the trial ansatz (2.3.24) to give the best approximation to Z in the given class. Calculating the derivative of the RHS of Eq. (2.3.25) with respect to $\bar\beta$ and taking into account that m depends on $\bar\beta$ according to Eq. (2.3.17), we find the maximum at $\bar\beta$ given by Eq. (2.3.18), which reproduces the simplest version of the mean-field method described above.

To restore Elitzur's theorem, a more sophisticated trial ansatz [Dro81] can be considered:

$$Z_0 \;=\; \int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{e}^{\,N_c \sum_{x,\mu} \mathrm{Re\,tr}\, B^\dagger_{x,\mu} U_{x,\mu}}\,, \tag{2.3.27}$$

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

$$B_{x,\mu} \;=\; \bar\beta\, \Omega_x\, \Omega^\dagger_{x+a\hat\mu}\,, \tag{2.3.28}$$

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

Problem 2.21 Perform the variational mean-field calculation with the ansatz (2.3.27).
Solution Let us denote

$$M^{ij}_{x,\mu} \;=\; \frac{\int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{e}^{\,N_c \sum_{x,\mu} \mathrm{Re\,tr}\, B^\dagger_{x,\mu} U_{x,\mu}}\, U^{ij}_{x,\mu}}{\int \prod_{x,\mu} dU_{x,\mu}\, \mathrm{e}^{\,N_c \sum_{x,\mu} \mathrm{Re\,tr}\, B^\dagger_{x,\mu} U_{x,\mu}}}\,. \tag{2.3.29}$$

Then the analog of Eq. (2.3.26) reads

$$\begin{aligned} &\left\langle \frac{\beta}{N_c} \sum_p \mathrm{Re\,tr}\, U_{\partial p} - N_c \sum_{x,\mu} \mathrm{Re\,tr}\, B^\dagger_{x,\mu} U_{x,\mu} \right\rangle_0 \\ &\qquad = \;\; \frac{\beta}{N_c} \sum_p \mathrm{Re\,tr}\, M_{\partial p} - N_c \sum_{x,\mu} \mathrm{Re\,tr}\, B^\dagger_{x,\mu} M_{x,\mu} \end{aligned} \tag{2.3.30}$$

so that the inequality (2.3.25) takes on the form

$$Z \;\geq\; Z_0\, \mathrm{e}^{\frac{\beta}{N_c} \sum_p \mathrm{Re\,tr}\, M_{\partial p} - N_c \sum_{x,\mu} \mathrm{Re\,tr}\, B^\dagger_{x,\mu} M_{x,\mu}}\,. \tag{2.3.31}$$

$B_{x,\mu}$ can now be determined by maximizing with respect to $B_{x,\mu}$ and taking into account Eq. (2.3.29).

It is easy to see that if $B_{x,\mu} = \bar\beta \cdot \mathrm{I}$ is a solution as before, then (2.3.28) is also a solution. Therefore, we get

$$\left\langle U^{ij}_{x,\mu} \right\rangle_0 \;=\; m \int d\Omega_{x+a\hat\mu}\, d\Omega_x\, \Omega_{x+a\hat\mu}\, \Omega^\dagger_x \;=\; 0\,, \tag{2.3.32}$$

where the integration over Ω takes into account different equivalent maxima. Thus, all gauge invariant quantities for the ansatz (2.3.27) are the same as for the ansatz (2.3.24) while gauge noninvariant quantities now vanish in agreement with Elitzur's theorem.

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 choosen as is explained in Fig. 2.16 but not when the free energy of both phases coincide (the standard Maxwell rule in statistical physics). Perhaps, the criterion of Fig. 2.16 should be choosen 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. [FLZ82], which take into account fluctuations around the mean-field solution (2.3.28), agree for the Maxwell-rule criterion with numerical data. These results are reviewed in Ref. [DZ83].

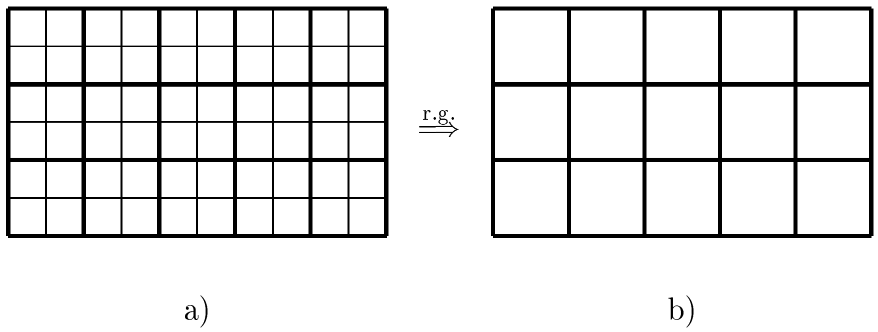


Fig. 2.17: Lattice renormalization group transformation (2.3.33). The thin lines of the old lattice a) represent links on which integration is performed. The new lattice b) has the lattice spacing $2a$ but the same spatial extent La.

2.3.4 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 [WK74]). The idea to apply a similar method to lattice gauge theories is due to Migdal [Mig75].

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. 2.17a. 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 $2a$,

$$a \stackrel{\text{r.g.}}{\Longrightarrow} 2a\,, \tag{2.3.33}$$

which is depicted in Fig. 2.17b. The space size of the lattice is L before the transformation and becomes $L/2$ after the transformation,

$$L \stackrel{\text{r.g.}}{\Longrightarrow} \frac{L}{2}\,, \tag{2.3.34}$$

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. 2.17a becomes a more general one

under the renormalization group transformation:

$$\begin{aligned} S[U] &= \sum_p \beta \frac{1}{N_c}\,\mathrm{tr}\,U_p \\ &\stackrel{\text{r.g.}}{\Longrightarrow}\; S'\,[U] = \sum_p \beta_1' \frac{1}{N_c}\,\mathrm{tr}\,U_p \\ &\qquad + \sum_{p_2} \beta_2' \frac{1}{N_c}\,\mathrm{tr}\,U_{p_2} + \sum_{p_3} \beta_3' \frac{1}{N_c}\,\mathrm{tr}\,U_{p_3} + \ldots\,. \end{aligned} \tag{2.3.35}$$

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

$$\beta \stackrel{\text{r.g.}}{\Longrightarrow} \beta' = \beta - \Delta\beta \tag{2.3.36}$$

under the renormalization group transformation on the lattice.

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

$$\int\limits_{\beta-\Delta\beta}^{\beta} \frac{dx}{x^{3/2}} \mathcal{B}^{-1}\left(\frac{6}{x}\right) \;=\; -\frac{2\ln 2}{\sqrt{6}}\,. \tag{2.3.37}$$

Here the $\ln 2$ on the RHS is due to Eq. (2.3.33) and the relation (2.2.32) between β and g^2 is used with $N_c = 3$.

For the pure SU(3) gauge theory, we get from Eq. (2.3.37)

$$\Delta\beta = 0,579 + \frac{0,204}{\beta} + O\left(\frac{1}{\beta^2}\right) \tag{2.3.38}$$

at asymptotically large β.

One can integrate over the thin links in Fig. 2.17a 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

$$e^{-S[U]} = \sum_r f_r d_r \chi_r(U) \tag{2.3.39}$$

where

$$d_r = \chi_r(\mathrm{I}) \tag{2.3.40}$$

is the dimension of a given representation r and f_r are the coefficients which depend on the form of $S[U]$.

Migdal [Mig75] proposed to approximate the new action, which appear after

$$a \Longrightarrow \rho\, a\,, \tag{2.3.41}$$

by the formula

$$e^{-S'[U']} = \left[\sum_r f_r{}^{\rho^2} d_r \chi_r(U')\right]^{\rho^{d-2}}, \tag{2.3.42}$$

which is exact in $d = 2$ dimensions. Kadanoff [Kad76] slightly modified the recursion relation (2.3.42).

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 [Ake93] for the $\Delta\beta$, which is defined by Eq. (2.3.36), versus β are depicted in Fig. 2.18. The solid line represents the asymptotics (2.3.38). 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.

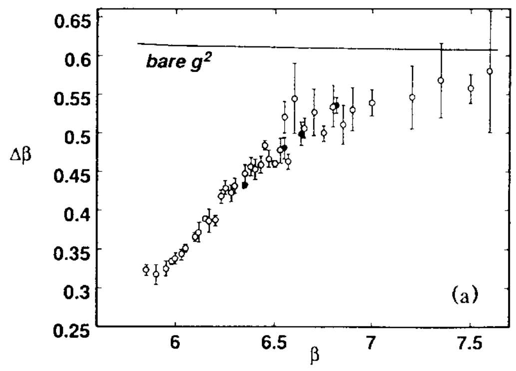


Fig. 2.18: Monte Carlo data by Akemi *et al.* [Ake93] for $\Delta\beta$. The errorbars represent statistical errors. The solid line represents the asymptotics (2.3.38).

2.3.5 Monte Carlo method

The idea of the Monte Carlo method is to calculate the partition function (2.2.31) and the averages (2.2.39) 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:

$$C = \{U^{ij}_{x,\mu}, \ldots, \ldots, \ldots, \ldots, \ldots, \ldots, \ldots, \ldots, \ldots, \ldots\}\,. \tag{2.3.43}$$

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

$$\int \prod_{x,\mu} dU_{x,\mu} \cdots = \sum_C \cdots\,. \tag{2.3.44}$$

The averages (2.2.39) can be rewritten as

$$\langle F(C) \rangle = \frac{\sum\limits_C e^{-\beta S(C)} F(C)}{\sum\limits_C e^{-\beta S(C)}}, \tag{2.3.45}$$

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

$$E = \{C_1, \ldots, C_N\} \tag{2.3.46}$$

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

$$P_{\text{Bol}}(C_n) = Z^{-1}(\beta)\, e^{-\beta S(C_n)}. \tag{2.3.47}$$

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

Given an equilibrium ensemble, the averages (2.3.45) take the form of the arithmetic mean

$$\langle F[U] \rangle = \frac{1}{N} \sum_{n=1}^{N} F(C_n) \tag{2.3.48}$$

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

$$W_{R\times\mathcal{T}} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{N_c} \operatorname{tr} U_{R\times\mathcal{T}}(C_n)\,. \tag{2.3.49}$$

If all configurations in the equilibrium ensemble are independent, then the RHS of Eq. (2.3.49) will approximate the exact value of $W_{R\times\mathcal{T}}$ 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 (2.3.46) is obtained from the previous one by a definite algorithm but stochastically. In other words, the random process is completely determined by the probability $P(C_{n-1} \to C_n)$ 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.*

$$P(C_{n-1} \to C_n) = P(C_{n-1}, C_n)\,. \tag{2.3.50}$$

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 (2.3.47). This is ensured if $P(C, C')$ satisfies the detailed balance condition

$$e^{-\beta S(C)} P(C, C') = e^{-\beta S(C')} P(C', C)\,. \tag{2.3.51}$$

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.

Problem 2.22 Prove the statements 1) and 2) listed in the previous paragraph using the detailed balance condition (2.3.51).

Solution Let a state C is encountered in ensembles E and E' with the probability densities $P(C)$ and $P'(C)$, respectively. Then a distance between the two ensembles can be defined as

$$\|E - E'\| = \sum_C |P(C) - P'(C)|\,. \tag{2.3.52}$$

For a Markov process when Eq. (2.3.50) holds, we have

$$P'(C) = \sum_{C'} P(C, C')\, P(C') \tag{2.3.53}$$

if E' is obtained from E by a Monte Carlo algorithm. The transition probability $P(C, C')$ is nonnegative and obeys

$$\sum_C P(C, C') = \sum_{C'} P(C, C') = 1 \tag{2.3.54}$$

since each new state is obtained from an old one and vice versa.

It is now easy to prove the statement 1). Summing the detailed balance condition (2.3.51) over C', we get

$$P_{\mathrm{Bol}}(C) \;=\; \sum_{C'} P(C,C') P_{\mathrm{Bol}}(C')\,, \tag{2.3.55}$$

i.e. the Boltzmann distribution is an eigenvector of $P(C,C')$. Comparing with Eq. (2.3.53), we see that the new distribution is again the Boltzmann one, which proves 1).

To prove the statement 2), let us compare the distances from E and E' to some equilibrium ensemble E_{Bol} associated with the Boltzmann distribution (2.3.47). We have the inequality

$$\begin{aligned} \|E' - E_{\mathrm{Bol}}\| &= \sum_{C} \left|P'(C) - P_{\mathrm{Bol}}(C)\right| \\ &= \sum_{C} \left|\sum_{C'} P(C,C')\left[P(C') - P_{\mathrm{Bol}}(C')\right]\right| \\ &\leq \sum_{CC'} P(C,C')\left|P(C') - P_{\mathrm{Bol}}(C')\right| \\ &= \sum_{C'} \left|P(C') - P_{\mathrm{Bol}}(C')\right| \\ &= \|E - E_{\mathrm{Bol}}\|\,, \end{aligned} \tag{2.3.56}$$

where Eqs. (2.3.53), (2.3.55) and (2.3.54) are used. Thus, 2) is proven.

Specific Monte Carlo algorithms differ in the choice of the transition probability $P(C,C')$ while the detailed balance condition (2.3.51) 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

$$P\left(U'_{x,\mu}\right) \;\propto\; \mathrm{e}^{-\beta S(C')}\,. \tag{2.3.57}$$

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 (suppose randomly on the group manifold).

b) The difference between the action for this trial configuration and that for the old one is calculated:

$$\Delta S = S\left(C'\right) - S\left(C\right)\,. \tag{2.3.58}$$

c) A random number $r \in [0,1]$ is generated.

d) If

$$\mathrm{e}^{-\beta\Delta S} \;>\; r\,, \tag{2.3.59}$$

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.

A new configuration C', which is obtained by applying either Monte Carlo algorithm to each links 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 (2.3.46) 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 [CJR83] and the book by Creutz [Cre83].

2.3.6 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 [Cre79] who evaluated the string tension for the SU(2) gauge group. His result is reproduced in Fig. 2.19 and looks very much like what is expected in Fig. 2.12. 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 [LR82]. They calculated equipotential

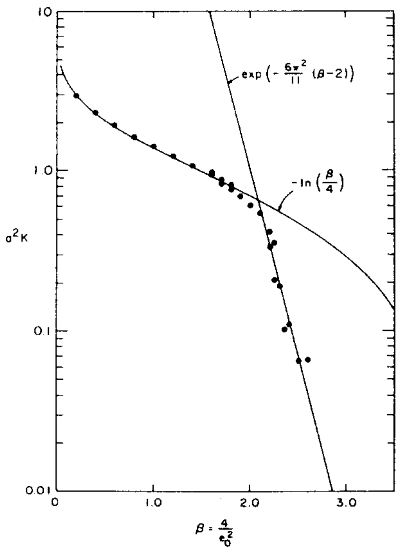


Fig. 2.19: Monte Carlo data by Creutz [Cre79] for the string tension in the SU(2) pure lattice gauge theory.

surfaces for the interaction between static quarks. In the strong coupling region $g^2 \to \infty$, they appear as in Fig. 2.20a since the interaction potential is given by

$$E(x,y,z) \;=\; K\,(\,|x| + |y| + |z|\,) \tag{2.3.60}$$

because the distance between the quarks is measured along the lattice. This is associated 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 [LR82] are shown in Figs. 2.20b, c. They demonstrate the restoration of rotational symmetry when passing from $\beta = 2$ (Fig. 2.20b) to $\beta = 2.25$ (Fig. 2.20c).

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.

As an example of modern Monte Carlo calculations, the Monte Carlo data by Bali and Schilling [BS92] for the interaction potential in the SU(3)

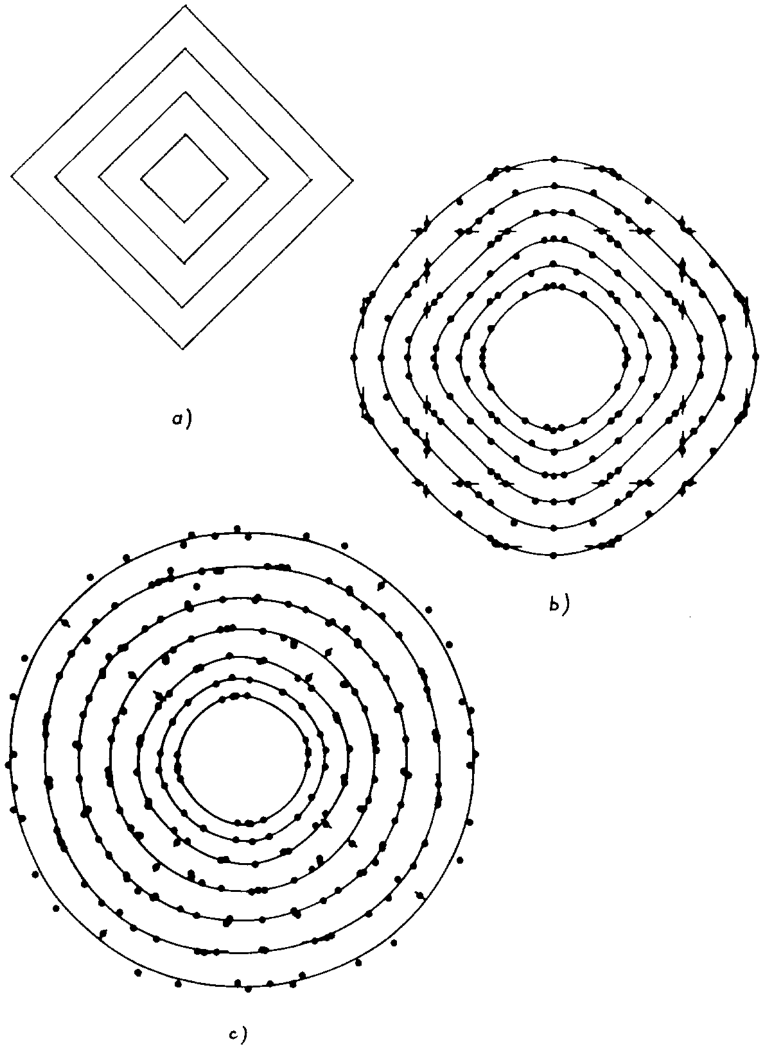


Fig. 2.20: Behavior of equipotential lines at different values of β: a) the strong coupling limit $\beta = 0$, b) $\beta = 2$, c) $\beta = 2.25$. Figs. b) and c), taken from the paper by Lang and Rebbi [LR82], show how the rotational symmetry is restored as β is increased.

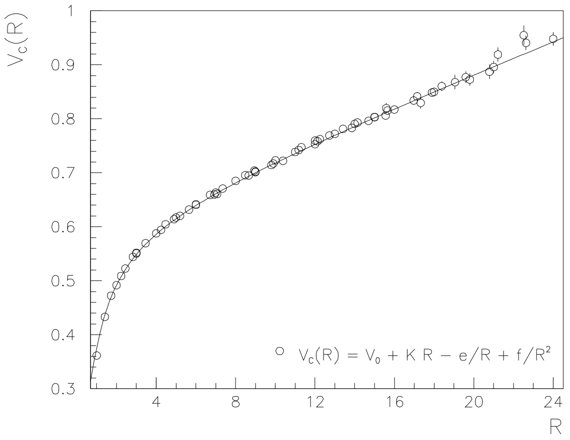


Fig. 2.21: Monte Carlo data [BS92] for the continuum interaction potential $E(R)$. The solid curve is a sum of linear and Coulomb potentials.

pure lattice gauge theory are shown in Fig. 2.21. The solid curve represents how the data are described by a sum of linear and Coulomb potentials. The maximal lattice size was 32^4 and the spatial extent of the lattice was up to 3.3 fm.

The Proceeding of the Lattice Conference [Lat94] are very useful, as is already mentioned, for a today's look at the subject.

2.4 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 [Wil75], 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.

2.4.1 Chiral fermions

The quark fields are generically matter fields, whose gauge transformation in the continuum is given by Eqs. (2.1.1) and (2.1.3), and can be put on a lattice according to Eq. (2.2.7). Then the lattice gauge transformation is

$$\begin{aligned} \psi_x &\xrightarrow{\text{g.t.}} \psi'_x = \Omega_x \psi_x \,, \\ \bar\psi_x &\xrightarrow{\text{g.t.}} \bar\psi'_x = \bar\psi_x \Omega^\dagger_x \,. \end{aligned} \tag{2.4.1}$$

The lattice analog of the QCD action (2.1.14) reads[14]

$$\begin{aligned} S\left[U,\bar\psi,\psi\right] &= \beta\, S_{\rm Lat}\left[U\right] + M \sum_x \bar\psi_x \psi_x \\ &+ \frac{1}{2} \sum_{x,\mu>0} \left[\bar\psi_x \gamma_\mu U^\dagger_{x,\mu} \psi_{x+a\hat\mu} - \bar\psi_{x+a\hat\mu} \gamma_\mu\, U_{x,\mu}\, \psi_x\right] . \end{aligned} \tag{2.4.2}$$

[14]A 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 (see the footnote on p. 91). It does not matter how to define $U_{x,\mu}$ since the Haar measure is invariant under Hermitean conjugation.

The first term on the RHS is the pure gauge lattice action (2.2.16). 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 Hermitean and invariant under the lattice gauge transformation (2.2.13), (2.4.1) at a finite lattice spacing.

The partition function of lattice QCD with fermions is defined by

$$Z(\beta, M) = \int \prod_{x,\mu} dU_{x,\mu} \prod_{x} d\bar{\psi}_x d\psi_x \, \mathrm{e}^{-S[U,\bar{\psi},\psi]}, \tag{2.4.3}$$

where the action is given by Eq. (2.4.2). The integration over $U_{x,\mu}$ is as in Eq. (2.2.31), and the integral over the quark field is the Grassmann one. The averages are defined by

$$\begin{aligned} \langle F[U,\bar{\psi},\psi] \rangle &= \\ & Z^{-1}(\beta, M) \int \prod_{x,\mu} dU_{x,\mu} \prod_{x} d\bar{\psi}_x d\psi_x \, \mathrm{e}^{-S[U,\bar{\psi},\psi]} F[U,\bar{\psi},\psi] \end{aligned} , \tag{2.4.4}$$

which extends Eq. (2.2.39) to the case of fermions. Since both the action and the measure in Eq. (2.4.4) 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 (2.4.2) reproduces (2.1.14) in the naive continuum limit $a \to 0$, let us assume that the lattice quark field ψ_x slowly varies from site to site and substitute

$$\begin{aligned} \psi_x &\to a^{3/2}\psi(x), \\ \psi_{x+a\hat{\mu}} &\to a^{3/2}(\psi(x) + a\partial_\mu\psi(x)) \end{aligned} \tag{2.4.5}$$

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 (2.4.5) together with Eq. (2.2.10) yields

$$\bar{\psi}_x \gamma_\mu U^\dagger_{x,\mu} \psi_{x+a\hat{\mu}} \to a^3 \bar{\psi}\gamma_\mu\psi + a^4 \bar{\psi} \nabla^{\mathrm{fun}}_\mu \gamma_\mu \psi + \mathcal{O}(a^5), \tag{2.4.6}$$

where there is no summation over μ in the second term as before in this Chapter. The first term cancels when substituted into Eq. (2.4.2) 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 (2.4.2) was proposed in Ref. [Wil74]. For $M = 0$ it is invariant under the global chiral transformation

$$\psi_x \xrightarrow{\text{c.t.}} \mathrm{e}^{i\alpha\gamma_5}\psi_x, \qquad \bar{\psi}_x \xrightarrow{\text{c.t.}} \bar{\psi}_x \mathrm{e}^{i\alpha\gamma_5}. \tag{2.4.7}$$

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 of Sect. 1.3.

Problem 2.23 Show that the lattice action (2.4.2) is invariant under

$$\psi_x \to i\gamma_4\gamma_5(-1)^{t/a}\psi_x. \tag{2.4.8}$$

Find 15 more similar transformations.
Solution Let us defined the generators T_α by

$$\psi_x \to T_\alpha \psi_x, \qquad \bar{\psi}_x \to \bar{\psi}_x T^\dagger_\alpha. \tag{2.4.9}$$

The transformation (2.4.8) can be performed for each of the $d = 4$ axes which gives

$$T_\alpha = i\gamma_\mu\gamma_5(-1)^{x_\mu/a}. \tag{2.4.10}$$

The other generators are given by products of (2.4.10). Their explicit form is [KS81a]

$$\begin{aligned} T_\alpha = {}& \mathrm{I},\ i\gamma_\mu\gamma_5(-1)^{x_\mu/a},\ i\gamma_\mu\gamma_\nu(-1)^{(x_\mu+x_\nu)/a}\ (\mu>\nu), \\ & \gamma_4(-1)^{(x+y+z)/a},\ \ldots,\ \gamma_1(-1)^{(y+z+t)/a}, \gamma_5(-1)^{(x+y+z+t)/a}. \end{aligned} \tag{2.4.11}$$

There are all together $1+4+6+4+1=16$ independent transformations which form a discrete subgroup of the U(4) group.

2.4.2 Fermion doubling

As is pointed out at the end of the previous Subsection, the lattice fermionic action (2.4.2) 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 (2.4.2) 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:

$$\psi_k = a^{5/2}\sum_x \psi_x \mathrm{e}^{-ikx}. \tag{2.4.12}$$

The free fermionic action then reads

$$S_0\left[\bar\psi,\psi\right] \;=\; \int\limits_{-\pi/a}^{\pi/a} \frac{d^4k}{(2\pi)^4}\bar\psi_k G^{-1}(k)\,\psi_k \tag{2.4.13}$$

with

$$G^{-1}(k) \;=\; \frac{1}{a}\sum_{\mu=1}^{4} i\gamma_\mu \sin k_\mu a \tag{2.4.14}$$

for $M=0$.

In the naive continuum limit, the sin in Eq. (2.4.14) can be expanded in the power series in a, which results in the free (inverse) continuum propagator

$$G^{-1}(k) \;\to\; i\sum_{\mu=1}^{4}\gamma_\mu k_\mu \;=\; i\,\hat k\,. \tag{2.4.15}$$

The Lorentz invariance has been restored after summing over μ.

When passing from the lattice expression (2.4.14) to the continuum one (2.4.15), 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

$$\sinh^2 Ea \;=\; \sum_{\mu=1}^{3}\sin^2 p_\mu a\,. \tag{2.4.16}$$

Let us look for solutions of Eq. (2.4.16) 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

$$p^{(1)} \;=\; (E,0,0,p_z) \tag{2.4.17}$$

are related by

$$\sinh Ea \;=\; \sin p_z a\,, \tag{2.4.18}$$

which follows from the substitution of (2.4.17) into the dispersion law.

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

$$p^{(2)} \;=\; \left(E,0,0,\frac{\pi}{a}-p_z\right) \tag{2.4.19}$$

is also a solution of Eq. (2.4.18) if (2.4.17) is. Quite analogously, the four-vectors

$$\begin{aligned} p^{(3)} &= \left(E,\frac{\pi}{a},0,p_z\right)\,, \\ \ldots &, \\ p^{(8)} &= \left(E,\frac{\pi}{a},\frac{\pi}{a},\frac{\pi}{a}-p_z\right)\,, \end{aligned} \tag{2.4.20}$$

which are obtained from $p^{(1)}$ and $p^{(2)}$ by changing zeros for π/a, also satisfy Eq. (2.4.18). Therefore, a quark state with the energy E is eightfold degenerate.

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

$$\Psi^{(j)}(t,x,y,z) \;\propto\; \exp\left(iEt - ip_x^{(j)}x - ip_y^{(j)}y - ip_z^{(j)}z\right). \tag{2.4.21}$$

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 $a\to 0$ 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 (2.4.2) is as good as for the slowly varying functions when Eq. (2.4.5) holds. In order to see that, let us rewrite the action (2.4.2) as

$$\begin{aligned} S\left[U,\bar\psi,\psi\right] &= \beta\, S_{\rm Lat}\left[U\right] + M\sum_x \bar\psi_x\psi_x \\ &+\frac{1}{2}\sum_{x,\mu>0}\left[\bar\psi_x\gamma_\mu\left(U^\dagger_{x,\mu}\psi_{x+a\hat\mu} - U^\dagger_{x,-\mu}\psi_{x-a\hat\mu}\right)\right]. \end{aligned} \tag{2.4.22}$$

Even if ψ_x has opposite signs at neighboring lattice sites along the μ-axis, as is illustrated by Fig. 2.22a, *i.e.*

$$\psi_{x+a\hat\mu} \;\to\; -\,\psi_x\,, \tag{2.4.23}$$

then the difference $\psi_{x+a\hat\mu}-\psi_{x-a\hat\mu}$ on the RHS of Eq. (2.4.22) is still of the correct order in a:

$$\psi_x \;\to\; a^{3/2}\,\psi(x)\,, \qquad \psi_{x+a\hat\mu}-\psi_{x-a\hat\mu} \;\to\; -\,2a^{5/2}\partial_\mu\psi(x)\,, \tag{2.4.24}$$

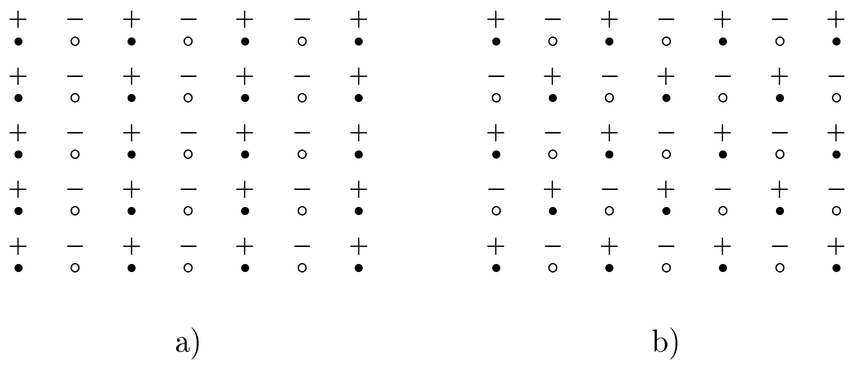


Fig. 2.22: Altering signs of ψ_x on a lattice along a) one axis and b) two axes.

so that the continuum fermionic action is reproduced except for the sign of the γ_μ-matrix which is opposite to that in Eq. (2.1.14).

This extra minus sign can be absorbed in the redefinition of the continuum fermionic field $\psi(x) \to i\gamma_\mu\gamma_5\psi(x)$, 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 even or odd number of axes (see Fig. 2.22). 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.

Problem 2.24 Calculate the vector and axial charges of the doublers deriving the vector and axial currents on a lattice.
Solution The vector and axial currents on a lattice can be derived using a lattice analog of the Noether theorem. The invariance of the lattice fermionic action under

$$\psi_x \to \mathrm{e}^{\,i\alpha_x}\,\psi_x\,, \qquad \bar\psi_x \to \bar\psi_x\,\mathrm{e}^{\,-i\alpha_x} \tag{2.4.25}$$

results in the lattice vector current

$$V_{x,\mu} \;=\; \frac{1}{2}\left[\bar\psi_x\gamma_\mu U^\dagger_{x,\mu}\psi_{x+a\hat\mu} + \bar\psi_{x+a\hat\mu}\gamma_\mu\,U_{x,\mu}\,\psi_x\right], \tag{2.4.26}$$

which is conserved in the sense that

$$\sum_{\mu>0}\left(V_{x,\mu} - V_{x-a\hat\mu,\mu}\right) \;=\; 0\,. \tag{2.4.27}$$

This can be proven using the lattice (quantum) Dirac equation

$$\frac{1}{2}\sum_{\mu>0}\gamma_\mu\left(U^\dagger_{x,\mu}\psi_{x+a\hat\mu} - U^\dagger_{x,-\mu}\psi_{x-a\hat\mu}\right) + M\psi_x \;\overset{\text{w.s.}}{=}\; \frac{\delta}{\delta\bar\psi_x}\,, \tag{2.4.28}$$

which is the lattice analog of Eq. (1.3.20).

Analogously, the lattice chiral transformation

$$\psi_x \to \mathrm{e}^{\,i\alpha_x\gamma_5}\,\psi_x\,, \qquad \bar\psi_x \to \bar\psi_x\,\mathrm{e}^{\,i\alpha_x\gamma_5} \tag{2.4.29}$$

results in the lattice axial current

$$A_{x,\mu} \;=\; \frac{i}{2}\left[\bar\psi_x\gamma_\mu\gamma_5 U^\dagger_{x,\mu}\psi_{x+a\hat\mu} + \bar\psi_{x+a\hat\mu}\gamma_\mu\gamma_5 U_{x,\mu}\psi_x\right], \tag{2.4.30}$$

which reproduces (1.3.11) as $a \to 0$. The current (2.4.30) is conserved for $M = 0$.

It is now easy to verify that 16 generators (2.4.11) commute with the lattice U(1) transformation (2.4.25) so that the lattice vector current (2.4.26) is left invariant. Analogously, the lattice axial U(1) transformation (2.4.29) commutes only with $1+6+1 = 8$ of 16 generators (2.4.11) which are built out of the products of an even number of the generators (2.4.10) and does not commute with the $4+4 = 8$ other generators. Therefore, the axial current (2.4.30) is invariant under the $1+6+1 = 8$ transformations, which are the products of an even number of the generators (2.4.10), and alters its sign under the other $4+4 = 8$ transformations, which are the products of an odd one. Thus, the vector charge of all the doublers is the same while the axial charge is positive for 8 and negative for 8 other doublers.

It is worth noting that the mass term in Eq. (2.4.2) 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 (2.4.13), 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. 2.23. The behavior around $k_4 = 0$ is prescribed by Eq. (2.4.15) and is just a straight line with a positive slop. 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. 2.24. There is no doubling of states in the bosonic case.

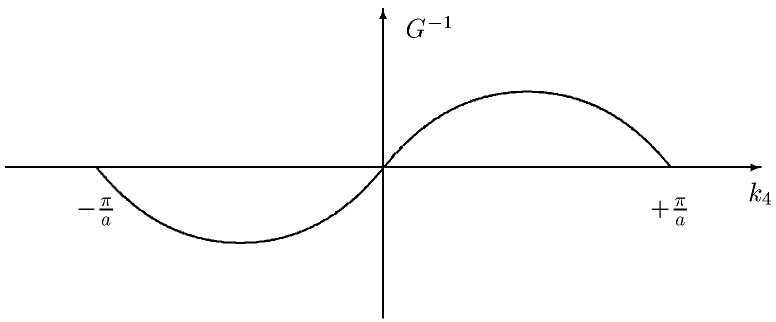


Fig. 2.23: Momentum dependence of G^{-1} for the chiral lattice fermions. The periodicity leads to an extra zero at $k_4 = \pi/a$.

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 and Ninomiya [NN81]. 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. (2.4.14) by $k_\mu a$ itself to get an expression which is similar to the continuum propagator (2.4.15). 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 [Lat94].

2.4.3 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. [KS75, Sus77] and elaborated for the Euclidean formulation in Refs. [STW81, KS81b].

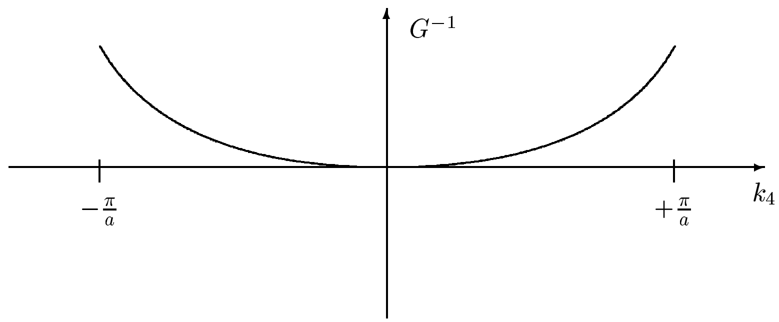


Fig. 2.24: Momentum dependence of G^{-1} for the lattice bosons. No doubling of states is associated with this behavior.

Let us substitute

$$\psi_x = \gamma_1^{x/a}\gamma_2^{y/a}\gamma_3^{z/a}\gamma_4^{t/a}\phi_x \tag{2.4.31}$$

into the free fermionic action (2.4.13). Then it takes the form

$$S_0\left[\bar\psi,\psi\right] = \frac{1}{2}\sum_x\sum_i\sum_{\mu>0}\eta_{x,\mu}\left(\left[\phi_x^\dagger\right]^i\left[\phi_{x+a\hat\mu}\right]^i - \left[\phi_{x+a\hat\mu}^\dagger\right]^i\left[\phi_x\right]^i\right), \tag{2.4.32}$$

which is diagonal with respect to the spinor indices since

$$\eta_{x,\mu} = (-1)^{(x_1+\ldots+x_{\mu-1})/a}, \tag{2.4.33}$$

or explicitly

$$\begin{aligned}\eta_{x,\mu} &= 1 \qquad \text{for } \mu=1\,,\\ \eta_{x,\mu} &= (-1)^{x_1/a} \qquad \text{for } \mu=2\,,\\ &\cdots\,,\\ \eta_{x,\mu} &= (-1)^{(x_1+\ldots+x_{d-1})/a} \qquad \text{for } \mu=d\,,\end{aligned} \tag{2.4.34}$$

does not depend on spinor indices.

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

$$\phi_x^i = \begin{pmatrix}\chi_x\\0\\0\\0\end{pmatrix}, \qquad \bar\phi_x^i = \begin{pmatrix}\bar\chi_x\\0\\0\\0\end{pmatrix}. \tag{2.4.35}$$

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. [KS75, Sus77].

The action of the Kogut–Susskind fermions is

$$\begin{aligned} S\left[U,\bar{\psi},\psi\right] &= \beta\, S_{\rm Lat}\left[U\right] + M\sum_x \bar{\chi}_x\chi_x \\ &+\frac{1}{2}\sum_{x,\mu>0}\eta_{x,\mu}\left(\bar{\chi}_x\, U^\dagger_{x,\mu}\,\chi_{x+a\hat{\mu}} - \bar{\chi}_{x+a\hat{\mu}}\, U_{x,\mu}\,\chi_x\right). \end{aligned} \tag{2.4.36}$$

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. [KMNP83]). 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.

2.4.4 Wilson fermions

The chiral lattice fermions were proposed by Wilson [Wil74]. Soon after that he recognized [Wil75] 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

$$\begin{aligned} S\left[U,\bar{\psi},\psi\right] &= \beta\, S_{\rm Lat}\left[U\right] + M\sum_x \bar{\psi}_x\psi_x \\ &-\frac{1}{2}\sum_{x,\mu>0}\left[\bar{\psi}_x\left(1-\gamma_\mu\right)U^\dagger_{x,\mu}\psi_{x+a\hat{\mu}} + \bar{\psi}_{x+a\hat{\mu}}\left(1+\gamma_\mu\right)U_{x,\mu}\,\psi_x\right]. \end{aligned} \tag{2.4.37}$$

The difference between this action and the action (2.4.2) for chiral fermions is due to the projector operators $(1\pm\gamma_\mu)$ which pick only one fermionic state.

Substituting the expansion (2.4.5) in the action (2.4.37), we get, in the naive continuum limit, the continuum fermionic action (2.1.14) with the mass being

$$m = \frac{M-4}{a}. \tag{2.4.38}$$

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

$$M \to 4 + ma\,. \tag{2.4.39}$$

In order to see that there are no other relativistic fermion states in the limit (2.4.39), let us consider the fermionic propagator which is given by

$$G^{-1}\left(k\right) = M - \frac{1}{2}\sum_{\mu=1}^{4}\left[\left(1-\gamma_\mu\right)\mathrm{e}^{\,ik_\mu a} + \left(1+\gamma_\mu\right)\mathrm{e}^{\,-ik_\mu a}\right]. \tag{2.4.40}$$

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

$$\cosh Ea = \frac{1+\left(M-\sum\limits_{\mu=1}^{3}\cos p_\mu a\right)^2 + \sum\limits_{\mu=1}^{3}\sin^2 p_\mu a}{2\left(M-\sum\limits_{\mu=1}^{3}\cos p_\mu a\right)}. \tag{2.4.41}$$

Let a particle be at rest, *i.e.* $p_1=p_2=p_3=0$ and $E=m>0$. Then Eq. (2.4.41) reduces for $ma\ll 1$ to the relation (2.4.39). It is easy to show that a particle at rest is the only solution to Eq. (2.4.41) whose energy is finite as $a\to 0$.

The difference between the dispersion laws for the chiral and Wilson fermions is because the function on the RHS of Eq. (2.4.41) is no longer periodic. It reduces for $a\to 0$ and $M\to 4$ to a usual relation

$$E = \sqrt{\vec{p}^{\,2}+m^2} \tag{2.4.42}$$

between the energy and momentum of a relativistic particle.

Problem 2.25 Show that a solution to the dispersion law (2.4.41) is unique for $M\approx 4$.

Solution For $M \approx 4$, we can replace the LHS of Eq. (2.4.41) by 1 and substitute $M = 4$ on the RHS. Then Eq. (2.4.41) reduces to the equation for spatial components of the four-momentum:

$$\left(3-\sum_{\mu=1}^{3}\cos p_\mu a\right)^2+\left(3-\sum_{\mu=1}^{3}\cos^2 p_\mu a\right) \;=\; 0\,, \qquad (2.4.43)$$

whose only solution is $p_1 = p_2 = p_3 = 0$ since both terms on the LHS are nonnegative.

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. (2.4.41) to be $\sim 1/a$ so that this state is inessential as $a \to 0$.

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 $a \to 0$.

Problem 2.26 Calculate the masses of all 16 fermionic states.
Solution Substituting Eqs. (2.4.23), (2.4.24) and so on into the action (2.4.37), we get

$$m \;=\; \frac{M-\sum_{\mu=1}^{4}s_\mu}{a} \qquad (2.4.44)$$

where

$$s_\mu \;=\; \mathrm{e}^{ip_\mu a}\begin{cases} +1 & p_\mu = 0 \\ -1 & p_\mu = \pi/a \end{cases}\,. \qquad (2.4.45)$$

Therefore, 1 state is relativistic as $M \to 4$ while 15 others have the masses $\sim 1/a$.

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

$$P^{\pm}_\mu \;=\; \frac{1\pm\gamma_\mu}{2} \qquad \boxed{\text{Wilson fermions}} \qquad (2.4.46)$$

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

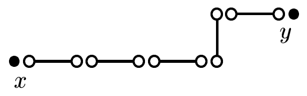


Fig. 2.25: A path Γ_{yx} made out of the string bits, which leads to a nonvanishing term of the hopping parameter expansion for the quark propagator (2.4.49) on a lattice. Each site involves at least two quark fields (depicted by the circles). Otherwise the Grassmann integral at a given site vanishes.

the negative ones. In particular, there are no backtrackings in the (lattice) sum over paths since

$$P^+_\mu P^-_\mu \;=\; 0\,. \qquad (2.4.47)$$

This removes the doubling.

Problem 2.27 Represent the fermion propagator in an external Yang–Mills field as a sum over paths on a lattice, performing an expansion in $1/M$.
Solution Let us rescale the fermion field, absorbing the parameter M in front of the mass term. The fermionic part of the new action reads

$$S_\psi \;=\; \sum_x \bar\psi_x\psi_x - \kappa\sum_{x,\mu>0}\left[\bar\psi_x P^-_\mu U^\dagger_{x,\mu}\psi_{x+a\hat\mu}+\bar\psi_{x+a\hat\mu}P^+_\mu U_{x,\mu}\psi_x\right], \qquad (2.4.48)$$

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. (2.4.48) by a string bit as in Fig. 2.4 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. (2.4.48) into the definition (2.4.4) and expanding the exponential in κ, we get a combination of terms constructed from the string bits. A nonvanishing contribution to the quark propagator

$$G^{ij}_{mn}(x,y;U) \;=\; \left\langle \psi^i_m(x)\bar\psi^j_n(y)\right\rangle_\psi\,, \qquad (2.4.49)$$

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. 2.25. Otherwise, the average over ψ and $\bar\psi$ vanishes due to the rules of integration over Grassmann variables described in Problem (1.17).

Therefore, we get

$$G^{ij}_{mn}(x,y;U) \;=\; \sum_{\Gamma_{yx}}\frac{1}{M^{L(\Gamma)+1}}U^{ij}[\Gamma_{yx}]\left[\prod_{\Gamma_{yx}}P^{\pm}_\mu\right]_{mn}\,, \qquad (2.4.50)$$

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. (2.4.46), while

$$P_\mu^\pm = \pm \frac{\gamma_\mu}{2} \qquad \boxed{\text{chiral fermions}} \tag{2.4.51}$$

for chiral fermions. The sum in Eq. (2.4.50) 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. (2.4.50) is a lattice analog of the continuum formula (2.1.43) for the case of fermions.

Problem 2.28 Represent the integral over fermions in Eq. (2.4.3) as a sum over closed paths on a lattice, performing an expansion in $1/M$.
Solution The calculation is analogous to that of the previous problem. The result reads

$$\int \prod_x d\bar\psi_x d\psi_x \,\mathrm{e}^{-S_\psi} = \mathrm{e}^{-S_{\mathrm{ind}}[U]} \tag{2.4.52}$$

with

$$S_{\mathrm{ind}}[U] = -\sum_\Gamma \frac{\operatorname{tr} U[\Gamma]}{L(\Gamma)\, M^{L(\Gamma)}} \operatorname{sp} \prod_\Gamma P_\mu^\pm , \tag{2.4.53}$$

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 (2.4.53) 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 (2.2.16) by introducing many flavors of lattice fermions [Ban83, Ham83].

2.4.5 Quark condensate

The lattice action (2.4.37) 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 $a \to 0$ 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. 2.26). $m_\pi = 0$ 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.

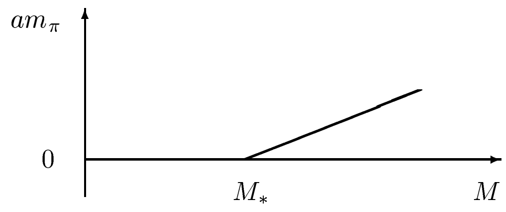


Fig. 2.26: Dependence of the π-meson mass on the lattice quark mass M. At $M = M_*$ the π-meson becomes massless and the chiral symmetry is restored.

For the chiral or Kogut–Susskind fermions with $M = 0$, the lattice action is invariant under the global chiral transformation (2.4.7). The order parameter for breaking the chiral symmetry is

$$\bar\psi\psi \xrightarrow{\text{c.t.}} \bar\psi\,\mathrm{e}^{2i\alpha\gamma_5}\psi , \tag{2.4.54}$$

which is not invariant under the chiral transformation. Therefore, the average of $\bar\psi\psi$ must vanish if the symmetry is not broken spontaneously[15]. Such spontaneous breaking results in

$$\left\langle \bar\psi\psi \right\rangle \neq 0 . \tag{2.4.55}$$

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 [HP81] are shown in Fig. 2.27. Its agreement with asymptotic scaling demonstrates that the chiral symmetry is spontaneously broken in the continuum QCD.

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 by using Eq. (1.2.15), one is

[15]Spontaneous symmetry breaking usually occurs when the vacuum state is not invariant under the symmetry transformation.

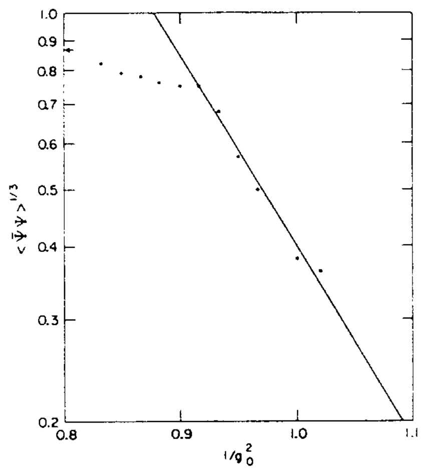


Fig. 2.27: Monte Carlo data by Hamber and Parisi [HP81] for the quark condensate in the quenched approximation.

left with the determinant, say for the Kogut–Susskind fermions, of the matrix

$$D\,[U] \;=\; M\,\delta_{x,y} + \frac{1}{2}\sum_{\mu>0}\left[\eta_{x,\mu}\,U^{\dagger}_{x,\mu}\delta_{x(x+a\hat{\mu})} - \eta_{x,\mu}\,U_{x,\mu}\delta_{x(x-a\hat{\mu})}\right] \quad (2.4.56)$$

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

$$\mathrm{e}^{\,-\beta\,S_{eff}[U]} \;=\; \det D\,[U]\;\mathrm{e}^{\,-\beta\,S[U]}\,. \quad (2.4.57)$$

The matrix that appears in this determinant has at least $N_cL^4 \times N_cL^4$ 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. [Lat94] for a review of the subject).

2.5 Finite temperatures

Finite-temperature quantum field theories at thermodynamic equilibrium are naturally described by Euclidean path integrals. The time-variable in this approach is compactified and varies between 0 and the inverse temperature $1/T$. Periodic boundary conditions are imposed on Bose fields while antiperiodic ones are imposed on Fermi fields in order to reproduce the standard Bose or Fermi statistics, respectively.

The lattice formulation of QCD at finite temperature is especially simple, since the Euclidean lattice has a finite extent in the temporal direction. The Wilson criterion of confinement is not applicable at finite temperatures and is replaced by another one, which is based on the thermal Wilson lines passing through the lattice in the temporal direction. They are closed due to the periodic boundary condition for the gauge field.

When the temperature increases, QCD undergoes [Pol78, Sus79] a deconfining phase transition which is associated with the liberation of quarks. At low temperatures below the phase transition, thermodynamical properties of the hadron matter are well described by a gas of noninteracting hadrons while at high temperatures above the phase transition these are well described by an ideal gas of quarks and gluons.

The situation with the deconfining phase transition becomes less definite when effects of virtual quarks are taken into account. The deconfining phase transition makes strict sense only for large values of the quark mass. For light quarks, a phase transition associated with the chiral symmetry restoration at high temperatures occurs with increasing the temperature. It makes strict sense only for massless quarks.

We first derive, in this Section, a path-integral representation of finite-temperature quantum field theories starting from the Boltzmann distribution. Then we apply this technique to QCD and discuss the confinement criterion at finite temperatures as well as the deconfining and chiral symmetry restoration phase transitions which occur with increasing the temperature.

2.5.1 Feynman–Kac formula

Thermodynamic properties of an equilibrium system in $3+1$ dimensions are determined by the partition function

$$\begin{aligned} Z\,(T,V) \;&=\; \sum_n \mathrm{e}^{\,-E_n/T} \\ &\equiv\; \mathrm{Sp}\,\mathrm{e}^{\,-H/T} \end{aligned} \quad (2.5.1)$$

which is associated with the Boltzmann distribution at the temperature T. Here H is a Hamiltonian of the system and Sp is calculated over any complete set of states, say, over eigenstates of the Hamiltonian whose eigenvalues are characterized by the energy E_n.

For a quantum theory of a single scalar field $\varphi(\vec{x},t)$, the (Schrödinger) states are described by the bra- and ket-vectors $\langle g|$ and $|f\rangle$:

$$\langle g|\,\vec{x}\rangle \;=\; g(\vec{x})\,, \qquad \langle\vec{x}\,|f\rangle \;=\; f(\vec{x})\,, \tag{2.5.2}$$

as is explained in Subsect. 1.1.1. The matrix element of the evolution operator $\exp(-H/T)$ is given by the formula

$$\langle g|\,\mathrm{e}^{-H/T}\,|f\rangle \;=\; \int\limits_{\substack{\varphi(\vec{x},0)=f(\vec{x})\\ \varphi(\vec{x},1/T)=g(\vec{x})}} D\varphi(\vec{x},t)\;\mathrm{e}^{-\int_0^{1/T}dt\,\mathcal{L}[\varphi]}\,, \tag{2.5.3}$$

where $\mathcal{L}$ is the proper Lagrangian, say for example,

$$\mathcal{L}[\varphi] \;=\; \int\limits_V d^3\vec{x}\left(\frac{1}{2}(\partial_\mu\varphi)^2+\frac{1}{2}m^2\varphi^2+\frac{\lambda}{3!}\varphi^3\right) \tag{2.5.4}$$

for the cubic self-interaction of φ. The derivation is quite analogous to that of Problem 1.8.

In order to calculate the trace over states, one should put $g(\vec{x})=f(\vec{x})$ and perform the additional integration over $f(\vec{x})$. This yields the Feynman–Kac formula[16]

$$\begin{aligned}\mathrm{Sp}\,\mathrm{e}^{-H/T} \;&=\; \int Df(\vec{x})\,\langle f|\,\mathrm{e}^{-H/T}\,|f\rangle\\ &=\; \int\limits_{\varphi(\vec{x},1/T)=\varphi(\vec{x},0)} D\varphi(\vec{x},t)\;\mathrm{e}^{-\int_0^{1/T}dt\,\mathcal{L}[\varphi]}\,.\end{aligned} \tag{2.5.5}$$

Notice that the path integral in Eq. (2.5.5) is taken with periodic boundary conditions for the field φ:

$$\varphi(\vec{x},1/T)=\varphi(\vec{x},0)\,. \tag{2.5.6}$$

It reproduces as $T\to 0$ the standard Euclidean formulation of quantum field theory which is discussed in Sect. 1.2. The point is that nothing depends on

[16]Its derivation in the modern context of non-Abelian gauge theories, which extends the Feynman derivation [Fey53] for statistical mechanics, is due to Bernard [Ber74].

real time for a system at equilibrium. The variable t in Eq. (2.5.5) is just the proper time of the disentangling procedure. This analogy between the partition functions of statistical systems and the Euclidean formulation of quantum field theory is already mentioned in the Remark to Subsect. 1.1.7.

Problem 2.29 Derive the Feynman–Kac formula for a quantum particle with non-relativistic Hamiltonian (1.1.103).
Solution The matrix element $\langle x\,|\exp(-H/T)|\,x\rangle$ is determined by Eq. (1.1.114) to be

$$\left\langle x\,\middle|\,\mathrm{e}^{-H/T}\middle|\,x\right\rangle \;=\; \int\limits_{\substack{z_\mu(0)=x_\mu\\ z_\mu(1/T)=x_\mu}} Dz_\mu(t)\;\mathrm{e}^{-\int_0^{1/T}dt\,\mathcal{L}(t)}\,, \tag{2.5.7}$$

where the Lagrangian $\mathcal{L}(t)$ is given by Eq. (1.1.115). Integrating over d^dx, we get [Fey53]

$$\begin{aligned}\mathrm{Sp}\,\mathrm{e}^{-H/T} \;&=\; \int\limits_V d^dx\left\langle x\,\middle|\,\mathrm{e}^{-H/T}\middle|\,x\right\rangle\\ &=\; \int\limits_{z_\mu(0)=z_\mu(1/T)} Dz_\mu(t)\;\mathrm{e}^{-\int_0^{1/T}dt\,\mathcal{L}(t)}\,.\end{aligned} \tag{2.5.8}$$

This integral is over the trajectories with periodic boundary conditions

$$z_\mu(0) \;=\; z_\mu(1/T)\,. \tag{2.5.9}$$

Problem 2.30 Calculate the partition function (2.5.8) for the free case.
Solution The Gaussian path integral with the boundary conditions

$$z_\mu(0) \;=\; z_\mu(1/T) \;=\; x_\mu \tag{2.5.10}$$

is calculated in Subsect. 1.1.5 with the result given by Eq. (1.1.86). In order to calculate the partition function (2.5.8), we need to integrate this expression over x_μ which yields [Fey53]

$$Z(T,V) \;=\; \int\limits_V d^dx\,\mathcal{F}(1/mT) \;=\; V\left(\frac{mT}{2\pi}\right)^{d/2}\,. \tag{2.5.11}$$

The formula (2.5.11) is to be compared with that given by the Boltzman distribution in classical statistics. Since the energy of a free non-relativistic particle is

$$E(\vec{p}) \;=\; \frac{\vec{p}^{\,2}}{2m}\,, \tag{2.5.12}$$

the Boltzmann distribution is given by the sum over positions of the particle in a box of volume V and the integration over its momentum $\vec{p}$:

$$Z(T,V) \;=\; V\int \frac{d^d\vec{p}}{(2\pi)^d}\,\mathrm{e}^{\,-E(\vec{p})/T} \;=\; V\left(\frac{mT}{2\pi}\right)^{d/2}, \tag{2.5.13}$$

which coincides with Eq. (2.5.11) derived from the path integral.

Problem 2.31 Calculate the partition function (2.5.5) for the free case.
Solution Since the path integral over $\varphi(\vec{x},t)$ is Gaussian, it can be represented as

$$\begin{aligned}
\ln Z(T,V) &= -\frac{1}{2}\ln\det\left(-\partial_\mu^2+m^2\right)\\
&= -\frac{1}{2}\mathrm{Sp}\,\ln\left(-\partial_\mu^2+m^2\right)\\
&= -\frac{1}{2}V\int\frac{d^d\vec{p}}{(2\pi)^d}\mathrm{Sp}_t\,\ln\left(-D^2+\omega^2\right),
\end{aligned} \tag{2.5.14}$$

where

$$\omega \;=\; \sqrt{\vec{p}^{\,2}+m^2}\,. \tag{2.5.15}$$

We have used the fact that the $\vec{x}$ variable is not restricted while the remaining trace of the one-dimensional operator is to be calculated with periodic boundary conditions.

We shall perform the calculation by expressing the trace via the diagonal resolvent of the same operator as it was already done in the Problem 1.29. The Green function $G_\omega(t-t')$ is no longer given by Eq. (1.1.38) because of the periodic boundary conditions. We get, instead, the sum over even frequencies:

$$G_\omega(t-t') \;=\; T\sum_{n=-\infty}^{+\infty}\frac{\mathrm{e}^{\,i2n\pi T(t'-t)}}{(2n\pi T)^2+\omega^2}, \tag{2.5.16}$$

which satisfies $G_\omega(1/T)=G_\omega(0)$, as it should be for periodic boundary conditions, and reproduces Eq. (1.1.38) as $T\to 0$. The diagonal resolvent is given by

$$G_\omega(0) \;=\; T\sum_{n=-\infty}^{+\infty}\frac{1}{(2n\pi T)^2+\omega^2} \;=\; \frac{1}{2\omega}\coth\frac{\omega}{2T}\,. \tag{2.5.17}$$

Therefore,

$$\begin{aligned}
\mathrm{Sp}_t\,\ln\left(-D^2+\omega^2\right) &= \int^{\omega^2} d\omega^2\int_0^{1/T} dt\,G_\omega(0)\\
&= \int^{\omega} d\omega\frac{1}{T}\coth\frac{\omega}{2T} \;=\; \frac{\omega}{T}+2\ln\left(1-\mathrm{e}^{\,-\omega/T}\right)
\end{aligned} \tag{2.5.18}$$

modulo an ω-independent constant. Substituting into Eq. (2.5.14), we get

$$\ln Z(T,V) \;=\; -V\int\frac{d^d\vec{p}}{(2\pi)^d}\left[\frac{\omega}{2T}+\ln\left(1-\mathrm{e}^{\,-\omega/T}\right)\right], \tag{2.5.19}$$

which is the standard result for an ideal Bose gas in quantum statistics modulo the first term on the RHS associated with the zero-point energy of the vacuum.

2.5.2 QCD at finite temperature

QCD at finite temperatures is described by the partition function

$$Z\,(T,V) \;=\; \int DA_\mu\,D\bar\psi\,D\psi\,\mathrm{e}^{\,-\int_0^{1/T}dt\,\int_V d^3\vec{x}\,\mathcal{L}[A_\mu,\psi,\bar\psi]}, \tag{2.5.20}$$

which is the proper analog of Eq. (2.5.5). The path integral is taken with the boundary conditions

$$\begin{aligned}
A_\mu\,(\vec{x},1/T) &= A_\mu\,(\vec{x},0)\,, &(2.5.21)\\
\psi\,(\vec{x},1/T) &= -\psi\,(\vec{x},0)\,, &(2.5.22)\\
\bar\psi\,(\vec{x},1/T) &= -\bar\psi\,(\vec{x},0)\,, &(2.5.23)
\end{aligned}$$

which are periodic for the gauge field (gluon) and antiperiodic for the Fermi fields (quarks). The antiperiodicity of the Fermi fields is related, roughly speaking, with the famous extra minus sign of fermionic loops in the vacuum energy.

Problem 2.32 Calculate the partition function for free massive one-dimensional fermions with antiperiodic boundary conditions

$$\psi\,(1/T) \;=\; -\,\psi\,(0)\,, \qquad \bar\psi\,(1/T) \;=\; -\,\bar\psi\,(0)\,. \tag{2.5.24}$$

Solution The calculation is analogous to that of Problem 2.31. We get

$$\ln Z(T,V) \;=\; \ln\det\left(D+m\right) \;=\; \mathrm{Sp}\,\ln\left(D+m\right). \tag{2.5.25}$$

The fermion Green function $G_m(t-t')$ is given by the sum over odd frequencies:

$$G_m(t-t') \;=\; T\sum_{n=-\infty}^{+\infty}\frac{\mathrm{e}^{\,i(2n+1)\pi T(t'-t)}}{i(2n+1)\pi T+m}, \tag{2.5.26}$$

which satisfies $G_m(1/T) \;=\; -G_m(0)$, as it should be for antiperiodic boundary conditions.

As $T \to 0$, we get

$$G_m(t-t') = \int_{-\infty}^{+\infty} \frac{d\epsilon}{2\pi} \frac{\mathrm{e}^{\,i\epsilon(t'-t)}}{i\epsilon+m} = \theta\left(t-t'\right) \tag{2.5.27}$$

since the contour of integration over ϵ can be closed for $t > t'$ $(t < t')$ in the lower (upper) half-plane. We have thus reproduced the fermionic Green function (2.1.35) from Problem 2.3.

The diagonal resolvent is given by

$$G_m(0) = T \sum_{n=-\infty}^{+\infty} \frac{1}{i(2n+1)\pi T + m} = \frac{1}{2} \tanh \frac{m}{2T}\,, \tag{2.5.28}$$

which differs from Eq. (2.5.17) by the change of the coth for the tanh. Therefore,

$$\ln Z(T,V) = \int^m dm \frac{1}{T} \tanh \frac{m}{2T} = \frac{m}{2T} + \ln\left(1 + \mathrm{e}^{\,-m/T}\right) \tag{2.5.29}$$

modulo an m-independent constant. The second term on the RHS involves the plus sign which characterizes the Fermi statistics (remember that $\omega = m$ if there is no spatial dimensions). If we were choose periodic boundary conditions instead of antiperiodic ones, we would get the minus sign as in Eq. (2.5.19) which is wrong for fermions. The first term on the RHS is associated again with the zero-point energy of the vacuum.

The discussion of the previous Subsection about the relation between the finite-temperature and Euclidean formulations explains why the latter allows to calculate in QCD only static quantities, say hadron masses or interaction potentials, which do not depend on the time. It is also worth noting that we did not add a gauge fixing term in Eq. (2.5.20) having in mind a lattice quantization as before.

The lattice formulation of finite-temperature QCD is especially simple. One should take an asymmetric lattice whose size along the temporal axis is much smaller than that along the spatial ones:

$$L_t = \frac{1}{Ta} \ll L\,. \tag{2.5.30}$$

This guarantees the system to be in a thermodynamic limit. Then the temperature is given by

$$T = \frac{1}{aL_t}\,, \tag{2.5.31}$$

i.e. coincide with the inverse extent of the lattice along the temporal axis. The periodic boundary conditions are usually imposed on the lattice by construction.

Since the lattice spacing a and the bare coupling constant g^2 are related by Eq. (2.2.85), the temperature (2.5.31) can be rewritten as

$$T = \frac{1}{L_t} \Lambda_{\mathrm{QCD}} \exp\left[\int \frac{dg^2}{\mathcal{B}(g^2)}\right]. \tag{2.5.32}$$

Therefore, one can change the temperature on the lattice varying either the size along temporal axis L_t or g^2.

2.5.3 Confinement criterion at finite temperature

The Wilson's criterion of confinement, which is discussed in Subsect. 2.2.6, is not applicable at finite temperatures. A proper criterion for confinement at finite temperatures was proposed by Polyakov [Pol78].

The Polyakov criterion of confinement at finite temperature uses the thermal Wilson loop which goes along the temporal direction:

$$L(\vec{x}) = \operatorname{tr} \mathrm{P}\, \mathrm{e}^{\int_0^{1/T} dt\, \mathcal{A}_d(\vec{x},t)}\,. \tag{2.5.33}$$

It is gauge invariant because of the periodic boundary conditions for the gauge field and is called the *Polyakov loop* or the Wilson line. One can imagine that the time-variable $t \equiv x_d$ is compactified so that the Polyakov loop winds around the torus in the temporal direction as is shown in Fig. 2.28.

The lattice Polyakov loop

$$L_{\vec{x}} = \operatorname{tr} \prod_{x_d} U_{x,d} \tag{2.5.34}$$

is just the trace of the product of the link variables along a line which goes in the temporal direction through the lattice with imposed periodic boundary conditions.

Using the lattice gauge transformation (2.2.13), almost all link variables, associated with links pointing in the temporal direction, can be put equal 1 except for one time slice since the gauge transformation is periodic:

$$\Omega\left(\vec{x},0\right) = \Omega\left(\vec{x},1/T\right). \tag{2.5.35}$$

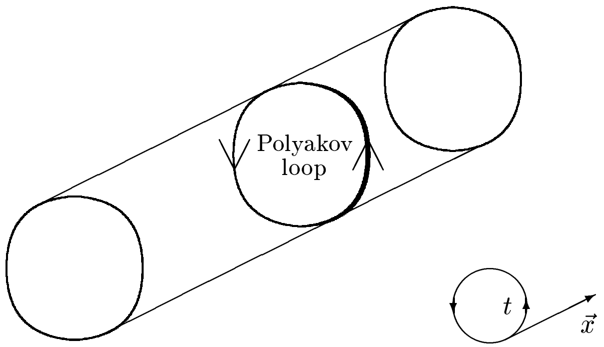


Fig. 2.28: Polyakov loop which winds around compactified temporal direction.

The average of the Polyakov loop is related to the free energy $F_0(\vec{x})$ of a single quark (minus that of the vacuum) nailed at the point $\vec{x}$ of a three-dimensional space by

$$\langle L(\vec{x})\rangle = \mathrm{e}^{-F_0/T} . \qquad (2.5.36)$$

If F_0 is infinite, which is associated with confinement, then

$$\langle L(\vec{x})\rangle = 0 \qquad \boxed{\text{confinement}} . \qquad (2.5.37)$$

On the contrary,

$$\langle L(\vec{x})\rangle \neq 0 \qquad \boxed{\text{deconfinement}} \qquad (2.5.38)$$

is associated with deconfinement. This is the Polyakov criterion of confinement at finite temperature.

This criterion establishes a connection on a lattice between confinement and the Z(3) symmetry — the center of the SU(3). The Z(3)-transformation of the link variables

$$U_{x,d} \rightarrow Z_{x_d} U_{x,d} \qquad (Z_{x_d} \in \mathrm{Z}(3)) \qquad (2.5.39)$$

leaves the lattice action to be invariant. This transformation is not of the type of the local gauge transformation (2.2.13) since only the temporal link-variables are transformed. The parameter Z_{x_d} of the transformation (2.5.39) depends on x_d but is independent of the spatial coordinates $\vec{x}$ so that the symmetry is a global one.

While the lattice action is invariant under the transformation (2.5.39), the Polyakov loop transforms as

$$L_{\vec{x}} \rightarrow Z\, L_{\vec{x}} \qquad (Z \in \mathrm{Z}(3)) , \qquad (2.5.40)$$

where

$$Z = \prod_{x_d} Z_{x_d} . \qquad (2.5.41)$$

Therefore, Eq. (2.5.37) holds if the symmetry is unbroken, while Eq. (2.5.38) signalizes about spontaneous breaking of the symmetry. Thus, confinement or deconfinement are associated with the unbroken or broken global Z(3) symmetry, respectively.

On a lattice of a finite volume, the number of degrees of freedom is finite and spontaneous breaking of the Z(3) symmetry is impossible. It is more convenient to use then the criterion which is based on the correlator of two Polyakov loops separated by a distance R along a spatial direction. This correlator determines the interaction energy $E(R)$ between a quark and an antiquark by the formula

$$\left\langle L(\vec{x}) L^\dagger(\vec{y}) \right\rangle_{\text{conn}} = \mathrm{e}^{-E(R)/T} . \qquad (2.5.42)$$

A finite correlation length is now associated with confinement while an infinite one corresponds to deconfined quarks.

More about the Z(3) symmetry in finite-temperature lattice gauge theories can be found in the review by Svetitsky [Sve86].

Problem 2.33 Calculate the correlator (2.5.42) to the leading order of the strong coupling expansion.
Solution The calculation is analogous to that of Subsect. 2.2.5. The group integral is nonvanishing when the plaquettes completely fill a cylinder, spanned by two Polyakov loop, whose area is R/T. This is an analog of the filling shown in Fig. 2.10. Contracting the indices, we get

$$\left\langle L_{\vec{x}} L^\dagger_{\vec{y}} \right\rangle_{\text{conn}} = (W_{\partial p})^{R/T} , \qquad (2.5.43)$$

where $W_{\partial p}$ is given by Eq. (2.2.72). This yields the same interaction potential $E(R)$ as before (see Eqs. (2.2.76) and (2.2.77)).

Remark on high temperatures

At high temperatures $T \to \infty$, the temporal direction shrinks and the partition function (2.5.20) reduces to a three-dimensional one with the coupling constant

$$g_{3d}^2 \;=\; g^2 T\,, \tag{2.5.44}$$

which has a dimension of [mass] in three dimensions. The three dimensional QCD and QED always confine. If we take a Wilson loop in the form of a rectangle along spatial directions in four-dimensional QCD at high temperature, its average coincides with that in three dimensions and obeys the area law. This does not mean, however, that we are in a confining phase since the confinement criterion at finite temperature is different [Pol78].

2.5.4 Deconfining transition

The effects of finite temperatures are negligible under normal circumstances in QCD where the typical energy scale is of the order of hundreds MeV while the temperature, say, of $T \approx 300$ °K is associated with the energy[17] $kT \approx 3\cdot 10^{-8}$ MeV. However, for the times $\approx \mu s$ in the very early universe, energies of thermal fluctuations were ~ 200 MeV, *i.e.* of the order of the mass of the π-meson. Therefore, π-mesons can be created out of the vacuum at those times while their density in a unit volume is described by thermodynamics of an ideal gas. Heavier hadrons are suppressed at these energies by the Boltzmann factor.

The energy density $\mathcal{E}(T)$ of the hadron matter is given by the standard thermodynamical relation

$$\mathcal{E}(T) \;=\; \frac{1}{V}\frac{\partial}{\partial(1/T)} \ln Z(T,V)\bigg|_V\,, \tag{2.5.45}$$

with $Z(T,V)$ given by Eq. (2.5.20).

When the density of hadrons is small, $\mathcal{E}(T)$ is given by the formula

$$\mathcal{E}_{\rm h}(T) \;=\; \frac{T}{2\pi^2}\sum_{i=\pi,\rho,\omega,\ldots} g_i\left[m_i^3 {\rm K}_1(m_i/T) + 3m_i^2{\rm K}_2(m_i/T)\right], \tag{2.5.46}$$

where $g_\pi = 3$, $g_\rho = 9$, $g_\omega = 3$, ... are the statistical weights of the π, ρ, ω, ... mesons while K_1 and K_2 are the modified Bessel functions.

[17] Here $k = 8.6\cdot 10^{-11}$ Mev/°K is Boltzmann constant.

Problem 2.34 Derive Eq. (2.5.46) starting from the partition function (2.5.19).
Solution For a dilute gas, the logarithm in Eq. (2.5.19) can be expanded in $\exp(-E/T)$. Therefore, we get

$$\begin{aligned}\ln Z(T,V) \;&=\; \frac{\rm const}{T} + V\int\frac{d^3\vec{p}}{(2\pi)^3}\,{\rm e}^{-\sqrt{\vec{p}^{\,2}+m^2}/T}\\ &=\; \frac{\rm const}{T} + \frac{VTm^2}{2\pi^2}{\rm K}_2(m/T)\,.\end{aligned} \tag{2.5.47}$$

The second term on the RHS describes the classical statistics of an ideal gas of relativistic particles. Equation (2.5.46) can now be derived by differentiating this formula with respect to $1/T$ according to Eq. (2.5.45) and taking into account statistical weights of the hadron states. The zero-point energy term gives a T-independent contribution to $\mathcal{E}_{\rm h}$, which only changes the reference level of energy.

At low temperatures, the hadron matter is in the confinement phase. However, when the temperature is increased, a phase transition associated with deconfinement occurs at some temperature $T = T_c$ as was first pointed out by Polyakov [Pol78] and Susskind [Sus79]. For $T < T_c$ the interaction potential between static quarks is linear, as is shown in Fig. 2.11a, while for $T > T_c$ the potential is deconfining, as is shown in Fig. 2.11b. The state of the hadron matter with deconfined quarks and gluons is often called the *quark-gluon plasma*.

There exists a very simple physical argument why the deconfining phase transition must occur in QCD when the temperature is increased. It is based on the string picture of confinement which was considered in Subsect. 2.2.6. The string is made of the gluon field between static quarks in the confining phase, which are associated with the string end points. With increasing the temperature, a condensation of strings of an infinite length will inevitably occur due to a large entropy of such states, which corresponds to a deconfining phase transition.

Problem 2.35 Derive the temperature of a phase transition for an elastic string analyzing the temperature dependence of its free energy.
Solution Let us consider the thermodynamics of an elastic string with nailed end points. For low temperatures, thermal fluctuations of the length of the string are suppressed by the Boltzmann factor since the energy is proportional to the length. Therefore, the string is tightened along the shortest distance between the quarks which leads to a linear potential.

When the temperature is increased, entropy effects associated with fluctuations of a shape of the string become essential. An increment of the string length l by Δl looses the energy

$$\Delta E \;=\; \frac{\partial E}{\partial l}\Delta l \;=\; K\Delta l\,, \tag{2.5.48}$$

where K is the string tension as before, but gains the entropy

$$\Delta S \;=\; \frac{\partial S}{\partial l}\Delta l\,. \tag{2.5.49}$$

The change of the free energy is given by

$$\Delta F \;=\; \Delta E - T\Delta S \;=\; \left[K - T\frac{\partial S}{\partial l}\right]\Delta l\,. \tag{2.5.50}$$

A phase transition occurs at the temperature

$$T_c \;=\; K\left(\frac{\partial S}{\partial l}\right)^{-1}, \tag{2.5.51}$$

when the changes of the energy and entropy compensate each other, so that the free energy ceases to depend on Δl. Therefore, the phase transition is associated with a condensation of arbitrary long strings.

The energy density $\mathcal{E}(T)$ is described by a free gas of hadrons for low temperatures, as is already mentioned, and by a free gas of quarks and gluons at high temperatures. The latter statement is due to asymptotic freedom, which says that the effective coupling constant describing strong interaction at the temperature T is given by

$$g^2(T) \;=\; \frac{8\pi^2}{b\ln\left(\dfrac{\Lambda_{\mathrm{QCD}}}{T}\right)} \tag{2.5.52}$$

with

$$b \;=\; -11 + \frac{2}{3}N_f \tag{2.5.53}$$

and N_f being the number of fermion species (or flavors) whose mass is much less than T. This formula has the same structure as the running constant $g^2(Q)$ which describes strong interaction at the momenta Q. Since $Q \sim T$ for thermal fluctuations, these two coupling constants coincide with logarithmic accuracy.[18]

The energy density $\mathcal{E}(T)$ of the quark-gluon plasma is given by Boltzmann law

$$\mathcal{E}_{\mathrm{p}}(T) \;=\; g_{\mathrm{p}}\frac{\pi^2}{30}T^4 + B \tag{2.5.54}$$

[18]The perturbative calculations in QCD at finite temperature are described in the book by Kapusta [Kap89].

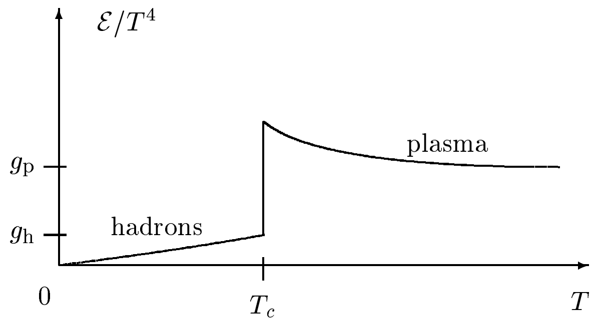


Fig. 2.29: Temperature dependence of the energy density of the hadron matter. $\mathcal{E}(T)$ for the hadron and plasma phases are given by Eqs. (2.5.46) and (2.5.54). The difference $\mathcal{E}_{\mathrm{p}} - \mathcal{E}_{\mathrm{h}}$ at the temperature T_c of the deconfinement phase transition is equal to the latent heat $\Delta\mathcal{E}$.

where

$$g_{\mathrm{p}} \;=\; 2\cdot 8 + \frac{7}{8}\cdot 2\cdot 2\cdot 3\cdot N_f \tag{2.5.55}$$

is the statistical weight, *i.e.* the number of independent internal degrees of freedom of the particles of the ideal gas. There are 2 spin and 8 color states of gluons, and 2 spin, 2 particle-antiparticle, 3 color and N_f flavor states of quarks ($N_f = 2$ for the u- and d-quarks). The factor 7/8 is a usual one for fermions.

The T-independent constant $B > 0$ in Eq. (2.5.54) is associated with the fact that the vacuum energy in the plasma phase is higher than in the hadron one. In other words, the energy density of the perturbative vacuum is larger by B than that of a non-perturbative one. It is because of this energy difference that hadrons are stable at low temperatures. Such a shift of energy densities between perturbative and non-perturbative vacua is typical for the bag models of hadrons.

Numerical Monte Carlo simulations of lattice gauge theory at finite temperature indicate that the deconfining phase transition is of the first order and occurs at $T_c \approx 200$ MeV. The actual dependence of the energy density on T, calculated by the Monte Carlo method, is well described by Eq. (2.5.46) for $T < T_c$ and Eq. (2.5.54) for $T > T_c$. This behavior is illustrated by Fig. 2.29.

Problem 2.36 Calculate T_c and the latent heat $\Delta\mathcal{E}$ approximating $\mathcal{E}_{\rm h}$ by an ideal gas of massless π-mesons.
Solution It is reasonable to disregard the mass of the π-mesons for $T \gtrsim 200$ Mev. Then

$$\mathcal{E}_{\rm h}(T) = g_{\rm h}\frac{\pi^2}{30}T^4 \tag{2.5.56}$$

with $g_{\rm h}=3$ being due to the three isotopic states (π^+, π^-, and π°). $\mathcal{E}_{\rm h}(T)$ for the plasma state is given by Eq. (2.5.54).

The pressure for the relativistic gases with the energy densities (2.5.56) and (2.5.54) is given, respectively, by

$$\mathcal{P}_{\rm h}(T) = g_{\rm h}\frac{\pi^2}{90}T^4 \tag{2.5.57}$$

and

$$\mathcal{P}_{\rm p}(T) = g_{\rm p}\frac{\pi^2}{90}T^4 - B\,. \tag{2.5.58}$$

The positive constant B in the energy density (2.5.54) leads to a negative pressure in the plasma state at low temperatures. Therefore, the hadron phase is preferable at low temperatures. This is in a spirit of the bag model of hadrons. At high energies the pressure is higher for the plasma phase, since

$$g_{\rm p} = 37 > g_{\rm h} = 3\,, \tag{2.5.59}$$

so that the plasma phase is realized. The pressure against T^4 is shown in Fig. 2.30 for both phases of the hadron matter.

The deconfinement phase transition occurs when the pressures in both phases coincide. Therefore, we get

$$T_c^4 = \frac{B}{\frac{\pi^2}{90}\left(g_{\rm p}-g_{\rm h}\right)} \tag{2.5.60}$$

and

$$\Delta\mathcal{E} \equiv \mathcal{E}_{\rm p}(T) - \mathcal{E}_{\rm h}(T) = 4B\,. \tag{2.5.61}$$

If we were put $g_{\rm h}=0$ in Eq. (2.5.60), this would change the value of T_c by few percents. This justifies the approximation of massless π-mesons.

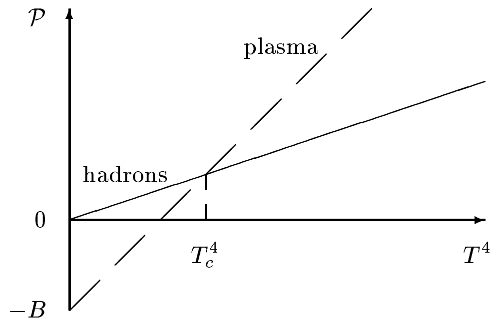


Fig. 2.30: Pressure against T^4 for the two phases of the hadron matter. The solid and dashed lines represent Eqs. (2.5.57) and (2.5.58), respectively. The hadron phase is stable for $T < T_c$, while the plasma phase is stable for $T > T_c$.

Remark on deconfining transition in cosmology

The confining phase transition from quark-gluon plasma to hadrons happened in the early universe when its age was $\approx \mu s$. The equation of state of the hadron matter is described by Eqs. (2.5.54), (2.5.58) before that time and by Eqs. (2.5.46), (2.5.57) after that time. There are presumably no cosmological consequences of this phase transition, which survive till our time, since it happened too long ago. For instance, fluctuations of the hadron matter density which might occur just after the phase transition were washed out by the further expansion. The confining phase transition in the early universe is considered in the review [CGS86], Section 6.

2.5.5 Restoration of chiral symmetry

The chiral symmetry is spontaneously broken in QCD at $T=0$, as is discussed in Subsect. 2.4.5. With increasing the temperature, the chiral symmetry should restore at some temperature T_{ch} (which not necessarily coincides with T_c) since perturbation theory is applicable at high T. This restoration occurs as a phase transition with $\langle\bar\psi\psi\rangle$ being the proper order parameter. Therefore, the quark condensate is destroyed at $T=T_{ch}$. Monte Carlo simulations indicates this phase transition to be of the first order.

However, there is a subtlety in the above string picture of quark confinement when virtual quarks are taken into account. The effects of virtual

Fig. 2.31: Breaking of the flux tube by creating a quark-antiquark pair (depicted by the open circles) out of the vacuum.

quarks are suppressed when their mass m is infinitely large and the picture of confinement is the same as in pure gluodynamics: quarks are permanently confined by strings made of the flux tubes of the gluon field. This is associated with a linear interaction potential.

For light virtual quarks, the flux tube can break creating a quark-antiquark pair out of the vacuum as is shown in Fig. 2.31. This happens when the energy saved in the flux tube is large enough to compensate the kinetic energy of the produced particles. Hence, the linear growth of the potential will stop at such distances.

The average of the Polyakov loop (2.5.33) is no longer criterion for quark confinement in the presence of virtual quarks. The test static quark can always be screened by an antiquark created out of the vacuum (a quark created at the same time will go to infinity). Therefore, the free energy F_0 in Eq. (2.5.36) is always finite so that $\langle L(\vec{x})\rangle \neq 0$ in both phases.

The effects of virtual quarks usually weaken a phase transition in a pure gauge theory. If the deconfining phase transition in the SU(3) pure gauge theory was of the second order rather than of the first order, it would presumably disappear for an arbitrary large but finite value of m. Such a phenomenon happens in the Ising model where an arbitrary small external magnetic field (which is an analog of the quark mass) destroys the second order phase transition. A discontinuity of $\langle L(\vec{x})\rangle$ at the first order deconfining transition continues in the T, m-plane as is illustrated by Fig. 2.32. It seems to terminate at some value m_c of the quark mass.

This situation with the order parameter for the deconfinement phase transition is somewhat similar to that for the chiral phase transition. $\langle\bar{\psi}\psi\rangle$ vanishes in the unbroken phase only for $m = 0$. If $m \neq 0$ but is small, there is a small explicit breaking of chiral symmetry due to the quark mass. Since the chiral phase transition is of the first order for $m = 0$, it is natural to expect that a discontinuity of $\langle\bar{\psi}\psi\rangle$ continues in the T, m-plane up to some value m_{ch} of the quark mass.

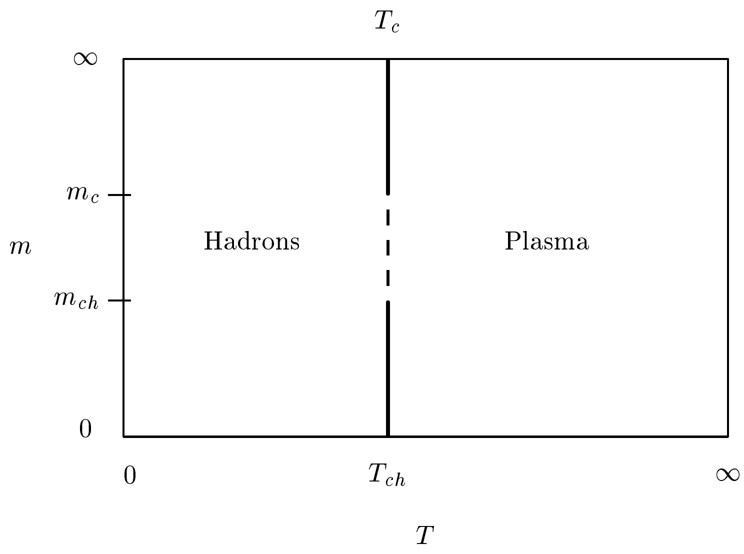


Fig. 2.32: Expected phase diagram of the hadron matter in the T, m-plane. The deconfining phase transition starts at $T = T_c$ for $m = \infty$. $\langle L(\vec{x})\rangle$ is its order parameter for $m > m_c$. The chiral phase transition starts at $T = T_{ch}$ for $m = 0$. $\bar{\psi}\psi$ is its order parameter for $m < m_{ch}$.

If $m_{ch} < m_c$, the phase diagram in the T, m-plane may look like that shown in Fig. 2.32. In the intermediate region $m_{ch} < m < m_c$, the behavior of neither $\langle L(\vec{x})\rangle$ nor $\langle\bar{\psi}\psi\rangle$ can answer the question whether a phase transition (or two separate transitions) occurs. A proper parameter, which signals about a phase transition is this region, could be the temperature dependence of the energy density $\mathcal{E}(T)$ that undergoes discontinuities at the points of first order phase transitions.

It is worth noting that an alternative behavior of the phase diagram in the T, m-plane, when $m_{ch} > m_c$, is not confirmed by Monte Carlo simulations.

2.6 Reference guide

There is a lot of very good introductory lectures/reviews on lattice gauge theories. For a perfect description of motivations and the lattice formulation, I would recommend the original paper [Wil74] and lectures [Wil75] by Wilson. Various topics of the lattice gauge theories are covered by the well-written book by Creutz [Cre83]. The book by Seiler [Sei82] contains some mathematically rigorous results. The recently published book [MM94] contains a contemporary look at the lattice theories.

I shall also list some of the reviews on lattice gauge theories which might be useful for deeper studies of the subject. The strong coupling expansion and the mean-field method are discussed in [DZ83]. The Monte Carlo method and some results of numerical simulations are considered in [CJR83, Mak84]. The fermion doubling problem and the Wilson fermions are discussed in the lectures [Wil75].

An introduction to quantum field theory at finite temperature is given in the book by Kapusta [Kap89], which contains, in particular, a discussion of perturbation theory in QCD at finite temperature. Lattice gauge theory aspects of the deconfinement phase transition at finite temperature is considered in the review [Sve86]. A description of the phases of the hadron matter at various temperatures, a comparison with results of the Monte Carlo simulations and a discussion of the deconfining phase transition in the early universe is contained in the review [CGS86].

Most of the reviews mentioned above were written in the beginning or middle of the eighties. Since that time a great progress happens in the Monte Carlo simulations of lattice gauge theories. New calculations are performed on larger lattices and with better statistics. A vast amount of papers on lattice gauge theories was published during the last ten years. There are more than 1000 subscribers to an electronic preprint archive `hep-lat@ftp.scri.fsu.edu`. The proceedings of the Lattice Conferences ([Lat94] and for the preceeding years) provide an updated information about this subject.

References to Chapter 2

[Ake93] K. Akemi *et al.*, *Scaling study of pure gauge lattice QCD by Monte Carlo renormalization group method*, Phys. Rev. Lett. **71** (1993) 3063.

[Are80] I.Ya. Aref'eva, *Non-Abelian Stokes theorem*, Theor. Math. Phys. **43** (1980) 353.

[Ban83] M. Bander, *Equivalence of lattice gauge and spin theories*, Phys. Lett. **126B** (1983) 463.

[BC81] G. Bhanot and M. Creutz, *Variant actions and phase structure in lattice gauge theory*, Phys. Rev. **D24** (1981) 3212.

[BCL77] A. Barducci, R. Casalbuoni and L. Lusanna, *Classical scalar and spinning particles interacting with external Yang–Mills field*, Nucl. Phys. **B124** (1977) 93.

[BDI74] R. Balian, J.M. Drouffe and C. Itzykson, *Gauge fields on a lattice. I. General outlook*, Phys. Rev. **D10** (1974) 3376.

[Ber74] C.W. Bernard, *Feynman rules for gauge theories at finite temperature*, Phys. Rev. **D9** (1974) 3312.

[Bra80] N. Bralić, *Exact computation of loop averages in two-dimensional Yang–Mills theory*, Phys. Rev. **D22** (1980) 3090.

[BS92] G.S. Bali and K. Schilling, *The running coupling from SU(3) gauge theory*, in Lattice 92, Proc. of the Int. Symp. on lattice field theory, eds. J. Smit and P. Van Baal, Nucl. Phys. B (Proc. Suppl.) **30**, 1993.

[BSSW77] A.P. Balachandran, P. Salomonson, B. Skagerstam and J. Winnberg, *Classical description of a particle interacting with a non-Abelian gauge field*, Phys. Rev. **D15** (1977) 2308.

[CGL81] P. Cvitanovich, J. Greensite and B. Lautrup, *The crossover point in lattice gauge theories with continuous gauge groups*, Phys. Lett. **105B** (1981) 197.

[CGS86] J. Cleymans, R.V. Gavai and E. Suhonen, *Quarks and gluons at high temperatures and densities*, Phys. Rep. **130** (1986) 217.

[CJR79] 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.

[CJR83] M. Creutz, L. Jacobs and C. Rebbi, *Monte Carlo computations in lattice gauge theories*, Phys. Rep. **95** (1983) 201.

[Cre79] 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.

[Cre80] M. Creutz, *Asymptotic-freedom scales*, Phys. Rev. Lett. **45** (1980) 313.

[Cre83] M. Creutz, *Quarks gluons and lattices*, Cambridge Univ. Press, 1983 (Russian transl.: Mir., M., 1987).

[Dro81] J.M. Drouffe, *The mean-field approximation in the SU(2) pure lattice gauge theory*, Phys. Lett. **105B** (1981) 46.

[DZ83] J.-M. Drouffe and J.-B. Zuber, *Strong coupling and mean field methods in lattice gauge theories*, Phys. Rep. **102** (1983) 1.

[Eli75] S. Elitzur, *Impossibility of spontaneously breaking local symmetries*, Phys. Rev. **D12** (1975) 3978.

[Fey53] R.P. Feynman, *Atomic theory of the λ transition in Helium*, Phys. Rev. **91** (1953) 1291.

[FLZ82] H. Flyvbjerg, B. Lautrup and J.B. Zuber, *Mean field with corrections in lattice gauge theory*, Phys. Lett. **110B** (1982) 279.

[GL81] J. Greensite and B. Lautrup, *Phase transition and mean-field methods in lattice gauge theories*, Phys. Lett. **104B** (1981) 41.

[GN80] J.L. Gervais and A. Neveu, *The slope of the leading Regge trajectory in quantum chromodynamics*, Nucl. Phys. **B163** (1980) 189.

[Gri78] V.N. Gribov, *Quantization of non-Abelian gauge theories*, Nucl. Phys. **B139** (1978) 1.

[Ham83] H.W. Hamber, *Lattice gauge theories at large N_F*, Phys. Lett. **126B** (1983) 471.

[HJS77] M.B. Halpern, A. Jevicki and P. Senjanović, *Field theories in terms of particle-string variables: spin, internal symmetries, and arbitrary dimension*, Phys. Rev. **D16** (1977) 2476.

[HP81] H. Hamber and G. Parisi, *Numerical estimates of hadron masses in a pure SU(3) gauge theory*, Phys. Rev. Lett. **47** (1981) 1792.

[Kad76] L.P. Kadanoff, *Recursion relations in statistical physics and field theory*, Ann. Phys. **100** (1976) 359.

[Kap89] J.I. Kapusta, *Finite-temperature field theory*, Cambridge Univ. Press, 1989.

[Kob79] D.H. Kobe, *Aharonov–Bohm effect revisited*, Ann. Phys. **123** (1979) 381.

[KM81] S.B. Khokhlachev and Yu.M. Makeenko, *Phase transition over the gauge group center and quark confinement in QCD*, Phys. Lett. **101B** (1981) 403.

[KMNP83] H. Kluberg-Stern, A. Morel, O. Napoly and B. Petersson, *Flavors of Lagrangian Susskind fermions*, Nucl. Phys. **B220** (1983) 447.

[KS75] J. Kogut and L. Susskind, *Hamiltonian formulation of Wilson's lattice gauge theories*, Phys. Rev. **D11** (1975) 395.

[KS81a] L.H. Karsten and J. Smit, *Lattice fermions: species doubling, chiral invariance and the triangle anomaly*, Nucl. Phys. **B183** (1981) 103.

[KS81b] N. Kawamoto and J. Smit, *Effective Lagrangian and dynamical symmetry breaking in strongly coupled lattice QCD*, Nucl. Phys. **B192** (1981) 100.

[Lat94] Lattice 94, Proc. of the XIIth Int. Symp. on lattice field theory, ed. F. Karsch *et al.*, Nucl. Phys. B (Proc. Suppl.) **42**, 1995.

[LN80] B. Lautrup and M. Nauenberg, *Phase transition in four-dimensional compact QED*, Phys. Lett. **95B** (1980) 63.

[Lon27] F. London, *Quantenmechanische Deutung der Theorie von Weyl*, Z. Phys. **42** (1927) 375.

[LR82] C.B. Lang and C. Rebbi, *Potential and restoration of rotational symmetry in SU(2) lattice gauge theory*, Phys. Lett. **115B** (1982) 137.

[Mak84] 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).

[Mig75] A.A. Migdal, *Recursion equations in gauge theories*, ZhETF **69** (1975) 810 (English transl.: Sov. Phys. JETP **42** (1975) 413).

[MM94] I. Montvay and G. Münster, *Quantum fields on a lattice*, Cambridge Univ. Press, 1994.

[NN81] 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.

[Pol78] A.M. Polyakov, *Thermal properties of gauge fields and quark liberation*, Phys. Lett. **72B** (1978) 477.

[Sei82] E. Seiler, *Gauge theories as a problem of constructive quantum field theory and statistical mechanics*, Lect. Notes in Physics 159, Springer-Verlag, 1982 (Russian translation: Mir., M., 1985).

[STW81] H.S. Sharatchandra, H.J. Thun and P. Weisz, *Susskind fermions on a Euclidean lattice*, Nucl. Phys. **B192** (1981) 205.

[Sus77] L. Susskind, *Lattice fermions*, Phys. Rev. **D16** (1977) 3031.

[Sus79] L. Susskind, *Lattice models of quark confinement at high temperature*, Phys. Rev. **D20** (1979) 2610.

[Sve86] B. Svetitsky, *Symmetry aspects of finite-temperature confinement transitions*, Phys. Rep. **132** (1986) 1.

[Weg71] F.J. Wegner, *Duality in generalized Ising models and phase transitions without local order parameter*, J. Math. Phys. **12** (1971) 2259.

[Wey19] H. Weyl, *Eine neue Erweiterung der Relativitätstheorie*, Ann. der Phys. **59** (1919) v.10, pp.101–133.

[Wil74] K.G. Wilson, *Confinement of quarks*, Phys. Rev. **D10** (1974) 2445.

[Wil75] 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.

[Wil77] 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.

[WK74] K.G. Wilson and J. Kogut, *The renormalization group and the ϵ-expansion*, Phys. Rep. **12C** (1974) 75 (Russian transl.: Mir., M., 1975).

[Won70] S.K. Wong, *Field and particle equations for the classical Yang–Mills field and particles with isotopic spin*, Nuovo Cim. **65A** (1970) 689.

[YM54] C.N. Yang and R.L. Mills, *Conservation of isotopic spin and isotopic gauge invariance*, Phys. Rev. **96** (1954) 191.

"Can I schedule parades and then call them off?"
"But just send out announcements *postponing* the parades. Don't even bother to schedule them."

J. Heller, *Catch-22*

Chapter 3

$1/N$ Expansion

In many physical problems, especially when fluctuations of scales of different orders of magnitude are essential, 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 [Hoo74a] 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, 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 shall start this Chapter by showing how the $1/N$-expansion works for the O(N)-vector models, and describing some applications to the four-Fermi interaction, the φ^4 theory and the nonlinear sigma model. Then we shall concentrate on multicolor QCD.

3.1 O(N) 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 [Sta68] in statistical mechanics and is known as the spherical model. The extension to quantum field theory was done by Wilson [Wil73] 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.

3.1.1 Four-Fermi theory

The action of the O(N)-symmetric four-Fermi theory in a d-dimensional Euclidean space[1] is defined by

$$S\left[\bar{\psi},\psi\right] \;=\; \int d^d x \left(\bar{\psi}\,\hat{\partial}\,\psi + m\,\bar{\psi}\psi - \frac{G}{2}\left(\bar{\psi}\psi\right)^2\right). \tag{3.1.1}$$

[1] In $d = 2$ this model was studied in the large-N limit in Ref. [GN74] and is often called the Gross–Neveu model.

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

$$\psi \;=\; (\psi_1, \ldots, \psi_N) \tag{3.1.2}$$

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

$$\bar{\psi}\psi \;=\; \sum_{i=1}^{N} \bar{\psi}_i \psi_i \,. \tag{3.1.3}$$

The dimension of the four-Fermi coupling constant G is

$$\dim\left[G\right] \;=\; m^{2-d}. \tag{3.1.4}$$

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 (3.1.1) can be equivalently rewritten as

$$S\left[\bar{\psi},\psi,\chi\right] \;=\; \int d^d x \left(\bar{\psi}\,\hat{\partial}\,\psi + m\,\bar{\psi}\psi - \chi\,\bar{\psi}\psi + \frac{\chi^2}{2G}\right), \tag{3.1.5}$$

where χ is an auxiliary field. The two forms of the action, (3.1.1) and (3.1.5), are equivalent due to the equation of motion which reads in the operator notation as

$$\boldsymbol{\chi} \;=\; G:\bar{\boldsymbol{\psi}}\boldsymbol{\psi}:, \tag{3.1.6}$$

where $:\ldots:$ stands for the normal ordering of operators. Equation (3.1.6) can be derived by varying the action (3.1.5) with respect to χ.

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

$$Z \;=\; \int D\chi D\bar{\psi} D\psi\, \mathrm{e}^{-S[\bar{\psi},\psi,\chi]} \tag{3.1.7}$$

with $S\left[\bar{\psi},\psi,\chi\right]$ given by Eq. (3.1.5), the action (3.1.1) appears after performing the Gaussian integral over χ. Therefore, one alternatively gets

$$Z \;=\; \int D\bar{\psi} D\psi\, \mathrm{e}^{-S[\bar{\psi},\psi]} \tag{3.1.8}$$

with $S[\bar{\psi},\psi]$ given by Eq. (3.1.1).

The perturbative expansion of the O(N)-symmetric four-Fermi theory can be conveniently represented using the formulation (3.1.5) 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 (3.1.5) that

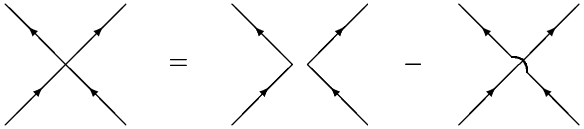
$$D_0\,(x-y) \equiv \langle\, \chi(x)\chi(y)\,\rangle_{\mathrm{Gauss}} \;=\; G\,\delta^{(d)}(x-y) \tag{3.1.9}$$

or

$$D_0\,(p) \equiv \langle\, \chi(-p)\chi(p)\,\rangle_{\mathrm{Gauss}} \;=\; G \tag{3.1.10}$$

in momentum space.

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



(3.1.11)

where the empty space inside the vertex is associated with the propagator (3.1.9) (or (3.1.10) 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. 3.1. The O(N) indices propagate through the solid lines so that the closed line in the diagram in Fig. 3.1b 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. 3.2.

Problem 3.1 Calculate the one-loop Gell-Mann–Low function of the four-Fermi theory in $d = 2$.

Solution Evaluating the diagrams in Fig. 3.1 which are logarithmically divergent in $d = 2$, and noting that the diagrams in Fig. 3.2 do not contribute to the wave-function renormalization of the ψ-field, which emerges to the next order in G, one gets

$$\mathcal{B}\,(G) \;=\; -\,\frac{(N-1)\,G^2}{2\pi}\,. \tag{3.1.12}$$

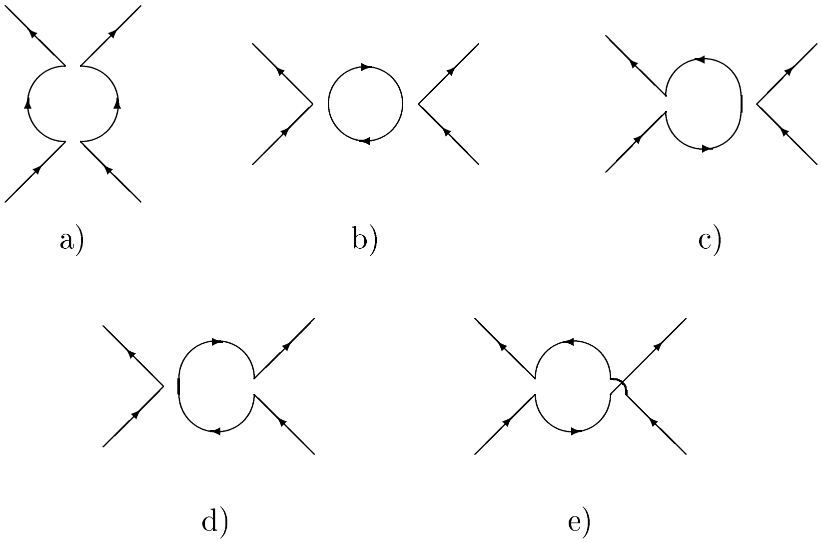


Fig. 3.1: Diagrams of the second order of perturbation theory for the four-Fermi vertex. The diagram b) involves the sum over the O(N) indices.

The four-Fermi theory in 2 dimensions is asymptotically free as was first noted by Anselm [Ans59] and rediscovered in Ref. [GN74].

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 $\left(\bar{\psi}\gamma_\mu\psi\right)^2$ 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 (3.1.1) 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 [Joh61] to all loops.

Remark on auxiliary fields

An introduction of the auxiliary field is often called, in English scientific terminology, the Hubbard–Stratonovich transformation in analogy with statistical mechanics. In spite of the original paper by Stratonovich [Str57] being published in a Russian journal, the proper Russian term is just "auxiliary

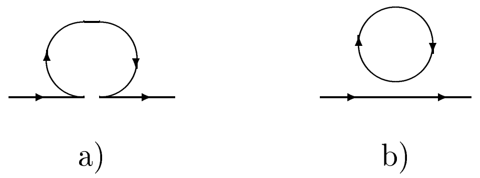


Fig. 3.2: One-loop diagrams for the propagator of the ψ-field. The diagram b) involves the sum over the O(N) indices.

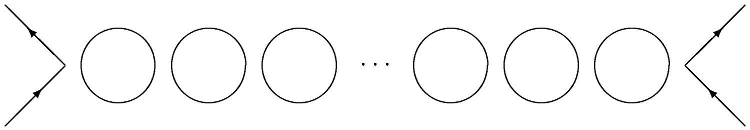


Fig. 3.3: Bubble diagram which survives the large-N limit of the O(N) vector models.

field".

3.1.2 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. 3.3 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

$$G^{n+1}N^n \sim G \tag{3.1.13}$$

as $N \to \infty$ since

$$G \sim \frac{1}{N} . \tag{3.1.14}$$

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

Let us denote

$$\sim\!\sim\!\sim\!\sim \;=\; G + \ldots + G^2 \bigcirc + G^{n+1} \overbrace{\bigcirc \cdots \bigcirc}^{n \text{ loops}} + \ldots \tag{3.1.15}$$

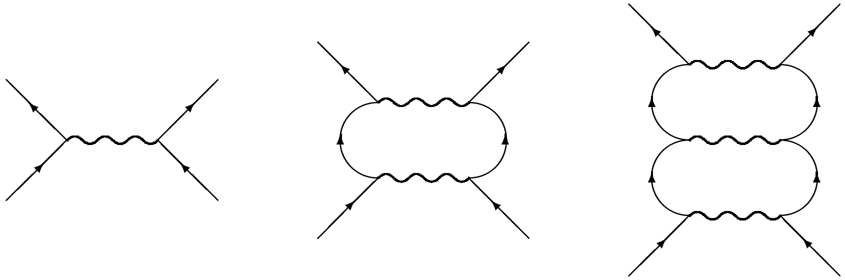


Fig. 3.4: Some diagrams of the $1/N$-expansion for the O(N) four-Fermi theory. The wavy line represents the (infinite) sum of the bubble graphs (3.1.15).

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. (3.1.15) is nothing but the free propagator (3.1.10).

Summing the geometric series of the fermion-loop chains on the RHS of Eq. (3.1.15), one gets analytically[2]

$$D^{-1}(p) \;=\; \frac{1}{G} - N \int \frac{d^d k}{(2\pi)^d} \frac{\mathrm{sp}\left(\hat{k} + im\right)\left(\hat{k} + \hat{p} + im\right)}{(k^2+m^2)\left((k+p)^2+m^2\right)} . \tag{3.1.16}$$

This determines the exact propagator of the χ field at large N. It is $\mathcal{O}\left(N^{-1}\right)$ 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 (3.1.15) which is given by Eq. (3.1.16). Some of the diagrams of the new expansion for the four-Fermi vertex are depicted in Fig. 3.4. 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.

[2]Recall that the free Euclidean fermionic propagator is given by $S_0(p) = (i\hat{p}+m)^{-1}$ due to Eqs. (3.1.5), (3.1.7) and the additional minus sign is associated with the fermion loop.

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 [Wil73] an existence of an ultraviolet-stable fixed point, *i.e.* a nontrivial zero of the Gell-Mann–Low function.

Problem 3.2 Show that the $1/N$-expansion of the four-Fermi theory is renormalizable in $2 \leq d < 4$ (but not in $d = 4$).
Solution In order to demonstrate renormalizability, let us analyze indices of the diagrams of the $1/N$-expansion.

First of all, we shall get rid of an ultraviolet divergence of the integral over the d-momentum k in Eq. (3.1.16). 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

$$G \;=\; \frac{g^2}{N}\,\Lambda^{2-d}\,, \tag{3.1.17}$$

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. (3.1.17) 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

$$D(p) \;\propto\; \frac{1}{|p|^{d-2}}\,. \tag{3.1.18}$$

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 $\bar\psi$-χ-ψ three-vertex. These divergencies can be removed by a renormalization of the coupling g, mass, and wave functions of ψ and χ.

This completes a demonstration of renormalizability of the $1/N$-expansion for the four-Fermi interaction in $2 \leq d < 4$. For more detail, see Ref. [Par75].

Problem 3.3 Calculate in $d = 3$ the value of g in Eq. (3.1.17).
Solution To calculate the divergent part of the integral in Eq. (3.1.16), we put $p = 0$ and $m = 0$. Remembering that the γ-matrices are 2×2 matrices in $d = 3$, we get

$$\int^{\Lambda} \frac{d^3k}{(2\pi)^3}\,\frac{\mathrm{sp}\,\hat k \hat k}{k^2 k^2} \;=\; 2\int^{\Lambda} \frac{d^3k}{(2\pi)^3}\,\frac{1}{k^2} \;=\; \frac{1}{\pi^2}\int^{\Lambda} d|k| \;=\; \frac{\Lambda}{\pi^2}\,. \tag{3.1.19}$$

Note that the integral is linearly divergent in $d = 3$ and Λ is the cutoff for the integration over $|k|$. This divergence can be cancelled by choosing G according to Eq. (3.1.17) with g equal to

$$g_* \;=\; \pi\,. \tag{3.1.20}$$

Problem 3.4 Calculate in $d = 3$ the coefficient of proportionality in Eq. (3.1.18).
Solution Let us choose $G = \pi^2/N\Lambda$ as is prescribed by Eqs. (3.1.17), (3.1.20) and put in Eq. (3.1.16) $m = 0$ since we are interested in the asymptotics $|p| \gg m$. Then the RHS of Eq. (3.1.16) can be rearranged as

$$\begin{aligned} D^{-1}(p) &= -2N\int \frac{d^3k}{(2\pi)^3}\left[\frac{k^2+kp}{k^2(k+p)^2} - \frac{1}{k^2}\right] \\ &= 2N\int \frac{d^3k}{(2\pi)^3}\,\frac{p^2+kp}{k^2(k+p)^2}\,. \end{aligned} \tag{3.1.21}$$

This integral is obviously convergent.

To calculate it, we apply the standard technique of the α-parametrization, which is based on the formula

$$\frac{1}{k^2} \;=\; \int_0^\infty d\alpha\,\mathrm{e}^{-\alpha k^2}\,. \tag{3.1.22}$$

We have

$$\int \frac{d^3k}{(2\pi)^3}\,\frac{p^2+kp}{k^2(k+p)^2} \;=\; \int_0^\infty d\alpha_1 \int_0^\infty d\alpha_2 \int \frac{d^3k}{(2\pi)^3}\left(p^2+kp\right)\mathrm{e}^{-\alpha_1 k^2 - \alpha_2 (k+p)^2} \tag{3.1.23}$$

after which the Gaussian integral over d^3k can easily be performed. We get then

$$D^{-1}(p) \;=\; \frac{N}{4\pi^{3/2}}\,p^2 \int_0^\infty d\alpha_1 \int_0^\infty d\alpha_2\,\frac{\alpha_1}{(\alpha_1+\alpha_2)^{5/2}}\,\mathrm{e}^{-\frac{\alpha_1\alpha_2}{\alpha_1+\alpha_2}p^2}\,. \tag{3.1.24}$$

The remaining integration over α_1 and α_2 can easily be done introducing the new variables $\alpha \in [0,\infty]$ and $x \in [0,1]$ so that

$$\begin{aligned} \alpha_1 \;=\; \alpha x\,, \qquad & \alpha_2 \;=\; \alpha(1-x)\,, \\ \frac{\partial\,(\alpha_1,\alpha_2)}{\partial\,(x,\alpha)} \;&=\; \alpha\,. \end{aligned} \tag{3.1.25}$$

This gives finally

$$D(p) \;=\; \frac{8}{N\,|p|}\,. \tag{3.1.26}$$

a) b)

Fig. 3.5: Diagrams for the $1/N$-correction to the ψ-field propagator (a) and the three-vertex (b).

Equation (3.1.26) (or (3.1.18) in d dimensions) is remarkable since it shows that the scale dimension of the field χ, which is defined in Subsect. 1.3.5 by Eq. (1.3.66), changes its value from $l_\chi = d/2$ in perturbation theory to $l_\chi = 1$ in the zeroth order of the $1/N$ expansion (remember that the momentum-space propagator of a field with the scale dimension l is proportional to $|p|^{2l-d}$). This appearance of scale invariance in the $1/N$-expansion of the four-Fermi theory at $2 < d < 4$ was first pointed out by Wilson [Wil73] and implies that the Gell-Mann–Low function $\mathcal{B}(g)$ has a zero at $g = g_*$ which is given in $d = 3$ by Eq. (3.1.20).

Problem 3.5 Find the (logarithmic) anomalous dimensions of the fields ψ, χ, and of the $\bar\psi$-χ-ψ three-vertex in $d = 3$ to order $1/N$.
Solution The $1/N$-correction to the propagator of the ψ-field is given by the diagram depicted in Fig. 3.5a). Since we are interested in an ultraviolet behavior, we can put again $m = 0$. Analytically, we have

$$S^{-1}(p) \;=\; i\hat p + \frac{8i}{N}\int^\Lambda \frac{d^3k}{(2\pi)^3}\frac{\hat k + \hat p}{|k|(k+p)^2}\,. \tag{3.1.27}$$

The (logarithmically) divergent contribution emerges from the domain of integration $|k| \gg |p|$ so we can expand the integrand in p. The p-independent term vanishes after integration over the directions of k so that we get

$$\begin{aligned} S^{-1}(p) \;&=\; i\hat p\left[1 + \frac{8}{N}\int^\Lambda \frac{d^3k}{(2\pi)^3}\frac{1}{|k|^3}\right] \\ &=\; i\hat p\left[1 + \frac{2}{3\pi^2 N}\ln\frac{\Lambda^2}{p^2} + \text{finite}\right]. \end{aligned} \tag{3.1.28}$$

The diagram, which gives a non-vanishing contribution to the three-vertex in order $1/N$, is depicted in Fig. 3.5b. It reads analytically

$$\Gamma(p_1,p_2) \;=\; 1 + \frac{8}{N}\int^\Lambda \frac{d^3k}{(2\pi)^3}\frac{(\hat k + \hat p_1)(\hat k + \hat p_2)}{|k|(k+p_1)^2(k+p_2)^2}\,, \tag{3.1.29}$$

where p_1 and p_2 the incoming and outgoing fermion momenta, respectively. The logarithmic domain is $|k| \gg |p|_{\max}$ with $|p|_{\max}$ being the largest of $|p_1|$ and $|p_2|$. This gives

$$\Gamma(p_1,p_2) \;=\; 1 - \frac{2}{\pi^2 N}\ln\frac{\Lambda^2}{p_{\max}^2} + \text{finite}\,. \tag{3.1.30}$$

An analogous calculation of the $1/N$ correction for the field χ is a bit more complicated since involves three two-loop diagrams (see Ref. [CMS93]). The resulting expression for $D^{-1}(p)$ reads

$$\Big(N\,D(p)\Big)^{-1} \;=\; \frac{\Lambda}{g^2} + \left[-\frac{\Lambda}{\pi^2} + \frac{|p|}{8}\right] + \frac{1}{\pi^2 N}\left[2\Lambda - |p|\left(\frac{2}{3}\ln\frac{\Lambda^2}{p^2} + \text{finite}\right)\right].\tag{3.1.31}$$

The linear divergence is cancelled to order $1/N$ providing g is equal to

$$g_* \;=\; \pi\left(1 + \frac{1}{N}\right), \tag{3.1.32}$$

which determines g_* to order $1/N$. After this $D^{-1}(p)$ takes the form

$$D^{-1}(p) \;=\; \frac{N\,|p|}{8}\left(1 - \frac{16}{3\pi^2 N}\ln\frac{\Lambda^2}{p^2}\right). \tag{3.1.33}$$

To make all three expressions (3.1.28), (3.1.30), and (3.1.33) finite, we need logarithmic renormalizations of the wave functions of ψ- and χ-fields and of the vertex Γ. This can be achieved by multiplying them by the renormalization constants

$$Z_i(\Lambda) \;=\; 1 - \gamma_i \ln\frac{\Lambda^2}{\mu^2} \tag{3.1.34}$$

where μ stands for a reference mass scale and γ_i are anomalous dimensions. The index i stands for ψ, χ, or v for the ψ- and χ-propagators or the three-vertex Γ, respectively. We have, therefore, calculated

$$\begin{aligned} \gamma_\psi \;&=\; \frac{2}{3\pi^2 N}\,, \\ \gamma_{\rm v} \;&=\; -\frac{2}{\pi^2 N}\,, \\ \gamma_\chi \;&=\; -\frac{16}{3\pi^2 N} \end{aligned} \tag{3.1.35}$$

to order $1/N$. Due to Eq. (3.1.6) γ_χ coincides with the anomalous dimension of the composite fields $\bar\psi\psi$

$$\gamma_{\bar\psi\psi} \;=\; \gamma_\chi\,. \tag{3.1.36}$$

Note, that

$$Z_\psi^2 Z_{\rm v}^{-2} Z_\chi \;=\; 1\,. \tag{3.1.37}$$

This implies that the effective charge is not renormalized and is given by Eq. (3.1.32). Thus, the nontrivial zero of the Gell-Mann–Low function persists to order $1/N$ (and, in fact, to all orders of the $1/N$-expansion).

Remark on scale invariance at fixed point

The renormalization group says that

$$\mu \;=\; \Lambda \exp\left[-\int \frac{dg^2}{\mathcal{B}(g^2)}\right], \tag{3.1.38}$$

which is the same as Eq. (2.2.86) since the correlation length $\sim \mu$. If $\mathcal{B}$ has a nontrivial fixed point g_*^2 near which

$$\mathcal{B}(g^2) \;=\; b\left(g^2-g_*^2\right) \tag{3.1.39}$$

with $b<0$, then the substitution into Eq. (3.1.38) gives

$$g^2 \;=\; g_*^2+\left(\frac{\mu}{\Lambda}\right)^{-b}. \tag{3.1.40}$$

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

$$\mathcal{B}(g^2) \;=\; bg^4\,. \tag{3.1.41}$$

The latter bahavior of $\mathcal{B}$ results, after the substitution into Eq. (3.1.38), in the logarithmic dependence

$$g^2 \;=\frac{1}{b\ln\frac{\mu}{\Lambda}} \tag{3.1.42}$$

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

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

$$\frac{\mu\, d\ln\Gamma_i}{d\mu} \;=\; \gamma_i\left(g^2\right)\,, \tag{3.1.43}$$

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

$$\Gamma_i \;\propto\; \mu^{\gamma_i(g_*^2)}\,. \tag{3.1.44}$$

This complements the heuristic consideration of Subsect. 1.3.5 on the relation between scale invariance and the vanishing of the Gell-Mann–Low function.

For the four-Fermi theory in $d=3$, Eq. (3.1.44) yields

$$S(p) \;=\; \frac{1}{i\hat p}\left(\frac{p^2}{\mu^2}\right)^{\gamma_\psi}, \tag{3.1.45}$$

$$D(p) \;=\; \frac{8}{N\,|p|}\left(\frac{p^2}{\mu^2}\right)^{\gamma_\chi}, \tag{3.1.46}$$

$$\Gamma(p_1,p_2) \;=\; \left(\frac{\mu^2}{p_1^2}\right)^{\gamma_{\rm v}} f\left(\frac{p_2^2}{p_1^2},\frac{p_1p_2}{p_1^2}\right), \tag{3.1.47}$$

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

$$\gamma_{\rm v} = \gamma_\psi + \frac{1}{2}\gamma_\chi \tag{3.1.48}$$

which guarantees that Eq. (3.1.37), implied by scale invariance, is satisfied.

The indices γ_i are given to order $1/N$ by Eqs. (3.1.35). When expanded in $1/N$, Eqs. (3.1.45) and (3.1.46) obviously reproduces Eqs. (3.1.28) and (3.1.33). Therefore, one gets the exponentiation of the logarithms which emerge in the $1/N$-expansion. The calculation of the next terms of the $1/N$-expansion for the indices γ_i is contained in Ref. [Gra91].

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. [MS69, GW70]. The conformal group in a d-dimensional space-time has $(d+1)(d+2)/2$ parameters as is illustrated by Table 3.1. More about the conformal group can be found in the lecture by Jackiw [Jac72].

A heuristic proof [MS69] of the fact that scale invariance implies conformal invariance is based on the explicit form of the conformal current K_μ^α, which is associated with the special conformal transformation, via the energy-momentum tensor:

$$K_\mu^\alpha \;=\; \left(2x_\nu x^\alpha - x^2\delta_\nu^\alpha\right)\theta_{\mu\nu}\,. \tag{3.1.49}$$

Group	Transformations		# Parameters
Lorentz	$\frac{d(d-1)}{2}$ rotations	$x'_\mu = \Omega_{\mu\nu} x_\nu$	$\frac{d(d-1)}{2}$
Poincaré	+ d translations	$x'_\mu = x_\mu + a_\mu$	$\frac{d(d+1)}{2}$
Weyl	+ 1 dilatation	$x'_\mu = \rho\, x_\mu$	$\frac{d^2+d+2}{2}$
Conformal	+ d special conformal	$\frac{x'_\mu}{(x')^2} = \frac{x_\mu}{x^2} + \alpha_\mu$	$\frac{(d+1)(d+2)}{2}$

Table 3.1: Contents and the number of parameters of groups of space-time symmetry.

Differentiating, we get

$$\partial_\mu K^\alpha_\mu \;=\; 2x^\alpha \theta_{\mu\mu}\,, \tag{3.1.50}$$

which is analogous to Eqs. (1.3.67) and (1.3.68) for the dilatation current. 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 [Pol70] for scalar theories. The proper formula for the four-Fermi theory reads [Mig71]

$$\begin{aligned}\Gamma(p_1,p_2) \;&=\; \mu^{2\gamma_{\rm v}}\frac{\Gamma(\frac{d}{2})\Gamma(\frac{d}{2}-\gamma_{\rm v})}{\Gamma(\gamma_{\rm v})} \\ &\times \int \frac{d^dk}{\pi^{d/2}}\frac{\hat{k}+\hat{p}_1}{[(k+p_1)^2]^{1+\gamma_\chi/2}}\frac{\hat{k}+\hat{p}_2}{[(k+p_2)^2]^{1+\gamma_\chi/2}}\frac{1}{|k|^{d-2+2\gamma_\psi-\gamma_\chi/2}}\,,\end{aligned} \tag{3.1.51}$$

where the coefficient in the form of the ratio of the Γ-functions is prescribed by the normalization (3.1.45) and (3.1.46) and the indices are related by Eq. (3.1.48) but can be arbitrary otherwise[3].

[3]The only restriction $\gamma_\psi \geq 0$ is imposed by the Källén–Lehmann representation of the propagator while there is no such restriction on γ_χ since it is a composite field.

Equation (3.1.51), which results from conformal invariance, unambiguously fixes the function f in Eq. (3.1.47). 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.

Problem 3.6 Calculate the integral on the RHS of Eq. (3.1.51) in $d = 3$ to order $1/N$.

Solution The integral on the RHS of Eq. (3.1.51) looks in $d = 3$ very much like that in Eq. (3.1.29) and can easily be calculated to the leading order in $1/N$ when only the region of integration over large momenta with $|k| \gtrsim |p|_{\max} \equiv \max\{|p_1|,|p_2|\}$ is essential to this accuracy.

Let us first note that the coefficient in front of the integral is $\propto \gamma_{\rm v} \sim 1/N$, so that one has to peak up the term $\sim 1/\gamma_{\rm v}$ in the integral for the vertex to be of order 1. This term comes from the region of integration with $|k| \gtrsim |p|_{\max}$. Recalling that $|p_1 - p_2| \lesssim |p|_{\max}$ in Euclidean space, one gets

$$\begin{aligned}&\int \frac{d^3k}{2\pi}\frac{\hat{k}+\hat{p}_1}{[(k+p_1)^2]^{1+\gamma_\chi/2}}\frac{\hat{k}+\hat{p}_2}{[(k+p_2)^2]^{1+\gamma_\chi/2}}\frac{1}{|k|^{1+2\gamma_\psi-\gamma_\chi/2}} \\ &\quad= \int_{p^2_{\max}}^{\infty}\frac{dk^2}{[k^2]^{1+\gamma_{\rm v}}} \;=\; \frac{1}{\gamma_{\rm v}\,(p^2_{\max})^{\gamma_{\rm v}}}\,,\end{aligned} \tag{3.1.52}$$

where Eq. (3.1.48) has been used and

$$\Gamma(p_1,p_2) \;=\; \left(\frac{\mu^2}{p^2_{\max}}\right)^{\gamma_{\rm v}}. \tag{3.1.53}$$

While the integral in Eq. (3.1.52) is divergent in the ultraviolet for $\gamma_{\rm v} < 0$, this divergence disappears after the renormalization.

Equation (3.1.30) is reproduced by Eq. (3.1.52) when expanding in $1/N$. This dependence of the three-vertex solely on the largest momentum is typical for logarithmic theories in the ultraviolet region where one can put, say, $p_1 = 0$ without changing the integral with logarithmic accuracy. This is valid if the integral is fast convergent in infrared regions which is our case.

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. (3.1.40), then scale invariance holds even in the massless case only for $|p| \gg \mu$ while it is broken if $|p| \lesssim \mu$.

3.1.3 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

$$S[\varphi^a] = \int d^d x \left[\frac{1}{2}(\partial_\mu \varphi^a)^2 + \frac{1}{2}m^2\varphi^a\varphi^a + \frac{\lambda}{8}(\varphi^a\varphi^a)^2\right] \tag{3.1.54}$$

where

$$\varphi^a = \left(\varphi^1, \ldots, \varphi^N\right). \tag{3.1.55}$$

The coupling λ in the action (3.1.54) must be positive for the theory to be well defined. The vertices of Feynman diagrams are associated with $-\lambda$.

Problem 3.7 Calculate the one-loop Gell-Mann–Low function of the O(N)-symmetric φ^4 theory in $d = 4$.
Solution The corresponding diagrams are similar to those of Fig. 3.1, 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. 3.1b has now an extra combinatoric factor of 1/2. The diagrams of Fig. 3.2 result in a mass renormalization and there is no wave-function renormalization of the φ-field in one loop so that one gets

$$\mathcal{B}(\lambda) = \frac{(N+8)\lambda^2}{16\pi^2}. \tag{3.1.56}$$

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 $N = 1$.

Introducing the auxiliary field $\chi(x)$ as in Subsect. 3.1.1, the action (3.1.54) can be rewritten as

$$S[\varphi^a, \chi] = \int d^d x \left[\frac{1}{2}\varphi^a\left(-\partial_\mu^2 + m^2 + \chi\right)\varphi^a - \frac{\chi^2}{2\lambda}\right]. \tag{3.1.57}$$

The two forms are equivalent due to the equation of motion

$$\chi = \frac{\lambda}{2} :\varphi^a\varphi^a: . \tag{3.1.58}$$

In other words χ is again a composite field.

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

$$Z[J^a, K] = \int_\uparrow D\chi(x) \int D\varphi^a(x) \times \mathrm{e}^{-S[\varphi^a,\chi] + \int d^dx J^a(x)\varphi^a(x) + \int d^dx K(x)\chi(x)}, \tag{3.1.59}$$

which is a functional of the sources J^a and K for the fields φ^a and χ and extends Eq. (1.2.49).

To make the path integral over $\chi(x)$ in Eq. (3.1.59) 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

$$D_0(p) = \langle\, \chi(-p)\chi(p)\,\rangle_{\mathrm{Gauss}} = -\lambda, \tag{3.1.60}$$

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

$$G(x, y; \chi) = \left\langle y \left| \frac{1}{-\partial_\mu^2 + m^2 + \chi} \right| x \right\rangle \tag{3.1.61}$$

as

$$Z[J^a, K] = \int_\uparrow D\chi(x)\, \mathrm{e}^{\int d^dx \frac{\chi^2}{2\lambda} + \frac{1}{2}\int d^dx d^dy J^a(x) G(x,y;\chi) J^a(y) + \int d^dx K(x)\chi(x)} \times \mathrm{e}^{-\frac{N}{2}\,\mathrm{Sp}\ln G^{-1}[\chi]}. \tag{3.1.62}$$

We have used here the obvious notation

$$G^{-1}[\chi] = -\partial_\mu^2 + m^2 + \chi. \tag{3.1.63}$$

It will also be convenient to use the short-hand notation

$$g \circ f = \langle g | f \rangle \equiv \int d^dx\, f(x) g(x). \tag{3.1.64}$$

Then, Eq. (3.1.62) can be rewritten as

$$Z[J^a, K] = \int_\uparrow D\chi(x)\, \mathrm{e}^{\frac{\chi\circ\chi}{2\lambda} + \frac{1}{2}J^a\circ G[\chi]\circ J^a + K\circ\chi - \frac{N}{2}\,\mathrm{Sp}\ln G^{-1}[\chi]}. \tag{3.1.65}$$

The exponent in Eq. (3.1.65) is $\mathcal{O}(N)$ at large N so the path integral can be evaluated as $N \to \infty$ by the saddle-point method. The saddle-point field configuration

$$\chi(x) = \chi_{\rm sp}(x) \tag{3.1.66}$$

is determined (implicitly) by the saddle-point equation

$$\chi_{\rm sp}(x) - \frac{\lambda N}{2} G(x,x;\chi_{\rm sp}) + \frac{\lambda}{2} J^a \circ G(\cdot,x;\chi_{\rm sp})\, G(x,\cdot;\chi_{\rm sp}) \circ J^a + \lambda K(x) = 0\,. \tag{3.1.67}$$

If $K \sim 1/\lambda$, each term here is $\mathcal{O}(1)$ since

$$\lambda \sim \frac{1}{N} \tag{3.1.68}$$

in analogy with Eq. (3.1.14).

When the sources J^a and K vanish so that the last two terms on the LHS of Eq. (3.1.67) equal zero, this equation reduces to

$$\chi_{\rm sp} - \frac{\lambda N}{2} G(x,x;\chi_{\rm sp}) = 0\,. \tag{3.1.69}$$

Its solution is x independent due to translational invariance and can be parametrized as

$$\chi_{\rm sp} = m_{\rm R}^2 - m^2 \tag{3.1.70}$$

where m and $m_{\rm R}$ are the bare and renormalized mass, respectively. Equation (3.1.69) then reduces to the standard formula [Wil73]

$$m^2 = m_{\rm R}^2 - \frac{\lambda N}{2} \int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2 + m_{\rm R}^2)} \tag{3.1.71}$$

for the mass renormalization at large N.

To take into account fluctuations around the saddle point, we expand

$$\chi(x) = \chi_{\rm sp} + \delta\chi(x) \tag{3.1.72}$$

where

$$\delta\chi(x) \sim \sqrt{\lambda} \sim N^{-1/2}\,. \tag{3.1.73}$$

The Gaussian integration over $\delta\chi(x)$ determines the pre-exponential factor in (3.1.65).

To construct the $1/N$ expansion of the generating functional (3.1.65), it is convenient to use the generating functional for connected Green's functions, which was already introduced in Eq. (1.2.52). It is usually denoted by $W[J^a, K]$ and is related to the partition function (3.1.59) by

$$Z[J^a, K] = {\rm e}^{W[J^a,K]}\,. \tag{3.1.74}$$

Then we get

$$W[J^a,K] = \frac{1}{2\lambda}\chi_{\rm sp}\circ\chi_{\rm sp} - \frac{N}{2}\,{\rm Sp}\ln G^{-1}[\chi_{\rm sp}] + \frac{1}{2} J^a \circ G[\chi_{\rm sp}] \circ J^a + K \circ \chi_{\rm sp} - \frac{1}{2}\,{\rm Sp}\ln\left(\lambda D^{-1}[\chi_{\rm sp}]\right) + \mathcal{O}\left(N^{-1}\right), \tag{3.1.75}$$

where

$$D^{-1}(x,y;\chi) = -\frac{1}{\lambda}\delta^{(d)}(x-y) - \frac{N}{2} G(x,y;\chi)\, G(y,x;\chi) + J^a \circ G(\cdot,x;\chi)\, G(x,y;\chi)\, G(y,\cdot;\chi) \circ J^a\,. \tag{3.1.76}$$

This operator emerges when integrating over the Gaussian fluctuations around the saddle point. The corresponding (last) term on the RHS of Eq. (3.1.75) 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 (3.1.72) in Eq. (3.1.65) and performing the perturbative expansion in $\delta\chi$.

If the sources J^a and K vanish so that the saddle-point value $\chi_{\rm sp}$ is given by the constant (3.1.70), than the RHS of Eq. (3.1.76) simplifies to

$$D^{-1}(x,y;\chi_{\rm sp}) = -\frac{1}{\lambda}\delta^{(d)}(x-y) - \frac{N}{2} G(x,y;\chi_{\rm sp})\, G(y,x;\chi_{\rm sp})\,. \tag{3.1.77}$$

Remembering the definition (3.1.61) of G and passing to the momentum-space representation, we get

$$D^{-1}(p) = -\frac{1}{\lambda} - \frac{N}{2}\int \frac{d^d k}{(2\pi)^d} \frac{1}{(k^2+m_{\rm R}^2)\,((k+p)^2+m_{\rm R}^2)}\,. \tag{3.1.78}$$

The sign of the first term on the RHS is consistent with Eq. (3.1.60).

Equation (3.1.78) is an analog of Eq. (3.1.16) in the fermionic case and can be alternatively obtained by summing bubble graphs of the type in Fig. 3.3 for

$$D(p) \;=\; \langle\, \chi(-p)\chi(p)\, \rangle\,. \tag{3.1.79}$$

The extra symmetry factor 1/2 in Eq. (3.1.78) is the usual combinatoric one for bosons. Therefore, the large-N saddle-point calculation of the propagator (3.1.79) results precisely in the zeroth-order of the $1/N$-expansion.

We see from Eq. (3.1.75) the difference between perturbation theory and the $1/N$-expansion. The perturbation theory in λ can be constructed as an expansion (3.1.72) around the saddle point $\chi_{\rm sp}$ given again by Eq. (3.1.67), 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. (3.1.75), which is associated with a one-loop diagram, appears in perturbation theory as a result of Gaussian fluctuations around this saddle-point.

Remark on the effective action

The effective action is a functional of the mean values of fields

$$\varphi^a_{cl}(x) \;=\; \frac{\delta W}{\delta J^a(x)}\,, \qquad \chi_{cl}(x) \;=\; \frac{\delta W}{\delta K(x)} \tag{3.1.80}$$

in the presence of the external sources. The effective action is defined as the Legendre transformation of $W[J^a, K]$ by

$$\Gamma\left[\varphi^a_{cl}, \chi_{cl}\right] \;\equiv\; -W + J^a \circ \varphi^a_{cl} + K \circ \chi_{cl}\,, \tag{3.1.81}$$

where the sources J^a and K, which are regarded as functionals of φ^a_{cl} and χ_{cl}, are to be determined by an inversion of Eq. (3.1.80). To the leading order in $1/N$ we get

$$\begin{aligned} J^a(x) &= G^{-1}\left[\chi_{cl}\right]\varphi^a_{cl}(x) + \mathcal{O}\left(N^{-1}\right)\,, \\ \chi_{cl}(x) &= \chi_{\rm sp}(x) + \mathcal{O}\left(N^{-1}\right)\,. \end{aligned} \tag{3.1.82}$$

When Eq. (3.1.82) (with the $1/N$ correction included) is substituted into Eq. (3.1.81) and account is taken of the $1/N$ terms, most of them cancel and we arrive at the relatively simple formula

$$\begin{aligned} \Gamma\left[\varphi^a_{cl}, \chi_{cl}\right] \;=\; & -\frac{1}{2\lambda}\chi_{cl}\circ\chi_{cl} + \frac{N}{2}\,{\rm Sp}\ln G^{-1}\left[\chi_{cl}\right] \\ & + \frac{1}{2}\varphi^a_{cl}\circ G^{-1}\left[\chi_{cl}\right]\circ\varphi^a_{cl} + \frac{1}{2}\,{\rm Sp}\ln\left(\lambda D^{-1}\left[\chi_{cl}\right]\right) + \mathcal{O}\left(N^{-1}\right)\,, \end{aligned} \tag{3.1.83}$$

where

$$\begin{aligned} D^{-1}\left(x, y; \chi_{cl}\right) \;=\; & -\frac{1}{\lambda}\delta^{(d)}(x-y) - \frac{N}{2}G\left(x, y; \chi_{cl}\right)G\left(y, x; \chi_{cl}\right) \\ & + \varphi^a_{cl}(x)G\left(x, y; \chi_{cl}\right)\varphi^a_{cl}(y) \end{aligned} \tag{3.1.84}$$

coinciding with (3.1.76) to the leading order in $1/N$.

The second and fourth terms on the RHS of Eq. (3.1.83), which involve Sp, are associated with one-loop diagrams of the fields φ^a and χ, respectively, in the classical background fields φ^a_{cl} and χ_{cl}. Higher orders in $1/N$ are given by diagrams which are one-particle irreducible with respect to both φ and χ.

It follows immediately from the definitions (3.1.80) and (3.1.81) that

$$\frac{\delta\Gamma}{\delta\varphi^a_{cl}(x)} \;=\; J^a(x)\,, \qquad \frac{\delta\Gamma}{\delta\chi_{cl}(x)} \;=\; K(x)\,. \tag{3.1.85}$$

Therefore, $\varphi^a_{cl}(x)$ and $\chi_{cl}(x)$ are determined in the absence of external sources by the equations

$$\frac{\delta\Gamma\left[\varphi^a_{cl}, \chi_{cl}\right]}{\delta\varphi^b_{cl}(x)} \;=\; 0 \tag{3.1.86}$$

and

$$\frac{\delta\Gamma\left[\varphi^a_{cl}, \chi_{cl}\right]}{\delta\chi_{cl}(x)} \;=\; 0\,. \tag{3.1.87}$$

Substituting (3.1.83) into Eqs. (3.1.86) and (3.1.87), we get to the leading order in $1/N$, respectively, the equations

$$\left(-\partial^2_\mu + m^2 + \chi_{cl}(x)\right)\varphi^a_{cl}(x) \;=\; 0 \tag{3.1.88}$$

and

$$\chi_{cl}(x) \;=\; \frac{\lambda}{2}\varphi^a_{cl}(x)\varphi^a_{cl}(x) + \frac{\lambda N}{2}G\left(x, x; \chi_{cl}\right)\,. \tag{3.1.89}$$

The first equation is just a classical equation of motion in an external field $\chi_{cl}(x)$ while the second one is just the average of the (quantum) equation (3.1.58). Equation (3.1.89) is often called the gap equation.

A solution to Eqs. (3.1.88) and (3.1.89) depends on what initial (or boundary) conditions are imposed.

Problem 3.8 Find translationally invariant solutions to Eqs. (3.1.88) and (3.1.89) and calculate the corresponding effective potential.

Solution The effective potential $V\left(\varphi^a_{cl},\chi_{cl}\right)$ is defined via the integrand in the effective action $\Gamma\left[\varphi^a_{cl},\chi_{cl}\right]$ for translationally invariant

$$\varphi^a_{cl}(x) \;=\; \bar{\varphi}^a\,, \qquad \chi_{cl} \;=\; \bar{\chi} \tag{3.1.90}$$

by the formula

$$\Gamma \;=\; \text{Vol.}\times V\left(\bar{\varphi}^a,\bar{\chi}\right)\,. \tag{3.1.91}$$

From Eq. (3.1.83) we get at large N

$$V \;=\; -\frac{1}{2\lambda}\bar{\chi}^2+\frac{N}{2}\int\frac{d^dk}{(2\pi)^d}\ln\left(k^2+m^2+\bar{\chi}\right)+\frac{1}{2}\left(m^2+\bar{\chi}\right)\bar{\varphi}^2\,, \tag{3.1.92}$$

which obviously recovers Eqs. (3.1.88) and (3.1.89) after varying with respect to constant $\bar{\varphi}^a$ and $\bar{\chi}$.

It is convenient to perform renormalization by introducing, in $d=4$, the renormalized coupling $\lambda_{\rm R}$ given by

$$\frac{1}{\lambda_{\rm R}} \;=\; \frac{1}{\lambda}+\frac{1}{2}\int^{\Lambda}\frac{d^4k}{(2\pi)^4}\frac{1}{k^2\left(k^2+m^2_{\rm R}\right)}=\frac{1}{\lambda}+\frac{N}{32\pi^2}\ln\frac{\Lambda^2}{m^2_{\rm R}} \tag{3.1.93}$$

and

$$\bar{\chi}_{\rm R} \;=\bar{\chi}+m^2\,. \tag{3.1.94}$$

Assuming that $\bar{\chi}_{\rm R}\ll\Lambda^2$ (also $m_{\rm R}\ll\Lambda$ as usual) and representing Eq. (3.1.71) in the form

$$\frac{m^2_{\rm R}}{\lambda_{\rm R}} \;=\; \frac{m^2}{\lambda}-\frac{N}{32\pi^2}\Lambda^2\,, \tag{3.1.95}$$

we rewrite Eqs. (3.1.88) and (3.1.89) as [Sch74, CJP74]

$$\bar{\chi}_{\rm R}\bar{\varphi}^a \;=\; 0\,, \tag{3.1.96}$$

$$\bar{\chi}_{\rm R}\left(1-\frac{\lambda_{\rm R}N}{32\pi^2}\ln\frac{\bar{\chi}_{\rm R}}{m^2_{\rm R}}\right) \;=\; m^2_{\rm R}+\frac{\lambda_{\rm R}}{2}\bar{\varphi}^2\,. \tag{3.1.97}$$

Equation (3.1.92) results then in the renormalized effective potential

$$V_{\rm R} \;=\; -\frac{1}{2\lambda_{\rm R}}\bar{\chi}^2_{\rm R}+\frac{m^2_{\rm R}\bar{\chi}_{\rm R}}{\lambda_{\rm R}}+\frac{N}{64\pi^2}\bar{\chi}^2_{\rm R}\left(-\frac{1}{2}+\ln\frac{\bar{\chi}_{\rm R}}{m^2_{\rm R}}\right)+\frac{1}{2}\bar{\chi}_{\rm R}\bar{\varphi}^2\,, \tag{3.1.98}$$

which obviously reproduces Eqs. (3.1.96) and (3.1.97).

Equations (3.1.96) and (3.1.97) possess the solution

$$\bar{\varphi}^a \;=\; 0\,, \qquad \bar{\chi}_{\rm R} \;=\; m^2_{\rm R} \qquad \text{for } m^2_{\rm R}>0\,, \tag{3.1.99}$$

$$\bar{\varphi}^2 \;=\; -\frac{2m^2_{\rm R}}{\lambda_{\rm R}}\,, \qquad \bar{\chi}_{\rm R} \;=\; 0 \qquad \text{for } m^2_{\rm R}<0\,. \tag{3.1.100}$$

The first of them is associated with an unbroken O(N) symmetry, while the second one corresponds to a spontaneous breaking of O(N) down to O($N-1$). Both formulas look like the proper tree-level ones while the only effect of loop corrections at large N is the renormalization of the coupling constant and mass.

A subtle point is a question of stability of these solutions. For small deviations of $\bar{\varphi}^2$ from the mean value given by Eqs. (3.1.99) and (3.1.100), the effective potential $V_{\rm R}$ is a monotonically increasing function of $\bar{\varphi}^2$, as can be shown for $\lambda_{\rm R}N<32\pi^2$ eliminating the auxiliary field $\bar{\chi}_{\rm R}$ from Eq. (3.1.98) by solving the gap equation (3.1.97) iteratively in $\bar{\varphi}^2$, and the solutions are locally stable. Both solutions are, however, unstable globally with respect to large fluctuations of the fields. This can be seen by eliminating $\bar{\varphi}^2$ from $V_{\rm R}$ by solving the gap equation (3.1.97) for $\bar{\varphi}^2$ which yields

$$V_{\rm R} \;=\; \frac{1}{2}\bar{\chi}^2_{\rm R}\left(\frac{1}{\lambda_{\rm R}}-\frac{N}{32\pi^2}\ln\frac{\bar{\chi}_{\rm R}}{m^2_{\rm R}}\right)-\frac{N}{128\pi^2}\bar{\chi}^2_{\rm R}\,. \tag{3.1.101}$$

This function is monotonically decreasing for very large

$$\bar{\chi}_{\rm R} \;>\; m^2_{\rm R}\,{\rm e}^{\frac{32\pi^2}{\lambda_{\rm R}N}}\,, \tag{3.1.102}$$

where the theory becomes unstable. This is related to the usual problem of "triviality" of the φ^4 theory which makes sense only for small couplings $\lambda_{\rm R}N$ as an effective theory and cannot be fundamental at very small distances of the order of

$$r \;\sim\; m^{-1}_{\rm R}\,{\rm e}^{-\frac{16\pi^2}{\lambda_{\rm R}N}}\,. \tag{3.1.103}$$

Problem 3.9 Find a solution to Eqs. (3.1.88) and (3.1.89) which exponentially decreases as

$$\varphi^a_{cl}(x) \;=\; \xi^a m_{\rm R}\,{\rm e}^{\,m_{\rm R}\tau} \qquad \text{for } \tau\to-\infty \tag{3.1.104}$$

where $\tau\equiv x_4$ and ξ^a is an O(N) vector.

Solution The difference from the previous problem is that φ_{cl} is no longer translationally invariant along the time-variable due to the initial condition (3.1.104). Let us denote

$$\varphi^a_{cl}(x) \;\equiv\; \Phi^a(\tau)\,,$$

$$\chi_{cl}(x) \;\equiv\; v(\tau)\,. \tag{3.1.105}$$

The the saddle-point equations (3.1.88) and (3.1.89) can be rewritten as

$$\left(-D^2+m^2+v(\tau)\right)\Phi^a(\tau) \;=\; 0 \tag{3.1.106}$$

and

$$v(\tau) \;=\; \frac{\lambda}{2}\Phi^a(\tau)\Phi^a(\tau)+\frac{\lambda N}{2}\int\frac{d^3k}{(2\pi)^3}G_\omega\left(\tau,\tau;v\right)\,, \tag{3.1.107}$$

where

$$D \;\equiv\; \frac{d}{d\tau}\,, \tag{3.1.108}$$

$$\omega \;=\; \sqrt{k^2+m^2} \tag{3.1.109}$$

and we have introduced the Fourier image of the Green function (3.1.61)

$$\begin{aligned} G_\omega\left(\tau,\tau;v\right) \;&\equiv\; \int d^3\vec{x}\,\mathrm{e}^{\,i\vec{k}\vec{x}}G\left((\tau,\vec{x}),(\tau,\vec{0});v\right) \\ &=\; \left\langle\tau\left|\frac{1}{-D^2+\omega^2+v}\right|\tau\right\rangle \end{aligned} \tag{3.1.110}$$

with respect to the spatial coordinate.

The solution to Eqs. (3.1.106) and (3.1.107) can easily be found to be

$$\Phi^a(\tau) \;=\; \frac{\xi^a m_{\rm R}\,\mathrm{e}^{\,m_{\rm R}\tau}}{1-\frac{\bar\lambda_{\rm R}\xi^2}{16}\,\mathrm{e}^{\,2m_{\rm R}\tau}}\,, \tag{3.1.111}$$

$$v(\tau) \;=\; \frac{\bar\lambda_{\rm R}}{2}\Phi^a(\tau)\Phi^a(\tau) \tag{3.1.112}$$

where the renormalized coupling

$$\bar\lambda_{\rm R} \;=\; \frac{\lambda_{\rm R}}{1+\frac{\lambda_{\rm R}N}{16\pi^2}} \tag{3.1.113}$$

differs from Eq. (3.1.93) only by an additional final renormalization and the renormalized mass $m_{\rm R}$ is defined in Eq. (3.1.71). This solution is nontrivial for $\xi^2\sim N$ and obviously satisfies the initial condition (3.1.104).

The solution is so simple because the diagonal resolvent (3.1.110) takes on the very simple form

$$G_\omega\left(\tau,\tau;v\right) \;=\; \frac{1}{2\omega}-\frac{v\left(\tau\right)}{4\omega(\omega^2-m_{\rm R}^2)} \tag{3.1.114}$$

for the potential $v(\tau)$ given by Eqs. (3.1.111), (3.1.112). This can be easily verified by substituting into the Gelfand–Dickey equation (1.1.123) with $\mathcal{G}=1$. This is a feature of an integrable potential which was already discussed in Problem 1.29.

The function $\Phi^a(\tau)$ given by Eq. (3.1.111) describes large-N amplitudes of multiparticle production at a threshold [Mak94].

3.1.4 Nonlinear sigma model

The nonlinear O(N) sigma model[4] in 2 Euclidean dimensions is defined by the partition function

$$Z \;=\; \int D\vec{n}\,\delta\left(\vec{n}^2-\frac{1}{g^2}\right)\mathrm{e}^{\,-\frac12\int d^2x(\partial_\mu\vec{n})^2} \tag{3.1.115}$$

where

$$\vec{n} \;=\; (n_1,\ldots,n_N) \tag{3.1.116}$$

is an O(N) vector. While the action in Eq. (3.1.115) is pure Gaussian, the model is not free due to the constraint

$$\vec{n}^2(x) \;=\; \frac{1}{g^2}\,, \tag{3.1.117}$$

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 [Pol75];

2) instantons for $N=3$ [BP75].

The action in Eq. (3.1.115) is $\sim N$ as $N\to\infty$ 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 $N\to\infty$, and rewrite the partition function (3.1.115) as

$$Z \;\propto\; \int_\uparrow Du\,(x)\int D\vec{n}\,(x)\,\mathrm{e}^{\,-\frac12\int d^2x\left[(\partial_\mu\vec{n})^2-u\left(\vec{n}^2-\frac{1}{g^2}\right)\right]}, \tag{3.1.118}$$

where the contour of integration over $u(x)$ is parallel to imaginary axis.

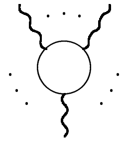
Doing the Gaussian integration over $\vec{n}$, we get

$$Z \;\propto\; \int_\uparrow Du\,(x)\,\mathrm{e}^{\,-\frac{N}{2}\mathrm{Sp}\ln\left(-\partial_\mu^2+u(x)\right)+\frac{1}{2g^2}\int d^2x\,u(x)}\,. \tag{3.1.119}$$

[4]The 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.

The first term in the exponent is as before nothing but the sum of one-loop diagrams in 2 dimensions

$$\frac{N}{2}\mathrm{Sp}\ln\left(-\partial_\mu^2+u\left(x\right)\right) \;=\; \sum_n \frac{1}{n} \tag{3.1.120}$$



where the auxiliary field u is denoted again by the wavy line. Equation (3.1.119) looks very much like Eq. (3.1.65) if we put there $J^a = K = 0$. The difference is that the exponent in (3.1.119) involves the term which is linear in u, while the analogous term in (3.1.65) is quadratic in χ.

Now the path integral over $u(x)$ in Eq. (3.1.119) 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

$$\frac{1}{g^2} - NG\left(x,x;u_{\mathrm{sp}}\right) \;=\; 0 \tag{3.1.121}$$

is quite analogous to Eq. (3.1.69) for the φ^4 theory while G is defined by

$$G\left(x,y;u\right) \;=\; \left\langle y\left|\frac{1}{-\partial_\mu^2+u}\right|x\right\rangle, \tag{3.1.122}$$

which is an analog of Eq. (3.1.61).

The coupling g^2 in Eq. (3.1.121) is $\sim 1/N$ as is prescribed by the constraint (3.1.117) which involves a sum over N terms on the LHS. This guarantees that a solution to Eq. (3.1.121) 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 (3.1.117), say, choosing

$$n_N \;=\; \frac{1}{g}\sqrt{1-g^2\sum_{a=1}^{N-1}n_a^2} \tag{3.1.123}$$

and expanding the square root in g^2. Only $N-1$ dynamical degrees of freedom are left so that the $\mathrm{O}(N)$-symmetry is broken in perturbation theory down to

$\mathrm{O}(N-1)$. The particles in perturbation theory are massless (like Goldstone bosons) and it suffers from infrared divergencies.

On the contrary, the solution to Eq. (3.1.121) has the form

$$u_{\mathrm{sp}} \;=\; m_{\mathrm{R}}^2 \equiv \Lambda^2\,\mathrm{e}^{-\frac{4\pi}{Ng^2}}, \tag{3.1.124}$$

where Λ is an ultraviolet cutoff. Therefore, all N particles acquires the same mass m_{R} in the $1/N$-expansion so that the $\mathrm{O}(N)$ 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.

Problem 3.10 Show that (3.1.124) is a solution to Eq. (3.1.121).
Solution Let us look for a translationally invariant solution

$$u_{\mathrm{sp}}(x) \;\equiv\; m_{\mathrm{R}}^2\,. \tag{3.1.125}$$

Then Eq. (3.1.121) in the momentum space reads

$$\frac{1}{g^2} \;=\; N\int^{\Lambda}\frac{d^2k}{(2\pi)^2}\frac{1}{k^2+m_{\mathrm{R}}^2} \;=\; \frac{N}{4\pi}\int_0^{\Lambda^2}\frac{dk^2}{k^2+m_{\mathrm{R}}^2} \;=\; \frac{N}{4\pi}\ln\frac{\Lambda^2}{m_{\mathrm{R}}^2}\,. \tag{3.1.126}$$

The exponentiation results in Eq. (3.1.124).

Equation (3.1.126) relates the bare coupling g^2 and the cutoff Λ and allows us to calculate the Gell-Mann–Low function yielding

$$\mathcal{B}(g^2) \;\equiv\; \frac{\Lambda\, dg^2}{d\Lambda} \;=\; -\frac{Ng^4}{2\pi}\,. \tag{3.1.127}$$

The analogous one-loop perturbation-theory formula for any N reads [Pol75]

$$\mathcal{B}(g^2) \;=\; -\frac{(N-2)g^4}{2\pi}\,. \tag{3.1.128}$$

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 $\mathrm{O}(2)$ is Abelian.

3.1.5 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 $\mathrm{O}(N)$ 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

$$\begin{aligned} &\langle\, u\,(x_1)\ldots u\,(x_k)\,\rangle \\ &\equiv \quad Z^{-1}\int_{\uparrow} Du\, \mathrm{e}^{-\frac{N}{2}\mathrm{Sp}\ln\left[-\partial_\mu^2+u\right]+\frac{1}{2g^2}\int d^2x\, u}\, u\,(x_1)\cdots u\,(x_k) \end{aligned} \tag{3.1.129}$$

in the 2-dimensional sigma model.

Since the path integral has a saddle point at some[5]

$$u\,(x) \;=\; u_{\mathrm{sp}}\,(x)\,, \tag{3.1.130}$$

we get to the leading order in $1/N$:

$$\langle\, u\,(x_1)\ldots u\,(x_k)\,\rangle \;=\; u_{\mathrm{sp}}\,(x_1)\ldots u_{\mathrm{sp}}\,(x_k)+\mathcal{O}\left(\frac{1}{N}\right), \tag{3.1.131}$$

which can be written in the factorized form

$$\langle\, u\,(x_1)\ldots u\,(x_k)\,\rangle \;=\; \langle\, u\,(x_1)\,\rangle\ldots\langle\, u\,(x_k)\,\rangle+\mathcal{O}\left(\frac{1}{N}\right). \tag{3.1.132}$$

Therefore, u becomes "classical" as $N\to\infty$ 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.

[5] This saddle point is, in fact, x-independent due to translational invariance.

3.2 Multicolor QCD

The method of the $1/N$-expansion can be applied to QCD. This was done by 't Hooft [Hoo74a] 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 $N_c\to\infty$ and the Veneziano limit [Ven76] when the ratio N_f/N_c is fixed as $N_c\to\infty$. 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.

3.2.1 Index or ribbon graphs

In order to describe the $1/N_c$-expansion of QCD, whose extension to N_c colors has already been considered in Subsect. 2.1.1, it is convenient to use the matrix-field representation (2.1.5). In this Section we shall use a slightly different definition

$$[A_\mu\,(x)]^{ij} \;=\; \sum_a A_\mu^a\,(x)\,[t^a]^{ij}\;, \tag{3.2.1}$$

which is similar to that used by 't Hooft [Hoo74a] and differs from (2.1.5) by a factor of ig:

$$\mathcal{A}_\mu^{ij}\,(x) \;=\; igA_\mu^{ij}\,(x)\,. \tag{3.2.2}$$

The matrix (3.2.1) is Hermitean.

The propagator of the matrix field $A^{ij}(x)$, in this notation, takes on the form

$$\left\langle A^{ij}_{\mu}(x)\, A^{kl}_{\nu}(y) \right\rangle_{\rm Gauss} = \frac{1}{2}\left(\delta^{il}\delta^{kj} - \frac{1}{N_c}\delta^{ij}\delta^{kl}\right) D_{\mu\nu}(x-y)\,, \tag{3.2.3}$$

where we have assumed, as usual, a gauge fixing to define the propagator in perturbation theory. For instance, one has

$$D_{\mu\nu}(x-y) = \frac{1}{4\pi^2}\frac{\delta_{\mu\nu}}{(x-y)^2} \tag{3.2.4}$$

in the Feynman gauge.

Equation (3.2.3) can be immediately derived from the standard formula

$$\left\langle A^{a}_{\mu}(x)\, A^{b}_{\nu}(y)\right\rangle_{\rm Gauss} = \delta^{ab} D_{\mu\nu}(x-y) \tag{3.2.5}$$

multiplying by the generators of the $\mathrm{SU}(N_c)$ gauge group according to the definition (2.1.5) and using the completeness condition

$$\sum_{a=1}^{N_c^2-1} (t^a)^{ij}(t^a)^{kl} = \frac{1}{2}\left(\delta^{il}\delta^{kj} - \frac{1}{N_c}\delta^{ij}\delta^{kl}\right) \qquad \boxed{\text{for }\ \mathrm{SU}(N_c)}\,, \tag{3.2.6}$$

where the factor of 1/2 is due to the normalization (2.1.6). Eq. (3.2.3) can be derived directly from a path integral over matrices.

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 $\mathrm{SU}(N_c)$ 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. (3.2.3), we depict the propagator by the double line

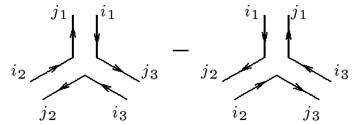
$$\left\langle A^{ij}_{\mu}(x)\, A^{kl}_{\nu}(y) \right\rangle_{\rm Gauss} \propto \delta^{il}\delta^{kj} = \begin{array}{ccc} i & \longrightarrow & l \\ j & \longleftarrow & k \end{array}\,. \tag{3.2.7}$$

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^{ij}_{μ} is Hermitean and its off-diagonal components are complex conjugate. The independent fields are, say, the complex fields A^{ij}_{μ} for $i>j$ and the diagonal real fields A^{ii}_{μ}. The arrow represents the direction of the propagation of the indices of the complex field A^{ij}_{μ} for $i>j$ while the complex-conjugate one, $A^{ji}_{\mu} = \left(A^{ij}_{\mu}\right)^*$, propagates in the opposite direction. For the real fields A^{ii}_{μ}, the arrows are not essential.

The double-line notation (3.2.7) looks similar to that of Subsect. 2.2.5. 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 (2.1.14), is depicted in the double-line notations as



$$\propto g\left(\delta^{i_1j_3}\delta^{i_2j_1}\delta^{i_3j_2} - \delta^{i_1j_2}\delta^{i_3j_1}\delta^{i_2j_3}\right) \tag{3.2.8}$$

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 (2.1.14). The color part of the three-vertex is antisymmetric under interchanging the gluons. The space-time structure, which reads in the momentum space as

$$\begin{aligned}&\gamma_{\mu_1\mu_2\mu_3}(p_1,p_2,p_3)\\ &\quad= \delta_{\mu_1\mu_2}(p_1-p_2)_{\mu_3} + \delta_{\mu_2\mu_3}(p_2-p_3)_{\mu_1} + \delta_{\mu_1\mu_3}(p_3-p_1)_{\mu_2}\,,\end{aligned} \tag{3.2.9}$$

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. (3.2.8) can alternatively be obtained by multiplying the standard vertex

$$\Gamma^{a_1a_2a_3}_{\mu_1\mu_2\mu_3}(p_1,p_2,p_3) = f^{a_1a_2a_3}\gamma_{\mu_1\mu_2\mu_3}(p_1,p_2,p_3) \tag{3.2.10}$$

by $(t^{a_1})^{i_1j_1}(t^{a_2})^{i_2j_2}(t^{a_3})^{i_3j_3}$, with f^{abc} being the structure constants of the $\mathrm{SU}(N_c)$ group, and using the formula

$$\begin{aligned}&f^{a_1a_2a_3}(t^{a_1})^{i_1j_1}(t^{a_2})^{i_2j_2}(t^{a_3})^{i_3j_3}\\ &\quad= \frac{1}{2}\left(\delta^{i_1j_3}\delta^{i_2j_1}\delta^{i_3j_2} - \delta^{i_1j_2}\delta^{i_3j_1}\delta^{i_2j_3}\right),\end{aligned} \tag{3.2.11}$$

which is a consequence of the completeness condition (3.2.6).

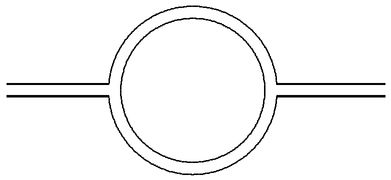
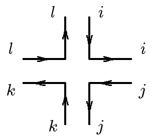


Fig. 3.6: Double-line representation of a one-loop diagram for the gluon propagator. The sum over the N_c indices is associated with the closed index line. The contribution of this diagram is $\sim g^2 N_c \sim 1$.

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



$$\propto g^2 \qquad (3.2.12)$$

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 [Hoo74a]. The simplest one which describes the one-loop correction to the gluon propagator is depicted in Fig. 3.6.[6] 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. (2.2.70). 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

$$g^2 \sim \frac{1}{N_c}\,. \qquad (3.2.13)$$

This dependence on N_c is similar to Eqs. (3.1.14) and (3.1.68) for the vector

[6]Here and in the most figures below the arrows of the index lines are omitted for simplicity.

models and is prescribed by the asymptotic-freedom formula

$$g^2 = \frac{24\pi^2}{11 N_c \ln\left(\Lambda/\Lambda_{QCD}\right)} \qquad (3.2.14)$$

of the pure $\mathrm{SU}(N_c)$ gauge theory.

Thus, the contribution of the diagram of Fig. 3.6 is of order

$$\text{Fig. 3.6} \sim g^2 N_c \sim 1 \qquad (3.2.15)$$

in the large-N_c limit.

The double lines of the diagram in Fig. 3.6 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 *ribbon* 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. (3.2.6) can be omitted at large N_c. Such a completeness condition emerges for the $\mathrm{U}(N_c)$ group whose generators T^A $(A = 1, \ldots, N_c^2)$ are

$$T^A = \left(t^a, \frac{\mathrm{I}}{\sqrt{2N}}\right), \qquad \mathrm{tr}\, T^A T^B = \frac{1}{2}\delta^{AB}\,. \qquad (3.2.16)$$

They obey the completeness condition

$$\sum_{A=1}^{N_c^2} \left(T^A\right)^{ij} \left(T^A\right)^{kl} = \frac{1}{2}\delta^{il}\delta^{kj} \qquad \boxed{\text{for } \mathrm{U}(N_c)}\,. \qquad (3.2.17)$$

The point is that elements of both the $\mathrm{SU}(N_c)$ group and the $\mathrm{U}(N_c)$ group can be represented in the form

$$U = \mathrm{e}^{iB}\,, \qquad (3.2.18)$$

where B is a general Hermitean matrix for $\mathrm{U}(N_c)$ and a traceless Hermitean matrix for $\mathrm{SU}(N_c)$.

Therefore, the double-line representation of perturbation-theory diagrams which is described in this Section holds, strictly speaking, only for the $\mathrm{U}(N_c)$ gauge group. However, the large-N_c limit of both the $\mathrm{U}(N_c)$ group and the $\mathrm{SU}(N_c)$ group is the same.

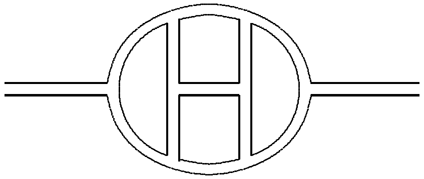


Fig. 3.7: Double-line representation of a four-loop diagram for the gluon propagator. The sum over the N_c indices is associated with each of the four closed index lines whose number is equal to the number of loops. The contribution of this diagram is $\sim g^8 N_c^4 \sim 1$.

3.2.2 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. (3.2.13).

Let us consider a typical diagram for the gluon propagator depicted in Fig. 3.7. 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

$$\text{Fig. 3.7} \;\sim\; \left(g^2 N_c\right)^4 \;\sim\; 1\,. \tag{3.2.19}$$

The diagrams of the type in Fig. 3.7, 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

$$n_2\text{-loop planar diagram} \;\sim\; \left(g^2 N_c\right)^{n_2} \;\sim\; 1\,, \tag{3.2.20}$$

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. 3.8. 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.

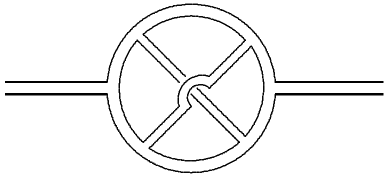


Fig. 3.8: Double-line representation of a three-loop non-planar diagram for the gluon propagator. The diagram has six three-gluon vertices but only one closed index line (while three loops!). The order of this diagram is $\sim g^6 N_c \sim 1/N_c^2$.

The order of this diagram in $1/N_c$ is

$$\text{Fig. 3.8} \;\sim\; g^6 N_c \;\sim\; \frac{1}{N_c^2}\,. \tag{3.2.21}$$

It is therefore suppressed at large N_c by $1/N_c^2$.

The non-planar diagram in Fig. 3.8 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. 3.6–3.8 can now be described by the unique formula

$$\text{genus-}h\text{ diagram} \;\sim\; \left(\frac{1}{N_c^2}\right)^{\text{genus}}\,. \tag{3.2.22}$$

Thus, the expansion in $1/N_c$ rearranges perturbation-theory diagrams according to their topology [Hoo74a]. For this reason, it is referred to as the *topological expansion* or the *genus expansion*. The general proof of Eq. (3.2.22) for an arbitrary diagram is given in Subsect. 3.2.4.

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 [Tut62, KNN77]

$$\#_{\rm p}\,(n_0) \equiv \#\text{ of planar graphs} \;\sim\; \text{const}^{n_0}\,, \tag{3.2.23}$$

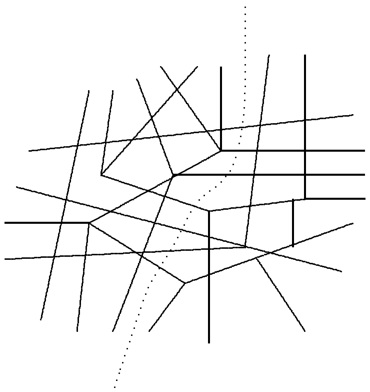


Fig. 3.9: Cutting a planar graph into two graphs. The cutting is along the dotted line. The numbers of vertices of each part and of the whole graph obey Eq. (3.2.24).

while the number of all the graphs grows with n_0 factorially. There is no dependence in Eq. (3.2.23) 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.

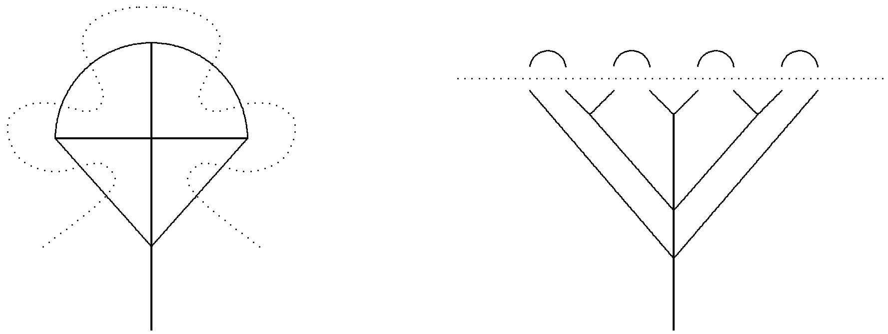
In the rest of these lecture notes, we shall discuss what is known about solving the problem of summing the planar graphs.

Problem 3.11 Show that Eq. (3.2.23) for the number of planar graphs is consistent with its independence of the number of external lines.
Solution Let us split a planar graph in two parts by cutting along some line as is depicted in Fig. 3.9. The numbers of vertices of each part, n_0' and n_0'', are obviously related to that of the original graph, n_0, by

$$n_0' + n_0'' \;=\; n_0\,. \tag{3.2.24}$$

We assume that both n_0' and n_0'' are large.

The number of cut lines is $\sim \sqrt{n_0}$ for a planar graph in contrast to that for a



a) b)

Fig. 3.10: Cutting a planar graph into trees and arches. The line of cutting is depicted in a) by the dotted line. The combination of tree and arches in b) is obtained from a) by a continuous distortion.

generic non-planar one, when it would be $\sim n_0$. Disregarding the cut lines, we get

$$\#_{\rm P}\,(n_0) \;=\; \#_{\rm P}\left(n_0'\right)\cdot \#_{\rm P}\left(n_0''\right), \tag{3.2.25}$$

which is obviously satisfied by the formula (3.2.23) accounting for Eq. (3.2.24).

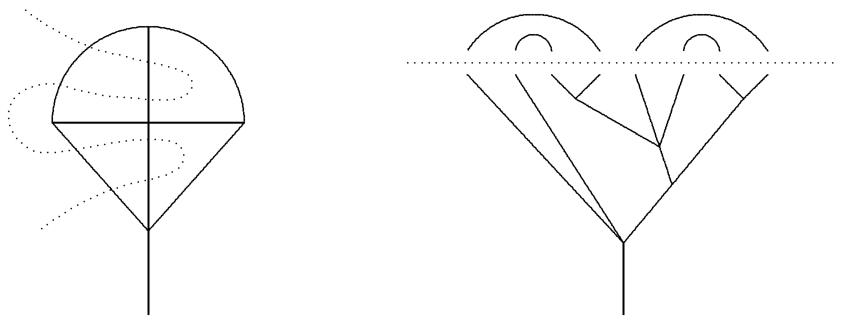
Problem 3.12 Cutting all loops of a planar graph, obtain the upper bound

$$\#_{\rm P} \le (1024)^{n_2} \tag{3.2.26}$$

for the number of planar graphs with n_2 loops.
Solution Since $\#_{\rm P}$ does not depend on the number of external lines (see Problem 3.11), let us consider a one particle irreducible planar graph with one external line and cut all the loops as is depicted in Fig. 3.10a. By a continuous distortion, it can be depicted like in Fig. 3.10b, where below the dotted line we have a tree with n_0 vertices and above the dotted line we have n_2 arches. The latter number coincides with the number of loops of the planar graph. The number of tips of the tree is $2n_2$.

Since each planar graph can be cut is several ways, $\#_{\rm P}$ is bounded from above by

$$\#_{\rm P} \;\le\; \#_{\rm A}\,(n_2)\;\#_{\rm T}\,(n_0,2n_2)\,, \tag{3.2.27}$$



a) b)

Fig. 3.11: Alternative cutting of the same planar graph as in Fig. 3.10 into trees and arches.

where $\#_{\mathrm{A}}(n_2)$ stands for the number of arches and $\#_{\mathrm{T}}(n_0, 2n_2)$ stands for the number of trees with n_0 vertices and $2n_2$ tips. An alternative way of cutting the same planar graph, which leads to a different combination of arches and trees, is depicted in Fig. 3.11.

The number of arches is well-known in mathematics and is given by the Catalan number of order n:

$$\#_{\mathrm{A}}(n) \;=\; \frac{2n!}{n!(n+1)!} \;\stackrel{n\to\infty}{\longrightarrow}\; 4^n\,. \tag{3.2.28}$$

The number of trees is not independent since a graph, dual to a tree graph, consists of arches as is illustrated by Fig. 3.12. The number of arches of this dual graph equals the sum of the number n_0 of vertices and the number $2n_2$ of tips, *i.e.* equals $n_0 + 2n_2$. Given the number n_2 of loops, the number n_0 of vertices is maximal when all vertices are trivalent, so that

$$n_0 \le 2n_2 - 1 \tag{3.2.29}$$

($n_0 = 2n_2$ for trivalent and $n_0 = n_2$ for fourvalent vertices when n_2 is large). Therefore, the number of arches of the dual graph is bounded by $4n_2$, so that

$$\#_{\mathrm{T}}(n_0, 2n_2) \;\le\; \#_{\mathrm{A}}(4n_2)\,. \tag{3.2.30}$$

Substituting in (3.2.27), we get [KNN77] the inequality (3.2.26). Finally, the inequality (3.2.23) can be obtained by noting that $n_0 \sim n_2$ for large n_2.

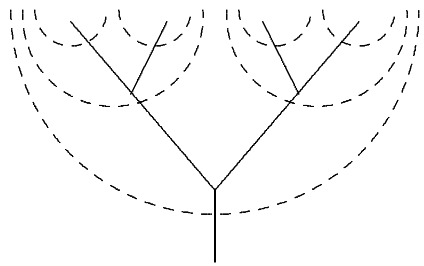


Fig. 3.12: A tree graph (the solid lines) and its dual (the dashed arches).

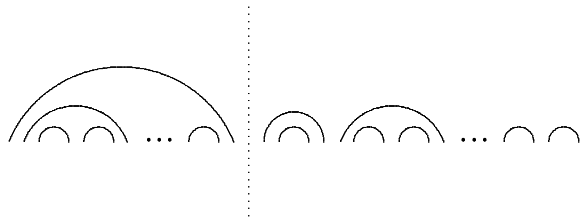


Fig. 3.13: Recurrence relation for the number of arches. The dotted line separates a configuration of n arches in two pieces: n' to the left and $n - n'$ to the right.

Problem 3.13 Derive Eq. (3.2.28) for the number of arches.
Solution Let us consider a general configuration of n arches like is depicted in Fig. 3.13. Let us pick up the leftmost arch, splitting the configuration in two pieces: n' arches to the left and $n - n'$ arches to the right of the dotted line. The number of arches obviously satisfies the recurrence relation

$$\#_{\mathrm{A}}(n) \;=\; \sum_{n'=1}^{n} \#_{\mathrm{A}}\left(n'-1\right) \#_{\mathrm{A}}\left(n-n'\right), \tag{3.2.31}$$

where the number of arches to the left of the dotted line is described by $\#_{\mathrm{A}}(n'-1)$ because one arch encircles $n'-1$ others. Eq. (3.2.31) expresses $\#_{\mathrm{A}}(n)$ recurrently via $\#_{\mathrm{A}}(0) = 1$.

Introducing the generating function

$$f_{\mathrm{A}}(g) \;=\; \sum_{n=0}^{\infty} g^{n+1} \#_{\mathrm{A}}(n)\,, \tag{3.2.32}$$

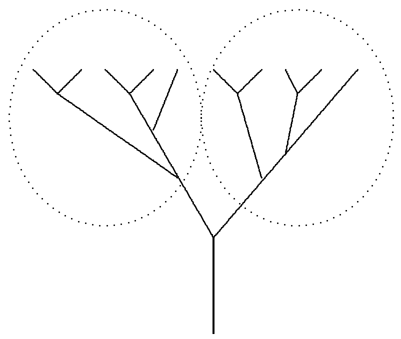


Fig. 3.14: Recurrence relation for the number of trees. The trees inside the left and right dotted circles have n' and $n-n'$ tips, respectively.

we rewrite Eq. (3.2.36) as the quadratic equation

$$f_{\mathrm{A}}(g) - g = f_{\mathrm{A}}^2(g)\,. \tag{3.2.33}$$

Its solution

$$f_{\mathrm{A}}(g) = \frac{1-\sqrt{1-4g}}{2} = \sum_{n=0}^{\infty} g^{n+1} \frac{(2n)!}{n!(n+1)!} \tag{3.2.34}$$

gives Eq. (3.2.28) for the number of arches.

Problem 3.14 Improve the inequality (3.2.26), calculating the number of trivalent tree graphs with n tips.
Solution Let us first note that the number of vertices of a trivalent tree graph with n tips equals $n-1$. Hence, we are interested in

$$\#_{\mathrm{T}}(n) \equiv \#_{\mathrm{T}}(n-1,n) \tag{3.2.35}$$

in the notation of Problem 3.12. Picking up the first vertex in a tree as is depicted in Fig. 3.14, we get the following recursion relation

$$\#_{\mathrm{T}}(n) = \sum_{n'=1}^{n-1} \#_{\mathrm{T}}\left(n'\right) \#_{\mathrm{T}}\left(n-n'\right), \tag{3.2.36}$$

which expresses $\#_{\mathrm{T}}(n)$ via $\#_{\mathrm{T}}(1) = 1$.

Introducing the generating function

$$f_{\mathrm{T}}(g) = \sum_{n=1}^{\infty} g^{n-1} \#_{\mathrm{T}}(n)\,, \tag{3.2.37}$$

where g^{n-1} corresponds to $n-1$ vertices of each tree, we rewrite Eq. (3.2.36) as the quadratic equation

$$f_{\mathrm{T}}(g) - 1 = g f_{\mathrm{T}}^2(g)\,. \tag{3.2.38}$$

Its solution

$$f_{\mathrm{T}}(g) = \frac{1-\sqrt{1-4g}}{2g} = \sum_{n=1}^{\infty} g^{n-1} \frac{(2n-2)!}{n!(n-1)!} \tag{3.2.39}$$

gives

$$\#_{\mathrm{T}}(n) = \frac{(2n-2)!}{n!(n-1)!}\,. \tag{3.2.40}$$

Returning to Problem (3.12), it is shown that

$$\#_{\mathrm{T}}(n_0, 2n_2) \le \#_{\mathrm{T}}(2n_2-1, 2n_2) = \#_{\mathrm{A}}(2n_2-1)\,. \tag{3.2.41}$$

The inequality here is due to (3.2.29) and the equality is because of the explicit formulas (3.2.28) and (3.2.40). We have improved, thus, the estimate (3.2.30) having calculated the number of tree graphs. The inequality (3.2.26) is now improved by

$$\#_{\mathrm{P}} \le (64)^{n_2}\,. \tag{3.2.42}$$

The actual number of the trivalent planar graphs was evaluated by Tutte [Tut62] and reads

$$\#_{\mathrm{P}} \approx \left(\frac{256}{27}\right)^{n_2} \tag{3.2.43}$$

for asymptotically large n_2.

3.2.3 Planar and non-planar graphs (the boundaries)

Equation (3.2.22) 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 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. 3.15 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

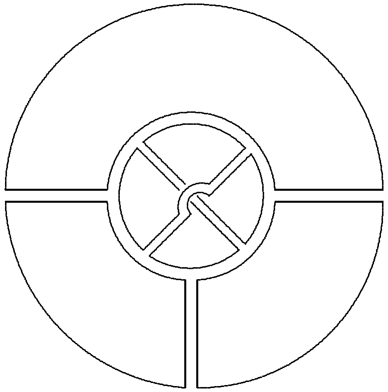


Fig. 3.15: Generic index diagram with $n_0 = 10$ vertices, $n_1 = 10$ gluon propagators, $n_2 = 4$ closed index lines, and $B = 1$ boundary. The color indices of the external lines are contracted by the Kronecker deltas (represented by the single lines) in a cyclic order. The extra factor of $1/N_c$ is due to the normalization (3.2.44). Its order in $1/N_c$ is $\sim 1/N_c^2$ in accord with Eq. (3.2.22).

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).

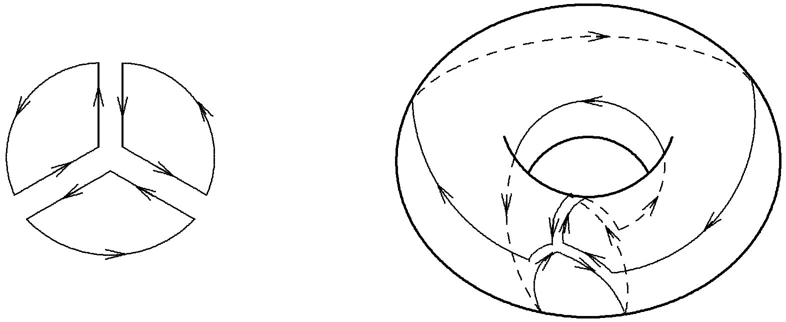
It is clear from the graphic representation that the diagram in Fig. 3.15 is associated with the trace over the color indices of the three-point Green's function

$$G^{(3)}_{\mu_1\mu_2\mu_3}(x_1,x_2,x_3) \;\equiv\; \frac{(ig)^3}{N_c}\left\langle \operatorname{tr}\left[A_{\mu_1}(x_1)\,A_{\mu_2}(x_2)\,A_{\mu_3}(x_3)\right]\right\rangle . \quad (3.2.44)$$

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. 3.15 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. (3.2.44) to multi-point Green's functions is obvious:

$$G^{(n)}_{\mu_1\cdots\mu_n}(x_1,\ldots,x_n) \;\equiv\; \frac{(ig)^n}{N_c}\left\langle \operatorname{tr}\left[A_{\mu_1}(x_1)\cdots A_{\mu_n}(x_n)\right]\right\rangle . \quad (3.2.45)$$



a) b)

Fig. 3.16: Planar a) and non-planar b) contributions of the two color structures in Eq. (3.2.8) for three-gluon vertex to $G^{(3)}$ in the lowest order of perturbation theory.

The factor $1/N_c$, which normalizes the trace, provides the natural normalization

$$G^{(0)} \;=\; 1 \quad (3.2.46)$$

of the averages.

Though the two terms in the index-space representation (3.2.8) of the three-gluon vertex look very similar, their fate in the topological expansion is quite different. When the color indices are contracted anti-clockwise, the first term leads to the planar contributions to $G^{(3)}$, the simplest of which is depicted in Fig. 3.16a. The anti-clockwise contraction of the color indices in the second term leads to a non-planar graph in Fig. 3.16b 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. 3.15 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

a) b)

Fig. 3.17: Example of a) connected and b) disconnected planar graphs.

of the lines is prescribed by the orientation of the external boundary which we choose to be, say, anti-clockwise like in Fig. 3.16a. 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 (3.2.45) naturally arise in the expansion of the trace of the non-Abelian phase factor for a closed contour, which was considered in Problem 2.2. One gets

$$\begin{aligned} &\left\langle \frac{1}{N_c} \operatorname{tr} \mathbf{P} \, \mathrm{e}^{ig \oint_\Gamma dx^\mu A_\mu(x)} \right\rangle \\ &\quad = \sum_{n=0}^{\infty} \oint_\Gamma dx_1^{\mu_1} \int_{x_1}^{x_1} dx_2^{\mu_2} \ldots \int_{x_1}^{x_{n-1}} dx_n^{\mu_n} \, G^{(n)}_{\mu_1 \cdots \mu_n}(x_1, \ldots, x_n) \,. \end{aligned} \tag{3.2.47}$$

The reason is because the ordering along a closed path implies the cyclic-ordering in the index space.

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. 3.17. The graph in Fig. 3.17a is connected

planar while the graph in Fig. 3.17b is disconnected planar.

The usual relation (1.2.52) between the generating functionals $W[J]$ and $Z[J]$ for connected graphs and all graphs, which is discussed in the Remark on p. 45, 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 (3.2.45) can give disconnected non-planar ones.

The generating functionals for all and connected planar graphs can be constructed [Cvi81] by means of introducing non-commutative sources $\boldsymbol{j}_\mu(x)$. "Non-commutative" means that there is no way to transform $\boldsymbol{j}_{\mu_1}(x_1)\boldsymbol{j}_{\mu_2}(x_2)$ into $\boldsymbol{j}_{\mu_2}(x_2)\boldsymbol{j}_{\mu_1}(x_1)$. This non-commutativity of the sources reflects the cyclic-ordered structure of the Green's functions (3.2.45) which possess only cyclic symmetry.

Using the short-hand notations (3.1.64) where the symbol $\circ$ includes the sum over the d-vector (or whatever available) indices except for the color ones:

$$\boldsymbol{j} \circ \mathcal{A} \;\equiv\; \sum_\mu \int d^d x \, \boldsymbol{j}_\mu(x) \mathcal{A}_\mu(x) \,, \tag{3.2.48}$$

we write down the definitions of the generating functionals for all planar and connected planar graphs, respectively, as

$$Z[\boldsymbol{j}] \;\equiv\; \sum_{n=0}^{\infty} \left\langle \frac{1}{N_c} \operatorname{tr} \, (\boldsymbol{j} \circ \mathcal{A})^n \right\rangle \tag{3.2.49}$$

and

$$W[\boldsymbol{j}] \;\equiv\; \sum_{n=0}^{\infty} \left\langle \frac{1}{N_c} \operatorname{tr} \, (\boldsymbol{j} \circ \mathcal{A})^n \right\rangle_{\text{conn}} \,. \tag{3.2.50}$$

The planar contribution to the Green's functions (3.2.45) and their connected counterparts can be obtained, respectively, from the generating functionals $Z[\boldsymbol{j}]$ and $W[\boldsymbol{j}]$ applying the non-commutative derivative which is defined by

$$\frac{\delta}{\delta \boldsymbol{j}_\mu(x)} \boldsymbol{j}_\nu(y) \, f(\boldsymbol{j}) \;=\; \delta_{\mu\nu} \delta^{(d)}(x-y) \, f(\boldsymbol{j}) \,, \tag{3.2.51}$$

where f is an arbitrary function of $\boldsymbol{j}$'s. In other words the derivative picks up only the leftmost variable.

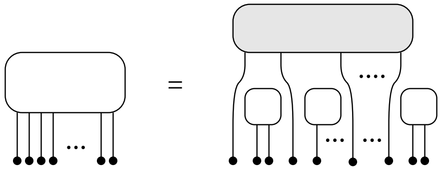


Fig. 3.18: Graphic derivation of Eq. (3.2.52): $Z[\boldsymbol{j}]$ is denoted by an empty box, $W[\boldsymbol{j}]$ is denoted by a shadow box, $\boldsymbol{j}$ is denoted by a filled circle. By picking a leftmost external line of a planar graph, we end up with a connected planar graph, whose remaining external lines are somewhere to the right interspersed by disconnected planar graphs. It is evident that $\boldsymbol{j}Z[\boldsymbol{j}]$ 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 $Z[\boldsymbol{j}]\boldsymbol{j}$, which results in Eq. (3.2.53).

The relation which replaces Eq. (1.2.52) for planar graphs is

$$Z[\boldsymbol{j}] \;=\; W[\boldsymbol{j}Z[\boldsymbol{j}]]\,, \tag{3.2.52}$$

while the cyclic symmetry says

$$W[\boldsymbol{j}Z[\boldsymbol{j}]] \;=\; W[Z[\boldsymbol{j}]\boldsymbol{j}]\,. \tag{3.2.53}$$

A graphic derivation of Eqs. (3.2.52) and (3.2.53) is given in Fig. 3.18. In other words, given $W[\boldsymbol{j}]$, one should construct an inverse function as the solution to the equation

$$\boldsymbol{j}_\mu(x) \;=\; \boldsymbol{J}_\mu(x)W[\boldsymbol{j}]\,, \tag{3.2.54}$$

after which Eq. (3.2.52) says

$$Z[\boldsymbol{j}] \;=\; W[\boldsymbol{J}]\,. \tag{3.2.55}$$

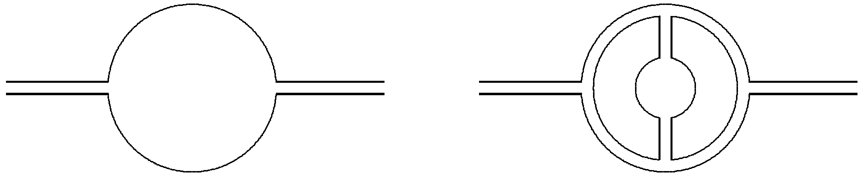
More about this approach to the generating functionals for planar graphs can be found in Ref. [CLS82].

Problem 3.15 Solve iteratively Eq. (3.2.54) for the Gaussian case.
Solution In the Gaussian case, only $G^{(2)}$ is nonvanishing which yields

$$W[\boldsymbol{j}] \;=\; 1 - g^2\boldsymbol{j}\circ D\circ\boldsymbol{j}\,, \tag{3.2.56}$$

where the propagator D is given by Eq. (3.2.4). Using Eq. (3.2.52), we get explicitly

$$Z[\boldsymbol{j}] \;=\; 1 - g^2\int d^dx\,d^dy\,D_{\mu\nu}(x-y)\,\boldsymbol{j}_\mu(x)\,Z[\boldsymbol{j}]\;\boldsymbol{j}_\nu(y)\,Z[\boldsymbol{j}]\,. \tag{3.2.57}$$



a) b)

Fig. 3.19: 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 $\sim g^6N_c^2 \sim 1/N_c$.

While this equation for $Z[\boldsymbol{j}]$ 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. (3.2.57) as

$$Z[\boldsymbol{j}] \;=\; \frac{1}{1+g^2\int d^dx\,d^dy\,D_{\mu\nu}(x-y)\,\boldsymbol{j}_\mu(x)\,Z[\boldsymbol{j}]\;\boldsymbol{j}_\nu(y)}\,, \tag{3.2.58}$$

whose iterative solution reads [Cvi81]

$$Z[\boldsymbol{j}] \;=\; \cfrac{1}{1+g^2\boldsymbol{j}\cfrac{\circ D\circ}{1+g^2\boldsymbol{j}\cfrac{\circ D\circ}{1+g^2\boldsymbol{j}\cfrac{\circ D\circ}{\vdots}\boldsymbol{j}}\boldsymbol{j}}\boldsymbol{j}}\,. \tag{3.2.59}$$

3.2.4 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

$$\left\langle\psi_i\bar\psi_j\right\rangle \;\propto\; \delta_{ij} \;=\; i \longrightarrow j\,. \tag{3.2.60}$$

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. 3.19a. It has two three gluon vertices and no closed index lines so that its order in $1/N_c$ is

$$\text{Fig. 3.19a} \sim g^2 \sim \frac{1}{N_c}. \tag{3.2.61}$$

Analogously, the order of a more complicated tree-loop diagram in Fig. 3.19b, which involves one quark loop and two closed index lines, is

$$\text{Fig. 3.19b} \sim g^6 N_c^2 \sim \frac{1}{N_c}. \tag{3.2.62}$$

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 diagram with L quark loops to have the same double-line representation as for pure gluon diagrams. Therefore, given Eq. (3.2.22), diagrams with L quark loops are suppressed at large N_c by

$$L \text{ quark loops} \sim \left(\frac{1}{N_c}\right)^{L+2\cdot\text{genus}}. \tag{3.2.63}$$

The single-line representation of the quark loops is similar to the one of the external boundary in Fig. 3.15. Moreover, such a diagram emerges when one calculates perturbative gluon corrections to the vacuum expectation value of the quark operator

$$O = \frac{1}{N_c}\bar{\psi}\psi, \tag{3.2.64}$$

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 (3.2.45) and (3.2.64).

In order to prove Eqs. (3.2.22) and its quark counterpart (3.2.63), 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. 3.20. Its order in $1/N_c$ is

$$\frac{1}{N_c^B} g^{n_0^{(3)}+2n_0^{(4)}} N_c^{n_2} \sim N_c^{n_2-n_0^{(3)}/2-n_0^{(4)}-B} \tag{3.2.65}$$

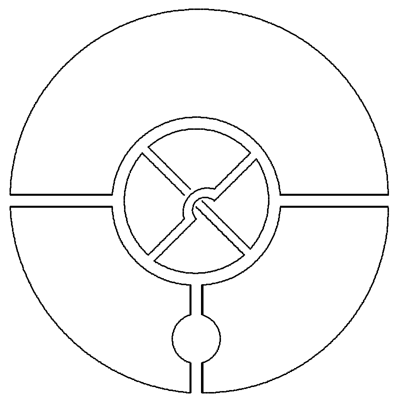


Fig. 3.20: Generic diagram in the index space which has $L = 1$ quark loop and $B = 1$ loop associated with the external boundary. Its order in $1/N_c$ is described by Eq. (3.2.71).

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

$$2n_1 = 3n_0^{(3)} + 4n_0^{(4)}, \tag{3.2.66}$$

since three- and four-point vertices emit three or four propagators, respectively, and each propagator connects two vertices. Using the relation (3.2.66), we rewrite the RHS of (3.2.65) as

$$N_c^{n_2-n_0^{(3)}/2-n_0^{(4)}-B} = N_c^{n_2-n_1+n_0-B}, \tag{3.2.67}$$

where the total number of vertices

$$n_0 = n_0^{(3)} + n_0^{(4)} \tag{3.2.68}$$

is introduced.

The exponent on the RHS of Eq. (3.2.67) 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. (3.2.63) from Eq. (3.2.22).

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

$$\chi \;\equiv\; 2 - 2h \;=\; n_2 + L + B - n_1 + n_0 \,. \tag{3.2.69}$$

Therefore the RHS of Eq. (3.2.67) can be rewritten as

$$N_c^{n_2 - n_1 + n_0 - B} \;=\; N_c^{2-2h-L-2B} \,. \tag{3.2.70}$$

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

$$\text{generic graph} \;\sim\; \left(\frac{1}{N_c}\right)^{2h+L+2(B-1)} \,. \tag{3.2.71}$$

For $B = 1$, we recover Eqs. (3.2.22) and (3.2.63).

Remark on the order of gauge action

We see from Eq. (3.2.45) 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. (3.2.2)). For these matrices

$$\frac{1}{N_c}\left\langle \mathrm{tr}\left[\mathcal{A}_{\mu_1}(x_1)\cdots\mathcal{A}_{\mu_n}(x_n)\right]\right\rangle \;=\; G^{(n)}_{\mu_1\cdots\mu_n}(x_1,\ldots,x_n) \tag{3.2.72}$$

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 (2.1.14) takes on the simple form

$$S \;=\; -\frac{1}{2g^2}\int d^d x \,\mathrm{tr}\,\mathcal{F}^2_{\mu\nu}(x) \,. \tag{3.2.73}$$

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 (1.3.63) 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.

Problem 3.16 Rederive the formula (3.2.71) using the calligraphic notations.
Solution 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. (3.2.73). The contribution of a generic graph is now of the order

$$\left(g^2\right)^{n_1 - n_0} N_c^{n_2 - B} \;\sim\; N_c^{n_2 - n_1 + n_0 - B} \tag{3.2.74}$$

for $g^2 \sim 1/N_c$. This coincides with the RHS of Eq. (3.2.67) which results in Eq. (3.2.71).

3.2.5 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. (3.2.60):

$$\left\langle \psi_i^\alpha \bar{\psi}_j^\beta \right\rangle \;\propto\; \delta^{\alpha\beta}\delta_{ij} \,. \tag{3.2.75}$$

Their contraction results in

$$\sum_{\alpha=1}^{N_f} \delta_{\alpha\alpha} = N_f \,. \tag{3.2.76}$$

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 $N_c \to \infty$, as was considered in the original paper by 't Hooft [Hoo74a], 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 on p. 164. 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

$$\frac{\Gamma_{\text{total}}}{M} \;\sim\; \frac{N_f}{N_c} \tag{3.2.77}$$

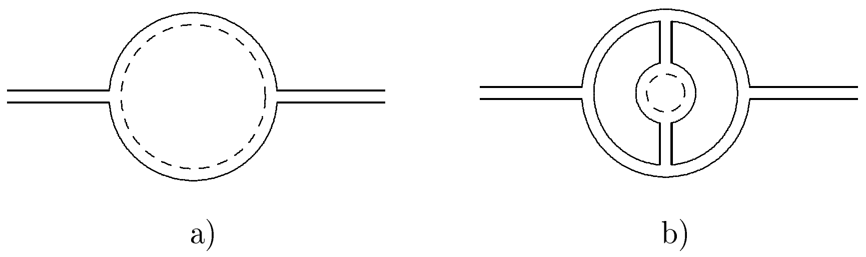


Fig. 3.21: Diagrams with quark loops in the Veneziano limit. The color and flavor indices of a quark loop which are represented by the solid and dashed single lines, respectively. The diagram a) is $\sim g^2 N_f \sim N_f/N_c$. The diagram b) is $\sim g^6 N_c^2 N_f \sim N_f/N_c$.

in the 't Hooft limit. The ratio on the LHS of Eq. (3.2.77) 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 $N_c \to \infty$ was proposed by Veneziano [Ven76]. Some diagrams for the gluon propagator, which involve one quark loop, are depicted in Fig. 3.21. 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. (3.2.76). This is analogous to the vector models which exactly describe the $\mathrm{O}(N_f)$ flavor symmetry in these notations.

The diagrams in Fig. 3.21 contribute, respectively,

$$\text{Fig. 3.21a} \sim g^2 N_f \sim \frac{N_f}{N_c} \tag{3.2.78}$$

and

$$\text{Fig. 3.21b} \sim g^6 N_c^2 N_f \sim \frac{N_f}{N_c} \tag{3.2.79}$$

in the limit (3.2.13). Likewise, a more general diagram with L quark loops will contribute

$$L \text{ quark loops} \sim \left(\frac{N_f}{N_c}\right)^L \left(\frac{1}{N_c^2}\right)^{\text{genus}} . \tag{3.2.80}$$

This formula obviously follows from Eq. (3.2.63) since each quark loop results in N_f.

We see from Eq. (3.2.80) that quark loops are not suppressed at large N_c in the Veneziano limit

$$N_f \sim N_c \to \infty \tag{3.2.81}$$

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 (3.2.81) 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. (3.2.77). I refer the reader to the original paper by Veneziano [Ven76] 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

$$\int D\bar\psi\, D\psi\, \mathrm{e}^{-\int d^4x \left(\bar\psi \widehat\nabla \psi + m\bar\psi\psi\right)} = \mathrm{e}^{\operatorname{Sp}\ln\left(\widehat\nabla + m\right)} \tag{3.2.82}$$

as is discussed in Subsect. 1.2.2. The trace in the exponent involves summation both over color and flavor indices, so that

$$\operatorname{Sp}\ln\left(\widehat\nabla + m\right) \sim N_c N_f . \tag{3.2.83}$$

The order in N_c of the pure gluon action is $\mathcal{O}\left(N_c^2\right)$ as is discussed in the Remark on p. 236. 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. (3.2.82) 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. (3.2.80). A diagram with two quark loops, which appears to the second order of this expansion, is depicted in Fig. 3.22.

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

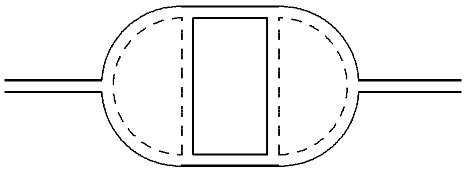


Fig. 3.22: Diagram with two quark loops in the Veneziano limit. The diagram is $\sim g^6 N_c N_f^2 \sim \left(N_f/N_c\right)^2$.

the $\mathcal{B}$-function of QCD with N_c colors and N_f flavors reads

$$b \;=\; -\frac{11}{3}N_c + \frac{2}{3}N_f\,. \tag{3.2.84}$$

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).

3.2.6 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. 3.1.5 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:

$$O\,(x) \;=\; \frac{1}{N_c^2}\,\mathrm{tr}\,F_{\mu\nu}^2\,(x)\,. \tag{3.2.85}$$

The normalizing factor provides the natural normalization

$$\left\langle \frac{1}{N_c^2}\,\mathrm{tr}\,F_{\mu\nu}^2\,(x)\right\rangle \;=\; \left\langle \frac{1}{2N_c^2}\,F_{\mu\nu}^a\,(x)\,F_{\mu\nu}^a\,(x)\right\rangle \;\sim 1\,. \tag{3.2.86}$$

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. 3.23.

The graph in Fig. 3.23a 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 (3.2.85). Its contribution is

$$\text{Fig. 3.23a} \;\sim\; \frac{1}{N_c^2}N_c^2 \times \frac{1}{N_c^2}N_c^2 \;\sim\; 1\,, \tag{3.2.87}$$

i.e. $\mathcal{O}\,(1)$ in accord with the general estimate (3.2.86).

The graph in Fig. 3.23b 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),

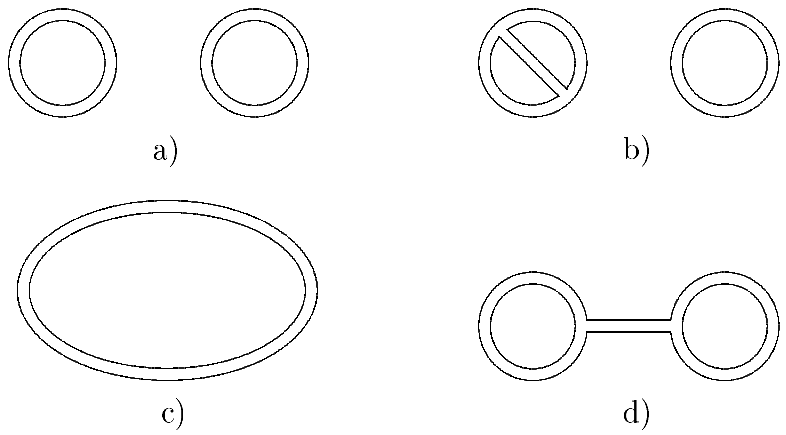


Fig. 3.23: Demonstration of the large-N_c factorization to the lowest orders of perturbation theory. The closed double line represents the average of the operator (3.2.85) to zeroth order in g. The diagrams a) and b), which are associated with the factorized part of the average on the LHS of Eq. (3.2.91), are $\mathcal{O}(1)$. The diagrams c) and d), which would violate the factorization, are suppressed by $1/N_c^2$.

the normalization factor $1/N_c^4$, and g^2 due to two three-gluon vertices. Its contribution is

$$\text{Fig. 3.23b} \sim g^2 N_c \sim 1\,, \tag{3.2.88}$$

i.e. also $\mathcal{O}(1)$ in the limit (3.2.13).

The graph in Fig. 3.23c is of the same type as the graph in Fig. 3.23a, 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

$$\text{Fig. 3.23c} \sim \frac{1}{N_c^2} \tag{3.2.89}$$

is suppressed by $1/N_c^2$.

The graph in Fig. 3.23d, which is of the same order in the coupling constant as the graph in Fig. 3.23b, involves only three closed index lines (the factor N_c^3) and is of order $1/N_c^2$:

$$\text{Fig. 3.23d} \sim g^2\frac{1}{N_c} \sim \frac{1}{N_c^2}\,. \tag{3.2.90}$$

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 $N_c \to \infty$. 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 $N_c \to \infty$:

$$\begin{aligned}
&\left\langle \frac{1}{N_c^2}\operatorname{tr} F^2(x_1)\,\frac{1}{N_c^2}\operatorname{tr} F^2(x_2)\right\rangle \\
&= \left\langle \frac{1}{N_c^2}\operatorname{tr} F^2(x_1)\right\rangle \left\langle \frac{1}{N_c^2}\operatorname{tr} F^2(x_2)\right\rangle + \mathcal{O}\left(\frac{1}{N_c^2}\right).
\end{aligned} \tag{3.2.91}$$

For a general set of gauge-invariant operators $O_1, \ldots, O_n$, the factorization property can be represented by

$$\langle O_1 \cdots O_n \rangle = \langle O_1 \rangle \cdots \langle O_n \rangle + \mathcal{O}\left(\frac{1}{N_c^2}\right). \tag{3.2.92}$$

This is analogous to Eq. (3.1.132) 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 [Wit79]. We shall discuss this in the next two Subsections.

The factorization property also holds for gauge-invariant operators constructed from quarks like in Eq. (3.2.64). For the case of several flavors N_f, we normalize these quark operators by

$$O_\Gamma = \frac{1}{N_f N_c}\bar\psi\Gamma\psi\,. \tag{3.2.93}$$

Here Γ stands for one of the combination of the gamma-matrices:

$$\Gamma = \mathrm{I},\ \gamma_5,\ \gamma_\mu,\ i\gamma_\mu\gamma_5,\ \Sigma_{\mu\nu} = \frac{1}{2i}[\gamma_\mu,\gamma_\nu]\,,\ldots\,. \tag{3.2.94}$$

The lowest-order diagrams of perturbation theory for the average of two quark operators (3.2.93) are depicted in Fig. 3.24. Estimation of their order in $1/N_c$ is analogous to that for the pure gluon graphs in Fig. 3.23.

The graph in Fig. 3.24a represents the zeroth order of perturbation theory for the average of two quark operators. It involves two closed color and

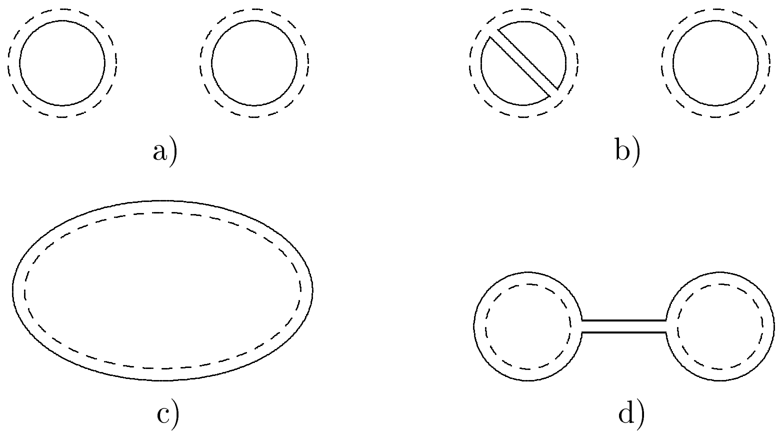


Fig. 3.24: Same as in Fig. 3.23 but for quark operators (3.2.93). The solid and dashed lines are associated with color and flavor indices, respectively. The diagram a), which contributes to the factorized part of the average on the LHS of Eq. (3.2.99), are $\mathcal{O}(1)$. The diagrams b) and c), which would violate the factorization, are suppressed by $1/N_c^2$ and $1/(N_f N_c)$, respectively.

two closed flavor index lines (the factor $N_f^2 N_c^2$) and the normalization factor $1/(N_f N_c)^2$ according to the definition (3.2.93). Its contribution is

$$\text{Fig. 3.24a} \sim \frac{1}{N_f^2 N_c^2} N_f^2 N_c^2 \sim 1\,. \tag{3.2.95}$$

This justifies the normalization factor in Eq. (3.2.93).

The graph in Fig. 3.24b involves a gluon line which is emitted and absorbed by the same quark operator. It has three closed color and two closed flavor index lines (the factor $N_f^2 N_c^3$), the normalization factor $1/(N_f N_c)^2$, and g^2 due to two quark-gluon vertices. Its contribution is

$$\text{Fig. 3.24b} \sim \frac{1}{N_f^2 N_c^2} N_f^2 g^2 N_c^3 \sim 1 \tag{3.2.96}$$

in full analogy with the pure gluon diagram in Fig. 3.23b.

The graph in Fig. 3.24c is similar to the graph in Fig. 3.23c — the lines connect two different quark operators. It has one closed color and one closed flavor index lines (the factor $N_f N_c$) and the normalization factor $1/(N_f N_c)^2$,

so that its contribution

$$\text{Fig. 3.24c} \sim \frac{1}{N_f N_c} \tag{3.2.97}$$

is suppressed by $1/(N_f N_c)$.

Finally, the graph in Fig. 3.24d involves a gluon line which is emitted by one quark operator and absorbed by the other. It has one closed color and two closed flavor index lines (factor $N_f^2 N_c$), the normalization factor $1/(N_f N_c)^2$, and g^2 due to two quark-gluon vertices. Its contribution

$$\text{Fig. 3.24d} \sim \frac{1}{N_f^2 N_c^2} N_f^2 g^2 N_c \sim \frac{g^2}{N_c} \sim \frac{1}{N_c^2} \tag{3.2.98}$$

is suppressed by $1/N_c^2$ in the limit (3.2.13).

We see that the factorization of the gauge-invariant quark operators holds both in the 't Hooft and Veneziano limits:

$$\langle\, O_{\Gamma_1} \cdots O_{\Gamma_n} \,\rangle = \langle\, O_{\Gamma_1} \,\rangle \cdots \langle\, O_{\Gamma_n} \,\rangle + \mathcal{O}\left(\frac{1}{N_f N_c}\right). \tag{3.2.99}$$

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 in item 3 on p. 241. The Veneziano limit is analogous to pure gluodynamics as is already mentioned.

It is worth noting that the factorization can be alternatively seen (at all orders of perturbation theory) from Eq. (3.2.71) 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. 3.2.4. The diagrams with gluon lines emitted and absorbed by the same operator like in Fig. 3.24b 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. 3.24d have two boundaries. According to Eq. (3.2.71), their contribution is suppressed by $1/N_c^2$. Alternatively, the diagrams like in Fig. 3.24c (including its planar dressing by gluons) have one boundary.[7] Their contribution is $\mathcal{O}(1)$ times $1/(N_f N_c)$ coming from the normalization of the operator (3.2.93). This proves the factorization property (3.2.99) at all orders of perturbation theory.

[7] In the dual-resonance model, they are associated with the meson-meson interaction due to an exchange of constituent quarks.

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 $\mathrm{SU}(N_c)$ 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).

Problem 3.17 Prove the factorization of the Wilson loop operators within the strong coupling expansion of the $\mathrm{SU}(N_c)$ lattice gauge theory as $N_c \to \infty$.
Solution Let us first estimate the order in N_c of the Wilson loop average (2.2.42). The explicit result to the leading order in β is given by Eqs. (2.2.73), (2.2.72), where

$$\beta \;\sim\; N_c^2 \tag{3.2.100}$$

in the limit (3.2.13) as is prescribed by Eq. (2.2.32). Therefore, $W_C \sim 1$ in the large-N_c limit.

To be precise, we first perform the strong coupling expansion in β and then set $N_c \to \infty$ in each term of the strong coupling expansion. As we shall see in a moment, the actual parameter is β/N_c^2, so that the limits $\beta \to 0$ and $N_c \to \infty$ are interchangeable.

It is easy to estimate the order in N_c of any graph of the strong coupling expansion for W_C, which looks like that in Fig. 2.10. Let the plaquettes fill an arbitrary surface enclosed by the loop C, with n_2, n_1, and n_0 being the number of plaquettes, links, and sites which belong to the surface. Each plaquette contributes β/N_c since it comes from the expansion of the exponential of the lattice action, each link contributes $1/N_c$ due to Eq. (2.2.60), and each site contributes N_c since it is associated with summing over the color indices due to Eq. (2.2.70). Accounting for the normalization factor $1/N_c^B$ where $B = 1$ is the number of boundaries, the contribution is of order

$$\begin{aligned} \left(\frac{\beta}{N_c}\right)^{n_2} N_c^{-n_1+n_0-B} \;&\sim\; \left(\frac{\beta}{N_c^2}\right)^{n_2} N_c^{n_2-n_1+n_0-B} \\ \sim\; \left(\frac{\beta}{N_c^2}\right)^{n_2} N_c^{n_2-n_1+n_0-B} \;&\sim\; \left(\frac{\beta}{N_c^2}\right)^{n_2} \left(\frac{1}{N_c}\right)^{2h+2(B-1)}, \end{aligned} \tag{3.2.101}$$

where we have used Euler's theorem (3.2.69). In the limit (3.2.100), the contribution does not depend on the order of the strong coupling expansion and is completely determined by the number B of boundaries and the genus h of the surface. This is analogous to the perturbation theory. For the minimal surface, we reproduce previous results.

We are now in a position to analyze the order in N_c of different terms in the strong coupling expansion of the average of two Wilson loop operators. The factorized part results from the surfaces of the type depicted in Fig. 3.25a which are

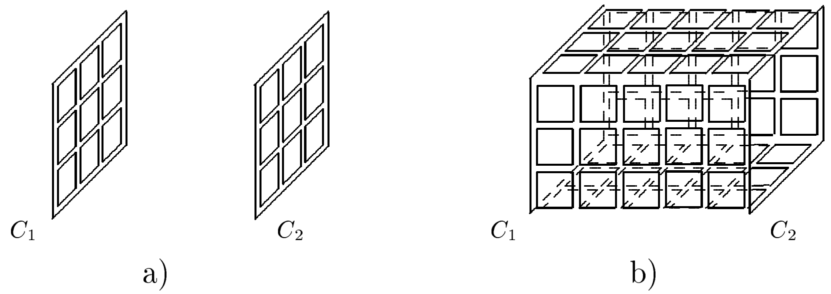


Fig. 3.25: Factorization of the Wilson loop operators in the strong coupling expansion as $N_c \to \infty$. The surfaces are constructed from plaquettes which come from the expansion of the exponential of the lattice action. Each link in the surface is passed at least twice: otherwise the result vanishes. a) involves two separate surfaces encloses by the Wilson loops. It contributes to the factorized part of the average on the LHS of Eq. (3.2.102). The surface in b) connects two different Wilson loops and would violate factorization. It has two boundaries and its contribution is suppressed by $1/N_c^2$ according to the general formula (3.2.101).

spanned by each individual loop. Its contribution is $\mathcal{O}(1)$ as $N_c \to \infty$. A non-factorized part emerges from surfaces of the type depicted in Fig. 3.25b, which connect two different Wilson loops. They look like a cylinder and have two boundaries. Their contribution is suppressed by $1/N_c^2$ according to the general formula (3.2.101).

Thus, we have proven the factorization property

$$\left\langle \frac{1}{N_c}\mathrm{tr}\, U_{C_1} \frac{1}{N_c}\mathrm{tr}\, U_{C_2} \right\rangle \;=\; \left\langle \frac{1}{N_c}\mathrm{tr}\, U_{C_1} \right\rangle \left\langle \frac{1}{N_c}\mathrm{tr}\, U_{C_2} \right\rangle + \mathcal{O}\left(\frac{1}{N_c^2}\right) \tag{3.2.102}$$

at all orders of the strong coupling expansion.

Problem 3.18 Find the relation between the Wilson loop averages in the fundamental and adjoint representations for a $\mathrm{SU}(N_c)$ pure gauge theory at large N_c.
Solution The characters in the fundamental and adjoint representations are related by Eq. (2.2.28). Using the factorization formula (3.2.102) with coinciding contours C_1 and C_2, we get

$$W_{\mathrm{adj}}(C) \;=\; \left[W_{\mathrm{fun}}(C)\right]^2 + \mathcal{O}\left(\frac{1}{N_c^2}\right). \tag{3.2.103}$$

As is discussed in Chapter 2, the Wilson loop average in the fundamental representation obeys area law (2.2.75). The same is true at $N_c = \infty$ for the Wilson loop average in the adjoint representation due to Eq. (3.2.103). In particular, the string tensions in the fundamental and adjoint representations at $N_c = \infty$ are related by

$$K_{\rm adj} \;=\; 2K_{\rm fun}\,. \tag{3.2.104}$$

On the other hand, the adjoint test quark can be screened at finite N_c by a gluon produced out of the vacuum. This is similar to the breaking of the flux tube in the fundamental representation by a quark-antiquark pair, which is discussed in Subsect. 2.5.5. Therefore, the perimeter law (2.2.79) must dominate for large contours. The point is that the perimeter law appears due to connected diagrams which are suppressed as $1/N_c^2$:

$$W_{\rm adj}(C) \stackrel{\text{large } C}{\longrightarrow} \;\; {\rm e}^{\,-2KA_{\rm min}(C)} + \frac{1}{N_c^2}\,{\rm e}^{\,-\mu\cdot L(C)}\,. \tag{3.2.105}$$

These properties of the adjoint representation were first pointed out in Ref. [KM81] cited in Chapter 2.

3.2.7 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. 3.1.5 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

$$Z \;=\; \int DA_\mu^a\,{\rm e}^{\,-\int d^4x\,\frac14 F_{\mu\nu}^a F_{\mu\nu}^a}\,. \tag{3.2.106}$$

The action, $\sim N_c^2$, is large as $N_c \to \infty$, but the "entropy" is also $\sim N_c^2$ due to the $N_c^2 - 1$ integrations over A_μ^a:

$$DA_\mu^a \;\sim\; {\rm e}^{\,N_c^2}\,. \tag{3.2.107}$$

Consequently, the saddle-point equation of the large-N_c Yang–Mills theory is *not* the classical one which reads[8]

$$\frac{\delta S}{\delta A_\mu^a} \;=\; \left(\nabla_\nu F_{\mu\nu}\right)^a = 0\,. \tag{3.2.108}$$

[8]It was already discussed in Problem 2.1.

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. 3.1.4 for the sigma-model. The expected new path-integral representation of the partition function (3.2.106) would be something like

$$Z \;\propto\; \int D\Phi\,\frac{1}{\frac{\partial\Phi[A]}{\partial A_\mu^a}}\,{\rm e}^{\,-N_c^2 S[\Phi]}\,. \tag{3.2.109}$$

The Jacobian

$$\frac{\partial\Phi\,[A]}{\partial A_\mu^a} \;\equiv\; {\rm e}^{\,-N_c^2 J[\Phi]} \tag{3.2.110}$$

in Eq. (3.2.109) is related to the old entropy factor, so that

$$J\,[\Phi] \;\sim\; 1 \tag{3.2.111}$$

in the large-N_c limit.

The original partition function (3.2.106) can be then rewritten as

$$Z \;\propto\; \int D\Phi\,{\rm e}^{\,N_c^2 J[\Phi] - N_c^2 S[\Phi]}, \tag{3.2.112}$$

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. (3.2.112) only in the exponent. Therefore, the saddle-point equation can be immediately written:

$$\frac{\delta S}{\delta \Phi} \;=\; \frac{\delta J}{\delta \Phi}\,. \tag{3.2.113}$$

Remembering that Φ is a functional of A_μ:

$$\Phi \;\equiv\; \Phi\,[A]\,, \tag{3.2.114}$$

we rewrite the saddle-point equation (3.2.113) as

$$\frac{\delta S}{\delta A_\nu^a} \;=\; \left(\nabla_\mu F_{\mu\nu}\right)^a \;=\; \frac{\delta J}{\delta A_\nu^a}\,. \tag{3.2.115}$$

It differs from the classical Yang–Mills equation (3.2.108) by the term on the RHS coming from the Jacobian (3.2.110).

Given $J[\Phi]$ which depends on the precise from of the variable $\Phi[A]$, Eq. (3.2.115) has a solution

$$A_\mu(x) \;=\; A_\mu^{\rm cl}(x)\,. \tag{3.2.116}$$

Let us first assume that there exists only one solution to Eq. (3.2.115). Then the path integral is saturated by a single configuration (3.2.116), so that the vacuum expectation values of gauge-invariant operators are given by their values at this configuration:

$$\langle\, O\,\rangle \;=\; O\left(A_\mu^{\rm cl}(x)\right). \tag{3.2.117}$$

The factorization property (3.2.92) will obviously be satisfied.

An existence of such a classical field configuration in multicolor QCD was conjectured by Witten [Wit79]. It was discussed in the lectures by Coleman [Col79] who called it the *master field*. Equation (3.2.115) 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. (3.2.115) 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. (3.2.115). However, this will not change Eq. (3.2.117) 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. (3.2.117) does. This implies that $A_\mu^{\rm cl}(x)$ must itself be Poincaré invariant up to a gauge transformation: a change of $A_\mu^{\rm cl}(x)$ under translations or rotations can be compensated by a gauge transformation. Moreover, there must exist a gauge in which $A_\mu^{\rm cl}(x)$ is space-time independent:

$$A_\mu^{\rm cl}(x) \;=\; A_\mu^{\rm cl}(0)\,. \tag{3.2.118}$$

In this gauge, rotations must be equivalent to a global gauge transformation, so that $A_\mu^{\rm cl}(0)$ 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 [Haa81]. The conjecture about an existence of only one solution to the master-field equation (3.2.115) seems to be somewhat incorrect. If several solutions exists, 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 $A_\mu^{\rm cl}(0)$ would be an operator in some Hilbert space rather than a c-valued function. This is simply because $A_\mu^{\rm cl}(0)$ is, in the matrix notation (3.2.1), a $N_c\times N_c$ matrix which becomes, as $N_c\to\infty$, 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. (3.2.115), 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 $A_\mu^{\rm cl}(0)$ 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 [VDN95]. 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. [Dou95, GG95].

3.2.8 $1/N_c$ 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

$$\Phi(C) \;\equiv\; \Phi[A] = \frac{1}{N_c}\,{\rm tr}\,{\bf P}\,{\rm e}^{\,ig\oint_C dx^\mu A_\mu(x)}. \tag{3.2.119}$$

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 [Man79]), and the Jacobian is huge. The reformulation was done [MM79] 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 are given by Eq. (1.2.47).

The RHS of Eq. (1.2.47) is proportional to the Planck's constant $\hbar$ as is explained in Subsect 1.2.5. In the semiclassical limit $\hbar \to 0$, we get

$$\left(-\partial_1^2 + m^2\right) \langle \varphi(x_1) \ldots \varphi(x_n) \rangle + \frac{\lambda}{2} \langle \varphi^2(x_1) \ldots \varphi(x_n) \rangle = 0\,, \quad (3.2.120)$$

whose solution is of the factorized form

$$\langle \varphi(x_1) \ldots \varphi(x_n) \rangle = \langle \varphi(x_1) \rangle \ldots \langle \varphi(x_n) \rangle + \mathcal{O}(\hbar) \quad (3.2.121)$$

provided that

$$\langle \varphi(x) \rangle \equiv \varphi_{\rm cl}(x) \quad (3.2.122)$$

obeys

$$\left(-\partial^2 + m^2\right) \varphi_{\rm cl}(x) + \frac{\lambda}{2} \varphi_{\rm cl}^2(x) = 0\,. \quad (3.2.123)$$

Equation (3.2.123) is nothing but the classical equation of motion for the φ^3 theory, which specifies extrema of the action (1.2.22) entering the path integral (1.2.2). Thus, we have reproduced, using the Schwinger–Dyson equations, the well-known fact that the path integral is dominated by a classical trajectory as $\hbar \to 0$. It is also clear how to perform the semiclassical expansion in $\hbar$ in the language of the Schwinger–Dyson equations: one should solve Eq. (1.2.47) 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

$$U[C_{xx}] = \Omega[C_{xx}]\, {\rm diag}\left({\rm e}^{\,ig\alpha_1(C)}, \ldots, {\rm e}^{\,ig\alpha_{N_c}(C)}\right) \Omega^\dagger[C_{xx}]\,. \quad (3.2.124)$$

Then $\Phi(C)$ reads

$$\Phi(C) = \frac{1}{N_c} \sum_{j=1}^{N_c} {\rm e}^{\,ig\alpha_j(C)}\,. \quad (3.2.125)$$

The phases $\alpha_j(C)$ are gauge invariant and normalized so that $\alpha_j(C) \sim 1$ as $N_c \to \infty$. For simplicity we omit below all the indices (including space ones) except color.

The commutator of Φ's can be estimated using the representation (3.2.125). Since

$$[\alpha_i, \alpha_j] \propto \delta_{ij}\,, \quad (3.2.126)$$

one gets

$$[\Phi(C), \Phi(C')] \sim g^2 \frac{1}{N_c} \sim \frac{1}{N_c^2} \quad (3.2.127)$$

in the limit (3.2.13), *i.e.* the commutator can be neglected as $N_c \to \infty$, and the field $\Phi(C)$ becomes classical.

Note that the commutator (3.2.127) is of order $1/N_c^2$. One factor $1/N_c$ is because of g in the definition (3.2.125) of $\Phi(C)$, while the other has a deep reason. Let us image the summation over j in Eq. (3.2.125) as some statistical averaging. It is well-known in statistics that such averages weakly fluctuate as $N_c \to \infty$, so that the dispersion is of order $1/N_c$. It is the factor which emerges in the commutator (3.2.127).

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. (3.2.125). 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 (3.2.126) 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. [Haa81].

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.

3.3 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 the 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 $N_c \to \infty$ and discuss its various properties, including a non-perturbative regularization. Finally, we briefly comment on what is known about solutions of the loop equation.

3.3.1 Observables in terms of Wilson loops

All observables in QCD can be expressed via the Wilson loops $\Phi(C)$ defined by Eq. (3.2.119). This property was first advocated by Wilson [Wil74] 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.

At finite N_c, observables are expressed via the n-loop averages

$$W_n(C_1,\ldots,C_n) \;=\; \langle\, \Phi(C_1)\cdots\Phi(C_n)\,\rangle\,, \qquad (3.3.1)$$

which are analogous to the n-point Green functions (1.2.45). The appropriate formulas for the continuum theory can be found in Ref. [MM81].

Great simplifications occur in these formulas at $N_c = \infty$, when all observables are expressed only via the one-loop average

$$W(C) \;=\; \langle\, \Phi(C)\,\rangle \equiv \left\langle \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\mathrm{e}^{ig\oint_C dx^\mu A_\mu} \right\rangle . \qquad (3.3.2)$$

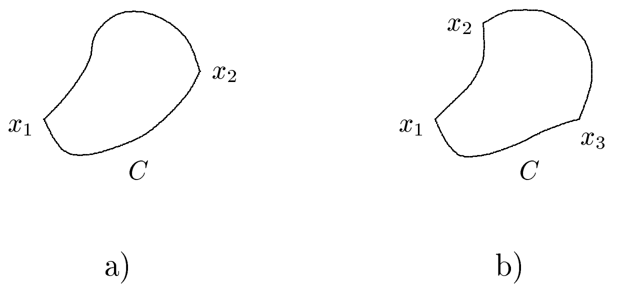


Fig. 3.26: Contours in the sum over paths representing observables: a) in Eq. (3.3.3) and b) in Eq. (3.3.4). The contour a) passes two nailed points x_1 and x_2. The contour b) passes three nailed points x_1, x_2, and x_3.

This is associated with the quenched approximation discussed on p. 164.

For example, the average of the product of two colorless quark vector currents (3.2.93) is given at large N_c by

$$\left\langle \bar\psi\gamma_\mu\psi(x_1)\,\bar\psi\gamma_\nu\psi(x_2)\right\rangle \;=\; \sum_{C\ni x_1,x_2} J_{\mu\nu}(C)\,\langle\,\Phi(C)\,\rangle\,, \qquad (3.3.3)$$

where the sum runs over contours C passing through the points x_1 and x_2 as is depicted in Fig. 3.26a. An analogous formula for the (connected) correlators of three quark scalar currents reads

$$\left\langle \bar\psi\psi(x_1)\,\bar\psi\psi(x_2)\,\bar\psi\psi(x_3)\right\rangle_{\mathrm{conn}} \;=\; \sum_{C\ni x_1,x_2,x_3} J(C)\,\langle\,\Phi(C)\,\rangle\,, \qquad (3.3.4)$$

where the sum runs over contours C passing through the three points x_1, x_2, and x_3 as is depicted in Fig. 3.26b. A general (connected) correlator of n quark currents is given by a similar formula with C passing through n points $x_1, \ldots, x_n$ (some of them may coincide).

The weights $J_{\mu\nu}(C)$ in Eq. (3.3.3) and $J(C)$ in Eq. (3.3.4) are completely determined by free theory. If quarks were scalars rather than spinors, then we would get

$$J(C) \;=\; \mathrm{e}^{-\frac12 m^2\tau-\frac12\int_0^\tau dt\,\dot z_\mu^2(t)} \;=\; \mathrm{e}^{-mL(C)} \qquad \boxed{\text{scalar quarks}}, \qquad (3.3.5)$$

where $L(C)$ is the length of the (closed) contour C, as was shown in Subsect. 1.1.6. Using the notation (1.1.152), we can rewrite Eq. (3.3.4) for scalar quarks as

$$\langle \psi^\dagger\psi(x_1)\,\psi^\dagger\psi(x_2)\,\psi^\dagger\psi(x_3)\rangle_{\rm conn} = \sum_{C\ni x_1,x_2,x_3}{}' \langle \Phi(C)\rangle . \tag{3.3.6}$$

Therefore, we get the sum over paths of the Wilson loop, likewise in Subsect. 1.1.7 and Problem 2.5.

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 $k_\mu(t)$ $(0\le t\le\tau)$ being an appropriate trajectory. The result reads[9]

$$J(C) = \int Dk(t)\;{\rm sp}\,\mathbf{P}\,{\rm e}^{-\int_0^\tau dt[ik_\mu(t)(\dot x_\mu(t)-\gamma_\mu(t))+m]} \tag{3.3.7}$$

and

$$J_{\mu\nu}(C) = \int Dk(t)\;{\rm sp}\,\mathbf{P}\left\{\gamma_\mu(t_1)\,\gamma_\nu(t_2)\,{\rm e}^{-\int_0^\tau dt[ik_\mu(t)(\dot x_\mu(t)-\gamma_\mu(t))+m]}\right\}, \tag{3.3.8}$$

where the values t_1 and t_2 of the parameter t are associated with the points x_1 and x_2 in Eq. (3.3.3), and the symbol of P–ordering puts the matrices γ_μ and γ_ν at a proper order.

Problem 3.19 Derive Eqs. (3.3.7) and (3.3.8).

Solution Since the spinor field ψ enters the QCD action quadratically, it can be integrated out in the correlators (3.3.3), (3.3.4), so that they can be represented, in the first quantized language, via the resolvent of the Dirac operator in the external field $\mathcal{A}_\mu$, with subsequent averaging over $\mathcal{A}_\mu$. Proceeding as in Sect. 1.1, we express the resolvent by

$$\left\langle y\left|\frac{1}{\widehat\nabla+m}\right|x\right\rangle = \int_0^\infty d\tau\,\left\langle y\left|{\rm e}^{-\tau\left(\widehat\nabla+m\right)}\right|x\right\rangle \tag{3.3.9}$$

and represent the matrix element of the exponential of the Dirac operator as

$$\left\langle y\left|{\rm e}^{-\tau\left(\widehat\nabla+m\right)}\right|x\right\rangle = {\rm e}^{-\tau m}\,\mathbf{P}\,{\rm e}^{-\int_0^\tau dt\,\widehat\nabla(t)}\delta^{(d)}(x-y)\,. \tag{3.3.10}$$

In order to disentangle the RHS, we insert the unity represented by

$$1 = \int_{z(0)=x} Dz(t)\int Dp(t)\,{\rm e}^{-i\int_0^\tau dt\,p_\mu(t)\dot z_\mu(t)}, \tag{3.3.11}$$

[9] See, *e.g.*, Ref. [BNZ79].

where the path integration over $p_\mu(t)$ is unrestricted, *i.e.* the integrals over $p_\mu(0)$ and $p_\mu(\tau)$ are included. Then we get

$$\begin{aligned}\left\langle y\left|{\rm e}^{-\tau\left(\widehat\nabla+m\right)}\right|x\right\rangle &= {\rm e}^{-\tau m}\int_{z(0)=x} Dz(t)\int Dp(t)\\ &\times\mathbf{P}\,{\rm e}^{-\int_0^\tau dt\left[ip_\mu(t)\dot z_\mu(t)-\left(ip_\mu(t)+\mathcal{A}_\mu(t)\right)\gamma_\mu(t)+\partial_\mu(t)\dot z_\mu(t)\right]}\delta^{(d)}(x-y)\,,\end{aligned} \tag{3.3.12}$$

whose equivalence to the original expression is obvious since everything commutes under the sign of the P–ordering (so that we can substitute $p_\mu(t)=-i\partial_\mu(t)$ in the integrand).

By making the change of the integration variable

$$p_\mu(t) = k_\mu(t)+i\mathcal{A}_\mu(t) \tag{3.3.13}$$

and proceeding as in Problem 1.12, we represent the RHS of Eq. (3.3.12) by

$$\begin{aligned}\left\langle y\left|{\rm e}^{-\tau\left(\widehat\nabla+m\right)}\right|x\right\rangle &= {\rm e}^{-\tau m}\int_{z(0)=x} Dz(t)\int Dk(t)\\ &\times\mathbf{P}\,{\rm e}^{-\int_0^\tau dt\left[ik_\mu(t)\left(\dot z_\mu(t)-\gamma_\mu(t)\right)-\dot z_\mu(t)\mathcal{A}_\mu(t)+\partial_\mu(t)\dot z_\mu(t)\right]}\delta^{(d)}(x-y)\\ &= {\rm e}^{-\tau m}\int_{\substack{z(0)=x\\ z(\tau)=y}} Dz(t)\int Dk(t)\,\mathbf{P}\,{\rm e}^{\int_x^y dz_\mu\,\mathcal{A}_\mu(z)}\\ &\times\mathbf{P}\,{\rm e}^{-i\int_0^\tau dt\,k_\mu(t)\left[\dot z_\mu(t)-\gamma_\mu(t)\right]},\end{aligned} \tag{3.3.14}$$

where the first P–exponential on the RHS depends only on color matrices (it is nothing but the non-Abelian phase factor), and the second one depends only on spinor matrices. In Ref. [BNZ79], Eq. (3.3.14) is derived by discretizing paths.

Equation (3.3.14) leads to Eqs. (3.3.7) and (3.3.8).

Remark on renormalization of Wilson loops

Perturbation theory for $W(C)$ can be obtained by expanding the path-ordered exponential in the definition (3.3.2) in g (see Eq. (3.2.47)) 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. 3.27a reads

$$W(C) = {\rm e}^{-g^2\frac{(N_c^2-1)}{8\pi N_c}\frac{L(C)}{a}}\,W_{\rm ren}(C)\,, \tag{3.3.15}$$

a) b)

Fig. 3.27: Examples of a) smooth contour and b) contour with a cusp. The tangent vector to the contour jumps through angle γ at the cusp.

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 on p. 117. 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. [GN80, Pol80, DV80].

If the contour C has a cusp (or cusps) but no self-intersections as is illustrated by Fig. 3.27b, then $W\,(C)$ is still multiplicatively renormalizable [BNS81]:

$$W\,(C) \;=\; Z\,(\gamma)\, W_{\text{ren}}\,(C)\,, \tag{3.3.16}$$

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}.

Problem 3.20 Calculate divergent parts of the Wilson loop average (3.3.2) for contours without self-intersections to order g^2. Consider the cases of a smooth contour C and a contour with a cusp.

Solution Expanding the Wilson loop average (3.3.2) in g^2 (see Eq. (3.2.47) and Problem 2.2), we get

$$W(C) \;=\; 1 + W^{(2)}(C) + \mathcal{O}(g^4) \tag{3.3.17}$$

with

$$W^{(2)}(C) \;=\; -\,g^2\frac{(N_c^2-1)}{4N_c}\oint\limits_C dx_\mu \oint\limits_C dy_\nu\, D_{\mu\nu}(x-y)\,, \tag{3.3.18}$$

where $D_{\mu\nu}(x-y)$ is the gluon propagator (3.2.4).

Since the contour integral in Eq. (3.3.18) diverges for $x = y$, we introduce the regularization by

$$D_{\mu\nu}(x-y) \;\stackrel{\text{Reg.}}{\Longrightarrow}\; \frac{1}{4\pi^2}\frac{\delta_{\mu\nu}}{(x-y)^2+a^2} \tag{3.3.19}$$

with a being the ultraviolet cutoff. Parametrizing the contour C by the function $z_\mu(\sigma)$, we rewrite the contour integral in Eq. (3.3.18) as

$$\oint\limits_C dx_\mu \oint\limits_C dy_\mu\, \frac{1}{(x-y)^2+a^2} \;=\; \int ds \int dt\, \frac{\dot z_\mu(s)\dot z_\mu(s+t)}{\left[z(s+t)-z(s)\right]^2+a^2}\,. \tag{3.3.20}$$

Choosing the proper-length parametrization (1.1.97) when

$$\dot z_\mu(s)\ddot z_\mu(s) \;=\; 0\,, \tag{3.3.21}$$

expanding in powers of t, and assuming that the contour C is smooth as is depicted in Fig. 3.27a, we get for the integral (3.3.20)

$$\int ds\, \dot x^2(s)\int dt\, \frac{1}{\dot x^2(s)t^2+a^2} = \frac{\pi}{a}\int ds\, \sqrt{\dot x^2(s)} = \frac{\pi}{a}L(C)\,. \tag{3.3.22}$$

Typical values of t in the last integral are $\sim a$, which justifies the expansion in t: next terms lead to a finite contribution as $a \to 0$.

Thus, we get

$$W^{(2)}(C) \;=\; -\,g^2\frac{(N_c^2-1)}{8\pi N_c}\frac{L(C)}{a} + \text{ finite as } a\to 0 \tag{3.3.23}$$

for a smooth contour. This is precisely the renormalization of the mass of a heavy test quark due to the interaction.

If the contour C is not smooth and has a cusp at some value s_0 of the parameter, as is depicted in Fig. 3.27b, then an extra divergent contribution in the integral (3.3.20) emerges when $s \approx s_0$, $t \approx t_0$. Introducing $\Delta s = s - s_0$ and $\Delta t = t - t_0$, we represent this extra divergent term by

$$\begin{aligned}&\dot x_\mu(s_0+0)\dot x_\mu(s_0-0)\int d\Delta s\int d\Delta t\, \frac{1}{\left[\dot x_\mu(s_0+0)\Delta s - \dot x_\mu(s_0-0)\Delta t\right]^2+a^2}\\ &= (\gamma\cot\gamma - 1)\ln\frac{L(C)}{a}\,,\end{aligned} \tag{3.3.24}$$

where γ is the angle of the cusp ($\cos\gamma \equiv \dot x_\mu(s_0+0)\dot x_\mu(s_0-0)$) and the upper limit of the integrations is chosen to be $L(C)$ with logarithmic accuracy. Collecting all together, we get finally for the divergent part of $W^{(2)}$:

$$\begin{aligned}W^{(2)} \;&=\; -g^2\frac{(N_c^2-1)}{8\pi N_c}\left(\frac{L(C)}{a} + \frac{1}{\pi}(\gamma\cot\gamma-1)\ln\frac{L(C)}{a}\right)\\ &\quad + \text{ finite as } a\to 0\,.\end{aligned} \tag{3.3.25}$$

The second term in this formula is associated with the bremsstrahlung radiation of a particle changing its velocity when passing the cusp. The answers in the Abelian and non-Abelian cases coincide to this order in g^2.

3.3.2 Schwinger–Dyson equations for Wilson loop

Dynamics of (quantum) Yang–Mills theory is described by the quantum equation of motion

$$\nabla^{ab}_\mu F^b_{\mu\nu}(x) \stackrel{\text{w.s.}}{=} \hbar \frac{\delta}{\delta A^a_\nu(x)} \tag{3.3.26}$$

which is analogous to Eq. (1.2.27) for the scalar field, and is again understood in the weak sense, *i.e.* for the averages

$$\left\langle \nabla^{ab}_\mu F^b_{\mu\nu}(x)\, Q[A] \right\rangle = \hbar \left\langle \frac{\delta}{\delta A^a_\nu(x)} Q[A] \right\rangle . \tag{3.3.27}$$

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. (3.2.45).

Strictly speaking, the last statement is incorrect, since we have not added, in Eqs. (3.3.26) and (3.3.27), 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 (2.1.5), when Eq. (3.3.26) for the Wilson loop takes on the form

$$\begin{aligned} &\left\langle \frac{1}{N_c} \operatorname{tr} \mathbf{P}\, \nabla_\mu \mathcal{F}_{\mu\nu}(x)\, \mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} \right\rangle \\ &= \left\langle \frac{g^2}{2N_c} \operatorname{tr} \frac{\delta}{\delta \mathcal{A}_\nu(x)} \mathbf{P}\, \mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} \right\rangle , \end{aligned} \tag{3.3.28}$$

where we have restored the units with $\hbar = 1$.

The variational derivative on the RHS can be calculated by virtue of the formula

$$\frac{\delta A^{ij}_\mu(y)}{\delta A^{kl}_\nu(x)} = \delta_{\mu\nu}\, \delta^{(d)}(x-y) \left(\delta^{il}\delta^{kj} - \frac{1}{N_c}\delta^{ij}\delta^{kl} \right) \tag{3.3.29}$$

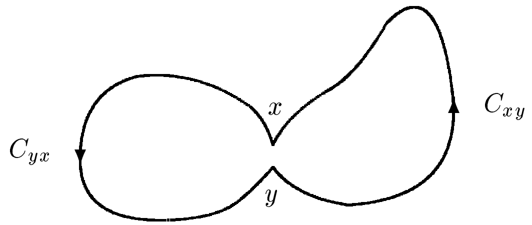


Fig. 3.28: Contours C_{yx} and C_{xy} which enter the RHS's of Eqs. (3.3.28) and (3.3.32).

which is a consequence of

$$\frac{\delta A^a_\mu(y)}{\delta A^b_\nu(x)} = \delta_{\mu\nu}\, \delta^{(d)}(x-y)\, \delta^{ab} . \tag{3.3.30}$$

The second term in the parentheses in Eq. (3.3.29) — same as in Eq. (3.2.6) — is because $\mathcal{A}_\mu$ is a matrix from the adjoint representation of SU(N_c).

By using Eq. (3.3.29), we get for the variational derivative on RHS of Eq. (3.3.28):

$$\begin{aligned} \operatorname{tr} \frac{\delta}{\delta \mathcal{A}_\nu(x)} \mathbf{P}\, \mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} &= \oint_C dy_\nu\, \delta^{(d)}(x-y)\, \times \\ &\left[\frac{1}{N_c} \operatorname{tr} \mathbf{P}\, \mathrm{e}^{\int_{C_{yx}} d\xi^\mu \mathcal{A}_\mu} \frac{1}{N_c} \operatorname{tr} \mathbf{P}\, \mathrm{e}^{\int_{C_{xy}} d\xi^\mu \mathcal{A}_\mu} - \frac{1}{N_c^3} \operatorname{tr} \mathbf{P}\, \mathrm{e}^{\int_C d\xi^\mu \mathcal{A}_\mu} \right] . \end{aligned} \tag{3.3.31}$$

The contours C_{yx} and C_{xy}, which are depicted in Fig. 3.28, 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. (3.3.28) as

$$\begin{aligned} &\left\langle \frac{1}{N_c} \operatorname{tr} \mathbf{P}\, \nabla_\mu \mathcal{F}_{\mu\nu}(x)\, \mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} \right\rangle \\ &= \lambda \oint_C dy_\nu\, \delta^{(d)}(x-y) \left[\langle \Phi(C_{yx})\, \Phi(C_{xy}) \rangle - \frac{1}{N^2} \langle \Phi(C) \rangle \right] \end{aligned} \tag{3.3.32}$$

where we have introduced

$$\lambda = \frac{g^2 N_c}{2} . \tag{3.3.33}$$

Notice that the RHS of Eq. (3.3.32) is completely represented via the (closed) Wilson loops.

Problem 3.21 Prove the cancellation of the contributions of the gauge-fixing and ghost terms in the Lorentz gauge.
Solution The Yang–Mills action, associated with the Lorentz gauge, reads

$$S_{\text{g.f.}} = -\frac{1}{g^2}\int d^d x \left[\frac{1}{2}\operatorname{tr}\mathcal{F}_{\mu\nu}^2 + \frac{1}{\alpha}\operatorname{tr}\left(\partial_\mu \mathcal{A}_\mu\right)^2\right]. \tag{3.3.34}$$

Since $\Phi(C)$ is gauge invariant, the (infinite) group-volume factors, in the numerator and denominator in the definition of the average, cancel when fixing the gauge (see the Remark on p. 113), and we get

$$\begin{aligned} W(C) &\equiv \frac{\int DA_\mu\, \mathrm{e}^{-S}\,\Phi(C)}{\int DA_\mu\, \mathrm{e}^{-S}} \\ &= \frac{\int DA_\mu\, \det\partial_\mu\nabla_\mu\, \mathrm{e}^{-S_{\text{g.f.}}}\,\Phi(C)}{\int DA_\mu\, \det\partial_\mu\nabla_\mu\, \mathrm{e}^{-S_{\text{g.f.}}}}, \end{aligned} \tag{3.3.35}$$

where $\det\partial_\mu\nabla_\mu$ is associated with ghosts.

The Schwinger–Dyson equation for the Yang–Mills theory in the Lorentz gauge is

$$\begin{aligned} \nabla_\mu^{ab} F_{\mu\nu}^b(x) &\overset{\text{w.s.}}{=} \hbar\frac{\delta}{\delta A_\nu^a(x)} \\ &+\frac{1}{\alpha}\partial_\nu\partial_\mu A_\mu^a + \frac{\partial}{\partial x_\nu} f^{abc} G^{bc}\left(x'=x,x;A\right), \end{aligned} \tag{3.3.36}$$

where $G^{bc}(x',x;A)$ is the Green function of the ghost in an external field A_μ. Applying this equation to the Wilson loop and using the gauge Ward identity (the Slavnov–Taylor identity), we transform the contribution from the second term on the RHS to

$$\begin{aligned} &\frac{1}{\alpha}\left\langle\frac{1}{N_c}\operatorname{tr}\partial_\nu\partial_\mu\mathcal{A}_\mu U(C_{xx})\right\rangle_{\text{g.f.}} \\ &= g^2\oint_C d\xi_\mu\left\langle\frac{1}{N_c}\operatorname{tr}\left[U(C_{\xi x})t^a U(C_{x\xi})t^b\right]\frac{\partial}{\partial x_\nu}\nabla^{bc}(\xi)\,G^{ca}(\xi,x;A)\right\rangle_{\text{g.f.}} \\ &= g^2\oint_C d\xi_\mu\frac{\partial}{\partial\xi_\mu}\left\langle\frac{1}{N_c}\operatorname{tr}\left[U(C_{\xi x})t^a U(C_{x\xi})t^b\right]\frac{\partial}{\partial x_\nu}G^{ba}(\xi,x;A)\right\rangle_{\text{g.f.}} \\ &= g^2\left\langle\frac{1}{N_c}\operatorname{tr}U(C_{xx})\left[t^a,t^b\right]\frac{\partial}{\partial x_\nu}G^{ba}\left(x'=x,x;A\right)\right\rangle_{\text{g.f.}} \end{aligned} \tag{3.3.37}$$

which exactly cancels the contribution from the ghost term in Eq. (3.3.36).

We have thus proven that the contribution of gauge-fixing and ghost terms in Eq. (3.3.36) are mutually cancelled, when applied to the Wilson loop (and in fact to any gauge-invariant functional).

3.3.3 Path and area derivatives

As we already mentioned, the RHS of Eq. (3.3.32) is completely represented via the (closed) Wilson loops. It is crucial for the loop-space formulation of QCD that the LHS of Eq. (3.3.32) 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 $x_\mu(\sigma) \in L_2$,[10] where $\sigma_0 \le \sigma \le \sigma_f$ and $\mu = 1,\ldots,d$, 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

$$s(\sigma) = \int_{\sigma_0}^{\sigma} d\sigma' \sqrt{\dot{x}_\mu^2(\sigma')} \tag{3.3.38}$$

where $\dot{x}_\mu(\sigma) = dx_\mu(\sigma)/d\sigma$.

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 $\sigma' = f(\sigma)$, $f'(\sigma) \ge 0$ describe the same loop — reparametrization invariance.

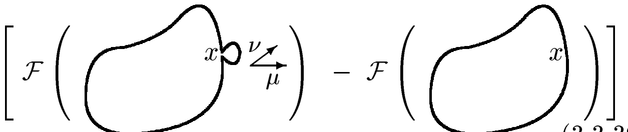
An example of functionals which are defined on the elements of loop space is the Wilson loop average (3.3.2) or, more generally, the n-loop average (3.3.1).

The differential calculus in loop space is built out of the path and area derivatives.

[10] Let us remind that L_2 stands for the Hilbert space of functions $x_\mu(\sigma)$ whose square is integrable over the Lebesgue measure: $\int_{\sigma_0}^{\sigma_f} d\sigma x_\mu^2(\sigma) < \infty$. We already mentioned this in the Remark on p. 19.

The *area derivative* of a functional $\mathcal{F}(C)$ is defined by the difference

$$\frac{\delta \mathcal{F}(C)}{\delta \sigma_{\mu\nu}(x)} \equiv \frac{1}{|\delta \sigma_{\mu\nu}|}$$



(3.3.39)

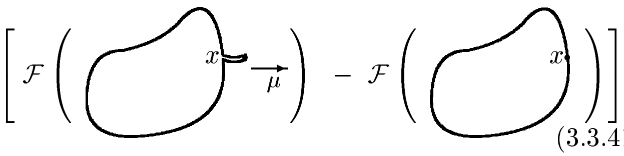
where an infinitesimal loop $\delta C_{\mu\nu}(x)$ is attached to a given loop at the point x in the $\mu\nu$-plane and $|\delta\sigma_{\mu\nu}|$ stands for the area enclosed by the $\delta C_{\mu\nu}(x)$. For a rectangular loop $\delta C_{\mu\nu}(x)$, one gets

$$\delta\sigma_{\mu\nu} \;=\; dx_\mu \wedge dx_\nu \,, \tag{3.3.40}$$

where the symbol $\wedge$ implies antisymmetrization.

Analogously, the *path derivative* is defined by

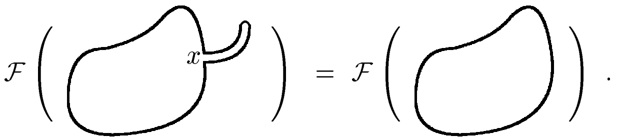
$$\partial^x_\mu \mathcal{F}(C_{xx}) \equiv \frac{1}{|\delta x_\mu|}$$



(3.3.41)

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:



(3.3.42)

This condition is equivalent to the Bianchi identity of Yang–Mills theory and is obviously satisfied by the Wilson loop (3.3.2) due to the properties of the non-Abelian phase factor (see Eq. (2.1.48)). Such functionals are known in mathematics as Chen integrals.[11]

A simple example of the Stokes functional is the area of the minimal surface, $A_{\min}(C)$. It obviously satisfies Eq. (3.3.42). Otherwise, the length

[11] See, *e.g.*, Ref. [Tav93] which contains definitions of path and area derivatives in this language.

$L(C)$ of the loop C is not a Stokes functional, since the lengths of contours on the LHS and RHS of Eq. (3.3.42) are different.

For the Stokes functionals, the variation on the RHS of Eq. (3.3.39) 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. (3.3.41) 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 (3.3.2)), the RHS of Eq. (3.3.41) vanishes due to the backtracking condition (3.3.42). 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

$$\Phi^a[C_{xx}] \equiv \frac{\mathrm{tr}}{N}\left(t^a\,\mathbf{P}\,\mathrm{e}^{\int_{C_{xx}} d\xi^\mu \mathcal{A}_\mu(\xi)}\right) \tag{3.3.43}$$

with the 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

$$\frac{\delta}{\delta\sigma_{\mu\nu}(x)}\frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} \;=\; \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\mathcal{F}_{\mu\nu}(x)\,\mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu}\,. \tag{3.3.44}$$

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 was done in Problem 2.8, and straightforwardly use the definition (3.3.39). The sense of Eq. (3.3.44) is very simple: $\mathcal{F}_{\mu\nu}$ is a curvature associated with the connection $\mathcal{A}_\mu$, as we discussed in the Remark on p. 99.

The functional on the RHS of Eq. (3.3.44) has a marked point x, and is of the type in Eq. (3.3.43). When the path derivative acts on such a functional according to the definition (3.3.41), the result reads

$$\partial^x_\mu \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,B(x)\,\mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} \;=\; \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\nabla_\mu B(x)\,\mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu}\,, \tag{3.3.45}$$

where

$$\nabla_\mu B \;=\; \partial_\mu B - [\mathcal{A}_\mu, B] \tag{3.3.46}$$

is the covariant derivative (2.1.11) in the adjoint representation (see also Problem 2.7).

Combining Eqs. (3.3.44) and (3.3.45), we finally represent the expression on the LHS of Eq. (3.3.28) (or Eq. (3.3.32)) as

$$\frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\nabla_\mu \mathcal{F}_{\mu\nu}(x)\,\mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} \;=\; \partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)}\frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu}\,, \tag{3.3.47}$$

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
∇^x_μ	covariant derivative	∂^x_μ	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 3.2: Vocabulary for translation of Yang–Mills theory from ordinary space in loop space.

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 3.2 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 (3.3.42) can be equivalently represented as

$$\epsilon_{\mu\nu\lambda\rho}\,\partial^x_\mu \frac{\delta}{\delta\sigma_{\nu\lambda}\,(x)}\Phi\,(C) \;=\; 0\,, \tag{3.3.48}$$

by choosing the appendix in Eq. (3.3.42) to be an infinitesimal straight line in the ρ-direction and geometrically applying the Stokes theorem. Using Eqs. (3.3.44) and (3.3.45), Eq. (3.3.48) can in turn be rewritten as

$$\epsilon_{\mu\nu\lambda\rho}\frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\nabla_\mu\mathcal{F}_{\nu\lambda}\,(x)\;\mathrm{e}^{\oint_C d\xi^\mu \mathcal{A}_\mu} \;=\; 0\,. \tag{3.3.49}$$

Therefore, Eq. (3.3.48) represents the Bianchi identity (2.1.19) 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

$$\frac{\delta}{\delta x_\mu(\sigma)} = \dot{x}_\nu(\sigma)\frac{\delta}{\delta\sigma_{\mu\nu}(x(\sigma))} + \sum_{i=1}^{m}\partial^{x_i}_\mu\delta(\sigma-\sigma_i)\,, \tag{3.3.50}$$

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 $x_1,\ldots,x_m$.

By using Eq. (3.3.50), the path derivative can be calculated as the limiting procedure

$$\partial^{x(\sigma)}_\mu = \int\limits_{\sigma-0}^{\sigma+0} d\sigma'\,\frac{\delta}{\delta x_\mu(\sigma')}\,. \tag{3.3.51}$$

The result is obviously nonvanishing only when ∂^x_μ 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 [Pol80]:

$$\frac{\delta}{\delta\sigma_{\mu\nu}(x(\sigma))} = \int\limits_{\sigma-0}^{\sigma+0} d\sigma'(\sigma'-\sigma)\frac{\delta}{\delta x_\mu(\sigma')}\frac{\delta}{\delta x_\nu(\sigma)}\,. \tag{3.3.52}$$

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 (3.3.48).

3.3.4 Loop equations

By virtue of Eq. (3.3.47), Eq. (3.3.32) can be represented completely in loop space:

$$\begin{aligned}&\partial^x_\mu\frac{\delta}{\delta\sigma_{\mu\nu}\,(x)}\left\langle\,\Phi\,(C)\,\right\rangle \\ &\quad= \lambda\oint\limits_C dy_\nu\,\delta^{(d)}\,(x-y)\left\langle\left[\Phi\,(C_{yx})\,\Phi\,(C_{xy})-\frac{1}{N_c^2}\Phi\,(C)\right]\right\rangle\,,\end{aligned} \tag{3.3.53}$$

or, using the definitions (3.3.1) and (3.3.2) of the loop averages, as

$$\partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W(C) = \lambda \oint_C dy_\nu\, \delta^{(d)}(x-y) \left[W_2(C_{yx}, C_{xy}) - \frac{1}{N_c^2} W(C) \right]. \qquad (3.3.54)$$

This equation is not closed. Having started from $W(C)$, we obtain another quantity, $W_2(C_1, C_2)$, so that Eq. (3.3.54) 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 (1.2.47) 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:

$$W_2(C_1, C_2) = W(C_1)\, W(C_2) + \mathcal{O}\left(\frac{1}{N_c^2}\right), \qquad (3.3.55)$$

as was discussed in Subsect. 3.2.6. Keeping the constant λ (defined by Eq. (3.3.33)) fixed in the large-N_c limit as is prescribed by Eq. (3.2.13), we get [MM79]

$$\partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W(C) = \lambda \oint_C dy_\nu\, \delta^{(d)}(x-y)\, W(C_{yx})\, W(C_{xy}) \qquad (3.3.56)$$

as $N_c \to \infty$.

Equation (3.3.56) 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. (3.3.56) should be solved in the class of Stokes functionals with the initial condition

$$W(0) = 1 \qquad (3.3.57)$$

for loops which are shrunk to points. This is a consequence of the obvious property of the Wilson loop

$$\mathrm{e}^{\oint_0 d\xi^\mu \mathcal{A}_\mu} = 1 \qquad (3.3.58)$$

and the normalization

$$\langle\, 1\, \rangle = 1 \qquad (3.3.59)$$

of the averages.

The factorization (3.3.55) can itself be derived from the chain of loop equations. Proceeding as before, we get

$$\begin{aligned} \frac{1}{\lambda}\partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W_n(C_1,\ldots,C_n) &= \oint_{C_1} dy_\nu\, \delta^{(d)}(x-y) \left[W_{n+1}(C_{xy}, C_{yx}, \ldots, C_n) - \frac{1}{N_c^2} W_n(C_1,\ldots,C_n) \right] \\ &+ \sum_{j\geq 2} \frac{1}{N_c^2} \oint_{C_j} dy_\nu\, \delta^{(d)}(x-y) \Big[W_{n-1}\left(C_1 C_j, \ldots, \underline{C_j}, \ldots, C_n\right) \\ &\qquad - W_n(C_1,\ldots,C_n) \Big]. \end{aligned} \qquad (3.3.60)$$

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 (3.3.60) looks like Eq. (1.2.47) for the φ^3-theory. Moreover, the number of colors N_c enters Eq. (3.3.60) simply as a scalar factor N_c^{-2}, likewise Plank's constant $\hbar$ enters Eq. (1.2.47). It is the major advantage of the use of loop space. What is said in Subsect. 3.2.8 about the "semiclassical" nature of the $1/N_c$-expansion of QCD is explicitly realized in Eq. (3.3.60). Its expansion in $1/N_c$ is straightforward.

At $N_c = \infty$, Eq. (3.3.60) is simplified to

$$\partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W_n(C_1,\ldots,C_n) = \lambda \oint_{C_1} dy_\nu\, \delta^{(d)}(x-y)\, W_{n+1}(C_{yx}, C_{xy}, \ldots, C_n). \qquad (3.3.61)$$

This equation possesses a factorized solution

$$\begin{aligned} W_n(C_1,\ldots,C_n) &= \langle\Phi(C_1)\rangle \ldots \langle\Phi(C_n)\rangle + \mathcal{O}\left(\frac{1}{N_c^2}\right) \\ &\equiv W(C_1)\cdots W(C_n) + \mathcal{O}\left(\frac{1}{N_c^2}\right) \end{aligned} \qquad (3.3.62)$$

provided $W(C)$ obeys Eq. (3.3.56) 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.

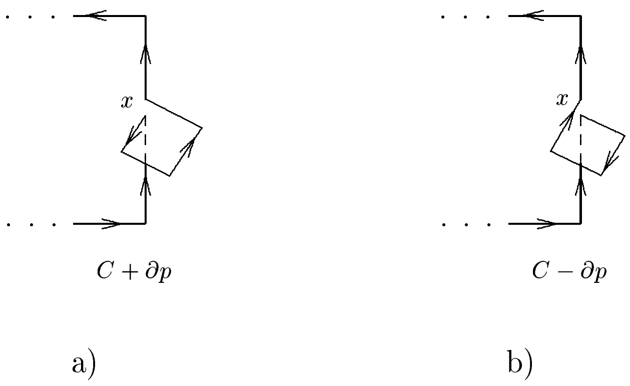


Fig. 3.29: Contours a) $C+\partial p$ and b) $C-\partial p$ on the RHS of the lattice loop equation (3.3.63).

Problem 3.22 Derive a lattice analog of the loop equation.
Solution The derivation is similar to that in Problem 2.12 for the classical case. We perform the shift (2.2.22) in the definition (2.2.42) of the lattice Wilson loop average. Likewise Eqs. (2.2.24) and (3.3.56), we get

$$\frac{\beta}{2N_c^2}\sum_p \left[W_{C+\partial p}-W_{C-\partial p}\right] \;=\; \sum_{l\in C}\tau_\nu(l)\,W_{C_{yx}}W_{C_{xy}}\,. \tag{3.3.63}$$

Here the contours $C+\partial p$ and $C-\partial p$ are obtained from C_{xx} by adding the boundary of the plaquette p ($-\partial p$ means that the orientation of the boundary is opposite) and the sum over p goes over the $2(d-1)$ plaquettes involving the link at which the shift of $U_{x,\nu}$ is performed. This contours are depicted in Fig. 3.29.

The sum on the RHS goes over the links belonging to the contour C. The unit vector $\tau_\nu(l)=0,\pm1$ stands for the projection of the (oriented) link $l\in C$ on the axis ν ($\tau_\nu(l)=1$, -1 or 0 when the directions are parallel, antiparallel or perpendicular, respectively). The point y is defined as the beginning of the link l if it has the positive direction, or as the end of l if it has the negative one. Such an asymmetry is due to the fact that we have performed the right shift (2.2.22) of $U_{x,\nu}$. The Kronecker symbol δ_{xy} guarantees that C_{yx} and C_{xy} are always closed.

Equation (3.3.63) is a lattice regularization of the continuum loop equation (3.3.56). The loop equation on the lattice was first discussed in Refs. [Foe79, Egu79].

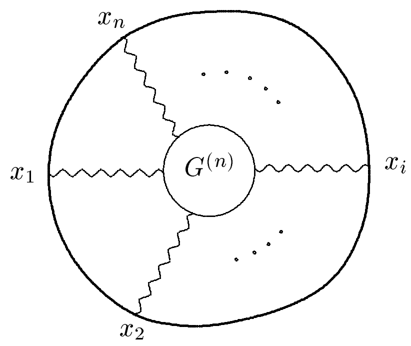


Fig. 3.30: Graphic representation of the terms on the RHS of Eq. (3.3.64).

3.3.5 Relation to planar diagrams

The perturbation-theory expansion of the Wilson loop average can be calculated from Eq. (3.2.47) which we represent in the form

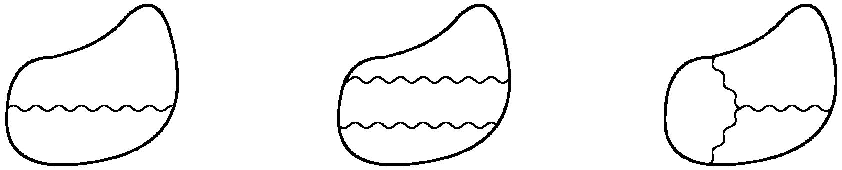
$$\begin{aligned} W(C) \;&=\; 1+\sum_{n=2}^{\infty}\oint\limits_C dx_1^{\mu_1}\oint\limits_C dx_2^{\mu_2}\ldots\oint\limits_C dx_n^{\mu_n} \\ &\quad\times\theta_c(1,2,\ldots,n)\,G^{(n)}_{\mu_1\mu_2\cdots\mu_n}(x_1,x_2,\ldots,x_n)\,, \end{aligned} \tag{3.3.64}$$

where $\theta_c(1,2,\ldots,n)$ orders the points x_1, …, x_n along contour in the cyclic order and $G^{(n)}_{\mu_1\cdots\mu_n}$ is given by Eq. (3.2.72). This θ-function has the meaning of the propagator of a test heavy particle which lives in the contour C (see Problem 2.2).

We assume, for definitiveness, the dimensional regularization throughout this Subsection to make all the integrals well-defined.

Each term on the RHS of Eq. (3.3.64) can be conveniently represented by the diagram in Fig. 3.30, 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. 3.2.3 with one external boundary — the Wilson loop in the given case. This was already mentioned in the Remark on p. 230. 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. 3.31. The diagram in Fig. 3.31a has been already considered in Problem 3.20 (see Eq. (3.3.18)).



a) b) c)

Fig. 3.31: 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.

The large-N_c loop equation (3.3.56) 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 (3.3.57) and some boundary conditions for asymptotically large contours are imposed.

Equation (3.3.64) can be viewed as an ansatz for $W(C)$ with some unknown functions $G^{(n)}_{\mu_1\cdots\mu_n}(x_1,\ldots,x_n)$ 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 $1,\ldots,n$ and depend only on $x_i - x_j$ (translational invariance). A main advantage of this ansatz is that it automatically corresponds to a Stokes functional, due to the properties of vector integrals, and the initial condition (3.3.57) is satisfied.

The action of the area and path derivatives on the ansatz (3.3.64) is easily calculable. For instance, the area derivative reads

$$
\begin{aligned}
\frac{\delta W(C)}{\delta\sigma_{\mu\nu}(z)} &= \sum_{n=1}^{\infty} \oint_C dx_1^{\mu_1} \ldots \oint_C dx_n^{\mu_n}\, \theta_c(1,2,\ldots,n) \\
&\times \Big[\left(\partial^z_\mu \delta_{\nu\alpha} - \partial^z_\nu \delta_{\mu\alpha}\right) G^{(n+1)}_{\alpha\mu_1\cdots\mu_n}(z,x_1,\ldots,x_n) \\
&\quad + \left(\delta_{\mu\beta}\delta_{\nu\alpha} - \delta_{\mu\alpha}\delta_{\nu\beta}\right) G^{(n+2)}_{\alpha\beta\mu_1\cdots\mu_n}(z,z,x_1,\ldots,x_n)\Big]. \qquad (3.3.65)
\end{aligned}
$$

The analogy with the Mandelstam formula (3.3.44) is obvious.

More about solving the loop equation by the ansatz (3.3.64) can be found in Refs. [MM81, BGSN82, Mig83].

Problem 3.23 Solve Eq. (3.3.56) by the ansatz (3.3.64) to order λ.

Solution To order λ, we can restrict ourselves by the $n = 2$ term in the ansatz (3.3.64). For the θ-function, we have

$$\theta_c(1,2) \equiv \frac{1}{2}\left[\theta(1,2)+\theta(2,1)\right] = \frac{1}{2}. \qquad (3.3.66)$$

The meaning of this formula is obvious: there is no cyclic ordering for two points. We rewrite therefore the ansatz as

$$W(C) = 1 - \frac{1}{2}\lambda \oint_C dx_\mu \oint_C dy_\nu\, D_{\mu\nu}(x-y) + \mathcal{O}\left(\lambda^2\right) \qquad (3.3.67)$$

with some unknown function $D_{\mu\nu}(x-y)$. Its tensor structure reads

$$D_{\mu\nu}(x-y) = \delta_{\mu\nu} D(x-y) + \partial_\mu\partial_\nu f(x-y)\,. \qquad (3.3.68)$$

The second (longitudinal) term in this formula does not contribute to $W(C)$ since the contour integral of this term vanishes in Eq. (3.3.67). We can thus write

$$W(C) = 1 - \frac{1}{2}\lambda \oint_C dx_\mu \oint_C dy_\mu\, D(x-y) + \mathcal{O}\left(\lambda^2\right). \qquad (3.3.69)$$

The area derivative can be easily calculated by using the Stokes theorem, which gives

$$
\begin{aligned}
&\frac{\delta}{\delta\sigma_{\mu\nu}(z)} \oint_C dx_\rho \oint_C dy_\rho D(x-y) \\
&= 2\left(\oint_C dy_\nu \partial_\mu D(z-y) - \oint_C dy_\mu \partial_\nu D(z-y) \right) \qquad (3.3.70)
\end{aligned}
$$

and

$$\partial^z_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(z)} \oint_C dx_\rho \oint_C dy_\rho D(x-y) = 2\oint_C dy_\nu \partial^2 D(z-y) \qquad (3.3.71)$$

since

$$\partial_\mu \oint dy_\mu \partial_\nu D(x-y) = 0\,. \qquad (3.3.72)$$

Substituting into the loop equation (3.3.56), we get

$$-\oint_C dy_\nu \partial^2 D(x-y) = \oint_C dy_\nu\, \delta^{(d)}(x-y) \qquad (3.3.73)$$

which is equivalent to

$$-\partial^2 D\,(x-y) \;=\; \delta^{(d)}\,(x-y) \tag{3.3.74}$$

since the contour C is arbitrary. The solution to Eq. (3.3.74) is unique, provided $D\,(x-y)$ decreases for large $x-y$, and recovers the propagator (3.2.4).

3.3.6 Loop-space Laplacian and regularization

The loop equation (3.3.56) 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. (3.3.56) is overspecified.

A practical difficulty in solving Eq. (3.3.56) 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. (3.3.56) over dx_ν along the same contour C, which yields

$$\begin{aligned}&\oint_C dx_\nu\,\partial_\mu^x\,\frac{\delta}{\delta\sigma_{\mu\nu}(x)}W(C)\\ &\quad=\;\lambda\oint_C dx_\mu\oint_C dy_\mu\,\delta^{(d)}(x-y)\,W(C_{yx})W(C_{xy})\,.\end{aligned} \tag{3.3.75}$$

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. (3.3.75) can be interpreted as an infinitesimal variation of elements of loop space.

Equations (3.3.56) and (3.3.75) are completely equivalent. A proof of equivalence of scalar Eq. (3.3.75) and original d-vector Eq. (3.3.56) is based on the important property of Eq. (3.3.56) whose both sides are identically annihilated by the operator ∂_ν^x. It is a consequence of the identity (see Subsect. 2.1.1)

$$\nabla_\mu\nabla_\nu\,\mathcal{F}_{\mu\nu}=-\frac{1}{2}\left[\mathcal{F}_{\mu\nu},\mathcal{F}_{\mu\nu}\right]\;=\;0 \tag{3.3.76}$$

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. (3.3.56) can in turn be deduced from Eq. (3.3.75).

Equation (3.3.75) is associated with the so-called second-order Schwinger–Dyson equation

$$\begin{aligned}&\int d^dx\,\nabla_\mu F^a_{\mu\nu}\,(x)\,\frac{\delta}{\delta A^a_\nu\,(x)}\\ &\quad\overset{\text{w.s.}}{=}\;\hbar\int d^dx\,d^dy\,\delta^{(d)}\,(x-y)\,\frac{\delta}{\delta A^a_\nu\,(y)}\,\frac{\delta}{\delta A^a_\nu\,(x)}\end{aligned} \tag{3.3.77}$$

in the same sense as Eq. (3.3.56) is associated with Eq. (3.3.26). It is called "second order" since the RHS involves two variational derivatives with respect to A_ν.

The operator on the LHS of Eq. (3.3.75) 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. (3.3.51) and (3.3.52), in an equivalent form

$$\begin{aligned}\Delta\;&\equiv\;\oint_C dx_\nu\,\partial_\mu^x\,\frac{\delta}{\delta\sigma_{\mu\nu}(x)}\\ &=\;\int_{\sigma_0}^{\sigma_f} d\sigma\int_{\sigma-0}^{\sigma+0} d\sigma'\,\frac{\delta}{\delta x_\mu(\sigma')}\,\frac{\delta}{\delta x_\mu(\sigma)}\,.\end{aligned} \tag{3.3.78}$$

As was first pointed out by Gervais and Neveu [GN79b], this operator is nothing but a functional extension of the Laplace operator, which is known in mathematics as the Levy operator.[12] Equation (3.3.75) can be represented in turn as an (inhomogeneous) functional Laplace equation

$$\Delta W(C)\;=\;\lambda\oint_C dx_\mu\oint_C dy_\mu\,\delta^{(d)}(x-y)\,W(C_{yx})W(C_{xy})\,. \tag{3.3.79}$$

We shall refer to this equation as the loop-space Laplace equation.

The form (3.3.79) of the loop equation is convenient for a non-perturbative ultraviolet regularization.

The idea is to start from the regularized version of Eq. (3.3.77), replacing the delta-function on the RHS by the kernel of the regularizing operator:

$$\delta^{ab}\delta^{(d)}(x-y)\;\overset{\text{Reg.}}{\Longrightarrow}\;\left\langle y\left|\boldsymbol{R}^{ab}\right|x\right\rangle=\boldsymbol{R}^{ab}\,\delta^{(d)}(x-y) \tag{3.3.80}$$

[12] See the book by Levy [Lev51], cited in the references to Chapter 1, and a recent review [Fel86].

with

$$\boldsymbol{R}^{ab} \;=\; \left(\mathrm{e}^{\,a^2\nabla^2/2}\right)^{ab}, \tag{3.3.81}$$

where ∇_μ is the covariant derivative in the adjoint representation. The regularized version of Eq. (3.3.77) is

$$\begin{aligned}&\int d^d x\, \nabla_\mu F^a_{\mu\nu}(x)\,\frac{\delta}{\delta A^a_\nu(x)}\\ &\overset{\text{w.s.}}{=}\; \hbar \int d^d x\, d^d y\, \left\langle y \left| \boldsymbol{R}^{ab} \right| x\right\rangle \frac{\delta}{\delta A^b_\nu(y)}\frac{\delta}{\delta A^a_\nu(x)}\,. \end{aligned}\tag{3.3.82}$$

To translate Eq. (3.3.82) in loop space, we use the path-integral representation (see Problem 2.5)

$$\left\langle y \left| \boldsymbol{R}^{ab} \right| x\right\rangle \;=\; \int\limits_{\substack{r(0)=x\\ r(a^2)=y}} Dr(t)\,\mathrm{e}^{-\frac{1}{2}\int_0^{a^2} dt\,\dot{r}^2(t)}\; 2\,\mathrm{tr}\,\left[t^a U(r_{yx}) t^b U(r_{xy})\right] \tag{3.3.83}$$

with

$$U(r_{yx}) \;=\; \mathbf{P}\,\mathrm{e}^{\int_x^y dr_\mu \mathcal{A}_\mu(r)}\,, \tag{3.3.84}$$

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 (3.3.83) so that

$$\begin{aligned}&\int\limits_{\substack{r(0)=x\\ r(a^2)=y}} Dr(t)\,\mathrm{e}^{-\frac{1}{2}\int_0^{a^2} dt\,\dot{r}^2(t)}\; 2\,\mathrm{tr}\,\left[t^a t^b\right]\\ &=\; \delta^{ab}\,\mathrm{e}^{\,a^2\partial^2/2}\delta^{(d)}(x-y) = \delta^{ab}\frac{1}{(2\pi a^2)^{d/2}}\,\mathrm{e}^{-\frac{(x-y)^2}{2a^2}}\,. \end{aligned}\tag{3.3.85}$$

Calculating the variational derivatives on the RHS of Eq. (3.3.82), using Eq. (3.3.83) and the completeness condition (3.2.6), we get as $N \to \infty$:

$$\begin{aligned}\int d^d x\, d^d y\, \left\langle y \left| \boldsymbol{R}^{ab} \right| x\right\rangle \frac{\delta}{\delta A^b_\nu(y)}\frac{\delta}{\delta A^a_\nu(x)}\Phi(C) \;&=\; \lambda \oint\limits_C dx_\mu \oint\limits_C dy_\mu\\ \times \int\limits_{\substack{r(0)=x\\ r(a^2)=y}} \mathcal{D}r(t)\,\mathrm{e}^{-\frac{1}{2}\int_0^{a^2} dt\dot{r}^2(t)}\;\Phi\left(C_{yx}r_{xy}\right)\Phi\left(C_{xy}r_{yx}\right)\,,\end{aligned}\tag{3.3.86}$$

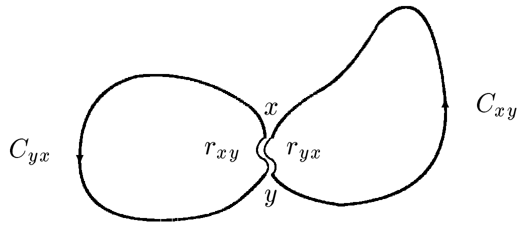


Fig. 3.32: Contours $C_{yx}r_{xy}$ and $C_{xy}r_{yx}$ which enter the RHS's of Eqs. (3.3.86) and (3.3.87).

where the contours $C_{yx}r_{xy}$ and $C_{xy}r_{yx}$ are depicted in Fig. 3.32. Averaging over the gauge field and using the large-N_c factorization, we arrive at the regularized loop-space Laplace equation [HM89]

$$\Delta W(C) \;=\; \lambda \oint\limits_C dx_\mu \oint\limits_C dy_\mu \int\limits_{\substack{r(0)=x\\ r(a^2)=y}} Dr(t)\,\mathrm{e}^{-\frac{1}{2}\int_0^{a^2} dt\,\dot{r}^2(t)}\; W(C_{yx}r_{xy})W(C_{xy}r_{yx}) \tag{3.3.87}$$

which manifestly recovers Eq. (3.3.79) when $a \to 0$.

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. (3.3.87) both are closed and do not have marked points if C does not have. Therefore, Eq. (3.3.87) 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. (3.3.78) 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 [Lev51], cited in the references to Chapter 1, uses solely the concept of the second variation of a functional $U[x]$, namely the term in the second variation which is proportional to $(\delta x_\mu(\sigma))^2$:

$$\delta^2 U[x] = \tfrac{1}{2}\int_{\sigma_0}^{\sigma_f} d\sigma\, U''_{xx}[x](\delta x_\mu(\sigma))^2 + \ldots\,. \tag{3.3.88}$$

The functional Laplacian Δ is then defined by the formula

$$\Delta U[x] = \int_{\sigma_0}^{\sigma_f} d\sigma \, U''_{xx}[x] \,. \tag{3.3.89}$$

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

$$\Delta\,(UV) \;=\; \Delta\,(U)\,V + U\Delta\,(V) \;. \tag{3.3.90}$$

The functional Laplacian can be approximated [Mak88] 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:

$$G\,(C,C') \;=\; \sum_{S_{C,C'}} \ldots \,, \tag{3.3.91}$$

which is analogous to the representation (1.1.98) of the Green function of the ordinary Laplacian. The standard perturbation theory can then be recovered by iterating Eq. (3.3.79) (or its regularized version (3.3.87)) in λ with the Green function (3.3.91).

3.3.7 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. [MM80] that area law

$$W\,(C) \equiv \langle \Phi\,(C) \rangle \;\propto\; \mathrm{e}^{\,-K\cdot A_{\min}(C)} \tag{3.3.92}$$

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

$$W\,(C) \;=\; \sum_{S:\partial S = C} \mathrm{e}^{\,-K\cdot A(S)} \,, \tag{3.3.93}$$

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. [GN79a, Nam79, Pol79]. Roughly speaking, the ansatz (3.3.93) is not consistent with the factorized structure on the RHS of Eq. (3.3.56).

Nevertheless, it was shown that if a free string satisfies Eq. (3.3.56), 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 or higher genera are associated with a string interaction. The coupling constant of this interaction is $\mathcal{O}\left(N_c^{-2}\right)$.

A formal solution of Eq. (3.3.56) for all loops was found by Migdal [Mig81] in the form of a fermionic string

$$W\,(C) \;=\; \sum_{S:\partial S = C} \int D\psi \, \mathrm{e}^{\,-\int d^2\xi \left[\bar\psi \sigma_k \partial_k \psi + \bar\psi\psi m^4 \sqrt{g}\right]}, \tag{3.3.94}$$

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 (3.3.92) is recovered. For small loops, the elves are necessary for asymptotic freedom. However, it is unclear whether or not the string solution (3.3.94) is practically useful for a 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 [EK82]. 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

$$Z_{\rm EK}(\beta) \;=\; \int \prod_{\rho=1}^{d} dU_\rho \, {\rm e}^{\beta \sum_{\mu>\nu} {\rm Re}\, \frac{1}{N_c} {\rm tr}\left(U_\mu U_\nu U_\mu^\dagger U_\nu^\dagger\right)} . \tag{3.3.95}$$

Here the matrices U_ρ ($\rho = 1, \ldots, d$) 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 will be said in the next Section.

3.3.8 Wilson loops in QCD$_2$

Two-dimensional QCD is popular since the paper by 't Hooft [Hoo74b] as a simplified model of QCD$_4$.

One can always choose the axial gauge

$$A_1 \;=\; 0 , \tag{3.3.96}$$

so that the commutator in the non-Abelian field strength (2.1.15) vanishes in two dimensions. Therefore, there is no gluon self-interaction in this gauge and the theory looks, at the first glance, like the Abelian one.

The Wilson loop average in QCD$_2$ can be straightforwardly calculated via the expansion (3.3.64) where only disconnected (free) parts of the correlators $G^{(n)}$ for even n should be left, since there is no interaction. Only the planar structure of color indices contributes at $N_c = \infty$. Diagrammatically, the diagrams of the type depicted in Fig. 3.31a and Fig. 3.31b are relevant for contours without self-intersections, while that in Fig. 3.31c should be omitted in two dimensions.

The color structure of the relevant planar diagrams can be reduced by the virtue of the formula

$$\sum_a \left(t^a\right)^{ik} \left(t^a\right)^{kj} \;=\; \frac{N_c}{2} \delta^{ij} , \tag{3.3.97}$$

which is a consequence of the completeness condition (3.2.6) at large N_c. We have

$$W(C) \;=\; 1 + \sum_k^\infty (-\lambda)^k \oint_C dx_1^{\mu_1} \oint_C dx_2^{\nu_1} \cdots \oint_C dx_{2k-1}^{\mu_k} \oint_C dx_{2k}^{\nu_k}$$
$$\times \theta_{\rm c}(1,2,\ldots,2k)\, D_{\mu_1\nu_1}(x_1 - x_2) \cdots D_{\mu_k\nu_k}(x_{2k-1} - x_{2k}) , \tag{3.3.98}$$

where the points $x_1, \ldots, x_{2k}$ are still cyclic ordered along the contour. Similar to Problem 2.2, we can exponentiate the RHS of Eq. (3.3.98) to get finally

$$W(C) \;=\; {\rm e}^{-\frac{\lambda}{2} \oint_C dx^\mu \oint_C dy^\nu D_{\mu\nu}(x-y)} . \tag{3.3.99}$$

This is the same formula as in the Abelian case if λ stands for e^2.

The propagator $D_{\mu\nu}(x,y)$ is, strictly speaking, the one in the axial gauge (3.3.96) which reads

$$D_{\mu\nu}(x-y) \;=\; \frac{1}{2} \delta_{\mu 2} \delta_{\nu 2} \left|x_1 - y_1\right| \delta^{(1)}(x_2 - y_2) . \tag{3.3.100}$$

However, the contour integral on the RHS of Eq. (3.3.99) is gauge invariant, and we can simply choose

$$D_{\mu\nu}(x-y) \;=\; \delta_{\mu\nu} D(x-y) . \tag{3.3.101}$$

In two dimensions[13] we have

$$D(x-y) \;=\; \frac{1}{4\pi} \ln \frac{\ell^2}{(x-y)^2} , \tag{3.3.102}$$

where ℓ is an arbitrary parameter of the dimension of length. Nothing depends on it because the contour integral of a constant vanishes.

The contour integral in the exponent on the RHS of Eq. (3.3.99) can be graphically represented as is depicted in Fig. 3.33, where $x_2 = y_2$ due to the delta-function in Eq. (3.3.100) and the bold line represents $|x_1 - y_1|$. This gives

$$\oint_C dx^\mu \oint_C dy^\nu D_{\mu\nu}(x-y) \;=\; A(C) \tag{3.3.103}$$

where $A(C)$ is the area enclosed by the contour C. We get finally

$$W(C) \;=\; {\rm e}^{-\frac{\lambda}{2} A(C)} \tag{3.3.104}$$

for the contours without self-intersections.

[13] In d dimensions

$$D(x-y) \;=\; \frac{1}{4\pi^{d/2}} \Gamma\left(\frac{d}{2} - 1\right) \frac{1}{\left[(x-y)^2\right]^{d/2-1}} .$$

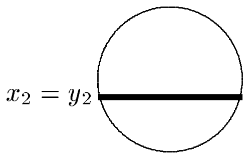


Fig. 3.33: Graphic representation of the contour integral on the LHS of Eq. (3.3.103) in the axial gauge. The bold line represents the gluon propagator (3.3.100) with $x_2 = y_2$ due to the delta-function.

Therefore, area law holds in two dimensions both in the non-Abelian and Abelian cases. This is, roughly speaking, because of the form of the two-dimensional propagator (3.3.102) which falls down with the distance only logarithmically in the Feynman gauge.

Problem 3.24 Prove Eq. (3.3.104) in the Feynman gauge.
Solution To prove Eq. (3.3.103) in the Feynman gauge (3.3.101), (3.3.102), we note that the area element in two dimensions can be represented by

$$d\sigma^{\mu\nu}(x) \;\equiv\; dx^\mu \wedge dx^\nu \;=\; e^{\mu\nu} d^2x\,, \tag{3.3.105}$$

where $e^{\mu\nu}$ is the antisymmetric tensor $e^{12} = -e^{21} = 1$. Therefore, the area can be represented by the double integral

$$A\,(C) \;=\; \frac{1}{2} \int\limits_{S(C)} d\sigma^{\mu\nu}(x) \int\limits_{S(C)} d\sigma^{\mu\nu}(y)\, \delta^{(2)}\,(x-y) \tag{3.3.106}$$

which goes along the surface $S(C)$ enclosed by the (non-intersecting) loop C.

Applying the Stokes' theorem, we get

$$\begin{aligned}
\oint\limits_C dx^\mu \oint\limits_C dy^\mu D\,(x-y) \;&=\; \int\limits_{S(C)} d\sigma^{\mu\nu}\,(x)\,\partial_\nu \oint\limits_C dy^\mu\, D\,(x-y) \\
&=\; -\int\limits_{S(C)} d\sigma^{\mu\nu}(x) \int\limits_{S(C)} d\sigma^{\mu\rho}(y)\,\partial_\nu\partial_\rho D\,(x-y) \\
&=\; -\frac{1}{2}\int\limits_{S(C)} d\sigma^{\mu\nu}(x) \int\limits_{S(C)} d\sigma^{\mu\nu}(y)\,\partial^2 D\,(x-y) \\
&=\; \frac{1}{2}\int\limits_{S(C)} d\sigma^{\mu\nu}(x) \int\limits_{S(C)} d\sigma^{\mu\nu}(y)\,\delta^{(2)}\,(x-y)\,.
\end{aligned} \tag{3.3.107}$$

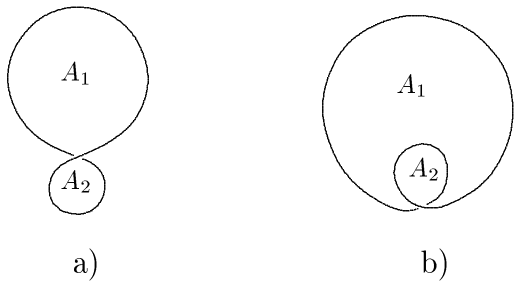


Fig. 3.34: Contours with one self-intersection: A_1 and A_2 stand for the areas of the proper windows. The total area enclosed by the contour in Fig. a) is $A_1 + A_2$. The areas enclosed by the exterior and interior loops in Fig. b) are $A_1 + A_2$ and A_2, respectively, while the total area of the surface with the folding is $A_1 + 2A_2$.

Using Eq. (3.3.106) we prove Eq. (3.3.104) in the Feynman gauge.

It is worth noting that Eq. (3.3.107) is based only on the Stokes' theorem and holds for contours with arbitrary self-intersections. On the contrary, Eq. (3.3.106) itself is valid only for non-intersecting loops.

The difference between the Abelian and non-Abelian cases shows up for the contours with self-intersections.

We first note that the simple formula (3.3.103) does *not* hold for contours with arbitrary self-intersections.

The simplest contours with one self-intersection are depicted in Fig. 3.34. There is nothing special about the contour in Fig. 3.34a. Equation (3.3.103) still holds in this case with $A\,(C)$ being the total area $A\,(C) = A_1 + A_2$.

The Wilson loop average for the contour in Fig. 3.34a coincides both for the Abelian and non-Abelian cases and equals

$$W\,(C) \;=\; \mathrm{e}^{-\frac{\lambda}{2}(A_1+A_2)}\,. \tag{3.3.108}$$

This is nothing but the exponential of the total area.

For the contour in Fig. 3.34b, we get

$$\oint\limits_C dx^\mu \oint\limits_C dy^\nu D_{\mu\nu}\,(x-y) \;=\; A_1 + 4A_2\,. \tag{3.3.109}$$

This is easy to understand in the axial gauge where the ends of the propagator line can lie both on the exterior and interior loops, or one end at the exterior

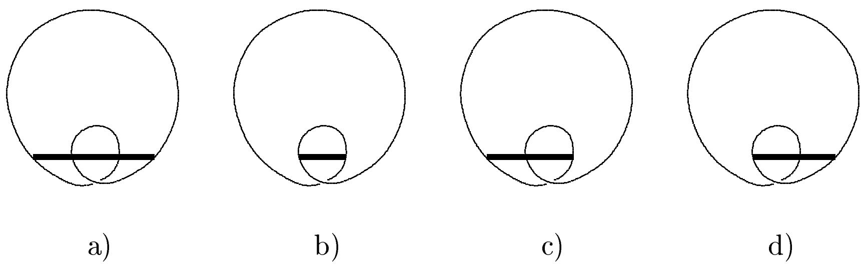


Fig. 3.35: Three type of contribution in Eq. (3.3.109) The ends of the propagator line lie both on a) exterior and b) interior loops, or c), d) one end on the exterior loop and another end on the interior loop.

loop and the other end on the interior loop. These cases are illustrated by Fig. 3.35. The contributions of the diagrams in Fig. 3.35a,b,c,d are $A_1 + A_2$, A_2, A_2, and A_2, respectively. The result given by Eq. (3.3.109) is obtained by summing over all four diagrams.

For the contour in Fig. 3.34b, the Wilson loop average is

$$W(C) \;=\; \mathrm{e}^{-\frac{\lambda}{2}(A_1+4A_2)} \tag{3.3.110}$$

in the Abelian case and

$$W(C) \;=\; (1-\lambda A_2)\,\mathrm{e}^{-\frac{\lambda}{2}(A_1+2A_2)} \tag{3.3.111}$$

in the non-Abelian case at $N_c = \infty$. They coincide only to the order λ as they should. The difference to the next orders is because only the diagrams with one propagator line connecting the interior and exterior loops are planar and, therefore, contribute in the non-Abelian case. Otherwise, the diagram is non-planar and vanishes as $N_c \to \infty$. Notice, that the exponential of the total area $A(C) = A_1 + 2A_2$ of the surface with the folding, which is enclosed by the contour C, appears in the exponent for the non-Abelian case. The additional pre-exponential factor could be associated with an entropy of foldings of the surface.

The Wilson loop averages (3.3.108) and (3.3.111) in QCD_2 at large N_c as well as the ones for contours with arbitrary self-intersections, which have a generic form

$$W(C) \;=\; P(A_1,\ldots,A_n)\,\mathrm{e}^{-\frac{\lambda}{2}Area} \tag{3.3.112}$$

where P is a polynomial of the areas of individual windows and $Area$ is the total area of the surface with foldings, were first calculated in Ref. [KK80] by solving the two-dimensional loop equation and in Ref. [Bra80] by applying the non-Abelian Stokes' theorem. The lattice version is given in Ref. [KK81].

Problem 3.25 Demonstrate that Eq. (3.3.99) satisfies the Abelian loop equation

$$\partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W(C) \;=\; \lambda \oint_C dy_\nu\, \delta^{(d)}(x-y)\, W(C)\,. \tag{3.3.113}$$

Solution The calculation is the same as in Problem 3.23. In $d = 2$ one can alternatively use [OP81] the expression on the RHS of Eq. (3.3.107).

Problem 3.26 Obtain Eqs. (3.3.108) and (3.3.111) for the contours with one self-intersection by solving the loop equation (3.3.56).
Solution Let us multiply Eq. (3.3.56) in $d = 2$ by $e_{\rho\nu}$ and integrate over dx^ρ along a small (open) piece C' of the contour C including the point of self-intersection. We get

$$e_{\rho\nu}\int_{C'} dx_\rho \partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W(C) \;=\; \lambda e_{\rho\nu}\int_{C'} dx_\rho \oint_C dy_\nu\, \delta^{(2)}(x-y)\, W(C_{yx})\, W(C_{xy})\,. \tag{3.3.114}$$

The RHS of Eq. (3.3.114) can be calculated analogously to the known representation for the number of self-intersections of a loop in two-dimensions. For the case of one self-intersection, we have

$$e_{\rho\nu}\int_{C'} dx_\rho \oint_C dy_\nu\, \delta^{(2)}(x-y)\, W(C_{yx})\, W(C_{xy}) \;=\; W(C_1)\, W(C_2)\,, \tag{3.3.115}$$

where C_1 and C_2 stands, respectively, for the upper and lower loops in Fig. 3.34a or the exterior and interior loops in Fig. 3.34b.

The LHS of Eq. (3.3.114) can be transformed as

$$e_{\rho\nu}\int_{C'} dx_\rho \partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W(C) \;=\; \int_{C'} dx_\nu \partial^x_\nu \frac{\delta}{\delta\sigma(x)} W(C)\,, \tag{3.3.116}$$

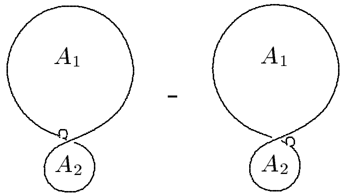
where $\delta/\delta\sigma(x)$ denotes the variational derivative with respect to the "scalar" area

$$\delta\sigma(x) \;=\; \frac{1}{2} e_{\mu\nu}\delta\sigma^{\mu\nu}(x)\,. \tag{3.3.117}$$

The integrand on the RHS of Eq. (3.3.116) is a total derivative and the contour integral reduces to the difference of the Ω-variations at the end points of the contour

C', which would vanish if no self-intersections. The RHS of Eq. (3.3.114) also vanishes if no self-intersections so $W(C)$ is determined in this case by Eq. (3.3.113) rather than Eq. (3.3.114).

For the contour in Fig. 3.34a, this gives

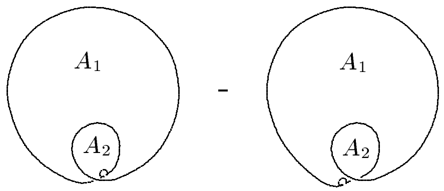


$$= \left(-\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) W(C)\,. \tag{3.3.118}$$

The Ω-variation of the contour on the LHS represents the variational derivative. The minus sign in front of $\partial/\partial A_1$ on the RHS is because adding the Ω-variation in the first term on the LHS decreases the area A_1 while that in the second term increases A_2. Then Eq. (3.3.114) takes for the contour in Fig. 3.34a the form

$$\left(-\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) W(C) \;=\; \lambda W(C_1)\, W(C_2)\,. \tag{3.3.119}$$

For the contour in Fig. 3.34b, we get quite similarly



$$= \left(2\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) W(C)\,. \tag{3.3.120}$$

Now adding the Ω-variation in the first term on the LHS increases A_1 and decreases A_2 while that in the second term decreases A_1. Equation (3.3.114) takes the form

$$\left(2\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) W(C) \;=\; \lambda W(C_1)\, W(C_2)\,. \tag{3.3.121}$$

The RHS's of Eqs. (3.3.119) and (3.3.121) are known since C_1 and C_2 have no self-intersections so that Eq. (3.3.104) holds for $W(C_1)$ and $W(C_2)$. Equations (3.3.119) and (3.3.121) take finally the explicit form [KK80]

$$\left(-\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) W(C) \;=\; \lambda\, \mathrm{e}^{-\frac{\lambda}{2}(A_1+A_2)}\,. \tag{3.3.122}$$

and

$$\left(2\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) W(C) \;=\; \lambda\, \mathrm{e}^{-\frac{\lambda}{2}(A_1+2A_2)}\,, \tag{3.3.123}$$

respectively. Their solution is uniquely given by Eqs. (3.3.108) and (3.3.111).

It is worth nothing that the linear Abelian loop equation (3.3.113) can be written for the contours in Fig. 3.34a,b as

$$\left(-\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) \ln W(C) \;=\; \lambda\,, \tag{3.3.124}$$

$$\left(2\frac{\partial}{\partial A_1} - \frac{\partial}{\partial A_2}\right) \ln W(C) \;=\; \lambda\,. \tag{3.3.125}$$

The operators on the LHS's are always the same for the non-Abelian and Abelian loop equations, which is a general property, but the RHS's generically differ: Eqs. (3.3.122) and (3.3.124) for the contour in Fig. 3.34a coincide while Eqs. (3.3.123) and (3.3.125) for the contour in Fig. 3.34b differ. The solution to Eq. (3.3.125) is given by (3.3.110).

Problem 3.27 Prove Eq. (3.3.115) for the contours with one self-intersection.
Solution Let the intersection corresponds to the values s_1 and s_2 of the parameter s, *i.e.* $x_\mu(s_1) = x_\mu(s_2)$. Noting that only the vicinities of s_1 and s_2 contribute to the integral on the LHS of Eq. (3.3.115), we get

$$\begin{aligned}
& e_{\rho\nu} \int\limits_{C'} dx_\rho \oint\limits_{C} dy_\nu\, \delta^{(2)}(x-y)\, W(C_{yx})\, W(C_{xy}) \\
&= e_{\rho\nu} \dot{x}_\rho(s_1)\, \dot{x}_\nu(s_2) \int ds \int dt\, \delta^{(2)}\left((s-s_1)\dot{x}(s_1) - (t-s_2)\dot{x}(s_2)\right) \\
&\quad \times W\left(C_{x(s_2)x(s_1)}\right) W\left(C_{x(s_1)x(s_2)}\right) \\
&= \frac{e_{\rho\nu} \dot{x}_\rho(s_1)\, \dot{x}_\nu(s_2)}{\sqrt{\dot{x}^2_\mu(s_1)\, \dot{x}^2_\nu(s_2) - \left(\dot{x}_\mu(s_1)\, \dot{x}_\mu(s_2)\right)^2}} W\left(C_{x(s_2)x(s_1)}\right) W\left(C_{x(s_1)x(s_2)}\right) \\
&= W\left(C_{x(s_2)x(s_1)}\right) W\left(C_{x(s_1)x(s_2)}\right)
\end{aligned} \tag{3.3.126}$$

which is precisely the RHS of Eq. (3.3.115).

Remark on the string representation

A nice property of QCD_2 at large N_c is that the exponential of the area enclosed by the contour C emerges[14] for the Wilson loop average $W(C)$. This

[14] This is not true, as is already discussed, in the Abelian case for contours with self-intersections.

is as it should for the Nambu–Goto string (3.3.93). However, the additional pre-exponential factors (like that in Eq. (3.3.111)) are very difficult to interpret in the stringy language. They may become negative for large loops which is impossible for a bosonic string. This explicitly demonstrates in $d = 2$ the statement of the previous subsection that the Nambu–Goto string is not a solution of the large-N_c loop equation.

3.3.9 Gross–Witten transition in lattice QCD$_2$

The lattice gauge theory on a two-dimensional lattice is defined by the partition function (2.2.31) with $d = 2$:

$$Z_{2d}(\beta) = \int \prod_x \prod_{\mu=1,2} dU_{x,\mu}\, \mathrm{e}^{-\beta S[U]}, \tag{3.3.127}$$

where the action is given by Eq. (2.2.16).

A specifics of two dimensions is that the number of the lattice sites is equal to the number of the plaquettes. For this reason, we can always perform such a gauge transformation that the link variables are chosen to be equal unity along one of the axes, say

$$U_{x,1} = 1\,. \tag{3.3.128}$$

Hence, the partition function (3.3.127) factorizes:

$$Z_{\mathrm{2d}} = (Z_{\mathrm{1p}})^{N_p}\,, \tag{3.3.129}$$

where N_p stands for the number of plaquettes of the lattice and Z_{1p} is the one-matrix integral

$$Z_{\mathrm{1p}}(\beta) = \int dU\, \mathrm{e}^{\beta\left(\frac{1}{N_c}\mathrm{Re\,tr}\,U-1\right)}. \tag{3.3.130}$$

In other words, the plaquette variables of the lattice gauge theory can be treated in two dimensions as independent.

The correct interpretation of Eq. (3.3.130) is that it is the partition function of the one-plaquette model, *i.e.* the lattice gauge theory on a single plaquette. This is consistent with the gauge invariance.

The one-matrix model (3.3.130) can be easily solved in the large-N_c limit by loop equations.

Let us first introduce the proper 'observables' for the one-matrix model:

$$W_n = \left\langle \frac{1}{N_c}\,\mathrm{tr}\,U^n \right\rangle_{\mathrm{1p}}, \tag{3.3.131}$$

where the average is taken with the same weight as in Eq. (3.3.130). The interpretation of W_n's in the language of the single-plaquette model is that these are the Wilson loop averages for contours which go along the boundary of n stacked plaquettes.

In order to derive the loop equation for the one-matrix model, we proceed quite analogous to the derivation of the loop equation in the lattice gauge theory (given in Problem 3.22). Let us consider the obvious identity

$$0 = \langle\, \mathrm{tr}\, t^a U^n \rangle_{\mathrm{1p}}\,, \tag{3.3.132}$$

and perform the (infinitesimal) change

$$U \to U\left(1 - it^a\epsilon^a\right), \quad U^\dagger \to \left(1 + it^a\epsilon^a\right)U^\dagger \tag{3.3.133}$$

of the integration variable on the RHS of Eq. (3.3.132). Since the Haar measure is invariant under the change (3.3.133), we finally get

$$\begin{aligned} \frac{\beta}{2N_c^2}\left(W_{n-1} - W_{n+1}\right) &= \sum_{k=1}^{n} W_k W_{n-k} \qquad \text{for } n \geq 1\,, \\ W_0 &= 1\,, \end{aligned} \tag{3.3.134}$$

where

$$\frac{\beta}{N_c^2} = \frac{1}{\lambda} \tag{3.3.135}$$

and $\lambda \sim 1$ as $N_c \to \infty$.

Equation (3.3.134) has the following exact solution

$$W_1 = \frac{1}{2\lambda}; \qquad W_n = 0 \qquad \text{for } n \geq 2 \tag{3.3.136}$$

which reproduces the strong coupling expansion. The leading order of the strong coupling expansion turns out to be exact at $N_c = \infty$.

However, the solution (3.3.136) can not be the desired solution at any values of the coupling constant. Since W_k are (normalized) averages of unitary matrices, they must obey

$$W_n \leq 1\,, \tag{3.3.137}$$

which is not the case for W_1, given by Eq. (3.3.136), at small enough values of λ.

In order to find all solutions to Eq. (3.3.134), let us introduce the generating function

$$f(z) \equiv \sum_{n=0}^{\infty} W_n z^n \tag{3.3.138}$$

and rewrite Eq. (3.3.134) as the quadratic equation

$$fz - \frac{1}{z}(f-1) + W_1 = 2\lambda\left(f^2 - f\right) . \tag{3.3.139}$$

A formal solution to Eq. (3.3.139) is

$$f(z) = -\frac{1-2\lambda z - z^2}{4\lambda z} + \frac{\sqrt{\left(1+2\lambda z + z^2\right)^2 + 4z^2\left(2\lambda W_1 - 1\right)}}{4\lambda z}, \tag{3.3.140}$$

where the positive sign of the square root is chosen to satisfy $f(0) = 1$.

The RHS of Eq. (3.3.140) depends on an unknown function $W_1(\lambda)$ which must guarantee for $f(z)$ to be a holomorfic function of the complex variable z inside the unit circle $|z| < 1$. This is a consequence of the inequality (3.3.137) which stems from the unitarity of U's.

There exist two solutions for which $f(z)$ is holomorfic inside the unit circle: the strong coupling solution given for $\lambda \geq 1$ by Eq. (3.3.136) and the weak coupling solution given for $\lambda \leq 1$ by

$$W_1 = 1 - \frac{\lambda}{2} . \tag{3.3.141}$$

A comparison with Eq. (2.3.1) for $d = 2$ shows that the leading order of the weak coupling expansion is now exact. Therefore, $f(z)$ is given by two *different* analitic functions for $\lambda > 1$ and $\lambda < 1$.

At the point $\lambda = 1$, a phase transition occurs as was discovered by Gross and Witten [GW80] who first solved lattice QCD_2 in the large-N_c limit. This phase transition is of the third order since both the first and second derivatives of the partition function are continuous at $\lambda = 1$. The discontinuity resides only in the third derivative. This phase transition is pretty unusual from the point of view of statistical mechanics where phase transitions usually occur in the limit of an infinite volume (otherwise the partition function is analytic in temperature). Now the Gross–Witten phase transition occurs even for the single-plaquette model (3.3.130) in the large-N_c limit. In other words, the number of degrees of freedom is now infinite due to the internal symmetry group rather than an infinite volume.

Finally, we mention that since plaquette veriables are independent in lattice QCD_2, the Wilson loop average for a non-intersecting lattice contour C takes the form

$$W_C = \left(W_1\right)^A \tag{3.3.142}$$

where A is the area (in the lattice units) enclosed by the contour C. W_1 in this formula is given by Eq. (3.3.136) in the strong coupling phase ($\lambda \geq 1$) and Eq. (3.3.141) in the weak coupling phase ($\lambda \leq 1$).

The continuum formula (3.3.104) is recovered for small λ from Eq. (3.3.142) as follows:

$$W_C = \left(1 - \frac{\lambda a^2}{2}\right)^{A/a^2} \stackrel{a\to 0}{\longrightarrow} \mathrm{e}^{-\frac{\lambda}{2}A}, \tag{3.3.143}$$

where we have restored the a-dependence as is prescribed by the dimensional analysis.

The solution of $N_c = \infty$ lattice QCD_2 by the loop equations, which is described in this Subsection, was given in Refs. [PR80, Fri81].

Problem 3.28 Calculate the density of eigenvalues for the matrix U in the one-matrix model (3.3.130).

Solution Let us reduce U to the diagonal form

$$U = \mathrm{diag}\left(\mathrm{e}^{i\alpha_1}, \ldots, \mathrm{e}^{i\alpha_j}, \ldots, \mathrm{e}^{i\alpha_{N_c}}\right) . \tag{3.3.144}$$

The density of eigenvalues (or the spectral density), $\rho(\alpha)$, is then defined as a fraction of the eigenvalues which lie in the interval $[\alpha, \alpha + d\alpha]$. In other words, introducing the continuum variable $x = j/N_c$ $(0 \leq x \leq 1)$ in the large-N_c limit, we have

$$\rho(\alpha) = \frac{dx}{d\alpha} \geq 0 \tag{3.3.145}$$

which obey the obvious normalization

$$\int_{-\pi}^{\pi} d\alpha\, \rho(\alpha) = \int_0^1 dx = 1 . \tag{3.3.146}$$

Given $\rho(\alpha)$, we can calculate W_n by

$$W_n = \int_{-\pi}^{\pi} d\alpha\, \rho(\alpha) \cos n\alpha . \tag{3.3.147}$$

It is now clear from the definition (3.3.131), (3.3.138) that

$$f(z) \;=\; \int_{-\pi}^{\pi} d\alpha\, \rho(\alpha)\, \frac{1}{1 - z\,e^{-i\alpha}}\,. \tag{3.3.148}$$

Choosing $z = \exp(i\omega)$, we rewrite Eq. (3.3.148) as

$$f\left(e^{i\omega}\right) \;=\; \frac{1}{2} + \frac{i}{2}\int_{-\pi}^{\pi} d\alpha\, \rho(\alpha) \cot\frac{\omega - \alpha}{2}\,. \tag{3.3.149}$$

The discontinuity of this analytic function at $\omega = \alpha \pm i0$ then determines $\rho(\alpha)$.

Using the explicit solution (3.3.140), we formally obtain

$$\rho(\alpha) \;=\; \frac{1}{2\lambda\pi}\sqrt{\left(\cos\alpha + \lambda + \sqrt{1 - 2\lambda W_1}\right)\left(\cos\alpha + \lambda - \sqrt{1 - 2\lambda W_1}\right)}\,. \tag{3.3.150}$$

For W_1 given by Eqs. (3.3.136) and (3.3.141) for the strong and weak coupling phases, we get finally

$$\rho(\alpha) \;=\; \frac{1}{2\pi}\left(1 + \frac{1}{\lambda}\cos\alpha\right) \qquad \text{for } \lambda \geq 1 \tag{3.3.151}$$

and

$$\rho(\alpha) \;=\; \frac{1}{\lambda\pi}\cos\frac{\alpha}{2}\sqrt{\lambda - \sin^2\frac{\alpha}{2}} \qquad \text{for } \lambda \leq 1 \tag{3.3.152}$$

for the strong and weak coupling solutions, respectively. Notice, that (3.3.151) is non-negative for $\lambda \geq 1$ as it should due to the inequality (3.3.145). For $\lambda < 1$, the strong coupling solution (3.3.151) becomes negative somewhere in the interval $[-\pi, \pi]$ which can not happen for a dynamical system. This is the reason why the other solution (3.3.152) is realized for $\lambda < 1$. It has the support on the smaller interval $[-\alpha_c, \alpha_c]$ where $0 < \alpha_c < \pi$ is determined by the equation

$$\sin^2\frac{\alpha_c}{2} \;=\; \lambda \tag{3.3.153}$$

which always has a solution for $\lambda < 1$. The weak coupling spectral density (3.3.152) is non-negative for $\lambda \leq 1$.

For small λ, $\alpha_c = 2\sqrt{\lambda}$ so that

$$\rho(\alpha) \;=\; \frac{1}{2\lambda\pi}\sqrt{4\lambda - \alpha^2}\,. \tag{3.3.154}$$

As $\lambda \to 0$, $\rho(\alpha) \to \delta(\alpha)$ and U freezes, modulo a gauge transformation, near a unit matrix. This guarantees the existence of the continuum limit of QCD_2.

The spectral densities (3.3.151) and (3.3.152) were first calculated in Ref. [GW80] by a direct solution of the saddle-point equation at large N_c.

Problem 3.29 Calculate the density of eigenvalues for the Gaussian Hermitean one-matrix model which is defined by the partition function

$$Z_{1h} \;=\; \int \prod_{i\leq j} d\Phi_{ij}\, e^{-N_c \frac{M}{2}\,\mathrm{tr}\,\Phi^2}\,. \tag{3.3.155}$$

Solution The difference from the previous problem is that Φ_{ij} is now a Hermitean $N_c \times N_c$ matrix whose eigenvalues p_i can take on values along the whole real axis.

Let us define

$$W_n \;\equiv\; \left\langle \frac{1}{N_c}\,\mathrm{tr}\,\Phi^n \right\rangle_{1h} \;=\; \int d\xi\, \rho(\xi)\, \xi^n\,. \tag{3.3.156}$$

The Schwinger–Dyson equations for the Hermitean one-matrix model can be derived [Wad81] using the invariance of the measure in (3.3.155) under the change

$$\Phi_{ij} \;\to\; \Phi_{ij} + \epsilon_{ij} \tag{3.3.157}$$

where ϵ_{ij} is (infinitesimal) Hermitean.

Proceeding as before and using the large-N_c factorization, we get the set of equations

$$\begin{aligned} M\,W_{n+1} \;&=\; \sum_{k=0}^{n-1} W_k W_{n-k} \qquad \text{for } n \geq 0\,, \\ W_0 \;&=\; 1\,. \end{aligned} \tag{3.3.158}$$

Introducing the generating function

$$W(p) \;\equiv\; \left\langle \frac{1}{N_c}\,\mathrm{tr}\,\frac{1}{p - \Phi} \right\rangle_{1h} \;=\; \sum_{n=0}^{\infty} \frac{W_n}{p^{n+1}}\,, \tag{3.3.159}$$

we rewrite Eq. (3.3.158) as the quadratic equation

$$M p\, W(p) - M \;=\; W^2(p) \tag{3.3.160}$$

whose solution reads

$$W(p) \;=\; \frac{Mp}{2} - \sqrt{\frac{M^2p^2}{4} - M}\,, \tag{3.3.161}$$

where we have chosen the minus sign of the root to satisfy the asymptotics $W(p) \to 1/p$ for large p as is prescribed by the definition (3.3.159).

$W(p)$ given by Eq. (3.3.161) is an analytic function of p with the cut from $p = -2/\sqrt{M}$ to $p = +2/\sqrt{M}$ along the real axis. The discontinuity of $W(p)$ at the cut determines the spectral density:

$$W(\xi \pm i0) \;=\; \frac{M\xi}{2} \mp i\pi\rho(\xi)\,. \tag{3.3.162}$$

We have

$$\rho(\xi) \;=\; \frac{M}{2\pi}\sqrt{\frac{4}{M}-\xi^2}. \qquad (3.3.163)$$

Notice that the spectral density is non-negative and has support on a finite interval $[-2/\sqrt{M}, 2/\sqrt{M}]$ in analogy with the unitary one-matrix model it the weak coupling regime. The spectral density (3.3.163) was first calculated by Wigner [Wig51] and is called Wigner's semicircle law.

We have already semicircle law in the previous Problem for the spectral density of the unitary one-matrix model at small λ (see Eq. (3.3.154)). This is because we can always substitute $U = \exp(i\Phi)$ where U is unitary and Φ is Hermitean and expand for small λ in Φ up to the quadratic term. We then obtain the Hermitean model (3.3.155) with $M = 1/\lambda$ from the unitary model (3.3.130).

3.4 Large-N_c reduction

The large-N_c reduction was first discovered by Eguchi and Kawai [EK82] who showed that the Wilson lattice gauge theory on a d-dimensional hypercubic lattice is equivalent at $N_c = \infty$ to the one on a hypercube with periodic boundary conditions. This construction is based on an extra $(Z_{N_c})^d$-symmetry which the reduced theory possesses to each order of the strong coupling expansion.

Soon after it was recognized that a phase transition occurs in the reduced model with decreasing the coupling constant, so that this symmetry is broken in the weak coupling regime. To cure the construction at weak coupling, the quenching prescription was proposed by Bhanot, Heller and Neuberger [BHN82] and elaborated by many authors. An elegant alternative reduction procedure based on twisting prescription was advocated by Gonzalez-Arroyo and Okawa [GAO83]. Each of these prescriptions results in the reduced model which is fully equivalent to multicolor QCD, both on the lattice and in the continuum.

While the reduced models look as a great simplification, since the space-time is reduced to a point, they still involve an integration over d infinite matrices which is in fact a continual path integral. It is not clear at the moment whether or not this is a real simplification of the original theory which can make it solvable. Nevertheless, the reduced models are useful and elegant representations of the original theory at large N_c.

We shall start this Section by a simplest example of a pure matrix scalar theory. The quenched reduced model for this case was proposed by Parisi [Par82] on the lattice end elaborated by Gross and Kitazawa [GK82] in the continuum, while the twisted reduced model was advocated by Eguchi and Nakayama [EN83]. Then we concentrate on the Eguchi–Kawai reduction of multicolor QCD both on the lattice and in the continuum.

3.4.1 Reduction of scalar field

Let us begin with a simplest example of a pure matrix scalar theory on a lattice whose partition function is defined by the path integral

$$Z = \int \prod_x \prod_{i\geq j} d\varphi_x^{ij}\, \mathrm{e}^{\sum_x N_c\, \mathrm{tr}\left(-V[\varphi_x]+\sum_\mu \varphi_x\varphi_{x+a\hat{\mu}}\right)}\,. \qquad (3.4.1)$$

Here φ_x is a $N_c \times N_c$ Hermitean matrix field and $V[\varphi]$ is some interaction potential, say

$$V[\varphi] \;=\; \frac{M}{2}\varphi^2+\frac{\lambda_3}{3}\varphi^3+\frac{\lambda_4}{4}\varphi^4\,. \tag{3.4.2}$$

The prescription of the large-N_c reduction is formulated as follows. We substitute

$$\varphi_x \;\to\; S_x \Phi S_x^\dagger\,, \tag{3.4.3}$$

where

$$\begin{aligned}[S_x]^{kj} \;&=\; \mathrm{e}^{\,ip_k^\mu x_\mu}\delta^{kj}\\ &=\; \mathrm{diag}\left(\mathrm{e}^{\,ip_1^\mu x_\mu},\ldots,\mathrm{e}^{\,ip_{N_c}^\mu x_\mu}\right)\end{aligned} \tag{3.4.4}$$

is a diagonal unitary matrix which eats the coordinate dependence, so that Φ does *not* depend on x.

The averaging of a functional $F[\varphi_x]$ which is defined with the same weight as in Eq. (3.4.1),

$$\Big\langle F[\varphi_x]\Big\rangle \equiv \frac{1}{Z}\int\prod_x d\varphi_x\,\mathrm{e}^{\sum_x N_c\,\mathrm{tr}\left(-V[\varphi_x]+\sum_\mu \varphi(x)\varphi(x+a\hat\mu)\right)}F[\varphi_x]\,, \tag{3.4.5}$$

can be calculated at $N_c=\infty$ by

$$\Big\langle F[\varphi_x]\Big\rangle \to a^{N_c d}\int_{-\frac{\pi}{a}}^{\frac{\pi}{a}}\prod_{\mu=1}^{d}\prod_{i=1}^{N_c}\frac{dp_i^\mu}{2\pi}\Big\langle F[S_x\Phi S_x^\dagger]\Big\rangle_{\mathrm{Reduced}} \tag{3.4.6}$$

where the average on the RHS is calculated [Par82] for the *quenched reduced model* whose averages are defined by

$$\Big\langle F[\Phi]\Big\rangle_{\mathrm{Reduced}} \equiv \frac{1}{Z_{\mathrm{Reduced}}}\times\int\prod_{i\geq j}d\Phi_{ij}\,\mathrm{e}^{-N_c\,\mathrm{tr}\,V[\Phi]+N_c\sum_{ij}|\Phi_{ij}|^2\sum_\mu\cos\left((p_i^\mu-p_j^\mu)a\right)}F[\Phi]\,. \tag{3.4.7}$$

The partition function of the reduced model reads

$$Z_{\mathrm{Reduced}} = \int\prod_{i\geq j}d\Phi_{ij}\,\mathrm{e}^{-N_c\,\mathrm{tr}\,V[\Phi]+N_c\sum_{ij}|\Phi_{ij}|^2\sum_\mu\cos\left((p_i^\mu-p_j^\mu)a\right)} \tag{3.4.8}$$

which can be deduced, modulo the volume factor, from the partition function (3.4.1) by the substitution (3.4.3).

Notice that the integration over the momenta p_i^μ on the RHS of Eq. (3.4.6) is taken *after* the calculation of averages in the reduced model. Such variables are usually called *quenched* in statistical mechanics which clarifies the terminology.

Since $N_c\to\infty$ it is not necessary to integrate over the quenched momenta in Eq. (3.4.6). The integral should be recovered if p_i^μ's would be uniformly distributed in a d-dimensional hypercube. Moreover, a similar property holds for the matrix integral over Φ as well, which can be substituted by its value at the saddle point configuration Φ_s:

$$\Big\langle F[\varphi_x]\Big\rangle \;\to\; F[S_x\Phi_s S_x^\dagger]\,, \tag{3.4.9}$$

where the momenta p_i^μ are uniformly distributed in the hypercube. Therefore, this saddle point configuration plays the role of a master field in the sense of Subsect. 3.2.7.

In order to show how Eq. (3.4.6) works, let us demonstrate how the planar diagrams of perturbation theory for the scalar matrix theory (3.4.1) are recovered in the quenched reduced model.

The quenched reduced model (3.4.8) is of the general type discussed in Sect. 3.2. The propagator is given by

$$\Big\langle \Phi_{ij}\Phi_{kl}\Big\rangle_{\mathrm{Gauss}} \;=\; \frac{1}{N_c}G\left(p_i-p_j\right)\delta_{il}\delta_{kj} \tag{3.4.10}$$

with

$$G\left(p_i-p_j\right) \;=\; \frac{1}{M-\sum_\mu\cos\left((p_i^\mu-p_j^\mu)a\right)}\,. \tag{3.4.11}$$

It is convenient to associate the momenta p_i and p_j in Eq. (3.4.11) with each of the two index lines representing the propagator and carrying, respectively, indices i and j. Remember, that these lines are oriented for a Hermitean matrix Φ and their orientation can be naturally associated with the direction of the flow of the momentum. The total momentum carried by the double line is p_i-p_j.

The simplest diagram which represents the correction of the second order in λ_3 to the propagator is depicted Fig. 3.36. The momenta p_i and p_j flows along the index lines i and j while the momentum p_k circulates along the

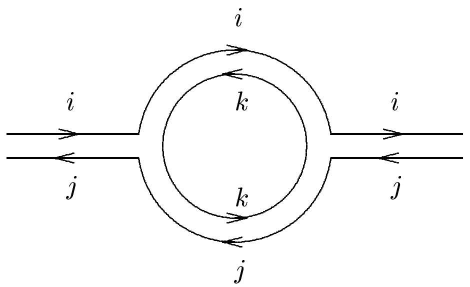


Fig. 3.36: Simplest planar diagram of the second order in λ_3 for the propagator in the quenched reduced model (3.4.8). The momentum p_i flows along the index line i. The momentum $p_i - p_j$ is associated with the double line ij.

index line k. The contribution of the diagram in Fig. 3.36 reads

$$\frac{\lambda_3^2}{N_c^2} G\left(p_i - p_j\right)^2 \sum_k G\left(p_i - p_k\right) G\left(p_k - p_j\right), \tag{3.4.12}$$

where the summation over the index k is just a standard one over indices forming a closed loop.

In order to show that the quenched-model result (3.4.12) reproduces the correction to the propagator in the original theory on an infinite lattice, we pass to the variables of the total momenta flowing along the double lines:

$$p_i - p_j = p; \quad p_k - p_j = q; \quad p_i - p_k = p - q\,, \tag{3.4.13}$$

which is obviously consistent with the momentum conservation at each of the two vertices of the diagram in Fig. 3.36. Since p_k's are uniformly distributed in the hypercube, the summation over k can be substituted as $N_c \to \infty$ by the integral

$$\frac{1}{N_c} \sum_k f\left(p_k\right) \Rightarrow a^d \int_{-\frac{\pi}{a}}^{\frac{\pi}{a}} \frac{d^d q}{(2\pi)^d} f\left(q\right). \tag{3.4.14}$$

The prescription (3.4.6) then gives the correct expression

$$a^d \frac{\lambda_3^2}{N_c} G\left(p\right)^2 \int_{-\frac{\pi}{a}}^{\frac{\pi}{a}} \frac{d^d q}{(2\pi)^d} G\left(q\right) G\left(p-q\right) \tag{3.4.15}$$

for the second-order contribution of the perturbation theory for the propagator on the lattice.

It is now clear how a generic planar diagram is recovered by the reduced model. We first represent the diagram by the double lines and associate the momentum p_i^μ with an index line carrying the index i. Then we write down the expression for the diagram in the reduced model with the propagator (3.4.11). Passing to the momenta flowing along the double lines, similar to Eq. (3.4.13), we get an expression which coincides with the integrand of the Feynman diagram for the theory on the whole lattice. It is crucial that such a change of variables can always be done for a planar diagram consistently with the momentum conservation at each vertex. The last step is that the summation over indices of closed index lines reproduces the integration over momenta associated with each of the loops according to Eq. (3.4.14). It is assumed that the number of loops is much less than N_c which is always true for a given diagram since N_c is infinite.

We thus have shown how planar diagrams of the lattice theory defined by the partition function (3.4.1) are recovered by the reduced model (3.4.8). The lattice was needed only as a regularization to make all integrals well-defined and was not crucial in the consideration. This construction can be formulated directly for the continuum theory [GK82, DW82] where the propagator turns into

$$G\left(p_i - p_j\right) = \frac{1}{\left(p_i - p_j\right)^2 + m^2} \tag{3.4.16}$$

and a Lorentz-invariant regularization can be achieved by choosing $p^2 < \Lambda^2$.

Remark on the twisted reduced model

An alternative reduction procedure is based on the twisting prescription [GAO83]. We again perform the unitary transformation (3.4.3) with the matrices S_x being expressed via a set of d (unitary) $N_c \times N_c$ matrices Γ_μ by

$$S_x = \Gamma_1^{x_1/a} \Gamma_2^{x_2/a} \Gamma_3^{x_3/a} \Gamma_4^{x_4/a} \tag{3.4.17}$$

where the coordinates of the (lattice) vector x_μ are measured in the lattice units. The matrices Γ_μ are explicitly constructed in Ref. [GAO83] and commute by

$$\Gamma_\mu \Gamma_\nu = Z_{\mu\nu} \Gamma_\nu \Gamma_\mu \tag{3.4.18}$$

with $Z_{\mu\nu} = Z_{\nu\mu}^\dagger$ being elements of Z_{N_c}.

For the twisting reduction prescription, Eq. (3.4.6) is valid providing the average on the RHS is calculated for the *twisted reduced model* which is defined by the partition function [EN83]

$$Z_{\mathrm{TRM}} = \int d\Phi\, \mathrm{e}^{-N_c \operatorname{tr} V[\Phi] + N_c \sum_\mu \operatorname{tr} \Gamma_\mu \Phi \Gamma_\mu^\dagger \Phi}. \tag{3.4.19}$$

Now the perturbation theory for the unreduced model (3.4.1) is recovered due to the explicit form of S_x given by Eq. (3.4.17).

We can change the order of Γ's in Eq. (3.4.17) defining a more general path-dependent factor

$$S_x = \mathbf{P} \prod_{l \in C_{x\infty}} \Gamma_\mu. \tag{3.4.20}$$

The path-ordered product in this formula runs over all links $l = (z, \mu)$ forming a path $C_{x\infty}$ from infinity to the point x.

Due to Eq. (3.4.18), changing the form of the path multiplies S_x by the Abelian factor

$$Z(C) = \prod_{\Box \in S : \partial S = C} Z_{\mu\nu}(\Box) \tag{3.4.21}$$

where (μ, ν) is the orientation of the plaquette $\Box$. The product runs over any surface spanned by the closed loop C which is obtained by passing the original path forward and the new path backward. Due to the Bianchi identity

$$\prod_{\Box \in cube} Z_{\mu\nu}(\Box) = 1 \tag{3.4.22}$$

where the product goes over six plaquettes forming a 3-dimensional cube on the lattice, the product on the RHS of Eq. (3.4.21) does not depend on the form of the surface S and is a functional of the loop C.

It is now easy to see that under this change of the path we get

$$[S_x]_{ij} [S_x^\dagger]_{kl} \to |Z(C)|^2 [S_x]_{ij} [S_x^\dagger]_{kl} \tag{3.4.23}$$

and the path-dependence is canceled because $|Z(C)|^2 = 1$. This is a general property which holds for the twisting reduction prescription of any even (*i.e.* invariant under the center Z_{N_c}) representation of $SU(N_c)$.

3.4.2 Reduction of Yang–Mills field

The statement of the Eguchi–Kawai reduction of the Yang–Mills field says that the theory on a d-dimensional space-time is equivalent at $N_c = \infty$ to the reduced model which is nothing but its reduction to a point. The action of the reduced model is given by

$$S_{\mathrm{EK}} = \frac{1}{2g^2 \Lambda^d} \operatorname{tr} [A_\mu, A_\nu]^2, \tag{3.4.24}$$

where A_μ are d space-independent matrices and Λ is a dimensionful parameter.

A naive statement of the Eguchi–Kawai reduction is that the averages coincide in both theories, for example,

$$\left\langle \frac{1}{N_c} \operatorname{tr} \mathbf{P}\, \mathrm{e}^{i \oint d\xi^\mu A_\mu(\xi)} \right\rangle_{\mathrm{d-dim}} = \left\langle \frac{1}{N_c} \operatorname{tr} \mathbf{P}\, \mathrm{e}^{i \oint d\xi^\mu A_\mu} \right\rangle_{\mathrm{EK}} \tag{3.4.25}$$

where the LHS is calculated with the action (2.1.14) and the RHS is calculated with the reduced action (3.4.24). Strictly speaking, this naive statement is valid only in $d = 2$ or supersymmetric case for the reason which will be explained in a moment.

The precise equivalence is valid only if the average of open Wilson loops vanish in the reduced model:

$$\left\langle \frac{1}{N_c} \operatorname{tr} \mathbf{P}\, \mathrm{e}^{i \int_{C_{yx}} d\xi^\mu A_\mu} \right\rangle_{\mathrm{EK}} = 0, \tag{3.4.26}$$

as it does in the d-dimensional theory due to the local gauge invariance under which

$$\left(\mathbf{P}\, \mathrm{e}^{i \int_{C_{yx}} d\xi^\mu A_\mu(\xi)} \right)_{ij} \to \left(\Omega^\dagger(y) \mathbf{P}\, \mathrm{e}^{i \int_{C_{yx}} d\xi^\mu A_\mu(\xi)} \Omega(x) \right)_{ij}. \tag{3.4.27}$$

The point is that this gauge invariance transforms in the reduced model into (global) rotation of the reduced field by constant matrices Ω:

$$A_\mu \to \Omega^\dagger A_\mu \Omega. \tag{3.4.28}$$

which does *not* guarantee such vanishing in the reduced model.

There exists, however, a symmetry of the reduced action (3.4.24) under the shift of A_μ by a unit matrix[15]:

$$A_\mu^{ij} \to A_\mu^{ij} + a_\mu \delta^{ij}, \tag{3.4.29}$$

[15]This symmetry is rigorously defined on a lattice where it is associated with a direction-dependent Z_{N_c} transformation.

which is often called the R^d-symmetry. Under the transformation (3.4.29), we get

$$\left(\mathbf{P}\,\mathrm{e}^{\,i\int_{C_{yx}} d\xi^\mu A_\mu}\right)_{ij} \to \mathrm{e}^{\,i(y^\mu - x^\mu)a_\mu}\left(\mathbf{P}\,\mathrm{e}^{\,i\int_{C_{yx}} d\xi^\mu A_\mu}\right)_{ij} \tag{3.4.30}$$

which guarantees, if the symmetry is not broken, the vanishing of the open Wilson loops

$$W_{\mathrm{EK}}\left(C_{yx}\right) \equiv \left\langle \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\mathrm{e}^{\,i\int_{C_{yx}} d\xi^\mu A_\mu}\right\rangle_{\mathrm{EK}} = 0 \tag{3.4.31}$$

in the reduced model.

The equivalence of the two theories can then be shown using the loop equation which reads for the reduced model

$$\begin{aligned}
\partial^x_\mu \frac{\delta}{\delta\sigma_{\mu\nu}(x)} W_{\mathrm{EK}}(C) &= \left\langle \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}[A_\mu,[A_\mu,A_\nu]]\,\mathrm{e}^{\,i\oint_{C_{xx}} d\xi^\mu A_\mu}\right\rangle_{\mathrm{EK}} \\
&= \lambda\Lambda^d \left\langle \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\frac{\partial}{\partial A_\nu}\,\mathrm{e}^{\,i\oint_{C_{xx}} d\xi^\mu A_\mu}\right\rangle_{\mathrm{EK}} \\
&= \lambda\Lambda^d \oint\limits_C dy_\nu\, W_{\mathrm{EK}}\left(C_{yx}\right) W_{\mathrm{EK}}\left(C_{xy}\right) .
\end{aligned} \tag{3.4.32}$$

The RHS is pretty much similar to the one in Eq. (3.3.56) while $\delta^{(d)}(x-y)$ is missing.

This delta function can be recovered if the R^d symmetry is not broken since

$$W_{\mathrm{EK}}\left(C_{yx}\right) \sim \frac{\delta^{(d)}(x-y)}{\delta^{(d)}(0)} W_{\mathrm{EK}}\left(C_{yx}\right) \tag{3.4.33}$$

due to Eq. (3.4.31) for the open loops.

This is not a rigorous argument since a regularization is needed. What actually happens is the following. If we smear the delta function introducing

$$\delta^{(d)}_\Lambda(x) = \left(\frac{\Lambda}{\sqrt{2\pi}}\right)^d \mathrm{e}^{\,-x^2\Lambda^2/2}\,, \tag{3.4.34}$$

then

$$\frac{1}{\delta^{(d)}_\Lambda(0)}\left(\delta^{(d)}_\Lambda(0)\right)^2 \propto \Lambda^d\,\mathrm{e}^{\,-x^2\Lambda^2} \to \delta^{(d)}(x)\,, \tag{3.4.35}$$

reproducing the delta function.

3.4.3 R^d-symmetry in perturbation theory

Since N_c is infinite, the R^d-symmetry can be broken spontaneously. The point is that the large-N_c limit plays the role of a statistical averaging, as is mentioned already in Subsection 3.2.8, and phase transitions are possible for infinite number of degrees of freedom. This phenomenon occurs in perturbation theory of the reduced model for $d \geq 3$.

The perturbation theory can be constructed expanding the fields around solutions of the classical equation

$$[A_\mu,[A_\mu,A_\nu]] = 0\,. \tag{3.4.36}$$

Any diagonal matrix

$$A^{\mathrm{cl}}_\mu \equiv p_\mu = \mathrm{diag}\left\{p^{(1)}_\mu,\ldots,p^{(N_c)}_\mu\right\} \tag{3.4.37}$$

is a solution to Eq. (3.4.36).

The perturbation theory of the reduced model can be constructed expanding around the classical solution (3.4.37):

$$A_\mu = A^{\mathrm{cl}}_\mu + gA^{\mathrm{q}}_\mu\,, \tag{3.4.38}$$

where A^{q}_μ is off-diagonal.

Substituting (3.4.38) into the action (3.4.24), we get

$$S_{\mathrm{EK}} = \mathrm{tr}\left\{\frac{1}{2}[p_\mu,A^{\mathrm{q}}_\nu]^2 - \frac{1}{2}[p_\mu,A^{\mathrm{q}}_\mu]^2\right\} + \text{higher orders}\,. \tag{3.4.39}$$

To fix the gauge symmetry (3.4.28), it is convenient to add

$$S_{\mathrm{g.f.}} = \mathrm{tr}\left\{\frac{1}{2}[p_\mu,A^{\mathrm{q}}_\mu]^2 + [p_\mu,b][p_\mu,c]\right\}\,, \tag{3.4.40}$$

where b and c are ghosts.

The sum of (3.4.39) and (3.4.40) gives

$$S_2 = \mathrm{tr}\left\{\frac{1}{2}[p_\mu,A^{\mathrm{q}}_\nu]^2 + [p_\mu,b][p_\mu,c]\right\} \tag{3.4.41}$$

up to quadratic order in A^{q}_μ.

Doing the Gaussian integral over A^{q}_ν, we get at the one-loop level:

$$\int dp_\mu dA^{\mathrm{q}}_\mu\,\mathrm{e}^{\,-S_2}\ldots = \int \prod_{k=1}^{N} dp^{(k)}_\mu \prod_{i<j}\left[(p^{(i)}_\mu - p^{(j)}_\mu)^2\right]^{1-d/2}\ldots\,, \tag{3.4.42}$$

where the integration over p_μ accounts for equivalent classical solutions.

For $d = 1$ the product on the RHS of Eq. (3.4.42) reproduces the Vandermonde determinant. For $d = 2$ it vanishes and does not affect dynamics. For $d \geq 3$ the measure is singular and the eigenvalues collapse. This leads us to a spontaneous breakdown of the R^d in perturbation theory.

The equivalence between the $N_c = \infty$ Yang–Mills theory on a whole space and the reduced model can be provided [BHN82] introducing a quenching prescription similar to the one described in Subsection 3.4.1. Then no collapse of eigenvalues happens and d-dimensional planar graphs are reproduced by the reduced model. More about the quenching prescription in Yang–Mills theory can be found in the reviews [Mig83, Das87] and cited there original papers.

Remark on supersymmetric case

In a supersymmetric gauge theory, there is an extra contribution from fermions to the exponent on the RHS of Eq. (3.4.42). Since the integration over fermions results in the extra factor $[(p_\mu^{(i)} - p_\mu^{(j)})^2]^{\,\mathrm{tr}\, I/2}$, this yields finally the exponent $1 - d/2 + \,\mathrm{tr}\, I/2$. It vanishes in $d = 4$ for either Majorana or Weyl fermions and in $d = 10$ for the Majorana–Weyl fermions. Therefore, the R^d-symmetry is not broken and no quenching is needed in the supersymmetric case [MK83, IKKT97].

3.4.4 Twisted reduced model

The continuum version of the twisted reduced model can be constructed [GAK83] by substituting $A_\mu \to A_\mu - \gamma_\mu$ into the action (3.4.24), where the matrices γ_μ obey the commutation relation

$$[\gamma_\mu, \gamma_\nu] = B_{\mu\nu} I \,, \tag{3.4.43}$$

where $B_{\mu\nu}$ is an antisymmetric tensor and d is even. This is possible only for infinite Hermitean matrices (operators). An example of such matrices is x and p operators in quantum mechanics. Eq. (3.4.43) is a continuum version of Eq. (3.4.18).

The Wilson loop averages in the twisted reduced model are defined by

$$W_{\mathrm{TEK}}\,(C_{yx}) = \left\langle \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\mathrm{e}^{-i\int_{C_{yx}} d\xi^\mu \gamma_\mu} \frac{1}{N_c}\,\mathrm{tr}\,\mathbf{P}\,\mathrm{e}^{\,i\int_{C_{yx}} d\xi^\mu A_\mu} \right\rangle_{\mathrm{TEK}} . \tag{3.4.44}$$

They vanish for open loops which is provided by the vanishing of the trace of the path-ordered exponential of γ_μ in this definition. For closed loops

this factor does not vanish and is needed to provide the equivalence with d-dimensional Yang–Mills perturbation theory, since the classical extrema of the twisted reduced model are $A_\mu^{\mathrm{cl}} = \gamma_\mu$ and the perturbation theory is constructed expanding around this classical solution.

The proof of the equivalence can be done using the loop equation quite similarly to that of Subsect. 3.4.2 for the Eguchi–Kawai model with an unbroken R^d symmetry.

References to Chapter 3

[Ans59] 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).

[BGSN82] R.A. Brandt, A. Gocksch, M. Sato, and F. Neri, *Loop space*, Phys. Rev. **D26** (1982) 3611.

[BHN82] G. Bhanot, U. Heller and H. Neuberger, *The quenched Eguchi–Kawai model*, Phys. Lett. **113B** (1982) 47.

[BNS81] R.A. Brandt, F. Neri and M. Sato, *Renormalization of loop functions for all loops*, Phys. Rev. **D24** (1981) 879.

[BNZ79] 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.

[BP75] 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).

[Bra80] N. Bralić, *Exact computation of loop averages in two-dimensional Yang–Mills theory*, Phys. Rev. **D22** (1980) 3090.

[BW93] E. Brézin and S. Wadia, *The large N expansion in quantum field theory and statistical physics: from spin systems to 2-dimensional gravity*, World Sci., 1993.

[CJP74] S. Coleman, R. Jackiw and H.D. Politzer, *Spontaneous symmetry breaking in the $O(N)$ model at large N*, Phys. Rev. **D10** (1974) 2491.

[CMS93] W. Chen, Y. Makeenko, and G.W. Semenoff, *Four-fermion theory and the conformal bootstrap*, Ann. Phys. **228** (1993) 341.

[CLS82] P. Cvitanović, P.G. Lauwers, and P.N. Scharbach, *The planar sector of field theories*, Nucl. Phys. **B203** (1982) 385.

[Col79] 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).

[Cvi81] P. Cvitanović, *Planar perturbation expansion*, Phys. Lett. **99B** (1981) 49.

[Das87] S.R. Das, *Some aspects of large-N theories*, Rev. Mod. Phys. **59** (1987) 235.

[Dou95] M.R. Douglas, *Stochastic master fields*, Phys. Lett. **B344** (1995) 117.

[DV80] V.S. Dotsenko and S.N. Vergeles, *Renormalizability of phase factors in non-Abelian gauge theory*, Nucl. Phys. **B169** (1980) 527.

[DW82] S. Das and S. Wadia, *Translation invariance and a reduced model for summing planar diagrams in QCD*, Phys. Lett. **B117** (1982) 228.

[Egu79] T. Eguchi, *Strings in U(N) lattice gauge theory*, Phys. Lett. **87B** (1979) 91.

[EK82] T. Eguchi and H. Kawai, *Reduction of dynamical degrees of freedom in the large-N gauge theory*, Phys. Rev. Lett. **48** (1982) 1063.

[EN83] T. Eguchi and R. Nakayama, *Simplification of Quenching Procedure for Large N Spin Models*, Phys. Lett. **122B** (1983) 59.

[Fel86] M.N. Feller, *Infinite-dimensional elliptic equations and operators of the type by P. Lèvy*, Sov. J. Usp. Mat. Nauk. **41** (1986) 97.

[Foe79] D. Foerster, *Yang–Mills theory — a string theory in disguise*, Phys. Lett. **87B** (1979) 87.

[Fri81] D. Friedan, *Some nonabelian toy models in the large N limit*, Commun. Math. Phys. **78** (1981) 353.

[GAO83] A. Gonzalez-Arroyo and M. Okawa, *The twisted Eguchi-Kawai model: A reduced model for large N lattice gauge theory*, Phys. Rev., **D27** (1983) 2397.

[GAK83] A. Gonzalez-Arroyo and C. P. Korthals Altes, *Reduced model for large N continuum field theories*, Phys. Lett. **B131** (1983) 396.

[GG95] R. Gopakumar and G.J. Gross, *Mastering the master field*, Nucl. Phys. **B451** (1995) 379.

[GK82] D.J. Gross and Y. Kitazawa, *A quenched momentum prescription for large-N theories*, Nucl. Phys. **B206** (1982) 440.

[GN74] D.J. Gross and A. Neveu, *Dynamical symmetry breaking in asymptotically free field theories*, Phys. Rev. **D10** (1974) 3235.

[GN79a] J.L. Gervais and A. Neveu, *The quantum dual string wave functional in Yang–Mills theories*, Phys. Lett. **80B** (1979) 255.

[GN79b] J.L. Gervais and A. Neveu, *Local harmonicity of the Wilson loop integral in classical Yang–Mills theory*, Nucl. Phys. **B153** (1979) 445.

[GN80] J.L. Gervais and A. Neveu, *The slope of the leading Regge trajectory in quantum chromodynamics*, Nucl. Phys. **B163** (1980) 189.

[Gra91] J.A. Gracey, *Calculation of exponent η to $O(1/N^2)$ in the $O(N)$ Gross Neveu model*, Int. J. Mod. Phys. **A6** (1991) 395, 2755(E).

[GW70] D. Gross and J. Wess, *Scale invariance, conformal invariance, and the high-energy behavior of scattering amplitudes*, Phys. Rev. **D2** (1970) 753.

[GW80] D. Gross and E. Witten, *Possible third-order phase transition in the large-N lattice gauge theory*, Phys. Rev. **D21** (1980) 446.

[Jac72] 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.

[Joh61] K. Johnson, *Solution of the equations for the Green's functions of a two dimensional relativistic field theory*, Nuovo Cim. **20** (1961) 773.

[Haa81] O. Haan, *Large N as a thermodynamic limit*, Phys. Lett. **106B** (1981) 207.

[HM89] M.B. Halpern and Yu.M. Makeenko, *Continuum-regularized loop-space equation*, Phys. Lett. **218B** (1989) 230.

[Hoo74a] G. 't Hooft, *A planar diagram theory for strong interactions*, Nucl. Phys. **B72** (1974) 461.

[Hoo74b] G. 't Hooft, *A two-dimensional model for mesons*, Nucl. Phys. **B75** (1974) 461.

[IKKT97] N. Ishibashi, H. Kawai, Y. Kitazawa and A. Tsuchiya, *A large-N reduced model as superstring*, Nucl. Phys. **B498** (1997) 467.

[KK80] V.A. Kazakov and I.K. Kostov, *Non-linear strings in two-dimensional $U(\infty)$ theory*, Nucl. Phys. **B176** (1980) 199.

[KK81] V.A. Kazakov and I.K. Kostov, *Computation of the Wilson loop functional in two-dimensional $U(\infty)$ lattice gauge theory*, Phys. Lett. **105B** (1981) 453.

[KNN77] J. Koplik, A. Neveu and S. Nussinov, *Some aspects of the planar perturbation series*, Nucl. Phys. **B123** (1977) 109.

[Mak83] Yu.M. Makeenko, *Large N*, in Gauge Theories of the Eighties, eds. R. Raitio and J. Lindfors, Lecture Notes in Physics **181**, Springer–Verlag, 1983, p. 67.

[Mak88] Yu.M. Makeenko, *Polygon discretization of the loop-space equation*, Phys. Lett. **212B** (1988) 221.

[Mak94] Y. Makeenko, *Exact multiparticle amplitudes at threshold in large-N component ϕ^4 theory*, Phys. Rev. **50** (1994) 4137.

[Man79] S. Mandelstam, *Charge-monopole duality and the phases of non-Abelian gauge theories*, Phys. Rev. **D19** (1979) 2391.

[Mig71] A.A. Migdal, *On hadronic interactions at small distances*, Phys. Lett. **37B** (1971) 98.

[Mig81] A.A. Migdal, *QCD = Fermi string theory*, Nucl. Phys. **189** (1981) 253.

[Mig83] A.A. Migdal, *Loop equations and 1/N expansion*, Phys. Rep. **102** (1983) 199.

[MK83] R.L. Mkrtchyan and S.B. Khokhlachev, *Reduction of the $U(\infty)$ theory to a model of random matrices*, JETP Lett. **37** (1983) 160.

[MM79] Yu.M. Makeenko and A.A. Migdal, *Exact equation for the loop average in multicolor QCD*, Phys. Lett. **88B** (1979) 135.

[MM80] Yu.M. Makeenko and A.A. Migdal, *Self-consistent area law in QCD*, Phys. Lett. **97B** (1980) 235.

[MM81] 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).

[MS69] G. Mack and A. Salam, *Finite-component field representation of the conformal group*, Ann. Phys. **53** (1969) 144.

[Nam79] Y. Nambu, *QCD and the string model*, Phys. Lett. **80B** (1979) 372.

[OP81] P. Olesen and J.L. Petersen, *The Makeenko–Migdal equation in a domained QCD vacuum*, Nucl. Phys. **181** (1981) 157.

[Par75] G. Parisi, *The theory of non-renormalizable interaction*, Nucl. Phys. **B100** (1975) 368.

[Par82] G. Parisi, *A simple expression for planar field theories*, Phys. Lett. **112B** (1982) 463.

[Pol70] A.M. Polyakov, *Conformal symmetry of critical fluctuations*, Pis'ma v ZhETF **12** (1970) 381) (English transl.: JETP Lett. **12** (1970) 381).

[Pol75] A.M. Polyakov, *Interaction of Goldstone particles in two dimensions. Applications to ferromagnets and massive Yang–Mills fields*, Phys. Lett. **59B** (1975) 79.

[Pol79] A.M. Polyakov, *String representations and hidden symmetries for gauge fields*, Phys. Lett. **82B** (1979) 247.

[Pol80] A.M. Polyakov, *Gauge fields as rings of glue*, Nucl. Phys. **B164** (1980) 171.

[PR80] G. Paffuti and P. Rossi, *A solution of Wilson's loop equation in lattice QCD_2*, Phys. Lett. **92B** (1980) 321.

[Sch74] H.J. Schnitzer, *Nonperturbative effective potential for $\lambda\varphi^4$ theory in the many-particle limit*, Phys. Rev. **D10** (1974) 1800.

[Sta68] H.E. Stanley, *Spherical model as the limit of infinite spin dimensionality*, Phys. Rev. **176** (1968) 718.

[Str57] R.L. Stratonovich, DAN SSSR **115** (1957) 1097.

[Tav93] J.N. Tavares, *Chen integrals, generalized loops and loop calculus*, preprint DF/IST 5.93 (May, 1993), hep-th/9305173.

[Tut62] W.T. Tuttle, *A census of planar triangulations*, Can. J. Math. **14** (1962) 21.

[VDN95] D.V. Voiculescu, K.J. Dykema and A. Nica, *Free Random Variables*, AMS, Providence 1992.

[Ven76] G. Veneziano, *Some aspects of a unified approach to gauge, dual and Gribov theories*, Nucl. Phys. **B117** (1976) 519.

[Wad81] S.R. Wadia, *Dyson–Schwinger equations approach to the large-N limit: Model systems and string representation of Yang–Mills theory*, Phys. Rev. **D24** (1981) 970.

[Wig51] E.P. Wigner, Ann. Math. **53**, 36 (1951); Proc. Cambridge Philos. Soc. **47**, 790 (1951).

[Wil73] K.G. Wilson, *Quantum field-theory models in less than 4 dimensions*, Phys. Rev. **D7** (1973) 2911.

[Wit79] 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.