Advanced Topics in Quantum Field Theory

FRANÇOIS GELIS

Institut de Physique Théorique, CEA-Saclay

MASTER HEP, ÉCOLE POLYTECHNIQUE

2017–2018

Foreword

This manuscript is made of lectures notes for the course "Quantum Field Theory 3" given at the Master "High Energy Physics" of Ecole Polytechnique in 2017–2018. Since it is a sequel of the course "Quantum Field Theory 2", whose lecture notes are printed in a separate volume, the present text contains some references to this previous volume. In order to avoid collisions in the numbering of sections and equations, the previous volume contains the chapters 1 through 7, and the chapters of the present volume are numbered 8 through 14.

Contents

Chapter 8

Effective field theories

Until now, we have discussed various quantum field theories (the electroweak theory and quantum chromodynamics) that are believed to provide a unified description of all particle physics up to the scale of electroweak symmetry breaking, i.e. roughly $\Lambda_{\rm EW} \sim 200$ GeV. However, it is hard to imagine that there isn't some kind of new physical phenomena (new particles, new interactions) at higher energy scales (so far out of reach of experimental searches). An interesting question is therefore to understand why the Standard Model is such a good description of physics below the electroweak scale, despite the fact that it does not contain any of the physics at higher scale. In other words, despite the fact that there is distinct physics on scales that span many orders of magnitude, why can "low energy" phenomena be described by ignoring most of the higher scales? The same question could be asked in other areas: for instance, why can chemistry (i.e. phenomena of atomic bonding in molecules) get away without any of the complications of quantum electrodynamics? The general question is that of the separation between various physical scales.

In the context of quantum field theory, such a low energy description is called an *effective theory*. The basic idea is that most of the details of an underlying more fundamental (i.e. valid at higher energy) description are not important at lower energies, except for a small number of parameters. As we shall see in this chapter, effective field theories may occur in several situations:

- *Top-down* : the quantum field theory which is valid at higher energy is known, but it is unnecessarily complicated to describe phenomena at lower energy scales. A typical example is that of a theory that contains particles that are much heavier than the energy scale of interest (e.g., the top quark in quantum chromodynamics, while one is interested in interactions at the GeV scale). In this case, the effective theory "integrates out" the higher mass particles in order to obtain a simpler theory.

- *Bottom-up* : we have a theory believed to be valid at a given energy scale, but have no clear idea of what may exist at higher scales. In this case, one may view the existing theory as an effective description of some (so far unknown) more fundamental theory at higher energy, and try to complete it by adding new (higher dimensional, and therefore usually non-renormalizable in four dimensions) local interactions to it.

- *Symmetry driven* : even when the underlying theory is known, its direct application may be rendered very impractical because the physics of interest involves some non-perturbative phenomena, such as the formation of bound states (for instance, in QCD at low energy, the quarks and gluons cease to be the relevant degrees of freedom and the physical excitations are the light hadrons). An effective theory for these bound states may be constructed from the requirement that it should be consistent with the symmetries of the underlying theory. This case differs from the top-down approach in the sense that the low energy description is not constructed by integrating out the high scales, but solely from symmetry considerations.

In the top-down approach, where the fundamental underlying theory is known, the goal of obtaining an effective description for low energy phenomena could in principle be achieved by the renormalization group. In particular, the functional renormalization group introduced in the section 7.5.3 allows to evolve from an ultraviolet classical action towards a low energy quantum effective action, by progressively integrating out layers of lower and lower momentum. There is nothing wrong with this approach, but one has to keep in mind that the effective action obtained in this way is usually extremely complicated and cumbersome to use in practical applications (in particular, it could have infinitely many effective interactions, all of which are in general non-local). In a sense, the quantum effective action that results from the RG evolution is much more complex that the original ultraviolet action, and the gain in terms of simplicity is rather dubious. In contrast, the concept of effective theory that we are aiming at in this chapter is a field theory in which the ultraviolet physics is encapsulated into a finite number of local operators, with coupling constants that may depend on the energy scale and on the properties of the degrees of freedom that have been integrated out.

8.1 General principles of effective theories

8.1.1 Low energy effective action

For the purpose of this general discussion, let us consider a quantum field theory in which the fields are collectively denoted ϕ (this may be a single field, or a collection of several fields) and a classical action $S[\phi]$. We view this theory as the high energy theory, and we wish to construct an alternative description applicable to low energy phenomena, below some energy scale Λ. To this effect, let us assume that we can split the field into a low frequency part (soft) and a high frequency part (hard),

$$\phi \equiv \phi_{\text{S}} + \phi_{\text{H}} \; . \tag{8.1}$$

This separation may be achieved by a cutoff in Fourier space, but the details of how this is done are not important at this level of discussion. The classical action of the original theory is thus a function of ϕ_{S} and ϕ_{H} and the path integration is over the soft and the hard components of the field. Now, assume that we are interested in calculating the expectation value of an observable that depends only on the soft component of the field, $\mathcal{O}(\phi_{\text{S}})$. Then, we may write

$$\langle \mathcal{O} \rangle = \int \left[\mathrm{D}\phi_{\text{S}} \mathrm{D}\phi_{\text{H}} \right] \, e^{\mathrm{i}S[\phi_{\text{S}},\phi_{\text{H}}]} \, \mathcal{O}(\phi_{\text{S}}) = \int \left[\mathrm{D}\phi_{\text{S}} \right] \, e^{\mathrm{i}S_\Lambda[\phi_{\text{S}}]} \, \mathcal{O}(\phi_{\text{S}}) \; , \tag{8.2}$$

where in the second equality we have defined

$$e^{i\mathcal{S}_\Lambda[\phi_S]} \equiv \int \left[D\phi_H\right] e^{i\mathcal{S}[\phi_S,\phi_H]} . \tag{8.3}$$

$\mathcal{S}_\Lambda[\phi_S]$ is the action of the low energy effective theory. Using the operator product expansion, it may be written as a sum of local operators, possibly infinitely many of them:

$$\mathcal{S}_\Lambda[\phi_S] \equiv \int d^d x \sum_n \lambda_n \mathcal{O}_n . \tag{8.4}$$

8.1.2 Power counting

The behavior of the couplings λ_n can be inferred from dimensional analysis. For the sake of this discussion, let us consider the case where ϕ is a scalar field, whose mass dimension is $\phi \sim (\text{mass})^{(d-2)/2}$ in d spacetime dimensions. If the operator $\mathcal{O}_n$ contains N_n powers of the field ϕ and D_n derivatives, its dimension is

$$\mathcal{O}_n \sim (\text{mass})^{d_n} \quad \text{with } d_n = D_n + N_n \tfrac{d-2}{2} , \tag{8.5}$$

and it must be accompanied with a coupling λ_n whose dimension is $\lambda_n \sim (\text{mass})^{d-d_n}$. Assuming that the cutoff Λ is the only dimensionful parameter that enters in the construction of the effective theory (except for the field operator and derivatives, that enter in the operators $\mathcal{O}_n$), we must have $\lambda_n = \Lambda^{d-d_n} g_n$, where g_n is a dimensionless constant, whose numerical value is typically of order one.

Consider now the application of this effective theory to the study of a phenomenon characterized by a single energy scale E. On dimensional grounds, we have

$$\int d^d x \, \mathcal{O}_n \sim E^{d_n - d} . \tag{8.6}$$

Combined with the corresponding coupling constant, the contribution of this operator would be of order

$$\lambda_n \int d^d x \, \mathcal{O}_n \sim g_n \left(\frac{\Lambda}{E}\right)^{d-d_n} . \tag{8.7}$$

This estimate is the basis of the following classification of the operators that may enter in the action of the effective theory:

- $d_n > d$: the contribution of these operators is suppressed at low energy, i.e. when $E \ll \Lambda$. For this reason, these operators are called *irrelevant*. This does not mean that their contribution is not important and interesting, since there may be observables for which they are the sole contribution. Note also that these operators are non-renormalizable by the standard power counting rules.
- $d_n = d$: the contribution of these operators does not depend on the ratio of scales E/Λ, except perhaps via logarithms. These operators are called *marginal*, and correspond to renormalizable operators.

- $d_n < d$: the contribution of these operators becomes more and more important as the energy scale decreases. These operators, called *relevant*, are super-renormalizable.

Recall also that a higher dimension d_n corresponds to operators of greater complexity (since in $d > 2$ the dimension increases with more powers of the field or more derivatives). Therefore, there is in general only a finite number of operators whose dimension is below a given value. For a given a cutoff Λ and an energy scale E, one must therefore only consider a finite number of operators in order to reach a given accuracy.

In a conventional quantum field theory, one usually insists on including only renormalizable operators, in order to avoid the proliferation of new couplings at each order or perturbation theory, and the usual statement of renormalizability amounts to saying that all infinities may be absorbed into the redefinition of a *finite* number of parameters of the theory, at every order of perturbation theory. In contrast, since a low energy effective theory may contain operators of dimension $d_n > d$, it is usually not renormalizable in this usual sense, but the cutoff Λ provides a natural way of keeping all the contributions finite. In this case, the power counting is organized by the fact that the cutoff Λ is also the dimensionful scale that enters in the couplings of negative mass dimension that come with operators of mass dimension greater than four. For instance, an operator of dimension 6 has a coupling constant that scales as Λ^{-2}, and physical observables may be expanded in powers of E/Λ, where E is some low energy scale. In the presence of such higher dimensional operators, the usual statement of renormalizability must now be replaced by a weaker assertion: namely, that all the ultraviolet divergences that occur at a given order in E/Λ can be absorbed into the redefinition of a finite number of parameters. More precisely, in order to calculate consistently effects of order Λ^{-r}, we must include all operators up to a mass dimension of $4 + r$. Thus, the number of constants that must be adjusted in the renormalization process grows as we go to higher order.

In the case of top-down effective theories, the renormalizability of the underlying field theory implies that the low energy physics depends on the ultraviolet only through the values of the relevant and marginal couplings. In addition, a small number of irrelevant couplings may matter in certain specific observables (e.g., if an irrelevant operator is the only one that contributes). In fact, if the cutoff of the effective theory is high enough compared to the physical energy scale of interest, the effective theory can have a very strong predictive power, despite the fact that it a priori contains an infinity of operators. But conversely, in a bottom-up approach where we try to extend a renormalizable theory by adding to it higher dimensional operators, the fact that the low energy theory is renormalizable implies that it is not sensitive to the scale of new physics (in other words, a renormalizable low energy theory cannot predict at which high energy scale it breaks down and is superseded by another theory).

8.1.3 Relevant operators

In fact, in an effective theory, the relevant operators (super-renormalizable) are often more troublesome than the relevant ones (non-renormalizable). Consider for instance the operator ϕ^2, that corresponds to the mass term in the effective Lagrangian and has dimension $\phi^2 \sim (\text{mass})^{d-2}$, and whose corresponding coupling has dimension $(\text{mass})^2$, i.e. $\lambda = g\,\Lambda^2$. Thus, small masses are not natural in a low energy effective theory: the natural scale of a mass is that of the cutoff Λ (the dimensionless coupling g is generically of order one). In order to obtain small masses in a low energy effective field theory, there must be some symmetry that prevents the corresponding mass term, such as:

- A gauge symmetry for spin 1 particles.
- A chiral symmetry for fermions.
- A spontaneous breaking of symmetry, so that some scalars are the corresponding massless Nambu-Goldstone bosons.
- Supersymmetry may also forbid certain types of mass terms (if unbroken, the mass must be strictly zero, and if broken, the mass will settle to a value close to the scale of supersymmetry breaking).

By that account, the Standard Model (without any supersymmetric extension) is not natural, since it does not contain any mechanism to prevent the mass of the Higgs scalar boson to be at a cutoff scale (possibly much higher than the electroweak scale) where the Standard Model is superseded by a more fundamental theory.

Likewise, relevant interaction terms have a large contribution to low energy observables, that scales like

$$\left(\frac{\Lambda}{E}\right)^{d-d_n} \gg 1 \quad \text{with } d > d_n \,. \tag{8.8}$$

Therefore, the existence of relevant interaction terms implies that the dynamics is strongly coupled at low energy. This may lead to the formation of bound states or condensates, which calls for a low energy effective theory that contains different degrees of freedom. An example is that of the identity operator, which is not forbidden by any symmetry and has mass dimension 0 (therefore, it is a relevant operator). Although this operator has no effect if added to the Lagrangian of a field theory (since it amounts to adding a constant to the potential energy), its coefficient becomes a cosmological constant if this field theory is minimally coupled to gravity[1]. From the power counting of the previous section, the natural value of the coupling constant in front of this operator is Λ^d. Thus, if we view the Standard Model as an effective theory, the cosmological constant should be at least as large as the fourth ($d = 4$) power of the cutoff at which the Standard Model is replaced by some other theory. This in sharp contrast with observations. Indeed, if the dark energy inferred from the measured acceleration of the expansion of the Universe is attributed to a cosmological constant, its value is many orders of magnitude below its natural value in quantum field theory (its corresponds to an energy density of the vacuum of the order of 10^{-47} GeV^4).

8.2 Example: Fermi theory of weak decays

As a first illustration of the concept of effective field theory, let us consider the case of Fermi's theory of weak interactions. Historically, this model was constructed before the advent of the electroweak gauge theory, and therefore it may be viewed as a bottom-up construction. Nowadays, since the electroweak theory provides us a more fundamental description of weak interactions, we may derive Fermi's theory in a top-down fashion, as a low energy approximation of a known high energy theory.

[1] This example illustrates an ambiguity one faces when coupling a field theory to gravity: only energy differences matter for the dynamics of the field theory, but the absolute value of the energy enters in the energy-momentum tensor that acts as a source in Einstein's equations.

8.2.1 Fermi theory as a phenomenological description

If we consider the Standard Model at a scale of the order of the nucleon mass, i.e. around a GeV, it contains only the leptons, the light quarks, and the massless gauge bosons (photon and gluons). Thus, this low energy truncation has no mechanism for weak decays. Nevertheless, one may write an effective coupling involving a proton, a neutron (here, we prefer to use hadrons, that are the states encountered in actual experimental situations), an electron and the corresponding neutrino. The most general local operator combining these four fields may be written as

$$\frac{g_{12}}{\Lambda^2}\,\big(\overline{\psi}_p\Gamma_1\psi_n\big)\big(\overline{\psi}_e\Gamma_2\psi_\nu\big)\;, \tag{8.9}$$

where g_{12} is a dimensionless constant, Λ is a dimensionful scale, and $\Gamma_{1,2}$ are matrices chosen in the following set

$$\Gamma_{1,2}\in\big\{1,\gamma_5,\gamma^\mu,\gamma^\mu\gamma_5,\underbrace{\tfrac{i}{4}[\gamma^\mu,\gamma^\nu]}_{\sigma^{\mu\nu}}\big\}\;. \tag{8.10}$$

Note that $\sigma^{\mu\nu}\gamma_5$ is not linearly independent from these matrices, since $\sigma^{\mu\nu}\gamma_5\propto\epsilon^{\mu\nu\rho\sigma}\sigma_{\rho\sigma}$, and therefore need not be included in this list. Thus, the most general Lorentz invariant Lagrangian involving these four fields reads

$$\begin{aligned}\mathcal{L}_{\text{eff}} \;=\;& \underbrace{\big(\overline{\psi}_p\gamma_\mu\psi_n\big)\big(\overline{\psi}_e\gamma^\mu(C_V+C'_V\gamma_5)\psi_\nu\big)+\big(\overline{\psi}_p\gamma_\mu\gamma_5\psi_n\big)\big(\overline{\psi}_e\gamma^\mu\gamma_5(C_A+C'_A\gamma_5)\psi_\nu\big)}_{\text{vector, axial}}\\
&+\underbrace{\big(\overline{\psi}_p\psi_n\big)\big(\overline{\psi}_e(C_S+C'_S\gamma_5)\psi_\nu\big)+\big(\overline{\psi}_p\gamma_5\psi_n\big)\big(\overline{\psi}_e\gamma_5(C_P+C'_P\gamma_5)\psi_\nu\big)}_{\text{scalar, pseudo-scalar}}\\
&+\underbrace{\big(\overline{\psi}_p\sigma_{\mu\nu}\psi_n\big)\big(\overline{\psi}_e\sigma^{\mu\nu}(C_T+C'_T\gamma_5)\psi_\nu\big)}_{\text{tensor}}\;.\end{aligned} \tag{8.11}$$

Note that the presence of certain terms violate some discrete symmetries. For instance, the primed terms $C'_{V,S,P,T}$ all violate parity, and T-invariance requires that the ratio C_i/C'_i be real for all $i\in\{V,A,S,P,T\}$. On the other hand, by confronting this effective Lagrangian with the existing data on weak decays, we learn that

$$\begin{aligned}&C_V=\Lambda^{-2}\quad\text{with }\;\Lambda\sim 350\,\text{GeV}\;,\\
&C_A\approx 1.25\times C_V\;,\\
&C_V\sim C'_V\;,\quad C_A\sim C'_A\;,\\
&\frac{C_{S,P,T}}{C_V}\;,\;\frac{C'_{S,P,T}}{C_V}\lesssim 1\%\;.\end{aligned} \tag{8.12}$$

The first of these results is an indication of the energy scale at which the Fermi theory breaks down and should be replaced by a more accurate microscopic description of weak decays, and the second one implies that this underlying theory is chiral . The fact that $C_{V,A}\sim C'_{V,A}$ is a sign of parity violation in weak interactions. Finally, the last property tells us that this microscopic interaction is not mediated by a scalar or a tensor with a mass less than $\sim$ 2 TeV. All these informations may be used in constraining the possible form of the theory that describes weak interactions at higher energies.

8.2.2 Fermi theory from the electroweak model

Let us now consider the opposite exercise: namely, start from the Lagrangian of the Standard Model and obtain the low energy effective theory of weak interactions by a matching procedure. We know that the $W^{\pm}$ bosons responsible for weak decays couple to left-handed fermions arranged in $SU(2)$ doublets:

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}_L \,, \quad \begin{pmatrix} d \\ u \end{pmatrix}_L \,, \tag{8.13}$$

where we have written only the relevant doublets for the decay $n \to pe\overline{\nu}_e$. In addition, we have to keep in mind that the mass eigenstates are misaligned with the weak interaction eigenstates in the quark sector. Thus, the vertex Wud contains a factor V_{ud} from the CKM matrix. With these ingredients, the tree level decay amplitude $d \to ue\overline{\nu}_e$ reads:

$$\mathcal{A} = \frac{g^2}{8} V_{ud} \frac{i}{k^2 - m_W^2} \left(\overline{\psi}_u \gamma^\mu (1-\gamma_5)\psi_d\right)\left(\overline{\psi}_e \gamma^\mu (1-\gamma_5)\psi_\nu\right) , \tag{8.14}$$

where k^μ is the 4-momentum carried by the intermediate W boson. In the low momentum limit, $k^2 \ll m_W^2$, this amplitude becomes independent of the momentum transfer and could have been generated by the following contact interaction

$$\mathcal{L}_{\text{eff}} = \frac{G_F}{\sqrt{2}} V_{ud} \left(\overline{\psi}_u \gamma^\mu (1-\gamma_5)\psi_d\right)\left(\overline{\psi}_e \gamma^\mu (1-\gamma_5)\psi_\nu\right) \quad \text{with} \quad \frac{G_F}{\sqrt{2}} \equiv \frac{g^2}{8 M_W^2} \,. \tag{8.15}$$

In order to obtain from this the physical decay amplitude $n \to pe\overline{\nu}_e$, we need the matrix element

$$\left\langle p \middle| \overline{\psi}_u \gamma^\mu (1-\gamma_5)\psi_d \middle| n \right\rangle \tag{8.16}$$

with initial and final nucleons instead of quarks. In the low momentum limit, it may be related to a similar matrix element with the spinors of the proton and neutron by

$$\left\langle p \middle| \overline{\psi}_u \gamma^\mu (1-\gamma_5)\psi_d \middle| n \right\rangle = \left\langle p \middle| \overline{\psi}_p \gamma^\mu (g_V - g_A \gamma_5)\psi_n \middle| n \right\rangle + \mathcal{O}(k^\mu) \,, \tag{8.17}$$

where $g_{V,A}$ are two constants that may be viewed as the zero momentum limit of some form factors. Then, by comparing the decay amplitudes obtained from the low energy effective theory guessed on the basis of phenomenological considerations, and the one obtained by starting from the electroweak theory, we obtain

$$\begin{aligned} C_V &= -C'_V = g_V \frac{g^2}{8 M_W^2} V_{ud} = \frac{1}{\Lambda^2} \,, \\ C_A &= -C'_A = -g_A \frac{g^2}{8 M_W^2} V_{ud} \,, \\ C_{S,P,T} &= C'_{S,P,T} = 0 \,. \end{aligned} \tag{8.18}$$

In this top-down approach, we see that the parity violation inferred from experimental evidence is in fact maximal in the electroweak theory, and that the scalar and tensor contributions are exactly zero. Note also that the scale Λ that we introduced by hand in the low energy effective theory does not coincide exactly with the mass of the heavy particle which is integrated out (in the present case, the W boson), but has the same order of magnitude. Finally, even though we performed here the matching at tree level, it is in principle possible to correct the coefficients of the low energy effective theories by electroweak and QCD loop corrections.

8.3 Standard model as an effective field theory

8.3.1 Standard Model

The Standard Model unifies the strong and electroweak interactions into a unique renormalizable field theory. Although it agrees with most observed phenomena[2], it is unreasonable to expect that the Standard Model remains an accurate description of particle physics to arbitrarily high energy scales. A more modest point of view is to consider the Standard Model as a low energy approximation of some more fundamental theory that we do not yet know. In this perspective, it would just be the zeroth order of some expansion,

$$\mathcal{L} = \underbrace{\mathcal{L}_{\text{SM}}}_{\Lambda^0} + \underbrace{\mathcal{L}^{(1)}}_{\Lambda^{-1}} + \underbrace{\mathcal{L}^{(2)}}_{\Lambda^{-2}} + \cdots \,, \tag{8.19}$$

and a natural endeavor is to construct the terms $\mathcal{L}^{(1,2,\cdots)}$, made of operators with mass dimension greater than four. By power counting, these operators must be suppressed by coupling constants that are inversely proportional to powers of some high energy scale Λ at which corrections to the Standard Model become important. In the construction of these corrections, one usually abides by the following constraints:

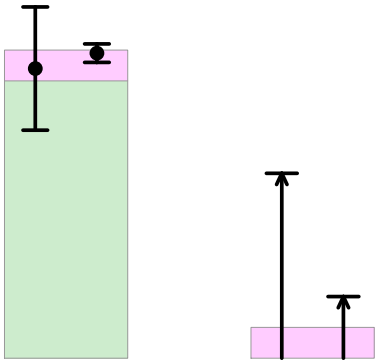
- Lorentz invariance is preserved to all orders in Λ^{-1},
- The $\mathrm{SU}(3) \times \mathrm{SU}(2) \times \mathrm{U}(1)$ gauge symmetry of the Standard Model remains a symmetry of the higher order corrections (the idea being that whatever is the more fundamental theory that underlies the Standard Model, it is more symmetric, not less),
- The corrections are built with the degrees of freedom of the Standard Model,
- The vacuum expectation value of the Higgs is not modified by the corrections.

As we have mentioned earlier, since the Standard Model is renormalizable, there is no way to determine the scale Λ within the Standard Model itself. Instead, one should enumerate the higher dimensional operators up to a certain mass dimension (which corresponds to a certain order in Λ^{-1}) and investigate their possible observable consequences. Experiments can then search for these effects, and either provide the values of some of the parameters introduced in $\mathcal{L}^{(1,2,\cdots)}$, or give lower bounds on the scale of new physics in case of a null observation. Note that there are two main classes of higher dimensional operators, illustrated in the figure 8.1:

- Operators that lead to corrections to processes already allowed in the Standard Model. These corrections may become potentially visible in more precise experiments.
- Operators that allow processes that were forbidden in the Standard Model. In this case, what is needed are more sensitive experiments, able to detect extremely rare events.

[2] One exception is the fact that neutrinos have masses, that does not have a very compelling explanation in the Standard Model – we shall return to this issue in the next subsection.

Figure 8.1: *Left: a higher dimensional operator provides a correction (magenta) to an observable which is non-zero in the Standard Model (green). Right: the higher dimensional operator allows a process that was impossible in the Standard Model. In the latter case, experiments usually provide an upper value for the yield of these very rare processes, that decreases as the sensitivity improves, thereby pushing higher up the energy scale of this new physics.*



8.3.2 Dimension 5 operators and neutrino masses

The right handed neutrinos are singlet under $SU(3)$ and $SU(2)$ and have a null electrical charge, which means that they do not feel any of the interactions of the Standard Model. As a consequence, all the neutrinos detected in experiments (via their weak interactions with the matter of the detector) are left handed neutrinos, implying that there is no direct evidence for the existence of right handed neutrinos. For this reasons, right handed neutrinos are usually not considered as a part of Standard Model.

The observation of neutrino oscillations, i.e. the fact that the flavor of a neutrino can change as it propagates, implies that there are non-zero mass differences between neutrinos[3]. Therefore, at most one of the neutrinos can be massless, and at least two of them must be massive.

Neutrino masses from the Higgs mechanism : Since the electroweak theory is chiral (right handed leptons are $SU(2)$ singlet, while the left handed ones belong to $SU(2)$ doublets), a naive Dirac mass term of the form $m_D \, \overline{\psi}_L \psi_R$ is not invariant under $SU(2)$. However, we may construct such a Dirac mass in the same way as for the other leptons, by starting from a Yukawa coupling involving the Higgs boson:

$$\lambda \left(\overline{\psi}_{L,i\alpha} \, \epsilon_{ij} \, \Phi^*_j \right) \psi_{R,\alpha} \, , \tag{8.20}$$

[3]Consider for instance a β decay: it produces an electron anti-neutrino (i.e. a weak interaction eigenstate) of definite momentum. If mass eigenstates are misaligned with the weak interaction eigenstates, then this neutrino may project on several mass eigenstates. Since the time evolution of the phase of a wavefunction depends on the mass of the particle, these mass eigenstates evolve slightly differently in time (unless all the neutrino masses are identical). At the detection time, this leads to a flavor decomposition which is different from the one at the time of production. Thus, the original electron anti-neutrino will be a mixture of electron, muon and tau anti-neutrinos. Conversely, the observation of this change of flavor implies mass differences in the neutrino sector.

where i, j are indices in the fundamental representation of $SU(2)$ and α is a Dirac index. The matrix $\epsilon \equiv it^2$ is the second generator of the fundamental algebra of $SU(2)$ Thanks to the contraction of the left handed spinor doublet with the Higgs field, we now have an $SU(2)$ invariant combination. Then, spontaneous symmetry breaking gives a non-zero expectation value v to the Higgs field, and this interaction term becomes a Dirac mass term for the neutrino, with a mass $m_{D} = \lambda v$. Generating the neutrino mass by this mechanism would place the neutrinos almost on the same footing as the other leptons, provided we add right handed neutrinos to the degrees of freedom of the Standard Model[4]. The only distinctive feature of the right handed neutrinos would be that they do not feel any of the gauge interactions of the Standard Model. For this reason, they are sometimes called *sterile neutrinos*. The main drawback of this solution is that it requires an even larger range of values of the Yukawa couplings, with no natural explanation.

Majorana neutrino masses : An alternative would be to have a Majorana mass for the left handed neutrinos of the Standard Model. Instead of introducing this mass term by hand, it can be generated via spontaneous symmetry breaking from a *Weinberg operator*:

$$\frac{c}{\Lambda} \left(\psi^{t}_{L,i\alpha} \epsilon_{ij} \Phi_j\right) \mathbf{C}_{\alpha\beta} \left(\Phi^{t}_{k} \epsilon_{kl} \psi_{L,l\beta}\right) , \tag{8.21}$$

where $\mathbf{C} \equiv \gamma^0\gamma^2$ is the charge conjugation operator. Firstly, note that this operator has mass dimension 5, hence the coupling constant proportional to Λ^{-1}. In fact, this operator is the only 5-dimensional operator that obeys the constraints listed in the previous section[5]. After spontaneous symmetry breaking, the Higgs field acquires a vacuum expectation value, leading to a Majorana mass term for the left handed neutrinos,

$$\frac{c\, v^2}{\Lambda} \nu^{t}_{L} \mathbf{C} \nu_{L} , \tag{8.22}$$

which corresponds to a Majorana mass $m_{M} = cv^2\Lambda^{-1}$. The appeal of this mechanism, is that a small mass of the neutrinos is naturally explained by a high scale Λ for the new physics. For instance, a neutrino mass of the order of 1 eV or below corresponds to $\Lambda \gtrsim 10^{13}$ GeV. Note that the operator in eq. (8.21) does not conserve lepton number, since it is not invariant under the following global transformation

$$\psi \to e^{i\alpha}\psi , \quad \psi^{\dagger} \to e^{-i\alpha}\psi^{\dagger} , \quad \psi^{t} \to e^{i\alpha}\psi^{t} . \tag{8.23}$$

For this reason, this alternative mechanism is clearly beyond the Standard Model. However, as long as gauge symmetries are preserved, the violation of lepton number is not considered particularly dramatic. In a sense, one may view the lepton conservation that exists in the Standard Model as accidental, being a consequence of the fact that only dimension-four operators are included.

[4] Whether this type of term is "beyond the Standard Model" is to a large extent a matter of definition. Before the observation of neutrino oscillations, the Standard Model was most often defined without right handed neutrinos, and therefore massless neutrinos. But it would have been equally acceptable to include right handed neutrinos from the start, with Yukawa couplings so small that their masses were too small to detect.

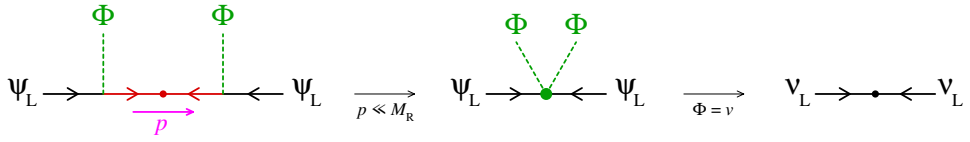
[5] $\psi^{t}_{L,i\alpha}\epsilon_{ij}\Phi_j$ and $\Phi^{t}_{k}\epsilon_{kl}\psi_{L,l\beta}$ are both $SU(2)$ invariant (but not Lorentz invariant), and the combination $\psi^{t}_{L,i\alpha}\mathbf{C}_{\alpha\beta}\psi_{L,l\beta}$ is Lorentz invariant. This combination is $SU(3)$ invariant only for the leptons (not for the quarks).

Weinberg operator from the low energy limit of another QFT : In the spirit of the bottom-up construction of an effective theory, the operator of eq. (8.21) can be obtained by exploring all the possibilities for dimension 5 operators built with the degrees of freedom of the Standard Model and some symmetry requirements. However, this operator can also be obtained in the low energy limit of a renormalizable quantum field theory. Consider an extension of the field content of the Standard Model, where we add a right handed neutrino ν_R with a very large Majorana mass M_R (much heavier than the electroweak scale), that also couples to the $SU(2)$ doublet containing the left handed neutrino and to the Higgs field via a Yukawa coupling,

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{SM} + \mathcal{L}_{\nu_R} \ , \\ \mathcal{L}_{\nu_R} &\equiv i\,\overline{\nu_R}\,\partial\!\!\!/\,\nu_R - y\,\left(\overline{\psi}_L\,\epsilon\Phi^*\right)\nu_R - y^*\,\overline{\nu_R}\left(\Phi^t\epsilon^\dagger\psi_L\right) \\ &\qquad + \frac{1}{2}\left(M_R\,\nu_R^t\,\mathbf{C}\,\nu_R + M_R^*\,\nu_R^{t*}\mathbf{C}\,\nu_R^*\right) . \end{aligned} \tag{8.24}$$

With two instances of the Yukawa coupling and a propagator of the heavy Majorana neutrino,

Figure 8.2: *Diagrammatic illustration of the see-saw mechanism. Left:* $\Phi\Phi\psi_L^t\psi_L$ *4-point function made with two Yukawa vertices and one insertion of the* ν_R *propagator. Middle:* $\Phi\Phi\psi_L^t\psi_L$ *local vertex obtained after integrating out the right handed neutrino. Right: Majorana mass term for* ν_L *obtained after spontaneous symmetry breaking.*



it is possible to build a (non-local) four-point function involving two Higgs fields and two left handed leptons (see the figure 8.2). At energies much lower than the mass M_R of the right handed neutrino, the intermediate propagator may be approximated by a constant[6]

$$\frac{i(\not{p} + M_R)\mathbf{C}}{p^2 - M_R^2} \underset{p \ll M_R}{\longrightarrow} -i\frac{\mathbf{C}}{M_R} \ , \tag{8.25}$$

which leads to the (local) Weinberg operator. The latter gives a Majorana mass for the left handed neutrino after the Higgs field has acquired a non-zero vacuum expectation value through spontaneous symmetry breaking. This mechanism is known as the *see-saw mechanism*[7].

8.3.3 Higher dimensional operators

The number of operators of mass dimension 6 is much larger, and lead to a broader array of possible phenomena. Even if we restrict to operators that conserve lepton and baryon number,

[6] The propagator of a Majorana fermion is that of a Dirac fermion multiplied by the charge conjugation matrix $\mathbf{C}$.

[7] More precisely, it corresponds to the *Type-I* see-saw mechanism. Type-II and Type-III see-saw mechanisms exist, that differ in the nature of the heavy particle that connects the $\Phi\Phi\psi_L^t\psi_L$ fields in the original four point function.

there are 80 independent operators. Instead of listing them, let us discuss an important result used to reduce the list of possible operators down to a smaller set of independent operators, thanks to the field equations of motion that result from the zeroth order Lagrangian (i.e. the Standard Model Lagrangian).

Operator removal from the Lagrangian : As an illustration of the principles at work in this reduction, consider the Lagrangian of a real scalar field with a quartic interaction, extended by two operators of dimension 6,

$$\mathcal{L} = \underbrace{\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4}_{\mathcal{L}^{(0)}} + \frac{1}{\Lambda^2}\left(\lambda_1\,\phi^6 + \lambda_2\,\phi^3\Box\phi\right) . \tag{8.26}$$

The equation of motion that follows from the zeroth order Lagrangian is

$$(\Box + m^2)\phi + \frac{\lambda}{6}\,\phi^3 = 0 . \tag{8.27}$$

Naively, it is tempting to replace the last term of the effective Lagrangian, $\phi^3\Box\phi$, by a sum of terms in ϕ^4 and ϕ^6. However, it is not totally clear that this is legitimate when this interaction term is inserted in a more complicated graph. A more robust justification goes as follows. Consider a new scalar field ψ related to ϕ by

$$\phi = \psi + \lambda_2\,\Lambda^{-2}\,\psi^3 . \tag{8.28}$$

Note that both terms in the right hand side have mass dimension 1 and transform as Lorentz scalars. Rewriting the terms of the above Lagrangian in terms of ψ, we obtain

$$\begin{aligned}
\frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) &= \frac{1}{2}(\partial_\mu\psi)(\partial^\mu\psi) - \lambda_2\,\Lambda^{-2}\,\psi^3\Box\psi + \mathcal{O}(\Lambda^{-4}) ,\\
m^2\,\phi^2 &= m^2\,\psi^2 + 2\lambda_2\,\Lambda^{-2}\,\psi^4 + \mathcal{O}(\Lambda^{-4}) ,\\
\frac{\lambda}{4!}\,\phi^4 &= \frac{\lambda}{4!}\,\psi^4 + 4\lambda_2\,\Lambda^{-2}\,\psi^6 + \mathcal{O}(\Lambda^{-4}) ,
\end{aligned} \tag{8.29}$$

and finally

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\psi)(\partial^\mu\psi) - \frac{1}{2}m^2\psi^2 - \frac{\lambda'}{4!}\psi^4 + \frac{1}{\Lambda^2}\,\lambda_1'\psi^6 + \mathcal{O}(\Lambda^{-4}) , \tag{8.30}$$

where λ', λ_1' are new coupling constants for the quartic and sextic terms. In the spirit of an effective field theory, we do not care about the terms of order Λ^{-4} since they come with operators of dimension 8, that we are not considering here. Thus, by the change of variable of eq. (8.28), we can eliminate the term that seemed redundant in the Lagrangian. More generally, any term of the form

$$\Lambda^{-2}\,f(\phi)\,\big(\underbrace{\Box\phi + m^2\phi + \tfrac{\lambda}{6}\phi^3}_{\text{l.h.s. of the EOM}}\big) , \tag{8.31}$$

where $f(\phi)$ is any local function of the fields of mass dimension 3 (e.g., ϕ^3, $m^2\phi$, $\Box\phi$), can be removed from the effective Lagrangian by the following field redefinition

$$\phi = \psi + \Lambda^{-2}\,f(\phi) . \tag{8.32}$$

Functional Jacobian : Having removed the operator $\phi^3\Box\phi$ from the Lagrangian is not enough, because the change of variable (8.28) has also implications elsewhere. Firstly, in the path integral representation of the generating functional, this change of variable introduces a Jacobian since the functional integration measure is modified as follows

$$\big[D\phi(x)\big] = \big[D\psi(x)\big]\ \det\left(\frac{\delta\phi(x)}{\delta\psi(y)}\right)\ . \tag{8.33}$$

For a transformation of the type (8.28), the determinant depends on the field since we have

$$\frac{\delta\phi(x)}{\delta\psi(y)} = \Lambda^{-2}\,\delta(x-y)\,\left[\Lambda^2 + 3\,\lambda_2\,\psi^2(x)\right]\ . \tag{8.34}$$

and therefore this determinant should not be disregarded. Like in the Fadeev-Popov quantization procedure, we may express it as an path integral over fictitious Grassmann fields $\chi, \overline{\chi}$, by writing:

$$\det\left(\frac{\delta\phi(x)}{\delta\psi(y)}\right) = \int\big[D\chi(x)D\overline{\chi}(x)\big]\ \exp\left\{i\int d^4x\,\overline{\chi}(x)\Big(\Lambda^2 + 3\,\lambda_2\,\psi^2(x)\Big)\chi(x)\right\}\ . \tag{8.35}$$

In the case of our simple example, the kinetic term of this ghost field is a bit peculiar since it does not contain any derivatives. However, it exhibits a feature which is completely generic, namely the fact that its mass is of order Λ. Since the ghosts can only appear in closed loops, their contribution is suppressed by inverse powers of Λ. In other words, the determinant depends on the field ψ, but this dependence is of higher order in Λ^{-2} and will not affect our effective theory.

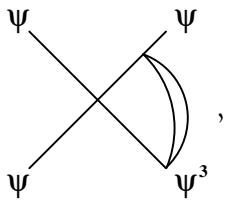
Modifications to the external points : Secondly, the change of variables (8.28) modifies the coupling to the fictitious source in the generating functional:

$$\int d^4x\ J(x)\phi(x) = \int d^4x\ J(x)\left(\psi(x) + \lambda_2\,\Lambda^{-2}\,\psi^3(x)\right)\ . \tag{8.36}$$

Thus, every functional derivative with respect to J brings a factor $\psi + \lambda_2\,\Lambda^{-2}\,\psi^3$ in the time-ordered product of interest:

$$\big\langle 0_{out}\big|T\,\phi(x_1)\cdots\big|0_{in}\big\rangle = \big\langle 0_{out}\big|T\left(\psi(x_1) + \lambda_2\,\Lambda^{-2}\,\psi^3(x_1)\right)\cdots\big|0_{in}\big\rangle\ . \tag{8.37}$$

At this point, we have shown that the only possible effect of the term we have removed from the effective Lagrangian is to modify the operators inside a time-ordered product of fields (in the form of extra terms that will appear on the external legs of the corresponding Feynman graphs). However, the physical quantities are not the above correlation functions themselves, but the on-shell transition amplitudes obtained with the LSZ reduction formulas, i.e. the residue of the 1-particle poles in the Fourier transform of eq. (8.37). For instance, in a 4-point function contributing to a $2 \to 2$ scattering amplitude, one would have a graph such as the following



,

where one of the operators in the T-product is a ψ^3 (in this diagrammatic representation, we have not yet amputated the external propagators.) We can readily see that one point of this function is not terminated by a propagator, and therefore does not exhibit a 1-particle pole. Thus, such a graph does not contribute to the on-shell transition amplitude when inserted in the LSZ reduction formula. Although we have used a very simple example to illustrate the chain of arguments leading to this result, it is in fact completely general: if a term of the effective Lagrangian can be rewritten as a linear combination of other operators thanks to the leading order equation of motion, then this term can be ignored in the effective theory without changing anything to the S matrix.

8.4 Effective theories in QCD

Quantum Chromodynamics is also an area where effective field theories are quite useful. Indeed, since QCD contains many dimensionful scales (the scale Λ_{QCD} at which the coupling constant becomes large, and the masses of the six families of quarks, that span a wide range of momentum scales), we may expect that some simplifications are possible if one is interested in processes in which some of these scales are irrelevant. Several effective theories have been developed in order to simplify the treatment of strong interactions in some special kinematical situations, and we shall discuss two of them in this section.

8.4.1 Heavy quark effective theory

Main ideas : There are six families of quarks in Nature, u, d, s, c, b, and t. The u, d and s quarks are light in comparison to other QCD scales (in particular the confinement scale Λ_{QCD}), while the c, b and t are considered heavy. Besides the well known nucleons (proton, neutron) and light mesons (pions, rho), that are made of u and d valence quarks, some hadrons contain heavy quarks (c and b only, since the t quark decays before a bound state can form). An obvious source of simplification in the presence of heavy quarks is asymptotic freedom, thanks to which the strong coupling constant at the scale m_Q is not very large and thus the strong interactions are more like electromagnetic interactions. In particular, hadrons made of a pair of heavy quark and antiquark $Q\overline{Q}$ have a size of order $(\alpha_s m_Q)^{-1}$. When this size is much smaller than $\Lambda_{\text{QCD}}^{-1}$, these bound states are quite similar to a hydrogen atom.

However, hadrons mixing heavy and light quarks are not as simple, because their size is of order $\Lambda_{\text{QCD}}^{-1}$ and the typical momentum transfer between the light and heavy quarks is of order Λ_{QCD}. Thus, in these heavy-light hadrons, on may view the heavy quark as surrounded by a non-perturbative cloud of light quarks and gluons. Such systems are characterized by two different scales:

- The heavy quark mass m_Q and the corresponding Compton wavelength $\lambda_Q = m_Q^{-1}$,
- The confinement scale Λ_{QCD}, that controls the typical size of bound states, $R_h \sim \Lambda_{\text{QCD}}^{-1}$.

For a heavy quark, one has $\lambda_Q \ll R_h$. Thus, in a certain sense, the heavy quark may be viewed as a point-like object inside a much larger hadron. Loosely speaking, the quantum numbers of the heavy quark (flavour, spin) are confined in a volume of order of its Compton wavelength λ_Q,

but the accompanying cloud of light quarks and gluons can only resolve distances as small as $\Lambda_{\rm QCD}^{-1}$. Therefore, the light degrees of freedom are totally insensitive to the heavy quark quantum numbers, and they only feel its color field. Moreover, for a heavy-light hadron, the rest frame of the hadron is almost equivalent to the rest frame of the heavy quark. In this frame, the color field of the heavy quark is the Coulomb electrical field produced by a static color charge, that does not depend on the heavy quark mass. Thus, we expect that the configuration of the light constituents is independent of m_Q when $m_Q \to \infty$. These observations constitute what is called *heavy quark symmetry*, that we shall derive more formally later in this section. Note that, unlike chiral symmetry for massless quarks, it is not a symmetry of the QCD Lagrangian, but rather an approximate symmetry that arises in special kinematical conditions (namely, when a heavy quark interacts only with light degrees of freedom via soft exchanges). Heavy quark symmetry provides relationships between bound states that differ only in the flavour and/or the spin[8] of the heavy quark, for instance the B, D, B^*, D^* mesons, or the Λ_b, Λ_c baryons. Heavy quark effective theory exploits this separation of scales in a systematic way in order to calculate the dependence of various physical quantities on the mass of the heavy quark, by an expansion in powers of m_Q^{-1}.

Spinor decomposition : Let us assume that there is a large gap between the confinement scale $\Lambda_{\rm QCD}$ and the heavy quark mass m_Q, and introduce an intermediate scale Λ such that $\Lambda_{\rm QCD} \ll \Lambda \ll m_Q$. Our goal is to construct an effective theory which is equivalent to QCD at long distance, i.e. for momenta below Λ (but may differ from QCD above Λ). Heavy quark effective theory is somewhat special in that we do not completely integrate out the heavy quarks (since one of its applications is to describe bound states that contain heavy quarks), but we rather integrate out only a part of the heavy quark degrees of freedom. This is done by writing the momentum of a heavy quark as follows:

$$p^\mu \equiv m_Q\, v^\mu + q^\mu \,, \tag{8.38}$$

where v^μ is the *hadron* velocity (satisfying $v_\mu v^\mu = 1$) and q^μ is a residual momentum whose components are much smaller than m_Q. This decomposition just highlights the fact that the heavy quark moves almost with the hadron velocity. By definition, the term $m_Q\, v^\mu$ does not change, while q^μ fluctuates due to the interactions of the heavy quark with the light degrees of freedom. However, the typical changes of q^μ are of order $\Lambda_{\rm QCD}$. Thus, the physical picture that emerges from this separation is that of a heavy quark that moves almost along a straight line, just undergoing little kicks from the surrounding cloud of light constituents.

By combining eq. (8.38) and the Dirac equation, we can see that the dominant spacetime dependence of spinors is a phase $\exp(\pm i\, m_Q\, v\cdot x)$. Moreover, the velocity v_μ can be used to construct two spin projectors,

$$P_\pm \equiv \frac{1 \pm \not{v}}{2} \,, \tag{8.39}$$

thanks to which we may decompose the spinor ψ of a heavy quark into

$$\begin{aligned} q_v(x) &\equiv e^{i\, m_Q\, v\cdot x}\, P_+ \psi(x) \,, \\ Q_v(x) &\equiv e^{i\, m_Q\, v\cdot x}\, P_- \psi(x) \,, \end{aligned} \tag{8.40}$$

[8] This is analogous to the fact that isotopes have almost identical chemistry, since the cloud of electrons surrounding the nucleus is almost independent of its mass (in a first approximation, it depends only on its electrical charge). Likewise, the independence with respect to the spin of the heavy quark is analogous to the near degeneracy of the hyperfine levels in atomic physics.

or conversely

$$\psi(x) = e^{-i\,m_Q\,v\cdot x}\,\Big[q_v(x) + Q_v(x)\Big]\,. \tag{8.41}$$

By introducing this decomposition in the Dirac Lagrangian, we obtain

$$\begin{aligned}\mathcal{L} &= \overline{\psi}\,\big(i\not{D} - m_Q\big)\,\psi \\ &= (\overline{q}_v + \overline{Q}_v)\big(i\not{D} - m_Q + m_Q\not{v}\big)(q_v + Q_v) \\ &= \overline{q}_v\big(i\,v\cdot D\big)q_v - \overline{Q}_v\big(i\,v\cdot D + 2m_Q\big)Q_v + \overline{q}_v\big(i\not{D}_\perp\big)Q_v + \overline{Q}_v\big(i\not{D}_\perp\big)q_v\,,\end{aligned} \tag{8.42}$$

where we have decomposed the covariant derivative as $D^\mu \equiv v^\mu\,(v\cdot D) + D^\mu_\perp$. From this form of the Lagrangian, we see that q_v is a massless spinor while Q_v has a mass $2m_Q$. Thus, the heavy quark effective theory will be obtained by integrating out Q_v.

Effective Lagrangian : Let us consider the generating function of the correlation functions of the "light" field q_v, defined as

$$Z[\eta,\overline{\eta}] \equiv \int \big[Dq_v D\overline{q}_v DQ_v D\overline{Q}_v\big]\; e^{i\int d^4x\;(\mathcal{L}+\overline{\eta}q_v+\overline{q}_v\eta)}\,. \tag{8.43}$$

The path integration over the heavy field Q_v is Gaussian and can be performed analytically, giving

$$Z[\eta,\overline{\eta}] = \int \big[Dq_v D\overline{q}_v\big]\;\Delta_v[A]\;e^{i\int d^4x\;(\mathcal{L}_{\rm eff}+\overline{\eta}q_v+\overline{q}_v\eta)}\,, \tag{8.44}$$

with the following effective Lagrangian

$$\mathcal{L}_{\rm eff} \equiv \overline{q}_v\big(i\,v\cdot D\big)q_v + \overline{q}_v\big(i\not{D}_\perp\big)\,\frac{1}{2m_Q + i v\cdot D}\,\big(i\not{D}_\perp\big)q_v\,, \tag{8.45}$$

and where $\Delta_v[A]$ is the functional determinant produced by the Gaussian integral:

$$\Delta_v[A] \equiv \Big(\det\big(2m_Q + i v\cdot D\big)\Big)^{1/2}\,. \tag{8.46}$$

Note that if one chooses the strict axial gauge $v\cdot A = 0$, then this determinant is constant and may be disregarded.

Derivative expansion : Because of the presence of derivatives in the denominator of eq. (8.45), the corresponding effective action is non-local. In order to obtain a local effective theory, we should expand this expression into a series of local operators. Such an expansion is legitimate because we have pulled out the fast phase $\exp(i\,m_Q v\cdot x)$ from the spinor. The resulting light field $q_v(x)$ has only a slow spacetime dependence associated with the residual momentum $q^\mu \sim \Lambda_{\rm QCD}$. Moreover, interactions with soft gluons involve a gauge field $A^\mu \sim \Lambda_{\rm QCD}$, and we thus have

$$v\cdot D \sim \Lambda_{\rm QCD} \ll m_Q\,, \tag{8.47}$$

which allows the following expansion

$$\frac{1}{2m_Q + i\,v\cdot D} = \frac{1}{2m_Q}\sum_{n=0}^{\infty}\left(-i\,\frac{v\cdot D}{2m_Q}\right)^n \ . \tag{8.48}$$

Up to the terms of order m_Q^{-1}, the effective Lagrangian therefore reads

$$\mathcal{L}_{\rm eff} = \underbrace{\bar{q}_v\big(i\,v\cdot D\big)q_v}_{\mathcal{L}_\infty} + \bar{q}_v\frac{\big(i\not{D}_\perp\big)^2}{2m_Q}q_v + \frac{g}{2m_Q}\,\bar{q}_v M^{\mu\nu}F_{\mu\nu}q_v + \mathcal{O}(m_Q^{-2})\ , \tag{8.49}$$

where $F_{\mu\nu}$ is the QCD field strength tensor, and $M^{\mu\nu} \equiv \frac{i}{4}[\gamma^\mu,\gamma^\nu]$ are the generators of the Poincaré algebra in the spin $1/2$ representation. The first term $\mathcal{L}_\infty$ is the only one that survives in the limit of infinite quark mass. The terms of order m_Q^{-1} can be interpreted respectively as the contribution of the transverse motion to the kinetic energy and the interaction between the spin of the quark and the chromo-magnetic field.

Heavy quark symmetry : The leading term in the effective Lagrangian, $\mathcal{L}_\infty$, corresponds to the following Feynman rules

Since there are no Dirac matrices in the expression of the vertex, the interactions with gluons do not alter the spin of the heavy quark at order m_Q^0. More formally, since $\mathcal{L}_\infty$ does not contain Dirac matrices, it is invariant under

$$q_v \to e^{i\,\theta^i S^i}\,q_v \quad , \qquad \text{with } S^i \equiv \frac{1}{2}\begin{pmatrix}\sigma^i & 0\\ 0 & \sigma^i\end{pmatrix}\ . \tag{8.50}$$

(The σ^i are the Pauli matrices.) Since we have $[S^i,S^j] = i\,\epsilon^{ijk}\,S^k$, this corresponds to an $SU(2)$ invariance of $\mathcal{L}_\infty$. Moreover, since $\mathcal{L}_\infty$ is independent of m_Q, all heavy quarks play the same role. With N_f flavors of heavy quarks, the leading effective Lagrangian

$$\mathcal{L}_\infty = \sum_{f=0}^{N_f}\bar{q}_{vf}\big(i\,v\cdot D\big)q_{vf} \tag{8.51}$$

has an $SU(2N_f)$ symmetry, that constitutes the spin-flavour heavy quark symmetry. These symmetries are broken by the corrections in m_Q^{-1}, since they depend explicitly on the mass and contain Dirac matrices.

8.4.2 Color Glass Condensate

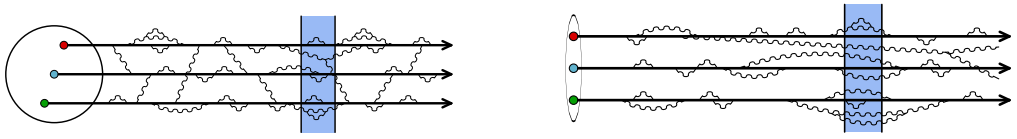
Kinematics of high energy collisions : Another area of strong interactions which is hardly tractable in QCD itself, but where progress can be made with the help of an effective description

is that of collisions between hadrons (or nuclei) at very high energy. Consider for instance a proton. A naive picture is that it is made of three valence quarks, bound by gluon exchanges. However, in a relativistic quantum description, these constituents can all fluctuate into virtual quarks, antiquarks and gluons. The important point is that these fluctuations are short-lived: roughly speaking, their lifetime spans all scales from zero to the proton size.

When a proton is probed in an experiment (for instance, by colliding it with another proton) characterized by a certain time resolution, the fluctuations of the proton whose lifetime is smaller than this resolution do not play an active role. Through renormalization, their only effect is to set the values of the parameters of the Lagrangian (in particular, the coupling constant) at the scale relevant for this experiment. On the other hand, the fluctuations whose lifetime is large compared to the characteristic timescale of the probe are seen as actual constituents of the proton. For instance, if a quark has fluctuated into a long-lived (compared to the timescale probed in the experiment) quark-gluon state, then the experiment will see a quark plus a gluon, both on-shell. Note that on-shellness, i.e. the fact that a momentum satisfies $P^2 = m^2$ for a particle of mass m, should not be viewed in a strict sense in this context. On-shellness in principle requires that the energy be *exactly* $p_0 = \sqrt{\mathbf{p}^2 + m^2}$. But by the uncertainty principle, one would have to observe this particle for an infinitely long time to check an exact equality. Thus, in a measurement of duration Δt, any particle whose energy p_0 satisfies $\Delta t \left| p_0 - \sqrt{\mathbf{p}^2 + m^2} \right| \lesssim 1$ cannot be distinguished from an exactly on-shell particle.

This discussion provides the physical justification of the parton model, in which bound states such as protons are described by means of distributions of quarks, antiquarks and gluons (generically called "partons"). Except for the valence quarks, these constituents are in fact quantum fluctuations, but their long lifetime (compared to the interaction time) allows to treat them as on-shell. Moreover, these partons distributions must vary with the resolution scale (in space and time) with which the proton is probed, since a smaller resolution scale will resolve more partons in the measurement.

Figure 8.3: *Cartoon of the fluctuations inside a nucleon. The blue band indicates the resolution in time of an external probe. Left: slow nucleon. Right: boosted nucleon.*



In particular, in a collision between two hadrons, the duration of the collision scales as the inverse of the collision energy (roughly speaking, this is the time necessary for the two Lorentz contracted hadrons to go through each other), and therefore the parton distributions grow with the collision energy. Let us now assume that in such a collision we are interested in processes characterized by a transverse momentum Q. This means that this measurement resolves a fixed transverse distance of the order of Q^{-1}, which may also be viewed as the minimal spatial extension of the wavefunction of the partons probed in this collision. Combining this fact with the increase of the number of partons with energy, we see that higher collision energy leads eventually to a situation where the partons fully pack the available volume in the hadron. This regime of strong interactions is known as *(gluon) saturation*. Note that the gluon occupation

number cannot grow above α_s^{-1}, because this is the value at which gluon splittings and gluon recombinations roughly balance each other.

Degrees of freedom : In order to discuss the relevant degrees of freedom, let us consider the point of view of an observer at rest, while the hadron moves with a very large momentum in the z direction. Due to the special kinematics of this problem, it is convenient to introduce *light-cone coordinates*, defined as

$$x^+ \equiv \frac{x^0 + x^3}{\sqrt{2}} , \quad x^- \equiv \frac{x^0 - x^3}{\sqrt{2}} . \tag{8.52}$$

(The remaining two coordinates are the transverse coordinates $\boldsymbol{x}_\perp$.) Similar definitions can be introduced for 4-momenta. These coordinates have the virtue of transforming very simply under boosts in the z direction, since $x^\pm$ just undergo a rescaling:

$$x^+ \to e^\omega \, x^+ , \quad x^- \to e^{-\omega} \, x^- , \quad \boldsymbol{x}_\perp \to \boldsymbol{x}_\perp . \tag{8.53}$$

In order to order the constituents by their longitudinal momentum, the most convenient variable is *rapidity*, defined as $y \equiv \frac{1}{2} \ln(p^+/p^-)$, since it is shifted by an additive constant under a boost in the z direction. By definition, $y = 0$ (i.e. $p_z = 0$) corresponds to objects with no longitudinal momentum in the observer's frame. Quantum fluctuations with a large positive rapidity appear to the observer as nearly on-shell constituents. At the largest rapidities (corresponding to the total p_z of the hadron), there are few constituents, mostly the valence quarks. Because of their large longitudinal momentum, the dynamics of these constituents is considerably slowed down by time dilation, and therefore they appear static to the observer. The only relevant information about these fast partons is the color current they carry. This current is longitudinal, and because these constituents are static, it does not depend on the light-cone variable[9] x^+ and takes the following form :

$$J_a^\mu(x) \equiv \delta^{\mu +} \, \rho_a(x^-, \boldsymbol{x}_\perp) , \tag{8.54}$$

where the function ρ_a is the spatial distribution of color charge. For a high energy hadron, Lorentz contraction implies that the x^- dependence of this function is very peaked around $x^- \approx 0$. On the other hand, the $\boldsymbol{x}_\perp$ dependence reflects the distribution of the constituents of the hadron in the plane transverse to the collision axis. Since this depends on the peculiar spatial arrangement of the constituents at the time of the collision, the function $\rho_a(x^-, \boldsymbol{x}_\perp)$ is not known and may be considered as a random variable with a probability distribution $W[\rho]$. When one repeats may similar collisions, the expectation value of an observable is obtained by a functional average

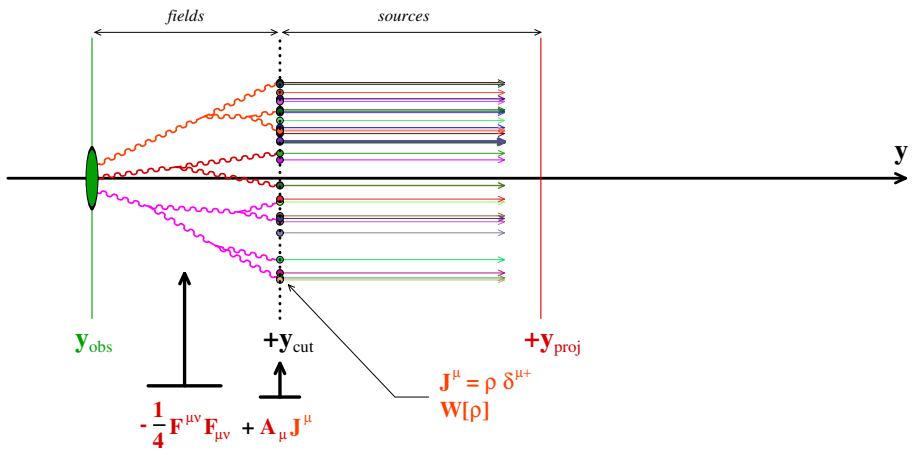
$$\big\langle \mathcal{O} \big\rangle = \int \big[D\rho \big] \; W[\rho] \; \mathcal{O}[\rho] , \tag{8.55}$$

where $\mathcal{O}[\rho]$ is the value of this observable calculated with an arbitrary instance of the distribution ρ_a.

In contrast, the constituents that lie at small rapidity in the observer's frame have a time evolution that cannot be neglected. These modes are thus described according to the original

[9] The evolution in x^+ is generated by the component P^- of the momentum. However, for massless on-shell modes, we have $P^- = \mathbf{P}_\perp^2/(2P^+) \to 0$ for the fast moving modes in the z direction.

Figure 8.4: *Degrees of freedom in the Color Glass Condensate effective description of a high energy hadron.*



Yang-Mills action, as illustrated in the figure 8.4. Moreover, due to the hierarchy between the longitudinal momenta of the modes described as a color current and those described as regular gauge fields, the coupling between them may be approximated as eikonal, i.e. by a term of the form $J_\mu A^\mu$, and therefore the action of the effective theory reads

$$S = \int d^4x \left(-\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu\, a} + J_{\mu\, a} A^\mu_a \right) . \tag{8.56}$$

This effective theory is called the *Color Glass Condensate*.

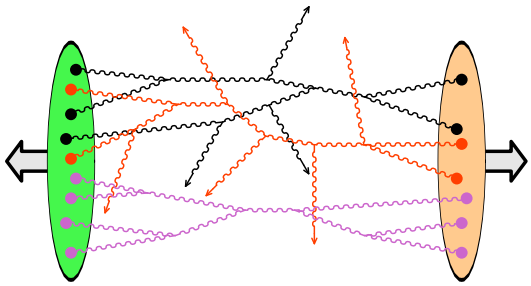
Power counting in the saturation regime : The power counting for the graphs that appear in this effective theory is a bit peculiar in the saturation regime. Indeed, this situation corresponds to a gluon occupation number of order g^{-2}, which is achieved with a color current of order g^{-1}. The order of a connected graph $\mathcal{G}$ with n_E external gluons, n_L loops and n_J insertions of the color current is given by

$$\mathcal{G} \sim g^{-2}\, g^{n_E}\, g^{2\, n_L}\, (gJ)^{n_J} , \tag{8.57}$$

where J denotes the typical magnitude of the current. Thus, in the saturated regime where $J \sim g^{-1}$, the magnitude of connected graphs does not depend on n_J, which means that all observables depend non-perturbatively on the color current. In contrast, the loop expansion still corresponds to an expansion in powers of g^2. Observables at tree-level are given by an infinite sum of tree diagrams (corresponding to an arbitrary number of insertions of J^μ), that can be expressed in terms of classical solutions of the Yang-Mills equations of motion in the presence of an external source:

$$\left[D_\mu, F^{\mu\nu} \right]_a = -J^\nu_a . \tag{8.58}$$

Figure 8.5: *Typical graph in a hadron-hadron collision, in which both hadrons are described with the CGC effective theory. The solid dots represent insertions of the color current* J_μ.



In order to have a unique solution, these equations must be supplemented with boundary conditions. One may show that in the case of inclusive observables[10], the appropriate boundary condition is a retarded one, in which the initial fields (and their time derivative) are zero in the remote past (i.e. long before the collision). The classical field obtained by solving the above equation of motion is a strong field,

$$\mathcal{A}^\mu \sim g^{-1} \ , \tag{8.59}$$

which leads to several technical complications. Some of these issues are discussed in the chapter 14. Higher order contributions correspond to loops evaluated in the presence of this classical field as a background.

Cutoff dependence : In addition, this effective description must be endowed with a cutoff (denoted y_{cut} in the figure 8.4) in rapidity that separates the two types of degrees of freedom. This cutoff does not appear explicitly in the above classical action (8.56), and therefore observables do not depend on it at tree level. But it enters in loop corrections as an upper limit in the integral over the longitudinal momentum circulating in the loop. Indeed, including in the loop modes that have a rapidity larger than y_{cut} would lead to a double counting, since these modes are already included via the color current J_μ. Generically, this cutoff introduces a linear dependence on y_{cut} in the 1-loop correction of observables. In fact, one may show that for all inclusive observables, the cutoff dependence at 1-loop can be written as

$$\delta\mathcal{O}_{\text{NLO}}[\rho] = y_{cut}\ \mathbf{H}\ \mathcal{O}_{\text{LO}}[\rho] + \text{terms that do not depend on } y_{cut} \ , \tag{8.60}$$

where $\mathbf{H}$ is a universal (i.e. the same for all inclusive observables) operator containing second order derivatives with respect to ρ_a. An important property of this operator is that it is self-adjoint:

$$\int [D\rho]\ A[\rho]\ \left(\mathbf{H}\, B[\rho]\right) = \int [D\rho]\ \left(\mathbf{H}\, A[\rho]\right)\ B[\rho] \ . \tag{8.61}$$

[10] Inclusive observables are measurements for which one sums over all the possible final states without excluding any of them. For instance, the average particle multiplicity in the final state is an inclusive observable, while the probability of producing exactly 3 particles is not.

However, the cutoff is not a physical parameter, since it was just introduced by hand in order to separate the two types of degrees of freedom, and therefore it should not appear in physical quantities. The way out of this situation is to realize that by changing the value of the cutoff, one is also modifying which modes are described by the color current J_μ. Consequently, the distribution $W[\rho]$ should in fact depend on $y_{\rm cut}$. Using eqs. (8.60) and (8.61), we see immediately that the cutoff dependence coming from the loop correction to observables can be canceled if we also change

$$W[\rho] \to W[\rho] - y_{\rm cut}\, \mathbf{H}\, W[\rho]\ . \tag{8.62}$$

More precisely, this substitution cancels the linear dependence on $y_{\rm cut}$. A more rigorous procedure is to apply it to an infinitesimal variation $\delta y_{\rm cut}$ of the cutoff, for which the quadratic terms are truly negligible. By doing so, the change of eq. (8.62) becomes a differential equation

$$\frac{\partial W[\rho]}{\partial y_{\rm cut}} = -\mathbf{H}\, W[\rho]\ , \tag{8.63}$$

that controls how the probability distribution $W[\rho]$ changes as one varies the cutoff (this equation is called the *JIMWLK equation*).

8.5 Effective theories of spontaneous symmetry breaking

8.5.1 Nambu-Goldstone bosons

Spontaneous breaking of a global continuous symmetry leads to the emergence of massless spin 0 bosons, one for each broken generator of the original symmetry, the Nambu-Goldstone bosons. The other fields of the theory remain massive. Therefore, at low energy, we expect that the physics is dominated by the fluctuations of the Nambu-Goldstone bosons, and that we may neglect all the other excitations. Non-linear sigma models[11] provide an effective description that contains only the massless particles.

Let our starting point be an action of the form

$$\mathcal{S} \equiv \int \mathrm{d}^{\mathrm{d}}x \left\{ \tfrac{1}{2}(\partial_\mu \phi(x))(\partial^\mu \phi(x)) - V(\phi(x)) \right\}\ , \tag{8.64}$$

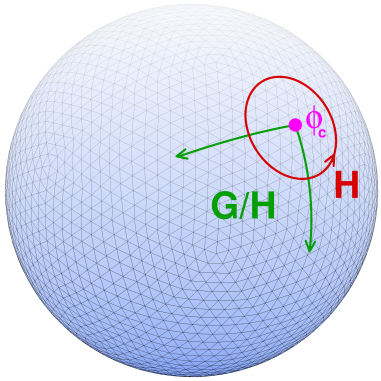
assumed to be invariant under the global action of a Lie group $\mathcal{G}$. The metric of d-dimensional spacetime is chosen to be Minkowskian (but this discussion is equally applicable to Euclidean space). In addition, the potential $V(\phi)$ has non trivial minima at some $\phi \neq 0$. Due to the $\mathcal{G}$-invariance of the action, the non trivial minima cannot be unique. Given a certain minimum $\phi_{\rm c}$, all the field configurations that may be reached from $\phi_{\rm c}$ by the action of $\mathcal{G}$ are also minima. If we assume that there are no accidental (i.e. not caused by the symmetry of the action) degeneracies of the minima, the set of all minima can therefore be written as

$$\mathcal{M}_0 \equiv \left\{ g\phi_{\rm c} \middle| g \in \mathcal{G} \right\}\ . \tag{8.65}$$

If $\mathcal{H}$ is the subgroup of $\mathcal{G}$ that leaves $\phi_{\rm c}$ invariant (sometimes called the *stabilizer* of $\phi_{\rm c}$), then $\mathcal{M}_0$ is also the coset $\mathcal{G}/\mathcal{H}$.

[11] Their name comes from early applications to the physics of pions, that may be viewed as Goldstone bosons of the (approximate for small but non-zero quark masses) chiral symmetry $\mathrm{SU}(2) \times \mathrm{SU}(2)$ that exists in the light quark (u and d) sector of quantum chromodynamics. This symmetry is spontaneously broken to a residual $\mathrm{SU}(2)$ symmetry in the vacuum of QCD, leading to the appearance of three nearly massless scalar particles.

Figure 8.6: *Illustration of the symmetry breaking pattern in the case of a model with* $G = O(3)$ *symmetry. The set* $\mathcal{M}_0$ *of the minima of the potential is a 2-dimensional sphere. The stabilizer of the minimum* ϕ_c *is* $H = O(2)$*. The green arrows show the field configurations obtained by applying* $\mathcal{G}/\mathcal{H}$ *to* ϕ_c*, i.e. the allowed values of the Nambu-Goldstone fields.*



8.5.2 Non-linear sigma model

Definition : At low energy, the quantum fluctuations of the fields that have remained massive can be neglected, and the massless components of ϕ may be obtained by acting on ϕ_c by a matrix representation of $\mathcal{G}$,

$$\phi_i = R_{ij}(g)\,\phi_{cj}\ . \tag{8.66}$$

Note that, given an element $h \in H$, we have

$$R_{ij}(gh)\,\phi_{cj} = R_{ij}(g)\,\underbrace{R_{jk}(h)\,\phi_{ck}}_{\phi_{cj}} = R_{ij}(g)\,\phi_{cj}\ . \tag{8.67}$$

Thus, the field ϕ given by eq. (8.66) is not really a function of the full group $\mathcal{G}$, but depends only on elements of the coset $\mathcal{G}/\mathcal{H}$. Let us now split the generators t^a of the Lie algebra $\mathfrak{g}$ into those (for $n < a$) that correspond to $\mathfrak{h}$, and the complement (for $1 \leq a \leq n$). From the definition of $\mathcal{H}$ as the stabilizer of ϕ_c, we have

$$\begin{aligned} n < a \quad &: \quad t^a_{ij}\,\phi_{cj} = 0\ , \\ 1 \leq a \leq n \quad &: \quad t^a_{ij}\,\phi_{cj} \neq 0\ . \end{aligned} \tag{8.68}$$

Thus, the matrix $R(g)$ can be written as

$$R(\theta) = \exp\left(i\sum_{a=1}^{n}\theta^a t^a\right)\ . \tag{8.69}$$

The value of the potential does not change under the action of $\mathcal{G}$ on ϕ_c, and we are free to choose the value of its minimum to be $V(\phi_c) = 0$. Thus, the action becomes

$$S = \frac{1}{2}\int d^dx\ \phi_{ci}\big(\partial_\mu R^{-1}_{ik}(\theta)\big)\,\big(\partial^\mu R_{kj}(\theta)\big)\phi_{cj} = -\frac{1}{2}\int d^dx\ \phi_{ci}\big(A_\mu(\theta)A^\mu(\theta)\big)_{ij}\phi_{cj}\ , \tag{8.70}$$

where in the second expression we have introduced $A_\mu \equiv R^{-1}\partial_\mu R$ (an element of the algebra).

Eq. (8.70) gives the action in terms of the "coordinates" θ^a on the coset $\mathcal{G}/\mathcal{H}$, corresponding to a certain choice of the generators t^a. However, it is interesting to express the action in terms of a completely arbitrary system of coordinates on $\mathcal{G}/\mathcal{H}$, that we may denote ϑ^m. Since eq. (8.70) has only two derivatives $\partial_\mu \cdots \partial^\mu$, the same must be true of its expression in any system of coordinates. On the other hand, it may contain terms of arbitrarily high degree in ϑ. Thus, the most general action is of the form

$$S = \frac{1}{2}\int d^d x\; g_{mn}(\vartheta)\left(\partial_\mu \vartheta^m\right)\left(\partial^\mu \vartheta^n\right)\,, \tag{8.71}$$

where the coefficients $g_{mn}(\vartheta)$ can be related to $R(\theta)$ as follows:

$$g_{mn}(\vartheta) \equiv -4\left[\phi_{ci}t^a_{ik}t^b_{kj}\phi_{cj}\right]\,\mathrm{tr}\left[t^a R^{-1}\frac{\partial R}{\partial \vartheta^m}\right]\,\mathrm{tr}\left[t^b R^{-1}\frac{\partial R}{\partial \vartheta^n}\right]\,. \tag{8.72}$$

They form a metric tensor on $\mathcal{G}/\mathcal{H}$, if the coset is viewed as a Riemannian manifold. Indeed, if we use a different system of coordinates ϖ_p on $\mathcal{G}/\mathcal{H}$, $g_{mn}(\vartheta)$ would be replaced by

$$\begin{aligned} g_{pq}(\varpi) &\equiv -4\left[\phi_{ci}t^a_{ik}t^b_{kj}\phi_{cj}\right]\,\mathrm{tr}\left[t^a R^{-1}\frac{\partial R}{\partial \varpi^p}\right]\,\mathrm{tr}\left[t^b R^{-1}\frac{\partial R}{\partial \varpi^q}\right] \\ &= \left(\frac{\partial \vartheta^m}{\partial \varpi^p}\right)\left(\frac{\partial \vartheta^n}{\partial \varpi^q}\right)\,g_{mn}(\vartheta)\,, \end{aligned} \tag{8.73}$$

which is indeed the expected transformation law of a metric tensor under a change of coordinates. The field theory described by the action (8.71) is called a *non-linear sigma model*. Note that the derivative $\partial_\mu \vartheta^m$ of the coordinate ϑ^m is a vector that lives on the tangent space to the manifold $\mathcal{G}/\mathcal{H}$ at the point ϑ. Therefore, the action (8.71), in which the tensor g_{mn} is contracted with two vectors, is a scalar – invariant under changes of coordinates on the manifold.

The Taylor expansion of the metric in powers of the field ϑ determines which couplings exist in the classical action. Interestingly, even though the kinetic term of the original action was quadratic in the fields, we now have a term with two derivatives and possibly arbitrarily high orders in the field. Loosely speaking, this is due to the fact that spontaneous symmetry breaking has restricted the fields from a space $\mathbb{R}^n$ in which the symmetry $\mathcal{G}$ was linearly realized, down to a curved manifold in which it is realized non-linearly. In addition, it is worth stressing that the final action is uniquely determined from eq. (8.69), but may take various explicit forms depending on the choice of coordinates ϑ^m on $\mathcal{G}/\mathcal{H}$. In other words, the non-linear sigma model has an intrinsic geometrical meaning, that does not depend on the system of coordinates one uses.

Path integral quantization : The quantization of the non-linear sigma model can be achieved via path integration. The action is quadratic in derivatives of the field, but with the unusual feature that these derivatives are multiplied by a function of the field. In order to ascertain the consequence of this property, it is necessary to start from the Hamilton formulation of the path integral, and to perform explicitly the integral over the conjugate momenta. For a Lagrangian density

$$\mathcal{L} = \frac{1}{2}\,g_{mn}(\vartheta)\left(\partial_\mu \vartheta^m\right)\left(\partial^\mu \vartheta^n\right)\,, \tag{8.74}$$

the conjugate momenta read

$$\pi_m \equiv \frac{\partial \mathcal{L}}{\partial \partial^0 \vartheta^m} = g_{mn}(\vartheta)\, \partial^0 \vartheta^n \ , \tag{8.75}$$

and the Hamiltonian is given by

$$\mathcal{H} = \pi_m \big(\partial^0 \vartheta^m\big) - \mathcal{L} = \frac{1}{2}\, g^{mn}(\vartheta)\, \pi_m \pi_n + \frac{1}{2}\, g_{mn}(\vartheta)\, \big(\boldsymbol{\nabla}\vartheta^m\big) \cdot \big(\boldsymbol{\nabla}\vartheta^n\big) \ , \tag{8.76}$$

where g^{mn} is the inverse of the metric tensor, $g^{mn} g_{np} = \delta^m{}_p$. The Hamiltonian is quadratic in the momenta, but since the coefficient in front of $\pi_m \pi_n$ depends on the field, the determinant produced in the Gaussian integration over the momenta cannot be disregarded. After this integral has been performed, the generating functional is given by the following formula

$$Z[j_m] = \int \Big[\sqrt{g(x)} \prod_m D\vartheta^m(x)\Big]\ \exp\left\{ \frac{i}{\hbar} \int d^d x\, \Big(\mathcal{L}(\vartheta) + j_m \vartheta^m \Big) \right\} , \tag{8.77}$$

where we denote $g(x) \equiv \det\big(g_{mn}(\vartheta(x))\big)$. Interestingly, the field dependence of $g_{mn}(\vartheta)$ alters the path integral in a rather natural way: the measure $\big[\prod_m D\vartheta^m\big]$ is replaced by $\big[\sqrt{g}\prod_m D\vartheta^m\big]$, which is invariant under changes of coordinates on the manifold $\mathcal{G}/\mathcal{H}$.

Note that in eq. (8.77), we have introduced an explicit $\hbar$, that will be useful later to keep track of the number of loops. The perturbative expansion in the non-linear sigma model corresponds to an expansion in powers of $\hbar$. From the path integral, we can infer that the typical field amplitudes scale as

$$\vartheta \sim \sqrt{\hbar} \ , \tag{8.78}$$

which means that the perturbative expansion is also an expansion around $\vartheta = 0$ (i.e. around $\phi = \phi_c$). For such small fields, the effects of the curvature of the manifold are perturbative, and we can expand the metric tensor in powers of the field (an explicit choice of coordinates must be made for this). The bare propagator of the ϑ fields is given by

$$G_{mn}(p) = \frac{i\,\delta_{mn}}{p^2 + i0^+} \ . \tag{8.79}$$

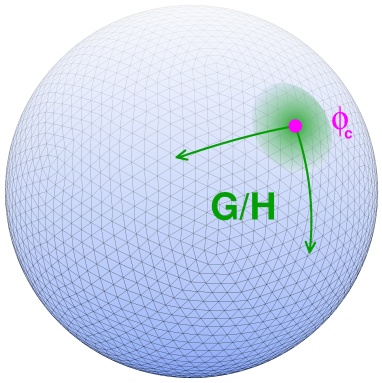
Renormalization : Dimensional analysis tells that the field ϑ has the dimension

$$\vartheta \sim (\text{mass})^{(d-2)/2} \tag{8.80}$$

(in a system of units where $\hbar = 1$). From this, we see that there are three cases regarding the ultraviolet power counting in the non-linear sigma model:

- $d < 2$: the Taylor coefficients of the metric tensor all have a positive mass dimension, and are therefore super renormalizable.
- $d = 2$: the Taylor coefficients are dimensionless, and the theory is renormalizable.

Figure 8.7: *Perturbative expansion in the non-linear sigma model: only field configurations near ϕ_c are explored.*



- $d > 2$: the Taylor coefficients have a negative mass dimension and are all non-renormalizable by power counting.

The most interesting situation is therefore the two-dimensional case. It differs somewhat from the renormalization of the quantum field theories we have encountered until now, since the action contains an infinite series of terms (of increasing degree in ϑ), and an important question is whether the action (8.71) conserves its structure under renormalization.

Recall that the fields ϑ^m transform under a non-linear representation of the group $\mathcal{G}$. Thus, their variation under an infinitesimal transformation of parameters ϵ_a may be written as

$$\delta\vartheta^m \equiv \epsilon^a \, T_a^m(\vartheta) \, , \tag{8.81}$$

where the $T_a^m(\vartheta)$ are smooth functions of the fields. Under the same transformation, the variation of the action reads

$$\delta\mathcal{S} = \epsilon^a \int d^2x \, T_a^m(\vartheta) \, \frac{\delta\mathcal{S}}{\delta\vartheta^m(x)} \, , \tag{8.82}$$

and the invariance of the action under $\mathcal{G}$ thus requires that

$$T_a^p \, \frac{\partial g_{mn}}{\partial\vartheta^p} + g_{pn} \, \frac{\partial T_a^p}{\partial\vartheta^m} + g_{pm} \, \frac{\partial T_a^p}{\partial\vartheta^n} = 0 \, . \tag{8.83}$$

In other words, the possible forms of the metric tensor are constrained by the symmetry $\mathcal{G}$. Indeed, the coset $\mathcal{G}/\mathcal{H}$ is an *homogeneous space*[12], i.e. a manifold that possesses additional symmetries that reduce the dimension of the space of allowed metrics. More precisely, an homogeneous space is such that given any pair of points ϑ and ϑ' on the manifold, there is an isometry (i.e. a distance preserving transformation) that maps ϑ to ϑ'. If in addition the

[12] Thanks to their connections to Lie algebras, a systematic classification of homogeneous spaces is possible.

space is isotropic, then it is said to be *maximally symmetric*[13]. In an N-dimensional maximally symmetric space, there is a particularly simple relationship between the metric and curvature tensors:

$$\begin{aligned} R_{mn} &= \frac{R}{N}\, g_{mn} \qquad (R \equiv R^m{}_m)\,, \\ R_{mnpq} &= \frac{R}{N(N-1)} \left(g_{mp}g_{nq} - g_{mq}g_{np}\right). \end{aligned} \tag{8.84}$$

(These two identities imply that the scalar curvature R is constant over the entire manifold for a dimension $N > 2$.)

A possible strategy for studying the renormalization of the sigma model is to introduce an analogue of the BRST transformation of non-Abelian gauge theories, and the associated Slavnov-Taylor identities obeyed by the quantum effective action. These identities, combined with dimensional and symmetry arguments that restrict the terms that may arise in the renormalized action, are sufficient to show that the renormalized action is structurally identical to eq. (8.71), with a group-invariant metric tensor that obeys a renormalized version of eqs. (8.83).

Example of $G = O(n)$: A scalar field ϕ^i with n components has an $O(n)$ symmetry if the action depends only on the combination $\phi^i\phi^i$. Potentials with non-trivial minima (i.e. at $\phi \neq 0$) in fact have infinitely degenerate minima that form a $(n-1)$-dimensional sphere $\mathcal{S}_{n-1}$ (see the figure 8.6 for an illustration in the case $n = 3$). Each minimum has a stabilizer subgroup $H = O(n-1)$ (the smaller group of rotations around the direction fixed by this minimum), and we indeed have $\mathcal{S}_{n-1} = O(n)/O(n-1)$. A possible explicit parameterization of the field ϕ consists in writing

$$\phi \equiv \left\{\sigma, \xi\right\}, \tag{8.85}$$

where σ has one component and ξ has $n-1$ components. Assuming the parameters of the potential are adjusted so that the sphere $\mathcal{S}_{n-1}$ of minima has radius $|\phi| = 1$, we must impose the constraint $\sigma^2 + \xi^2 = 1$, which means that σ may be viewed as a dependent field that depends non-linearly on ξ. Usually, these coordinates are chosen in such a way that the symmetry-breaking vacuum is $\phi_c = (\sigma = 1, \xi = 0)$. In the vicinity of ϕ_c, σ is the "radial" massive field, while the ξ^i are the "angular" variables corresponding to the massless Nambu-Goldstone bosons.

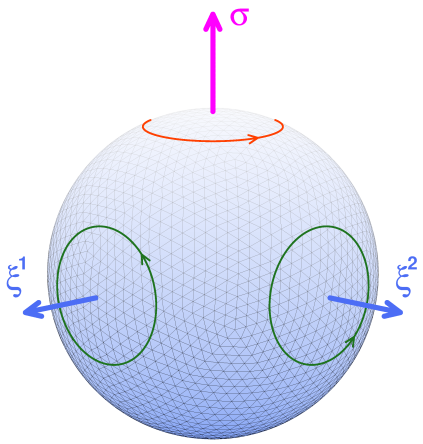
Then, we may split the generators of the $o(n)$ algebra into those of the stabilizer $o(n-1)$ and the complementary set of generators:

- The generators of $o(n-1)$ act linearly on ξ. More precisely, they leave $\sigma^2 + \xi^2$ invariant by leaving both σ and ξ^2 unchanged (thus simply rotating the $n-1$ components of ξ).
- In contrast, the generators of the complementary set preserve $\sigma^2 + \xi^2$, but mix σ and ξ as follows:

$$\begin{aligned} \sigma &\to \sigma - \epsilon^i\, \xi^i\,, \\ \xi^i &\to \xi^i + \epsilon^i \sqrt{1 - \xi^2}\,, \end{aligned} \tag{8.86}$$

[13] A maximally symmetric manifold of dimension N has $N(N+1)/2$ distinct isometries. In Euclidean space, this corresponds to N translations and $N(N-1)/2$ rotations, but this maximal number of isometries is the same in N-dimensional manifolds with curvature.

Figure 8.8: *Illustration of the (σ, ξ) coordinates for an $O(3)$ model. The red circle corresponds to the transformations that preserve σ and act linearly on ξ (as an $O(2)$ rotation). The green circles are the transformations that mix σ and ξ (and transform the latter non-linearly).*



and therefore they act non-linearly on ξ.

The most general $O(n)$-invariant action with $\sigma^2 + \xi^2 = 1$ reads

$$\begin{aligned} \mathcal{S} &= \frac{1}{2}\int d^d x \left\{ (\partial_\mu \sigma)(\partial^\mu \sigma) + (\partial_\mu \xi)(\partial^\mu \xi) \right\} \\ &= \frac{1}{2}\int d^d x \; g_{ij}(\xi)\; (\partial_\mu \xi^i)(\partial^\mu \xi^j)\;, \end{aligned} \tag{8.87}$$

where in the second line we have eliminated σ and we have defined

$$g_{ij}(\xi) \equiv \delta_{ij} + \frac{\xi^i \xi^j}{1-\xi^2}\;. \tag{8.88}$$

The tensor g_{ij} is the metric on the $\mathcal{S}_{n-1}$ sphere, in the system of coordinates provided by the ξ^i. The couplings of this theory are determined by the Taylor expansion of the metric tensor, which in this case is completely specified by the choice of the coordinates and by the symmetries of the problem. In $d = 2$ dimensions, this theory is renormalizable by power counting. Although it contains an infinite number of couplings, it is not necessary to renormalize each of them individually. Instead, the renormalization preserves the structure of the action (8.87) with a metric tensor that remains dictated by the $O(n)$ symmetry.

8.5.3 Nonlinear sigma model on a generic Riemannian manifold

We have derived the non-linear sigma model as the effective action that describes the dynamics of the massless Nambu-Goldstone bosons after a spontaneous breaking of symmetry. In this

case, the fields of the non-linear sigma model live on a manifold which is also a homogeneous space thanks to the symmetries of the original problem. These symmetries severely constrain the possible forms of the metric, and play an important role in constraining the form of the loop corrections.

However, it is possible to consider an action of the form (8.71) for fields ϑ^m living on a generic smooth Riemannian manifold that does not possess any special symmetry. The power counting argument made earlier is unchanged, and we expect that this more general kind of sigma model is also renormalizable in 2 dimensions. For these generalized models, it has been shown that the dependence of the metric tensor (i.e. the function that defines all the couplings of the model) on the renormalization scale μ is governed by the following Callan-Symanzik equation:

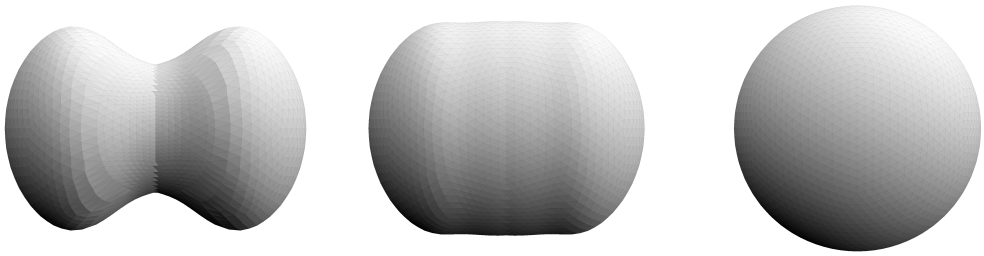
$$\mu\frac{\partial}{\partial\mu}g^{mn} = -\frac{1}{2\pi}R^{mn} - \frac{1}{8\pi^2}R^{mpqr}R^{n}{}_{pqr} + \text{higher orders}\ . \tag{8.89}$$

Note that if we apply this equation in the case of a maximally symmetric space, for which the curvature tensors have simple expressions in terms of the metric tensor, it reduces to

$$\mu\frac{\partial}{\partial\mu}g^{mn} = -\frac{R}{2\pi N}\,g^{mn}\left[1 + \frac{R}{2\pi N(N-1)} + \cdots\right]. \tag{8.90}$$

Thus, in this special case, the metric is rescaled but retains its form under changes of scale (because it is constrained by the isometries of the manifold). On a generic manifold, the scale evolution governed by eq. (8.89) explores a much broader space of metrics. Generally speaking, the renormalization flow tends to expand the regions of negative curvature and to shrink those of positive curvature.

Figure 8.9: *Left to right : successive stages of the Ricci flow on a 2-dimensional manifold.*



There is an interesting analogy between the renormalization group equation (8.89) and the *Ricci flow*,

$$\partial_\tau g^{mn} = -2\,R^{mn}\ , \tag{8.91}$$

introduced independently in mathematics by Hamilton in 1981 as a tool for studying the geometrical classification of 3-dimensional manifolds[14]. In a sketchy way, the idea is to start with a

[14] In 2 dimensions, connected manifolds are known to fall into three geometrical classes: flat, spherical or hyperbolic,

generic metric tensor on the manifold, and to smoothen this metric by evolution with the Ricci flow (the Ricci flow is somewhat analogous to a heat equation, that tends to uniformize the temperature distribution). For instance, if the metric evolves into one that has a constant positive curvature, one would have proved that the original manifold is homeomorphic to a sphere. For 2-dimensional manifolds, this is indeed what happens: the Ricci flow evolves the metric tensor into one that has a constant scalar curvature, corresponding to one of the three possible geometries. Applications of Ricci flow to 3-dimensional manifolds turned out to be complicated by singularities that develop as the metric evolves, and required additional steps known as "surgery" to excise the singularities. There is nowadays some speculation about whether the additional terms in eq. (8.89) compared to eq. (8.91) have a regularizing effect that may prevent the appearance of these singularities and thus make the surgical steps unnecessary.

depending on their curvature. More precisely, any such 2-dimensional manifold can be endowed with a metric that has a constant scalar curvature, either null, positive or negative. *Thurston geometrization conjecture* proposed a similar –but much more complicated– classification of 3-dimensional manifolds. In particular, this conjecture contains as a special case *Poincaré's conjecture*, stating that every closed simply connected 3-dimensional manifold is homeomorphic to a 3-sphere. The geometrization conjecture was proved in 2003 by Perelman, with techniques in which the Ricci flow plays a central role.

Chapter 9

Quantum anomalies

Noether's theorem states that for each continuous symmetry of a classical Lagrangian, there exists a corresponding conserved current. By construction, this conservation law holds at tree level, and a very important question is whether it is preserved by quantum corrections in higher orders of the theory. *Quantum anomalies* are situations where a classical symmetry is violated by quantum effects. We have already encountered anomalies in the section 3.5, where we saw that the fermionic functional measure is not invariant under chiral transformations of massless fermions, which had interesting connections with the index of the Dirac operator (its zero modes in the presence of an external field).

When such an anomaly arises in a global symmetry like chiral symmetry, its effect is just to introduce a corrective term into the conservation equation of the corresponding current (which may have some physical consequences, however). But when it affects a local gauge symmetry, its effects are devastating, since the renormalizability and unitarity of gauge theories relies on the validity to all orders of the gauge symmetry. In general, gauge theories with an anomalous gauge symmetry do not make sense, and it is therefore of utmost importance to check that no such gauge anomaly is present in theories of phenomenological relevance.

9.1 Axial anomalies in a gauge background

9.1.1 Two dimensional example: Schwinger model

The simplest example of theory that exhibits a quantum anomaly is quantum electrodynamics in two dimensions with massless fermions, also known as the Schwinger model. The Lagrangian of this theory reads:

$$\mathcal{L} \equiv i\,\bar{\Psi}\not{D}\,\Psi - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}\,, \tag{9.1}$$

where $D_\mu \equiv \partial_\mu - ie\,A_\mu$ and $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$. This Lagrangian is invariant under (local) $U(1)$ transformations,

$$\begin{aligned} \Psi(x) \quad &\to \quad e^{ie\chi(x)}\,\Psi(x)\,, \\ A^\mu(x) \quad &\to \quad A^\mu(x) + \partial^\mu\chi(x)\,, \end{aligned} \tag{9.2}$$

which, by Noether's theorem, implies the existence of a conserved electromagnetic current:

$$J^\mu \equiv -ie\,\overline{\Psi}\gamma^\mu\Psi \quad , \qquad \partial_\mu J^\mu = 0\,. \tag{9.3}$$

(In the following, this current will be called a *vector current*.) Being a gauge symmetry, this invariance is crucial for the unitarity of the theory, since it ensures that longitudinal photons do not contribute as initial or final states of physical amplitudes.

Because the fermions are massless, this theory has another symmetry. In order to see it, let us introduce[1] a matrix γ^5,

$$\gamma^5 = \frac{1}{2}\,\epsilon_{\mu\nu}\,\gamma^\mu\gamma^\nu = \gamma^0\gamma^1\,, \tag{9.4}$$

where $\epsilon_{\mu\nu}$ is the 2-dimensional completely antisymmetric tensor, normalized by $\epsilon_{01} = +1$. Using γ^5, one may decompose Ψ in its left and right handed components:

$$\Psi = \Psi_{_R} + \Psi_{_L}\,, \quad \Psi_{_R} \equiv \frac{1+\gamma^5}{2}\,\Psi\,, \quad \Psi_{_L} \equiv \frac{1-\gamma^5}{2}\,\Psi\,, \tag{9.5}$$

and the fermionic part of the Lagrangian can be rewritten as

$$i\,\overline{\Psi}\not{D}\,\Psi = i\,\Psi^\dagger_{_R}\,\gamma^0\not{D}\,\Psi_{_R} + i\,\Psi^\dagger_{_L}\,\gamma^0\not{D}\,\Psi_{_L}\,. \tag{9.6}$$

In other words, the kinetic term does not mix the left and right components (this would not be true with a mass term). As a consequence, the Lagrangian is invariant if we multiply the left and right components by independent phases,

$$\Psi_{_R} \quad\to\quad e^{i\alpha}\,\Psi_{_R} \quad , \qquad \Psi_{_L} \quad\to\quad e^{i\beta}\,\Psi_{_L}\,. \tag{9.7}$$

Note that this is a global invariance, unlike the gauge symmetry discussed previously. Equivalently, the massless Dirac Lagrangian is invariant under the following global transformation,

$$\Psi \quad\to\quad e^{i\theta\gamma^5}\,\Psi\,, \tag{9.8}$$

that amounts to multiplying by conjugate phases the left and right components (because of the γ^5 in the exponential). Since this is a continuous symmetry, Noether's theorem also applies here and tells us that the *axial current* is conserved:

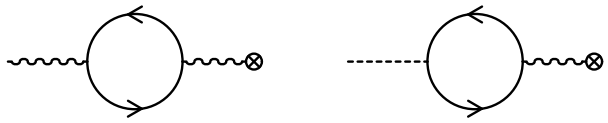
$$J^\mu_5 \equiv -ie\,\overline{\Psi}\gamma^5\gamma^\mu\Psi \quad , \qquad \partial_\mu J^\mu_5 = 0\,. \tag{9.9}$$

The conservation laws (9.3) and (9.9) have been obtained with Noether's theorem, from the fact that the *classical* Lagrangian possesses certain continuous symmetries. Let us now study how the vector and axial currents are modified at 1-loop. Here, we consider a fixed configuration of the gauge potential $A_\mu(x)$, that acts as a background external field (this also means that

[1] It is possible to define γ^5 in any even space-time dimension $D = 2r$, as follows

$$\gamma^5 \equiv \frac{i^{r-1}}{(2r)!}\,\epsilon_{\mu_1\mu_2\cdots\mu_{2r}}\,\gamma^{\mu_1}\gamma^{\mu_2}\cdots\gamma^{\mu_{2r}}\,.$$

Figure 9.1: *Left:* 1*-loop contribution to the vector current in a background gauge potential (the wavy line terminated by a cross represents the background field). Right:* 1*-loop contribution to the axial current.*



the photon kinetic term plays no role in this discussion). The lowest order 1-loop graphs that contribute to these currents are shown in the figure 9.1. The expectation values of these currents resulting from these graphs can be written as

$$\left\langle \widetilde{J}^{\mu}(q) \right\rangle = \Pi^{\mu\nu}(q)\, \widetilde{A}_{\nu}(q) \ , \quad \left\langle \widetilde{J}^{\mu}_{5}(q) \right\rangle = \Pi^{\mu\nu}_{5}(q)\, \widetilde{A}_{\nu}(q) \ , \tag{9.10}$$

(the tilde denotes the Fourier transform of the external field) where the self-energies $\Pi^{\mu\nu}$ and $\Pi^{\mu\nu}_5$ are given by

$$\begin{aligned} i\,\Pi^{\mu\nu}(q) &\equiv e^2 \int \frac{d^D k}{(2\pi)^D} \ \frac{\mathrm{tr}\left(\gamma^{\mu} \not{k} \gamma^{\nu} (\not{k} + \not{q})\right)}{(k^2 + i0^+)((k+q)^2 + i0^+)} \ , \\ i\,\Pi^{\mu\nu}_5(q) &\equiv e^2 \int \frac{d^D k}{(2\pi)^D} \ \frac{\mathrm{tr}\left(\gamma^5 \gamma^{\mu} \not{k} \gamma^{\nu} (\not{k} + \not{q})\right)}{(k^2 + i0^+)((k+q)^2 + i0^+)} \ . \end{aligned} \tag{9.11}$$

(The only difference between them is the γ^5 inside the trace, that comes from the definition of the axial current). In order to secure the subsequent manipulations, let us assume that some regularization has been performed on the momentum integrals, without specifying it for now. The denominators can be arranged into a single factor by using Feynman's parameterization,

$$\frac{1}{(k^2 + i0^+)((k+q)^2 + i0^+)} = \int_0^1 dx \ \frac{1}{(l^2 + \Delta(x))^2} \ , \tag{9.12}$$

where we have introduced $l \equiv k + x\,q$ and $\Delta(x) \equiv x(1-x)\,q^2$. After calculating the trace, the vector-vector self-energy can be written as follows:

$$\Pi^{\mu\nu}(q) = A(q^2)\, g^{\mu\nu} - B(q^2)\, \frac{q^{\mu} q^{\nu}}{q^2} \ , \tag{9.13}$$

where the coefficients $A(q^2)$ and $B(q^2)$ are given by the following integrals:

$$\begin{aligned} A(q^2) &\equiv -iDe^2 \int \frac{d^D l}{(2\pi)^D} \int_0^1 dx \ \frac{\Delta(x) + \left(\frac{2}{D} - 1\right) l^2}{(l^2 + \Delta(x))^2} \ , \\ B(q^2) &\equiv -iDe^2 \int \frac{d^D l}{(2\pi)^D} \int_0^1 dx \ \frac{2\,\Delta(x)}{(l^2 + \Delta(x))^2} \ . \end{aligned} \tag{9.14}$$

In $D = 2$ spacetime dimensions, the second integral is finite and gives:

$$B(q^2) \underset{D=2}{=} \frac{e^2}{\pi} \ , \tag{9.15}$$

while the first integral is ambiguous. Indeed, the term in l^2 in the numerator leads to an ultraviolet divergence, but it is multiplied by the factor $\frac{2}{D}-1$ that vanishes precisely when $D=2$. If we use a cutoff as ultraviolet regulator, this term would vanish and we would have $A=B/2$, which would violate the conservation of the vector current at one-loop. In dimensional regularization, in contrast, the factor $\frac{2}{D}-1$ compensates a pole in $1/(D-2)$ that comes from evaluating the integral in D dimensions, leaving a finite but non-zero result. In fact, in dimensional regularization we obtain $A=B$, and the conservation of the vector current holds at one-loop. No matter which regularization procedure we adopt, it must give $A=B$ for vector current conservation, i.e. for preserving gauge symmetry at 1-loop.

Let us now turn to the axial-vector self-energy $\Pi_5^{\mu\nu}$. Using the definition of γ^5, we obtain

$$\mathrm{tr}\left(\gamma^5\gamma^\mu\gamma^\nu\right) = -D\,\epsilon^{\mu\nu}\,, \tag{9.16}$$

and

$$\begin{aligned}\mathrm{tr}\left(\gamma^5\gamma^\mu\not{A}\gamma^\nu\not{B}\right) &= A^\nu\,\mathrm{tr}\left(\gamma^5\gamma^\mu\not{B}\right) + B^\nu\,\mathrm{tr}\left(\gamma^5\gamma^\mu\not{A}\right) - A\cdot B\,\mathrm{tr}\left(\gamma^5\gamma^\mu\gamma^\nu\right)\\ &= -D\,\epsilon^{\mu\sigma}\Big[B_\sigma A^\nu + A_\sigma B^\nu - (A\cdot B)\,g_\sigma{}^\nu\Big]\,.\end{aligned} \tag{9.17}$$

This identity leads to

$$\Pi_5^{\mu\nu}(q) = -\epsilon^{\mu\sigma}\,\Pi_\sigma{}^\nu(q) = -\epsilon^{\mu\sigma}\left[A(q^2)\,g_\sigma{}^\nu - B(q^2)\frac{q_\sigma q^\nu}{q^2}\right]\,, \tag{9.18}$$

where A and B are the same coefficients as in eq. (9.13). Therefore, the divergence of the axial current is given by

$$q_\mu\left\langle\widetilde{J}_5^\mu(q)\right\rangle = -A(q^2)\,\epsilon^{\mu\nu}\,q_\mu\,\widetilde{A}_\nu(q)\,. \tag{9.19}$$

If we have adopted a regularization that preserves gauge symmetry, i.e. such that $A=B$, this divergence is non-zero and reads

$$q_\mu\left\langle\widetilde{J}_5^\mu(q)\right\rangle = -\frac{e^2}{\pi}\,\epsilon^{\mu\nu}\,q_\mu\,\widetilde{A}_\nu(q)\,, \tag{9.20}$$

or, going back to coordinate space:

$$\partial_\mu\left\langle J_5^\mu(x)\right\rangle = -\frac{e^2}{\pi}\,\epsilon^{\mu\nu}\,\partial_\mu\,A_\nu(x) = -\frac{e^2}{2\pi}\,\epsilon^{\mu\nu}\,F_{\mu\nu}(x)\,. \tag{9.21}$$

The non-conservation of the axial current at one loop is the unavoidable conclusion in any regularization scheme that preserves the conservation of the vector current. Moreover, since when this is the case A becomes equal to the ultraviolet finite coefficient B, it does not suffer from any scheme dependence, and the above result may thus be viewed as a scheme-free result. The result (9.21) is known as an *axial anomaly*. A somewhat milder conclusion of this 2-dimensional exercise is that it not possible to preserve both vector and axial current conservation at one-loop. We could in principle adopt a regularization scheme that conserves the axial current, which requires $A=0$. But the price to pay would be the loss of gauge invariance at 1-loop. Since gauge invariance is deemed more fundamental (in particular, it ensures the unitarity of the theory), this route is generally not considered further.

Note that ultraviolet divergences are necessary[2] for the existence of this anomaly. Indeed, at the classical level, the Lagrangian density is invariant under the global transformation:

$$\Psi \to e^{i\theta\gamma^5}\Psi \quad , \qquad \Psi^\dagger \to \Psi^\dagger e^{-i\theta\gamma^5} \ . \tag{9.22}$$

The Feynman graphs that contribute to the expectation value of the axial current in a background electromagnetic field have an equal number of Ψ's and $\Psi^\dagger$'s (this statement is true to all orders of perturbation theory). Since the axial symmetry is global, when we apply the above axial transformation to a graph, all the factors $\exp(\pm i\theta\gamma^5)$ should naively cancel, leaving a result that does not depend on θ. This conclusion would indeed be correct if all the integrals were finite, but may be invalidated by the subtraction procedure necessary to obtain finite results in the presence of divergences. In the explicit example that we have studied, the ultraviolet regularizations that are consistent with gauge symmetry all spoil axial symmetry.

Beyond one loop, a graph contributing to the expectation value of the axial current may contain subgraphs that are ultraviolet divergent. However, since QED is renormalizable, these sub-divergences will all have been made finite thanks to counterterms calculated in the previous orders of the perturbative expansion. Thus, we need only to study the intrinsic ultraviolet divergence of the graph under consideration, an indicator of which is given by its superficial degree of divergence. For the sake of definiteness, let us assume that the graph $\mathcal{G}$ has n_ψ fermion propagators, n_γ photon propagators, n_V photon-fermion-fermion vertices, n_A insertions of the external electromagnetic field and n_L loops (plus one extra vertex where the axial current is attached). These quantities are not independent, but obey the following identities:

$$\begin{aligned} &2n_\gamma = n_V \ , \\ &2n_\psi = 2 + 2(n_V + n_A) \ , \\ &n_L = n_\psi + n_\gamma - n_A - n_V \ . \end{aligned} \tag{9.23}$$

Using these relations, the superficial degree of divergence of the graph reads:

$$\omega(\mathcal{G}) \equiv 2n_L - n_\psi - 2n_\gamma = 2 - n_\psi \ . \tag{9.24}$$

The simplest graph that contributes to the axial current, shown in the figure 9.1, has $n_\psi = 2$ and therefore has a logarithmic ultraviolet divergence. More complicated graphs, either with more insertions of the external field or with more than one loop, all have $n_\psi > 2$ and are therefore convergent after all their sub-divergences have been subtracted. This argument, although it lacks some rigor, indicates that the axial anomaly does not receive any correction beyond the one-loop result, and that eq. (9.21) is therefore an exact result. An alternate justification of this property is based on the derivation of the axial anomaly from the fermionic path integral, which gives the determinant of the Dirac operator in the background field. Indeed, as we have seen in the section 2.5, functional determinants correspond to 1-loop diagrams.

9.1.2 Axial anomaly in four dimensions

γ^5 in four dimensions : Let us now turn to a more realistic 4-dimensional example, that also has some relevance in understanding the decay of pseudo-scalar mesons like the π^0. The setup

[2] In a certain sense, the axial anomaly is also an infrared effect since it exists only for massless fermions (for massive fermions, there is no axial symmetry to begin with). Moreover, as we have already seen when discussing the Atiyah-Singer index theorem, the axial anomaly is related to the zero modes of the Dirac operator in a background field.

is exactly the same as in the previous section, except that we consider now four space-time dimensions. The main modification is the definition of the γ^5 matrix,

$$\gamma^5 \underset{D=4}{=} \frac{i}{4!}\,\epsilon_{\mu\nu\rho\sigma}\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma = i\,\gamma^0\gamma^1\gamma^2\gamma^3 \,. \tag{9.25}$$

The traces of a γ^5 with any odd number of ordinary Dirac matrices are all zero,

$$\mathrm{tr}\left(\gamma^5\gamma^{\mu_1}\cdots\gamma^{\mu_{2n+1}}\right) = 0 \,. \tag{9.26}$$

In order to evaluate the traces of γ^5 with an even number of Dirac matrices, let us firstly recall the general formula for a trace of an even number of Dirac matrices:

$$\mathrm{tr}\left(\gamma^{\mu_1}\cdots\gamma^{\mu_{2n}}\right) = D\sum_{\text{pairings }\mathcal{P}} \mathrm{sign}\,(\mathcal{P})\prod_{s\in\mathcal{P}} g^{\mu_{s_1}\mu_{s_2}} \,, \tag{9.27}$$

where a pairing $\mathcal{P}$ is a set of pairs $\mathcal{P} = \left\{(s_1 s_2), (s_1' s_2'), \cdots\right\}$ made of the integers in $[1, 2n]$. The signature of $\mathcal{P}$, denoted $\mathrm{sign}\,(\mathcal{P})$, is the signature of the permutation that reorders the sequence $s_1 s_2 s_1' s_2' \cdots$ into $1234\cdots$. Since the Minkowski metric tensor $g^{\mu\nu}$ is diagonal, each Lorentz index carried by one of the Dirac matrices must coincide with the Lorentz index of another matrix in order to obtain a non vanishing result. Hence, we have

$$\mathrm{tr}\left(\gamma^5\right) = i\,\mathrm{tr}\left(\gamma^0\gamma^1\gamma^2\gamma^3\right) = 0 \,. \tag{9.28}$$

The same is true if the γ^5 is accompanied by only two ordinary Dirac matrices,

$$\mathrm{tr}\left(\gamma^5\gamma^\mu\gamma^\nu\right) = i\,\mathrm{tr}\left(\gamma^0\gamma^1\gamma^2\gamma^3\gamma^\mu\gamma^\nu\right) = 0 \,, \tag{9.29}$$

and the simplest non-zero trace has a γ^5 and four ordinary Dirac matrices, $\mathrm{tr}\left(\gamma^5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\right)$. By the previous argument, each of the indices $\mu\nu\rho\sigma$ must match one of the indices 0123 hidden in $\gamma^5 = i\,\gamma^0\gamma^1\gamma^2\gamma^3$. Therefore, $\mu\nu\rho\sigma$ must be a permutation of 0123. Since the four Dirac matrices are all distinct, they all anticommute, and the result is completely antisymmetric in $\mu\nu\rho\sigma$, so that we have

$$\mathrm{tr}\left(\gamma^5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\right) = A\,\epsilon^{\mu\nu\rho\sigma} \,. \tag{9.30}$$

In order to calculate the prefactor, we just need to evaluate the trace for a particular assignment of the indices, for instance $\mu\nu\rho\sigma = 3210$,

$$A\underbrace{\epsilon^{3210}}_{+1} = \mathrm{tr}\left(\gamma^5\gamma^3\gamma^2\gamma^1\gamma^0\right) = i\,\mathrm{tr}\left(\gamma^0\,\gamma^1\,\gamma^2\underbrace{\gamma^3\gamma^3}_{-1}\gamma^2\,\gamma^1\,\gamma^0\right) = -4\,i \,. \tag{9.31}$$

$$\underbrace{\underbrace{\underbrace{\qquad\qquad}_{-1}}_{+1}}_{-1}$$

This gives $A = -4\,i$, i.e.

$$\mathrm{tr}\left(\gamma^5\gamma^\mu\gamma^\nu\gamma^\rho\gamma^\sigma\right) = -4\,i\,\epsilon^{\mu\nu\rho\sigma} \,. \tag{9.32}$$

Order 1 in the external field : Let us now turn to the calculation of the expectation value of the axial current in four dimensions. The simplest graph to consider is again the graph on the right of the figure 9.1. Its contribution to axial current is

$$\left\langle \widetilde{J}_5^\mu(q) \right\rangle = \Pi_5^{\mu\nu}(q)\, \widetilde{A}_\nu(q)\ , \tag{9.33}$$

with

$$\begin{aligned} i\,\Pi_5^{\mu\nu}(q) &\equiv e^2 \int \frac{d^D k}{(2\pi)^D}\, \frac{\mathrm{tr}\left(\gamma^5\gamma^\mu \not{k}\gamma^\nu(\not{k}+\not{q})\right)}{(k^2+i0^+)((k+q)^2+i0^+)} \\ &= e^2 \int \frac{d^D l}{(2\pi)^D} \int_0^1 dx\, \frac{\mathrm{tr}\left(\gamma^5\gamma^\mu(\not{l}-x\not{q})\gamma^\nu(\not{l}+(1-x)\not{q})\right)}{(l^2+\Delta(x))^2}\ , \end{aligned} \tag{9.34}$$

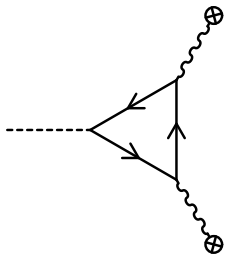
where we have introduced the Feynman parameterization in the second line, and the new integration variable $l \equiv k + xq$. The trace that appears in the numerator is proportional to

$$\epsilon^{\mu\alpha\nu\beta}(l-xq)_\alpha(l+(1-x)q)_\beta \propto \epsilon^{\mu\alpha\nu\beta} l_\alpha q_\beta\ , \tag{9.35}$$

and is therefore odd in the momentum l. Therefore, the momentum integral vanishes, and this graph does not contribute to the axial current.

Order 2 in the external field : At second order in the external field, we encounter the graph of the figure 9.2. Its contribution to the expectation value of the axial current reads

Figure 9.2: *Graph contributing to the chiral anomaly in a gauge background in four space-time dimensions.*



$$\left\langle \widetilde{J}_5^\mu(q) \right\rangle = \frac{1}{2!}\int \frac{d^4k_1\, d^4k_2}{(2\pi)^8}\, (2\pi)^4\delta(q+k_1+k_2)\, \Gamma_5^{\mu\nu\rho}(q,k_1,k_2)\, \widetilde{A}_\nu(k_1)\widetilde{A}_\rho(k_2)\ , \tag{9.36}$$

where we have introduced the following three-point function:

$$\begin{aligned} &i\,\Gamma_5^{\mu\nu\rho}(q,k_1,k_2) \equiv \\ &\quad \equiv e^3 \int \frac{d^D k}{(2\pi)^D}\, \frac{\mathrm{tr}\left(\gamma^5\gamma^\mu(\not{k}+\not{a}+\not{k}_1)\gamma^\nu(\not{k}+\not{a})\gamma^\rho(\not{k}+\not{a}-\not{k}_2)\right)}{((k+a+k_1)^2+i0^+)((k+a)^2+i0^+)((k+a-k_2)^2+i0^+)} \\ &\quad + e^3 \int \frac{d^D k}{(2\pi)^D}\, \frac{\mathrm{tr}\left(\gamma^5\gamma^\mu(\not{k}+\not{b}+\not{k}_2)\gamma^\rho(\not{k}+\not{b})\gamma^\nu(\not{k}+\not{b}-\not{k}_1)\right)}{((k+b+k_2)^2+i0^+)((k+b)^2+i0^+)((k+b-k_1)^2+i0^+)}\ . \end{aligned} \tag{9.37}$$

The two terms correspond to the two ways of attaching the fields with momenta k_1 and k_2 to the external photon lines. For a reason that will become clear later, we have taken the freedom to introduce independent shifts a and b of the integration variables in the two terms. Such shifts would of course have no effect on convergent integrals, since they just correspond to a linear change of variable. However, we are here in the presence of linearly divergent integrals, and these shifts have a nontrivial interplay with the ultraviolet regularization. Note that since $\{\gamma^5,\gamma^\alpha\}=0$, we may move the γ^5 just before the matrices γ^ν or γ^ρ without changing the integrand, as if the axial current was attached at the other edges of the triangle (where the momenta k_1 or k_2 enter, respectively).

Next, in order to test the conservation of the axial current, we contract this amplitude with q_μ, that we may rewrite as follows:

$$\begin{aligned} q_\mu &= -(k_1+k_2)_\mu \\ &= (k+a-k_2)_\mu-(k+a+k_1)_\mu \\ &= (k+b-k_1)_\mu-(k+b+k_2)_\mu\,. \end{aligned} \tag{9.38}$$

This leads to

$$\begin{aligned} &q_\mu\Gamma_5^{\mu\nu\rho}(q,k_1,k_2)=4e^3\int\frac{d^Dk}{(2\pi)^D}\,\epsilon^{\alpha\nu\beta\rho} \\ &\times\left\{\frac{(k_1)_\alpha(k+a)_\beta}{((k+a)^2+i0^+)((k+a+k_1)^2+i0^+)}+\frac{(k_2)_\alpha(k+a)_\beta}{((k+a)^2+i0^+)((k+a-k_2)^2+i0^+)}\right. \\ &\left.-\frac{(k_1)_\alpha(k+b)_\beta}{((k+b)^2+i0^+)((k+b-k_1)^2+i0^+)}-\frac{(k_2)_\alpha(k+b)_\beta}{((k+b)^2+i0^+)((k+b+k_2)^2+i0^+)}\right\}. \end{aligned} \tag{9.39}$$

By taking $a=b=0$, and assuming a regularization that preserves Lorentz invariance, each term leads to a vanishing integral. Consider for instance the first term. Since k_1 is the only 4-vector that enters in the integrand besides the integration variable k, the result of its integral is proportional to $\epsilon^{\alpha\nu\beta\rho}(k_1)_\alpha(k_1)_\beta=0$. Since the same reasoning applies to the four terms, we would therefore naively conclude that the axial current is conserved. However, we should make sure that the vector currents are also conserved. For this, we also need to calculate $(k_1)_\nu\Gamma_5^{\mu\nu\rho}$ and $(k_2)_\rho\Gamma_5^{\mu\nu\rho}$. The same method as above gives

$$\begin{aligned} &(k_1)_\nu\Gamma_5^{\mu\nu\rho}(q,k_1,k_2)=-4e^3\int\frac{d^Dk}{(2\pi)^D}\,\epsilon^{\alpha\mu\beta\rho} \\ &\times\left\{\frac{(k+a)_\alpha(k+a-k_2)_\beta}{((k+a)^2+i0^+)((k+a-k_2)^2+i0^+)}-\frac{(k+a+k_1)_\alpha(k+a-k_2)_\beta}{((k+a+k_1)^2+i0^+)((k+a-k_2)^2+i0^+)}\right. \\ &\left.+\frac{(k+b+k_2)_\alpha(k+b-k_1)_\beta}{((k+b+k_2)^2+i0^+)((k+b-k_1)^2+i0^+)}-\frac{(k+b+k_2)_\alpha(k+b)_\beta}{((k+b)^2+i0^+)((k+b+k_2)^2+i0^+)}\right\}. \end{aligned} \tag{9.40}$$

and

$$(k_2)_\rho \Gamma_5^{\mu\nu\rho}(q,k_1,k_2) = -4e^3 \int \frac{d^D k}{(2\pi)^D}\, \epsilon^{\alpha\mu\beta\nu} \times \left\{ \frac{(k+a+k_1)_\alpha (k+a-k_2)_\beta}{((k+a+k_1)^2+i0^+)((k+a-k_2)^2+i0^+)} - \frac{(k+a+k_1)_\alpha (k+a)_\beta}{((k+a+k_1)^2+i0^+)((k+a)^2+i0^+)} + \frac{(k+b)_\alpha (k+b-k_1)_\beta}{((k+b)^2+i0^+)((k+b-k_1)^2+i0^+)} - \frac{(k+b+k_2)_\alpha (k+b-k_1)_\beta}{((k+b+k_2)^2+i0^+)((k+b-k_1)^2+i0^+)} \right\} . \tag{9.41}$$

It turns out that the choice $a = b = 0$ leads to non vanishing results for the conservation of the vector currents. Consider for instance $(k_1)_\nu \Gamma_5^{\mu\nu\rho}$. With $a = b = 0$ and a regularization that preserves Lorentz invariance as well as reflection symmetry $k \to -k$, we have:

$$\begin{aligned} (k_1)_\nu \Gamma_5^{\mu\nu\rho}(q,k_1,k_2) &= -8e^3 \int \frac{d^D k}{(2\pi)^D}\, \epsilon^{\alpha\mu\beta\rho} \frac{(k+k_2)_\alpha (k-k_1)_\beta}{((k+k_2)^2+i0^+)((k-k_1)^2+i0^+)} \\ &\propto \epsilon^{\alpha\mu\beta\rho}(k_2)_\alpha (k_1)_\beta \neq 0 . \end{aligned} \tag{9.42}$$

A systematic search indicates that the only choice of a and b that gives a null result for both eqs. (9.40) and (9.41) is

$$a = -b = k_2 - k_1 . \tag{9.43}$$

Since the conservation of the vector current is necessary in order to preserve gauge symmetry, and that the latter is a requirement for unitarity, we must adopt this choice. Returning to eq. (9.39) for the axial current with these values of a and b, we obtain:

$$q_\mu \Gamma_5^{\mu\nu\rho}(q,k_1,k_2) = 16e^3 \int \frac{d^D k}{(2\pi)^D}\, \epsilon^{\alpha\nu\beta\rho} \frac{(k_1)_\alpha}{(k+k_2)^2+i0^+} \frac{(k+k_2-k_1)_\beta}{(k+k_2-k_1)^2+i0^+} . \tag{9.44}$$

Let us define

$$F^{\nu\rho}(k) \equiv \epsilon^{\alpha\nu\beta\rho} \frac{(k_1)_\alpha}{k^2+i0^+} \frac{(k-k_1)_\beta}{(k-k_1)^2+i0^+} , \tag{9.45}$$

and note that

$$\int \frac{d^D k}{(2\pi)^D}\, F^{\nu\rho}(k) = 0 . \tag{9.46}$$

(because with a Lorentz invariant regularization the result can only depend on the vector k_1, which would unavoidably give zero when contracted with the two free slots of the $\epsilon^{\alpha\nu\beta\rho}$.) Therefore, we can write

$$\begin{aligned} q_\mu \Gamma_5^{\mu\nu\rho}(q,k_1,k_2) &= 16e^3 \int \frac{d^D k}{(2\pi)^D} \left[F^{\nu\rho}(k+k_2) - F^{\nu\rho}(k) \right] \\ &= 16e^3 \int \frac{d^D k}{(2\pi)^D} \left[k_2^\sigma \frac{\partial F^{\nu\rho}(k)}{\partial k^\sigma} + \frac{k_2^\sigma k_2^\tau}{2} \frac{\partial^2 F^{\nu\rho}(k)}{\partial k^\sigma \partial k^\tau} + \cdots \right] . \end{aligned} \tag{9.47}$$

Since the integrand now contains only derivatives, we can use Stoke's formula in order to rewrite the divergence of the axial current as a surface integral on the boundary at infinity of momentum space. If we view this boundary as the limit $k_* \to \infty$ of a sphere of radius k_*, the "area" of this boundary grows like k_*^3 in $D = 4$. On the other hand, the function $F^{\nu\rho}(k)$ behaves as k^{-3}, and each subsequent derivative decreases faster by one additional power of k^{-1}. Therefore, the result is given in full by the first term of the expansion:

$$\begin{aligned} q_\mu \Gamma_5^{\mu\nu\rho}(q,k_1,k_2) &= 16e^3 \int \frac{d^D k}{(2\pi)^D}\, k_2^\sigma \frac{\partial F^{\nu\rho}(k)}{\partial k^\sigma} \\ &= \frac{16ie^3}{(2\pi)^4} \epsilon^{\alpha\nu\beta\rho} (k_1)_\alpha (k_2)^\sigma \lim_{k_*\to\infty} \underbrace{\int_{\mathcal{S}_3(k_*)} d^3\mathcal{S}\, \frac{k_\sigma}{k}\frac{k_\beta}{k^4}}_{\frac{\pi^2 g_{\sigma\beta}}{2}} \\ &= -i\,\frac{e^3}{2\pi^2}\, \epsilon^{\nu\rho\alpha\beta} (k_1)_\alpha (k_2)_\beta \,, \end{aligned} \tag{9.48}$$

In the second line, $\mathcal{S}_3(k_*)$ is the 3-sphere of radius k_* (i.e. the boundary of a 4-ball of radius k_*), k_σ/k is the unit vector normal to the sphere, and the factor i arises when going to Euclidean momentum space. Note that we have anticipated the limit $k \to \infty$ in order to simplify the function $F^{\nu\rho}(k)$. Therefore, the contribution of the triangle graph to the divergence of the axial current reads

$$q_\mu \big\langle \widetilde{J}_5^\mu(q) \big\rangle = -i\,\frac{e^3}{4\pi^2}\, \epsilon^{\nu\rho\alpha\beta} \int \frac{d^4k_1 d^4k_2}{(2\pi)^4}\, \delta(q+k_1+k_2)\, (k_1)_\alpha (k_2)_\beta\, \widetilde{A}_\nu(k_1)\widetilde{A}_\rho(k_2) \,, \tag{9.49}$$

or in coordinate space

$$\partial_\mu \big\langle J_5^\mu(x) \big\rangle = -\frac{e^3}{16\pi^2}\, \epsilon^{\alpha\nu\beta\rho}\, F_{\alpha\nu}(x)\, F_{\beta\rho}(x) \,. \tag{9.50}$$

This is the main result of this section, namely the existence of an anomalous divergence of the axial current in the presence of a background electromagnetic field. In the course of the calculation, we have seen that depending on the labeling of the integration momentum, we can make the anomaly appear in any of the three external currents. In the situation considered here, with one axial current corresponding to a global symmetry, and two vector currents stemming from a local gauge symmetry, we must enforce the conservation of the vector currents and therefore assign in full the anomaly to the axial one. But the same calculation would arise in the context of a chiral gauge theory (where the left and right handed fermions belong to different representations of the gauge group). In this case, the natural choice would be to regularize the triangle so that the symmetry among the three currents is preserved, and the anomaly would then be equally shared by the three currents.

Corrections : Let us now discuss potential corrections to the result (9.50). Firstly, we should examine one-loop graphs with more than two photons in addition to the insertion of the axial current. A simple dimensional argument can exclude that such graphs contribute to the divergence of the axial current. Indeed, $\partial_\mu J_5^\mu$ has mass dimension 4. In an abelian gauge theory, each external photon must appear in the right hand side in the form of the field strength $F^{\mu\nu}$, that has

mass dimension 2. A term with n photons would thus have mass dimension $2n$, and require a prefactor of mass dimension $4-2n$ to be a valid contribution to the divergence of the axial current. But since the fermions we are considering are massless and the coupling constant is dimensionless in four dimensions, there is no dimensionful parameter in the theory for making up such a prefactor.

Let us now consider higher loop corrections. From the calculation that led to eq. (9.50), the anomaly results from the integration over the momentum that runs in the fermion loop, provided that the integrand has mass dimension 4 or higher. Note that some of the higher order corrections just renormalize the objects that appear in the right hand side of eq. (9.50), such as the photon field strength and the coupling constant, without changing the structure of the anomaly (including the numerical prefactor). Quite generally however, adding an internal photon line requires to add more fermion propagators in the main loop, which reduces its degree of ultraviolet divergence. Of course, the integration over the momentum of this internal photon may itself be ultraviolet divergent, but it can be regularized in a way that does not interfere with axial symmetry and thus does not contribute to the anomaly.

9.2 Generalizations

9.2.1 Axial anomaly in a non-abelian background

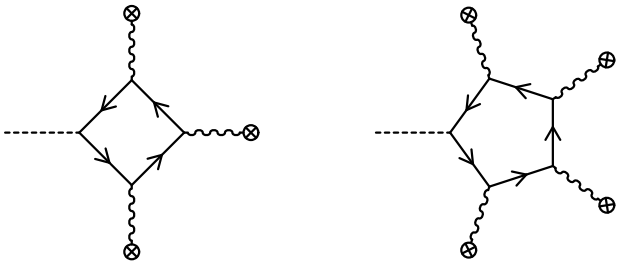
In the previous section, we have discussed axial anomalies in an abelian gauge theory. However, a similar anomaly arises in the presence of a non-abelian background gauge field. Let us assume that the fermions are in a representation of the gauge algebra where the generators are t^a. The calculation of the triangle graph proceeds almost in the same way as in the abelian case, except for the Lie algebra generators, and eq. (9.50) becomes

$$\partial_\mu \langle J_5^\mu(x)\rangle = -\frac{e^3}{4\pi^2}\,\mathrm{tr}\left(t^a t^b\right)\epsilon^{\alpha\nu\beta\rho}\left(\partial_\alpha A_\nu^a(x)\right)\left(\partial_\beta A_\rho^b(x)\right)\,. \tag{9.51}$$

This is not gauge invariant, but it is easy to guess what should be the right hand side to restore gauge invariance:

$$\partial_\mu \langle J_5^\mu(x)\rangle = -\frac{e^3}{16\pi^2}\,\mathrm{tr}\left(t^a t^b\right)\epsilon^{\alpha\nu\beta\rho}\,F_{\alpha\nu}^a(x)\,F_{\beta\rho}^b(x)\,. \tag{9.52}$$

The same dimensional argument that we have used in the abelian case also applies here: there cannot be contributions to the anomaly of degree higher than two in the field strength. Note that when expanded in terms of the gauge potential A_μ^a, eq. (9.52) contains terms of degree 3 and 4, that exist only in a non-abelian background. Diagrammatically, they correspond to contributions coming from the following two diagrams:



(But the direct extraction of the anomaly contained in these graphs would be very cumbersome, due to the numerous terms arising from permutations of the external gauge fields.)

9.2.2 Axial anomaly in a gravitational background

Another situation where an axial anomaly is present is the case of a gravitational background. Of course, this is to a large extent an academic exercise since the resulting anomaly is extremely small, due to the weakness of the gravitational coupling at the usual scales of particle physics. Nevertheless, since every field is in principle coupled to gravity, the anomalies caused by a gravitational background are unavoidable unless the matter fields of the theory are arranged in a specific way. Interestingly, the calculation of this gravitational anomaly can be performed even if we do not have a consistent quantum theory of gravity, since it does not involve quantum fluctuations of the gravitational field (the only loop is a fermion loop).

At tree level, the couplings between gravity and ordinary fields are determined from the principle of general covariance. Let us sketch here how such a calculation is done, without entering into too many technical detail. The first step is to obtain a generally covariant generalization of the Dirac operator, for an arbitrary metric tensor $g^{\mu\nu}$, from which we can read off the coupling of the fermion to the background gravitational field. In a curved spacetime, we wish to generalize the Dirac matrices so that they satisfy

$$\left\{\gamma^\mu(x),\gamma^\nu(x)\right\} = 2\,g^{\mu\nu}(x)\;. \tag{9.53}$$

(In this section, we use the Greek letters μ,ν,ρ,σ for indices related to curved coordinates, and Greek letters from the beginning of the alphabet $\alpha,\beta,\gamma,\delta$ for indices related to flat Minkowski coordinates.) In a curved spacetime, the covariant derivative of the metric tensor vanishes, and it is therefore natural to request the same for the Dirac matrices. However, this requires that we introduce a *spin connection*, which is a matrix Γ_μ defined so that

$$\nabla_\mu\gamma_\nu \equiv \partial_\mu\gamma_\nu - \Gamma^\lambda_{\mu\nu}\gamma_\lambda - \Gamma_\mu\gamma_\nu + \gamma_\nu\Gamma_\mu = 0\;, \tag{9.54}$$

where $\Gamma^\lambda_{\mu\nu}$ is the usual Christoffel's symbol. The covariant derivative acting on a spinor is $(\partial_\mu - \Gamma_\mu)\,\Psi$ and the generally covariant Dirac equation for a massless fermion reads

$$i\,\gamma^\mu\,(\partial_\mu - \Gamma_\mu)\,\Psi = 0\;. \tag{9.55}$$

In order to construct a Lagrangian that transforms as a scalar, we need a matrix Γ such that $\psi^\dagger\Gamma\psi$ is a real scalar. This is the case if the following conditions are satisfied

$$\begin{aligned}
&\Gamma = \Gamma^\dagger\;,\\
&\Gamma\,\gamma^\mu = \gamma^{\mu\dagger}\Gamma\;,\\
&\nabla_\mu\Gamma = \partial_\mu\Gamma + \Gamma^\dagger_\mu\Gamma + \Gamma\,\Gamma_\mu\;.
\end{aligned} \tag{9.56}$$

We then define $\overline{\Psi} \equiv \Psi^\dagger\Gamma$, and the Lagrangian density is

$$\mathcal{L} \equiv i\,\sqrt{-g}\,\overline{\Psi}\gamma^\mu\,\nabla_\mu\,\Psi\;. \tag{9.57}$$

(g is the determinant of the metric tensor.) The vector current and its conservation law generalize into

$$J^\mu \equiv \overline{\Psi}\gamma^\mu\,\Psi \quad , \qquad \nabla_\mu\,J^\mu = 0\;. \tag{9.58}$$

In the massless case, we can in addition define a conserved axial current:

$$J_5^\mu \equiv \overline{\Psi}\,\gamma^5\,\gamma^\mu\,\Psi \quad , \qquad \nabla_\mu\, J_5^\mu = 0\ . \tag{9.59}$$

However, as we shall see, this conservation law suffers from an anomaly in a curved spacetime. Firstly, let us introduce a representation of the Dirac matrices for a generic curved spacetime, that makes an explicit connection with the metric tensor. This is achieved by introducing four vector fields $e^\alpha{}_\mu(x)$ (called a *vierbein*, or *tetrad*) such that[3]

$$g_{\mu\nu}(x) = \eta_{\alpha\beta}\ e^\alpha{}_\mu(x)\, e^\beta{}_\nu(x)\ , \tag{9.60}$$

where in this section we use the notation $\eta_{\alpha\beta}$ for the Minkowski metric tensor. This is equivalent to introducing at each point x a local Minkowski frame with coordinates y^α. Note that $e^\alpha{}_\mu$ transforms as a vector under diffeomorphisms (a *coordinate vector*) with respect to the index μ, and as an ordinary 4-vector under Lorentz transformations (called a *tetrad vector* in this context) with respect to the index α. The indices $\alpha, \beta, \cdots$ are raised and lowered with the Minkowski metric tensor, while the indices $\mu, \nu, \cdots$ are raised and lowered with the curved space metric $g_{\mu\nu}(x)$. Since in the right hand side of eq. (9.60) the indices α and β are contracted with the Lorentz tensor $\eta_{\alpha\beta}$, the result is a scalar under Lorentz transformations, but a rank-2 tensor under diffeomorphisms. The Dirac matrices in curved spacetime ($\gamma^\mu(x)$) can then be related to those in flat spacetime (γ^α) by

$$\gamma^\mu(x) = e_\alpha{}^\mu(x)\,\gamma^\alpha\ , \tag{9.61}$$

and a spin connection Γ_μ that satisfies eq. (9.54) (and reduces to zero in flat spacetime) is given by

$$\Gamma_\mu(x) = -\frac{1}{4}\,\gamma_\alpha\,\gamma_\beta\ e^{\alpha\rho}(x)\,\nabla_\mu\, e^\beta{}_\rho(x)\ , \tag{9.62}$$

with $\nabla_\mu e^\beta{}_\rho = \partial_\mu e^\beta{}_\rho - \Gamma^\nu_{\mu\rho} e^\beta{}_\nu$ (since $e^\beta{}_\rho$ is a coordinate vector with respect to the index ρ). A matrix Γ that fulfills eqs. (9.56) is the flat spacetime γ^0, and the matrix γ^5 is still given in terms of the flat spacetime Dirac matrices by $\gamma^5 = i\,\gamma^0\,\gamma^1\gamma^2\,\gamma^3$.

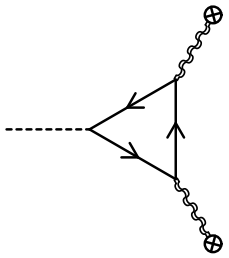
We have now a representation of the Dirac operator in an arbitrary curved spacetime, expressed in terms of the vierbein $e^\alpha{}_\mu$ that encodes the curved metric, from which we may read off the coupling between a spin $1/2$ field and the external gravitational field. We will not go into the technology required for calculating a fermion loop like the graph of the figure 9.3, and just quote the final result for the divergence of the axial current:

$$\nabla_\mu \left\langle J_5^\mu(x)\right\rangle = \frac{1}{384\pi^2}\ \epsilon^{\alpha\beta\gamma\delta}\ R_{\alpha\beta}{}^{\mu\nu}(x)\, R_{\gamma\delta\mu\nu}(x)\ , \tag{9.63}$$

where $R_{\mu\nu\rho\sigma}$ is the curvature tensor (it plays in gravity the same role as the field strength $F_{\mu\nu}$ in a non-abelian gauge theory). This formula indicates that a curved spacetime, i.e. an external gravitational field, leads to an anomalous contribution to the divergence of the axial current. This effect is of course tiny in ordinary situations where gravity is weak. But it should in principle be kept in mind when attempting to construct an anomaly free chiral gauge theory, if one wishes this theory to remain consistent all the way up to the Planck scale.

[3] In this section, we denote $\eta_{\alpha\beta} \equiv \text{diag}\,(1,-1,-1,-1)$ the flat spacetime Minkowski metric, in order to distinguish it from $g_{\mu\nu}$.

Figure 9.3: *Graph contributing to the chiral anomaly in a gravitational background in four space-time dimensions.*



9.2.3 Gauge anomalies and their cancellations

In all the examples that we have considered until now in this chapter, the anomaly appeared in the conservation of a current associated to a global symmetry such as chiral symmetry. Although it indicates a violation of this symmetry by quantum corrections, the anomaly does not make the theory inconsistent in this case. However, in graphs mixing the axial current and insertions of external gauge fields, we made sure that the ultraviolet regularization does not spoil the Ward identity associated to the gauge symmetry.

But we may also consider chiral gauge theories, in which the left and right handed components of the fermions belong to different representations of the gauge algebra. This is for instance the case in the Standard Model, where the electroweak interaction is chiral (the left handed fermions form $\mathrm{SU}(2)$ doublets, while the right handed fermions are singlet under $\mathrm{SU}(2)$). In such a theory, the gauge coupling between fermions and gauge fields involve the left or right projectors $P_{R,L} \equiv (1 \pm \gamma^5)/2$, and the corresponding current contains a γ^5, very much like the current of a global axial symmetry.

In this case, the triangle diagram that gave the axial anomaly in four dimensions is replaced by a graph with three external gauge bosons, with chiral couplings to the fermion loop. With a massless fermion in the loop, all the projectors P_L along the loop can be brought together, where they reduce to a single projector since $P_L^2 = P_L$. The γ^5 contained in this projector leads to an anomaly. The calculation is almost identical to the case of a global axial symmetry, except that now we should choose the shifts a and b so that the resulting 3-point function is symmetric in the external fields, since they play identical roles. But this choice does not eliminate the anomaly; it just distributes it evenly among the three external currents, leading to an anomaly proportional to $\mathrm{tr}\left(t^a\{t^b, t^c\}\right)$.

Unlike anomalies of global symmetries, an anomaly of a gauge symmetry makes it immediately inconsistent because it would for instance spoil its unitarity and renormalizability. For this reason, most chiral gauge theories do not make sense. The only ones that actually do are those for which the fermion fields are arranged in representations of the gauge group such that $\mathrm{tr}\left(t^a\{t^b, t^c\}\right) = 0$. This turns out to be the case for the Standard Model with its known matter fields: all the gauge anomalies cancel (within each generation of fermions) thanks in particular to the peculiar values of the electrical charges of the quarks and leptons. Interestingly, gravitational anomalies also cancel in the Standard Model.

9.3 Wess-Zumino consistency conditions

9.3.1 Consistency conditions

In the subsection 9.2.1, where we have derived the axial anomaly in a non-abelian background field, we first obtained a partial answer with only the terms quadratic in the external field, and then we used gauge symmetry in order to reconstruct the missing terms (of order 3 and 4 in the external field). However, how to promote such a partial result into the full expression of the anomaly is not always so obvious, for instance in the case of chiral gauge theories where the gauge symmetry itself is anomalous (in this case, we cannot invoke gauge invariance to restore the full answer). The *Wess-Zumino consistency conditions* are a set of equations satisfied by the anomaly function, that are powerful enough to allow reconstructing the anomaly from the knowledge of its lowest order in the gauge fields.

Even in the case where the anomalous symmetry is global, it is convenient to couple a (fictitious in that case) gauge field A^μ to the corresponding current J^μ whose conservation is violated by the anomaly. By doing this, we promote the symmetry to a local gauge invariance (violated by the anomaly), and we may return to a global symmetry by letting the gauge coupling go to zero. Let us denote $\Gamma[A]$ the effective action for the gauge field (i.e. the effective action in which the fermions are included only in the form of loop corrections). In the absence of anomaly, $\Gamma[A]$ would be invariant under gauge transformations of the field A^μ,

$$\begin{aligned} 0 \underset{\text{no anomaly}}{=} \delta_\theta \Gamma[A] &= \int d^4x \left(\left(D_\mu^{\text{adj}}\right)_{ab} \theta_b(x) \right) \frac{\delta \Gamma[A]}{\delta A_\mu^a(x)} \\ &= -\int d^4x \, \theta_b(x) \underbrace{\left(D_\mu^{\text{adj}}\right)_{ba} \frac{\delta}{\delta A_\mu^a(x)}}_{i \mathcal{T}_b(x)} \Gamma[A] \, . \end{aligned} \tag{9.64}$$

When this symmetry is spoiled by an anomaly, the effective action is no longer invariant, and we may write

$$\mathcal{T}_a(x) \, \Gamma[A] \equiv G_a[x;A] \, , \tag{9.65}$$

where the function $G_a[x;A]$ encodes the anomaly. This function is closely related to the non-zero right hand side of the anomalous conservation law for the current associated to the symmetry, since the effective action and the current are related by

$$J^{\mu a}(x) + \frac{\delta \Gamma[A]}{\delta A_\mu^a(x)} = 0 \, , \tag{9.66}$$

which implies

$$\left(D_\mu^{\text{adj}}\right)_{ba} J^{\mu a}(x) = -G_b[x;A] \, . \tag{9.67}$$

Since the anomaly is local, $G_b[x;A]$ should be a local (at the point x) polynomial in the gauge field and its derivatives. One may then check that the operators $\mathcal{T}_a(x)$ obey the following commutation relation,

$$\left[\mathcal{T}_a(x), \mathcal{T}_b(y)\right] = i \, g \, f^{abc} \, \delta(x-y) \, \mathcal{T}_c(x) \, , \tag{9.68}$$

where the f^{abc} are the structure constants of the gauge group. From this, we deduce the following identity

$$\mathcal{T}_a(x)\, G_b[y;A] - \mathcal{T}_b(y)\, G[x;A] = i\, g\, f^{abc}\, \delta(x-y)\, G_c[x;A]\ , \tag{9.69}$$

called the *Wess-Zumino consistency conditions*. Since this identity is linear in the anomaly function G_a, it cannot constrain its overall normalization (for this, it is usually necessary to compute the triangle diagram). However, this equation is strong enough to fully constrain its dependence on the gauge field from the term of lowest order in A.

9.3.2 BRST form of the Wess-Zumino condition

The consistency condition can be recasted into a more convenient form that involves BRST symmetry. Let us introduce a ghost field χ_a, and recall that the BRST transformation reads:

$$\mathbf{Q}_{\text{BRST}} A^a_\mu(x) = \left(D^{\text{adj}}_\mu\right)_{ab} \chi_b(x) \quad , \qquad \mathbf{Q}_{\text{BRST}} \chi_a(x) = -\frac{g}{2}\, f^{abc}\, \chi_b(x)\, \chi_c(x)\ . \tag{9.70}$$

Then, let us encapsulate the anomaly function into the following local functional of ghost number $+1$:

$$\mathcal{G}[A,\chi] \equiv \int d^4x\ \chi_a(x)\, G_a[x;A]\ . \tag{9.71}$$

We obtain:

$$\begin{aligned}
\mathbf{Q}_{\text{BRST}} \mathcal{G}[A,\chi] &= i\int d^4x d^4y\ \chi_a(x)\chi_b(y)\, \mathcal{T}_b(y) G_a[x;A] \\
&\qquad\qquad -\frac{g}{2}\int d^4x\ f^{abc}\chi_a(x)\chi_b(x)\, G_c[x;A] \\
&= \frac{i}{2}\int d^4x d^4y\ \chi_a(x)\chi_b(y)\Big\{\mathcal{T}_b(y) G_a[x;A] - \mathcal{T}_a(x) G_b[y;A] \\
&\qquad\qquad \underbrace{+i\, g\, \delta(x-y)\, f^{abc} G_c[x;A]}_{=0}\Big\}\ .
\end{aligned} \tag{9.72}$$

Therefore, the Wess-Zumino consistency conditions are equivalent to the statement that the functional $\mathcal{G}[A,\chi]$ is BRST-invariant:

$$\mathbf{Q}_{\text{BRST}} \mathcal{G}[A,\chi] = 0\ . \tag{9.73}$$

Since $\mathbf{Q}_{\text{BRST}}$ is nilpotent, a trivial solution of this equation is of course

$$\mathcal{G}[A,\chi] = \mathbf{Q}_{\text{BRST}}\, h[A]\ , \tag{9.74}$$

where $h[A]$ does not depend on the ghost field (indeed, Q_{BRST} increases the ghost number by one unit, and $\mathcal{G}[A,\chi]$ must have ghost number unity). But since $h[A]$ is a local functional of the gauge field, it may be subtracted from the action to cancel the anomaly. Thus, genuine anomalies are given by local functionals $\mathcal{G}[A,\chi]$ of ghost number $+1$ that satisfy the consistency condition

(9.73), modulo a term obtained by acting with $\mathbf{Q}_{\text{BRST}}$ on a functional of A only. Note that if we write $\mathcal{G}[A,\chi]$ as the integral of a local density,

$$\mathcal{G}[A,\chi] \equiv \int d^4x\, \mathcal{G}(x)\ , \tag{9.75}$$

then the BRST action on the density should be a total derivative

$$\mathbf{Q}_{\text{BRST}}\, \mathcal{G}(x) = \partial_\mu \zeta^\mu\ . \tag{9.76}$$

9.3.3 Solution of the consistency condition

In order to determine how the Wess-Zumino equation constrains $\mathcal{G}(x)$, the language of *differential forms* introduced in the section 4.7.3 is very handy, as a way to encapsulate both Lorentz and group indices in compact objects. The 1-forms dx^μ anticommute among themselves under the exterior product $\wedge$. In addition, they also anticommute with the ghost field and the BRST generator $\mathbf{Q}_{\text{BRST}}$. The volume element weighted by the fully antisymmetric tensor $\epsilon^{\mu\nu\rho\sigma}$ can therefore be written as

$$d^4x\ \epsilon^{\mu\nu\rho\sigma} = dx^\mu \wedge dx^\nu \wedge dx^\rho \wedge dx^\sigma\ . \tag{9.77}$$

Then, given a vector V_μ and the corresponding 1-form

$$\boldsymbol{V} \equiv V_\mu dx^\mu\ , \tag{9.78}$$

we may write in a compact manner

$$\int d^4x\ \epsilon^{\mu\nu\rho\sigma}\, V_\mu V_\nu V_\rho V_\sigma = \int \boldsymbol{V}\wedge\boldsymbol{V}\wedge\boldsymbol{V}\wedge\boldsymbol{V}\ . \tag{9.79}$$

The exterior derivative $\mathbf{d} \equiv \partial_\mu\, dx^\mu\wedge$ satisfies

$$\mathbf{d}^2 = 0 \quad , \qquad \mathbf{Q}_{\text{BRST}}\mathbf{d} + \mathbf{d}\mathbf{Q}_{\text{BRST}} = 0\ . \tag{9.80}$$

If we also denote

$$\mathbf{A} \equiv ig\, A^a_\mu t^a dx^\mu \quad , \qquad \boldsymbol{\chi} \equiv ig\,\chi_a t^a\ , \tag{9.81}$$

(for later convenience, we absorb a factor i in the definitions of $\mathbf{A}$ and $\boldsymbol{\chi}$) the BRST transformations take the following form

$$\begin{aligned} &\mathbf{Q}_{\text{BRST}}\,\mathbf{A} = -\mathbf{d}\boldsymbol{\chi} + \{\mathbf{A},\boldsymbol{\chi}\}\ ,\\ &\mathbf{Q}_{\text{BRST}}\,\boldsymbol{\chi} = \boldsymbol{\chi}^2\ . \end{aligned} \tag{9.82}$$

On dimensional grounds, the anomaly function $\mathcal{G}[A,\chi]$ may contain the following terms:

$$\begin{aligned} \mathcal{G}[A,\chi] \;=\; & -iC\int d^4x\ \epsilon^{\mu\nu\rho\sigma}\,\chi_a\ \mathrm{tr}\Big\{t^a\Big((\partial_\mu A_\nu)(\partial_\rho A_\sigma)\\ & + ia_1\ (\partial_\mu A_\nu)A_\rho A_\sigma + ia_2\ A_\mu(\partial_\nu A_\rho)A_\sigma + ia_3\ A_\mu A_\nu(\partial_\rho A_\sigma)\\ & - b\ A_\mu A_\nu A_\rho A_\sigma\Big)\Big\}\ . \end{aligned} \tag{9.83}$$

The term on the first line comes from the triangle diagram, whose explicit calculation gives the overall coefficient C. The terms of the second and third lines come from the square and pentagon diagrams, respectively. Alternatively, they can be obtained from the consistency conditions. Firstly, the previous equation may be rewritten as a sum of forms:

$$\begin{aligned}\mathcal{G}[A,\chi] &= \gamma\int \mathrm{tr}\left\{\chi\left((\mathbf{dA})\wedge(\mathbf{dA})\right)+\alpha_1\,(\mathbf{dA})\wedge\mathbf{A}\wedge\mathbf{A}+\alpha_2\,\mathbf{A}\wedge(\mathbf{dA})\wedge\mathbf{A}\right.\\ &\qquad\left.+\alpha_3\,\mathbf{A}\wedge\mathbf{A}\wedge(\mathbf{dA})+\beta\,\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\right)\Big\}\,, \end{aligned}\tag{9.84}$$

where $\gamma,\alpha_{1,2,3},\beta$ are constants related to $C,a_{1,2,3},b$. Consider first the BRST transform of the last term,

$$\mathbf{Q}_{\text{BRST}}\,\mathrm{tr}\left\{\chi\,\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\right\}=\mathrm{tr}\left\{\chi^2\,\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\right\}+\text{terms in }\chi\,(\mathbf{d}\chi)\wedge\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\,.\tag{9.85}$$

Since $\mathbf{Q}_{\text{BRST}}$ cannot increase the degree in $\mathbf{A}$, the term in $\chi^2\,\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}$ cannot be canceled by the terms in $\alpha_{1,2,3}$, and therefore we must have $\beta=0$. We need then to evaluate the BRST transformation of the other terms. For instance,

$$\begin{aligned}\mathbf{Q}_{\text{BRST}}\,\mathrm{tr}\left\{\chi\,(\mathbf{dA})\wedge(\mathbf{dA})\right\} &= \mathrm{tr}\Big\{-\chi^2\,(\mathbf{dA})\wedge(\mathbf{dA})+\chi\,(\mathbf{d}\chi)\wedge\mathbf{A}\wedge(\mathbf{dA})\\ &\qquad-(\mathbf{d}\chi)\chi\wedge(\mathbf{dA})\wedge\mathbf{A}-\mathbf{A}\chi\wedge(\mathbf{dA})\wedge(\mathbf{d}\chi)\\ &\qquad-\chi\,\mathbf{A}\wedge(\mathbf{d}\chi)\wedge(\mathbf{dA})\Big\}\,.\end{aligned}\tag{9.86}$$

By evaluating similarly the BRST transforms of the other terms, one can check that when $\alpha_1=-\alpha_2=\alpha_3=-1/2$ the BRST transform of the anomaly functional is the integral of an exact form and therefore vanishes:

$$\mathbf{Q}_{\text{BRST}}\,\mathcal{G}[A,\chi]=\gamma\int_{\mathbb{R}^4}\mathbf{d}\mathcal{F}=\gamma\int_{\partial\mathbb{R}^4}\mathcal{F}=0\,.\tag{9.87}$$

This is in fact the only possibility. Introducing the field strength 2-form,

$$\mathbf{F}\equiv\mathbf{dA}-\mathbf{A}\wedge\mathbf{A}=\frac{ig}{2}\,t^a\,F^a_{\mu\nu}\,dx^\mu dx^\nu\,,\tag{9.88}$$

the anomaly functional for these values of the coefficients can then be rewritten as

$$\mathcal{G}[A,\chi]=\gamma\int\mathrm{tr}\left\{\chi\;\mathbf{d}\left[\mathbf{A}\wedge\mathbf{F}+\frac{1}{2}\mathbf{A}\wedge\mathbf{A}\wedge\mathbf{A}\right]\right\}.\tag{9.89}$$

Therefore, except for the prefactor C whose determination requires to calculate the triangle diagram, the consistency relations completely determine the dependence of the anomaly function on the gauge field.

9.4 't Hooft anomaly matching

Some models of physics beyond the Standard Model conjecture that the quarks and leptons are bound states of more fundamental degrees of freedom, confined by some strong gauge interaction at a scale $\Lambda\gg\Lambda_{\text{electroweak}}$. A difficulty with this picture is to explain the fact that quarks

and leptons are light (in fact, massless, if it were not for electroweak symmetry breaking), while being bound states of some strong interaction at a much higher scale. Indeed, the naive mass of these confined states is naturally of order Λ (the Goldstone mechanism cannot give light *fermions*, only scalar particles).

As shown by 't Hooft, one way this may happen is to have in the underlying fundamental theory a global chiral symmetry with generators T^a, such that the anomaly function $\mathrm{tr}\,(\mathsf{T}^a\{\mathsf{T}^b,\mathsf{T}^c\})$ is non-zero. In the low energy sector of the spectrum of this theory, there must be spin $1/2$ massless bound states, on which this chiral symmetry acts with generators $\mathbb{T}^a$, and whose anomaly coefficients are identical to the high energy ones:

$$\mathrm{tr}\left(\mathbb{T}^a\left\{\mathbb{T}^b,\mathbb{T}^c\right\}\right)=\mathrm{tr}\left(\mathsf{T}^a\left\{\mathsf{T}^b,\mathsf{T}^c\right\}\right)\,. \tag{9.90}$$

The proof of this assertion goes as follows. Let us first couple a fictitious weakly coupled gauge boson to the generators T^a. We also introduce additional fictitious massless fermions coupled only to the fictitious gauge boson, but not to the strongly interacting gauge bosons responsible for the confinement, tuned so that their contribution exactly cancels the anomaly:

$$\left[\mathrm{tr}\left(\mathsf{T}^a\left\{\mathsf{T}^b,\mathsf{T}^c\right\}\right)\right]_{\substack{\text{physical}\\ \text{high energy}}}+\left[\mathrm{tr}\left(\mathsf{T}^a\left\{\mathsf{T}^b,\mathsf{T}^c\right\}\right)\right]_{\substack{\text{fictitious}\\ \text{fermions}}}=0\,. \tag{9.91}$$

Let us now examine the low energy part of the spectrum of this theory, i.e. at energies much lower than the strong scale Λ. Since they are not coupled in any way to the strong sector, this low energy spectrum contains the fictitious gauge bosons and massless fermions, unmodified compared to what we have introduced at high energy. In addition, this spectrum contains the bound states made of the trapped fermions and strongly interacting gauge bosons. For consistency, this low energy description must also be anomaly-free, which means that the bound states must transform under the chiral symmetry with generators $\mathbb{T}^a$, such that

$$\left[\mathrm{tr}\left(\mathbb{T}^a\left\{\mathbb{T}^b,\mathbb{T}^c\right\}\right)\right]_{\substack{\text{physical}\\ \text{bound states}}}+\left[\mathrm{tr}\left(\mathsf{T}^a\left\{\mathsf{T}^b,\mathsf{T}^c\right\}\right)\right]_{\substack{\text{fictitious}\\ \text{fermions}}}=0\,. \tag{9.92}$$

The crucial point in this argument is that the contribution of the fictitious fermions is the same in the equations (9.91) and (9.92), because these fermions are not coupled to the strongly interacting sector. Eqs. (9.91) and (9.92) immediately give (9.90). In other words, *the anomalies of the trapped elementary fermions must be mimicked by those of the massless spin* $1/2$ *bound states they are confined into.*

Chapter 10

Localized field configurations

All the applications of quantum field theory we have encountered so far amount to study situations that may be viewed as small perturbations above the vacuum state; i.e. interactions involving states that contain only a few particles. Besides the fact that these situations are actually encountered in scattering experiments, their importance stems from the stability of the vacuum, that makes it a natural state to expand around.

In this chapter, we will study other field configurations, classically stable, that may also be sensible substrates for expansions that differ from the standard perturbative expansion that we have studied until now. However, under normal circumstances, a localized "blob" of fields is not stable: it will usually decay into a field which is zero everywhere. As we shall see, the stability of the field configurations considered in this chapter is due to topological obstructions that prevent a smooth transformation between the field configuration of interest and the null field that corresponds to the vacuum. These field configurations can be classified according to their space-time structure:

- *Event-like :* localized both in time and space (e.g., instantons). These may be viewed as local extrema of the 4-dimensional action, and therefore may give a (non-perturbative) contribution to path integrals.
- *Worldline-like :* localized in space, independent of time (e.g., skyrmions, monopoles). These field configurations behave very much like stable particles (at least classically), and their non-trivial topology confers them conserved charges.
- *Strings, Domain walls :* extended in one or two spatial dimensions, independent of time.

10.1 Domain walls

A domain wall is a 2-dimensional[1] interface between two regions of space where a discrete symmetry is broken in different ways. Their simplest realization arises in a real scalar field

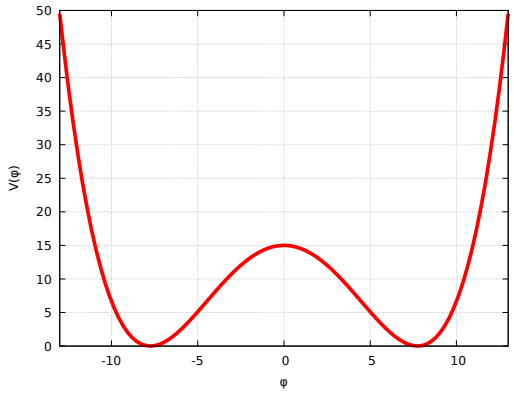
[1] This is for 4-dimensional spacetime. In D-dimensional spacetime, domain walls have dimension $D-2$.

theory, symmetric under $\phi \to -\phi$, but with a potential that leads to spontaneous symmetry breaking, such as

$$V(\phi) \equiv V_0 - \frac{\mu^2}{2}\phi^2 + \frac{\lambda}{4!}\phi^4 \,, \tag{10.1}$$

where the constant shift V_0 is chosen so that the minima of this potential are 0. There are two

Figure 10.1: *Quartic potential (10.1) exhibiting spontaneous symmetry breaking.*



such minima, at field values

$$\phi = \pm\phi_* \,, \quad \phi_* \equiv \sqrt{\frac{6\,\mu^2}{\lambda}} \,. \tag{10.2}$$

In order to simplify the discussion, let us consider field configurations that depend only on x, and are independent of time, as well as of the transverse coordinates y, z. We seek field configurations that obey the classical field equation of motion,

$$-\partial_x^2\phi + V'(\phi) = 0 \,, \tag{10.3}$$

and have a finite energy (per unit of transverse area),

$$\frac{d\mathcal{E}}{dy\,dz} = \int_{-\infty}^{+\infty} dx \left\{ \tfrac{1}{2}\big(\partial_x\phi(x)\big)^2 + V(\phi(x)) \right\} < \infty \,. \tag{10.4}$$

This energy density is the sum of two positive definite terms (since we have adjusted the potential so that its minima are $V(\pm\phi_*) = 0$. For the integral over x to converge when $x \to \pm\infty$, it is necessary that $\phi(x)$ becomes constant when $|x| \to \infty$, and that this constant be $+\phi_*$ or $-\phi_*$. There are therefore four possibilities for the values of the field at $x = \pm\infty$:

$$\begin{aligned}
(\mathbf{i}) \quad &: \quad \phi(-\infty) = +\phi_* \,, \quad \phi(+\infty) = +\phi_* \,, \\
(\mathbf{ii}) \quad &: \quad \phi(-\infty) = -\phi_* \,, \quad \phi(+\infty) = -\phi_* \,, \\
(\mathbf{iii}) \quad &: \quad \phi(-\infty) = -\phi_* \,, \quad \phi(+\infty) = +\phi_* \,, \\
(\mathbf{iv}) \quad &: \quad \phi(-\infty) = +\phi_* \,, \quad \phi(+\infty) = -\phi_* \,.
\end{aligned} \tag{10.5}$$

The first two of these possibilities do not lead to stable field configurations of positive energy, because they can be continuously deformed (while holding the asymptotic values unchanged) into the constant fields $\phi(x) = +\phi_*$, or $\phi(x) = -\phi_*$, respectively, that have zero energy. Physically, this means that if one creates a field configuration with these boundary values, it will decay into a constant field (i.e., the regions where the field was excited to values different from $\pm\phi_*$ will dilute away to $|x| = \infty$).

The interesting cases are encountered when the field takes values corresponding to opposite minima at $x = -\infty$ and $x = +\infty$. If one holds the asymptotic values of the field fixed, then it is not possible to deform continuously such a field configuration into one that would have zero energy. Thus, there must be stable field configurations of positive energy with these boundary values. A very handy trick, due to Bogomol'nyi, is to rewrite the energy density as follows:

$$\frac{d\mathcal{E}}{dydz} = \frac{1}{2}\int_{-\infty}^{+\infty} dx \left(\partial_x\phi(x) \pm \sqrt{2V(\phi(x))}\right)^2 \mp \int_{\phi(-\infty)}^{\phi(+\infty)} d\phi \sqrt{2V(\phi)} \,. \tag{10.6}$$

In the cases **i**, **ii**, the second term vanishes, and the energy density is allowed to be zero, by having a constant field equal to $\pm\phi_*$. Let us consider now the case **iii**. In this case, it is convenient to choose the minus sign in the first term, so that

$$\frac{d\mathcal{E}}{dydz} = \frac{1}{2}\int_{-\infty}^{+\infty} dx \left(\partial_x\phi(x) - \sqrt{2V(\phi(x))}\right)^2 + \underbrace{\int_{-\phi_*}^{+\phi_*} d\phi \sqrt{2V(\phi)}}_{>0} \,. \tag{10.7}$$

The second term is now strictly positive, and does not depend on the details of $\phi(x)$ (except its boundary values). Since the first term is the integral of a square, this implies that there is no field configuration of zero energy with this boundary condition. The minimal energy density possible with this boundary condition is

$$\left.\frac{d\mathcal{E}}{dydz}\right|_{\min} = \int_{-\phi_*}^{+\phi_*} d\phi \sqrt{2V(\phi)} \,, \tag{10.8}$$

reached for a field configuration that obeys

$$\partial_x\phi(x) = \sqrt{2V(\phi(x))} \,. \tag{10.9}$$

Taking one more derivative implies that

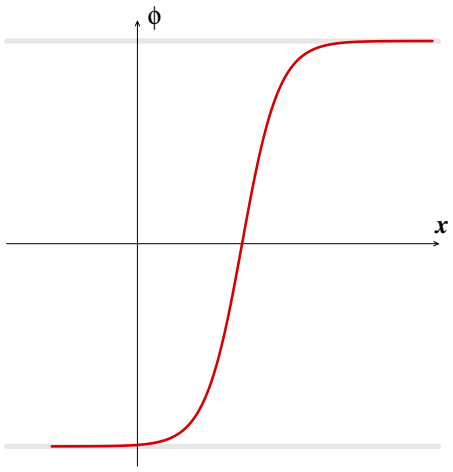
$$\partial_x^2\phi = \frac{(\partial_x\phi)\,V'(\phi)}{\sqrt{2V(\phi)}} = V'(\phi) \,, \tag{10.10}$$

which is nothing but the classical equation of motion (10.3). Since this is a second order equation, it admits in general a unique solution with prescribed boundary values $\pm\phi_*$ at $x = \pm\infty$. Such a solution interpolates between the two ground states of the potential of the figure 10.1. The ground state $\phi = +\phi_*$ is realized at $x \to +\infty$, while the other ground state is realized at $x \to -\infty$. Since these two vacua correspond to two different ways to spontaneously break the $\phi \to -\phi$ symmetry, there must exist an interface between the two phases, called a *domain wall*. From eq. (10.9), we may write

$$x(\phi) = x_0 + \int_0^{\phi} \frac{d\xi}{\sqrt{2V(\xi)}} \,, \tag{10.11}$$

where x_0 is an integration constant that can be interpreted as the coordinate where the field ϕ is zero. In other words, x_0 is the location of the center of the domain wall that separates the regions of different vacua. The domain wall is a local minimum of the energy density (and the absolute

Figure 10.2: *Domain wall profile corresponding to the potential of the figure 10.1.*



minimum for the mixed boundary conditions **iii**). Moreover, it is separated from the (lower energy) configurations **i**, **ii** that have a constant field by an infinite energy barrier[2]. Indeed, going from **iii** to **i** implies shifting the value of the field from $-\phi_*$ to $+\phi_*$ in the (infinite) vicinity of $x = -\infty$. In the middle of this process, the field in this region will be $\phi = 0$, at which $V(\phi) = V_0 > 0$, a configuration that has an infinite energy density. Thus, the domain wall solution is stable, except for shifts of x_0 (since the energy density is independent of x_0): the domain wall may move along the x axis, but cannot disappear.

Let us finish by a note on the y, z dependence that has been neglected so far. Reintroducing the transverse dependence adds the term $\frac{1}{2}\left((\partial_y\phi)^2 + (\partial_z\phi)^2\right)$ to the integrand of the energy density in eq. (10.4). These terms are positive, or zero for fields that do not depend on y and z. Therefore, the minimum of energy density is reached for domain walls that are invariant by translation in the transverse directions. Domain walls that are not translation invariant are not stable, but will relax to this y, z-invariant configuration. Physically, one may view the term $\frac{1}{2}\left((\partial_y\phi)^2 + (\partial_z\phi)^2\right)$ as a surface tension energy, and the energetically favored configurations are those for which the interface has the lowest curvature.

[2] From this fact, we may infer that domain walls are also stable quantum mechanically.

10.2 Skyrmions

Skyrmions are field configurations that arise in model resulting from a spontaneous symmetry breaking, such as a non-linear sigma model. Consider for instance the following action,

$$S[\xi] = \int d^D x \left\{ \frac{1}{2} \sum_{a,b} g_{ab}(\xi) \left(\partial_i \xi^a\right) \left(\partial_i \xi^b\right) + \cdots \right\} , \tag{10.12}$$

where the fields ξ^a are the Nambu-Goldstone bosons of a broken symmetry from the symmetry group $\mathcal{G}$ down to $\mathcal{H}$. The matrix $g_{ab}(\xi)$ is positive definite, and in general field dependent. The dots represent terms with higher derivatives, that we have not written explicitly. In such a model, the Nambu-Goldstone fields ξ^a may be viewed as elements of the coset $\mathcal{G}/\mathcal{H}$.

In order to have a finite action, the derivatives of the fields should decrease faster than $|x|^{-D/2}$ at large distance,

$$\left|\partial_i \xi^a(x)\right| \underset{|x|\to\infty}{\lesssim} |x|^{-D/2} , \tag{10.13}$$

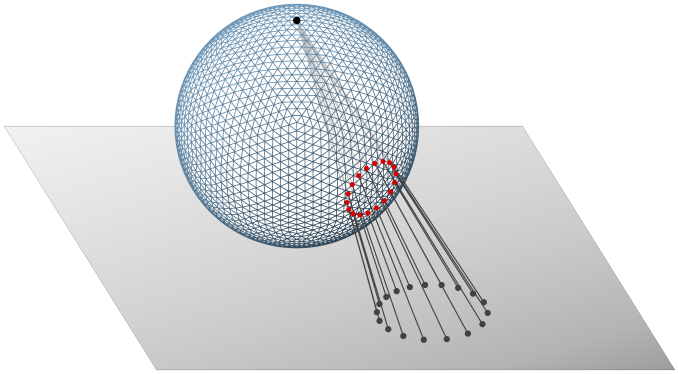
which means that the field $\xi^a(x)$ should go to a constant, with a remainder that decreases faster than $|x|^{1-D/2}$.

The constant value of ξ^a at infinity can be chosen to be some fixed predefined element of $\mathcal{G}/\mathcal{H}$. Thus, we may view the field $\xi^a(x)$ as a mapping

$$\xi^a \ : \quad \mathcal{S}_D \mapsto \mathcal{G}/\mathcal{H} , \tag{10.14}$$

where $\mathcal{S}_D$ is the D-dimensional sphere, which is topologically equivalent to the euclidean space $\mathbb{R}^D$ with all the points $|x| = \infty$ identified as a single point. This equivalence may be made manifest by a stereographic projection, illustrated for $D = 2$ in the figure 10.3.

Figure 10.3: *Stereographic projection that maps the plane $\mathbb{R}^2$ to the sphere $\mathcal{S}_2$. All the points at infinity in the plane are identified, and mapped to the north pole of the sphere.*



These mappings, taking a fixed value at $|x| = \infty$, can be organized into topological classes containing functions that can be continuously deformed into one another. The set of these classes is a group, known as the D-th homotopy group of $\mathcal{G}/\mathcal{H}$, denoted $\pi_D(\mathcal{G}/\mathcal{H})$.

The original version of this model was intended to describe nucleons as a topologically stable configuration of the pion field. In this case, there are $D = 3$ spatial dimensions, and the chiral symmetry $SU(2) \times SU(2)$ is spontaneously broken to $SU(2)$. The coset in which ξ^a lives is $SU(2)$, and the relevant homotopy group is $\pi_3(SU(2)) = \mathbb{Z}$. The integer that enumerates the topological classes is then identified with the baryon number.

Note that the model defined by eq. (10.12), with only second order derivatives, cannot have stable solutions, a result known as *Derricks' theorem*. In order to see this, consider a skyrmion solution $\xi^a(x)$, and construct another field by a rescaling:

$$\xi^a_R(x) \equiv \xi^a(x/R) \; . \tag{10.15}$$

The action becomes $S[\xi_R] = R^{D-2} S[\xi]$. In $D > 2$ dimensions, we may make it decrease continuously to zero, despite the fact that ξ^a and ξ^a_R have the same topology. Such a solution may be stabilized by adding a term with higher derivatives, such as

$$V[\xi] \equiv \int d^D x \, h_{abcd}(\xi) \Big(\partial_i \xi^a \partial_i \xi^b \Big) \Big(\partial_j \xi^c \partial_j \xi^d \Big) \; . \tag{10.16}$$

Under the same rescaling, we now have $V[\xi_R] = R^{D-4} V[\xi]$. In $D = 3$ spatial dimensions, the term with second derivatives decreases to zero when $R \to 0$, while the above quartic term increases to $+\infty$. Their sum therefore exhibits an extremum at some finite scale R_*. Although we obtain in this way non-trivial stable solutions, there is a priori no reason to limit ourselves to terms with four derivatives, and therefore the predictive power of such a model is limited by the many possible choices for these higher order terms.

10.3 Monopoles

10.3.1 Dirac monopole

Magnetic monopoles are not forbidden in quantum electrodynamics, but their existence would automatically lead to the quantization of electrical charge, as first noted by Dirac. Let us reproduce here this argument. Consider the magnetic field of a would-be monopole:

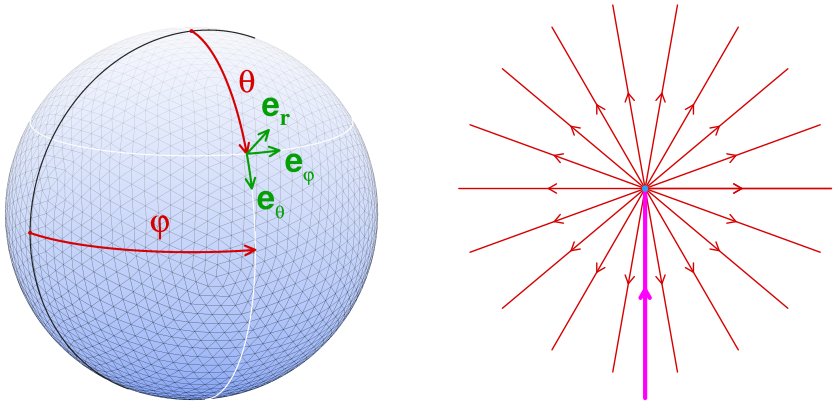
$$\mathbf{B} = g \, \frac{\widehat{x}}{|x|^2} \; . \tag{10.17}$$

Maxwell's equation $\boldsymbol{\nabla} \cdot \mathbf{B} = 0$ implies that we cannot find a vector potential $\mathbf{A}$ for this magnetic field in all space. But it is possible to find one that works almost everywhere, for instance

$$\mathbf{A}(x) = g \, \frac{1 - \cos\theta}{|x| \sin\theta} \, \boldsymbol{e}_\varphi \; , \tag{10.18}$$

where θ is the polar angle, ϕ the azimuthal angle, and $\boldsymbol{e}_\varphi$ is the unit vector tangent to the circle of constant $|x|$ and θ. This vector potential is not defined on the semi-axis $\theta = \pi$ (i.e. the semi-axis of negative z). One may argue that on this semi-axis, we have in addition to the monopole field a singular B_z whose magnetic flux precisely cancels the magnetic flux of the monopole, so that the total flux on any closed surface containing the origin is zero, as illustrated

Figure 10.4: *Left: notations for the polar coordinates local frame used in eq. (10.18). Right: magnetic field lines of the Dirac monopole, corresponding to the vector potential of eq. (10.18).*



in the figure 10.4. Thus, in this solution, the magnetic flux $\Phi_m \equiv 4\pi g$ of the monopole is "brought from infinity" by an infinitely thin "solenoid". Even if it is infinitely thin, such a solenoid may in principle be detected by looking for interferences between the wavefunctions of charged particles that have propagated left and right of the solenoid (this corresponds to the *Aharonov-Bohm effect*). For a particle of electrical charge e, the corresponding phase shift is $e\Phi_m = 4\pi eg$. Dirac pointed out that this interference is absent when the phase shift is a multiple of 2π, i.e. when the electric and magnetic charges are related by

$$ge = \frac{n}{2}\,, n \in \mathbb{Z}\,. \tag{10.19}$$

Thus, electrodynamics can perfectly accommodate genuine magnetic monopoles, provided this condition is satisfied, since the annoying solenoid that comes with the above vector potential is totally undetectable. In particular, this implies that electrical charges should be multiples of some elementary quantum of electrical charge if monopoles exist. Note that in quantum electrodynamics, while the electric and magnetic charges must be related by eq. (10.19), there is no constraint a priori on the mass of monopoles and it should be regarded as a free parameter.

Let us mention briefly an alternative argument, that does not involve discussing the detectability of Dirac's solenoid. Instead of the vector potential of eq. (10.18), one could instead have chosen

$$\boldsymbol{A}'(x) = -g\,\frac{1+\cos\theta}{|x|\sin\theta}\,\boldsymbol{e}_\varphi\,, \tag{10.20}$$

that has a singularity on the semi-axis $\theta = 0$. When Dirac's quantization condition is satisfied, one may patch eqs. (10.18) and (10.20) in order to obtain a vector potential which is regular in all space (except at the origin, where the monopole is located). To see this, consider a region Ω_1 corresponding to $0 \leq \theta \leq 3\pi/4$ and a region Ω_2 corresponding to $\pi/4 \leq \theta \leq \pi$. Then, we choose $\boldsymbol{A}$ in Ω_1 and $\boldsymbol{A}'$ in Ω_2. In the overlap of the two regions, $\pi/4 \leq \theta \leq 3\pi/4$, we have

$$(\boldsymbol{A} - \boldsymbol{A}') \cdot dx = 2g\,d\phi\,, \tag{10.21}$$

and we can write

$$\mathbf{A} - \mathbf{A}' = \boldsymbol{\nabla}\chi \quad \text{with } \chi(\phi) \equiv 2g\,\phi\ . \tag{10.22}$$

For this to be an acceptable gauge transformation, the phase by which it multiplies the wave-function of a charged particle should be single-valued, i.e.

$$e^{ie\,\chi(\phi+2\pi)} = e^{ie\,\chi(\phi)}\ , \tag{10.23}$$

which is precisely the case when the condition (10.19) is satisfied.

This argument can even be made without any reference to the explicit solutions (10.18) and (10.20). Let us consider a large sphere surrounding the origin, divide it in an upper and lower hemispheres (see the figure 10.7), and denote $\mathbf{A}$ and $\mathbf{A}'$ the vector potentials that represent the monopole in these two hemispheres. On the equator, their difference should be a pure gauge,

$$\mathbf{A} - \mathbf{A}' = \frac{i}{e}\,\Omega^\dagger(x)\left(\boldsymbol{\nabla}\,\Omega(x)\right)\ . \tag{10.24}$$

Along the equator, we have

$$\Omega(\phi) = \Omega(0)\ \exp\left\{-ie\int_{\gamma[0,\phi]}\left(\mathbf{A}-\mathbf{A}'\right)\cdot dx\right\}\ , \tag{10.25}$$

where the integration path $\gamma[0,\phi]$ is the portion of the equator that extends between the azimuthal angles 0 and ϕ. After a complete revolution, we have

$$\begin{aligned}\Omega(2\pi) &= \Omega(0)\,\exp\left\{-ie\oint_{\text{Equator}}\left(\mathbf{A}-\mathbf{A}'\right)\cdot dx\right\}\\ &= \Omega(0)\,\exp\left\{-ie\big(\underbrace{\Phi_U+\Phi_L}_{\text{flux}\,=4\pi g}\big)\right\} = \Omega(0)\,e^{-4\pi i\,eg}\ .\end{aligned} \tag{10.26}$$

To obtain the first equality on the second line, we use Stokes's theorem to rewrite the contour integrals of $\mathbf{A}$ and $\mathbf{A}'$ as surface integrals of the corresponding magnetic field. Therefore, we obtain the magnetic fluxes through the upper and lower hemispheres, respectively, whose sum is the total flux $4\pi g$ of the monopole. Requesting the single-valuedness of Ω leads to Dirac's condition on eg.

10.3.2 Monopoles in non-Abelian gauge theories

There are also non-abelian field theories that exhibit $U(1)$ magnetic monopoles, as classical solutions whose stability is ensured by topology. The simplest example is an $SU(2)$ gauge theory coupled to a Higgs field in the adjoint representation[3], whose Lagrangian density reads

$$\mathcal{L} \equiv -\frac{1}{4}\,F^a_{\mu\nu}F^{a,\mu\nu} + \frac{1}{2}\left(D_\mu\Phi^a\right)\left(D^\mu\Phi^a\right) - V(\Phi)\ , \tag{10.27}$$

[3]This model is known as the Georgi-Glashow model. It was considered at some point as a possible candidate for a field theory of electroweak interactions, until the neutral vector boson Z^0 was discovered. Here, we use it as a didactical example of a theory with classical solutions that are magnetic monopoles.

with

$$\begin{aligned} V(\Phi) &\equiv \frac{\lambda}{8}\left(\Phi^a\Phi^a - v^2\right)^2 , \\ D_\mu\Phi^a &= \partial_\mu\Phi^a - e\,\epsilon_{abc}A^b_\mu\Phi^c , \\ F^a_{\mu\nu} &= \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - e\,e_{abc}\,A^b_\mu A^c_\nu , \end{aligned} \tag{10.28}$$

where we have written explicitly the structure constants of the $\mathfrak{su}(2)$ algebra. In order to study static classical solutions, it is simpler to consider the minima of the energy:

$$\mathcal{E} \equiv \int d^3\boldsymbol{x}\left\{\frac{1}{2}\left(E^a_i E^a_i + B^a_i B^a_i + \left(D_i\Phi^a\right)\left(D_i\Phi^a\right)\right) + V(\Phi)\right\} , \tag{10.29}$$

where $E^a_i \equiv F^a_{0i}$ is the (non-abelian) electrical field and $B^a_i \equiv \frac{1}{2}\,\epsilon_{ijk}F^a_{jk}$ is the magnetic field.

It is possible to choose a gauge (called the unitary gauge) in which the Higgs field triplet takes the form

$$\Phi^a = \left(0, 0, v + \varphi\right) . \tag{10.30}$$

In this equation, we have anticipated spontaneous symmetry breaking, that will give to the Higgs field a vacuum expectation value v, and we have made a specific choice about the orientation of the vacuum in $SU(2)$. The field φ is thus the quantum fluctuation of the Higgs about its expectation value. In this process, the fields $A^{1,2}_\mu$ will become massive (with a mass $M_W = e\,v$), as well as the Higgs field (with mass $M_H = \sqrt{\lambda}\,v$), while the field A^3_μ remains massless (it corresponds to a residual unbroken $U(1)$ symmetry). The classical vacuum of this theory corresponds to

$$\varphi = 0 , \quad A^a_\mu = 0 . \tag{10.31}$$

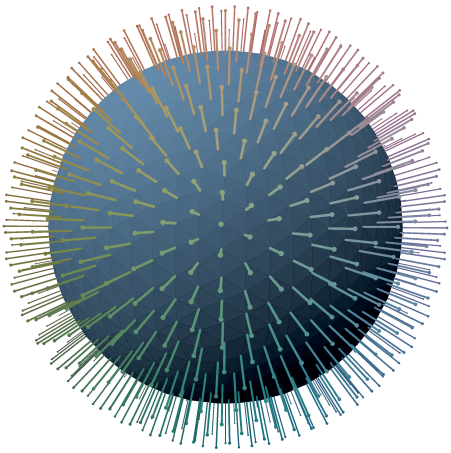
Now, we seek stable classical field configurations that are local minima of the energy, but are not equivalent to the vacuum in the entire space. To prove the existence of such fields, it is sufficient to exhibit a field configuration of non-zero energy that cannot be continuously deformed into the null fields of eq. (10.31) (up to a gauge transformation). In order to have a finite energy, the scalar field Φ^a should reach a minimum of the potential $V(\Phi)$ at large distance $|x| \to \infty$ (we have shifted the potential so that its minimum is zero), but it may approach different minima depending on the direction $\widehat{x}$ in space. The allowed asymptotic behaviors of Φ^a define a mapping from the sphere S_2 (the orientations $\widehat{x}$, for three spatial dimensions) to the sphere $\Phi^a\Phi^a = v^2$ of the minima of $V(\Phi)$. Since $\mathfrak{su}(2)$ is 3-dimensional, the set of zeroes of the Higgs potential is also a sphere S_2, and it is natural to consider the following example[4]:

$$\Phi^a(\widehat{x}) \equiv v\,\widehat{x}^a , \tag{10.32}$$

sometimes called a "hedgehog" configuration because the direction of color space pointed to by the Higgs field is locked to the spatial direction, as shown in the figure 10.5. Any smooth classical field Φ^a that obeys this boundary condition at infinite spatial distance must vanish at some point in the interior of the sphere. Therefore, is cannot simply be a gauge transform of

[4]Here, we see that it is crucial that the Higgs potential has non-trivial minima. If $\Phi^a \equiv 0$ was the only minimum, it would not be possible to construct solutions of finite energy that are not topologically equivalent to the vacuum.

Figure 10.5: *Cartoon of the hedgehog configuration of eq. (10.32). Each needle indicates the color orientation of Φ^a at the corresponding point on the sphere.*



the constant field $\Phi^a = v\,\delta_{3a}$ (the expectation value of the Higgs field in the vacuum). Once again, the classes of fields that can be continuously deformed into one another are given by a homotopy group, in this case the group $\pi_2(\mathcal{M}_0)$ where $\mathcal{M}_0$ is the manifold of the minima of the Higgs potential. For the $SU(2)$ group, $\mathcal{M}_0$ is topologically equivalent to the 2-sphere $\mathcal{S}_2$, and the equivalence classes of the mappings $\mathcal{S}_2 \mapsto \mathcal{S}_2$ are indexed by the integers, since $\pi_2(\mathcal{S}_2) = \mathbb{Z}$. The hedgehog field of eq. (10.32) has topological number $+1$, while the vacuum has topological number 0.

At spatial infinity, the hedgehog configuration (10.32) is gauge equivalent to the standard Higgs vacuum aligned with the third color direction, $\Phi^a = v\,\delta_{3a}$. In order to see this, let us introduce the following $SU(2)$ transformation, that depends on the polar angle θ and azimuthal angle ϕ as follows[5]:

$$\Omega(\theta,\phi) \equiv -\cos\frac{\theta}{2}\sin\phi + 2i\left(\sin\frac{\theta}{2}\,t_f^1 + \cos\frac{\theta}{2}\cos\phi\,t_f^3\right), \tag{10.33}$$

where the t_f^a are the generators of the fundamental representation of $\mathfrak{su}(2)$. Then, one may check explicitly that

$$\delta_{3a}\,\Omega^\dagger t^a \Omega = \sin\theta\Big(\cos\phi\,t_f^1 + \sin\phi\,t_f^2\Big) + \cos\theta\,t_f^3 = \hat{x}_a t_f^a\ . \tag{10.34}$$

Thus, Ω transforms the usual Higgs vacuum into the hedgehog configuration at infinity. Note that (10.33) is not a valid gauge transformation over the entire space because it is not well defined at the origin.

[5] When an $SU(2)$ transformation in the fundamental representation is written as

$$\Omega \equiv u_0 + 2i\,u_a\,t_f^a\ ,$$

its unitarity ($\Omega^\dagger\Omega = 1$) is equivalent to $u_0^2 + u_1^2 + u_2^2 + u_3^2 = 1$.

The choice of eq. (10.32) for the asymptotic behavior of the Higgs field was motivated by the requirement that the potential $V(\Phi)$ gives a finite contribution to the energy. The term in $\left(D_i\Phi^a\right)^2$ should also give a finite contribution. However, note that

$$\partial_i\Phi^a(\widehat{x}) = \frac{v}{|x|}\left(\delta^{ia} - \widehat{x}^i\widehat{x}^a\right) \tag{10.35}$$

is not square integrable. We must therefore adjust the asymptotic behavior of the gauge potential in order to cancel this term in the covariant derivative, by requesting that

$$\epsilon_{abc}\, A_i^b\, \widehat{x}^c \underset{|x|\to\infty}{=} \frac{\delta^{ia} - \widehat{x}^i\widehat{x}^a}{e\,|x|}\ , \tag{10.36}$$

which is satisfied if

$$A_i^b \underset{|x|\to\infty}{=} \frac{\epsilon_{ibd}\,\widehat{x}^d}{e\,|x|} + \text{term in } \widehat{x}^b\ . \tag{10.37}$$

The corresponding field strength and magnetic field are given by

$$\begin{aligned} F_{ij}^a &\underset{|x|\to\infty}{=} \frac{1}{e\,|x|^2}\left(2\,\epsilon_{ija} + 2\left(\epsilon_{iad}\widehat{x}^j - \epsilon_{jad}\widehat{x}^i\right)\widehat{x}^d - \epsilon_{ijd}\,\widehat{x}^a\widehat{x}^d\right)\ , \\ B_i^a &\underset{|x|\to\infty}{=} \frac{\widehat{x}^i\widehat{x}^a}{e\,|x|^2}\ . \end{aligned} \tag{10.38}$$

Therefore, at large distance (these considerations do not give the precise form of the fields at finite distance) there is a purely radial magnetic field that vanishes like $|x|^{-2}$, i.e. according to Coulomb's law, thus suggesting that a magnetic monopole is present at the origin. For a more robust interpretation, we should apply a gauge transformation that maps the asymptotic Hedgehog Higgs field into the usual Higgs vacuum, aligned with the third color direction. Thanks to eq. (10.34), we see that in this process the magnetic field of eq. (10.38), proportional to $\widehat{x}^a$, will become proportional to δ_{3a}. But the third color direction precisely corresponds to the gauge potential that remains massless in the spontaneous symmetry breaking $SU(2) \to U(1)$. Therefore, eq. (10.38) is indeed the magnetic field of a $U(1)$ magnetic monopole. Its flux through a sphere surrounding the origin is

$$\Phi_m = \frac{4\pi}{e}\ , \tag{10.39}$$

equivalent to that of a magnetic charge $g \equiv e^{-1}$ at the origin.

Until now, we have only discussed the implications of requiring a finite energy on the asymptotic form of the Higgs field and of the gauge potentials. In order to obtain their values at finite distance, one may make the following ansatz:

$$\Phi^a(x) = v\,\widehat{x}^a\,f(|x|)\ , \quad A_i^a(x) = \epsilon_{iab}\,\frac{\widehat{x}^b}{e\,|x|}\,g(|x|)\ , \tag{10.40}$$

where f, g are two functions that can be determined from the classical equations of motion. From this solution over the entire space, one sees that the monopole is an extended object made of two parts:

- A compact core, of radius $R_m \sim M_W^{-1}$, in which the $SU(2)$ symmetry is unbroken and the vector bosons are all massless. One may view the core as a cloud of highly virtual gauge bosons and Higgses.
- Beyond this radius, a halo in which the $SU(2)$ symmetry is spontaneously broken. In this halo, up to a gauge transformation, the Higgs field is that of the ordinary broken vacuum, the vector bosons $A_{1,2}$ are massive, and the A_3 field is massless, with a tail that corresponds to a radial $U(1)$ magnetic field.

Given these fields, the total energy of the field configuration can be identified with the mass (in contrast with Dirac's point-like monopole in quantum electrodynamics, whose mass is not constrained) of the monopole (since it is static). It takes the form

$$M_m = \frac{4\pi}{e^2} M_W \, C(\lambda/e^2) \,, \tag{10.41}$$

where $C(\lambda/e^2)$ is a slowly varying function of the ratio of coupling constants, of order unity. Note that the core and the halo contribute comparable amounts to this mass. Interestingly, the size M_W^{-1} of this monopole is much larger (by a factor $\alpha^{-1} = 4\pi/e^2$) than its Compton wavelength M_m^{-1}. Therefore, when $\alpha \ll 1$, the monopole receives very small quantum corrections and is essentially a classical object.

We have argued earlier that the topologically non-trivial configurations of the scalar field that lead to a finite energy can be classified according to the homotopy group $\pi_2(S_2)$. Since this group is the group $\mathbb{Z}$ of the integers, there are monopole solutions with any magnetic charge multiple of e^{-1} (the solution we have constructed explicitly above has topological number 1), i.e.

$$ge = n \,, \quad n \in \mathbb{Z} \,. \tag{10.42}$$

Therefore, in this field theoretical monopole solution, the electrical charge would also be naturally quantized. At first sight, eqs. (10.42) and (10.19) appear to differ by a factor $1/2$. Note however, that in the $SU(2)$ model we are considering in this section, it is possible to introduce matter fields in the fundamental representation[6] that carry a $U(1)$ electrical charge $\pm e/2$ (this is the smallest possible electrical charge in this model). Thus, if rewritten in terms of this minimal electrical charge, the monopole quantization condition (10.42) is in fact identical to Dirac's condition. Although the Georgi-Glashow model studied in this section is no longer considered as phenomenologically relevant, theories that unify the strong and electroweak interactions into a unique compact Lie group (such as $SU(5)$ for instance) do have magnetic monopoles.

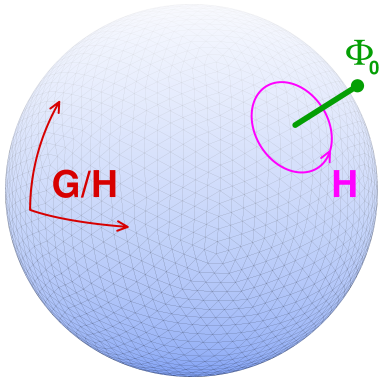
10.3.3 Topological considerations

In the previous two subsections, we have encountered two seemingly different topological classifications of magnetic monopoles. The Dirac monopole appeared closely related to the mappings from a circle (the equator between the two hemispheres in the figure 10.7) to the group $U(1)$, whose classes are the elements of the homotopy group $\pi_1(U(1)) = \mathbb{Z}$. In contrast, the

[6] If Ψ is a doublet that lives in this representation, the covariant derivative acting on it reads:

$$D_\mu \Psi = \partial_\mu \Psi - i\, e\, A^a_\mu t^a_f \Psi = \cdots - i\, \frac{e}{2}\, A^3_\mu \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \Psi \,.$$

Figure 10.6: *Illustration of the symmetry breaking pattern. $\mathcal{H}$ is the residual invariance after choosing a minimum Φ_0. The coset $\mathcal{G}/\mathcal{H}$ is the manifold that holds the minima of $V(\Phi)$ (a 2-sphere in the case of the Georgi-Glashow model).*



monopole discussed in the Georgi-Glashow model was related to the behavior of the Higgs field at large distance, i.e. to mappings from the 2-sphere $\mathcal{S}_2$ to the manifold $\mathcal{M}_0$ of the minima of the Higgs potential $V(\Phi)$, whose equivalence classes are the elements of the homotopy group $\pi_2(\mathcal{M}_0) = \mathbb{Z}$.

Let us now argue that these two ways of viewing monopoles are in fact equivalent. In order to make this discussion more general, consider a gauge theory with internal group $\mathcal{G}$, coupled to a Higgs boson, spontaneously broken to a residual gauge symmetry of group $\mathcal{H}$. Let us denote $\mathcal{M}_0$ the manifold of the minima of the Higgs potential. This manifold is invariant under transformations of $\mathcal{G}$. Given a minimum Φ_0, the other minima can be obtained by multiplying Φ_0 by the elements of $\mathcal{G}$:

$$\mathcal{M}_0 = \left\{ \Phi \;\middle|\; \Phi = \Omega\Phi_0; \Omega \in \mathcal{G} \right\} . \tag{10.43}$$

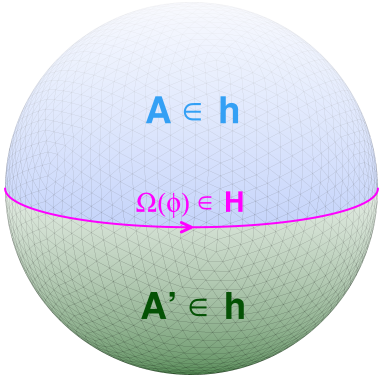
(Here, we are assuming that there are no accidental degeneracies among the minima, i.e. minima Φ and Φ' that are not related by a gauge transformation.) The manifold defined in eq. (10.43) is in fact the coset $\mathcal{G}/\mathcal{H}$,

$$\mathcal{M}_0 = \mathcal{G}/\mathcal{H} . \tag{10.44}$$

This pattern of spontaneous symmetry breaking is illustrated in the figure 10.6.

The first way of classifying monopoles is to consider the gauge field on a sphere, as was done in the subsection 10.3.1. At large distance compared to the inverse mass of the bosons that became massive due to spontaneous symmetry breaking, only the massless gauge bosons contribute, and the corresponding gauge fields live in the algebra $\mathfrak{h}$ of the residual group $\mathcal{H}$. We can reproduce the argument made at the end of the subsection 10.3.1. The gauge potentials in the upper and lower hemispheres are related on the equator by a gauge transformation $\Omega(\phi) \in \mathcal{H}$, that must be single-valued as the azimuthal angle ϕ wraps around the equator. $\Omega(\phi)$ is therefore a mapping from the circle $\mathcal{S}_1$ to the residual gauge group $\mathcal{H}$. These mappings can be grouped

Figure 10.7: *Decomposition of the sphere into two hemispheres with gauge potentials* $\mathbf{A}$ *and* $\mathbf{A}'$.



into classes that differ by their winding number. In this general setting, we may adopt the winding number as the *definition* of the product eg of the electric charge by the magnetic charge comprised within the sphere. Note that $\pi_1(\mathcal{H})$ is discrete, and therefore the winding number can vary only by finite jumps[7]. Moreover, the mapping $\Omega(\phi)$ on the equator is a smooth function of the azimuthal angle ϕ and of the radius R of the sphere. Consequently, the winding number must be independent of the radius R. From this fact, two different situations may arise:

- The relevant gauge fields belong to $\mathfrak{h}$ all the way down to zero radius. In this case, the magnetic charge is independent of the radius of the sphere at all R, which means that the monopole is a point-like singularity at the origin, like the original Dirac monopole.
- There exists a short-distance core in which the gauge fields live in an algebra which is larger than $\mathfrak{h}$ (possibly the algebra $\mathfrak{g}$ before symmetry breaking). Inside this core, the above argument is no longer valid, and the magnetic charge inside the sphere may vary continuously with the radius. In this case, the monopole is an extended object whose size is the radius of the core (its magnetic charge is spread out in the core).

Alternatively, we may construct a monopole as a non-trivial classical field configuration that minimizes the energy, by starting from the behavior at infinity of the Higgs field. In order to have a finite energy, the Higgs field should go to a minimum of $V(\Phi)$ when $|x| \to \infty$. The asymptotic Higgs field is therefore a mapping from the 2-sphere S_2 to $\mathcal{M}_0 = \mathcal{G}/\mathcal{H}$, and it leads to a classification of the classical field configurations based on the homotopy group $\pi_2(\mathcal{G}/\mathcal{H})$. The correspondence between the two points of view is based on the following relationship,

$$\pi_2(\mathcal{G}/\mathcal{H}) = \pi_1(\mathcal{H})/\pi_1(\mathcal{G}) \,. \tag{10.45}$$

For a simply connected Lie group $\mathcal{G}$ (e.g., all the $SU(N)$), the first homotopy group is trivial, $\pi_1(\mathcal{G}) = \{0\}$, and we have

$$\pi_2(\mathcal{G}/\mathcal{H}) = \pi_1(\mathcal{H}) \,, \tag{10.46}$$

[7] Therefore, it must be conserved by time evolution. Indeed, time evolution is continuous, and the only way for a discrete quantity to evolve continuously is to be constant.

hence the equivalence between the two ways of classifying monopoles.

10.4 Instantons

Until now, all the extended field configurations we have encountered were time independent. After integration over time, their action is infinite, and therefore they do not contribute to path integrals. In this section, we will discuss field configurations of finite action, called *instantons*, that are localized both in space and in time. Consider a Yang-Mills theory in D-dimensional Euclidean space, whose action reads

$$S[A] \equiv \frac{1}{4}\int d^D x \; F^{ij}_a(x) F^{ij}_a(x) \; . \tag{10.47}$$

(We use latin indices $i, j, k, \cdots$ for Lorentz indices in Euclidean space.) Instantons are non-trivial (i.e. not pure gauges in the entire spacetime) gauge field configurations that realize local minima of this action.

10.4.1 Asymptotic behavior

In order to have a finite action, these fields must go to a pure gauge when $|x| \to \infty$,

$$A^i_a t^a \underset{|x|\to\infty}{\to} \frac{i}{g}\,\Omega^\dagger(\widehat{x})\,\partial^i \Omega(\widehat{x}) \; , \tag{10.48}$$

where $\Omega(\widehat{x})$ is an element of the gauge group that depends only on the orientation $\widehat{x}$. Since multiplying $\Omega(\widehat{x})$ by a constant group element Ω_0 does not change the asymptotic gauge potential, we can always arrange that $\Omega(\widehat{x}_0) = 1$ for some fixed orientation $\widehat{x}_0$. Note that a gauge potential such as (10.48), that becomes a pure gauge at large distance, must decrease at least as fast as $|x|^{-1}$. More precisely, we may write

$$A^i_a(x) t^a = \underbrace{\frac{i}{g}\,\Omega^\dagger(\widehat{x})\,\partial^i\Omega(\widehat{x})}_{|x|^{-1}} + \underbrace{a^i(x)}_{\ll |x|^{-1}} \; . \tag{10.49}$$

The field strength associated to such a field decreases faster than $|x|^{-2}$, and therefore the corresponding action is finite in $D = 4$ dimensions. There is in fact a scaling argument showing that instanton solutions can only exist in four dimensions. Given an instanton field configuration $A^i(x)$ and a scaling factor R, let us define

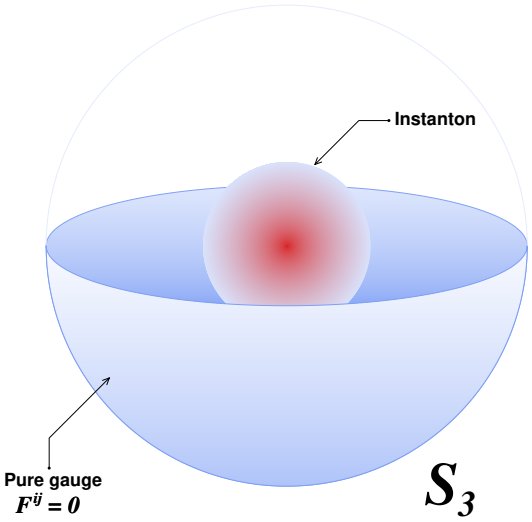
$$A^i_R(x) \equiv \frac{1}{R}\,A^i(x/R) \; . \tag{10.50}$$

Since classical Yang-Mills theory is scale invariant, the field A^i_R is also an extremum of the action (i.e. a solution of the classical Yang-Mills equations) if A^i is. The action of this rescaled field is given by

$$S[A_R] = R^{D-4}\,S[A] \; . \tag{10.51}$$

Therefore, given an instanton $A^i(x)$, we may continuously deform it into another field configuration $A^i_R(x)$ whose action is multiplied by R^{D-4}. Unless $D = 4$, this action has a higher or lower value, in contradiction with the fact that A^i was a local extremum[8]. Thus, non-trivial local extrema of the classical Euclidean Yang-Mills action can only exist in $D = 4$. In four dimensions, if A^i is an instanton, then A^i_R is also an instanton (with the same value of the action). Thus, classical instantons can exist with any size. But this degeneracy is lifted by quantum corrections, that introduce a scale into Yang-Mills theory via the running coupling.

Figure 10.8: *Cartoon of an instanton (the illustration is for $D = 3$, although instantons actually exist in $D = 4$). The sphere S_3 is in fact infinitely far away from the center of the instanton.*



10.4.2 Bogomol'nyi inequality and self-duality condition

In the study of instantons, a useful variant of Bogomol'nyi trick is to start from the following obvious inequality,

$$0 \leq \int d^4x \, (F^a_{ij} \mp \frac{1}{2}\epsilon_{ijkl}F^a_{kl})^2 \, , \tag{10.52}$$

which leads to

$$\begin{aligned} 0 \quad &\leq \quad \int d^4x \left(F^a_{ij}F^a_{ij} \mp \epsilon_{ijkl}F^a_{ij}F^a_{kl} + \frac{1}{4}\epsilon_{ijkl}\epsilon_{ijmn}F^a_{kl}F^a_{mn} \right) \\ &= \int d^4x \left(F^a_{ij}F^a_{ij} \mp \epsilon_{ijkl}F^a_{ij}F^a_{kl} + \frac{1}{2}(\delta_{km}\delta_{ln} - \delta_{kn}\delta_{lm})F^a_{kl}F^a_{mn} \right) \\ &= \int d^4x \left(2F^a_{ij}F^a_{ij} \mp \epsilon_{ijkl}F^a_{ij}F^a_{kl} \right) \, . \end{aligned} \tag{10.53}$$

[8] The only exception to this reasoning occurs if $S[A] = 0$. But this happens only in the trivial situation where A^i is a pure gauge in the entire spacetime.

By choosing appropriately the sign, this can be rearranged into a lower bound for the action:

$$S[A] \geq \frac{1}{8}\left|\epsilon_{ijkl}\int d^4x\, F^a_{ij}F^a_{kl}\right| , \tag{10.54}$$

known as *Bogomol'nyi's inequality*. Interestingly, we recognize in the right hand side an integral identical to the one that enters in the θ-term of Yang-Mills theories (see the section 4.7) or in the anomaly function (see the section 3.5 and the chapter 9). This equality becomes an equality when:

$$F^a_{ij} = \pm\frac{1}{2}\epsilon_{ijkl}F^a_{kl} . \tag{10.55}$$

A solution that obeys this condition is by construction a minimum of the Euclidean action $S[A]$, and therefore a solution of the classical Yang-Mills equations. But like in the case of domain walls, finding field configurations that fulfill this self-duality condition is somewhat simpler than solving directly the Yang-Mills equations. Thus, from now on, we will look for gauge fields that fulfill eq. (10.55) and go to a pure gauge as $|x| \to \infty$.

10.4.3 Topological classification

In $D = 4$, the functions $\Omega(\hat{x})$ that define the asymptotic behavior of instantons map the 3-sphere $\mathcal{S}_3$ into the gauge group $\mathcal{G}$,

$$\Omega : \quad \mathcal{S}_3 \mapsto \mathcal{G} , \tag{10.56}$$

with a fixed value $\Omega(\hat{x}_0) = 1$. These functions can be grouped into topological classes, such that mappings belonging to the same class can be continuously deformed into one another. The set of these classes can be endowed with a group structure, called the third homotopy group of $\mathcal{G}$ and denoted $\pi_3(\mathcal{G})$ (for any $SU(N)$ group with $N \geq 2$, we have $\pi_3(\mathcal{G}) = \mathbb{Z}$). Note that the asymptotic forms of the fields A^i and A^i_R are identical, implying that these two instantons belong to the same topological class. Since their actions are identical in four dimensions, this scaling provides a continuous family of instantons that belong to the same topological class and have the same action. This is in fact more general: we will show later that the action of an instanton depends only on the topological class of the instanton, and therefore can only vary by discrete amounts.

10.4.4 Minimal action

Let us assume that we have found a self-dual gauge field configuration, that realizes the equality in eq. (10.54). In order to calculate its action, we can use the fact that $\epsilon^{ijkl}F^1_{ij}F^a_{kl}$ is a total derivative,

$$\frac{1}{2}\,\epsilon^{ijkl}F^1_{ij}F^a_{kl} = \partial_i\Big[\underbrace{\epsilon^{ijkl}\Big(A^a_j F^a_{kl} - \frac{g}{3}\,f^{abc}\,A^a_j A^b_k A^c_l\Big)}_{K^i}\Big] . \tag{10.57}$$

(This property was derived in the section 4.7.) The vector K^i can also written as a trace of objects belonging to the fundamental representation:

$$K^i = 2\,\epsilon^{ijkl}\,\mathrm{tr}\left(A_j F_{kl} + \frac{2ig}{3}\,A_j A_k A_l\right) . \tag{10.58}$$

Since the integrand in the right hand side of eq. (10.54) is a total derivative, one may use Stokes's theorem in order to rewrite the integral as a 3-dimensional integral extended to a spherical hypersurface $\mathcal{S}_R$ of radius $R \to \infty$:

$$\mathcal{S}_{\rm min}[A] \equiv \frac{1}{8}\epsilon_{ijkl}\int d^4x\, F^a_{ij}F^a_{kl} = \lim_{R\to\infty}\frac{1}{4}\int_{\mathcal{S}_R} d^3S_i\, K^i\,. \tag{10.59}$$

Thus, the minimum of the action depends only on the behavior of the gauge field at large distance (this does not mean that the action does not depend on details of the gauge field in the interior, but more simply that the gauge fields that realize the minima are fully determined in the bulk by their asymptotic behavior). From the earlier discussion of the asymptotic behavior of instanton solutions, we know that

$$A_i(x) \underset{|x|\to\infty}{\sim} |x|^{-1}\,, \quad F_{ij}(x) \underset{|x|\to\infty}{\ll} |x|^{-2}\,. \tag{10.60}$$

Therefore, in the current K^i, the term A_jF_{kl} is negligible in front of the term $A_jA_kA_l$ at large distance, and we can also write

$$\begin{aligned}\mathcal{S}_{\rm min}[A] &= \lim_{R\to\infty}\frac{ig}{3}\int_{\mathcal{S}_R} d^3S_i\,\epsilon^{ijkl}\,{\rm tr}\left(A_jA_kA_l\right)\\ &= \lim_{R\to\infty}\frac{1}{3\,g^2}\int_{\mathcal{S}_R} d^3S_i\,\epsilon^{ijkl}\,{\rm tr}\left(\Omega^\dagger(\partial_j\Omega)\Omega^\dagger(\partial_k\Omega)\Omega^\dagger(\partial_l\Omega)\right)\,,\end{aligned} \tag{10.61}$$

where $\Omega(\widehat{x})$ is the group element that defines the asymptotic pure gauge behavior of the gauge potential in the direction $\widehat{x}$. In this expression, each derivative brings a factor R^{-1}, while the domain of integration scales as R^3. The result is therefore independent of the radius of the sphere and we can ignore the limit $R \to \infty$.

On this sphere, let us choose a system of coordinates made of three variables $(\theta_1,\theta_2,\theta_3)$, such that the volume element in $\mathcal{S}_R$ is $d\theta_1 d\theta_2 d\theta_3$. To rewrite the previous integral more explicitly in terms of these variables, it is convenient to introduce a fourth –radial– coordinate $\theta_0 \equiv |x|$. The coordinates $(\theta_0,\theta_1,\theta_2,\theta_3)$ are thus coordinates in $\mathbb{R}^4$, and $d^4x = d\theta_0 d\theta_1 d\theta_2 d\theta_3$. The volume element on the sphere $\mathcal{S}_R$ is $d\theta_1 d\theta_2 d\theta_3 = d^4x\,\delta(\theta_0 - R)$. Noting that $\widehat{x}_i = \partial\theta_0/\partial x_i$, we can write

$$d^3S_i = \widehat{x}_i\; d\theta_1 d\theta_2 d\theta_3 = \frac{\partial\theta_0}{\partial x_i}\; d\theta_1 d\theta_2 d\theta_3 = d^4x\,\delta(\theta_0 - R)\,\frac{\partial\theta_0}{\partial x_i} \tag{10.62}$$

and the minimal action becomes

$$\begin{aligned}\mathcal{S}_{\rm min}[A] &= \frac{1}{3\,g^2}\int d^4x\,\delta(\theta_0 - R)\,\epsilon^{ijkl}\,\frac{\partial\theta_0}{\partial x_i}\frac{\partial\theta_a}{\partial x_j}\frac{\partial\theta_b}{\partial x_k}\frac{\partial\theta_c}{\partial x_l}\\ &\qquad\times{\rm tr}\left\{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_a}\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_b}\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_c}\right\}\,,\end{aligned} \tag{10.63}$$

where we have rewritten the derivatives with respect to x_i in terms of derivatives with respect to θ_a (the implicit sums on a,b,c run over the indices $1,2,3$ only, because the group element Ω depends only on the orientation $\widehat{x}$). Finally, we may use:

$$\epsilon^{lijk}\frac{\partial\theta_0}{\partial x_i}\frac{\partial\theta_a}{\partial x_j}\frac{\partial\theta_b}{\partial x_k}\frac{\partial\theta_c}{\partial x_l} = \det\left(\frac{\partial(\theta_0\theta_1\theta_2\theta_3)}{\partial(x_1x_2x_3x_4)}\right)\underbrace{\epsilon_{0abc}}_{=\epsilon_{abc}}\,. \tag{10.64}$$

The determinant is nothing but the Jacobian of the coordinate transformation $\{x_i\} \to \{\theta_a\}$. Therefore, we obtain

$$\mathcal{S}_{\rm min}[A] = \frac{1}{3\,g^2}\int d\theta_1\,d\theta_2 d\theta_3\;\epsilon_{abc}\;{\rm tr}\left\{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_a}\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_b}\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_c}\right\}\,. \tag{10.65}$$

10.4.5 Cartan-Maurer invariant

Definition : In order to calculate the integral that appears in eq. (10.65), let us make a mathematical diggression. Consider a d-dimensional manifold $\mathcal{S}$, of coordinates $(\theta_1,\theta_2,\cdots,\theta_d)$, a manifold $\mathcal{M}$ that may be viewed as a matrix representation of a Lie group, and a mapping Ω from $\mathcal{S}$ to $\mathcal{M}$:

$$(\theta_1,\theta_2,\cdots,\theta_d)\in\mathcal{S}\longrightarrow\Omega(\theta_1,\theta_2,\cdots,\theta_d)\in\mathcal{M}\,. \tag{10.66}$$

The *Cartan-Maurer form* $\mathcal{F}[\Omega]$ is an integral that generalizes the one encountered earlier:

$$\mathcal{F}[\Omega]\equiv\int d\theta_1\cdots d\theta_d\;\epsilon^{i_1\cdots i_d}\;{\rm tr}\left\{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_1}}\cdots\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_d}}\right\}\,, \tag{10.67}$$

where $\epsilon^{i_1\cdots i_d}$ is the d-dimensional completely antisymmetric tensor, normalized by $\epsilon^{12\cdots d}=+1$. In d dimensions, this tensor tranforms as follows under circular permutations:

$$\epsilon^{i_1\cdots i_d}=-(-1)^d\epsilon^{i_2\cdots i_d i_1}\,. \tag{10.68}$$

Using the cyclicity of the trace, we conclude that $\mathcal{F}[\Omega]=0$ if the dimension d is even. In the following, we thus restrict the discussion to the case where d is odd[9].

Coordinate independence : Consider now another system of coordinates on $\mathcal{S}$, that we denote θ'_i. We have:

$$\epsilon^{i_1\cdots i_d}\,\frac{\partial\theta'_{j_1}}{\partial\theta_{i_1}}\cdots\frac{\partial\theta'_{j_d}}{\partial\theta_{i_d}}=\det\left(\frac{\partial(\theta'_i)}{\partial(\theta_j)}\right)\;\epsilon^{j_1\cdots j_d}\,, \tag{10.69}$$

and the determinant in the right hand side is the Jacobian of the coordinate transformation. We thus obtain

$$\mathcal{F}[\Omega]=\int d\theta'_1\cdots d\theta'_d\;\epsilon^{j_1\cdots j_d}\;{\rm tr}\left\{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta'_{j_1}}\cdots\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta'_{j_d}}\right\}\,, \tag{10.70}$$

which is identical to eq. (10.67), except for the fact that it is expressed in terms of the new coordinates θ'_i. This proves that $\mathcal{F}[\Omega]$ is independent of the choice of the coordinate system on $\mathcal{S}$, and is a property of the manifold $\mathcal{S}$ itself.

[9] This is the case in the study of instantons, since in this case the manifold $\mathcal{S}$ is the 3-sphere $\mathcal{S}_3$.

Change under a small variation of Ω : Let us now study the change of $\mathcal{F}[\Omega]$ when we vary the mapping Ω by $\delta\Omega$. Thanks to the cyclicity of the trace, the variation of each factor $\Omega^\dagger \partial g/\partial\theta_i$ gives the same contribution to the variation of $\mathcal{F}[\Omega]$. Therefore, it is sufficient to consider one of these variations, and to multiply its contribution by the number of factors, d:

$$\delta\mathcal{F}[\Omega] = d\int d\theta_1\cdots d\theta_d\ \epsilon^{i_1\cdots i_d}\ \mathrm{tr}\left\{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_1}}\cdots\delta\left(\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_d}}\right)\right\}\ . \quad (10.71)$$

The variation of the last factor inside the trace can be written as

$$\begin{aligned}\delta\left(\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_d}}\right) &= -\Omega^\dagger(\theta)\delta\Omega(\theta)\underbrace{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_d}}}_{-\frac{\partial\Omega^\dagger(\theta)}{\partial\theta_{i_d}}\Omega(\theta)} + \Omega^\dagger(\theta)\frac{\partial\delta\Omega(\theta)}{\partial\theta_{i_d}} \\ &= \Omega^\dagger(\theta)\frac{\partial\delta\Omega(\theta)\Omega^\dagger(\theta)}{\partial\theta_{i_d}}\Omega(\theta)\ . \end{aligned} \quad (10.72)$$

Then, integrating by parts with respect to θ_{i_d}, we obtain:

$$\begin{aligned}\delta\mathcal{F}[\Omega] = -d\int d\theta_1\cdots d\theta_d\ \epsilon^{i_1\cdots i_d} \\ \times\ \mathrm{tr}\left\{\frac{\partial}{\partial\theta_{i_d}}\left(\frac{\partial\Omega(\theta)}{\partial\theta_{i_1}}\Omega^\dagger(\theta)\cdots\frac{\partial\Omega(\theta)}{\partial\theta_{i_{d-1}}}\Omega^\dagger(\theta)\right)\delta\Omega(\theta)\Omega^\dagger(\theta)\right\}\ . \end{aligned} \quad (10.73)$$

All the terms containing a factor $\partial^2\Omega/\partial\theta_{i_d}\partial\theta_{i_a}$ vanish because the second derivative is symmetric under the exchange of of the indices i_d and i_a, while the prefactor $\epsilon^{i_1\cdots i_d}$ is antisymmetric. The remaining terms are those where the derivative with respect to θ_d act on one of the factors $\Omega^\dagger$. There are $d-1$ such terms, all identical up to a circular permutation of the indices $i_1,\cdots,i_{d-1}$. Using eq. (10.68) with $d-1$ instead of d, and the fact that $d-1$ is even, we see that these terms have alternating signs. Since there is an even number of them $(d-1)$, their sum is zero,

$$\delta\mathcal{F}[\Omega] = 0\ . \quad (10.74)$$

Therefore, $\mathcal{F}[\Omega]$ is invariant under small changes of Ω, which implies that $\mathcal{F}[\Omega]$ can only vary by discrete jumps. In particular, when $\mathcal{S}$ is the d-sphere $\mathcal{S}_d$, $\mathcal{F}[\Omega]$ depends only on the homotopy class of Ω. These classes form a group $\pi_d(\mathcal{M})$. Moreover, $\mathcal{F}[\Omega]$ provides a representation of $\pi_d(\mathcal{M})$: if $\overline{\Omega}$ denotes the homotopy class to which Ω belongs, we have

$$\mathcal{F}[\overline{\Omega}_1\times\overline{\Omega}_2] = \mathcal{F}[\overline{\Omega}_1] + \mathcal{F}[\overline{\Omega}_2]\ . \quad (10.75)$$

(We denote by $\times$ the group composition in $\pi_d(\mathcal{M})$.) As a consequence, if there exists an Ω for which the Cartan-Maurer invariant is nonzero, then all its integer multiples can also be obtained, thereby proving that the homotopy group $\pi_d(\mathcal{M})$ contains $\mathbb{Z}$.

Case of a Lie group target manifold : Let us now specialize to the case where the target manifold $\mathcal{M}$ is a d-dimensional Lie group $\mathcal{H}$, and exploit its group structure in order to obtain simpler expressions. In this case, the θ_a's can also be used as coordinates on $\mathcal{H}$. Consider two elements Ω_1 and Ω_2 of $\mathcal{H}$, represented respectively by the coordinates θ_a and ϕ_a. Their product $\Omega_2\Omega_1$ is an element of $\mathcal{H}$ of coordinates $\psi(\theta,\phi)$ (the group multiplication determines how ψ depends on θ and ϕ). Since we have shown that the choice of cordinates on $\mathcal{S}$ is irrelevant, we may choose them in such a way that the function $\Omega(\theta)$ is a representation of the group $\mathcal{H}$, i.e.

$$\Omega(\phi)\Omega(\theta) = \Omega(\psi(\theta,\phi)) \ . \tag{10.76}$$

By differentiating this equality with respect to ψ_j at fixed ϕ, we obtain

$$\Omega(\phi)\,\frac{\partial\Omega(\theta)}{\partial\theta_i}\,\frac{\partial\theta_i}{\partial\psi_j} = \frac{\partial\Omega(\psi)}{\partial\psi_j} \ , \tag{10.77}$$

and after left multiplication by $\Omega^\dagger(\psi)$, this leads to

$$\Omega^\dagger(\theta)\,\frac{\partial\Omega(\theta)}{\partial\theta_i} = \frac{\partial\psi_j}{\partial\theta_i}\;\Omega^\dagger(\psi)\,\frac{\partial\Omega(\psi)}{\partial\psi_j} \ . \tag{10.78}$$

Using (10.69), the integrand of $\mathcal{F}[\Omega]$ at the point θ can be expressed as

$$\begin{aligned}\epsilon^{i_1\cdots i_d}\ \mathrm{tr}\left\{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_1}}\cdots\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_{i_d}}\right\}\\ = \det\left(\frac{\partial(\psi)}{\partial(\theta)}\right)\epsilon^{j_1\cdots j_d}\ \mathrm{tr}\left\{\Omega^\dagger(\psi)\frac{\partial\Omega(\psi)}{\partial\psi_{j_1}}\cdots\Omega^\dagger(\psi)\frac{\partial\Omega(\psi)}{\partial\psi_{j_d}}\right\} \ ,\end{aligned} \tag{10.79}$$

where ψ can be any fixed reference point in the group. In the right side, the integration variable θ now appears only inside the determinant.

The Lie group $\mathcal{H}$ being a smooth manifold, it can be endowed with a metric tensor $\gamma_{ij}(\theta)$, that transforms as follows in a change of coordinates

$$\gamma_{ij}(\psi) = \frac{\partial\theta_k}{\partial\psi_i}\,\frac{\partial\theta_l}{\partial\psi_j}\gamma_{kl}(\theta) \ . \tag{10.80}$$

Given a mapping $\Omega(\theta)$ between coordinates and group elements, a possible choice for the metric is given by[10]

$$\gamma_{ij}(\theta) = -\frac{1}{2}\ \mathrm{tr}\left\{\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_i}\Omega^\dagger(\theta)\frac{\partial\Omega(\theta)}{\partial\theta_j}\right\} \ . \tag{10.81}$$

Moreover, for any such metric $\gamma_{ij}(\theta)$, we have:

$$\det\left(\frac{\partial(\psi)}{\partial(\theta)}\right) = \sqrt{\frac{\det\gamma(\theta)}{\det\gamma(\psi)}} \ . \tag{10.82}$$

[10] In the algebra of a compact Lie group, the Killing form $\mathbf{K}(X,Y) \equiv \mathrm{tr}\left(\mathrm{ad}_X\,\mathrm{ad}_Y\right)$ is a negative definite inner product, from which one can define a distance on the group manifold in the vicinity of the origin. Eq. (10.81) extends this definition globally to the entire group, in a way which is invariant under left and right group action.

Therefore, the Cartan-Maurer invariant $\mathcal{F}[\Omega]$ takes the following form

$$\mathcal{F}[\Omega] = \epsilon^{j_1 \cdots j_d} \, \mathrm{tr} \left\{ \Omega^\dagger(\psi) \frac{\partial \Omega(\psi)}{\partial \psi_{j_1}} \cdots \Omega^\dagger(\psi) \frac{\partial \Omega(\psi)}{\partial \psi_{j_d}} \right\} \frac{1}{\sqrt{\det \gamma(\psi)}} \int \mathrm{d}^d\theta \, \sqrt{\det \gamma(\theta)} \, , \tag{10.83}$$

in which all the terms that do not depend on θ have been factored out in front of the integral. In fact, $\mathrm{d}^d\theta \, \sqrt{\det \gamma(\theta)}$ is an invariant measure on the Lie group, and the integral is therefore the volume of the group. In other words, the previous formula exploits the group invariance in order to rewrite the Cartan-Maurer invariant as the product of the integrand evaluated at a fixed point by the volume of the group. Since ψ is arbitray in this expression, we may choose the value ψ_0 that corresponds to the group identity. Furthermore, groups elements in the vicinity of the identity may be written as

$$\Omega(\psi) \underset{\psi \to \psi_0}{\approx} 1 + 2i \, (\psi - \psi_0)_a \, t^a \, , \tag{10.84}$$

where the t^a's are the generators of the Lie algebra $\mathfrak{h}$. Then, the derivatives read simply

$$\left. \frac{\partial \Omega(\psi)}{\partial \psi_a} \right|_{\psi_0} = 2i \, t^a \, . \tag{10.85}$$

From this, we obtain the following compact expression for $\mathcal{F}[\Omega]$:

$$\mathcal{F}[\Omega] = (2i)^d \, \epsilon^{i_1 \cdots i_d} \, \mathrm{tr} \left\{ t^{i_1} \cdots t^{i_d} \right\} \frac{1}{\sqrt{\det \gamma(0)}} \int \mathrm{d}^d\theta \, \sqrt{\det \gamma(\theta)} \, . \tag{10.86}$$

Cartan-Maurer invariant for $\mathcal{H} = \mathrm{SU}(2)$: Consider the following mapping from the 3-sphere $\mathcal{S}_3$ to $\mathrm{SU}(2)$:

$$\Omega(\theta) = \begin{pmatrix} \theta_4 + i\theta_3 & \theta_2 + i\theta_1 \\ -\theta_2 + i\theta_1 & \theta_4 - i\theta_3 \end{pmatrix} = \theta_4 + 2i \, \theta_a t^a \, , \tag{10.87}$$

with $t^{1,2,3}$ the generators of the $\mathfrak{su}(2)$ algebra (for the fundamental representation, the Pauli matrices divided by 2) and $\theta_1^2 + \theta_2^2 + \theta_3^2 + \theta_4^2 = 1$. The following identities hold:

$$\begin{aligned} &\det \Omega(\theta) = 1 \, , \quad \Omega^\dagger(\theta) = \theta_4 - 2i \, \theta_a t^a \, , \\ &\forall i \in \{1,2,3\}, \quad \frac{\partial \Omega(\theta)}{\partial \theta_i} = 2i \, t^i - \frac{\theta_i}{\sqrt{1 - \theta^2}} \, . \end{aligned} \tag{10.88}$$

(We denote $\theta^2 \equiv \theta_1^2 + \theta_2^2 + \theta_3^2$.) In the evaluation of eq. (10.81), we need traces of products of up to four t^a matrices. In the fundamental representation, they can all be obtained from

$$\begin{aligned} &\mathrm{tr} \, (t^i) = 0 \, , \\ &t^i t^j = \frac{i}{2} \epsilon^{ijk} t^k + \frac{1}{4} \delta^{ij} \, , \end{aligned} \tag{10.89}$$

which leads to

$$\begin{aligned}
&\mathrm{tr}\,(t^i t^j) = \frac{1}{2}\delta^{ij}\ ,\\
&\mathrm{tr}\,(t^i t^j t^k) = \frac{i}{4}\epsilon^{ijk}\ ,\\
&\mathrm{tr}\,(t^i t^j t^k t^l) = \frac{1}{8}(\delta^{ij}\delta^{kl} + \delta^{il}\delta^{jk} - \delta^{ik}\delta^{jl})\ .
\end{aligned} \tag{10.90}$$

Then, the metric tensor of eq. (10.81) reads

$$\gamma_{ij}(\theta) = \delta_{ij} + \frac{\theta_i\theta_j}{1-\theta^2}\ , \tag{10.91}$$

and its determinant is

$$\det\gamma(\theta) = \frac{1}{1-\theta^2}\ . \tag{10.92}$$

Combining the above results, we obtain the following expression for the Cartan-Maurer invariant of the homotopy class of Ω in $\pi_3(SU(2))$

$$\mathcal{F}[\Omega] = (2i)^3\,\epsilon^{abc}\,\mathrm{tr}\,(t^a t^b t^c)\int\frac{2\,d^3\theta}{\sqrt{1-\theta^2}}\ . \tag{10.93}$$

The factor 2 comes from the fact that there are two allowed values of θ_4 for each $\theta_{1,2,3}$. Finally, we have

$$\mathcal{F}[\Omega] = 96\pi\int_0^1\frac{d\theta\,\theta^2}{\sqrt{1-\theta^2}} = 24\pi^2\ . \tag{10.94}$$

In fact, the mapping of eq. (10.87) wraps only once in $SU(2)$, and the above result therefore corresponds to the topological index $+1$. Since $24\pi^2$ is non-zero, there are other classes of Ω's whose Cartan-Maurer invariants are the integer multiples of this result, and the second homotopy group is $\pi_3(SU(2)) = \mathbb{Z}$. Note also that this result extends to any Lie group that contains an $SU(2)$ subgroup.

10.4.6 Explicit instanton solution

In a gauge theorie whose gauge group contains an $SU(2)$ subgroup, the mapping of eq. (10.87)) can be used to construct the asymptotic form of an instanton of topological index $+1$,

$$A_i(x) \underset{|x|\to\infty}{=} \frac{i}{g}\,\Omega^\dagger(\hat{x})\partial_i\Omega(\hat{x})\ , \tag{10.95}$$

with $\Omega(\hat{x}) \equiv \hat{x}_4 + 2i\,\hat{x}_i t^i$. One may then prove that the self-dual field configuration in the bulk that has this large distance behavior is given by

$$A_i(x) = \frac{i}{g}\,\frac{r^2}{r^2+R^2}\,\Omega^\dagger(\hat{x})\partial_i\Omega(\hat{x})\ , \tag{10.96}$$

with an arbitrary radius R. From the result (10.94) of the previous subsection, we find that the minimum of the action that corresponds to this solution is:

$$S_{\rm min}[A] = \frac{8\pi^2}{g^2} \ . \tag{10.97}$$

Up to translations, dilatations or gauge transformations, this is the only field configuration that gives this action. The field strength corresponding to eq. (10.96) is localized in Euclidean spacetime, with a size of order R. One may also superimpose several such solutions. Provided that their centers are separated by distances much larger than R, this sum is also a solution of the classical equations of motion, and its action is a multiple of $8\pi^2/g^2$.

10.4.7 Instantons and the θ-term in Yang-Mills theory

Since we have uncovered classical field configurations of non-zero topological index with finite action, a legitimate question is their role in an Euclidean path integral, since functional integration a priori sums over all classical field configurations. For more generality, we may assume that in the path integral the fields of topological index n are weighted with a factor $P(n)$ that may vary with n (this generalization would allow for instance to exclude fields of topological index different from zero). Thus, the expectation value of an observable $\mathcal{O}$ may be written as

$$\langle \mathcal{O} \rangle = Z^{-1} \sum_{n\in\mathbb{Z}} P(n) \int [DA]_n \ \mathcal{O}[A] \ e^{-S[A]} \ , \tag{10.98}$$

where $[DA]_n$ is the functional measure restricted to gauge fields of topological index n. The normalization factor Z is given by the same path integral without the observable.

The dependence of $P(n)$ on the topological index cannot be arbitrary. In order to see this, let us consider two spacetime subvolumes Ω_1 and Ω_2, non overlapping and such that $\Omega_1 \cup \Omega_2 = \mathbb{R}^4$. Assume further that the support of the observable $\mathcal{O}$ is entirely inside Ω_1. The topological number, that may be obtained as the integral over spacetime of $\epsilon_{ijkl} F^a_{ij} F^a_{kl}$, is additive[11] and we may define topological numbers n_1 and n_2 for Ω_1 and Ω_2, respectively. The total topological number n is given by $n = n_1 + n_2$. In the expectation value of eq. (10.98), we can therefore split the integration into the domains Ω_1 and Ω_2 as follows

$$\langle \mathcal{O} \rangle = Z^{-1} \sum_{n_1,n_2\in\mathbb{Z}} P(n_1+n_2) \int [DA]_{n_1} \ \mathcal{O}[A] \ e^{-S_{\Omega_1}[A]} \int [DA]_{n_2} \ e^{-S_{\Omega_2}[A]} \ , \tag{10.99}$$

where $[DA]_{n_i}$ is the functional measure for gauge fields with topological number n_i in the domain Ω_i. Since the observable is localized inside the domain Ω_1, we should be able to remove any dependence on the domain Ω_2 from its expectation value. This dependence cancels between the numerator and the factor Z^{-1} in the previous expression provided that the weight $P(n_1+n_2)$ factorizes as follows:

$$P(n_1+n_2) = P(n_1)P(n_2) \ , \tag{10.100}$$

[11] But note that this integral does not have to be an integer when the integration domain is not the entire spacetime. However, it is approximately an integer when the size of the domain is much larger than the instanton size.

which implies that

$$P(n) = e^{-n\theta} \ , \tag{10.101}$$

where θ is an arbitrary constant. From the previous results, the topological number of a field configuration is given by the integral

$$n = \frac{g^2}{64\pi^2} \int d^4x \ \epsilon^{ijkl} \, F^a_{ij} F^a_{kl} \ . \tag{10.102}$$

Therefore, we may capture the effect of the topological weight $P(n)$ by adding to the Lagrangian density the following term

$$\mathcal{L}_\theta \equiv \frac{\theta \, g^2}{64\pi^2} \ \epsilon^{ijkl} \, F^a_{ij} F^a_{kl} \ . \tag{10.103}$$

After this term has been added, it is no longer necessary to split the path integral into separate topological sectors. The previous Lagrangian is nothing but the θ-term that we have already encountered in the discussion of non-Abelian gauge theories. There, it appeared as a term that cannot be excluded on the grounds of gauge symmetry. In the present discussion, we see that the θ-term results from a non-uniform weighting of the field configurations of different topological index ($\theta = 0$ corresponds to a path integration where all the fields are weighted equally, regardless of their topological index).

10.4.8 Quantum fluctuations around an instanton

Consider an instanton solution $A^\mu_{n,\alpha}(x)$, that provides a local minimum of the Euclidean action, where the subscript n is the topological index of the instanton, and α collectivey denotes all the other parameters that characterize the instanton (its center, its size, its orientation in color space). The expectation value of an observable reads

$$\langle \mathcal{O} \rangle = Z^{-1} \int [DA] \ e^{-S[A]} \ \mathcal{O}(A) = Z^{-1} \int [Da] \ e^{-S[A_{n,\alpha}+a]} \ \mathcal{O}(A_{n,\alpha} + a) \ , \tag{10.104}$$

where we denote a^μ the difference $A^\mu - A^\mu_{n,\alpha}$. Since the instanton is an extremum of the action, the dependence of the action on a^μ begins with quadratic terms:

$$S[A_{n,\alpha} + a] = \frac{8\pi^2 \, |n|}{g^2} + \frac{1}{2} \int d^4x d^4y \ G^{-1}_{n\alpha,m\beta}(x,y) \, a(x) a(y) + \cdots \tag{10.105}$$

It is important to note that the action has *flat directions* in the space of field configurations, that correspond to changing the parameters of the instanton inside its topological class. For instance, changing the center coordinates of the instanton does not modify the value of its action. Along these directions, the second derivative of the action vanishes. This means that the matrix of second-order coefficient $G^{-1}_{n\alpha,m\beta}(x,y)$ has a number of vanishing eigenvalues, corresponding to these flat directions.

If we expand the action only to quadratic order in a^μ, which amounts to a one-loop approximation in the background of the instanton, a typical contribution to the expectation value of

eq. (10.104) is a product of "dressed propagators" $\mathcal{G}_{n\alpha,m\beta}(x,y)$ connecting pairwise the gauge fields contained in the observable $\mathcal{O}$ and a determinant:

$$\big\langle \mathcal{O} \big\rangle = Z^{-1}\ e^{-8\pi^2|n|/g^2} \left\{ \Big(\det \mathcal{G} \Big)^{1/2} \prod \mathcal{G} + \cdots \right\} . \tag{10.106}$$

Our goal here is simply to extract the dependence of such an expectation value on the topological index n. Besides the obvious exponential prefactor, a dependence on n hides in the determinant. Le us rewrite it as a product on the spectrum of $\mathcal{G}^{-1}$

$$\Big(\det \mathcal{G} \Big)^{1/2} = \prod_s \lambda_s^{-1/2} , \tag{10.107}$$

where the λ_s are the eigenvalues of $\mathcal{G}^{-1}$. If we rescale the gauge fields by a power of the coupling g, $g\,A \to A$, the only dependence on g in the Yang-Mills action is a prefactor g^{-2}, and all the eigenvalues λ_s are also proportional to g^{-2}. Moreover, as explained above, we should remove the zero modes from this product, since they do not give a quadratic term in eq. (10.105). If we are interested only in the powers of g, we may write

$$\Big(\det \mathcal{G} \Big)^{1/2} \sim \prod_{\text{all modes}} g \prod_{\text{zero modes}} g^{-1} . \tag{10.108}$$

The first factor, that involves a (continuous) infinity of modes, is not well defined but it does not depend on the details of the instanton background. In contrast, the second factor brings one factor of g^{-1} for each collective coordinate of the instanton. For an instanton of topological number $n = 1$, these collective coordinates are:

- the 4 coordinates of the center of the instanton,
- the size R of the instanton,
- 3 angles that determine the orientation of the instanton,
- for $SU(2)$, 3 parameters defining a global gauge rotation.

Of the last 6 parameters, 3 correspond to simultaneous spatial and color rotations that produce the same instanton solution, and they should not be counted. There are therefore 8 collective coordinates for the $n = 1$ $SU(2)$ instanton[12], and its contribution to expectation values scales as

$$\big\langle \mathcal{O} \big\rangle_{n=1} \sim e^{-8\pi^2/g^2}\ g^{-8} . \tag{10.109}$$

Because of the exponential factor that contains the inverse coupling, all the Taylor coefficients of this function are vanishing at $g = 0$. Thus, such a contribution never shows up in perturbation theory.

[12] This counting is more involved for an $SU(3)$ instanton. In this case, there are 7 collective coordinates corresponding to rotations and gauge transformations, hence a total of 12 collective coordinates.

Chapter 11

Modern tools for amplitudes

11.1 Shortcomings of the usual approach

Transition amplitudes play a central role in quantum field theory, since they are the building blocks of most observables. Their square gives transition probabilities, that enter in measurable cross-sections. Until now, we have exposed the traditional way of calculating these amplitudes. Starting from a classical action that encapsulates the bare couplings of a given quantum field theory, one can derive Feynman rules for propagators and vertices (listed in the figure 5.2 for Yang-Mills theory in covariant gauges), whose application provides a straightforward algorithm for the evaluation of amplitudes. However, the use of these Feynman rules is very cumbersome for the following reasons:

- Even at tree level, the number of distinct graphs contributing to a given amplitude increases very rapidly with the number of external lines. For instance, for amplitudes with only gluons, we have

# of gluons	# of diagrams	$n!$
4	4	24
5	25	120
6	220	720
7	2,485	5,040
8	34,300	40,320
9	559,405	362,880
10	10,525,900	3,628,800

  (The third column indicates the values of $n!$, for comparison. We see that the number of graphs grows faster than the factorial of the number of external gluons.)

- The internal gluon propagators of these diagrams carry unphysical degrees of freedom, which contributes to the great complexity of each individual diagram.

- The Feynman rules are sufficiently general to compute amplitudes with arbitrary external momenta (not necessarily on-shell) and polarizations (not necessarily physical), although this is not useful for amplitudes that will be used in cross-sections. One may hope for a leaner formalism, that only calculates what is strictly necessary for physical quantities.

The situation becomes even worse with loop diagrams. Another situation with an even higher degree of complexity, even at tree-level, is that of gravity. It would be desirable to be able to calculate tree-level amplitudes with gravitons, since they enter for instance in the study of the scattering of gravitational waves by a distribution of masses, but because the graviton has spin 2, the corresponding Feynman rules are considerably more complicated (especially the self-couplings of the graviton) than those of Yang-Mills theory.

It turns out that physical on-shell amplitudes in gauge theories are considerably simpler than on may expect from the Feynman rules and the intermediate steps of their calculation by the usual perturbation theory, and a legitimate query is whether there is a more direct route to reach this final answer. The goal of this chapter is to give a glimpse (in particular, our discussion will be restricted to tree-level amplitudes, but a significant part of the many recent developments deal with loop corrections) of some of the recent developments that led to powerful new methods for calculating amplitudes. A recurring theme of these methods is to avoid as much as possible references to the Lagrangian, which may be viewed as the main source of the complications in standard perturbation theory (for instance, the gauge invariance of the Lagrangian is the reason why non-physical gluon polarizations appear in the Feynman rules). Instead, these methods try to gather as much information as possible on amplitudes based on symmetries and kinematics.

11.2 Color ordering of gluonic amplitudes

Let us firstly focus on the color structure of tree Feynman diagrams, in order to organize and simplify it. Although the techniques we expose here can be extended to quarks, we consider tree amplitudes that contain only gluons for simplicity, in the case of the $SU(N)$ gauge group. The structure constants f^{abc} of the group appear in the three-gluon and four-gluon vertices. The first step is to rewrite the structure constants in terms of the generators t_f^a of the fundamental representation of $\mathfrak{su}(N)$. Using the following relations among the generators,

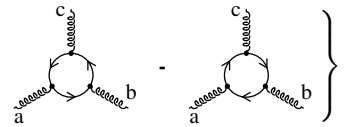
$$\left[t_f^a, t_f^b\right] = i\, f^{abc}\, t_f^c \quad , \qquad \mathrm{tr}\,(t_f^a t_f^b) = \frac{\delta^{ab}}{2}\,, \tag{11.1}$$

we can write

$$i\, f^{abc} = 2\,\mathrm{tr}\,(t_f^a t_f^b t_f^c) - 2\,\mathrm{tr}\,(t_f^b t_f^a t_f^c)\,, \tag{11.2}$$

which has also the following diagrammatic representation

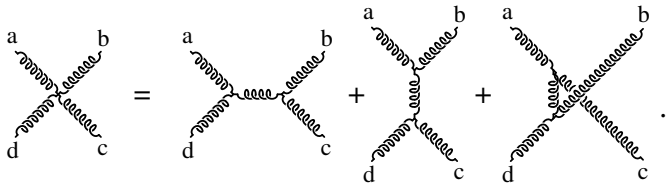
$$i\, f^{abc} = 2 \left\{ \quad \right. \tag{11.3}$$



$$\left. \right\} .$$

The black dots indicate the fundamental representation generators t_f^a. Note that the "loops" in this representation are not actual fermion loops, they are just a graphical cue indicating how

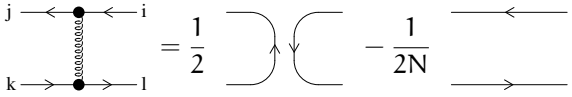
the indices carried by the t_f^a's are contracted in the traces. We may also apply this trick to the 4-gluon vertex, which from the point of view of its color structure (but not for what concerns its momentum dependence) is equivalent to a sum of three terms with two 3-gluon vertices,



(11.4)

Since the gluon propagators are diagonal in color (i.e proportional to a δ_{ab}), the t_f^a that are attached to the endpoints of the internal gluon propagators have their color indices contracted and summed over. The result of this contraction is given by the following $\mathfrak{su}(N)$ Fierz identity:

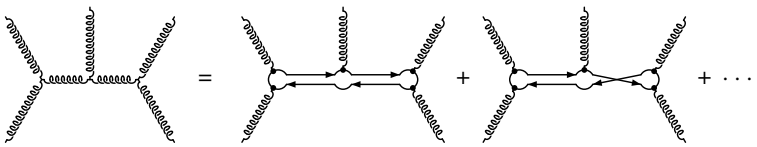
$$(t_f^a)_{ij}(t_f^a)_{kl} =$$



(11.5)

Thus, it seems that these contractions produce 2^n terms for n internal gluon propagators, but this can in fact be simplified tremendously by noticing that the second term of the Fierz identity corresponds to the exchange of a colorless object[1], that does not couple to gluons. All these terms in $1/N$ must therefore cancel in purely gluonic amplitudes (this is not true anymore if quarks are involved, either as external lines of via loop corrections).

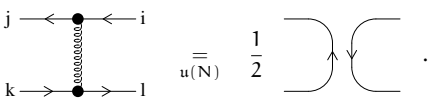
We illustrate in the following equation a few of the color structures generated by this procedure in the case of a tree-level five-gluon diagram:



(11.6)

Each of the terms contains a *single trace* of five t_f^a, one for each external gluon (the color matrices attached to the internal gluon lines have all disappeared when using the Fierz identity). The terms in the right hand side correspond to the various ways of choosing the clockwise or counterclockwise loop for each f^{abc} (see eq. (11.3)). "Twists" such as the one appearing in the second term of the previous equation arise when two such adjacent loops have opposite orientations.

[1] A more rigorous justification is to note that $SU(N) \times U(1) = U(N)$, where $U(N)$ is the group of the $N \times N$ unitary matrices. For the fundamental generators of the $u(N)$ algebra, the Fierz identity is



The $U(N)$ gauge theory differs from the $SU(N)$ one by the extra $U(1)$, and the comparison of their Fierz identities indicates that the term in $1/2N$ in eq. (11.5) is due to this $U(1)$ factor. Being Abelian, this extra factor corresponds to a photon-like mode that does not couple to gluons.

Quite generally, any n-gluon tree amplitude $\mathcal{M}_n(1\cdots n)$ can be decomposed as a sum of terms corresponding to the allowed color structures. These color structures are single traces of fundamental representation color matrices carrying the color indices of the external gluons. A priori, these matrices could be reshuffled by an arbitrary permutation in $\mathfrak{S}_n$, but thanks to the cyclic invariance of the trace we can reduce the sum to the quotient set $\mathfrak{S}_n/\mathbb{Z}_n$ of permutations modulo a cyclic permutation[2]:

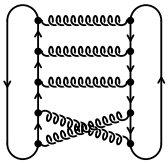
$$\mathcal{M}_n(1\cdots n) \equiv 2 \sum_{\sigma\in\mathfrak{S}_n/\mathbb{Z}_n} \mathrm{tr}\left(t_f^{a_{\sigma(1)}}\cdots t_f^{a_{\sigma(n)}}\right) \mathcal{A}_n(\sigma(1)\cdots\sigma(n)) \,, \tag{11.7}$$

where the prefactor 2 combines the factors 2 from eq. (11.2) and the factors $\frac{1}{2}$ from the first term of the Fierz identity (11.5). The object $\mathcal{A}_n(\sigma(1)\cdots\sigma(n))$ is called a *color-ordered partial amplitude*. By construction, it depends only on the momenta and polarizations of the external gluons, but not on their colors since they have already been factored out in the trace. Therefore, the partial amplitudes are gauge invariant. From eq. (11.7), the squared amplitude summed over all colors can be written as

$$\begin{aligned}\sum_{\text{colors}} \left|\mathcal{M}_n(1\cdots n)\right|^2 \;=\; 4 \sum_{\sigma,\rho\in\mathfrak{S}_n/\mathbb{Z}_n} \sum_{\text{colors}} \mathrm{tr}\left(t_f^{a_{\sigma(1)}}\cdots t_f^{a_{\sigma(n)}}\right) \mathrm{tr}^*\left(t_f^{a_{\rho(1)}}\cdots t_f^{a_{\rho(n)}}\right) \\ \times \mathcal{A}_n(\sigma(1)\cdots\sigma(n))\,\mathcal{A}_n^*(\rho(1)\cdots\rho(n)) \,.\end{aligned} \tag{11.8}$$

The sum over colors of the product of two traces that appears in the first line can be performed using the $\mathfrak{su}(N)$ Fierz identity (11.5). For instance

$$\mathrm{tr}\left(t^a t^b t^c t^d t^e\right) \mathrm{tr}^*\left(t^b t^a t^c t^d t^e\right) = \tag{11.9}$$



which can be then expressed as a function of N by repeated use of the Fierz identity.

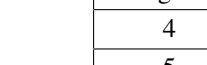
At this point, we have isolated the color dependence of the amplitude, from its momentum and polarization dependences that are factorized into the partial amplitudes. Of course, calculating the latter is still not easy, but the task is significantly reduced for two reasons:

- The color-ordered partial amplitudes only receive contributions from planar graphs where the gluons are cyclic-ordered, whose number grows much slower than the total number of graphs:

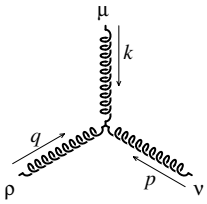
[2] This is equivalent to considering permutations that have the fixed point $\sigma(1) = 1$, i.e. permutations that only reshuffle the set $\{2\cdots n\}$. For n external gluons, there are $(n-1)!$ independent color structures. The basis provided by these traces is over-complete, and there exist linear relationships among the tree-level partial amplitudes, known as the *Kleiss-Kuijf relations*. These relations reduce the number of partial amplitudes from $(n-1)!$ to $(n-2)!$. Additional relationships known as the *Bern-Carrasco-Johansson relations* further reduce this number to $(n-3)!$,

# of gluons	# of graphs	# of cyclic-ordered graphs
4	4	3
5	25	10
6	220	38
7	2,485	154
8	34,300	654
9	559,405	2,871
10	10,525,900	12,925

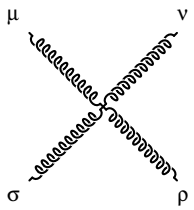
- The Feynman rules for calculating the cyclic color-ordered amplitudes are much simpler than the original Yang-Mills Feynman rules, because the vertices are stripped of all their color factors. Now, we have:



$$= \frac{-i\,g^{\mu\nu}}{p^2+i0^+} + \frac{i}{p^2+i0^+}\left(1-\frac{1}{\xi}\right)\frac{p^\mu p^\nu}{p^2} \tag{11.10}$$



$$= \begin{array}{l} g\left\{g^{\mu\nu}\,(k-p)^\rho \right. \\ \quad \left. + g^{\nu\rho}\,(p-q)^\mu + g^{\rho\mu}\,(q-k)^\nu\right\} \end{array} \tag{11.11}$$

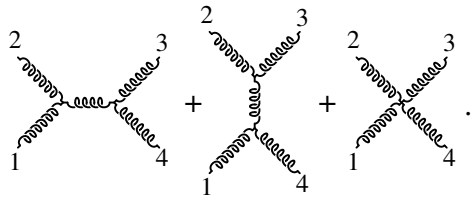


$$= -i\,g^2\,(2\,g^{\mu\rho}g^{\nu\sigma} - g^{\mu\sigma}g^{\nu\rho} - g^{\mu\nu}g^{\rho\sigma}) \tag{11.12}$$

  In the case of the 4-gluon vertex, we have included only the terms that correspond to the cyclic ordering $\mu\nu\rho\sigma$ (note that it is invariant under cyclic permutations, i.e. the Feynman rule is the same for the vertices $\nu\rho\sigma\mu$, $\rho\sigma\mu\nu$ and $\sigma\mu\nu\rho$). We can already see a considerable simplification of the Feynman rules, since all the color factors have disappeared, and the Lorentz structure of the 4-gluon vertex is also much simpler than in the original Feynman rules.

Even after having isolated the color structure, the remaining color-ordered amplitudes are still complicated. As an illustration of the color-ordered Feynman rules, let us consider the partial amplitude $\mathcal{A}_4(1,2,3,4)$ that contributes to one of the color structures in the $gg \to gg$

amplitude. Because of color ordering, only three graphs contribute to this partial amplitude:

$$\mathcal{A}_4(1,2,3,4) = $$  $$.$$ (11.13)

For definiteness, let us assume that the external momenta $p_1 \cdots p_4$ are defined as incoming, and denote $\epsilon_1 \cdots \epsilon_4$ the four polarization vectors. Using the rules listed in eqs. (11.10-11.12), we obtain:

$$\begin{aligned}
&\mathcal{A}_4(1,2,3,4) = \\
&= \frac{-i\,g^2}{(p_1+p_2)^2}\Big[(2p_2+p_1)\cdot\epsilon_1\epsilon_2^\lambda - (2p_1+p_2)\cdot\epsilon_2\epsilon_1^\lambda + \epsilon_1\cdot\epsilon_2(p_1-p_2)^\lambda\Big] \\
&\qquad\times\Big[(p_3+2p_4)\cdot\epsilon_3\epsilon_{4\lambda} - (2p_3+p_4)\cdot\epsilon_4\epsilon_{3\lambda} + \epsilon_3\cdot\epsilon_4(p_3-p_4)_\lambda\Big] \\
&+\frac{-i\,g^2}{(p_2+p_2)^2}\Big[(p_2+2p_3)\cdot\epsilon_2\epsilon_3^\lambda - (2p_2+p_3)\cdot\epsilon_3\epsilon_2^\lambda + \epsilon_2\cdot\epsilon_3(p_2-p_3)^\lambda\Big] \\
&\qquad\times\Big[(2p_1+p_4)\cdot\epsilon_4\epsilon_{1\lambda} - (p_1+2p_4)\cdot\epsilon_1\epsilon_{4\lambda} + \epsilon_1\cdot\epsilon_4(p_4-p_1)_\lambda\Big] \\
&-i\,g^2\Big[2\epsilon_1\cdot\epsilon_3\epsilon_2\cdot\epsilon_4 - \epsilon_1\cdot\epsilon_4\epsilon_2\cdot\epsilon_3 - \epsilon_1\cdot\epsilon_2\epsilon_3\cdot\epsilon_4\Big]\,.
\end{aligned} \tag{11.14}$$

Although this is considerably simpler than the full 4-gluon amplitude, it remains quite difficult to extract physical results from such an expression.

11.3 Spinor-helicity formalism

11.3.1 Motivation

Part of the complexity of eq. (11.14) lies in the fact that this formula still contains a large amount of redundant and unnecessary information, since each polarization may be shifted by a 4-vector proportional to the momentum of the corresponding external gluon, thanks to gauge invariance. For instance, the transformation

$$\epsilon_1^\mu \to \epsilon_1^\mu + \kappa\, p_1^\mu\,, \tag{11.15}$$

leaves the amplitude unchanged. However, it is not clear how to optimally choose the polarization vectors in order to simplify an expression such as eq. (11.14). In other words, the question is how to represent the spin degrees of freedom of the external particles in order to make the amplitude as simple as possible. In the traditional approach to the calculation of amplitudes, one usually refrains from introducing any explicit form for the polarization vectors. Instead, one first squares the amplitude written in terms of generic polarization vectors, such as eq. (11.14), and then the sum over the polarizations of the external gluons is performed by using

$$\sum_{\text{physical pol.}} \epsilon^{\mu *}(p)\epsilon^\nu(p) = -g^{\mu\nu} + \frac{p^\mu n^\nu}{p\cdot n} + \frac{n^\mu p^\nu}{p\cdot n}\,, \tag{11.16}$$

where n^μ is some arbitrary light-like vector. Note that this is the formula for summing over all physical polarizations, which is necessary when calculating unpolarized cross-sections. For cross-sections involving polarized particles, one would perform only a partial sum, which leads to a different projector in the right hand side. If the amplitude is a sum of N_t terms, then this process generates $3N_t^2$ terms in the squared amplitude summed over polarizations. In contrast, the spinor-helicity method that we shall expose below aims at obtaining the amplitude in terms of explicit polarization vectors, for a given assignment of the helicities $\{h_1 = \pm, \cdots, h_n = \pm\}$ of the external gluons, in the form of an expression made of N_t terms that can be easily evaluated (numerically at least). The sum of these N_t terms is done first, and then squared, which is an $\mathcal{O}(1)$ computational task (simply squaring a complex number). Thus, the total cost scales as $2^n N_t$ in this approach. Since N_t grows very quickly with n, this is usually better.

11.3.2 Representation of 4-vectors as bi-spinors

In the previous section, we have seen how the adjoint color degrees of freedom may be represented in terms of the smaller fundamental representation. Likewise, we will now represent the Lorentz structure associated to spin-1 particles in terms of spin-1/2 variables. From a mathematical standpoint, this representation exploits the fact that elements of the Lorentz group $SO(3,1)$ can be mapped to 2×2 complex matrices of unit determinant, i.e. elements of the group $SL(2,\mathbb{C})$. Likewise, 4-momenta can be mapped to 2×2 complex matrices. In order to make this mapping explicit, let us introduce a set of four matrices σ^μ defined by

$$\sigma^\mu \equiv (\mathbf{1}, \boldsymbol{\sigma}^i) \,, \tag{11.17}$$

where $\boldsymbol{\sigma}^{1,2,3}$ are the usual Pauli matrices. In terms of these matrices, a 4-vector p^μ can be mapped into

$$p^\mu \quad \to \quad \mathbf{P} \equiv p_\mu \sigma^\mu = \begin{pmatrix} p_0 + p_3 & p_1 - ip_2 \\ p_1 + ip_2 & p_0 - p_3 \end{pmatrix} \,. \tag{11.18}$$

(In the second equality, we have used the explicit representation of the Pauli matrices.) For amplitudes involving only external gluons, the momentum p^μ has a vanishing invariant norm, $p_\mu p^\mu = 0$, which translates into

$$\begin{aligned} 0 &= p_\mu p^\mu = p_0^2 - p_1^2 - p_2^2 - p_3^2 \\ &= (p_0 + p_3)(p_0 - p_3) - (p_1 + ip_2)(p_1 - ip_2) = \det(\mathbf{P}) \,. \end{aligned} \tag{11.19}$$

Thus, the massless on-shell condition is equivalent to the determinant of the matrix $\mathbf{P}$ being zero. For a 2×2 matrix, a null determinant means that the matrix can be factorized as the direct product of two vectors:

$$\mathbf{P}_{ab} = \lambda_a \xi_b \,, \tag{11.20}$$

where λ, ξ are complex vectors known as *Weyl spinors*. An explicit representation of these vectors is

$$\lambda_a \equiv \begin{pmatrix} \sqrt{p_0 + p_3} \\ \frac{p_1 + ip_2}{\sqrt{p_0 + p_3}} \end{pmatrix} \quad , \qquad \xi_b \equiv \begin{pmatrix} \sqrt{p_0 + p_3} \\ \frac{p_1 - ip_2}{\sqrt{p_0 + p_3}} \end{pmatrix} \tag{11.21}$$

For a real valued 4-vector, λ_a and ξ_a are mutual complex conjugates. However, when we later analytically continue the external momenta in the complex plane, this will no longer be the case. To make the notations more compact, it is customary to introduce the following notations:

$$\big|p\big] = \lambda_a \quad , \qquad \big\langle p\big| = \xi_a \ , \tag{11.22}$$

so that the matrix $\mathbf{P}$ may be written as:

$$\mathbf{P} = \big|p\big]\big\langle p\big| \ . \tag{11.23}$$

It is also convenient to define spinors with raised indices, related to the previous ones as follows,

$$\lambda^a \equiv \epsilon^{ab}\lambda_b = \big[p\big| \quad , \qquad \xi^a \equiv \epsilon^{ab}\xi_b = \big|p\big\rangle \ , \tag{11.24}$$

where ϵ^{ab} is the completely antisymmetric tensor in two dimensions, normalized with $\epsilon^{12} = +1$. From these spinors with raised indices, we may define a 2×2 matrix representation of the 4-vector p^μ with raised indices:

$$\overline{\mathbf{P}} \equiv \big|p\big\rangle\big[p\big| \ . \tag{11.25}$$

Note that this alternative representation corresponds to the definition[3]

$$\overline{\mathbf{P}} \equiv p_\mu\overline{\sigma}^\mu \ , \tag{11.26}$$

with $\overline{\sigma}^\mu \equiv (\mathbf{1}, -\boldsymbol{\sigma}^i)$. In the Weyl representation, where the Dirac matrices read

$$\gamma^\mu = \begin{pmatrix} 0 & \sigma^\mu \\ \overline{\sigma}^\mu & 0 \end{pmatrix} \ , \tag{11.27}$$

we thus have

$$\not{p} \equiv p_\mu\gamma^\mu = \begin{pmatrix} 0 & \mathbf{P} \\ \overline{\mathbf{P}} & 0 \end{pmatrix} \ . \tag{11.28}$$

The fact that we are dealing with on-shell momenta is already built in the factorized representation of eq. (11.20). Amplitudes depend on kinematical invariants such as $(p+q)^2$, for which it is straightforward to check that[4]

$$(p+q)^2 = 2\,p\cdot q = \big\langle pq\big\rangle\big[pq\big] \ , \tag{11.29}$$

where the brackets are defined by contracting upper and lower spinor indices, as in

$$\big\langle pq\big\rangle \equiv \big\langle p\big|_a\big|q\big\rangle^a \ . \tag{11.30}$$

These brackets are antisymmetric ($\big\langle pq\big\rangle = -\big\langle qp\big\rangle$), since they may also be written as:

$$\big\langle pq\big\rangle = \epsilon^{ab}\xi_a(p)\xi_b(q) \ . \tag{11.31}$$

[3] We may use $\epsilon^{ac}\epsilon^{bd}\delta_{dc} = \delta_{ab}$ and $\epsilon^{ac}\epsilon^{bd}\sigma^i_{dc} = -\sigma^i_{ab}$.

[4] For real momenta, angle and square brackets are mutual complex conjugates, and $(p+q)^2$ is a real quantity.

Note that the mixed brackets are zero, $\langle pq] = 0$, as well as the angle and square brackets with twice the same momentum, $\langle pp\rangle = [pp] = 0$.

It is useful to work out the form of momentum conservation in the spinor formalism. For an amplitude with external momenta $\{p_i\}$, chosen to be all incoming, let us denote $|i\rangle, |i], \cdots$ the corresponding spinors. For any arbitrary momenta p and q, we may then write

$$0 = \langle p| \sum_i \overline{P}_i |q] = \sum_i \langle p|i\rangle [i|q] \ . \tag{11.32}$$

Another interesting identity follows from the fact that three 2-component spinors cannot be linearly independent. Thus, given $|p\rangle$, $|q\rangle$ and $|r\rangle$, we must have a relationship of the form:

$$|r\rangle = \alpha\,|p\rangle + \beta\,|q\rangle \ . \tag{11.33}$$

Contracting this equation with $\langle p|$ and $\langle q|$ gives the explicit expression of the coefficients α and β:

$$\alpha = \frac{\langle qr\rangle}{\langle qp\rangle} \quad , \qquad \beta = \frac{\langle pr\rangle}{\langle pq\rangle} \ . \tag{11.34}$$

This leads to

$$|p\rangle\langle qr\rangle + |q\rangle\langle rp\rangle + |r\rangle\langle pq\rangle = 0 \ , \tag{11.35}$$

known as the *Schouten identity*. A similar identity holds with square brackets:

$$|p][qr] + |q][rp] + |r][pq] = 0 \ . \tag{11.36}$$

11.3.3 Polarization vectors

At this point, we have a representation in terms of spinors for the on-shell momenta that appear on the external legs of amplitudes. We also need a similar representation for the polarization vectors. The polarization vectors for a gluon of momentum p with positive and negative helicities may be represented as follows:

$$\epsilon_+^\mu(p;q) \equiv -\frac{\langle q|\overline{\sigma}^\mu|p]}{\sqrt{2}\,\langle qp\rangle} \quad , \qquad \epsilon_-^\mu(p;q) \equiv -\frac{\langle p|\overline{\sigma}^\mu|q]}{\sqrt{2}\,[pq]} \ , \tag{11.37}$$

where q is an arbitrary reference momentum, whose presence is due to the gauge invariance (eq. (11.15)). It does not have to correspond to any of the physical momenta upon which the amplitude depends, and can be chosen in such a way that it simplifies the amplitude. This auxiliary vector can be different for each external line, but it must be the same in each contribution to a given process (this is because a single graph usually does not give a gauge invariant contribution when considered alone). Let us mention a useful Fierz identity for contracting two of the numerators that appear in the above polarization vectors[5]:

$$\langle 1|\overline{\sigma}^\mu|2]\langle 3|\overline{\sigma}_\mu|4] = 2\,\langle 13\rangle\,[24] \ , \tag{11.38}$$

[5] We may use $(\overline{\sigma}^\mu)_{ab}(\overline{\sigma}_\mu)_{cd} = 2(\delta_{ab}\delta_{cd} - \delta_{ad}\delta_{bc}) = 2\,\epsilon^{ac}\epsilon^{bd}$.

from which we obtain the contractions between polarization vectors

$$\begin{aligned}
\epsilon_+(p;q)\cdot\epsilon_+(p';q') &= \frac{[pp']\langle qq'\rangle}{\langle qp\rangle\langle q'p'\rangle}\,,\\
\epsilon_-(p;q)\cdot\epsilon_-(p';q') &= \frac{\langle pp'\rangle[qq']}{[qp][q'p']}\,,\\
\epsilon_+(p;q)\cdot\epsilon_-(p';q') &= \frac{\langle qp'\rangle[pq']}{\langle qp\rangle[p'q']}\,.
\end{aligned} \tag{11.39}$$

Using eq. (11.28), we also obtain the following identities:

$$\begin{aligned}
&p\cdot\epsilon_\pm(p;q) = q\cdot\epsilon_\pm(p;q) = 0\,,\\
&k\cdot\epsilon_+(p;q) = -\frac{\langle qk\rangle[kp]}{\sqrt{2}\langle qp\rangle}\quad,\qquad k\cdot\epsilon_-(p;q) = -\frac{\langle pk\rangle[kq]}{\sqrt{2}[pq]}\,.
\end{aligned} \tag{11.40}$$

11.3.4 Three-point amplitudes in Yang-Mills theory

Let us now discuss the very important case of 3-particle amplitudes in the massless case, since they will appear later as the building blocks of more complicated amplitudes. Such an amplitude depends on three on-shell momenta $p_{1,2,3}$ such that $p_1+p_2+p_3=0$. This implies that

$$\langle 12\rangle[12] = 2\,p_1\cdot p_2 = (p_1+p_2)^2 = p_3^2 = 0\,. \tag{11.41}$$

Therefore, either $\langle 12\rangle = 0$ or $[12]=0$. Let us assume that $\langle 12\rangle\neq 0$. We also have:

$$\langle 12\rangle[23] = \langle 1|\overline{P}_2|3] = -\langle 1|\overline{P}_1+\overline{P}_3|3] = -\underbrace{\langle 11\rangle}_{0}[13] - \langle 13\rangle\underbrace{[33]}_{0} = 0\,, \tag{11.42}$$

which implies that $[23]=0$. Likewise, $[13]=0$. Therefore, all the square brackets are zero if $\langle 12\rangle\neq 0$. Conversely, all the angle brackets would be zero if instead we had assumed that $[12]\neq 0$. From this discussion, we conclude that massless on-shell 3-point amplitudes may depend either on square brackets or on angle brackets, but not on a mixture of both. Recall now that, for real momenta, angle and square brackets are related by complex conjugation. Thus, 3-point amplitudes can only exist for complex momenta. This is of course a trivial consequence of kinematics: momentum conservation $p_1+p_2+p_3=0$ is impossible for three real-valued light-like momenta, except on a measure-zero subset of exceptional configurations.

Let us now be more explicit and calculate the 3-point amplitudes in Yang-Mills theory. For generic polarization vectors $\epsilon_{1,2,3}$, the Feynman rule (11.11) leads to

$$\mathcal{A}_3(123) = 2g\Big[(\epsilon_1\cdot\epsilon_2)(p_1\cdot\epsilon_3)+(\epsilon_2\cdot\epsilon_3)(p_2\cdot\epsilon_1)+(\epsilon_3\cdot\epsilon_1)(p_3\cdot\epsilon_2)\Big]\,, \tag{11.43}$$

where we have used $p_i\cdot\epsilon_i=0$ to cancel several terms. Consider first the helicities $--+$. Using eqs. (11.38) and (11.40), we obtain

$$\begin{aligned}
\mathcal{A}_3(1^-2^-3^+) &= -\frac{\sqrt{2}\,g}{[q_11][q_22]\langle q_33\rangle}\\
&\times\Big\{\langle 12\rangle[q_1q_2]\langle q_31\rangle[13] + \langle 2q_3\rangle[q_23]\langle 12\rangle[2q_1] + \langle q_31\rangle[3q_1]\langle 23\rangle[3q_2]\Big\}\,.
\end{aligned} \tag{11.44}$$

Each of the three terms contains in the numerator an angle bracket between the external momenta (respectively $\langle 12\rangle$, $\langle 12\rangle$ and $\langle 23\rangle$). Therefore, for this amplitude to be non-zero, we must adopt the choice of spinor representation where it is the square brackets that are zero. With this choice, the first term vanishes since it contains $[13]$:

$$\mathcal{A}_3(1^-2^-3^+) = -\sqrt{2}\,g\,\frac{\langle 2q_3\rangle[q_23]\langle 12\rangle[2q_1] + \langle q_31\rangle[3q_1]\langle 23\rangle[3q_2]}{[q_11][q_22]\langle q_33\rangle}\,. \tag{11.45}$$

Using momentum conservation (11.32) in the form of

$$\underbrace{\langle 11\rangle}_{0}[1q_1] + \langle 12\rangle[2q_1] + \langle 13\rangle[3q_1] = 0\,, \tag{11.46}$$

and the Schouten identity (11.35), we arrive at

$$\mathcal{A}_3(1^-2^-3^+) = \sqrt{2}\,g\,\langle 12\rangle\,\frac{[q_13][q_23]}{[q_11][q_22]}\,. \tag{11.47}$$

Momentum conservation also implies

$$\frac{[q_13]}{[q_11]} = \frac{\langle 12\rangle}{\langle 23\rangle}\quad,\qquad \frac{[q_23]}{[q_22]} = \frac{\langle 12\rangle}{\langle 31\rangle}\,, \tag{11.48}$$

which leads to a form of the amplitude that does not contain the auxiliary vectors $q_{1,2}$ anymore:

$$\mathcal{A}_3(1^-2^-3^+) = \sqrt{2}\,g\,\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 31\rangle}\,. \tag{11.49}$$

We have thus obtained a remarkably compact expression of the 3-point amplitude in terms of spinor variables, which is explicitly independent of all the auxiliary vectors q_i. Likewise, a similar calculation would give the following answer for the $++-$ amplitude:

$$\mathcal{A}_3(1^+2^+3^-) = \sqrt{2}\,g\,\frac{[12]^3}{[23][31]}\,. \tag{11.50}$$

(The $+++$ and $---$ amplitudes are zero in Yang-Mills theory, as argued in the next subsection.) Eqs. (11.49) and (11.50) are both much simpler than the Feynman rule for the 3-gluon vertex. This is the simplest illustration of an assertion we made at the beginning of this section, namely that on-shell amplitudes with physical polarizations are much simpler than one may expect from the traditional perturbative expansion. In the case of the 3-gluon amplitude, we may think that the simplicity comes from the fact that it receives contributions from a single diagram. However, this is not true. As a teaser for the next section, let us give the answers for some 4-gluon and 5-gluon amplitudes in the spinor-helicity formalism:

$$\begin{aligned}
\mathcal{A}_4(1^-2^-3^+4^+) &= i(\sqrt{2}\,g)^2\,\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 34\rangle\langle 41\rangle}\,,\\
\mathcal{A}_4(1^-2^-3^-4^+5^+) &= i^2(\sqrt{2}\,g)^3\,\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 34\rangle\langle 45\rangle\langle 51\rangle}\,,
\end{aligned} \tag{11.51}$$

that appear to generalize trivially eq. (11.49) although they result from the sum of 3 and 10 Feynman graphs, respectively. In this section, we have followed a pedestrian approach that consists in starting from the usual Feynman rules, and translating all their building blocks in the spinor-helicity language. However, the simplicity of the results provides an important hint: there must be a better way to obtain them, that bypasses the traditional Feynman rules and provides the answer much more directly.

11.3.5 Little group scaling

It turns out that massless on-shell 3-point amplitudes are completely constrained by a scaling argument, except for an overall prefactor. Thus, the Lagrangian is in a sense useless in specifying their form (it only plays a marginal role in setting their normalization). From eqs. (11.23) and (11.25), it is clear that the representation of massless on-shell 4-momenta as bi-spinors is invariant under the following rescaling:

$$\big|\mathbf{p}\big\rangle \to \lambda \big|\mathbf{p}\big\rangle \quad , \qquad \big|\mathbf{p}\big] \to \lambda^{-1} \big|\mathbf{p}\big] \, , \tag{11.52}$$

known as *little group scaling*. The terminology follows from the fact that there is a one-parameter $SO(2)$ subgroup (the rotations in the plane transverse to $\mathbf{p}$) of the Lorentz group that leaves invariant the vector p^μ. Such a residual symmetry that leaves a vector invariant is called *little group*. In the spinor formulation, this residual symmetry precisely corresponds to the transformation of eq. (11.52).

Under little group scaling of $\big|\mathbf{p}\big\rangle$ and $\big|\mathbf{p}\big]$, the polarization vectors of eq. (11.37) scale as follows:

$$\epsilon^\mu_+(\mathbf{p};\mathbf{q}) \to \lambda^{-2}\, \epsilon^\mu_+(\mathbf{p};\mathbf{q}) \quad , \qquad \epsilon^\mu_-(\mathbf{p};\mathbf{q}) \to \lambda^{2}\, \epsilon^\mu_-(\mathbf{p};\mathbf{q}) \, , \tag{11.53}$$

i.e. a scaling by a factor λ^{-2h} for a helicity h. Note that the polarization vectors are invariant under little group scaling of the auxiliary vector $\mathbf{q}$. In an amplitude, the internal ingredients (propagators and vertices) are not affected by little group scaling. Therefore, if we apply the little group scaling λ_i to an external momentum i of an amplitude, its expression in terms of square and angle spinors must transform as

$$\mathcal{A}_n(1\cdots i^{h_i}\cdots n) \quad \to \quad \lambda_i^{-2h_i}\, \mathcal{A}_n(1\cdots i^{h_i}\cdots n) \, , \tag{11.54}$$

where h_i is the helicity of the external line i (we do not need to specify the helicities of the other external lines).

It turns out that the structure of all 3-point amplitudes is completely fixed by this property. Let us start from the following generic expression

$$\mathcal{A}_3(1^{h_1}2^{h_2}3^{h_3}) = C\,\big\langle 12\big\rangle^\alpha \big\langle 23\big\rangle^\beta \big\langle 31\big\rangle^\gamma \, , \tag{11.55}$$

with α,β,γ undetermined exponents and C a numerical prefactor. Little group scaling implies that

$$-2\,h_1 = \alpha+\gamma \, , \quad -2\,h_2 = \alpha+\beta \, , \quad -2\,h_3 = \beta+\gamma \, , \tag{11.56}$$

whose solution is

$$\alpha = h_3 - h_1 - h_2 \,, \quad \beta = h_1 - h_2 - h_3 \,, \quad \gamma = h_2 - h_3 - h_1 \,. \tag{11.57}$$

Therefore the 3-point amplitude must have the following structure

$$\mathcal{A}_3(1^{h_1}2^{h_2}3^{h_3}) = C\,\langle 12\rangle^{h_3-h_1-h_2}\langle 23\rangle^{h_1-h_2-h_3}\langle 31\rangle^{h_2-h_3-h_1} \,. \tag{11.58}$$

In other words, only the numerical prefactor remains to be determined. The $--+$ 3-gluon amplitude derived in the previous subsection indeed has this structure.

Note that instead of eq. (11.55), we could have chosen an ansatz that involves the square brackets,

$$\mathcal{A}_3(1^{h_1}2^{h_2}3^{h_3}) = C\,[12]^{\alpha'}[23]^{\beta'}[31]^{\gamma'} \,. \tag{11.59}$$

(This is the only alternative, since we are not allowed to mix square and angle brackets in a 3-point amplitude for massless particles.) Little group scaling would now lead to

$$\alpha' = -h_3 + h_1 + h_2 \,, \quad \beta' = -h_1 + h_2 + h_3 \,, \quad \gamma' = -h_2 + h_3 + h_1 \,, \tag{11.60}$$

and consequently

$$\mathcal{A}_3(1^{h_1}2^{h_2}3^{h_3}) = C\,[12]^{-h_3+h_1+h_2}[23]^{-h_1+h_2+h_3}[31]^{-h_2+h_3+h_1} \,. \tag{11.61}$$

The expected dimension of the amplitude is sufficient to choose between eqs. (11.58) and (11.61). Indeed, both angle and square brackets have mass dimension 1, while the 3-gluon amplitude should have dimension 1 in 4-dimensional Yang-Mills theory (for which the coupling constant is dimensionless). Since all the kinematical dependence is carried by the brackets, the prefactor C can only be made of coupling constants and numerical factors, and must therefore be dimensionless in Yang-Mills theory. Consider first the $--+$ amplitude: eq. (11.58) gives a mass dimension $+1$, while eq. (11.61) gives a mass dimension -1. Therefore, the $--+$ amplitude must be expressed by eq. (11.58) in terms of angle brackets. The same argument tells us that the $++-$ amplitude must be given by eq. (11.61), in terms of square brackets.

Let us consider now the $---$ amplitude, for which the little group scaling tells us that

$$\mathcal{A}_3(1^-2^-3^-) = C\,\langle 12\rangle\langle 23\rangle\langle 31\rangle \,. \tag{11.62}$$

Therefore, the prefactor C should have mass dimension -2, which cannot be constructed from the dimensionless coupling constant of Yang-Mills theory, unless $C = 0$ (the same conclusion holds if we try to construct this amplitude with square brackets). Likewise, we conclude that the $+++$ amplitude is zero as well.

11.3.6 Maximally Helicity Violating amplitudes

Let us consider a tree Feynman diagram contributing to a n-point amplitude, with n_3 3-gluon vertices, n_4 4-gluon vertices and n_I internal propagators. These quantities are related by:

$$\begin{aligned} n + 2\,n_I &= 3\,n_3 + 4\,n_4 \,, \\ n_I &= n_3 + n_4 - 1 \,. \end{aligned} \tag{11.63}$$

The second equation is the statement that this graph has no loops. From these equation we get the following identities:

$$n = n_3 + 2\,n_4 + 2 \quad , \qquad n_3 - 2\,n_{_I} = 4 - n \; . \tag{11.64}$$

The contribution of this Feynman graph to the amplitude is made of n polarization vectors, n_I denominators coming from the internal propagators, and n_3 powers of momentum in the numerator, that come from the 3-gluon vertices[6]:

$$\mathcal{A}_n(1\cdots n) \sim \frac{\Big[\prod_{i=1}^{n}\epsilon_i^{\mu_i}\Big]\Big[\prod_{j=1}^{n_3} L_j^{\nu_j}\Big]}{\prod_{k=1}^{n_I} K_k^2} \sim \frac{\big(\text{mass}\big)^{n_3}}{\big(\text{mass}\big)^{2n_{_I}}} \sim \big(\text{mass}\big)^{4-n} \; . \tag{11.65}$$

Firstly, we see that the mass dimension of the n-point amplitude is $4-n$. Moreover, the amplitude $\mathcal{A}_n$ does not carry any Lorentz index. Therefore, in the numerator all the Lorentz indices μ_i and ν_j must be contracted pairwise. These contractions lead to three type of factors:

$$\epsilon_i \cdot \epsilon_{i'} \; , \quad \epsilon_i \cdot L_j \; , \quad L_j \cdot L_{j'} \; . \tag{11.66}$$

Only-$+$ amplitude : Now, consider an amplitude with only $+$ helicities. From eqs. (11.39), we see that all contractions between polarization vectors are proportional to

$$\epsilon_+(i;q_i)\cdot\epsilon_+(i';q_{i'}) \propto \big\langle q_i q_{i'}\big\rangle \; . \tag{11.67}$$

By choosing the auxiliary momenta q_i to be all equal to q, we make all these contractions vanish. Therefore, to obtain a non-zero contribution, it is necessary to contract all the polarization vectors with momenta from the 3-gluon vertices, $\epsilon_i \cdot k_j$. But from the first of eqs. (11.64), we see that $n > n_3$, which means that it is impossible to contract all the n polarization vectors with the n_3 momenta from the vertices. Thus, the all-plus amplitude is zero:

$$\mathcal{A}_n(1^+2^+\cdots n^+) = 0 \; . \tag{11.68}$$

By the same reasoning, we conclude that the all-minus amplitude is also zero. We can see here the power that stems from the freedom of choosing the auxiliary vectors q_i; for generic q_i's, this amplitude would still be zero (since it does not depend on the q_i's), but this zero would result from intricate cancellations among the many graphs that contribute to $\mathcal{A}_n$. Instead, with a smart choice of the auxiliary vectors, we can make this cancellation happen graph by graph.

$-+\cdots+$ amplitude : Consider now an amplitude with one $-$ helicity carried by the first external leg, and $n-1$ $+$ helicities carried by the external legs 2 to n. Now, it is convenient to choose the auxiliary vectors as follows

$$q_2 = q_3 = \cdots = q_n = p_1 \; . \tag{11.69}$$

Again, all the contractions of polarization vectors cancel, since $\epsilon_+(i;q_i)\cdot\epsilon_+(i';q_{i'}) = 0$ for $i,i' \geq 2$, and $\epsilon_-(1;q_1)\cdot\epsilon_+(i;q_i) = 0$ for $i \geq 2$. Since $n > n_3$, it is not possible to contract all the polarization vectors with momenta from the 3-gluon vertices, and these amplitudes also vanish at tree level:

$$\mathcal{A}_n(1^-2^+\cdots n^+) = 0 \; . \tag{11.70}$$

(We also have $\mathcal{A}_n(1^+2^-\cdots n^-) = 0$ at tree level.)

[6] We assume for simplicity Feynman gauge, in which the numerator of the gluon propagator does not depend on momentum.

Maximally Helicity Violating amplitudes : Let us flip one more helicity, e.g. with the assignment $1^-2^-3^+\cdots n^+$. This time, a useful choice of auxiliary vectors is

$$q_1 = q_2 = p_n \quad , \qquad q_3 = q_4 = \cdots = q_n = p_1 \ . \tag{11.71}$$

With this choice, all the contractions of polarization vectors are zero, except:

$$\epsilon_-(2;q_2)\cdot\epsilon_+(i;q_i) \neq 0 \quad \text{for } i = 3,\cdots,n-1 \ . \tag{11.72}$$

Thus, this time, we need to contract the remaining $n-2$ polarization vectors with the n_3 momenta from the 3-gluon vertices, which is possible (provided that $n_4 = 0$, which means that diagrams containing 4-gluon vertices do not contribute to the $--+\cdots+$ amplitude for our choice of auxiliary vectors). Therefore, this assignment of helicities gives a non-zero amplitude:

$$\mathcal{A}_n(1^-2^-3^+\cdots n^+) \neq 0 \ . \tag{11.73}$$

These amplitudes, called the *Maximally Helicity Violating (MHV) amplitudes*, are the simplest non-zero amplitudes. As we shall see later, they are given at tree level by very compact formulas in terms of square and angular brackets (note that up to $n = 5$ external lines, all the non-zero amplitudes are MHV amplitudes). Generically, the complexity of amplitudes increases with the number of $-$ helicities, culminating with amplitudes that have comparable numbers of $-$ and $+$ helicities (increasing further the number of $-$ helicities then reduces the complexity).

11.4 Britto-Cachazo-Feng-Witten on-shell recursion

11.4.1 Main idea

As we have seen, the main obstacle to the calculation of amplitudes by the usual Feynman rules is the proliferation of graphs as one increases the number of external legs. This problem remains true even after one has factorized the color factors, even if it is somewhat mitigated by the fact that the number of cyclic-ordered graphs grows at a slower pace.

This issue could be avoided if there was a way to break down a tree amplitude into smaller pieces (themselves tree amplitudes) that have a smaller number of external legs. It turns out that an amplitude naturally factorizes into two sub-amplitudes when one of its internal propagators goes on-shell. The physical reason of such a factorization is that on-shell momenta correspond to infinitely long-lived particles. Thus, the two sub-amplitudes on each side of this on-shell propagator do not talk to one another. The other advantage of this situation is that the two sub-amplitudes would themselves be on-shell, and therefore we may use for them spinor-helicity formulas that could have been previously obtained for amplitudes with fewer external legs. If this were possible, we would thus obtain a recursive relationship (in the number of external legs) for on-shell amplitudes.

11.4.2 Analytical properties of amplitudes with shifted momenta

Unfortunately, with fixed generic external momenta, tree amplitudes do not have internal on-shell propagators. The trick is to consider a one-parameter *complex* deformation of the external momenta, adjusted in order to make an internal denominator vanish:

$$\mathcal{A}_n(12\cdots n) \quad \to \quad \mathcal{A}_n(12\cdots n;z) \ , \tag{11.74}$$

where z is a complex variable that controls the deformation. The singularities of tree Feynman graphs come from the zeroes of the denominators of its internal propagators, which give poles in z. Our goal will be to choose this deformation in such a way that the total momentum remains conserved, and the deformed external momenta are still on-shell. With such a choice, we will be able to reuse the on-shell formulas obtained for smaller amplitudes.

Let us consider the ratio $\mathcal{A}_n(\cdots;z)/z$. Besides the poles coming from the internal propagators, the ratio also has a simple pole at $z=0$. Let us assume that $\mathcal{A}_n(\cdots;z)$ vanishes when $|z|\to\infty$, so that the integral of $\mathcal{A}_n(\cdots;z)/z$ on a contour at infinity in the complex plane vanishes. Then, we may write

$$0=\oint_\gamma \frac{dz}{2\pi i}\,\frac{\mathcal{A}_n(\cdots;z)}{z}=\mathcal{A}_n(\cdots;0)+\sum_{z_i\in\{\text{poles of }\mathcal{A}_n\}}\text{Res}\,\frac{\mathcal{A}_n(\cdots;z)}{z}\Big|_{z_i}\,. \tag{11.75}$$

The first term, $\mathcal{A}_n(\cdots;z=0)$, is nothing but the amplitude we aim at calculating. This formula therefore expresses it in terms of the residues of $\mathcal{A}_n(\cdots;z)/z$ at the simple poles corresponding to the internal propagators of the amplitude. Moreover, these residues will be factorizable into smaller on-shell amplitudes, precisely because the poles z_i correspond to the on-shellness of some internal propagator.

11.4.3 Minimal momentum shifts

There are many ways to implement a complex shift of the external momenta, but all of them must fulfill the following conditions:

- The sum of the shifted incoming momenta should remain zero. Therefore, we must shift at least two momenta (and the simplest is to shift only two).
- The shifted momenta should stay on-shell at all z.
- The amplitude evaluated at the shifted momenta should go to zero as $|z|\to\infty$.

The condition of momentum conservation is trivially satisfied by choosing two momenta i,j to be shifted, and by giving them opposite shifts:

$$\begin{aligned} p_i \quad &\to \quad \widehat{p}_i=p_i(z)\equiv p_i+z\,k\,,\\ p_i \quad &\to \quad \widehat{p}_j=p_j(z)\equiv p_j-z\,k\,, \end{aligned} \tag{11.76}$$

where we denote with a hat the shifted momenta. All the momenta p_k for $k\neq i,j$ are left unmodified. The on-shell conditions for $\widehat{p}_{i,j}$ are satisfied provided that

$$k^2=0\,,\quad p_i\cdot k=0\,,\quad p_j\cdot k=0\,. \tag{11.77}$$

It turns out that these equations have two solutions (up to an arbitrary prefactor), provided we allow complex momenta. In the spinor notation, the first condition is automatically satisfied if $\mathbf{K}$ can be factorized as in eq. (11.20), while the second and third conditions become

$$\langle ik\rangle[ik]=0\,,\quad \langle jk\rangle[jk]=0\,. \tag{11.78}$$

This explains why we need a complex momentum k^μ. Indeed, for a real k^μ, $|\mathbf{k}]$ and $|\mathbf{k}\rangle$ are related by complex conjugation, and the above conditions reduce to $\langle i\mathbf{k}\rangle = \langle j\mathbf{k}\rangle = 0$. With two-component spinors, this implies $|\mathbf{k}\rangle \propto |i\rangle$ and $|\mathbf{k}\rangle \propto |j\rangle$, which is in general impossible. By allowing a complex momentum k^μ, we let $|\mathbf{k}]$ and $|\mathbf{k}\rangle$ be independent, which allows to solve the above conditions by having for instance:

$$|\mathbf{k}\rangle = |i\rangle \quad , \qquad |\mathbf{k}] = |j] \ . \tag{11.79}$$

(The other independent solution consists in exchanging the roles of i and j.) The bi-spinors corresponding to the shifted momenta are

$$\widehat{\mathbf{P}}_i = |i]\langle i| + z\,|j]\langle i| = \big(|i] + z|j]\big)\,\langle i| \ , \quad \widehat{\mathbf{P}}_j = |j]\langle j| - z\,|j]\langle i| = |j]\,\big(\langle j| - z\langle i|\big) \ , \tag{11.80}$$

from which we read the shifted spinors:

$$\begin{aligned} |\hat{i}\rangle &= |i\rangle \ , \quad |\hat{j}\rangle = |j\rangle - z|i\rangle \ , \\ |\hat{i}] &= |i] + z|j] \ , \quad |\hat{j}] = |j] \ . \end{aligned} \tag{11.81}$$

11.4.4 Behavior at $|z| \to \infty$

Until now, our description of this method has been completely generic and applicable to all sorts of quantum field theories, since no reference has been made to the details of its Lagrangian. These details become important when discussing the condition that $\mathcal{A}_n(\cdots; z)$ vanishes at infinity. Let us discuss the behavior at large z in the case of Yang-Mills theory. Firstly, a z dependence enters in the polarization vectors of the external lines i and j:

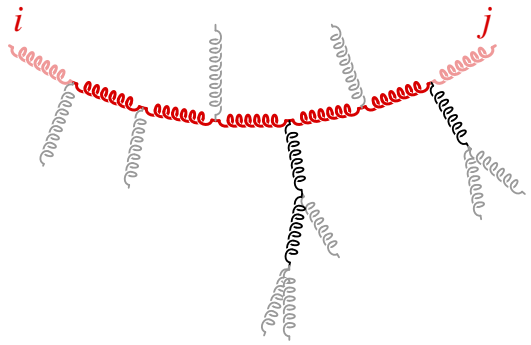
$$\begin{aligned} \epsilon_+^\mu(\hat{i}; q) &= -\frac{\langle q|\overline{\sigma}^\mu|\hat{i}]}{\sqrt{2}\,\langle q\hat{i}\rangle} \sim z \quad , \qquad \epsilon_+^\mu(\hat{j}; q) = -\frac{\langle q|\overline{\sigma}^\mu|\hat{j}]}{\sqrt{2}\,\langle q\hat{j}\rangle} \sim z^{-1} \ , \\ \epsilon_-^\mu(\hat{i}; q) &= -\frac{\langle \hat{i}|\overline{\sigma}^\mu|q]}{\sqrt{2}\,[\hat{i}q]} \sim z^{-1} \quad , \qquad \epsilon_-^\mu(\hat{j}; q) = -\frac{\langle \hat{j}|\overline{\sigma}^\mu|q]}{\sqrt{2}\,[\hat{j}q]} \sim z \ . \end{aligned} \tag{11.82}$$

Inside a graph contributing to this n-point amplitude, we can follow a string of propagators that all carry shifted momenta, from the external line i to the external line j, as illustrated in the figure 11.1. For all these propagators, since $k^2 = 0$, the denominators are linear in z. In addition, the 3-gluon vertices along this string of propagators are linear in the momenta, and therefore scale as z. Along this string, there are s vertices, and $s-1$ propagators, hence a global behavior $\sim z$ at large z. For the assignment $\{h_i = -, h_j = +\}$ of polarizations, we thus find an overall behavior in z^{-1}, valid graph by graph. For the polarizations $\{h_i = +, h_j = +\}$ and $\{h_i = -, h_j = -\}$, the amplitude also goes to zero when $|z| \to \infty$, but this is no longer true graph by graph and a more involved argument is necessary in order to reach this conclusion. Finally, for $\{h_i = +, h_j = -\}$, the amplitude does not go to zero, and we cannot use this shift in eq. (11.75).

11.4.5 Recursion formula

Since all-$+$ amplitudes are zero, the assignment of helicities that we shall consider is generically of the form $1^- \cdots r^-(r+1)^+ \cdots n^+$, and the shift applied to the lines $i = 1, j = n$ leads to a

Figure 11.1: *Propagators affected by the momentum shift (shown in red) in a tree amplitude. The black and gray lines do not depend on z. The lighter propagators, on the external lines, are not actually part of the expression of the amplitude.*



vanishing amplitude when $|z| \to \infty$. We can therefore apply eq. (11.75) and write the amplitude in the following way:

$$\mathcal{A}_n(\cdots) = - \sum_{z_i \in \{\text{poles of } \mathcal{A}_n\}} \text{Res}\, \frac{\mathcal{A}_n(\cdots;z)}{z}\Big|_{z_i} . \tag{11.83}$$

As explained earlier, the poles z_i come from the vanishing denominators of the internal propagators, i.e. one of the propagators along the red line in the figure 11.2. Let us denote K_I the momentum (before the shift) carried by the propagator producing the pole, with the convention that it is oriented in the same direction as p_1. The shift changes this momentum into

$$K_I \quad \to \quad \widehat{K}_I \equiv K_I + z\,k\,, \tag{11.84}$$

and the condition that the denominator of the propagator vanishes after the shift is

$$0 = \widehat{K}_I^2 = K_I^2 + 2\,z_I\,K_I\cdot k\,, \quad \text{i.e. } z_I = -\frac{K_I^2}{2\,K_I\cdot k}\,. \tag{11.85}$$

The singular propagator divides the amplitude into left and right sub-amplitudes, so that we may write:

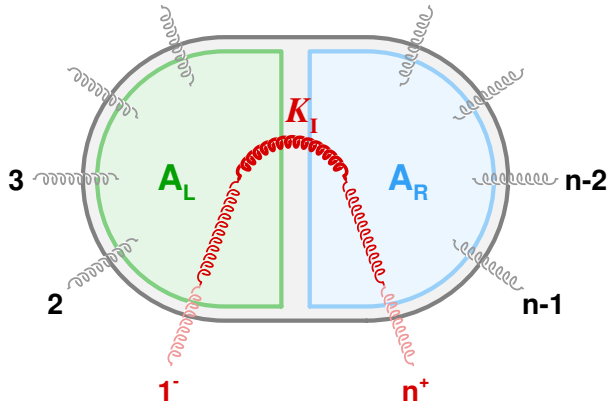
$$\mathcal{A}_n(\hat{1}\,2\cdots(n-1)\hat{n};z) \equiv \sum_{h=\pm} \mathcal{A}_L(\hat{1}\,2\cdots -\widehat{K}_I^{+h};z)\,\frac{i}{\widehat{K}_I^2}\,\mathcal{A}_R(\widehat{K}_I^{-h}\cdots(n-1)\hat{n};z)\,, \tag{11.86}$$

with a sum over the helicity h of the intermediate gluon[7]. From this expression, the residue at the pole z_I of $\mathcal{A}_n(\cdots;z)/z$ takes the form

$$\text{Res}\,\frac{\mathcal{A}_n(\cdots;z)}{z}\Big|_{z_I} = -\sum_{h=\pm} \mathcal{A}_L(\hat{1}\,2\cdots -\widehat{K}_I^{+h};z_I)\,\frac{i}{K_I^2}\,\mathcal{A}_R(\widehat{K}_I^{-h}\cdots(n-1)\hat{n};z_I)\,. \tag{11.87}$$

[7] Both $\mathcal{A}_L$ and $\mathcal{A}_R$ are defined with all gluons incoming. This is why one has argument $-\widehat{K}_I$ and the other one $+\widehat{K}_I$. For the same reason, the helicity is $+h$ on one side and $-h$ on the other side.

Figure 11.2: *Setup for the BCFW recursion formula with shifts applied to the external lines* 1 *and* n. *The pole comes from the propagator carrying the momentum* K_I, *highlighted in dark red. This singular propagator divides the graph into left and right sub-amplitudes,* $\mathcal{A}_L$ *and* $\mathcal{A}_R$.



Both $\mathcal{A}_L$ and $\mathcal{A}_R$ have strictly less than n external lines, which means that the formula is recursive: it expresses an amplitude in terms of smaller amplitudes, eventually breaking it down to 3-point amplitudes. Moreover, the crucial point here is that, when evaluated at the value z_I that gives $\widehat{K}_I^2 = 0$, the left and right sub-amplitudes have only *on-shell* (but complex) external momenta. Therefore, this recursion never requires off-shell amplitudes, which is of utmost importance for keeping out of the calculation unnecessarily complicated kinematics and unphysical degrees of freedom. Since each internal propagator along the red line can be singular for some z, eq. (11.83) contains one term for each such propagator. There are at most $n-3$ terms in this sum, corresponding to the partitions of $[2, n-1] = [2, l] \cup [l+1, n-1]$ with $2 \leq l \leq n-2$.

11.4.6 Parke-Taylor formula for MHV amplitudes

MHV recursion formula : As an illustration of the BCFW recursion formula, let us determine the explicit expression of the MHV amplitudes $\mathcal{A}_n(1^-2^-3^+\cdots n^+)$. We show all the helicity assignments, including those of the singular propagator, in the figure 11.3. In order to avoid having an all-$+$ sub-amplitude on the right, we must choose $h = +$. This choice makes $\mathcal{A}_R$ an $-+\cdots+$ amplitude, which is also zero unless it is a 3-point amplitude. Thus, the BCFW formula reduces to a single term:

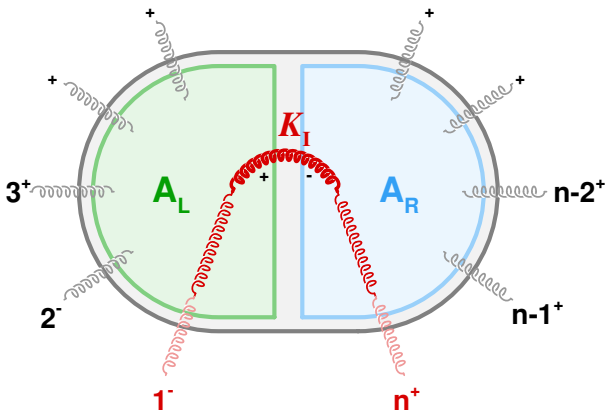
$$\mathcal{A}_n(1^-2^-3^+\cdots n^+) = \mathcal{A}_{n-1}(\widehat{1}^-\,2^-3^+\cdots(n-2)^+ - \widehat{K}_I^+; z_I)\,\frac{i}{K_I^2}\,\mathcal{A}_3(\widehat{K}_I^-(n-1)^+\widehat{n}^+; z_I)\,, \tag{11.88}$$

where the momentum carried by the singular propagator is (before the shift)

$$K_I = -(p_{n-1} + p_n)\,. \tag{11.89}$$

In the right hand side of eq. (11.88), the factor on the right is an already known 3-point amplitude, and the factor on the left is an MHV amplitude with $n-1$ external legs.

Figure 11.3: *Setup for applying the BCFW recursion formula to the calculation of the $--+\cdots+$ MHV amplitude. We have indicated explicitly all the helicities. Note that only $h = +$ is allowed in the sum over the helicity of the singular propagator (otherwise the right-side sub-amplitude would be a vanishing all-+ amplitude).*



Four-point MHV amplitude : Let us now calculate the first few iterations of this recursion, in order to guess a formula for the MHV amplitude that will be our hypothesis for an inductive proof. Firstly, consider the $--++$ 4-point MHV amplitude, In this case, the BCFW recursion formula gives

$$\mathcal{A}_4(1^-2^-3^+4^+) = \mathcal{A}_3(\hat{1}^-\,2^-\,-\,\hat{K}_I^+;z_I)\,\frac{i}{K_I^2}\,\mathcal{A}_3(\hat{K}_I^-3^+\hat{4}^+;z_I)\,, \tag{11.90}$$

and both amplitudes in the right hand side are known. This gives:

$$\mathcal{A}_4(1^-2^-3^+4^+) = 2\,i\,g^2\,\frac{\langle\hat{1}2\rangle^3}{\langle 2\hat{K}_I\rangle\langle\hat{K}_I\hat{1}\rangle}\,\frac{1}{\langle 12\rangle[12]}\,\frac{[3\hat{4}]^3}{[\hat{4}\hat{K}_I]\,[\hat{K}_I 3]}\,. \tag{11.91}$$

Using the fact that

$$|\hat{1}\rangle = |1\rangle\,,\quad |\hat{1}] = |1] + z|4]\,,\quad |\hat{4}\rangle = |4\rangle - z|1\rangle\,,\quad |\hat{4}] = |4]\,, \tag{11.92}$$

we obtain

$$\begin{aligned}
&|\hat{K}_I\rangle[\hat{K}_I| = |1\rangle[1| + |2\rangle[2| + z_I\,|1\rangle[4|\,,\\
&\langle 2\hat{K}_I\rangle[\hat{K}_I 4] = \langle 21\rangle[14]\,,\\
&\langle 1\hat{K}_I\rangle[\hat{K}_I 3] = \langle 12\rangle[23]\,,
\end{aligned} \tag{11.93}$$

which leads to

$$\mathcal{A}_4(1^-2^-3^+4^+) = 2\,i\,g^2\,\frac{[34]^3}{[41]\,[12]\,[23]}\,. \tag{11.94}$$

This formula, that depends only on square brackets, can also be expressed in terms of angle brackets. Let us multiply the numerator and denominator by $\langle 12\rangle^3$. Then, momentum conservation leads to

$$\begin{aligned}
&[41]\langle 12\rangle = -[43]\langle 32\rangle \,, \\
&\langle 12\rangle[23] = -\langle 14\rangle[43] \,, \\
&[12]\langle 12\rangle = (p_1+p_2)^2 = (p_3+p_4)^2 = [34]\langle 34\rangle \,,
\end{aligned} \tag{11.95}$$

and we finally obtain

$$\mathcal{A}_4(1^-2^-3^+4^+) = 2\,i\,g^2\,\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 34\rangle\langle 41\rangle}\,. \tag{11.96}$$

This formula could in principle have been obtained from eq. (11.14), by putting the external lines on-shell and by using $--++$ polarization vectors, at the cost of considerable effort. We see here the power of on-shell recursion: since one only manipulates on-shell sub-amplitudes with physical polarizations, the complexity of all the intermediate expressions is comparable to that of the final result, unlike with the standard method.

Five-point MHV amplitude : Consider now the amplitude $\mathcal{A}_5(1^-2^-3^+4^+5^+)$. The BCFW recursion formula (11.88) now reads:

$$\begin{aligned}
\mathcal{A}_5(1^-2^-3^+4^+5^+) &= \mathcal{A}_4(\hat{1}^-2^-3^+ - \widehat{K}_I^+; z_I)\,\frac{i}{K_I^2}\,\mathcal{A}_3(\widehat{K}_I^- 4^+\hat{5}^+; z_I) \\
&= (\sqrt{2}\,g)^3\,i^2\,\frac{\langle\hat{1}2\rangle^3[4\hat{5}]^3}{\langle 23\rangle\langle 3\widehat{K}_I\rangle\langle\widehat{K}_I 1\rangle\,[45]\langle 45\rangle\,[\hat{5}\widehat{K}_I][\widehat{K}_I 4]}\,,
\end{aligned} \tag{11.97}$$

where we have chosen to express K_I^2 as $(p_4+p_5)^2 = [45]\langle 45\rangle$. This time, we use

$$\begin{aligned}
&|\hat{1}\rangle = |1\rangle\,, \quad |\hat{1}] = |1] + z|5]\,, \quad |\hat{5}\rangle = |5\rangle - z|1\rangle\,, \quad |\hat{5}] = |5]\,, \\
&|\widehat{K}_I\rangle[\widehat{K}_I| = -|4\rangle[4| - |5\rangle[5| + z_I\,|1\rangle[5|\,, \\
&\langle 3\widehat{K}_I\rangle[\widehat{K}_I\hat{5}] = -\langle 34\rangle[45]\,, \\
&\langle\hat{1}\widehat{K}_I\rangle[\widehat{K}_I 4] = -\langle 51\rangle[45]\,,
\end{aligned} \tag{11.98}$$

which gives

$$\mathcal{A}_5(1^-2^-3^+4^+5^+) = (\sqrt{2}\,g)^3\,i^2\,\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 34\rangle\langle 45\rangle\langle 51\rangle}\,. \tag{11.99}$$

This remarkably simple formula, that encapsulates the sum of 10 cyclic-ordered Feynman diagrams (in QCD, this corresponds to 25 diagrams before color ordering), in fact exhausts all the possibilities for 5-point functions (the $++---$ amplitude is given by the same formula with square brackets instead of angle brackets).

Parke-Taylor formula : The previous results for 3, 4 and 5-point MHV amplitudes lead us to conjecture the following general formula:

$$\mathcal{A}_n(1^-2^-3^+\cdots n^+) = (\sqrt{2}\,g)^{n-2}\,i^{n-3}\,\frac{\langle 12\rangle^3}{\langle 23\rangle\langle 34\rangle\cdots\langle (n-1)n\rangle\langle n1\rangle}\,, \tag{11.100}$$

known as the *Parke-Taylor formula*. Let us assume the formula to be true for all $p < n$, and consider now the case of the n-point MHV amplitude. The BCFW recursion formula reads:

$$\begin{aligned}\mathcal{A}_n(1^-2^-3^+\cdots n^+) &= i\,\frac{\mathcal{A}_{n-1}(\hat{1}^-\,2^-3^+\cdots(n-2)^+ - \widehat{K}_I^+;z_I)\mathcal{A}_3(\widehat{K}_I^-(n-1)^+\hat{n}^+;z_I)}{K_I^2}\\ &= (\sqrt{2}\,g)^{n-2}\,i^{n-3}\,\frac{\langle\hat{1}2\rangle^3}{\langle 23\rangle\cdots\langle(n-2)\widehat{K}_I\rangle\langle\widehat{K}_I 1\rangle}\\ &\quad\times\frac{1}{[(n-1)n]\langle(n-1)n\rangle}\,\frac{[(n-1)\hat{n}]^3}{[\hat{n}\widehat{K}_I]\,[\widehat{K}_I(n-1)]}\,,\end{aligned} \tag{11.101}$$

where we have used our induction hypothesis for the $(n-1)$-point MHV amplitude that appears in the left sub-amplitude. The spinor manipulations that are necessary to simplify this expression are the same as in the case of the 5-point amplitude, and lead to:

$$\begin{aligned}\langle(n-2)\widehat{K}_I\rangle[\widehat{K}_I\hat{n}] &= -\langle(n-2)(n-1)\rangle[(n-1)n]\,,\\ \langle\hat{1}\widehat{K}_I\rangle[\widehat{K}_I(n-1)] &= -\langle n1\rangle[(n-1)n]\,,\end{aligned} \tag{11.102}$$

thanks to which we obtain eq. (11.100) for n points. Up to 5-points, all amplitudes are MHV (or anti-MHV, i.e. $++---$). Beyond 5-points, there exist non-MHV amplitudes, that are not given by the Parke-Taylor formula. However, multiple MHV amplitudes can be sewed together in order to construct the non-MHV ones, with a set of rules known as the *Cachazo-Svrcek-Witten (CSW) rules*. Such an expansion is much more efficient that the textbook perturbation theory, because it is in terms of on-shell gauge-invariant building blocks (the MHV amplitudes) that already encapsulate a lot of the underlying complexity.

11.5 Gravitational amplitudes

11.5.1 Textbook approach for amplitudes with gravitons

In the previous section, we have derived the BCFW recursion formula and applied it to the calculation of the tree-level MHV amplitudes in Yang-Mills theory. However, the validity of this recursion is by no means limited to a gauge theory with spin-1 bosons such as gluons. It may in fact be applied to any quantum field theory provided that:

- we have expressions for the on-shell 3-point amplitudes,
- the shifted amplitudes vanish when $|z| \to \infty$.

In particular, it could be interesting to apply it to the calculation of scattering amplitudes that involve gravitons[8]. The Feynman rules for Einstein gravity can be obtained from the Hilbert-Einstein action,

$$S_{HE} \equiv \int d^4x\, \sqrt{-g}\, \left\{ \frac{2}{\kappa^2} R - \frac{g^{\mu\nu} g^{\rho\sigma}}{4} F_{\mu\rho} F_{\nu\sigma} + \frac{g^{\mu\nu}}{2} (\partial_\mu \phi)(\partial_\nu \phi) - \frac{m^2}{2} \phi^2 \right\} , \quad (11.103)$$

where $g^{\mu\nu}$ is the metric tensor, R is the Ricci curvature and κ is a coupling constant related to Newton's constant by $\kappa^2 = 32\pi G_N$. In this action, we have also added the minimal coupling to a gauge field and to a scalar field, in order to investigate gravitational interactions with light and matter. The rules for the propagators and vertices involving gravitons are obtained by expanding the metric around flat space:

$$g^{\mu\nu} = \eta^{\mu\nu} + \kappa\, h^{\mu\nu} . \quad (11.104)$$

($\eta^{\mu\nu}$ is the flat space Minkowski metric.) Let us make a remark on dimensions: Newton's constant has mass dimension -2, κ has mass dimension -1, the Ricci curvature has mass dimension 2, and $h^{\mu\nu}$ has mass dimension $+1$ (like the scalar ϕ and the photon A_μ). The expansion in powers of $h^{\mu\nu}$ leads to an infinite series of terms (because the Ricci tensor contains the inverse $g_{\mu\nu}$ of the metric tensor, and also because of the expansion of the square root $\sqrt{-g}$). Schematically, the expansion of the Hilbert-Einstein action starts with the following terms:

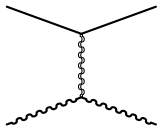
$$S_{HE} \sim \int d^4x \left\{ h\partial^2 h + \kappa\, h^2 \partial^2 h + \kappa^2\, h^3 \partial^2 h + \cdots + \kappa\, h\phi\partial^2\phi + \kappa\, h F^2 + \cdots \right\} . \quad (11.105)$$

This sketch only indicates the number of powers of h and the number of derivatives contained in each term, but of course the actual structure of these terms is much more complicated. For instance, the vertex describing the coupling $\phi\phi h$ between two scalars and a graviton reads:

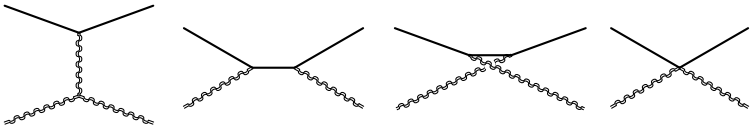
$$\Gamma^{\mu\nu}(p_1, p_2) = -\frac{i\kappa}{2} \left[p_1^\mu p_2^\nu + p_1^\nu p_2^\mu - \eta^{\mu\nu} (p_1 \cdot p_2 - m^2) \right] , \quad (11.106)$$

where $p_{1,2}$ are the momenta carried by the two scalar lines (since the graviton has spin 2, the graviton attached to this vertex carries two Lorentz indices). But the $\gamma\gamma h$ coupling is far more complicated, and the hhh tri-graviton vertex is even more complex, leading to extremely cumbersome perturbative calculations if performed by the traditional approach.

It turns out that tree amplitudes in Einstein gravity have a simple form in the spinor-helicity formalism, very much like their Yang-Mills analogue. The goal of this section is to illustrate on two examples the use of the spinor-helicity formalism, combined to the BCFW recursion, in order to calculate some amplitudes that have a relevance in gravitational physics: (1) gravitational bending of light by a mass, and (2) scattering of a gravitational wave by a mass. In both examples, the mass acting as a source of gravitational field is taken to be a scalar particle. In the approach based on conventional Feynman perturbation theory, these processes are given by the following diagrams:

$\mathcal{A}_{\phi\gamma\to\phi\gamma} \sim$ 

[8] At tree-level, these amplitudes are completely prescribed by the equivalence principle and general relativity, and their calculation does not require to have a consistent theory of gravitational quantum fluctuations.

$\mathcal{A}_{\phi h\to\phi h}\sim$ 

In particular, the second example (bending of a gravitational wave by a mass) would be an extremely difficult calculation, because of the complexity of the 3-graviton vertex.

11.5.2 Three-point amplitudes with gravitons

In order to obtain these amplitudes with the formalism previously exposed in the case of Yang-Mills theory, the first step is to obtain the 3-point amplitudes involving scalars, photons and gravitons. External scalar particles must have helicity $h = 0$, photons can have helicities $h = \pm 1$ and gravitons can have helicities $h = \pm 2$ with polarization vectors that are "squares" of the gluon polarization vectors:

$$\epsilon^{\mu\nu}_{2h}(p;q) = \epsilon^{\mu}_{h}(p;q)\,\epsilon^{\nu}_{h}(p;q)\,. \tag{11.107}$$

For 3-point amplitudes that involve only massless particles (photons and gravitons), little group scaling is sufficient to constrain completely their form. We obtain:

$$\begin{aligned}
&\mathcal{A}_{h\gamma\gamma}(1^{\pm 2}2^{+}3^{+}) = \mathcal{A}_{h\gamma\gamma}(1^{\pm 2}2^{-}3^{-}) = 0\,,\\
&\mathcal{A}_{h\gamma\gamma}(1^{+2}2^{+}3^{-}) = -\tfrac{\kappa}{2}\,\big[12\big]^{4}\big[23\big]^{-2}\,,\\
&\mathcal{A}_{h\gamma\gamma}(1^{+2}2^{-}3^{+}) = -\tfrac{\kappa}{2}\,\big[23\big]^{-2}\big[31\big]^{4}\,,\\
&\mathcal{A}_{h\gamma\gamma}(1^{-2}2^{+}3^{-}) = -\tfrac{\kappa}{2}\,\big\langle 23\big\rangle^{-2}\big\langle 31\big\rangle^{4}\,,\\
&\mathcal{A}_{h\gamma\gamma}(1^{-2}2^{-}3^{+}) = -\tfrac{\kappa}{2}\,\big\langle 12\big\rangle^{4}\big\langle 23\big\rangle^{-2}\,.
\end{aligned} \tag{11.108}$$

In order to obtain the zeroes of the first line, and to choose between square and angle brackets for the non-zero results, we use the fact that the 3-point amplitude must have mass dimension $+1$, with a prefactor made up only of numerical constants and one power of κ (that has mass dimension -1). The value of the prefactor is obtained by inspecting the term of order κ in the expansion of $\sqrt{-g}\,F^2$. For the 3-graviton amplitudes, little group scaling leads to

$$\begin{aligned}
&\mathcal{A}_{hhh}(1^{+2}2^{+2}3^{+2}) = \mathcal{A}_{hhh}(1^{-2}2^{-2}3^{-2}) = 0\,,\\
&\mathcal{A}_{hhh}(1^{-2}2^{-2}3^{+2}) \propto \kappa\,\big\langle 12\big\rangle^{6}\big\langle 23\big\rangle^{-2}\big\langle 31\big\rangle^{-2}\,,\\
&\mathcal{A}_{hhh}(1^{+2}2^{+2}3^{-2}) \propto \kappa\,\big[12\big]^{6}\big[23\big]^{-2}\big[31\big]^{-2}\,.
\end{aligned} \tag{11.109}$$

Interestingly, the kinematical part of the non-zero 3-graviton amplitudes is simply the square[9] of that of the 3-gluon amplitudes with like-sign helicities (see eqs. (11.49) and (11.50)), despite a considerably more complicated Feynman rule for the 3-graviton vertex. This is yet another illustration of the fact that traditional Feynman rules carry a lot of unnecessary information that disappears in on-shell amplitudes with physical polarizations.

[9] This property of 3-point purely gravitational amplitudes has a generalization for n-point amplitudes, known as the *Kawai-Lewellen-Tye (KLT) relations*. These relations were also recently interpreted as a form of *color-kinematics duality* by Bern, Carrasco and Johansson.

For the $\phi\phi h$ amplitude, we cannot rely on little group scaling because the scalar field is massive. Instead, we simply contract eq. (11.106) with the polarization vector (11.107) of the graviton, and take the external momenta on mass-shell. For a graviton of helicity $+2$, we have

$$\begin{aligned}\mathcal{A}_{\phi\phi h}(1^0 2^0 3^{+2}) &= -i\kappa\,(p_1\cdot\epsilon_+(p_3;q))(p_2\cdot\epsilon_+(p_3;q)) \\ &= -\frac{i\kappa}{2}\,\frac{\langle q|\overline{P}_1|p_3]\langle q|\overline{P}_2|p_3]}{\langle qp_3\rangle^2}\,,\end{aligned} \tag{11.110}$$

where $p_3+p_1+p_2=0$. With a graviton of helicity -2, we have

$$\mathcal{A}_{\phi\phi h}(1^0 2^0 3^{-2}) = -\frac{i\kappa}{2}\,\frac{\langle p_3|\overline{P}_1|q]\langle p_3|\overline{P}_2|q]}{[qp_3]^2}\,. \tag{11.111}$$

Note that,

$$\begin{aligned}\langle p_3|\overline{P}_2|q] &= -\langle p_3|\overline{P}_1|q] - \underbrace{\langle p_3|\overline{P}_3}_{0}|q] = -\langle p_3|\overline{P}_1|q]\,, \\ \langle q|\overline{P}_2|p_3] &= -\langle q|\overline{P}_1|p_3] - \langle q|\underbrace{\overline{P}_3|p_3]}_{0} = -\langle q|\overline{P}_1|p_3]\,,\end{aligned} \tag{11.112}$$

which allows the following simplification of the above 3-point amplitudes:

$$\mathcal{A}_{\phi\phi h}(1^0 2^0 3^{+2}) = \frac{i\kappa}{2}\,\frac{\langle q|\overline{P}_1|p_3]^2}{\langle qp_3\rangle^2}\,,\quad \mathcal{A}_{\phi\phi h}(1^0 2^0 3^{-2}) = \frac{i\kappa}{2}\,\frac{\langle p_3|\overline{P}_1|q]^2}{[qp_3]^2}\,. \tag{11.113}$$

Note that since $p_1^2 = m^2 \neq 0$, the bi-spinor $\overline{P}_1$ does not admit a factorized form, and this cannot be simplified further.

11.5.3 Gravitational bending of light

Shifted momenta : Consider now the amplitude $\mathcal{A}_{\gamma\gamma\phi\phi}(1^+2^-3^04^0)$, and apply the shift to the lines 2 and 3:

$$\begin{aligned}&\widehat{p}_2 \equiv p_2 + z\,k\,,\quad \widehat{p}_3 \equiv p_3 - z\,k\,, \\ &k^2 = 0\,,\quad k\cdot p_2 = k\cdot p_3 = 0\,.\end{aligned} \tag{11.114}$$

Since p_2 is massless, the condition $p_2\cdot k=0$ can be satisfied by choosing for instance:

$$|k\rangle = |2\rangle \tag{11.115}$$

However, since $p_3^2 = m^2$, the bi-spinor P_3 that represents the momentum p_3 cannot be factorized. Instead, we may write

$$0 = 2\,k\cdot p_3 = \langle k|\overline{P}_3|k]\,, \tag{11.116}$$

which can be solved by[10]

$$\big|\mathbf{k}\big] = \mathbf{P}_3\big|2\big\rangle\,. \tag{11.117}$$

The shifted bi-spinors read

$$\begin{aligned}
\widehat{\mathbf{P}}_2 &= \big|2\big]\big\langle 2\big| + z\,\big|\mathbf{k}\big]\big\langle\mathbf{k}\big| = \big(\underbrace{\big|2\big] + z\,\mathbf{P}_3\big|2\big\rangle}_{\big|\hat{2}\big]}\big)\,\big\langle 2\big|\,,\\
\widehat{\mathbf{P}}_3 &= \big|3\big]\big\langle 3\big| - z\,\big|\mathbf{k}\big]\big\langle\mathbf{k}\big| = \big|3\big]\big\langle 3\big| - z\,\mathbf{P}_3\big|2\big\rangle\big\langle 2\big|\,.
\end{aligned} \tag{11.118}$$

Note that the second one is not factorizable, because the line 3 carries a massive particle.

Scattering amplitude : With this choice of shifts, the BCFW recursion formula can be written as follows[11]

$$\mathcal{A}_{\gamma\gamma\phi\phi}(1^+2^-3^04^0) = \frac{i}{P_I^2}\sum_{h=\pm 2}\mathcal{A}_{\gamma\gamma h}(1^+\hat{2}^- - \widehat{P}_I^{+h};z_I)\,\mathcal{A}_{h\phi\phi}(\widehat{P}_I^{-h}\hat{3}^04^0;z_I)\,, \tag{11.119}$$

where the shifted momenta in the 3-point amplitudes are evaluated at the z_I for which the shifted momentum $\widehat{P}_I$ of the intermediate graviton is on-shell. The condition for the intermediate momentum to be on-shell reads

$$0 = \widehat{P}_I^2 = (p_1+\widehat{p}_2)^2 = 2\,p_1\cdot\widehat{p}_2 = \big\langle 12\big\rangle\Big(\underbrace{\big[12\big] + z_I\,\big[1\big|\mathbf{P}_3\big|2\big\rangle}_{\big[1\hat{2}\big]=0}\Big)\,, \tag{11.120}$$

whose solution is $z_I = -\big[12\big]/\big[1\big|\mathbf{P}_3\big|2\big\rangle$. Plugging in the results for the 3-point amplitudes and summing explicitly over the two helicities of the intermediate graviton, the $\gamma\gamma\phi\phi$ amplitude can be written as

$$\mathcal{A}_{\gamma\gamma\phi\phi}(1^+2^-3^04^0) = \frac{\kappa^2}{4}\,\frac{1}{\big\langle 12\big\rangle\big[12\big]}\left\{\frac{\big[\widehat{P}_I 1\big]^4}{\big[1\hat{2}\big]^2}\,\frac{\big\langle\widehat{P}_I\big|\overline{\mathbf{P}}_4\big|\mathbf{q}\big]^2}{\big[\mathbf{q}\widehat{P}_I\big]^2} + \frac{\big\langle\widehat{P}_I\hat{2}\big\rangle^4}{\big\langle 1\hat{2}\big\rangle^2}\,\frac{\big\langle\mathbf{q}\big|\overline{\mathbf{P}}_4\big|\widehat{P}_I\big]^2}{\big\langle\mathbf{q}\widehat{P}_I\big\rangle^2}\right\}. \tag{11.121}$$

For the first term, we may write

$$\frac{\big[\widehat{P}_I 1\big]^4}{\big[1\hat{2}\big]^2} = \frac{\big[\widehat{P}_I 1\big]^4\big\langle\widehat{P}_I 1\big\rangle^4}{\big[1\hat{2}\big]^2\big\langle\widehat{P}_I 1\big\rangle^4} = \frac{\big(2\,p_1\cdot\widehat{p}_2\big)^4}{\big[1\hat{2}\big]^2\big\langle\widehat{P}_I 1\big\rangle^4} = \frac{\big\langle 1\hat{2}\big\rangle^4\big[1\hat{2}\big]^2}{\big\langle\widehat{P}_I 1\big\rangle^4} = 0\,. \tag{11.122}$$

[10] We use $\big\langle\mathbf{k}\big|\overline{\mathbf{P}}_3\big|\mathbf{k}\big] = \big\langle 2\big|\overline{\mathbf{P}}_3\mathbf{P}_3\big|2\big\rangle = m^2\big\langle 22\big\rangle = 0$. When p_3 is massless, $\mathbf{P}_3$ factorizes as $\mathbf{P}_3 = \big|3\big]\big\langle 3\big|$, and this solution becomes $\big|\mathbf{k}\big] = \big|3\big]\big\langle 32\big\rangle$. Up to a rescaling, this is the solution we have previously used in the massless case.

[11] Note that the factorization with one scalar and one photon on each side of the singular propagator is not allowed: indeed, the intermediate propagator would need to carry a scalar, and we would have two $\phi\phi\gamma$ sub-amplitudes, that are zero per our assumption that the scalar field is not electrically charged.

The final zero occurs when we evaluate the expression at z_1, as a consequence of eq. (11.120). Therefore, the amplitude reduces to a single term. Furthermore, we are still free to choose the auxiliary vector $\mathbf{q}$. A convenient choice turns out to be $\mathbf{q} = \mathbf{p}_2$, which leads to:

$$\mathcal{A}_{\gamma\gamma\phi\phi}(1^+2^-3^04^0) = \frac{\kappa^2}{4}\,\frac{\langle\widehat{P}_1 2\rangle^2\langle 2|\overline{\mathbf{P}}_4|\widehat{P}_1]^2}{\langle 12\rangle^3[12]}\,. \tag{11.123}$$

Then, notice that

$$\langle 2|\overline{\mathbf{P}}_4|\widehat{P}_1]\langle\widehat{P}_1 2\rangle = \langle 2|\overline{\mathbf{P}}_4|1]\langle 12\rangle\,, \tag{11.124}$$

which gives the following extremely compact form for the amplitude:

$$\mathcal{A}_{\gamma\gamma\phi\phi}(1^+2^-3^04^0) = \frac{\kappa^2}{4}\,\frac{\langle 2|\overline{\mathbf{P}}_4|1]^2}{\langle 12\rangle[12]}\,. \tag{11.125}$$

Cross section and deflection angle : Using $\langle 2|\overline{\mathbf{P}}_4|1]^\dagger = \langle 1|\overline{\mathbf{P}}_4|2]$, the modulus square of the amplitude is

$$\left|\mathcal{A}_{\gamma\gamma\phi\phi}(1^+2^-3^04^0)\right|^2 = \frac{\kappa^4}{16}\,\frac{\langle 2|\overline{\mathbf{P}}_4|1]^2\langle 1|\overline{\mathbf{P}}_4|2]^2}{\langle 12\rangle^2[12]^2} = \frac{\kappa^4}{16}\,\frac{(s_{13}s_{14}-m^4)^2}{s_{12}^2}\,, \tag{11.126}$$

where we have introduced the Lorentz invariants $s_{ij} \equiv (p_i+p_j)^2$. The differential cross-section with respect to the solid angle of the outgoing photon is given by

$$\frac{d\sigma}{d\Omega} = \frac{1}{64\pi^2\,s_{14}}\left|\mathcal{A}_{\gamma\gamma\phi\phi}(1^+2^-3^04^0)\right|^2\,. \tag{11.127}$$

Let us now consider the limit of long wavelength photons, namely $\omega = |\mathbf{p}_{1,2}| \ll m$. In this limit, the Lorentz invariants that appear in the cross-section simplify into

$$\begin{aligned} s_{12} &\approx 4\,\omega^2\,\sin^2\tfrac{\theta}{2}\,,\\ s_{13} &\approx m^2 - 2\,m\,\omega - 4\,\omega^2\,\sin^2\tfrac{\theta}{2}\,,\\ s_{14} &\approx m^2 + 2\,m\,\omega\,, \end{aligned} \tag{11.128}$$

where ω is the photon energy and θ its deflection angle in the center of mass frame (which is also the frame of the massive scalar particle in this limit). For large enough impact parameters, the deflection angle is small, $\theta \ll 1$. Thus we obtain in this limit

$$\frac{d\sigma}{d\Omega} \approx \frac{16\,G_N^2\,m^2}{\theta^4}\,. \tag{11.129}$$

In order to determine the deflection angle as a function of the impact parameter b, consider a flux $\mathcal{F}$ of photons along the z direction, with the massive scalar at rest at the origin. Out of this flux, consider specifically the incoming photons in a ring of radius b and width db. The number of photons flowing per unit time through this ring is

$$2\pi\,b\,\mathcal{F}\,db\,. \tag{11.130}$$

All these photons are scattered in the range of polar angles $[\theta(b) + d\theta, \theta(b)]$ (note that $d\theta$ is negative for $db > 0$, because the deflection angle decreases at larger b), which corresponds to a solid angle:

$$d\Omega = -2\pi \sin\big(\theta(b)\big)\, d\theta \,. \tag{11.131}$$

By definition, the number of scattering events is the flux times the cross-section, i.e.

$$2\pi b\, \mathcal{F}\, db = \mathcal{F}\, \frac{d\sigma}{d\Omega}\, d\Omega \,, \tag{11.132}$$

that can be integrated for small angles into

$$\theta(b) = \frac{4\, G_{_N}\, m}{b} \,. \tag{11.133}$$

(The integration constant is chosen so that the deflection vanishes when $b \to \infty$.) This is indeed the standard formula from general relativity, that can be derived by considering geodesics in the Schwarzchild metric.

11.5.4 Scattering of gravitational waves by a mass

Let us now study the scattering amplitude between a scalar and a graviton, whose low energy limit will provide us information about the scattering of a long wavelength gravitational wave by a mass. A priori, each of the two gravitons may have a helicity ± 2, but the cases $\{+2,+2\}$ and $\{-2,-2\}$ correspond to a helicity flip of the graviton, which is suppressed at low frequency. Therefore, let us consider the amplitude $\mathcal{A}_{hh\phi\phi}(1^{-2}2^{+2}3^0 4^0)$. When writing the BCFW recursion for this amplitude, the simplest shift is one that affects the lines 1 and 2, more specifically:

$$\begin{aligned} &|\hat{2}] = |2] \,, \quad |\hat{2}\rangle = |2\rangle - z\,|1\rangle \,, \\ &|\hat{1}] = |1] + z\,|2] \,, \quad |\hat{1}\rangle = |1\rangle \,, \end{aligned} \tag{11.134}$$

Because the polarization vectors of the gravitons are squares of the spin-1 ones, this shift can be proven to lead to a vanishing amplitude when $|z| \to \infty$ simply by power counting. With the shift (11.134), the intermediate propagator carries a scalar, and therefore it has only the $h = 0$ helicity. The BCFW recursion formula contains two terms,

$$\begin{aligned} \mathcal{A}_{hh\phi\phi}(1^{-2}2^{+2}3^0 4^0) \;=\; & \mathcal{A}_{h\phi\phi}(\hat{1}^{-2}4^0 \widehat{P}^0_{23}) \frac{i}{P^2_{23} - m^2} \mathcal{A}_{h\phi\phi}(\hat{2}^{+2}3^0 - \widehat{P}^0_{23}) \\ & + \mathcal{A}_{h\phi\phi}(\hat{1}^{-2}3^0 \widehat{P}^0_{24}) \frac{i}{P^2_{24} - m^2} \mathcal{A}_{h\phi\phi}(\hat{2}^{+2}4^0 - \widehat{P}^0_{24}) \,, \end{aligned} \tag{11.135}$$

that differ by a permutation of the external scalars. In the above equation, we have made explicit the intermediate momentum, $\widehat{P}_{23} \equiv \widehat{p}_2 + p_3$ in the first term and $\widehat{P}_{24} \equiv \widehat{p}_2 + p_4$ in the second one. The explicit forms of the first and second terms are

$$\begin{aligned} i\, \frac{\mathcal{A}_{h\phi\phi}(\hat{1}^{-2}4^0 \widehat{P}^0_{23}) \mathcal{A}_{h\phi\phi}(\hat{2}^{+2}3^0 - \widehat{P}^0_{23})}{P^2_{23} - m^2} &= \frac{-i\kappa^2 \,\langle \hat{1}|\overline{\mathbf{P}}_4|\mathbf{q}]^2 \langle \mathbf{q}'|\overline{\mathbf{P}}_3|\hat{2}]^2}{4(P^2_{23} - m^2)\,[\hat{1}\mathbf{q}]^2 \langle \mathbf{q}'\hat{2}\rangle^2} \,, \\ i\, \frac{\mathcal{A}_{h\phi\phi}(\hat{1}^{-2}3^0 \widehat{P}^0_{24}) \mathcal{A}_{h\phi\phi}(\hat{2}^{+2}4^0 - \widehat{P}^0_{24})}{P^2_{24} - m^2} &= \frac{-i\kappa^2 \,\langle \hat{1}|\overline{\mathbf{P}}_3|\mathbf{q}]^2 \langle \mathbf{q}'|\overline{\mathbf{P}}_4|\hat{2}]^2}{4(P^2_{24} - m^2)\,[\hat{1}\mathbf{q}]^2 \langle \mathbf{q}'\hat{2}\rangle^2} \,. \end{aligned} \tag{11.136}$$

A convenient choice of auxiliary vectors is $\mathbf{q} = \mathbf{p}_2$ and $\mathbf{q}' = \mathbf{p}_1$, which leads to

$$\begin{aligned}\mathcal{A}_{hh\phi\phi}(1^{-2}2^{+2}3^04^0) &= -i\frac{\kappa^2}{4}\frac{\langle 1|\overline{\mathbf{P}}_3|2]^4}{\langle 12\rangle^2[12]^2}\left\{\frac{1}{P_{23}^2-m^2}+\frac{1}{P_{24}^2-m^2}\right\} \\ &= i\frac{\kappa^2}{16}\frac{\langle 1|\overline{\mathbf{P}}_3|2]^4}{\langle 12\rangle[12]}\frac{1}{(p_2\cdot p_3)(p_2\cdot p_4)}\,. \end{aligned} \tag{11.137}$$

The square of this amplitude can be related to that of photon-scalar gravitational scattering by

$$\left|\mathcal{A}_{hh\phi\phi}(1^{-2}2^{+2}3^04^0)\right|^2 = \left|\mathcal{A}_{\gamma\gamma\phi\phi}(1^-2^+3^04^0)\right|^2\left\{1-\frac{m^2\,s_{12}}{(s_{13}-m^2)(s_{14}-m^2)}\right\}. \tag{11.138}$$

In the limit of a graviton of small energy (i.e. a gravitational wave of long wavelength) and small deflection angle (i.e. at large impact parameter), the second factor in the right hand side becomes equal to 1, and we have

$$\left|\mathcal{A}_{hh\phi\phi}(1^{-2}2^{+2}3^04^0)\right|^2 \underset{\substack{\omega\ll m\\ \theta\ll 1}}{\approx} \left|\mathcal{A}_{\gamma\gamma\phi\phi}(1^-2^+3^04^0)\right|^2. \tag{11.139}$$

This implies that in this limit, the bending of a gravitational wave by a mass is the same as the bending of a light ray (but there are some differences beyond this limit).

Chapter 12

Lattice field theory

12.1 Discretization of space-time

We have seen earlier that the running coupling in an $SU(N)$ non-Abelian gauge theories decreases at large energy (provided the number of quark flavors is less than $11N/2$). The counterpart of asymptotic freedom is that the coupling increases towards lower energies, precluding the use of perturbation theory to study phenomena in this regime. Among such properties is that of color confinement, i.e. the fact that colored states cannot exist as asymptotic states. Instead the quarks and gluons arrange themselves into color neutral bound states, that can be mesons (e.g. pions, kaons) made of a quark and an antiquark or baryons (e.g. protons, neutrons) made of three quarks[1]. A legitimate question would be to determine the mass spectrum of the asymptotic states of QCD from its Lagrangian.

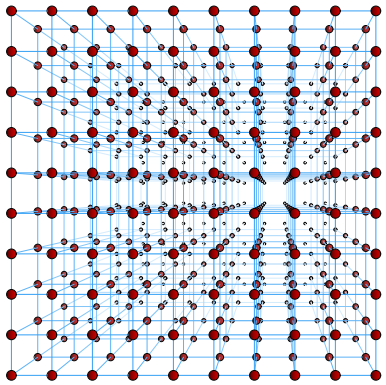
Since the perturbative expansion is not applicable for this type of problem, one would like to be able to attack it via some non-perturbative approach. By *non-perturbative*, we mean a method by which observables would directly be obtained to all orders in the coupling constant, without any expansion. One such method, known as *lattice field theory*, consists in discretizing space-time in order to evaluate numerically the path integral. The continuous space-time is replaced by a discrete grid of points, the simplest arrangement being a hyper-cubic lattice such as the one shown in the figure 12.1. The distance between nearest neighbor sites is called the lattice spacing, and usually denoted a. The lattice spacing, being the smallest distance that exists in this setup, therefore provides a natural ultraviolet regularization. Indeed, on a lattice of spacing a, the largest conjugate momentum is of order a^{-1}. Moreover, one usually uses periodic boundary conditions; if the lattice has N spacings in a given direction, then we have $\phi(x + Na) = \phi(x)$ for bosonic fields and $\phi(x + Na) = -\phi(x)$ for fermionic fields.

12.2 Scalar field theory

As an illustration of some of the issues involved in the discretization of a quantum field theory, let us consider a simple scalar field theory with a local interaction in ϕ^4, whose action in

[1] More exotic bound states made of four (tetraquarks) or five (pentaquarks) have also been speculated, but the experimental evidence for these states is so far not fully conclusive. Likewise, there may exist bound states without valence quarks, the *glueballs*.

Figure 12.1: *Discretization of Euclidean space-time on a hyper-cubic lattice (here shown in three dimensions).*



continuous space-time is

$$S = \int d^4x \left\{ -\frac{1}{2}\phi(x)\big(\partial_\mu\partial^\mu + m^2\big)\phi(x) - \frac{\lambda}{4!}\phi^4(x)\right\} . \tag{12.1}$$

A natural choice is to replace the integral over space-time by a discrete sum over the sites of the lattice, weighted by the volume a^4 of the elementary cells of the lattice,

$$a^4 \sum_{x\in \text{ lattice}} \underset{a\to 0}{\to} \int d^4x . \tag{12.2}$$

Then we replace the continuous function $\phi(x)$ by a discrete set of real numbers that live on the lattice nodes. For simplicity, we keep denoting $\phi(x)$ the value of the field on the lattice site x. The discretization of the mass and interaction terms is trivial, but the discretization of the derivatives that appear in the D'Alembertian operator is not unique. Using only two nearest neighbors, one may define forward or backward finite differences,

$$\begin{aligned} \nabla_{\text{F}} f(x) &\equiv \frac{f(x+a)-f(x)}{a} \\ \nabla_{\text{B}} f(x) &\equiv \frac{f(x)-f(x-a)}{a} , \end{aligned} \tag{12.3}$$

that both go to the continuum derivative in the limit $a \to 0$. However, unlike the continuous derivative, ∇_{F} and ∇_{B} are not anti-adjoint. Instead, assuming periodic boundary conditions, we have

$$\sum_{x\in \text{ lattice}} f(x)\Big(\nabla_{\text{F}} g(x)\Big) = -\sum_{x\in \text{ lattice}} \Big(\nabla_{\text{B}} f(x)\Big)\, g(x) . \tag{12.4}$$

In other words, $\nabla_{\text{F}}^\dagger = -\nabla_{\text{B}}$. From this, we may construct a self-adjoint discrete second derivative as follows:

$$\nabla_{\text{B}}\nabla_{\text{F}} f(x) = \frac{f(x+a)+f(x-a)-2f(x)}{a^2} \underset{a\to 0}{\to} f''(x) . \tag{12.5}$$

Thus, a self-adjoint discretization of the scalar Lagrangian leads to

$$S_{\text{lattice}} = a^4 \sum_{x \in \text{ lattice}} \Big\{ -\frac{1}{2}\phi(x)\big(\nabla_{B\,\mu}\nabla_F^{\mu} + m^2\big)\phi(x) - \frac{\lambda}{4!}\phi^4(x)\Big\} \,. \tag{12.6}$$

Let us make a few remarks concerning the errors introduced by the discretization. Firstly, the continuous spacetime symmetries (translation and rotation invariance) of the underlying theory are now reduced to the subgroup of the discrete symmetries of a cubic lattice. They are recovered in the limit $a \to 0$. Another source of discrepancy between the continuum and discrete theories is the dispersion relation that relates the energy and momentum of an on-shell particle. In the continuum theory, this relation is of course

$$E^2 = \mathbf{p}^2 + m^2 \,, \tag{12.7}$$

where $-\mathbf{p}^2$ is an eigenvalue of the Laplacian. In order to find its counterpart with the above discretization, we must determine the spectrum of the finite difference operator $\nabla_B \nabla_F$. On a lattice with $\mathcal{N}$ sites and periodic boundary conditions, its eigenfunctions are given by

$$\phi_k(x) \equiv e^{2i\pi \frac{kx}{\mathcal{N}a}} \qquad \text{with } k \in \mathbb{Z} \,, \quad -\tfrac{\mathcal{N}}{2} \le k \le \tfrac{\mathcal{N}}{2} \,. \tag{12.8}$$

The associated eigenvalue is

$$\lambda_k \equiv \frac{2}{a^2}\big(\cos\frac{2\pi k}{\mathcal{N}} - 1\big) = -\frac{4}{a^2}\,\sin^2\frac{\pi k}{\mathcal{N}} \,. \tag{12.9}$$

Thus, the one dimensional discrete analogue of the continuum $\mathbf{p}^2 + m^2$ is

$$m^2 + \frac{4}{a^2}\,\sin^2\frac{\pi k}{\mathcal{N}} \,. \tag{12.10}$$

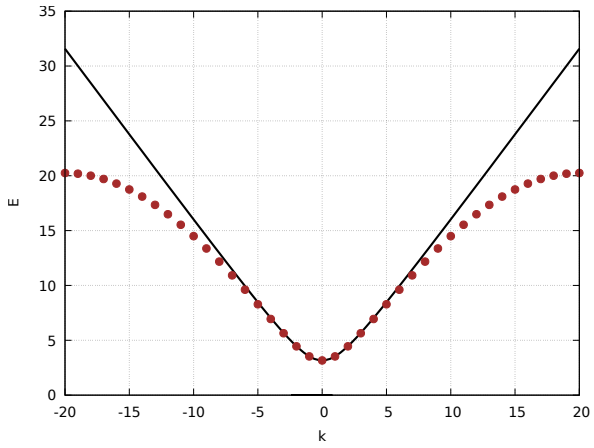
As long as $k \ll \mathcal{N}$, this agrees quite well with the continuum dispersion relation, but the agreement is not good for larger values of k. This discrepancy is illustrated in the figure 12.2. This mismatch does not improve by increasing the number of lattice points: only the center of the Brillouin zone has a dispersion relation that agrees with the continuum one. In order to mitigate this problem, one should choose the parameters of the lattice in such a way that the physically relevant scales correspond to values of k for which the distortion of the dispersion curve is small.

12.3 Gluons and Wilson action

Non-Abelian gauge theories pose an additional difficulty: since the local gauge invariance plays a central role in their properties, any attempt at discretizing gauge fields should preserve this symmetry. It turns out that there exists a discretization of the Yang-Mills action that goes to the continuum action in the limit where $a \to 0$, and has an exact gauge invariance. The main ingredient in this construction is eq. (4.140), that relates the trace of a Wilson loop along a small square,

$$[\Box]_{x;\mu\nu} \equiv U_\nu^\dagger(x)\, U_\mu^\dagger(x+\hat{\nu})\, U_\nu(x+\hat{\mu})\, U_\mu(x) \,, \tag{12.11}$$

Figure 12.2: *Discrepancy between the continuous (solid curve) and discrete (points) dispersion relations, on a one-dimensional lattice with* $N = 40$.



to the squared field strength. These elementary lattice Wilson loops are called *plaquettes*. In the fundamental representation, we have

$$\mathrm{tr}\left([\Box]_{x;\mu\nu}\right) = N - \frac{g^2 a^4}{4} F_a^{\mu\nu}(x) F^a_{\mu\nu}(x) + \mathcal{O}(a^6) \,. \tag{12.12}$$

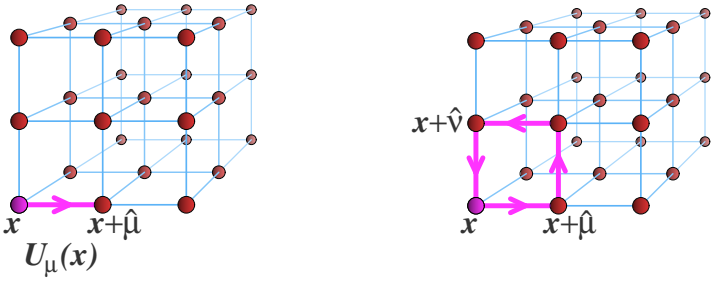
Note that, although the first two terms in the right hand side are real valued, the remainder (terms of order a^6 and beyond) may be complex. Therefore, it is convenient to take the real part of the trace of the Wilson loop in order to construct a real valued discrete action. By summing this equation over all the lattice points x and all the pairs of distinct directions (μ,ν), we obtain

$$a^4 \sum_{x\in \text{ lattice}} \left(-\frac{1}{4} F_a^{\mu\nu}(x) F^a_{\mu\nu}(x)\right) = \underbrace{\frac{N}{g^2} \sum_{x\in \text{ lattice}} \sum_{(\mu,\nu)} \left(N^{-1}\mathrm{tr}\left(\mathrm{Re}\,[\Box]_{x;\mu\nu}\right) - 1\right)}_{\text{Wilson action, denoted } \frac{1}{g^2} S_W[U]} + \mathcal{O}(a^2) \,. \tag{12.13}$$

Note that the error term of order a^6 becomes a term of order a^2 after summation over the lattice sites, since the number of sites grows like a^{-4} if the volume is held fixed. Thus, the sum of the traces of the Wilson loops over all the elementary plaquettes of the lattice provides a discretization of the Yang-Mills action. In this discrete formulation, the natural variables are not the gauge potentials $A^\mu(x)$ themselves, but the Wilson lines $U_\mu(x)$ that live on the edges of the lattice, called *link variables*. In this notation, x is the starting point and μ the direction of the Wilson line, as illustrated in the left panel of figure 12.3. The Wilson line oriented in the $-\hat{\mu}$ direction, i.e. starting at the point $x+\hat{\mu}$ and ending at the point x, is simply the Hermitic conjugate of $U_\mu(x)$. Under a local gauge transformation, the link variables are changed as follows:

$$U_\mu(x) \quad \to \quad \Omega^\dagger(x+\hat{\mu})\, U_\mu(x)\, \Omega(x) \,. \tag{12.14}$$

Figure 12.3: *Left: link variable. Right: plaquette on an elementary square of the lattice.*



The plaquette variable, shown in the right panel of figure 12.3, can then be obtained by multiplying four link variables, as indicated by eq. (12.11), and its trace is obviously invariant under the transformation of eq. (12.14).

At this stage, the discrete analogue of the path integral that gives the expectation value of a gauge invariant operator reads,

$$\langle \mathcal{O} \rangle = \int \prod_{x,\mu} dU_\mu(x)\ \mathcal{O}[U]\ \exp\left\{ i \frac{N}{g^2} \sum_x \sum_{(\mu,\nu)} \left(N^{-1} \mathrm{tr}\left(\mathrm{Re}\,[\Box]_{x;\mu\nu} \right) - 1 \right) \right\} . \quad (12.15)$$

Since there exists a left- and right-invariant[2] group measure $dU_\mu(x)$, the left hand side of this formula is gauge invariant. Moreover, it goes to the expectation value of the continuum theory in the limit of zero lattice spacing.

12.4 Monte-Carlo sampling

Thanks to the discretization, the path integral of the original theory is replaced by an ordinary integral over each of the link variables $U_\mu(x)$, whose number is finite. A non-perturbative answer could be obtained if one were able to evaluate these integrals numerically. However, because of the prefactor i inside the exponential in eq. (12.15), the integrand is a strongly oscillating function, whose numerical evaluation is practically impossible except on lattices with a very small number of sites. In order to be amenable to a numerical calculation, this integral must be transformed into an Euclidean one,

$$\langle \mathcal{O} \rangle_E = \int \prod_{x,\mu} dU_\mu(x)\ \mathcal{O}[U]\ \exp\left\{ \frac{N}{g^2} \sum_x \sum_{(\mu,\nu)} \left(N^{-1} \mathrm{tr}\left(\mathrm{Re}\,[\Box]_{x;\mu\nu} \right) - 1 \right) \right\} . \quad (12.16)$$

The exponential under the integral is now real-valued, and thus positive definite. Note that numerical quadratures such as Simpson's rule, are not practical for this problem, given the

[2] This means that:

$$\int dU\ f[U] = \int dU\ f[\Omega U] = \int dU\ f[U\Omega] .$$

Such a measure, known as the Haar measure, exists for compact Lie groups, such as $SU(N)$.

huge number of dimensions of the integral to be evaluated. For instance, for the 8-dimensional Lie group $SU(3)$, in 4 space-time dimensions, on a lattice with N^4 points, this dimension is $8 \times 4 \times N^4$. For $N = 32$, the path integral is thus transformed into a 2^{25}-dimensional ($2^{25} \sim 3.10^7$) ordinary integral. Instead, one views the exponential of the Wilson action as a probability distribution (up to a normalization constant) for the link variables, that may be sampled by a Monte-Carlo algorithm (e.g. the *Metropolis algorithm*) in order to estimate the integral.

In this approach, as long as one is evaluating the expectation value of gauge invariant observables, it is not necessary to fix the gauge in lattice QCD calculations. Gauge fixing is necessary when calculating non-gauge invariant quantities, such as propagators for instance. The Landau gauge is the most commonly used, because the Landau gauge condition is realized at the extrema of a functional of the link variables, However, the comparison between gauge-fixed lattice calculations and analytical calculations is very delicate, because of the existence of Gribov copies (the problem stems from the fact that the two setups may not select the same Gribov copy).

Although considering the Euclidean path-integral instead of the Minkowski one allows for a numerical evaluation by Monte-Carlo sampling, this leads to a serious limitation: only quantities that can be expressed as an Euclidean expectation value are directly calculable. Others could in principle be reached by an analytic continuation from imaginary to real time, but this turns out to be practically impossible numerically. For instance, the masses of hadrons are accessible to lattice QCD calculations (see the section 12.6 for an example), while scattering amplitudes cannot be calculated by this method.

12.5 Fermions

12.5.1 Discretization of the Dirac action

Consider now the Dirac action, whose expression in continuum space reads

$$S_D = \int d^4x\, \overline{\psi}(x) \left(i\gamma^\mu D_\mu - m \right) \psi(x) \,. \tag{12.17}$$

In the discretization, we assign a spinor $\psi(x)$ to each site of the lattice. Under a gauge transformation $\Omega(x)$, these spinors transform in the same way as in the continuous theory,

$$\psi(x) \;\to\; \Omega^\dagger(x)\,\psi(x) \quad , \qquad \overline{\psi}(x) \;\to\; \overline{\psi}(x)\,\Omega(x) \,. \tag{12.18}$$

The main difficulty in defining a discrete covariant derivative that transforms appropriately under a gauge transformation is that $\psi(x)$ and $\psi(x \pm a)$ transform differently when $\Omega(x)$ depends on space-time. This problem can be remedied by using a link variable between the point x and its neighbors. Like with the ordinary derivatives, one may define forward and backward discrete derivatives,

$$\begin{aligned} D^\mu_{F}\psi(x) &\equiv \frac{U^\dagger_\mu(x)\psi(x+a) - \psi(x)}{a} \,, \\ D^\mu_{B}\psi(x) &\equiv \frac{\psi(x) - U_\mu(x-a)\psi(x-a)}{a} \,, \end{aligned} \tag{12.19}$$

that both transform like a spinor at the point x. However, none of these two operators is anti-adjoint, and therefore they would not give a Hermitean Lagrangian density. This may be

achieved by using instead $\frac{1}{2}\left(D_F^\mu + D_B^\mu\right)$, which corresponds to a symmetric forward-backward difference

$$\frac{1}{2}\left(D_F^\mu + D_B^\mu\right)\psi(x) = \frac{U_\mu^\dagger(x)\psi(x+a) - U_\mu(x-a)\psi(x-a)}{2a} \,. \tag{12.20}$$

12.5.2 Fermion doublers

Let us now study how the dispersion relation of fermions is modified by this discretization. This can easily be done in the vacuum, i.e. by setting all the link variables to the identity. In this case, the eigenfunctions of the operator $\frac{1}{2}\left(D_F^\mu + D_B^\mu\right)$ are

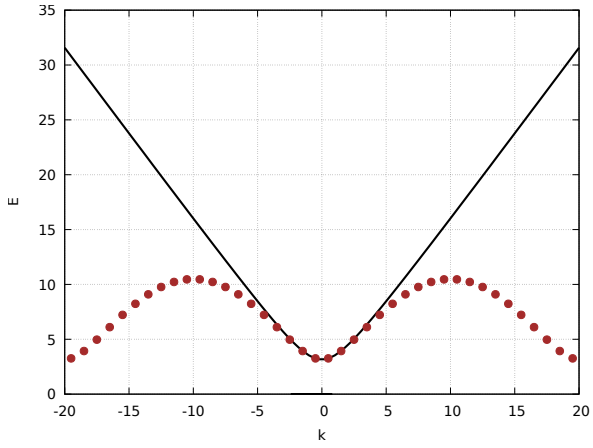
$$\psi_k(x) = e^{2i\pi\frac{(k+1/2)x}{Na}} \qquad \text{with } k \in \mathbb{Z}\,, \quad -\tfrac{N}{2} \le k \le \tfrac{N}{2}\,, \tag{12.21}$$

and the corresponding eigenvalue is

$$\lambda_k = \frac{i}{a}\,\sin\frac{2\pi(k+1/2)}{N}\,, \tag{12.22}$$

and the corresponding dispersion relation is $E^2 = |\lambda_k|^2 + m^2$. This dispersion relation is shown in the figure 12.4. Like in the bosonic case, the discrete dispersion relation agrees with the

Figure 12.4: *Discrepancy between the continuous (solid curve) and discrete (points) dispersion curves for fermions, on a one-dimensional lattice with* $N = 40$.



continuous one only for small enough k. However, the discrepancy at large k is now much more serious, because the discrete dispersion curve has another minimum at the edge of the Brillouin zone. This additional minimum indicates the existence of a second propagating mode of mass m. This spurious mode is called a *fermion doubler*. In d dimensions, the number of these fermionic modes is 2^d, while our goal was to have only one. This problem is quite serious, because it affects all quantities that depend on the number of quark flavors. In particular, this is the case of the running of the coupling constant, whenever quark loops are included.

12.5.3 Wilson term

Various modifications of the discretized Dirac action have been proposed to remedy the problem of fermion doublers. One of these modifications, known as the *Wilson term*, consists in adding to the Lagrangian the following term (written here for the direction μ),

$$-\frac{1}{2a}\,\overline{\psi}(x)\left[U_\mu^\dagger(x)\,\psi(x+a)+U_\mu(x-a)\,\psi(x-a)-2\psi(x)\right], \tag{12.23}$$

which is nothing but a D'Alembertian (or a Laplacian in the Euclidean theory) constructed with covariant derivatives. The corresponding operator in the continuum theory is

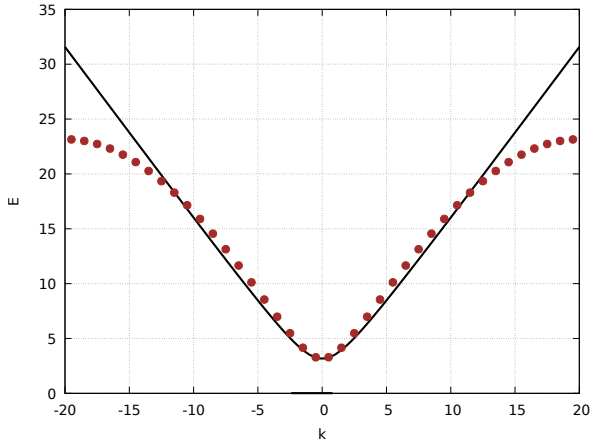
$$\frac{a}{2}\,\overline{\psi}\left(D_\mu D^\mu\right)\psi\,. \tag{12.24}$$

Note that the denominator in eq. (12.23) has a single power of the lattice spacing a, hence the prefactor a in the previous equation. Therefore, this term goes to zero in the limit $a\to 0$, and it should have no effect in the continuum limit. In the absence of gauge field ($U_\mu(x)\equiv 1$), the functions of eq. (12.21) are still eigenfunctions after adding the Wilson term, but with modified eigenvalues,

$$\lambda_k=\frac{i}{a}\,\sin\frac{2\pi\,(k+1/2)}{\mathcal{N}}+\frac{1}{a}\left(1-\cos\frac{2\pi\,(k+1/2)}{\mathcal{N}}\right). \tag{12.25}$$

Thus, the Wilson term does not modify the spectrum at small k, but lifts the spurious minimum

Figure 12.5: *Discrepancy between the continuous (solid curve) and discrete (points) dispersion curves in the fermionic case, on a one-dimensional lattice with $\mathcal{N}=40$, after inclusion of the Wilson term.*



that existed at the edge of the Brillouin zone, as shown in the figure 12.5. Roughly speaking, the Wilson term gives a mass of order a^{-1} to the fermion doublers, making them decouple from the rest of the degrees of freedom when $a\to 0$.

However, the Wilson term has an important drawback: there is no Dirac matrix γ^μ in eqs. (12.23) and (12.24) since the Lorentz indices are contracted directly between the two covariant derivatives. Therefore, the Wilson term –like an ordinary mass term– breaks explicitly the chiral symmetry of the Dirac Lagrangian in the case of massless fermions. The fermion doublers are in fact intimately related to chiral symmetry. Without the Wilson term, lattice QCD with massless quarks has an exact chiral symmetry unbroken by the lattice regularization, and therefore there cannot be a chiral anomaly. In fact, this absence of anomaly is precisely due to a cancellation of anomalies among the multiple copies (the doublers) of the fermion modes. This argument is completely general and not specific to the Wilson term: any mechanism that lifts the degeneracy among the doublers will spoil the anomaly cancellation and thus break chiral symmetry. For this reason, the study of phenomena related to chiral symmetry is always delicate in lattice QCD.

12.5.4 Evaluation of the fermion path integral

The path integral representation for fermions uses anti-commuting Grassmann variables. However, such variables are not representable as ordinary numbers in a numerical implementation. To circumvent this difficulty, one exploits the fact that the Dirac action is quadratic in the fermion fields (this remains true after adding the Wilson term to remove the fermion doublers). Therefore, the path integral over the fermion fields can be done exactly. In addition to the fermion fields contained in the action, there may be ψ's and $\overline{\psi}$'s (in equal numbers) in the operator whose expectation value is being evaluated. The result of such a fermionic path integral is given by

$$\int \left[D\psi D\overline{\psi}\right] \left(\psi(x_1)\overline{\psi}(x_2)\right) e^{iS_D[\psi,\overline{\psi}]} = S(x_1,x_2) \times \det\left(i\gamma^\mu D_\mu - m\right); \qquad (12.26)$$

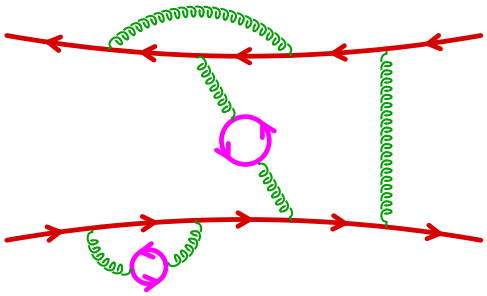
where $S(x_1,x_2)$ is the inverse of the Dirac operator $i\not{D} - m$ between the points x_1 and x_2. When there is more than one $\psi\overline{\psi}$ pair in the operator, one must sum over all the ways of connecting the ψ's and the $\overline{\psi}$'s by the fermion propagators $S(x,y)$. The same can be done in the lattice formulation. In this case, the Dirac operator is simply a (very large) matrix that depends on the configuration of link variables. Therefore one needs the inverse of this matrix, and its determinant.

In eq. (12.26), the determinant provides closed quark loops, while the propagator $S(x_1,x_2)$ connects the external points of the operator under consideration. This observation, illustrated in the figure 12.6, clarifies the meaning of the *quenched approximation*, in which the determinant of the Dirac operator is replaced by 1. This approximation, motivated primarily by the computational difficulty of evaluating the Dirac determinant, was widely used in lattice QCD computations until advances in algorithms and computer hardware made it unnecessary. Note that, although quark loops are not included in the quenched approximation, gluon loops are present to their full extent. In contrast, lattice QCD calculations that include the Dirac determinant, and thus the effect of quark loops, are said to use *dynamical fermions*.

12.6 Hadron mass determination on the lattice

Let us consider a hadronic state $|h\rangle$. Any operator $\mathcal{O}$ that carries the same quantum numbers as this hadron leads to a non-zero matrix element $\langle h|\mathcal{O}|0\rangle$. The vacuum expectation value of the

Figure 12.6: *Illustration of the two types of quark contributions. In red: quark propagators (i.e. inverse of the Dirac operator) that connect the ψ's and $\overline{\psi}$'s in the operator being evaluated. In purple: quark loops coming from the determinant of the Dirac operator.*



product of two $\mathcal{O}$ at different times 0 and T can be rewritten as follows,

$$\begin{aligned}
\langle 0|\mathcal{O}^\dagger(0)\mathcal{O}(T)|0\rangle &= \sum_n \langle 0|\mathcal{O}^\dagger(0)|\Psi_n\rangle\langle\Psi_n|\mathcal{O}(T)|0\rangle \\
&= \sum_n \langle 0|\mathcal{O}^\dagger(0)|\Psi_n\rangle\langle\Psi_n|\mathcal{O}(0)|0\rangle\, e^{-M_n T} \\
&= \sum_n \left|\langle\Psi_n|\mathcal{O}(0)|0\rangle\right|^2 e^{-M_n T}\,. \qquad (12.27)
\end{aligned}$$

In the first equality, we have inserted a complete basis of eigenstates of the QCD Hamiltonian, and the second equality follows from the fact that $\langle\Psi_n|$ is an eigenstate of rest energy M_n (there is no factor i inside the exponential because of the Euclidean time used in lattice QCD). The sum in the last equality receives non-zero contributions from all the states Ψ_n that possess the quantum numbers carried by the operator $\mathcal{O}$. However, taking the limit $T \to \infty$ selects the one among these eigenstates that has the smallest mass. This observation can be turned into a method to determine hadron masses in lattice QCD:

1. Choose an operator $\mathcal{O}$ that has the quantum numbers of the hadron of interest. The choice of the operator is not crucial, as long as the overlap $\langle h|\mathcal{O}|0\rangle$ is not zero. However, eq. (12.27) suggests that a better, i.e. less noisy with limited statistics, result may be obtained by trying to maximize this overlap.

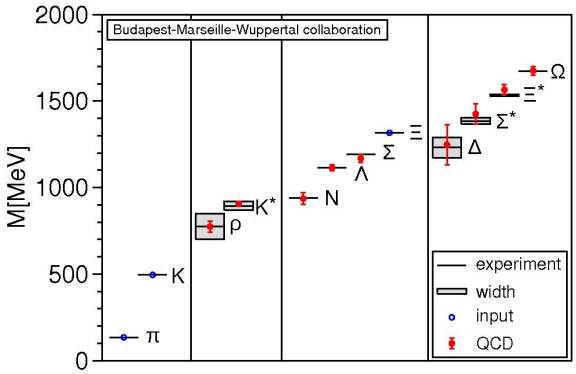
2. Evaluate the vacuum expectation value of $\mathcal{O}^\dagger(0)\mathcal{O}(T)$ by Monte-Carlo sampling, as a function of T.

3. Fit the large T tail of this expectation value. The slope of the exponential gives the mass of the lightest hadron that possesses these quantum numbers.

The discretized QCD Lagrangian contains several dimensionful parameters: the lattice spacing a and the quark masses m_f (one for each quark flavor), whose values need to be fixed before

novel predictions can be made. One must choose (at least) an equal number of physical quantities that are known experimentally. Their computed values depend on a, m_f, and one should adjust these parameters so that they match the experimental values. After this has been done, quantities computed in lattice QCD do not contain any free parameter anymore and are thus predictions. The figure 12.7 shows results of the determination of hadron masses using lattice

Figure 12.7: *Hadron mass determination from lattice QCD. Blue: masses used as input in order to set the lattice parameters. Red: predictions of lattice QCD. Boxes: experimental values.*



QCD.

12.7 Wilson loops and confinement

12.7.1 Strong coupling expansion

While perturbation theory is an expansion in powers of g^2, it is possible to use the lattice formulation of a Yang-Mills theory in order to perform an expansion in powers of the quantity $\beta \equiv g^{-2}$ that appears as a prefactor in the Wilson action. This is called a *strong coupling expansion*, since it becomes exact in the limit of infinite coupling.

This expansion produces integrals over the gauge group such as

$$\int dU\, U_{i_1 j_1} \cdots U_{i_n j_n}\, U^\dagger_{k_1 l_1} \cdots U^\dagger_{k_m l_m} \,. \tag{12.28}$$

The simplest of these integrals,

$$\int dU = 1 \tag{12.29}$$

is simply a choice of normalization of the invariant group measure. From the unitarity of the

group elements, one then obtains[3]

$$\int dU\, U_{ij} U^\dagger_{kl} = \frac{1}{N}\,\delta_{jk}\delta_{il}\ . \tag{12.30}$$

In these integrals, the link variables on different edges of the lattice are independent variables, and there is a separate integral for each of them. This is completely general: integrals of the form (12.28) are non-zero only if the integrands contains an equal number of U's and $U^\dagger$'s, i.e. for $n = m$. Therefore, each link variable U that appears in such a group integral must be matched by a corresponding $U^\dagger$. For instance, the group integral of the Wilson loop defined on an elementary plaquette is zero,

$$\int \prod_{x,\mu} dU_\mu(x)\ \mathrm{tr}\Big(\underbrace{U^\dagger_\nu(x)\,U^\dagger_\mu(x+\hat\nu)\,U_\nu(x+\hat\mu)\,U_\mu(x)}_{[\Box]_{x;\mu\nu}}\Big) = 0\ , \tag{12.31}$$

because the four link variables live on four distinct edges of the lattice. In contrast, the integral of the trace of a plaquette time the trace of the conjugate plaquette is non-zero:

$$\int \prod_{x,\mu} dU_\mu(x)\ \Big(\mathrm{tr}\,[\Box]_{x;\mu\nu}\Big)\ \Big(\mathrm{tr}\,[\Box]^\dagger_{x;\mu\nu}\Big) = 1\ . \tag{12.32}$$

Using these results, we can calculate to order β the expectation value of the trace of a plaquette:

$$\begin{aligned}\big\langle \mathrm{tr}\,[\Box]_{x;\mu\nu}\big\rangle &\equiv \frac{\displaystyle\int \prod_{x,\mu} dU_\mu(x)\ \Big(\mathrm{tr}\,[\Box]_{x;\mu\nu}\Big)\ \exp\Big\{\beta N \sum_{y;\rho\sigma}\Big(N^{-1}\mathrm{tr}\,\mathrm{Re}\,[\Box]_{y;\rho\sigma}-1\Big)\Big\}}{\displaystyle\int \prod_{x,\mu} dU_\mu(x)\ \exp\Big\{\beta N \sum_{y;\rho\sigma}\Big(N^{-1}\mathrm{tr}\,\mathrm{Re}\,[\Box]_{y;\rho\sigma}-1\Big)\Big\}} \\ &= \frac{\beta}{2}+\mathcal{O}(\beta^2)\ .\end{aligned} \tag{12.33}$$

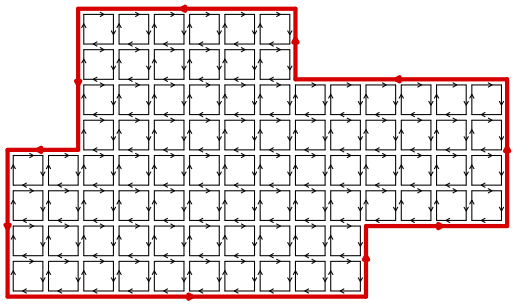
Consider now the trace of a more general Wilson loop along a path γ (planar, to simplify the discussion). Each U and $U^\dagger$ in the Wilson loop must be compensated by a link variable coming from the β expansion of the exponential of the Wilson action. The lowest order term in β corresponds to a minimal tiling of the Wilson loop by elementary plaquettes, as illustrated in the figure 12.8. The corresponding contribution is

$$\langle \mathrm{tr}\, W_\gamma\rangle = \left(\frac{\beta}{2}\right)^{\mathrm{Area}\,(\gamma)} + \cdots\ , \tag{12.34}$$

where the dots are terms of higher order in β (that can be constructed from non-minimal tilings of the contour γ, such that all the U's and $U^\dagger$'s are still paired appropriately).

[3] For $SU(2)$, one may parameterize the group elements in the fundamental representation by $U = \theta_0 + 2i\,\theta_a\,t^a_f$ with $\theta_0^2+\theta_1^2+\theta_2^2+\theta_3^2 = 1$, and the invariant group measure normalized according to eq. (12.29) is $dU = d\theta_1\,d\theta_2\,d\theta_3/(\pi^2\sqrt{1-\theta^2})$ (with $\theta_0 = \sqrt{1-\theta^2}$). By using this measure and the Fierz identity satisfied by the generators t^a_f, an explicit calculation leads easily to eq. (12.30).

Figure 12.8: *Tiling of a closed loop by elementary plaquettes.*



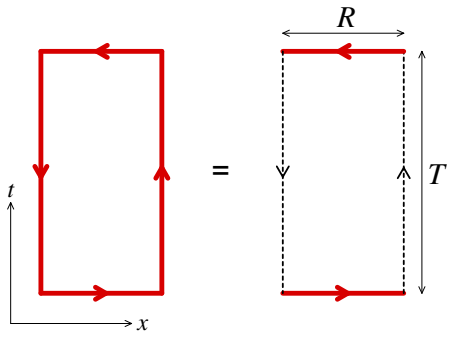
12.7.2 Heavy quark potential

Let us consider now a rectangular loop, with an extent R in the spatial direction 1 and an extent T in the Euclidean time direction 4. The previous result indicates that the expectation value of the trace of the corresponding Wilson loop has the following form,

$$\langle \mathrm{tr}\, W_\gamma \rangle \sim e^{-\sigma R T} + \cdots \,, \tag{12.35}$$

where σ is a constant. Although it is gauge invariant, this expectation value is easier to interpret in an axial gauge where $A_4 \equiv 0$. Indeed, in this gauge, the Wilson loop receives only contributions from gauge links along the spatial direction, as shown in the figure 12.9. Note that the

Figure 12.9: *Rectangular Wilson loop in the $A_4 \equiv 0$ gauge.*



remaining Wilson lines are precisely those that are needed to make a non-local gauge invariant operator with a quark at $x = R$ and an antiquark at $x = 0$,

$$\mathcal{O}_{q\bar{q}}(t) \equiv \overline{\psi}(t,0)\; W_{[0,R]}\; \psi(t,R) \,, \tag{12.36}$$

where $W_{[0,R]}$ is a (spatial) Wilson line going from (t,R) to $(t,0)$. Consider now the vacuum expectation value $\langle 0|\mathcal{O}^\dagger_{q\overline{q}}(0)\mathcal{O}_{q\overline{q}}(T)|0\rangle$. In this expectation value, the fermionic path integral produces two quark propagators that connect the ψ's to the $\overline{\psi}$'s. However, in the limit of infinite quark mass, the quarks propagate on straight lines at constant velocity and their propagator reduces to a Wilson line along this trajectory. For the propagation between $(0,x)$ and (T,x), this is a temporal Wilson line, that reduces to the identity in the $A_4 = 0$ gauge (represented by the dotted lines in the figure 12.9). Thus, we have

$$\langle \mathrm{tr}\, W_\gamma\rangle \propto \lim_{M\to\infty} \langle 0|\mathcal{O}^\dagger_{q\overline{q}}(0)\mathcal{O}_{q\overline{q}}(T)|0\rangle \,. \tag{12.37}$$

Inserting a complete basis of eigenstates of the Hamiltonian in the right hand side of eq. (12.37) and taking the limit $T\to\infty$, we find a result dominated by the quark-antiquark state of lowest energy E_0,

$$\lim_{T\to\infty} \langle 0|\mathcal{O}^\dagger_{q\overline{q}}(0)\mathcal{O}_{q\overline{q}}(T)|0\rangle = \left|\langle 0|\mathcal{O}^\dagger_{q\overline{q}}(0)|\Psi_0\rangle\right|^2 e^{-E_0 T} \,. \tag{12.38}$$

Moreover, in the limit of large mass, the energy E_0 of this state is dominated by the potential energy $V(R)$ between the quark and the antiquark (the quark and antiquark are non-relativistic, and their kinetic energy behaves as $\mathbf{P}^2/2M \to 0$),

$$\lim_{M,T\to\infty} \langle 0|\mathcal{O}^\dagger_{q\overline{q}}(0)\mathcal{O}_{q\overline{q}}(T)|0\rangle = \left|\langle 0|\mathcal{O}^\dagger_{q\overline{q}}(0)|\Psi_0\rangle\right|^2 e^{-V(R)\,T} \tag{12.39}$$

By comparing this result with that of the strong coupling expansion, eq. (12.35), we conclude that

$$V(R) = \sigma R \,. \tag{12.40}$$

This linear potential indicates that the force between the quark and antiquark is constant at large distance, in sharp contrast with a Coulomb potential in electrodynamics. This is a consequence of the *color confinement* property of QCD.

Chapter 13

Quantum field theory at finite temperature

13.1 Canonical thermal ensemble

13.1.1 Motivations

Historically, the main realm of developments and applications of Quantum Field Theory has been high energy particle physics. This corresponds to situations where the background is the vacuum, only perturbed by the presence of a few excitations whose interactions one aims at studying. Consequently, most of the QFT tools we have encountered so far are adequate for calculating transition amplitudes between pure states that contain only a few particles.

However, there are interesting physical problems that depart from this idealized situation. For instance, in the early universe, particles are believed to be in thermal equilibrium and form a hot and dense plasma. The typical energy of a particle in this thermal bath is of the order of the temperature[1], which implies that this surrounding medium may have an influence on all processes whose energy scale is comparable or lower. Thus, these problems contain some element of many body physics that was not present in applications of QFT to scattering reactions.

Another class of problems where many body effects are important is condensed matter physics. When studied at some sufficiently large distance scale, where the atomic discreteness is no longer important, these problems may be described in terms of (non relativistic) quantum fields where collective effects are usually important.

13.1.2 Canonical ensemble

Usually, the system one would like to study is a little part of a much larger system (this is quite obvious in the case of the early universe, but is also generally true in condensed matter physics).

[1] In this chapter, we extend the natural system of units we have used so far to also set the Boltzmann constant $k_B = 1$. Therefore, the temperature has the dimension of an energy.

Thus, its energy and other conserved quantities are not fixed. Instead, they fluctuate due to exchanges with the surroundings, that play the role of a thermal reservoir. The appropriate statistical ensemble for describing this situation is the (grand) canonical ensemble, in which the system is described by the following density operator

$$\rho \equiv \exp\Big\{ -\beta\, \mathcal{H} \Big\} \,, \tag{13.1}$$

where $\beta = T^{-1}$ is the inverse temperature and $\mathcal{H}$ is the Hamiltonian. Given an operator $\mathcal{O}$, one is usually interested in calculating its expectation value in the above statistical ensemble,

$$\big\langle \mathcal{O} \big\rangle \equiv \frac{\mathrm{Tr}\,(\rho\,\mathcal{O})}{\mathrm{Tr}\,(\rho)} \,. \tag{13.2}$$

13.1.3 Expression on a basis of eigenstates

Let us span the Hilbert space by states $\big|n\big\rangle$ that are eigenstates of $\mathcal{H}$,

$$\mathcal{H}\,\big|n\big\rangle = E_n\,\big|n\big\rangle \,. \tag{13.3}$$

In terms of these states, the trace of $\rho\,\mathcal{O}$ can be represented as follows

$$\mathrm{Tr}\,\big(\rho\,\mathcal{O}\big) = \sum_n e^{-\beta\,E_n}\,\big\langle n\big|\mathcal{O}\big|n\big\rangle \,. \tag{13.4}$$

From this representation, it is easy to see that the zero temperature limit selects the state of lowest energy, i.e. the ground state of the Hamiltonian. Assuming that this state is non-degenerate, this corresponds to a vacuum expectation value:

$$\lim_{T\to 0} \mathrm{Tr}\,\big(\rho\,\mathcal{O}\big) = \big\langle 0\big|\mathcal{O}\big|0\big\rangle \,. \tag{13.5}$$

In this sense, eq. (13.2) should be viewed as an extension of the formalism we already know, rather than something entirely different. In this chapter, we discuss various aspects of these thermal averages, starting with the necessary extensions to the formalism in order to perform their perturbative calculation.

13.2 Finite-T perturbation theory

13.2.1 Naive approach

The extension of ordinary perturbation theory to calculate expectation values such as (13.2) is usually called *Quantum Field Theory at finite temperature*. A first approach for evaluating such an expectation value could be to use the representation of the trace provided by eq. (13.4), and a similar formula for the denominator, which would fall back to the perturbative rules we already know (since the temperature and chemical potential appear only in the form of numerical prefactors). Note however a peculiarity of the matrix elements that appear in eq. (13.4): $\big\langle n\big|$ and $\big|n\big\rangle$ are identical states since they come from a trace (they are both *in*-states, since ρ defines

the initial state of the system). This is a bit different from the transition amplitudes that enter in scattering cross-sections, where the matrix elements are between an *in*-state and an *out*-state. The perturbative rules to compute these in-in expectation values are provided by the Schwinger-Keldysh formalism introduced in the section 1.14.5.

A difficulty with this naive approach is that the number of states that contribute significantly to the sum in eq. (13.4) is large at high temperature, especially when the temperature is large compared to the masses of the fields (and even more so with massless particles like photons). In fact, it is possible to encapsulate the sum over the eigenstates $|n\rangle$ and the canonical weight of these states $\exp(-\beta E_n)$ directly into the Schwinger-Keldysh rules, by a modest modification of its propagators.

13.2.2 Thermal time contour

To mimic closely the derivation of the Feynman rules at zero temperature, let us consider an observable made of the time-ordered product of elementary fields:

$$\mathcal{O} \equiv T\,\phi(x_1)\cdots\phi(x_n)\ . \tag{13.6}$$

Each Heisenberg representation field ϕ can be related to a field in the interaction representation by

$$\phi(x) = U(-\infty, x^0)\,\phi_{in}(x)\,U(x^0, -\infty)\ , \tag{13.7}$$

where U is the *time evolution* operator defined by:

$$U(t_2, t_1) \equiv T\,\exp i\int_{t_1}^{t_2} dx^0 d^3\mathbf{x}\ \mathcal{L}_I(\phi_{in}(x))\ , \tag{13.8}$$

with $\mathcal{L}_I$ the interaction term in the Lagrangian. Thanks to eq. (13.7), we remove all the dependence of the field ϕ on the interactions, that are now relegated to the evolution operator where it can easily be Taylor expanded.

In the canonical ensemble at non-zero temperature, there is another source of dependence on the interactions, hidden in the Hamiltonian inside the density operator. Indeed, for the system to be in statistical equilibrium, the canonical density operator should be defined with the same Hamiltonian as the one that drives the time evolution, i.e. a Hamiltonian that also contains the interactions of the system[2]. If we decompose the full Hamiltonian as $\mathcal{H} \equiv \mathcal{H}_0 + \mathcal{H}_I$, we have

$$e^{-\beta\mathcal{H}} = e^{-\beta\mathcal{H}_0}\ \underbrace{T\,\exp i\int_{-\infty}^{-\infty-i\beta} dx^0 d^3\mathbf{x}\ \mathcal{L}_I(\phi_{in}(x))}_{U(-\infty-i\beta,-\infty)}\ .$$

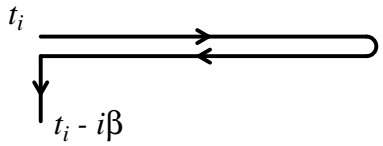
This formula can be proven by noticing that right and left hand sides are equal for $\beta = 0$, and by checking that their derivatives with respect to β are also equal (for this, we use the fact that the derivative of the time evolution with respect to its final time is known).

[2] An alternative point of view is to decide that ρ is the density operator of the system at $x^0 = -\infty$. There, we may turn off adiabatically the interactions, and therefore use only the free Hamiltonian inside ρ. In this section, we derive the formalism for an initial equilibrium state specified at a finite time $x^0 = t_i$.

From the previous formulas, we can write

$$e^{-\beta \mathcal{H}} \, T\, \phi(x_1)\cdots\phi(x_n) = \\ = e^{-\beta \mathcal{H}_0} \, P\, \phi_{in}(x_1)\cdots\phi_{in}(x_n) \, \exp i\int_{\mathcal{C}} dx^0 d^3\mathbf{x}\, \mathcal{L}_I(\phi_{in}(x)) \,, \qquad (13.9)$$

where the symbol P indicates a *path ordering*, and where the time integration contour is $\mathcal{C} = [t_i, +\infty] \cup [+\infty, t_i] \cup [t_i, t_i - i\beta]$:

$\mathcal{C}$ = 

. (13.10)

In this contour, t_i is the time at which the system is prepared in thermal equilibrium. As we shall see shortly, all observables are independent of this time, which physically means that a system in equilibrium has no memory of when it was put in equilibrium. Note that in eq. (13.9), the times $x_1^0, \cdots, x_n^0$ are understood to be on the upper branch of the contour.

13.2.3 Generating functional

The time contour (13.10) is very similar to the contour of the figure (1.1), with the addition of a vertical part that captures the interactions hidden in the density operator. Since we had to extend the real time axis into the contour $\mathcal{C}$, it is natural to extend also the observable of eq. (13.6) to allow the operators to be located anywhere on $\mathcal{C}$, with a path ordering instead of a time ordering,

$$\mathcal{O} \equiv P\, \phi(x_1)\cdots\phi(x_n) \,. \qquad (13.11)$$

The expectation values of these operators can be encapsulated in the following generating functional,

$$Z[j] \equiv \frac{\mathrm{Tr}\left(\rho \;\; P\, \exp\left\{ i\int_{\mathcal{C}} d^4x\, j(x)\phi(x)\right\}\right)}{\mathrm{Tr}\,(\rho)} \,, \qquad (13.12)$$

where the fictitious source $j(x)$ lives on the contour $\mathcal{C}$. In order to bring this generating functional to a useful form, we can follow very closely the derivation of the section 1.5.2, by first pulling out a factor that contains the interactions, and by rearranging the ordering of the free factor with two applications of the Campbell-Baker-Hausdorff formula. This leads to

$$Z[j] = \exp i\int_{\mathcal{C}} d^4x\, \mathcal{L}_I\left(\frac{\delta}{i\delta j(x)}\right) \;\; \exp\Big\{ -\frac{1}{2}\int_{\mathcal{C}} d^4x d^4y\, j(x)\, j(y)\, G^0(x,y)\Big\} \,, \qquad (13.13)$$

where the free propagator $G^0(x,y)$, defined on the contour $\mathcal{C}$, is given by

$$G^0(x,y) \equiv \frac{\mathrm{Tr}\left(e^{-\beta\,\mathcal{H}_0}\, P\, \phi_{in}(x)\phi_{in}(y)\right)}{\mathrm{Tr}\left(e^{-\beta\,\mathcal{H}_0}\right)} \,. \qquad (13.14)$$

13.2.4 Expression of the free propagator

In order to calculate the free propagator, we need the free Hamiltonian expressed in terms of creation and annihilation operators[3],

$$\mathcal{H}_0 = \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_\mathbf{p}}\, E_\mathbf{p}\; a^\dagger_{\mathbf{p},\text{in}} a_{\mathbf{p},\text{in}}\;, \tag{13.15}$$

and the canonical commutation relation of the latter:

$$\left[a_{\mathbf{p},\text{in}}, a^\dagger_{\mathbf{p}',\text{in}}\right] = (2\pi)^3\, 2E_\mathbf{p}\, \delta(\mathbf{p}-\mathbf{p}')\;. \tag{13.16}$$

From this, we get

$$\begin{aligned}
\left[e^{-\beta\mathcal{H}_0}, a_{\mathbf{p},\text{in}}\right] &= e^{-\beta\mathcal{H}_0}\left(1-e^{-\beta E_\mathbf{p}}\right) a_{\mathbf{p},\text{in}} \\
\text{Tr}\left(e^{-\beta\mathcal{H}_0}\, a_{\mathbf{p},\text{in}}\right) &= 0 \\
\text{Tr}\left(e^{-\beta\mathcal{H}_0}\, a^\dagger_{\mathbf{p},\text{in}} a_{\mathbf{p}',\text{in}}\right) &= (2\pi)^3\, 2E_\mathbf{p}\, n_B(E_\mathbf{p})\, \delta(\mathbf{p}-\mathbf{p}')\;,
\end{aligned} \tag{13.17}$$

where n_B is the Bose-Einstein distribution:

$$n_B(E) \equiv \frac{1}{e^{\beta E}-1}\;. \tag{13.18}$$

This leads to the following formula for the free propagator:

$$\begin{aligned}
G^0(x,y) = \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_\mathbf{p}} \Big[&\left(\theta_c(x^0-y^0)+n_B(E_\mathbf{p})\right) e^{-ip\cdot(x-y)} \\
&+\left(\theta_c(y^0-x^0)+n_B(E_\mathbf{p})\right) e^{+ip\cdot(x-y)}\Big]\;,
\end{aligned}$$

where θ_c generalizes the step function to the contour $\mathcal{C}$ (i.e. $\theta_c(x^0-y^0)$ is non-zero if x^0 is posterior to y^0 according to the contour ordering). This expression of the propagator generalizes to a non-zero temperature the formula (1.82) (the Bose-Einstein distribution goes to zero when $T \to 0$). Let us postpone a bit the calculation of the propagator in momentum space. For now, we just note the following rules for the perturbative expansion in coordinate space:

1. Draw all the graphs (with vertices corresponding to the interactions of the theory under consideration) that connect the n points of the observable. Graphs containing disconnected subgraphs should be ignored. Each graph should be weighted by its symmetry factor.
2. Each line of a graph brings a free propagator $G^0(x,y)$.
3. Each vertex brings a factor $-i\lambda$. The space-time coordinate of this vertex is integrated out, but the time integration runs over the contour $\mathcal{C}$.

Thus, the only differences with the zero temperature Feynman rules are the explicit form of the free propagator, and the fact the time integrations are over the contour $\mathcal{C}$ instead of the real axis.

[3] We can drop the zero point energy here. It would simply multiply the density operator by a constant factor, that would be canceled since all expectation values are normalized by the factor $1/\text{Tr}(\rho)$.

13.2.5 Kubo-Martin-Schwinger symmetry

The canonical density operator $\exp(-\beta H)$ can be viewed as an evolution operator for an imaginary time shift, which implies the following formal identity

$$e^{-\beta H}\,\phi(x^0-i\beta,\mathbf{x})\,e^{\beta H} = \phi(x^0,\mathbf{x})\;. \tag{13.19}$$

Let us now consider the following correlator

$$\mathcal{G}(t_i,\cdots) \equiv \mathrm{Tr}\left(e^{-\beta H}\,\mathrm{P}\,\phi(t_i,\mathbf{x})\cdots\right)\,, \tag{13.20}$$

that contains a field whose time argument is the initial time t_i (the other fields it contains need not be specified in this discussion). Since t_i is the "smallest" time on the contour $\mathcal{C}$, the operator that carries it is placed to the rightmost position by the path ordering. Thus, we have

$$\mathcal{G}(t_i,\cdots) = \mathrm{Tr}\left(e^{-\beta H}\left[\mathrm{P}\cdots\right]\phi(t_i,\mathbf{x})\right)\,, \tag{13.21}$$

where the path ordering now applies only to the remaining (unwritten) fields. Using the cyclic invariance of the trace and eq. (13.19), we then get

$$\begin{aligned}\mathcal{G}(t_i,\cdots) &= \mathrm{Tr}\left(e^{-\beta H}\,\phi(t_i-i\beta,\mathbf{x})\left[\mathrm{P}\cdots\right]\right)\\ &= \mathrm{Tr}\left(e^{-\beta H}\,\mathrm{P}\,\phi(t_i-i\beta,\mathbf{x})\cdots\right)\\ &= \mathcal{G}(t_i-i\beta,\cdots)\,,\end{aligned} \tag{13.22}$$

where in the second line we have used the fact that $t_i - i\beta$ is the "latest" time on the contour $\mathcal{C}$ in order to reinclude the operator carrying it inside the path ordering. This equality is one of the forms of the *Kubo-Martin-Schwinger (KMS) symmetry*: all bosonic time ordered correlators take identical values at the two endpoints of the contour $\mathcal{C}$. Note that, although we have singled out the first field in the correlator, this identity applies equally to all the fields.

The KMS symmetry is very closely tied to the fact that the system is in thermal equilibrium, since it is satisfied only when the density operator is the canonical equilibrium one. One of its consequences is that all the correlation functions are independent of the initial time t_i. In order to prove this assertion, let us first note that the free propagator satisfies the KMS symmetry, and does not contain t_i explicitly. A generic Feynman graph leads to time integrations that have the following structure:

$$\mathcal{G}(x_1,\cdots,x_n) = \int_{\mathcal{C}} dy_1^0\cdots dy_p^0\;\mathcal{F}(y_1^0,\cdots,y_p^0\,|\,x_1,\cdots,x_n)\;. \tag{13.23}$$

(We assume that the integrals over the positions at every vertex have already been performed.) Since the free propagator does not depend on t_i, the derivative of the integral with respect to t_i comes only from the endpoints of the integration contour, and we can write

$$\begin{aligned}\frac{\partial\mathcal{G}(x_1,\cdots,x_n)}{\partial t_i} &= \sum_{i=1}^{p}\int_{\mathcal{C}}\prod_{j\neq i}dy_j^0\,\Big[\mathcal{F}(\cdots,y_i^0=t_i,\cdots|\,x_1,\cdots,x_n)\\ &\qquad\qquad -\mathcal{F}(\cdots,y_i^0=t_i-i\beta,\cdots|\,x_1,\cdots,x_n)\Big]\\ &= 0\;.\end{aligned} \tag{13.24}$$

The vanishing result follows from the fact that the bracket in the integrand is zero, since it is built from objects that obey the KMS symmetry. The independence with respect to t_i merely reflects the fact that, in a system in thermal equilibrium, no measurement can tell at what time the system was prepared in equilibrium. From the analyticity properties of the integrand, the result of the integrations in eq. (13.23) is in fact invariant under all the deformations of the contour $\mathcal{C}$ that preserve the spacing $-i\beta$ between its endpoints.

13.2.6 Conserved charges

Until now, we have considered only the simplest case of a boson field coupled to thermal bath. Although energy is conserved, the system under consideration may exchange energy with the environment, which translates into the canonical density operator $\exp(-\beta\,\mathcal{H})$.

Let us now consider a Hermitian operator $\mathcal{Q}$ that commutes with the Hamiltonian, i.e. that corresponds to a conserved quantity. A field ϕ is said to carry a charge q if it obeys the following commutation relation:

$$\big[\mathcal{Q}, \phi(x)\big] = -q\,\phi(x)\ . \tag{13.25}$$

Note that if ϕ is a real field, then $q = -q^*$. Therefore, in order to have a non-zero real valued charge, the field should be complex.

When there are additional conserved quantities such as $\mathcal{Q}$, their conservation constrains in a similar fashion how they may be exchanged with the heat bath. The canonical equilibrium ensemble must be generalized into the grand canonical ensemble, in which the density operator of the subsystem is given by

$$\rho \equiv \exp\Big\{-\beta\,\big(\mathcal{H} + \mu\,\mathcal{Q}\big)\Big\}\ , \tag{13.26}$$

where μ is the chemical potential associated to the charge $\mathcal{Q}$. Although we have introduced a single such charge, there could be any number of them, each accompanied by its chemical potential. A first consequence of this generalization is that the KMS symmetry is modified by the conserved charge. Now it reads:

$$\mathcal{G}(t_i, \cdots) = e^{\beta\mu q}\,\mathcal{G}(t_i - i\beta, \cdots)\ , \tag{13.27}$$

where q is the charge carried by the field on which the identity applies. Thus, the values of correlation functions at the endpoints are equal up to a twist factor that depends on the chemical potential.

The simplest field that can carry a non trivial charge is a complex scalar field. In the interaction picture, it can be decomposed as follows on a basis of creation and annihilation operators:

$$\phi_{\text{in}}(x) = \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}}\,\Big[a_{\mathbf{p},\text{in}}\,e^{-ip\cdot x} + b^\dagger_{\mathbf{p},\text{in}}\,e^{+ip\cdot x}\Big]\ .$$

(Such a field requires two sets $\{a_{\mathbf{p},\text{in}}, b_{\mathbf{p},\text{in}}\}$ of such operators, because it describes a particle which is distinct from its anti-particle.) From this field, it is possible to construct a theory that has a global $U(1)$ symmetry, corresponding to the conservation of the following charge

$$\mathcal{Q} \equiv \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}}\,\Big\{b^\dagger_{\mathbf{p},\text{in}} b_{\mathbf{p},\text{in}} - a^\dagger_{\mathbf{p},\text{in}} a_{\mathbf{p},\text{in}}\Big\}\ . \tag{13.28}$$

It is then easy to obtain the following grand-canonical averages:

$$\begin{aligned}
\mathrm{Tr}\left(e^{\beta(\mathcal{H}_0+\mu Q)}\; a^\dagger_{\mathbf{p},\mathrm{in}} a_{\mathbf{p}',\mathrm{in}}\right) &= (2\pi)^3\, 2E_{\mathbf{p}}\, n_{\scriptscriptstyle B}(E_{\mathbf{p}}-\mu q)\,\delta(\mathbf{p}-\mathbf{p}') \\
\mathrm{Tr}\left(e^{\beta(\mathcal{H}_0+\mu Q)}\; b^\dagger_{\mathbf{p},\mathrm{in}} b_{\mathbf{p}',\mathrm{in}}\right) &= (2\pi)^3\, 2E_{\mathbf{p}}\, n_{\scriptscriptstyle B}(E_{\mathbf{p}}+\mu q)\,\delta(\mathbf{p}-\mathbf{p}')\;,
\end{aligned} \tag{13.29}$$

and finally obtain the free propagator for a complex scalar field carrying the charge q:

$$\begin{aligned}
G^0(x,y) = \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}} \Big[&\left(\theta_c(x^0-y^0) + n_{\scriptscriptstyle B}(E_{\mathbf{p}}-\mu q)\right) e^{-ip\cdot(x-y)} \\
&+\left(\theta_c(y^0-x^0) + n_{\scriptscriptstyle B}(E_{\mathbf{p}}+\mu q)\right) e^{+ip\cdot(x-y)}\Big]\;.
\end{aligned} \tag{13.30}$$

13.2.7 Fermions

Consider now spin $1/2$ fermions, whose interaction picture representation reads

$$\psi_{\mathrm{in}}(x) = \sum_{s=\pm}\int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}} \left\{ a^\dagger_{s\mathbf{p},\mathrm{in}}\, v_s(\mathbf{p}) e^{+ip\cdot x} + b_{s\mathbf{p},\mathrm{in}}\, u_s(\mathbf{p}) e^{+ip\cdot x}\right\}, \tag{13.31}$$

where the creation and annihilation operators obey canonical *anticommutation* relations (see eqs. (1.175)). Because they are anticommuting fields, a minus sign appears in the derivation of the KMS identity:

$$\mathcal{G}(t_i,\cdots) = -e^{\beta\mu q}\,\mathcal{G}(t_i-i\beta,\cdots)\;. \tag{13.32}$$

Moreover, the eqs. (13.29) are modified into

$$\begin{aligned}
\mathrm{Tr}\left(e^{\beta(\mathcal{H}_0+\mu Q)}\; a^\dagger_{\mathbf{p},\mathrm{in}} a_{\mathbf{p}',\mathrm{in}}\right) &= (2\pi)^3\, 2E_{\mathbf{p}}\, n_{\scriptscriptstyle F}(E_{\mathbf{p}}-\mu q)\,\delta(\mathbf{p}-\mathbf{p}') \\
\mathrm{Tr}\left(e^{\beta(\mathcal{H}_0+\mu Q)}\; b^\dagger_{\mathbf{p},\mathrm{in}} b_{\mathbf{p}',\mathrm{in}}\right) &= (2\pi)^3\, 2E_{\mathbf{p}}\, n_{\scriptscriptstyle F}(E_{\mathbf{p}}+\mu q)\,\delta(\mathbf{p}-\mathbf{p}')\;,
\end{aligned} \tag{13.33}$$

where $n_{\scriptscriptstyle F}$ is the Fermi-Dirac distribution,

$$n_{\scriptscriptstyle F}(E) \equiv \frac{1}{e^{\beta E}+1}\;, \tag{13.34}$$

and the free propagator reads

$$\begin{aligned}
S^0(x,y) = \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}} \Big[&(\not{p}_+ + m)\left(\theta_c(x^0-y^0) - n_{\scriptscriptstyle F}(E_{\mathbf{p}}-\mu q)\right) e^{-ip\cdot(x-y)} \\
&+(\not{p}_- + m)\left(\theta_c(y^0-x^0) - n_{\scriptscriptstyle F}(E_{\mathbf{p}}+\mu q)\right) e^{+ip\cdot(x-y)}\Big]\;,
\end{aligned} \tag{13.35}$$

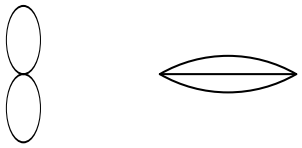
with the notation $\not{p}_\pm \equiv \pm E_{\mathbf{p}}\gamma^0 - \mathbf{p}\cdot\boldsymbol{\gamma}$.

13.2.8 Examples of physical observables

Thermodynamical quantities : A central quantity that encapsulates most of the thermodynamics of a system is its partition function, defined in the canonical ensemble as

$$Z \equiv \mathrm{Tr}\left(e^{-\beta \mathcal{H}}\right) . \tag{13.36}$$

In perturbation theory, the logarithm of Z is obtained as the sum of all the connected vacuum graphs at finite temperature. For instance, for a scalar field, one may have the following diagrams:



From Z, one may access various thermodynamical quantities as follows:

$$\begin{aligned} \text{Energy :} \qquad & E = -\frac{\partial Z}{\partial \beta} , \\ \text{Entropy :} \qquad & S = \beta E + \ln(Z) , \\ \text{Free energy :} \qquad & F = E - TS = -\frac{1}{\beta}\ln(Z) . \end{aligned} \tag{13.37}$$

These quantities encode the bulk properties of the system, such as its equation of state or the existence of phase transitions.

Production rates of weakly coupled particles : In a system at high temperature, it is sometimes interesting to calculate the production rate of a given species of particles. Firstly, note that this quantity is not interesting for particles that are in thermal equilibrium, since by definition they are produced and destroyed in equal amounts, so that their net production rate is zero. The real interest arises for weakly coupled particles that are not in thermal equilibrium with the bulk of the system. For instance, in a hot plasma of quarks and gluons interacting via the strong nuclear force, photons are also produced. However, since they interact only electromagnetically, they may not be thermalized. This is the case for instance when the system size is small compared to the mean free path of photons (i.e. the average distance between two interactions of a photon), because in this situation the produced photons escape without re-interactions. Their production rate therefore may be viewed as the "black body" radiation of the plasma of quarks and gluons.

A pedestrian method for calculating a production rate is the following formula:

$$\omega \frac{dN_\gamma}{dt\, d^3x\, d^3p} \propto \int_{\left(\substack{\text{unobserved}\\ \text{particles}}\right)} \left| \quad \omega \quad \right|^2 \; n(\omega_1)\cdots n(\omega_n) \times (1 \pm n(\omega'_1))\cdots(1 \pm n(\omega'_p)) , \tag{13.38}$$

where the integration is over the invariant phase-space of the unobserved incoming and outgoing particles, weighted by the appropriate occupation factor (n_B or n_F for a particle in the initial

state, and $1+n_{\text{B}}$ or $1-n_{\text{F}}$ for a particle in the final state). In this formula, the gray blob should be calculated with the finite-T Feynman rules.

The previous approach becomes rapidly cumbersome as the number of initial and final state particles increase. The bookkeeping may be simplified by using a finite-T generalization of the formula that relates the decay rate of a particle to the imaginary part of its self-energy:

$$\omega \frac{dN_\gamma}{dt d^3\mathbf{x} d^3\mathbf{p}} \propto \frac{1}{e^{\omega/T}-1} \ \text{Im} \underbrace{\Pi^\mu{}_\mu(\omega,\mathbf{p})}_{\text{photon self-energy}} \ . \tag{13.39}$$

Moreover, there exists a finite-T generalization of the Cutkosky's cutting rules, in order to organize the perturbative calculation of the imaginary part that appears in the right hand side.

Transport coefficients : Let us now discuss the case of *transport coefficients*. As their name suggest, these quantities characterize the ability of the system to move certain (locally conserved) quantities. For instance, the electrical conductivity encodes the properties of the system with respect to the transport of electrical charges, the shear viscosity tells us about how the system reacts to a shear stress (this coefficient is related to the transport of momentum), etc. Note that in their simplest version, these quantities do not depend on frequency (in fact, they are the zero frequency limit of a 2-point function), and therefore they describe the response of the system to an infinitely slow perturbation. They can be generalized into frequency dependent quantities that also contain information about the response to a dynamical disturbance.

The standard approach for evaluating transport coefficients is to use the *Green-Kubo formula*, that relates the transport coefficient to the 2-point correlation function of a current J that couples to the quantity of interest (electrical charge, momentum, etc):

$$\left[\begin{smallmatrix}\text{transport}\\ \text{coefficient}\end{smallmatrix}\right] \sim \lim_{\omega\to 0} \frac{1}{\omega} \,\text{Im} \int_0^{+\infty} dt d^3\mathbf{x}\, e^{-i\omega t} \,\left\langle \left[J(t,\mathbf{x}),J(0,\mathbf{0})\right]\right\rangle \ . \tag{13.40}$$

The physical meaning of this formula is that the system is perturbed at the origin by a current J, and one measures the linear response by evaluating the same current at a generic point $(t,\mathbf{x})$. The transport coefficient is proportional to the Fourier transform of this correlation function at zero energy and momentum. Note that this formula contains the commutator of the two currents, since one wants the two points to be causally connected.

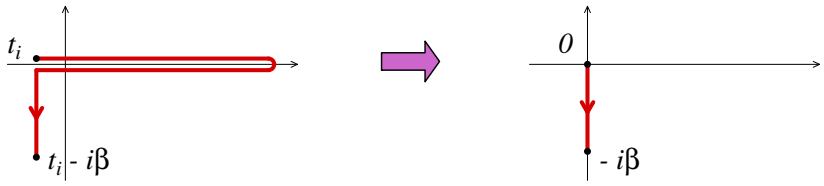
13.2.9 Matsubara formalism

The perturbative rules that we have derived so far are expressed in coordinate space, which is usually not very appropriate for explicit calculations. The standard way of turning them into a set of rules in momentum space is to Fourier transform all the propagators, and to rely on the fact that the Fourier transform of a convolution product is the ordinary product of the Fourier transforms, i.e. symbolically

$$\text{FT}\left(F * G\right) = \text{FT}\left(F\right) \times \text{FT}\left(G\right) \ . \tag{13.41}$$

However, the main difficulty in doing this at finite temperature is that the time integration in the "convolution product" involves an integration over the complex-shaped contour $\mathcal{C}$, which makes it unclear whether we may use the above identity.

Two main solutions to this problem have been devised. The first one is the *imaginary time formalism*, also known as the *Matsubara formalism*, that we have already presented superficially in the section 2.7.2. The main motivation of this formulation is that the quantities that describe the thermodynamics of a system in thermal equilibrium are time independent. Therefore, one may exploit the freedom to deform the contour $\mathcal{C}$ in order to simplify it, as shown in the following figure:



It is customary to denote $x^0 = -i\tau$, so that the variable τ is real and spans the range $[0, \beta)$ (the point $\tau = \beta$ should be removed – indeed, because of the KMS symmetry, it is redundant with the point $\tau = 0$). The imaginary time formalism corresponds to the Feynman rules derived earlier, specialized to this purely imaginary time contour. Note that one could in principle use this formalism in order to calculate time-dependent quantities. One would first obtain them as a function of imaginary times $\tau_1, \tau_2, \cdots$ and their dependence upon real times $x_1^0, x_2^0, \cdots$ may then be obtained by an analytic continuation.

From the KMS symmetry, we see that the propagator, and more generally the integrand of any Feynman diagram, is periodic (for bosons) in the variable τ with period β. Therefore, one can go to Fourier space by decomposing the time dependence in the form of a Fourier series and by doing an ordinary Fourier transform in space :

$$G^0(\tau_x, \mathbf{x}, \tau_y, \mathbf{y}) \equiv T \sum_{n=-\infty}^{+\infty} \int \frac{d^3\mathbf{p}}{(2\pi)^3} \, e^{i\omega_n(\tau_x - \tau_y)} e^{-i\mathbf{p}\cdot(\mathbf{x}-\mathbf{y})} \, G^0(\omega_n, \mathbf{p}) \,, \tag{13.42}$$

with $\omega_n \equiv 2\pi n T$. These discrete frequencies are called *Matsubara frequencies*. Note that for fermions, the propagator is antiperiodic with period β, and the discrete frequencies that appears in the Fourier series are $\omega_n = 2\pi(n + \frac{1}{2})T$. Moreover, if the line carries a conserved charge q, the discrete frequencies are shifted by $-i\mu q$, i.e. $\omega_n \to \omega_n - i\mu q$ (μ is the chemical potential associated to this conservation law). In the case of scalar fields, an explicit calculation gives the following free propagator in Fourier space,

$$G^0(\omega_n, \mathbf{p}) = \frac{1}{\omega_n^2 + \mathbf{p}^2 + m^2} \,. \tag{13.43}$$

Note that, up to a factor i, this propagator is the usual free zero temperature Feynman propagator in which one has substituted $p^0 \to i\omega_n$. Let us list here the Feynman rules for perturbative calculations in this formalism:

- Propagators :

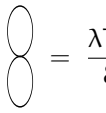
  $$G^0(\omega_n, \mathbf{p}) = 1/(\omega_n^2 + \mathbf{p}^2 + m^2) \,,$$

- Vertices : each vertex brings a factor λ. Moreover, the sum of the ω_n's and of the $\mathbf{p}$'s that enter into each vertex are zero,

- Loops :

$$T\sum_{n\in\mathbb{Z}}\int\frac{d^3\mathbf{p}}{(2\pi)^3}\,.$$

As an illustration of the use of this formalism, let us give two examples of vacuum graphs:



$$= \frac{\lambda T^2}{8}\sum_{m,n\in\mathbb{Z}}\int\frac{d^3\mathbf{p}}{(2\pi)^3}\frac{d^3\mathbf{q}}{(2\pi)^3}\,\frac{1}{(\omega_m^2+\mathbf{p}^2)(\omega_n^2+\mathbf{q}^2)}\,, \tag{13.44}$$

$$= \frac{g^2T^2}{6}\sum_{m,n\in\mathbb{Z}}\int\frac{d^3\mathbf{p}}{(2\pi)^3}\frac{d^3\mathbf{q}}{(2\pi)^3}\,\frac{1}{(\omega_m^2+\mathbf{p}^2)(\omega_n^2+\mathbf{q}^2)(\omega_{m+n}^2+(\mathbf{p}+\mathbf{q})^2)}\,. \tag{13.45}$$

(The integrands have been written for massless fields here.)

The Fourier space version of the Matsubara formalism is structurally very similar to the zero temperature Feynman rules, which makes it quite appealing. There is one caveat however: the continuous integrations over energies are now replaced by discrete sums, which are considerably harder to calculate. Let us expose here two general methods for evaluating these sums. The first one is based on the following representation of the propagator of eq. (13.43):

$$G^0(\omega_n,\mathbf{p}) = \frac{1}{2E_\mathbf{p}}\int_0^\beta d\tau\, e^{-i\omega_n\tau}\Big[(1+n_B(E_\mathbf{p}))\,e^{-E_p\tau}+n_B(E_\mathbf{p})\,e^{E_p\tau}\Big]\,, \tag{13.46}$$

where the integrand in the right hand side is a mixed representation that depends on the momentum $\mathbf{p}$ and the imaginary time τ. By replacing each propagator of a given graph by this formula, the discrete sums can be easily performed since they are all of the form

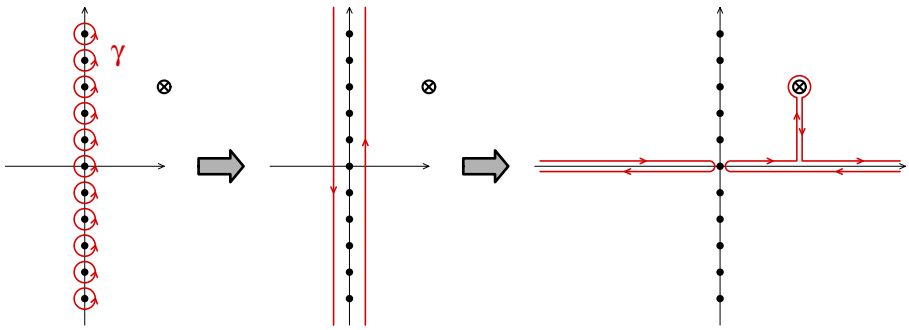
$$\sum_{n\in\mathbb{Z}} e^{i\omega_n\tau} = \beta\sum_{n\in\mathbb{Z}}\delta(\tau-n\beta)\,. \tag{13.47}$$

(The left hand side is obviously periodic in τ with period β, which is ensured in the right hand side by the sum over infinitely many shifted copies of the delta function.) At this point, one has to integrate over the τ's that have been introduced when replacing the propagators by (13.46), but these integrals are straightforward since the dependence on these times is in the form of delta functions (moreover, only a finite number of the delta functions that appear in the right hand side of eq. (13.47) actually contribute, due to the constraint that each τ must be in the range $[0,\beta)$). As an illustration, consider the evaluation of the 1-loop tadpole in a scalar theory with quartic coupling:

$$\begin{aligned}
&= \frac{\lambda T}{2}\sum_{n\in\mathbb{Z}}\int\frac{d^3\mathbf{p}}{(2\pi)^3}\frac{1}{\omega_n^2+\mathbf{p}^2}\\
&= \frac{\lambda}{2}\int\frac{d^3\mathbf{p}}{(2\pi)^3 2E_\mathbf{p}}\int_0^\beta d\tau\sum_n\delta(\tau-n\beta)\Big[(1+n_B(E_\mathbf{p}))e^{-E_p\tau}+n_B(E_\mathbf{p})e^{E_p\tau}\Big]\\
&= \frac{\lambda}{2}\int\frac{d^3\mathbf{p}}{(2\pi)^3 2E_\mathbf{p}}\big[1+2\,n_B(E_\mathbf{p})\big]\\
&= \lambda\left[\frac{\Lambda^2}{16\pi^2}+\frac{T^2}{24}+\cdots\right]\,,
\end{aligned} \tag{13.48}$$

where Λ is an ultraviolet cutoff that restricts the interaction range $|\mathbf{p}| \leq \Lambda$ (the final formula assumes that $\Lambda \gg T$, and we have not written the terms that depend on the mass). The first term is the usual zero temperature ultraviolet divergence, while the term coming from the Bose-Einstein distribution exists only at non-zero temperature. This second term is ultraviolet finite, thanks to the exponential suppression of the Bose-Einstein distribution at large energy. We can already note on this example that the ultraviolet divergences are identical to the zero temperature ones. This is a general property: if the action has already been renormalized at zero temperature, there are no additional ultraviolet divergences at finite temperature. This is quite clear on physical grounds: being at finite temperature means that one has a dense medium in which the average inter-particle distance is T^{-1}. However, in the ultraviolet limit, one probes distance scales that are much smaller than the inverse temperature, for which the effects of the surrounding medium are irrelevant.

Figure 13.1: *Successive deformations of the contour in order to calculate the discrete sums over Matsubara frequencies. The cross denotes a pole of the function $f(z)$, while the solid dots on the imaginary axis are the poles of $\mathcal{P}(z)$.*



An alternate method for evaluating the sums over the discrete Matsubara frequencies is to note that the function

$$\mathcal{P}(z) \equiv \frac{\beta}{e^{\beta z} - 1} \tag{13.49}$$

has simple poles of residue 1 at all the $z = i\omega_n$. Therefore, we can write

$$\sum_{n \in \mathbb{Z}} f(i\omega_n) = \oint_\gamma \frac{dz}{2i\pi}\, \mathcal{P}(z)\, f(z)\ , \tag{13.50}$$

where γ is an integration contour made of infinitesimal circles around each pole of $\mathcal{P}(z)$, as shown in the left part of the figure 13.1. The second step is to deform the contour γ as shown in the middle of the figure 13.1. For this transformation to hold as is, with no extra term, the function $f(z)$ should not have any pole on the imaginary axis, which is usually the case. Finally, a second deformation brings the contour along the real axis. If the function $f(z)$ has poles, the new contour should wrap around these poles, which an additional contribution. Thus, after these transformations, the discrete sum over the Matsubara frequencies has been rewritten as a continuous integral along the real axis (and the weight $\mathcal{P}(z)$ becomes an ordinary Bose-Einstein distribution), plus some isolated contributions coming from poles of the summand.

13.2.10 Momentum space Schwinger-Keldysh formalism

The imaginary time formalism is particularly well suited to calculate the time-independent thermodynamical properties of a system at finite temperature. However, interesting dynamical information is also contained in time-dependent objects. In principle, one could first evaluate them in the Matsubara formalism in terms of imaginary times τ (or imaginary frequencies $i\omega_n$), and then perform an analytic continuation to real times or energies. Beyond 2-point functions (i.e. for functions that depend on more than one energy, taking into account energy conservation), this analytic continuation is usually extremely complicated and for this reason it is desirable to be able to obtain the result directly in terms of real energies.

In fact, we may ignore[4] the vertical part of the contour $\mathcal{C}$. A heuristic justification is to let the initial time t_i go to $-\infty$ and turn off adiabatically the interactions in this limit. Therefore, the canonical density operator becomes $\exp(-\beta\,\mathcal{H}_0)$ and there is no need for the vertical part of the time contour. Let us call $+$ and $-$ respectively the upper and lower horizontal branches of the contour. We may then break down the free propagator $G^0(x,y)$ into four propagators G^0_{++}.G^0_{--}, G^0_{+-} and G^0_{-+} depending on where x, y are located, and Fourier transform each of them separately. For a scalar field, this gives:

$$\begin{aligned}
G^0_{++}(p) &= \frac{i}{p^2-m^2+i\epsilon} + 2\pi\, n_{\scriptscriptstyle B}(E_{\mathbf{p}})\,\delta(p^2-m^2)\,,\\
G^0_{+-}(p) &= 2\pi\,(\theta(-p^0)+n_{\scriptscriptstyle B}(E_{\mathbf{p}}))\,\delta(p^2-m^2)\,,\\
G^0_{--}(p) &= \left[G^0_{++}(p)\right]^* \quad , \qquad G^0_{-+}(p) = G^0_{+-}(-p)\,.
\end{aligned} \tag{13.51}$$

Note that these propagators are very closely related to those of the Schwinger-Keldysh formalism at zero temperature (see eqs. (1.288)), since we have

$$\forall \epsilon,\epsilon' = \pm\,, \qquad G^0_{\epsilon\epsilon'}(p) = \left[G^0_{\epsilon\epsilon'}(p)\right]_{T=0} + 2\pi\, n_{\scriptscriptstyle B}(E_{\mathbf{p}})\,\delta(p^2-m^2)\,. \tag{13.52}$$

The rules for the vertices and loops are identical to those of the Schwinger-Keldysh formalism at zero temperature, namely:

- One must assign types $+$ and $-$ to the vertices of a diagram in all the possible ways,
- Each vertex of type $+$ brings a factor $-i\lambda$ and each type $-$ vertex a factor $+i\lambda$,
- A vertex of type ϵ and a vertex of type ϵ' are connected by the free propagator $G^0_{\epsilon\epsilon'}$,
- Each loop momentum must be integrated with the measure $d^4p/(2\pi)^4$.

Since this formalism is a simple extension of the zero temperature Schwinger-Keldysh formalism (the only difference being the propagators in eq. (13.52)), it makes the connection with perturbation theory at zero temperature more transparent.

[4] A more careful treatment of the vertical part of the contour indicates that its effect it to replace the statistical distribution $n_{\scriptscriptstyle B}(E_{\mathbf{p}})$ by $n_{\scriptscriptstyle B}(|p^0|)$ in the equations (13.51). Because these distributions are accompanied by a delta function $\delta(p^2-m^2)$, this has no incidence except in self-energy insertions on a propagator. In this case, the delta function may be multiplied by a principal value of the same argument, which is nothing but a $\delta'(p^2-m^2)$, and $\delta'(p^2-m^2)\,n_{\scriptscriptstyle B}(E_{\mathbf{p}})$ is not the same distribution as $\delta'(p^2-m^2)\,n_{\scriptscriptstyle B}(|p^0|)$.

In the Matsubara formalism, the KMS symmetry is trivially encoded in the fact that all the objects depend only on the discrete frequencies ω_n. In the Schwinger-Keldysh formalism, it is somewhat more obfuscated. A generic n-point function $\Gamma_{\epsilon_1\cdots\epsilon_n}(p_1\cdots p_n)$, amputated of its eternal legs, obeys the following two identities:

$$\begin{aligned}
&\sum_{\epsilon_1\cdots\epsilon_n=\pm} \Gamma_{\epsilon_1\cdots\epsilon_n}(k_1,\cdots,k_n) = 0\,,\\
&\sum_{\epsilon_1\cdots\epsilon_n=\pm} \Big[\prod_{\{i|\epsilon_i=-\}} e^{-\beta k_i^0}\Big]\,\Gamma_{\epsilon_1\cdots\epsilon_n}(k_1,\cdots,k_n) = 0\,.
\end{aligned} \tag{13.53}$$

It is the second of these identities that reflects the KMS symmetry.

Finally, let us note for later use that the four propagators of eqs. (13.51) can be related to the zero temperature Feynman propagator and its complex conjugate by the following formula:

$$\begin{pmatrix} G^0_{++} & G^0_{+-} \\ G^0_{-+} & G^0_{--} \end{pmatrix} = U \begin{pmatrix} G^0_F & 0 \\ 0 & G^0_{\bar F} \end{pmatrix} U \tag{13.54}$$

with

$$U(p) \equiv \begin{pmatrix} \sqrt{1+n_B(E_p)} & \frac{\theta(-p^0)+n_B(E_p)}{\sqrt{1+n_B(E_p)}} \\ \frac{\theta(+p^0)+n_B(E_p)}{\sqrt{1+n_B(E_p)}} & \sqrt{1+n_B(E_p)} \end{pmatrix} \tag{13.55}$$

and

$$G^0_F(p) \equiv \frac{i}{p^2-m^2+i\epsilon}\,. \tag{13.56}$$

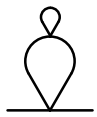
13.3 Long distance effective theories

13.3.1 Infrared divergences

Quantum field theories with massless bosons at non-zero temperature suffer from pathologies in the infrared sector, due to the low energy behavior of the Bose-Einstein distribution:

$$n_B(E) \underset{E\ll T}{\approx} \frac{T}{E} \gg 1\,. \tag{13.57}$$

As we shall see now, using a massless ϕ^4 scalar field theory as a playground, this leads to loop contributions that exhibit soft divergences. The simplest graph that suffers from this problem is the following 2-loop graph,



,

that has two nested tadpoles. Let us assume that the uppermost tadpole has already been combined with the corresponding 1-loop ultraviolet counterterm, so that only the finite part remains, and denote μ^2 the finite remainder. From eq. (13.48), its expression is given by

$$\mu^2 \equiv \frac{\lambda T^2}{24} \,. \tag{13.58}$$

(This is the exact result for the finite part in a massless theory.) With this shorthand, we have

$$\begin{aligned} &= \frac{\lambda T \mu^2}{2} \int \frac{d^3\mathbf{p}}{(2\pi)^3} \sum_{n\in\mathbb{Z}} \frac{1}{(\omega_n^2+\mathbf{p}^2)^2} \\ &= \frac{\lambda \mu^2}{2} \int \frac{d^3\mathbf{p}}{(2\pi)^3} \underbrace{\frac{n_{\scriptscriptstyle B}(p)(1+n_{\scriptscriptstyle B}(p))}{4p^2} \left\{ \frac{2}{T} + \frac{e^{\beta p}-e^{-\beta p}}{p} \right\}}_{\underset{p\ll T}{\approx} \frac{T}{p^4}} \\ &= \frac{\lambda \mu^2}{4\pi^2} \frac{T}{\Lambda_{\scriptscriptstyle IR}} + \text{infrared finite terms} \,, \end{aligned} \tag{13.59}$$

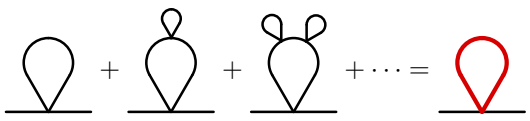
where in the last line we have introduced an *infrared cutoff* $\Lambda_{\scriptscriptstyle IR}$ in order to prevent a divergence at the lower end of the integration range. A similar calculation would indicate an even worse infrared singularity in the following 3-loop graph:

$$\sim \lambda T \mu \, \frac{\mu^3}{\Lambda_{\scriptscriptstyle IR}^3} + \text{infrared finite terms} \,, \tag{13.60}$$

and more generally for n insertions of the base tadpole on the main loop,

$$\sim \lambda T \mu \left(\frac{\mu}{\Lambda_{\scriptscriptstyle IR}} \right)^{2n-1} + \text{infrared finite terms} \,. \tag{13.61}$$

Unlike ultraviolet divergences that can, in a renormalizable theory, be disposed of systematically by a redefinition of the couplings in front of a few *local* operators in the Lagrangian, it is not possible to handle infrared divergences in this manner because they correspond to long distance phenomena. Fortunately, there is a simple way out in the present case, since the series of graphs that we have started evaluating are the first few terms of a geometrical series, since the repeated insertions of a tadpole equal to μ^2 (after subtraction of the appropriate counterterm) merely amounts to dressing by a mass μ^2 an originally massless propagator. Namely, we have



$$= \frac{\lambda}{2} \int \frac{d^3\mathbf{p}}{(2\pi)^3 2\sqrt{\mathbf{p}^2+\mu^2}} \left[1 + 2 n_{\scriptscriptstyle B}\left(\sqrt{\mathbf{p}^2+\mu^2}\right) \right] . \tag{13.62}$$

(The red propagator indicates a massive scalar with mass μ^2.) The procedure used here, that consists in summing an infinite subset of (individually divergent) perturbative contributions, is called a *resummation*. We can readily see that it leads to an infrared finite sum, since now the quantity μ^2 plays the role of a cutoff at small momentum.

Let us now estimate the contribution of the infrared sector to this integral. At weak coupling, we have $\mu \sim gT \ll T$. Therefore, for momenta $p \sim \mu$, we have

$$\lambda \int dp\ p^2\ \frac{1 + 2\,n_B(\sqrt{\mathbf{p}^2+\mu^2})}{\sqrt{\mathbf{p}^2+\mu^2}} \sim \lambda \int dp\ p^2\ \frac{T}{\mathbf{p}^2+\mu^2} \sim \lambda T\,\mu \sim \lambda^{3/2} T^2\ . \tag{13.63}$$

This contribution comes in addition to the ultraviolet divergence $\lambda\Lambda^2$ and the contribution λT^2 that are both contained in the first diagram of the resummed series (these terms come from momenta of order T or above). We observe here an unexpected feature; the appearance of half powers of the coupling constant λ. On the surface, this is quite surprising since the power counting indicates that one power of λ should come with each loop. This oddity is in fact a consequence of the infrared divergence of the integral, combined with the fact that μ introduced in the resummation is of order $\lambda^{1/2}$. Although the loop expansion generates a series which is analytic in λ, this property may be broken if some parameters in the integrands depend on $\lambda^{1/2}$.

13.3.2 Screened perturbation theory

The resummation of the finite part of the 1-loop tadpole is sufficient in order to screen the infrared divergences in the graphs corresponding to a strict loop expansion. However, since such a resummation amounts to a reorganization of perturbation theory (here, by already including an infinite set of graphs into the propagator), it should be done in a careful way that avoids any double countings, and ensures that we are not modifying the original theory. This can be achieved by a method, called *screened perturbation theory*, that consists in adding and subtracting a mass term to the Lagrangian,

$$\mathcal{L} = \frac{1}{2}\big(\partial_\mu\phi\big)\big(\partial^\mu\phi\big) - \frac{\lambda}{4!}\phi^4 - \frac{1}{2}\mu^2\phi^2 + \frac{1}{2}\mu^2\phi^2\ . \tag{13.64}$$

This manipulation clearly ensures that nothing is changed to the original theory. The reorganization of perturbation theory allowed by this trick comes from treating the two mass terms on different footings: the term $-\frac{1}{2}\mu^2\phi^2$ is treated non-perturbatively by including it directly into the definition of the free propagator, while the term $+\frac{1}{2}\mu^2\phi^2$ is treated order by order, as a finite counterterm.

In this reorganization, the value of μ^2 has so far been left unspecified, and it could a priori be chosen arbitrarily. A general rule governing this choice is to include in μ^2 as much as possible of the large contributions coming from loop corrections to the propagator. The 1-loop contribution in λT^2 is an obvious candidate for including in μ^2, since for momenta $p^2 \lesssim \lambda T^2$ this is indeed a large correction to the denominator of the propagator. At small coupling $\lambda \ll 1$, this is the dominant one. However, when the coupling increases, the propagator may receive additional large corrections from higher order loop corrections, and an improved resummation scheme could include these additional corrections.

A further improvement, sometimes considered in some applications, it to let μ^2 free and to use some reasonable condition to choose an "optimal" value. For instance, this condition may be

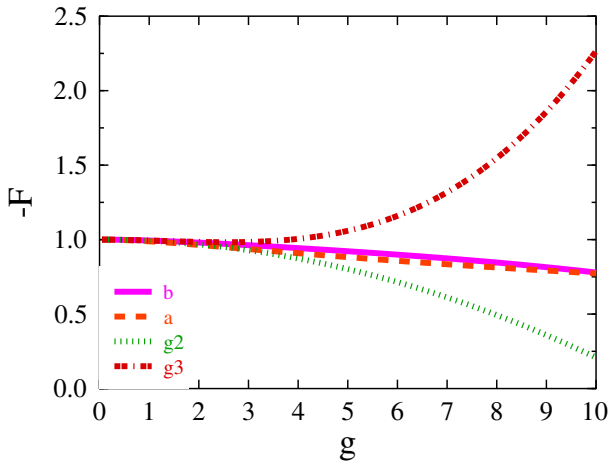
the minimization of the 1-loop correction, which in a sense would indicate that the resummation has shifted most of this loop contribution into the free propagator. The quantity to minimize is

$$\text{(loop)} + \text{counterterms} = \frac{\lambda}{2}\int^{\Lambda}\frac{d^3\mathbf{p}}{(2\pi)^3}\frac{1+2\,n_B(\sqrt{p^2+\mu^2})}{2\sqrt{p^2+\mu^2}} - \frac{\lambda\Lambda^2}{16\pi^2} - \mu^2\,, \qquad (13.65)$$

where the two subtractions are respectively the ultraviolet counterterm and the finite counterterm necessary in order not to overcount the mass μ^2. Such an equation is called a *gap equation*[5]. Because this equation is non-linear in μ^2, its solution contains all orders in λ, but at small λ it is dominated by the 1-loop result $\mu^2 = \lambda T^2/24$.

We show an application of this method to the calculation of the free energy F in the figure 13.2. In this figure, the results obtained at 1-loop and 2-loops in screened perturbation theory

Figure 13.2: *Free energy at non-zero temperature in the ϕ^4 scalar field theory (normalized to the free energy of the non-interacting theory). The horizontal axis is the coupling strength $g \equiv \lambda^{1/2}$. Curves "g2" and "g3": orders λ and $\lambda^{3/2}$ in the original perturbative expansion. Curves "a" and "b": screened perturbation theory at 1-loop and 2-loops, with the mass μ^2 determined as the exact solution of the gap equation (13.65).*



are compared to the first two orders (λ and $\lambda^{3/2}$) of the ordinary perturbative expansion. Firstly, we can see that the latter is quite unstable except at low coupling: the two subsequent orders differ substantially, and even the sign of the correction due to the interactions flips. In contrast, screened perturbation theory leads to a remarkably stable result, with very small changes when going from 1-loop to 2-loops. To a large extent, this success is due to the non trivial coupling dependence of the mass μ^2, acquired by solving the gap equation (13.65) (screened perturbation theory with only the 1-loop mass, would be better than strict perturbation theory, but would encounter some difficulties at large coupling).

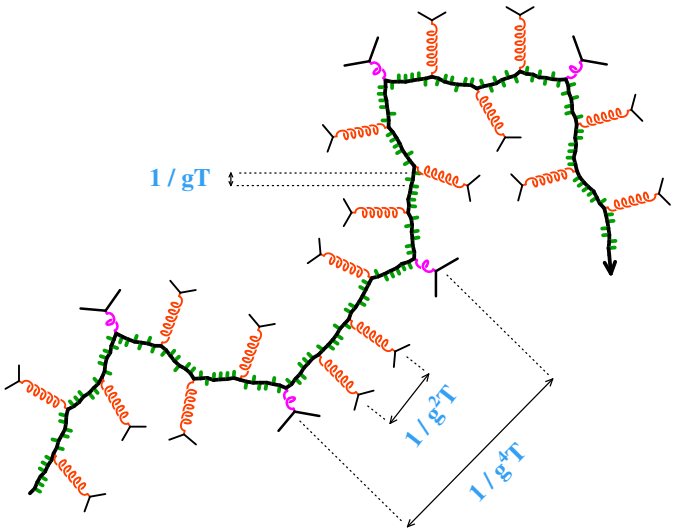
[5]The terminology comes from the fact that the solution of this equation usually shifts the energy of a particle, generating a "gap" in the spectrum, and thus requiring a non-zero energy to create such a particle.

13.3.3 Relevant physical scales

The mass generation that we have observed earlier in this section is due to the interactions with the surrounding dense medium: even if the asymptotic isolated particles are massless, their dispersion relation is modified by their collisions with the particles of the thermal bath[6]. In theories that have more structure, such as non-Abelian gauge theories, the presence of a hot medium is responsible for several phenomena, that we list here in terms of the corresponding characteristic length scale:

- $\ell = T^{-1}$. T is the typical momentum of a particle in this system, and its inverse is the typical separation between two neighboring particles. At shorter distance scales, a particle behaves exactly as if it were in the vacuum. This is why ultraviolet renormalization at non-zero temperature can be done with the zero-temperature counterterms.

- $\ell = (gT)^{-1}$, where g is the gauge coupling (that plays the same role as $\sqrt{\lambda}$ in the scalar theory we have discussed so far). This is the typical distance over which a particle "feels" modifications of its dispersion relation. In gauge theories, this modification is not a mere mass term, but may take the form of a momentum dependent self-energy, possibly with a non-zero imaginary part. Besides the appearance of a thermal gap in the spectrum of gauge bosons and matter fields, these self-energies also encode effects such as *Debye screening* and *Landau damping*.

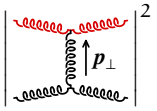


- $\ell = (g^2T)^{-1}$. This is the mean distance between scatterings with a soft color exchange. These are forward scatterings, since the momentum transfer (of order gT, the scale of the infrared cutoff provided by the dressing of the gluon propagator) is much smaller than the momentum of the incoming particles (typically T). A gross way to obtain this scale is by

[6]The same happens to electrons in a crystal, due to their interactions with photons.

estimating the corresponding scattering rate:

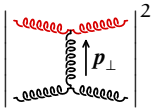
$$\Gamma_{\substack{\text{soft}\\\text{collisions}}} = \left| \quad \right|^2 \sim g^4 T^3 \int_{gT \lesssim p_\perp} \frac{d^2\mathbf{p}_\perp}{p_\perp^4} \sim g^2 T \,, \tag{13.66}$$



where $p_\perp$ is the momentum transfer transverse to the momentum of the incoming particles. Although these scatterings do not lead to an appreciable transport of momentum, they reshuffle the color of the particles and hence contribute to the color conductivity.

- $\ell = (g^4 T)^{-1}$. This is the mean distance between scatterings with a momentum transfer of order T, i.e. those that scatter particles at large angles. Estimating this scale is done as above, but with a lower limit of order T for the momentum transfer:

$$\Gamma_{\substack{\text{hard}\\\text{collisions}}} = \left| \quad \right|^2 \sim g^4 T^3 \int_{T \lesssim p_\perp} \frac{d^2\mathbf{p}_\perp}{p_\perp^4} \sim g^4 T \,. \tag{13.67}$$



This scale is usually called the *mean free path*. This is the relevant scale for all transport phenomena that require significant momentum exchanges, for instance the viscosity. Beyond this scale is the realm of collective effects such as sound waves (on these scales, it is more appropriate to describe the system as a fluid rather than in terms of elementary field excitations).

13.3.4 Perturbative and non-perturbative modes

Although it is in principle possible to study any phenomenon at finite temperature in terms of the bare Lagrangian, this becomes increasingly difficult at large distance because of non-trivial in-medium effects. In order to circumvent this difficulty, various resummation schemes and effective descriptions have been devised, one of which is the screened perturbation theory that we have discussed earlier.

Our goal here is not to give a detailed account of these various techniques, but to provide general principles regarding what can and cannot be treated perturbatively, focusing on gauge bosons. Let us first recall that a mode is perturbative if its kinetic energy dominates its interaction energy. For a mode of momentum k, the kinetic energy of a gauge field can be estimated as

$$K \sim \left\langle (\partial A)^2 \right\rangle \sim k^2 \left\langle A^2 \right\rangle \,. \tag{13.68}$$

For the interaction energy, we have

$$I \sim g^2 \left\langle A^4 \right\rangle \sim g^2 \left\langle A^2 \right\rangle^2 \,. \tag{13.69}$$

(The second part of the equation is of course not exact, but it gives the correct order of magnitude.) Thus, a mode of momentum k is perturbative if $k^2 \gg g^2 \langle A^2 \rangle$. When discussing the order of magnitude of $\langle A^2 \rangle$, it is useful to distinguish the contribution of the various momentum scales by defining

$$\left\langle A^2 \right\rangle_{\kappa^*} \sim \int^{\kappa^*} \frac{d^3\mathbf{p}}{E_\mathbf{p}} \, n_B(E_\mathbf{p}) \,,$$

the contribution of all the thermal modes up to the scale κ^*. From these considerations, we can now distinguish three types of modes:

- **Hard modes :** $k \sim T$**.** For these modes, we have $\langle A^2 \rangle_T \sim T^2$, and $K \gg I$. They are therefore fully perturbative.
- **Soft modes :** $k \sim gT$**.** For these modes, $k^2 \sim g^2 \langle A^2 \rangle_T$, which implies that the soft modes interact strongly with the hard modes. However, we also have $\langle A^2 \rangle_{gT} \sim gT^2$, so that $k^2 \gg g^2 \langle A^2 \rangle_{gT}$. Thus, the soft modes interact perturbatively among themselves. Consequently, it is possible to describe perturbatively the soft modes, provided one has performed first a resummation of the contribution of the hard modes. Screened perturbation theory is a realization of this idea.
- **Ultrasoft modes :** $k \sim g^2 T$**.** For these modes, we have $\langle A^2 \rangle_{g^2 T} \sim g^2 T^2$, so that $k^2 \sim g^2 \langle A^2 \rangle_{g^2 T}$. Therefore, the ultrasoft modes interact non perturbatively among themselves, and there is no way to treat them in a perturbative approach. A non perturbative approach, such as lattice field theory, is necessary for this.

13.4 Out-of-equilibrium systems

Until now, we have discussed only systems in equilibrium, whose initial state is described by the canonical density operator $\rho \equiv \exp(-\beta \mathcal{H})$. However, many interesting questions could also be asked for a system which is not initially in thermal equilibrium, the prime of them being to describe its relaxation towards equilibrium. In this section, we discuss a few aspects of the quantum field theory treatment of out-of-equilibrium systems.

13.4.1 Pathologies of the naive approach

Firstly, let us note that the Matsubara formalism does not seem prone to a simple out-of-equilibrium generalization, since the KMS symmetry (that encodes into the correlation functions the fact that the system is in equilibrium) is in a sense hardwired into the discrete Matsubara frequencies.

The Schwinger-Keldysh formalism appears to be a more adequate starting point for such a generalization. Let us first discuss a simple extension that does not work, because the reasons of its failure will teach us a useful lesson. Since in eqs. (13.52), the only reference to the statistical state of the system is contained in the Bose-Einstein distribution $n_B(E_{\mathbf{p}})$, we may try to replace it by an arbitrary distribution $f(\mathbf{p})$ that describes the particle distribution in an out-of-equilibrium system[7]:

$$\forall \epsilon, \epsilon' = \pm \, , \qquad G^0_{\epsilon\epsilon'}(p) = \Big[G^0_{\epsilon\epsilon'}(p) \Big]_{T=0} + 2\pi \, f(\mathbf{p}) \, \delta(p^2 - m^2) \, . \tag{13.70}$$

[7] Note firstly that this would not encompass the most general initial states, only those for which the initial correlations are only 2-point correlations.

Consider now the insertion of a self-energy Σ on the bare propagator,

$$= \sum_{\epsilon,\epsilon'=\pm} G^0_{+\epsilon}(p)\,\Sigma_{\epsilon\epsilon'}(p)\,G^0_{\epsilon'+}(p)\,. \tag{13.71}$$

Such an expression is delicate to expand, because it involves products of distributions that are notoriously ill-defined, such as $\delta^2(p^2-m^2)$. Let us first determine which of these products are well defined and which are not. For this, let us write

$$\begin{aligned}\left[iP\left(\frac{1}{z}\right)+\pi\delta(z)\right]^2 &= \pi^2\delta^2(z)-\left[P\left(\frac{1}{z}\right)\right]^2+2i\pi\delta(z)P\left(\frac{1}{z}\right)\\ &= \left[\frac{i}{z+i0^+}\right]^2=-i\frac{d}{dz}\left[\frac{i}{z+i0^+}\right]\\ &= -i\frac{d}{dz}\left[iP\left(\frac{1}{z}\right)+\pi\delta(z)\right]=\left[P\left(\frac{1}{z}\right)\right]'-i\pi\delta'(z)\,.\end{aligned} \tag{13.72}$$

From this exercise, we obtain the following two identities:

$$\begin{aligned}\pi^2\delta^2(z)-\left[P\left(\frac{1}{z}\right)\right]^2 &= \left[P\left(\frac{1}{z}\right)\right]'\\ 2\delta(z)P\left(\frac{1}{z}\right) &= -\delta'(z)\,.\end{aligned} \tag{13.73}$$

Since the derivative of a distribution is well-defined, this indicates that certain products (or combinations of products) of delta functions and principal values are well defined, but not all of them (for instance, the product $\delta^2(z)$ makes no sense).

Returning now to eq. (13.71) and expanding the propagators, we see that it contains terms that are ill-defined:

$$= \begin{bmatrix}\text{well defined}\\ \text{distributions}\end{bmatrix}+\pi^2\delta^2(p^2-m^2)\left[(1+f(\mathbf{p}))\Sigma_{+-}-f(\mathbf{p})\Sigma_{-+}\right], \tag{13.74}$$

where we have used the first of eqs. (13.53) in order to simplify the combination of self-energies that appear in the square bracket. Note that the square bracket vanishes *in equilibrium* thanks to the KMS symmetry. We are thus facing a very peculiar pathology, that exists only out-of-equilibrium.

We may learn a bit more about this issue by formally resumming the self-energy Σ on the propagator. Let us introduce the following notations:

$$\mathbb{G}^0\equiv\begin{pmatrix}G^0_{++} & G^0_{+-}\\ G^0_{-+} & G^0_{--}\end{pmatrix}\,,\quad \mathbb{D}\equiv\begin{pmatrix}G^0_{\scriptscriptstyle F} & 0\\ 0 & G^{0*}_{\scriptscriptstyle F}\end{pmatrix}\,,\quad \mathbb{S}\equiv\begin{pmatrix}\Sigma_{++} & \Sigma_{+-}\\ \Sigma_{-+} & \Sigma_{--}\end{pmatrix}\,, \tag{13.75}$$

and consider the resummed propagator defined by

$$\mathbb{G}\equiv\sum_{n=0}^{\infty}\left[\mathbb{G}^0(-i\mathbb{S})\right]^n\mathbb{G}^0\,. \tag{13.76}$$

A straightforward calculation shows that

$$\mathbb{G} = U \begin{pmatrix} G_F & G_F \widetilde{\Sigma} G_F^* \\ 0 & G_F^* \end{pmatrix} U \tag{13.77}$$

where U is the matrix defined in eq. (13.55), but with $f(\mathbf{p})$ instead of the Bose-Einstein distribution, and where we have used the following notations

$$\begin{aligned} &G_F(p) \equiv \frac{i}{p^2 - m^2 - \Sigma_F + i\epsilon} \,, \\ &\Sigma_F \equiv \Sigma_{++} + \Sigma_{+-} \,, \\ &\widetilde{\Sigma} \equiv \frac{1}{1 + f(\mathbf{p})} \Big[(1 + f(\mathbf{p}))\Sigma_{+-} - f(\mathbf{p})\Sigma_{-+} \Big] \,. \end{aligned} \tag{13.78}$$

Note that the Feynman propagator and its complex conjugate have mirror poles on each side of the real energy axis. If the self-energy Σ_F has no imaginary part, then these poles "pinch" the real axis and lead to a singularity (this is in fact a pathology of the same nature as the product δ^2 in eq. (13.74)). By performing explicitly the multiplication with the matrix U, we obtain the resummed propagator in the following form:

$$\begin{aligned} G_{\epsilon\epsilon'}(p) \;\; = \;\; & \Big[G^0_{\epsilon\epsilon'}(p) \Big]_{T=0} + 2\pi f(\mathbf{p})\, \delta(p^2 - m^2) \\ & + \Big[(1 + f(\mathbf{p}))\Sigma_{+-} - f(\mathbf{p})\Sigma_{-+} \Big] G_F(p) G_F^*(p) \,. \end{aligned} \tag{13.79}$$

Since it does not depend on the indices $\epsilon\epsilon'$, the pathological term (on the second line) appears on the same footing as the second term, that contains the distribution $f(\mathbf{p})$. Thus, the lesson of this calculation is that one may consider hiding this pathology into a redefinition of the distribution $f(\mathbf{p})$. However, the naive formalism that we have tried to use so far is not adequate for doing this consistently, and must be amended in a number of ways:

- The initial time t_i should not be taken to $-\infty$, as is done when using the Schwinger-Keldysh formalism in momentum space. Indeed, this is the time at which the system was prepared in an out-of-equilibrium state. If it were equal to $-\infty$, the system would have had an infinite amount of time for relaxing to equilibrium at the finite time where a measurement is performed. Note that observables will in general depend on the initial time t_i, in contrast with what happens in equilibrium.
- The Schwinger-Keldysh formalism in momentum space assumes that the system is invariant by translation, in particular in the time direction. This is clearly not the case when the system starts out-of-equilibrium, since it is expected to evolve towards equilibrium. Thus, one should stick to the formalism in coordinate space.

13.4.2 Kadanoff-Baym equations

The Kadanoff-Baym equations, that we shall derive now, may be viewed as a kind of *quantum kinetic equations*. These equations are exact, but contain a self-energy that must be truncated

to a manageable number of diagrams in order to be usable in practical applications. In the next subsection, we will show how the traditional kinetic equations can be derived from the Kadanoff-Baym equations.

The starting point is the Dyson-Schwinger equation, written in coordinate space, that expresses the resummation of a self-energy on the propagator:

$$\begin{aligned} G(x,y) &= G^0(x,y) + \int_{\mathcal{C}} d^4u d^4v \; G^0(x,u)\Big(-i\Sigma(u,v)\Big) G(v,y) \\ G(x,y) &= G^0(x,y) + \int_{\mathcal{C}} d^4u d^4v \; G(x,u)\Big(-i\Sigma(u,v)\Big) G^0(v,y) \;, \end{aligned} \tag{13.80}$$

where G^0 is the free propagator and G is the resummed one. Note that the time integrations run over the Schwinger-Keldysh contour $\mathcal{C}$. Here, we have written the equation in two ways, depending on whether the self-energy is inserted on the right or on the left of the bare propagator (in the end, the resulting propagator G is the same in both cases). Next, we apply the operator $\Box_x + m^2$ on the first equation and $\Box_y + m^2$ on the second equation. This eliminates the bare propagators, and we obtain:

$$\begin{aligned} (\Box_x + m^2) G(x,y) &= -i\delta_c(x-y) - \int_{\mathcal{C}} d^4v \; \Sigma(x,v) \; G(v,y) \;, \\ (\Box_y + m^2) G(x,y) &= -i\delta_c(x-y) - \int_{\mathcal{C}} d^4v \; G(x,v) \; \Sigma(v,y) \;, \end{aligned} \tag{13.81}$$

where $\delta_c(x-y)$ is the generalization of the delta function to the contour $\mathcal{C}$. This is one of the forms of the Kadanoff-Baym equations.

13.4.3 From QFT to kinetic theory

Kinetic theory is an approximation of the underlying dynamics in terms of a space-time dependent distribution of particles $f(x,\mathbf{p})$. One may note right away that this is necessarily an approximate description, because it is not possible to define simultaneously the position and momentum of a particle.

In the Kadanoff-Baym equations (13.81), the dressed propagator G and the self-energy Σ are in general not invariant under translations, precisely because the system is out-of-equilibrium. Therefore, one may not Fourier transform them in the usual way. Instead, one uses a *Wigner transform*, defined as follows

$$F(X,p) \equiv \int d^4s \; e^{ip\cdot s} \; F\Big(X+\frac{s}{2}, X-\frac{s}{2}\Big) \;, \tag{13.82}$$

where $F(x,y)$ is a generic 2-point function (we use the same symbol for its Wigner transform, since the arguments are sufficient to distinguish them). In other words, the Wigner transform is an usual Fourier transform with respect to the separation $s \equiv x-y$, and the result still depends on the mid-point $X \equiv (x+y)/2$. Note that in eq. (13.82), the time integration is over the real axis, not over the contour $\mathcal{C}$. Wigner transforms do not share with the Fourier transform their properties with respect to convolution. Given two 2-point functions F and G, let us define:

$$H(x,y) \equiv \int d^4z \; F(x,z) \, G(z,y) \;. \tag{13.83}$$

The Wigner transform of H is given by

$$H(X,p) = F(X,p)\,\exp\left\{\frac{i}{2}\left[\,\overleftarrow{\partial}_x\overrightarrow{\partial}_p - \overrightarrow{\partial}_x\overleftarrow{\partial}_p\,\right]\right\} G(X,p)\;, \tag{13.84}$$

where the arrows indicate on which side the corresponding derivative acts. The right hand side of this formula reduces to the ordinary product of the transforms when there is no X dependence, i.e. when the functions F and G are translation invariant. The first correction to the translation invariant case is proportional to the Poisson bracket of F and G,

$$H(X,p) = F(X,p)G(X,p) + \frac{i}{2}\big\{F(X,p),G(X,p)\big\} + \cdots\;. \tag{13.85}$$

The derivatives with respect to x and y that appear in the Kadanoff-Baym equations can be written in terms of derivatives with respect to X and s :

$$\begin{aligned}
&\partial_x = \frac{1}{2}\partial_X + \partial_s \quad , \quad \partial_y = \frac{1}{2}\partial_X - \partial_s \\
&\Box_x = \frac{1}{4}\Box_X + \partial_X\cdot\partial_s + \Box_s \quad , \quad \Box_y = \frac{1}{4}\Box_X - \partial_X\cdot\partial_s + \Box_s\;.
\end{aligned} \tag{13.86}$$

In these operators, the Wigner transform just amounts to a substitution

$$\partial_s \;\;\to\;\; -ip \quad , \qquad \Box_s \;\;\to\;\; -p^2\;. \tag{13.87}$$

In order to go from the Kadanoff-Baym equations to kinetic equations, two approximations are necessary:

1. **Gradient approximation :** $p \sim \partial_s \gg \partial_X$. The derivatives with respect to the mid-point X characterize the space and time scales over which the properties of the system (e.g. its particle distribution) change significantly. This approximation therefore means that this scale, that characterize the off-equilibriumness of the system, should be much larger than the De Broglie wavelength of the particles. Another way to state this approximation is that the mean free path in the system should be much larger than the wavelength of the particles, which amounts to a certain diluteness of the system. Using this approximation in the two Kadanoff-Baym equations (13.81), taking their difference, and breaking it down into its $++$, $--$, $+-$ and $-+$ components, one obtains

$$\begin{aligned}
&-2ip\cdot\partial_x\,(G_{+-}(X,p) - G_{-+}(X,p)) = 0\;, \\
&-2ip\cdot\partial_x\,(G_{+-}(X,p) + G_{-+}(X,p)) = 2\Big[G_{-+}\Sigma_{+-} - G_{+-}\Sigma_{-+}\Big]\;.
\end{aligned} \tag{13.88}$$

2. **Quasi-particle approximation :** This approximation consists in assuming that the dressed propagators $G_{\epsilon\epsilon'}$ can be written in terms of a local particle distribution $f(X,\mathbf{p})$ as in eqs. (13.70). This is equivalent to

$$\begin{aligned}
&G_{-+}(X,p) = (1 + f(X,p))\;\rho(X,p)\;, \\
&G_{+-}(X,p) = f(X,p)\;\rho(X,p)\;,
\end{aligned} \tag{13.89}$$

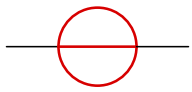
where $\rho(X,p) \equiv G_{-+}(X,p) - G_{+-}(X,p)$. This would be exact for non-interacting, infinitely long-lived, particles. In the presence of interactions, the approximation is justified when the time between two collisions of a particle is large compared to its wavelength.

Using eqs. (13.88) and (13.89), we obtain and equation for $f(X,\mathbf{p})$, which is nothing but a *Boltzmann equation*:

$$\Big[\partial_t + \boldsymbol{v}_{\mathbf{p}}\cdot\boldsymbol{\nabla}_{\boldsymbol{x}}\Big]\, f(X,p) = \underbrace{\frac{i}{2E_{\mathbf{p}}}\Big[(1+f(X,p))\Sigma_{+-} - f(X,p)\Sigma_{-+}\Big]}_{\mathbb{C}_{\mathbf{p}}[f;X]}\,, \tag{13.90}$$

where $\boldsymbol{v}_{\mathbf{p}} \equiv \mathbf{p}/E_{\mathbf{p}}$ is the velocity vector for particles of momentum $\mathbf{p}$. Note that the Boltzmann equation is spatially local since all the objects it contains are evaluated at the coordinate X, but its right hand side is non local in momentum. The right hand side, $\mathbb{C}_{\mathbf{p}}[f;X]$, is called the *collision term*. The combination $\partial_t + \boldsymbol{v}_{\mathbf{p}}\cdot\boldsymbol{\nabla}_{\boldsymbol{x}}$ that appears in the left hand side is called the transport derivative. It is zero on any function whose t and $\boldsymbol{x}$ dependence arise only in the combination $\boldsymbol{x} - \boldsymbol{v}_{\mathbf{p}}t$ (this is the case for a distribution of non-interacting particles, that move at the constant velocity $\boldsymbol{v}_{\mathbf{p}}$ prescribed by their momentum).

In order to obtain an explicit expression of the collision term, it is necessary to truncate the self-energies to a certain order (usually, the lowest order that gives a non-zero result) in the loop expansion. In a scalar theory with a ϕ^4 interaction, the self-energies should be evaluated at two-loops,

$\Sigma =$  . (13.91)

Using the Feynman rules of the Schwinger-Keldysh formalism, this diagram leads to the following collision term

$$\begin{aligned}\mathbb{C}_{\mathbf{p}}[f;X] \;=\; & \frac{\lambda^2}{4E_{\mathbf{p}}}\int\frac{d^3\mathbf{p}_1}{(2\pi)^3 2E_1}\int\frac{d^3\mathbf{p}_2}{(2\pi)^3 2E_2}\int\frac{d^3\mathbf{p}_3}{(2\pi)^3 2E_3}(2\pi)^4\delta(p-p_1-p_2-p_3)\\ & \times\Big[f(X,p_1)f(X,p_2)(1+f(X,p_3))(1+f(X,p))]\\ & \quad -f(X,p_3)f(X,p)(1+f(X,p_1))(1+f(X,p_2))\Big]\,.\end{aligned}$$

The expression describes the rate of change of the particle distribution, under the effect of 2-body elastic collisions. It is the difference between a production rate (coming from the term in which the particle of momentum $\mathbf{p}$ is produced, and thus weighted by a factor $1+f(X,\mathbf{p})$) and a destruction rate (from the term in the particle of momentum $\mathbf{p}$ is destroyed, and has a weight $f(x,\mathbf{p})$).

To close this section, let us mention an additional term that arises when the self-energy contains a local part, i.e. a term proportional to a delta function in space-time:

$$\Sigma(u,v) = \Phi(u)\delta_c(u-v) + \Pi(u,v) \tag{13.92}$$

The difference of the two Kadanoff-Baym equations will contain the combination $\Phi(y)G(x,y) - \Phi(x)G(x,y)$, whose Wigner transform at lowest order in the gradient approximation is

$$i\left(\partial_X\Phi(X)\right)\cdot\left(\partial_p G(X,p)\right)\,. \tag{13.93}$$

This extra term leads to a somewhat modified Boltzmann equation,

$$\Big[\partial_t + \boldsymbol{v}_{\mathbf{p}}\cdot\boldsymbol{\nabla}_{\boldsymbol{x}}\Big]\, f + \underline{\frac{1}{2E_{\mathbf{p}}}\partial_X\Phi\cdot\partial_p f} = \frac{i}{2E_{\mathbf{p}}}\Big[(1+f)\,\Pi_{+-} - f\,\Pi_{-+}\Big]\,. \tag{13.94}$$

In the new term (underlined), one may interpret $\partial_X \Phi$ as a mean force field acting on the particles. Under the action of this force, the particles accelerate which implies a change of their momentum. The left hand side of the above equation thus describes the change of the distribution of particles under the effect of this mean field, in the absence of any collisions (that are described by the right hand side).

Chapter 14

Strong fields and semi-classical methods

14.1 Introduction

Until now, all our discussion of quantum field theory has been centered on an expansion about the vacuum, i.e. on situations involving a system with few particles. This is also a regime in which the fields are in a certain sense[1] small. The connection between the field amplitude and the density of particles in a state may be grasped by writing the LSZ reduction formula that gives the expectation value of the number operator for a system whose initial state is $\big|\Phi_{\rm in}\big\rangle$. By mimicking the derivation of the section (1.4), one obtains easily

$$\begin{aligned}
\big\langle\Phi_{\rm in}\big|a^\dagger_{\mathbf{p},\rm out}a_{\mathbf{p},\rm out}\big|\Phi_{\rm in}\big\rangle &= \frac{1}{Z}\int d^4x d^4y\; e^{ip\cdot(x-y)}\;(\Box_x+m^2)(\Box_y+m^2) \\
&\qquad\qquad \times\big\langle\Phi_{\rm in}\big|\phi(x)\phi(y)\big|\Phi_{\rm in}\big\rangle \\
\big\langle\Phi_{\rm in}\big|\phi(x)\phi(y)\big|\Phi_{\rm in}\big\rangle &= \int\big[D\phi_\pm(z)\big]\;\phi_-(x)\phi_+(y)\;e^{i\,(S[\phi_+]-S[\phi_-])}\,, \qquad (14.1)
\end{aligned}$$

where in the second line we have sketched the path integral representation of the matrix element that appears in the reduction formula. Note that, since there is no time ordering in this matrix element, the Schwinger-Keldysh formalism must be used here. This formula is only a sketch, because the boundary conditions of the path integral at the initial time should be precised in order to properly account for the initial state $\big|\Phi_{\rm in}\big\rangle$. However, what we want to illustrate with these formulas is the direct relationship between large particle occupation numbers (the left hand side of the first equation), and large fields in a path integral. Moreover, in the path integral, the magnitude of the fields is controlled by the boundary conditions (this is the only thing that depends on the initial state of the system in the right hand side of the second equation).

There is an implicit assumption of weak fields in the perturbative machinery that we have studied so far, which is best viewed in the path integral formalism. For instance, in the second

[1] When we talk of small or large fields, we are referring to the magnitude of the c-number field in a path integral (it does not make sense to apply these qualifiers to the field operator itself).

of eqs. (14.1), the perturbative expansion amounts to writing $\mathcal{S} = \mathcal{S}_0 + \mathcal{S}_{\rm int}$, and to expand the exponentials in powers of $\mathcal{S}_{\rm int}$. In a scalar field theory with a quartic coupling, the interaction part of the action reads

$$\mathcal{S}_{\rm int}[\phi] = -\frac{\lambda}{4!}\int d^4x\ \phi^4(x)\ , \tag{14.2}$$

while the free action (that we keep inside the exponential) is given by

$$\mathcal{S}_0[\phi] = \frac{1}{2}\int d^4x\ \Big[(\partial_\mu\phi)(\partial^\mu\phi) - m^2\phi^2\Big]\ . \tag{14.3}$$

The common justification of the perturbative expansion is that, when the coupling constant λ is small, we have $\mathcal{S}_{\rm int} \ll \mathcal{S}_0$. However, since $\mathcal{S}_0[\phi]$ is quadratic in the field while $\mathcal{S}_{\rm int}[\phi]$ contains higher powers of ϕ, this inequality may not be true if the field is large, even at weak coupling. In order to make this statement more precise, we must account for the fact that the field has mass dimension 1. Let us denote by Q the typical momentum scale in the problem under consideration (for simplicity we assume that there is only one), and then we write

$$\phi(x) \sim \vartheta\, Q\ , \tag{14.4}$$

where ϑ is a dimensionless number that encodes the order of magnitude of the field. Naive dimensional analysis tells us that

$$\begin{aligned}(\partial_\mu\phi)(\partial^\mu\phi) &\sim \vartheta^2\, Q^4\ ,\\ \lambda\phi^4 &\sim \lambda\,\vartheta^4\, Q^4\ .\end{aligned} \tag{14.5}$$

For the interaction term to be small compared to the kinetic term, we must have

$$\lambda\,\vartheta^2 \ll 1\ , \tag{14.6}$$

which is slightly different from the usual criterion of small λ, since this condition depends on the field magnitude via ϑ. The purpose of this chapter is to explore situations of weak coupling (i.e. $\lambda \ll 1$) where the inequality (14.6) is not satisfied because of strong fields. We call this the *strong field regime* of quantum field theory. We will discuss two main situations where strong fields may occur:

- The initial state is a highly occupied state, such as a coherent state.
- The initial state is the ground state, but the system is driven by a strong external source.

As we shall see, since the coupling constant is assumed to be small, there is nevertheless a loop expansion, but each loop order (including the tree level approximation) is non-perturbative in a sense that we will clarify in the rest of the chapter.

14.2 Expectation values in a coherent state

In the section 1.14.5, we have presented the Schwinger-Keldysh formalism, that allows the evaluation of expectation values of an observable in the in- vacuum state, $\langle 0_{\rm in}|\mathcal{O}|0_{\rm in}\rangle$. In the

previous chapter, we have generalized this technique to expectation values in a thermal state, i.e. a mixed state whose density matrix is the canonical equilibrium one, $\rho \equiv \exp(-\beta\,\mathcal{H})$. Another generalization, that we shall consider in this section, is to consider an expectation value in a coherent state, which may be defined from the perturbative in-vacuum as follows

$$\big|\chi_{in}\big\rangle \equiv \mathcal{N}_\chi\ \exp\Big\{\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_\mathbf{k}}\ \chi(\mathbf{k})\ a^\dagger_{\mathbf{k},in}\Big\}\ \big|0_{in}\big\rangle\ , \tag{14.7}$$

where $\chi(\mathbf{k})$ is a function of 3-momentum and $\mathcal{N}_\chi$ a normalization constant adjusted so that $\big\langle\chi_{in}\big|\chi_{in}\big\rangle = 1$. From the canonical commutation relation

$$\big[a_{\mathbf{p},in},\, a^\dagger_{\mathbf{q},in}\big] = (2\pi)^3\ 2E_\mathbf{p}\ \delta(\mathbf{p}-\mathbf{q})\ , \tag{14.8}$$

it is easy to check the following identity

$$\begin{aligned} &a_{\mathbf{p},in}\ \big|\chi_{in}\big\rangle = \chi(\mathbf{p})\ \big|\chi_{in}\big\rangle\ ,\\ &\big|\mathcal{N}_\chi\big|^2 = \exp\Big\{-\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_\mathbf{k}}\big|\chi(\mathbf{k})\big|^2\Big\}\ . \end{aligned} \tag{14.9}$$

The first equation tells us that $\big|\chi_{in}\big\rangle$ is an eigenstate of annihilation operators, which is another definition of coherent states, and the second one provides the value of the normalization constant. The occupation number in the initial state is closely related to the function $\chi(\mathbf{k})$. Indeed, we have

$$\big\langle\chi_{in}\big|a^\dagger_{\mathbf{p},in}a_{\mathbf{p},in}\big|\chi_{in}\big\rangle = |\chi(\mathbf{p})|^2\ . \tag{14.10}$$

In other words, the number of particles in the mode of momentum $\mathbf{p}$ is the squared modulus of the function $\chi(\mathbf{p})$. A large χ thus corresponds to a highly occupied initial state (at the opposite, $\chi(\mathbf{p}) \equiv 0$ corresponds to the vacuum).

Consider now the generating functional for the extension of the Schwinger-Keldysh formalism in this coherent state,

$$\begin{aligned} Z_\chi[j] \ &\equiv\ \big\langle\chi_{in}\big|P\,\exp i\int_\mathcal{C} d^4x\ j(x)\phi(x)\big|\chi_{in}\big\rangle\\ &=\ \big\langle\chi_{in}\big|P\,\exp i\int_\mathcal{C} d^4x\ \Big[\mathcal{L}_{int}(\phi_{in}(x)) + j(x)\phi_{in}(x)\Big]\big|\chi_{in}\big\rangle\ , \end{aligned} \tag{14.11}$$

where $j(x)$ is a fictitious source that lives on the closed-time contour $\mathcal{C}$ introduced in the figure 1.1. As usual, the first step is to factor out the interactions as follows:

$$Z_\chi[j] = \exp i\int_\mathcal{C} d^4x\ \mathcal{L}_{int}\Big(\frac{\delta}{i\delta j(x)}\Big)\ \underbrace{\big\langle\chi_{in}\big|P\,\exp i\int_\mathcal{C} d^4x\ j(x)\phi_{in}(x)\big|\chi_{in}\big\rangle}_{Z_{\chi 0}[j]}\ . \tag{14.12}$$

A first application of the Baker-Campbell-Hausdorff formula enables one to remove the path ordering, which gives

$$\begin{aligned} Z_{\chi 0}[j] &= \big\langle\chi_{in}\big|\exp i\int_\mathcal{C} d^4x\ j(x)\phi_{in}(x)\big|\chi_{in}\big\rangle\\ &\quad\times\exp\Big\{-\frac{1}{2}\int_\mathcal{C} d^4x d^4y\ j(x)j(y)\ \theta_c(x^0-y^0)\ \big[\phi_{in}(x),\phi_{in}(y)\big]\Big\}\ , \end{aligned} \tag{14.13}$$

where $\theta_c(x^0 - y^0)$ generalizes the step function to the ordered contour $\mathcal{C}$. Note that the factor on the second line is a commuting number and thus can be removed from the expectation value. A second application of the Baker-Campbell-Hausdorff formula allows to normal-order the first factor. Decomposing the in-field as follows,

$$\phi_{\rm in}(x) \equiv \underbrace{\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\, a_{\mathbf{k},\rm in}\, e^{-ik\cdot x}}_{\phi^{(-)}_{\rm in}(x)} + \underbrace{\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\, a^\dagger_{\mathbf{k},\rm in}\, e^{+ik\cdot x}}_{\phi^{(+)}_{\rm in}(x)}\,, \tag{14.14}$$

we obtain the following expression for the free generating functional

$$\begin{aligned} Z_{\chi 0}[j] = {} & \langle\chi_{\rm in}| \exp\left\{ i\int_{\mathcal{C}} d^4x\; j(x)\phi^{(+)}_{\rm in}(x)\right\} \exp\left\{ i\int_{\mathcal{C}} d^4y\; j(y)\phi^{(-)}_{\rm in}(y)\right\} |\chi_{\rm in}\rangle \\ & \times \exp\left\{ +\frac{1}{2}\int_{\mathcal{C}} d^4x d^4y\; j(x)j(y)\left[\phi^{(+)}_{\rm in}(x), \phi^{(-)}_{\rm in}(y)\right]\right\} \\ & \times \exp\left\{ -\frac{1}{2}\int_{\mathcal{C}} d^4x d^4y\; j(x)j(y)\;\theta_c(x^0-y^0)\;\left[\phi_{\rm in}(x), \phi_{\rm in}(y)\right]\right\}. \end{aligned} \tag{14.15}$$

The factor of the first line can be evaluated by using the fact that the coherent state is an eigenstate of annihilation operators:

$$\begin{aligned} & \langle\chi_{\rm in}| \exp\left\{ i\int_{\mathcal{C}} d^4x\; j(x)\phi^{(+)}_{\rm in}(x)\right\} \exp\left\{ i\int_{\mathcal{C}} d^4y\; j(y)\phi^{(-)}_{\rm in}(y)\right\} |\chi_{\rm in}\rangle \\ & \quad = \exp\left\{ i\int_{\mathcal{C}} d^4x\; j(x)\;\underbrace{\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\left(\chi(\mathbf{k})e^{-ik\cdot x} + \chi^*(\mathbf{k})e^{+ik\cdot x}\right)}_{\Phi_\chi(x)}\right\}. \end{aligned} \tag{14.16}$$

We denote $\Phi_\chi(x)$ the field obtained by substituting the creation and annihilation operators of the in-field by $\chi^*(\mathbf{k})$ and $\chi(\mathbf{k})$ respectively. Note that this is no longer an operator, but a (real valued) c-number field. Moreover, because it is a linear superposition of plane waves, this field is a free field:

$$(\Box_x + m^2)\,\Phi_\chi(x) = 0\,. \tag{14.17}$$

The second and third factors of eq. (14.15) are commuting numbers, provided we do not attempt to disassemble the commutators. Using the decomposition of the in-field in terms of creation and annihilation operators, and the canonical commutation relation of the latter, we obtain

$$\begin{aligned} & \theta_c(x^0-y^0)\;\left[\phi_{\rm in}(x), \phi_{\rm in}(y)\right] - \left[\phi^{(+)}_{\rm in}(x), \phi^{(-)}_{\rm in}(y)\right] \\ & \quad = \underbrace{\theta_c(x^0-y^0)\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\, e^{-ik\cdot(x-y)} + \theta_c(y^0-x^0)\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\, e^{+ik\cdot(x-y)}}_{G^0_c(x,y)}\,, \end{aligned} \tag{14.18}$$

which is nothing but the usual bare path-ordered propagator $G^0_c(x,y)$. Collecting all the factors, the generating functional for path-ordered Green's functions in the Schwinger-Keldysh formalism with an initial coherent state reads

$$\begin{aligned} Z_\chi[j] &= \exp\Big\{i\int_{\mathcal{C}} d^4x\, \mathcal{L}_{\rm int}\Big(\frac{\delta}{i\delta j(x)}\Big)\Big\}\, \underline{\exp\Big\{i\int_{\mathcal{C}} d^4x\, j(x)\,\Phi_\chi(x)\Big\}} \\ &\quad\times \exp\Big\{-\frac{1}{2}\int_{\mathcal{C}} d^4x d^4y\, j(x)j(y)\; G^0_c(x,y)\Big\}\,. \end{aligned} \tag{14.19}$$

It differs from the corresponding functional with the perturbative vacuum[2] as initial state only by the second factor, that we have underlined. This generating functional is also equal to[3]

$$\begin{aligned} Z_\chi[j] &= \exp\Big\{i\int_{\mathcal{C}} d^4x\, j(x)\,\Phi_\chi(x)\Big\}\, \exp\Big\{i\int_{\mathcal{C}} d^4x\, \mathcal{L}_{\rm int}\Big(\Phi_\chi(x)+\frac{\delta}{i\delta j(x)}\Big)\Big\} \\ &\quad\times \exp\Big\{-\frac{1}{2}\int_{\mathcal{C}} d^4x d^4y\, j(x)j(y)\; G^0_c(x,y)\Big\}\,. \end{aligned} \tag{14.20}$$

The first factor has the effect of shifting the fields by $\Phi_\chi(x)$. The simplest way to see this is to write

$$\phi \equiv \Phi_\chi + \zeta\,. \tag{14.21}$$

In the definition (14.11), this leads to

$$Z_\chi[j] = \exp\Big\{i\int_{\mathcal{C}} d^4\, j(x)\,\Phi_\chi(x)\Big\}\,\big\langle\chi_{\rm in}\big|P\,\exp i\int_{\mathcal{C}} d^4x\, j(x)\,\zeta(x)\big|\chi_{\rm in}\big\rangle\,, \tag{14.22}$$

where the second factor in the right hand side is the generating functional for correlators of ζ. Comparing with eq. (14.20), we see that the generating functional for ζ is identical to the vacuum one, except that the argument ϕ of the interaction Lagrangian is replaced by $\Phi_\chi+\zeta$:

$$\mathcal{L}_{\rm int}(\phi) \quad\to\quad \mathcal{L}_{\rm int}(\Phi_\chi+\zeta)\,. \tag{14.23}$$

In other words, the field ζ appears to be coupled to a background field Φ_χ. For instance, for a ϕ^4 interaction term, we have

$$\mathcal{L}_{\rm int}(\Phi_\chi+\zeta) = -\lambda\left\{\frac{\zeta^4}{4!}+\frac{\zeta^3\Phi_\chi}{8}+\frac{\zeta^2\Phi_\chi^2}{4}+\frac{\zeta\Phi_\chi^3}{8}+\frac{\Phi_\chi^4}{4!}\right\}. \tag{14.24}$$

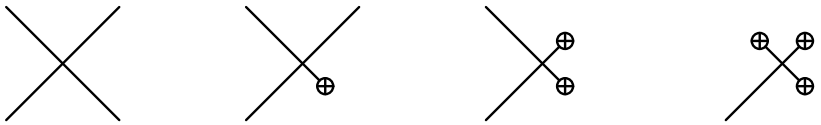
The first term, in ζ^4, gives the usual four-leg vertex in the Feynman rules, and the following terms describe the interactions of ζ with the background field Φ_χ. The last term plays no role since it does not contain the quantum field ζ. Except for the appearance of these new vertices that involve a background field, the Feynman rules are the same as in the Schwinger-Keldysh formalism for a vacuum initial state, with $+$ and $-$ vertices, and bare propagators G^0_{++}, G^0_{+-},

[2] The vacuum initial state corresponds to the function $\chi(k)\equiv 0$, i.e. to $\Phi_\chi(x)=0$.

[3] In this transformation, we use the functional analogue of

$$F(\partial_x)\,e^{\alpha x}\,G(x) = e^{\alpha x}\,F(\alpha+\partial_x)\,G(x)\,.$$

Figure 14.1: *Vertices that appear in the perturbative expansion for the calculation of expectation values with a coherent initial state. The circled cross denotes the field Φ_{χ}.*



G^0_{-+} and G^0_{--} to connect them. In summary, replacing the vacuum initial state by a coherent state amounts to extend the usual Schwinger-Keldysh formalism with a background field Φ_{χ}.

As in eq. (14.4), let us assume for the purpose of power counting that

$$\Phi_{\chi} \sim \vartheta Q \,, \tag{14.25}$$

and consider a connected graph $\mathcal{G}$ made of n_{E} external lines, n_{I} internal lines, n_1 vertices $\zeta\Phi_{\chi}^3$, n_2 vertices $\zeta^2\Phi_{\chi}^2$, n_3 vertices $\zeta^3\Phi_{\chi}$, n_4 vertices ζ^4, and n_{L} loops. These parameters are related by the following two identities:

$$\begin{aligned} n_{\text{E}} + 2n_{\text{I}} &= 4n_4 + 3n_3 + 2n_2 + n_1 \,, \\ n_{\text{L}} &= n_{\text{I}} - (n_1 + n_2 + n_3 + n_4) + 1 \,. \end{aligned} \tag{14.26}$$

Then, the order in λ and ϑ of this graph is given by

$$\begin{aligned} \mathcal{G} &\sim \lambda^{n_1+n_2+n_3+n_4}\, \vartheta^{3n_1+2n_2+n_3} \\ &\sim \lambda^{n_{\text{L}}-1+n_{\text{E}}/2}\, \left(\sqrt{\lambda}\,\vartheta\right)^{3n_1+2n_2+n_3} \,. \end{aligned} \tag{14.27}$$

The first factor is nothing but the usual order in λ of a connected graph with n_{E} external lines and n_{L} loops. The second factor counts the number of insertions $(3n_1 + 2n_2 + n_3)$ of the background field Φ_{χ}. Interestingly, it involves only the combination $\sqrt{\lambda}\,\vartheta$, that appears also in the inequality (14.6) that delineates the strong field regime. From eq. (14.27), we can draw the following conclusions:

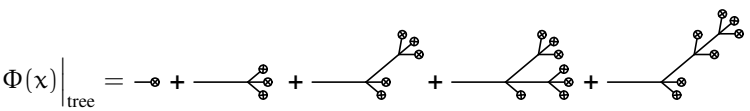
- When $\lambda\vartheta^2 \ll 1$, i.e. in the weak field regime, we can make a double perturbative expansion in λ and in ϑ (i.e. in the occupation of the initial coherent state). Leading order results correspond to tree diagrams with zero (or the minimal number necessary for the observable under consideration to be non-zero) insertions of the background field.
- When $\lambda\vartheta^2 \gtrsim 1$, i.e. in the strong field regime, the expansion in powers of λ is still possible (and is organized by the number of loops in the graphs). But the expansion in powers of the background field becomes illegitimate, and one should instead treat Φ_{χ} to all orders. As we shall see now, this leads to important modifications in the calculation of observables in the strong field regime.

Note that for a system prepared in a coherent initial state, it is the function $\chi(\mathbf{k})$ that defines the coherent that determines whether we are in the weak or strong field regime.

In order to illustrate the changes to the perturbative expansion in the strong field regime, let us consider a very simple observable, the expectation value of the field operator,

$$\Phi(x) \equiv \langle \chi_{in} | \phi(x) | \chi_{in} \rangle = \Phi_\chi(x) + \langle \chi_{in} | \zeta(x) | \chi_{in} \rangle \, . \tag{14.28}$$

The beginning of the diagrammatic representation of $\Phi(x)$ at tree level reads:

$\Phi(x)\Big|_{\text{tree}} =$  (14.29)

In fact, at tree level, $\Phi(x)$ is the sum of all the tree diagrams (weighted by the appropriate symmetry factor) whose root is the point x and whose leaves are the coherent field Φ_χ. This infinite set of trees can be generated recursively by the following integral representation:

$$\Phi(x) \;=\; \Phi_\chi(x) + i \int d^4x \, \Big[\underbrace{G^0_{++}(x,y) - G^0_{+-}(x,y)}_{G^0_R(x,y)} \Big] \Big(\underbrace{-\frac{\lambda}{6} \Phi^3(y)}_{U'(\Phi(y))} \Big) \, . \tag{14.30}$$

Interestingly, after one has summed over the $+$ and $-$ indices carried by the vertices, the propagators G^0_{++} and G^0_{+-} of the Schwinger-Keldysh diagrammatic rules always appear via their difference, which is nothing but the bare retarded propagator:

$$G^0_{++}(x,y) - G^0_{+-}(x,y) = G^0_{-+}(x,y) - G^0_{--}(x,y) = G^0_R(x,y) \, . \tag{14.31}$$

Since this propagator obeys

$$(\Box_x + m^2) \, G^0_R(x,y) = -i \, \delta(x-y) \, , \quad G^0_R(x,y) = 0 \quad \text{if } x^0 < y^0 \, , \tag{14.32}$$

the expectation value $\Phi(x)$ at tree level satisfies

$$\begin{aligned} &(\Box_x + m^2) \, \Phi(x) + U'(\Phi(x)) = 0 \, , \\ &\lim_{x^0 \to -\infty} \Phi(x) = \Phi_\chi(x) \, . \end{aligned} \tag{14.33}$$

In other words, at tree level, the field expectation value obeys the classical field equation of motion, with the boundary value $\Phi_\chi(x)$ at the initial time. The non-linearity of this equation of motion is crucial in the strong field regime, and all the terms of the series (14.29) have the same magnitude when $\lambda \vartheta^2 \sim 1$. Nevertheless, the representation of this series as the solution of the classical field equation of motion with a retarded boundary condition is very useful, since it turns the problem of summing an infinite series of Feynman graphs into the much simpler (at least numerically) problem of solving a partial differential equation.

This result for the expectation value of $\phi(x)$ generalizes to the expectation value of any observable built from the field operator: at tree level, its expectation value is obtained by replacing the operator $\phi(x)$ by the c-number classical field $\Phi(x)$ inside the observable:

$$\langle \chi_{in} | \mathcal{O}\big(\phi(x)\big) | \chi_{in} \rangle \underset{\text{tree level}}{=} \mathcal{O}\big(\Phi(x)\big) \, . \tag{14.34}$$

We will defer the study of loop corrections to these expectation values until the section 14.4, because this discussion will be common with another strong field situation that we shall discuss first, namely the case of quantum field theories coupled to a strong external source.

14.3 Quantum field theory with external sources

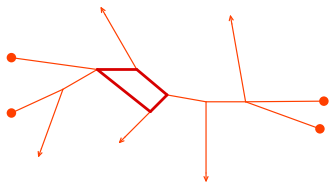
Let us now consider a second way to reach the large field regime. This time, the initial state of the system is the vacuum, but the field is coupled to an external source that drives the system away from the ground state. When the external source is large, the field expectation value will eventually become large itself, and the system will again be in the strong field regime. Let us consider a scalar field theory with quartic interaction coupled to a source J, whose Lagrangian is

$$\mathcal{L} \equiv \frac{1}{2}(\partial_\mu\phi)(\partial^\mu\phi) - \frac{1}{2}m^2\phi^2 - \underbrace{\frac{\lambda}{4!}\phi^4}_{U(\phi)} + J\phi \,. \tag{14.35}$$

Although we consider here the example of a ϕ^4 interaction term, we will often write the equations for a generic potential $U(\phi)$, and sometimes diagrammatic illustrations will be given for a cubic interaction for simplicity. These more general interactions terms will be defined as $\lambda^{-1+n/2}Q^{n-4}\phi^n$, where Q is an object of mass dimension 1. The Feynman rules for this theory are the usual ones, with the addition of a special rule for the external current J. In momentum space, a source j attached to the end of a propagator of momentum p contributes a factor $i\tilde{J}(p)$ (where $\tilde{J}$ is the Fourier transform of J).

The source $J(x)$ is a given function of space-time, fixed once for all. As we shall see shortly, the strong field regime corresponds to large sources $J \sim \lambda^{-1/2}$ – we all call this situation the *strong source dense*. In contrast, the situation where the external source J is small is called the *weak source dilute*. Consider a simply connected diagram (see figure 14.2), with n_E external

Figure 14.2: *Generic connected graph in the strong source regime. In this example,* $n_E = 5$, $n_I = 11$, $n_J = 4$, $n_L = 1$, $n_3 = 5$ *and* $n_4 = 2$.



legs, n_I internal lines, n_L independent loops, n_J sources, and n_3 cubic vertices, $n_4^{(4)}$ quartic vertices, etc... These parameters are not all independent. First, the number of propagator endpoints should match the available sites to which they can be attached. This leads to a first identity,

$$n_E + 2n_I = n_J + 3n_3 + 4n_4 + 5n_5 + \cdots \tag{14.36}$$

A second identity expresses the number of independent loops in terms of the other parameters,

$$n_L = n_I - (n_3 + n_4 + n_5 + \cdots) - n_J + 1 \,. \tag{14.37}$$

Thanks to these two relations, the order of a diagram $\mathcal{G}$ can be written as

$$\mathcal{G} \sim J^{n_J}\, \lambda^{\frac{1}{2}n_3+n_4+\frac{3}{2}n_5+\cdots} = \lambda^{n_L-1+N_E/2}\left(\sqrt{\lambda}\,J\right)^{n_J} . \tag{14.38}$$

This formula is very similar to eq. (14.27). First, it does not depend on the number of vertices and on the number of internal lines; only the number of external legs, the number of loops and the number of sources appear in the result. The strong source regime is the regime where it is not legitimate to expand in powers of J because the factor $\sqrt{\lambda}\,J$ is not small. In this case, the order of a diagram does not depend on its number of sources, and an infinite number of diagrams –with fixed n_E and n_L but arbitrary n_J– contribute at each order.

14.4 Observables at LO and NLO

Leading order : Let us consider an observable $\mathcal{O}(\phi)$, possibly non-local but with fields only at the same time t_f (the discussion could be generalized to fields with only space-like separations). At leading order in the strong field regime. As we have seen in the previous sections, this can be achieved by the presence of strong external sources, or by starting from a highly occupied coherent state. In both case, the calculation of expectation values is done with the Schwinger-Keldysh formalism. Note that since the field operators in the observable are taken at equal times, they commute and the result does not depend on the $+$ or $-$ assignments for those fields. But it is crucial to sum over all the $\pm$ indices in the internal vertices of the graphs.

At leading order in λ, its expectation value is obtained by simply replacing the field operator ϕ by the solution Φ of the classical equations of motion,

$$\left\langle \mathcal{O}(\phi)\right\rangle_{LO} = \mathcal{O}(\Phi) , \tag{14.39}$$

with

$$\begin{aligned} &(\Box_x + m^2)\Phi + U'(\Phi) = J , \\ &\lim_{x^0\to t_i} \Phi(x) = \Phi_\chi(x) . \end{aligned} \tag{14.40}$$

(We have combined in a single description the two situations, with an external source J and starting from a non-trivial coherent state $|\chi_{in}\rangle$.) Note that it is the internal sums over the $\pm$ indices of the Schwinger-Keldysh formalism that lead to retarded boundary conditions, by virtue of eq. (14.31).

Next-to-leading order : For such an observable, the corresponding next-to-leading order correction can be formally written as follows,

$$\left\langle \mathcal{O}(\phi)\right\rangle_{NLO} = \int_{t_f} d^3\mathbf{x}\; \delta\Phi(x)\, \frac{\delta\mathcal{O}(\Phi)}{\delta\Phi(x)} + \frac{1}{2}\int_{t_f} d^3\mathbf{x} d^3\mathbf{y}\; \mathcal{G}(x,y)\, \frac{\delta^2\mathcal{O}(\Phi)}{\delta\Phi(x)\delta\Phi(y)} , \tag{14.41}$$

where $\delta\Phi$ is the 1-loop correction to the classical field Φ, and $\mathcal{G}(x,y)$ is the propagator dressed by the background field Φ. The two contributions of eq. (14.41) are illustrated in the figure 14.3. Since the fields operators in the observable $\mathcal{O}(\phi)$ are all separated by space-like intervals, it is

Figure 14.3: *The two contributions to observables at NLO in the strong field regime.*

not necessary to indicate the $\pm$ indices in $\delta\Phi$ and $\mathcal{G}$, and we have in fact:

$$\begin{aligned}
&\delta\Phi_+(x) = \delta\Phi_-(x)\,, \\
&\mathcal{G}_{++}(x,y) = \mathcal{G}_{--}(x,y) = \mathcal{G}_{-+}(x,y) = \mathcal{G}_{+-}(x,y) \quad \text{if } (x-y)^2 < 0\,.
\end{aligned} \tag{14.42}$$

Let us start with $\delta\Phi_\pm$. The propagators in the diagram on the left of the figure 14.3 are the Schwinger-Keldysh propagators in the presence of a background field Φ, i.e. the propagators $\mathcal{G}_{\epsilon\epsilon'}$. For a generic interaction potential, we can write $\delta\Phi_\pm(x)$ as follows:

$$\delta\Phi_\epsilon(x) = -\frac{i}{2}\sum_{\epsilon'=\pm}\int d^4z\;\epsilon'\;\mathcal{G}_{\epsilon\epsilon'}(x,z)\,U'''(\Phi(z))\,\mathcal{G}_{\epsilon'\epsilon'}(z,z)\,. \tag{14.43}$$

In this formula, the $1/2$ is a symmetry factor, the factor ϵ' in the integrand takes into account the fact that vertices of type $-$ have an opposite sign in the Schwinger-Keldysh formalism, and the factor $-i\,U'''(\Phi(z))$ is the general form of the 3-particle vertex in the presence of an external field (for an arbitrary interaction potential U).

Thus, we have reduced the calculation to that of the 2-point functions $\mathcal{G}_{\pm\pm}$. These four propagators are defined recursively by the following equations :

$$\mathcal{G}_{\epsilon\epsilon'}(x,y) = G^0_{\epsilon\epsilon'}(x,y) - i\sum_{\eta=\pm}\eta\int d^4z\;G^0_{\epsilon\eta}(x,z)\,U''(\Phi(z))\,\mathcal{G}_{\eta\epsilon'}(z,y)\,. \tag{14.44}$$

Here, $-i\,U''(\Phi(z))$ is the general form for the insertion of a background field on a propagator in a theory with potential $U(\Phi)$. From these equations, we obtain the following equations :

$$\begin{aligned}
\left[\Box_x + m^2 + U''(\Phi(x))\right]\mathcal{G}_{+-}(x,y) &= \left[\Box_y + m^2 + U''(\Phi(y))\right]\mathcal{G}_{+-}(x,y) = 0\,, \\
\left[\Box_x + m^2 + U''(\Phi(x))\right]\mathcal{G}_{-+}(x,y) &= \left[\Box_y + m^2 + U''(\Phi(y))\right]\mathcal{G}_{-+}(x,y) = 0\,.
\end{aligned} \tag{14.45}$$

In addition to these equations of motion, these propagators must become equal to their free counterparts G^0_{+-} and G^0_{-+} when $x^0, y^0 \to -\infty$. From the definition of the various components of the Schwinger-Keldysh propagators, $\mathcal{G}_{++}$ and $\mathcal{G}_{--}$ are given in terms of $\mathcal{G}_{+-}$ and $\mathcal{G}_{-+}$ by the following expressions:

$$\begin{aligned}
\mathcal{G}_{++}(x,y) &= \theta(x^0-y^0)\,\mathcal{G}_{-+}(x,y) + \theta(y^0-x^0)\,\mathcal{G}_{+-}(x,y)\,, \\
\mathcal{G}_{--}(x,y) &= \theta(x^0-y^0)\,\mathcal{G}_{+-}(x,y) + \theta(y^0-x^0)\,\mathcal{G}_{-+}(x,y)\,.
\end{aligned} \tag{14.46}$$

The above conditions determine $\mathcal{G}_{+-}$ and $\mathcal{G}_{-+}$ uniquely. In order to find these propagators, let us recall the following representation of their bare counterparts :

$$\begin{aligned} G^0_{+-}(x,y) &= \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}}\, a_{-\mathbf{p}}(x) a_{+\mathbf{p}}(y)\,, \\ G^0_{-+}(x,y) &= \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}}\, a_{+\mathbf{p}}(x) a_{-\mathbf{p}}(y)\,, \end{aligned} \tag{14.47}$$

where

$$(\Box_x + m^2)\, a_{\pm\mathbf{p}}(x) = 0\,, \quad \lim_{x^0\to-\infty} a_{\pm\mathbf{p}}(x) = e^{\mp ip\cdot x}\,. \tag{14.48}$$

It is trivial to generalize this representation of the off-diagonal propagators to the case of a non zero background field, by writing

$$\begin{aligned} \mathcal{G}_{+-}(x,y) &= \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}}\, a_{-\mathbf{p}}(x) a_{+\mathbf{p}}(y)\,, \\ \mathcal{G}_{-+}(x,y) &= \int \frac{d^3\mathbf{p}}{(2\pi)^3 2E_{\mathbf{p}}}\, a_{+\mathbf{p}}(x) a_{-\mathbf{p}}(y)\,, \end{aligned} \tag{14.49}$$

with

$$\left[\Box_x + m^2 + U''(\Phi(x))\right] a_{\pm\mathbf{p}}(x) = 0\,, \quad \lim_{x^0\to-\infty} a_{\pm\mathbf{p}}(x) = e^{\mp ip\cdot x}\,. \tag{14.50}$$

By construction, these expressions of $\mathcal{G}_{+-}$ and $\mathcal{G}_{-+}$ obey the appropriate equations of motion, and go to the correct limit in the remote past. The functions $a_{\pm\mathbf{p}}(x)$ are sometimes called *mode functions*. They provide a complete basis for the linear space of solutions of the equation (14.50), i.e. the space of linearized perturbations to the classical solution of the field equation of motion.

Relationship between LO and NLO : At this point, we have all the building blocks in order to obtain the single inclusive spectrum at NLO. One can go further and obtain a formal relationship between the LO and NLO inclusive spectra. A key observation for this is that the functions $a_{\mathbf{k}}$ that appear in the dressed propagators $\mathcal{G}_{\pm\mp}$ can be obtained from the classical field Φ as follows:

$$a_{\pm\mathbf{k}}(x) = \mathbb{T}_{\pm\mathbf{k}}\, \Phi(x)\,, \tag{14.51}$$

where the operator $\mathbb{T}_{\pm\mathbf{k}}$ is defined by

$$\mathbb{T}_{\pm\mathbf{k}} \cdots \equiv \int\limits_{u^0=-\infty} d^3\mathbf{u}\, e^{\mp ik\cdot u} \left[\frac{\delta}{\delta\Phi_{\rm ini}(u)} \mp iE_{\mathbf{k}} \frac{\delta}{\delta(\partial^0\Phi_{\rm ini}(u))}\right] \cdots \Big|_{\Phi_{\rm ini}\equiv\Phi_{\chi}}\,. \tag{14.52}$$

In words, the operator $\mathbb{T}_{\pm\mathbf{k}}$ in eq. (14.51) differentiates the classical field Φ with respect to its initial condition $\Phi_{\rm ini}$, and replaces it by the initial condition of $a_{\pm\mathbf{k}}$. Since $a_{\pm\mathbf{k}}$ is a *linear* perturbation to Φ, this indeed gives the correct result.

Thus, the propagator $\mathcal{G}_{+-}(x,y)$ that enters at NLO can be written as

$$\mathcal{G}_{+-}(x,y) = \int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}} \left[\mathbb{T}_{-\mathbf{k}}\,\Phi(x)\right]\left[\mathbb{T}_{+\mathbf{k}}\,\Phi(y)\right] . \tag{14.53}$$

In the rest of our NLO calculation, we only need this propagator for a space-like separation between x and y, which implies that $\mathcal{G}_{+-}(x,y) = \mathcal{G}_{-+}(x,y)$. In this case, we can symmetrize the expression of the propagator as follows:

$$\mathcal{G}_{+-}(x,y) = \frac{1}{2}\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}} \left\{\left[\mathbb{T}_{-\mathbf{k}}\,\Phi(x)\right]\left[\mathbb{T}_{+\mathbf{k}}\,\Phi(y)\right] + \left[\mathbb{T}_{+\mathbf{k}}\,\Phi(x)\right]\left[\mathbb{T}_{-\mathbf{k}}\,\Phi(y)\right]\right\} . \tag{14.54}$$

As we shall see now, a similar expression can be obtained for $\delta\Phi_{\pm}$. Let us start from eq. (14.43). Since the propagators $\mathcal{G}_{++}$ and $\mathcal{G}_{--}$ are equal when the two endpoints are evaluated at equal times, we have

$$\delta\Phi_\epsilon(x) = -\frac{i}{2}\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}} d^4z \underbrace{\left[\mathcal{G}_{\epsilon+}(x,z) - \mathcal{G}_{\epsilon-}(x,z)\right]}_{\mathcal{G}_{\text{R}}(x,z)} U'''(\Phi(z))\; a_{-\mathbf{k}}(z) a_{+\mathbf{k}}(z) , \tag{14.55}$$

where $\mathcal{G}_{\text{R}}$ is the retarded propagator in the presence of the background field Φ. By writing more explicitly the interactions with the background field,

$$\begin{aligned}\delta\Phi_\epsilon(x) \;\;=\;\; -i\int d^4y\; G^0_{\text{R}}(x,y)\Big[U''(\Phi(y))\,\delta\Phi_\epsilon(y) \\ + \frac{1}{2}U'''(\Phi(y))\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\; a^*_{\mathbf{k}}(y) a_{\mathbf{k}}(y)\Big] ,\end{aligned} \tag{14.56}$$

(with G^0_{R} the bare retarded propagator), one may prove that

$$\delta\Phi_\epsilon(x) = \frac{1}{2}\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\; \mathbb{T}_{+\mathbf{k}}\mathbb{T}_{-\mathbf{k}}\;\Phi(x) . \tag{14.57}$$

By inserting this expression, as well as eq. (14.54), in eq. (14.41), we can write the NLO expectation value as follows,

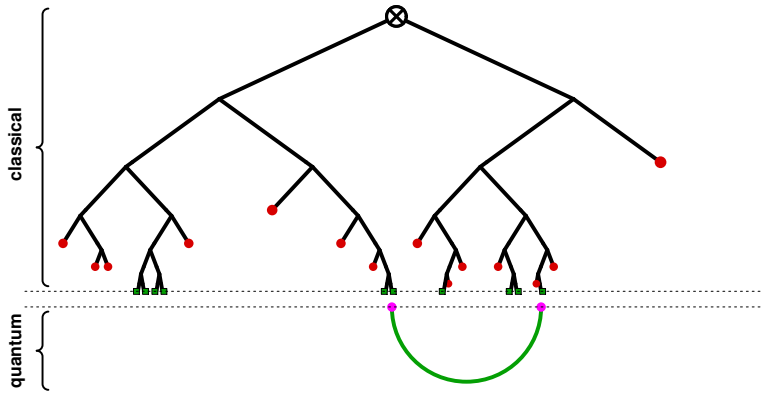
$$\left\langle \mathcal{O}\right\rangle_{\text{NLO}} = \left[\frac{1}{2}\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\; \mathbb{T}_{+\mathbf{k}}\mathbb{T}_{-\mathbf{k}}\right]\left\langle \mathcal{O}\right\rangle_{\text{LO}} . \tag{14.58}$$

This central result is illustrated in the figure 14.4. Some remarks should be made about this formula:

i. In this formula, the LO observable that appears in the right hand side must be considered as a functional of the initial classical field.

ii. The LO and NLO observables cannot be obtained in closed analytical form, because they contain the classical field Φ – retarded solution of a non-linear partial differential equation that cannot be solved analytically in general. Nevertheless, eq. (14.58) is an exact relationship between the two.

Figure 14.4: *Illustration of eq. (14.58). The pink dots represent the operator $\mathbb{T}_k(u)\mathbb{T}_{-k}(v)$. Their action is to remove two instances of the initial classical field (the green squares), and to connect them with the green link to form a loop.*



Why is the NLO "almost classical"? : In a sense, eq. (14.58) indicates that observables at NLO in the strong field regime are almost classical, since they can be obtained from the LO result (that depends only on the classical field Φ) by acting with the operators $\mathbb{T}_{\pm k}$ (i.e. derivatives with respect to the initial value of the classical field). If one had kept track of the powers of $\hbar$, the $\hbar$ that comes at NLO would just be an overall prefactor (the prefactor $1/2$ in eq. (14.58) would become $\hbar/2$), but all the rest of the formula would not contain any $\hbar$.

This is in fact not specific to the strong field regime nor to quantum field theory, but is a general property of quantum mechanics. To see this, consider a generic quantum system of Hamiltonian H and density operator ρ_t. The latter evolves according to the Liouville-von Neumann equation:

$$i\hbar \frac{\partial \rho_t}{\partial t} = \big[H, \rho_t\big] \ . \tag{14.59}$$

The next step is to introduce the Wigner transforms of the density operator:

$$W_t(\mathbf{x}, \mathbf{p}) \equiv \int ds \ e^{i\mathbf{p}\cdot\mathbf{s}} \ \big\langle \mathbf{x} + \frac{\mathbf{s}}{2} \big| \rho_t \big| \mathbf{x} - \frac{\mathbf{s}}{2} \big\rangle \ . \tag{14.60}$$

The Wigner transform of an operator is a Fourier transform of the matrix elements of the operator in the position basis with respect to the difference of coordinates. The function $W_t(\mathbf{x}, \mathbf{p})$ may be viewed in a loose sense[4] as a probability distribution in the classical phase-space of the system ($\mathbf{x}$ and $\mathbf{p}$ are classical variables, not operators). Note that the Wigner transform of the Hamiltonian operator H is the classical Hamiltonian $\mathcal{H}$. On may show that the Liouville-von

[4] W_t is not a bona fide probability distribution, because it is not positive definite in general. But the regions of phase-space where it is negative are small, typically of order $\hbar$. After being integrated either over $\mathbf{x}$ or over $\mathbf{p}$, it becomes a genuine probability distribution for the expectation values of $\mathbf{p}$ or $\mathbf{x}$, respectively.

Neumann equation is equivalent to

$$\frac{\partial W_t}{\partial \tau} = \mathcal{H}(\mathbf{x},\mathbf{p})\,\frac{2}{i\hbar}\,\sin\left(\frac{i\hbar}{2}\Big(\overleftarrow{\partial}_{\mathbf{p}}\overrightarrow{\partial}_{\mathbf{x}} - \overleftarrow{\partial}_{\mathbf{x}}\overrightarrow{\partial}_{\mathbf{p}}\Big)\right) W_t(\mathbf{x},\mathbf{p}) \tag{14.61}$$

$$= \underbrace{\{\mathcal{H}, W_t\}}_{\text{Poisson bracket}} + \mathcal{O}(\hbar^2) \tag{14.62}$$

The first line is an exact equation, known as the *Moyal-Groenewold equation*. In the second line, we have performed an expansion in powers of $\hbar$, and one can readily see that the order zero in $\hbar$ is nothing but the classical Liouville equation (it thus describes a system whose time evolution is classical). The first quantum correction to the time evolution arises only at the order $\hbar^2$. Therefore, at the order $\hbar$ (i.e. NLO in the language of quantum field theory), the time evolution of the system remains purely classical. This does not mean that there is no quantum correction of order $\hbar$, but that these corrections can only come from the initial state of the system (in particular, from the fact that a quantum system cannot have well defined $\mathbf{x}$ and $\mathbf{p}$ at the same time, and the Wigner distribution $W_t(\mathbf{x},\mathbf{p})$ must have a width of order $\hbar$ at least). The effect of the operator in $\mathbb{T}_{+k}\mathbb{T}_{-k}$ that acts on the LO in eq. (14.58) is precisely to restore this quantum width of the initial state.

14.5 Multi-point correlation functions at tree level

14.5.1 Generating functional for local measurements

Definition : In the previous section, we have studied a generic observable at leading and next-to-leading orders in λ, and we have established a general functional relationship that relates them. In a sense, this relationship reflects the fact the first $\hbar$ correction in a quantum theory is not fully quantum: at this order only the initial state contains quantum effects, but the time evolution of the system is still classical.

Let us consider now the case of observables that involve multiple points $x_1, \cdots, x_n$, corresponding to n simultaneous measurements. For simplicity, we assume that the points x_i where the measurements are performed lie on the same surface of constant time $x^0 = t_f$, but the final results are valid for any locally space-like surface (this ensures that there is no causal relation between the points x_i, and also that the ordering between the operators in the correlator does not matter). In this case, the leading order is a completely disconnected contribution made of n separate factors, that does not contain any correlation between the n measurements. However, the physically interesting information lies in the correlation between these measurements,

$$\mathcal{C}_{\{1\cdots n\}} \equiv \Big\langle \mathcal{O}(x_1)\cdots\mathcal{O}(x_n)\Big\rangle_c \,, \tag{14.63}$$

where the subscript c indicates that we retain only the connected part of the correlator. From the generic power counting arguments developed in the previous sections, these connected correlators are all of order λ^{-1} in the strong field regime. It is also important to realize that the

connected part of these correlators is subleading compared to their fully disconnected part, since

$$\begin{aligned}\big\langle \mathcal{O}(x_1)\cdots\mathcal{O}(x_n)\big\rangle &= \underbrace{\big\langle \mathcal{O}(x_1)\big\rangle\cdots\big\langle \mathcal{O}(x_n)\big\rangle}_{\lambda^{-n}} \\ &\quad + \underbrace{\sum_{i<j}\big\langle \mathcal{O}(x_i)\mathcal{O}(x_j)\big\rangle_c \prod_{k\neq i,j}\big\langle \mathcal{O}(x_k)\big\rangle}_{\lambda^{1-n}} + \cdots + \underbrace{\big\langle \mathcal{O}(x_1)\cdots\mathcal{O}(x_n)\big\rangle_c}_{\lambda^{-1}}\end{aligned} \tag{14.64}$$

We see in this formula that, in the strong field regime, the fully connected part of a n-point correlator is suppressed by λ^{n-1} compared to the trivial disconnected term. Thus, even at tree level, the correlated part of a n-point function is not a leading order quantity, but arises only at order $n-1$ in the expansion in powers of λ.

One can encapsulate all the correlation functions (14.63) into a generating functional defined as follows[5]:

$$\mathcal{F}[z(\mathbf{x})] \equiv \big\langle 0_{in}\big| \exp\int_{t_f} d^3\mathbf{x}\; z(\mathbf{x})\,\mathcal{O}(\phi(x))\big|0_{in}\big\rangle\;, \tag{14.65}$$

where the argument of the field in $\mathcal{O}$ is $x \equiv (t_f,\mathbf{x})$. From this generating functional, the correlation functions are obtained by differentiating with respect to $z(\mathbf{x}_1),\cdots,z(\mathbf{x}_n)$ and by setting $z\equiv 0$ afterwards. In order to remove the uncorrelated part of the n-point function, we should differentiate the logarithm of $\mathcal{F}$, i.e.

$$\mathcal{C}_{\{1\cdots n\}} = \left.\frac{\delta^n \ln\mathcal{F}}{\delta z(\mathbf{x}_1)\cdots\delta z(\mathbf{x}_n)}\right|_{z\equiv 0} \tag{14.66}$$

The observable $\mathcal{O}(\phi(x))$ is made of the field in the Heisenberg picture, $\phi(x)$, that can be related to the field $\phi_{in}(x)$ of the interaction picture as follows:

$$\phi(x) = U(-\infty,x^0)\,\phi_{in}(x)\,U(x^0,-\infty)\;, \tag{14.67}$$

where $U(t_1,t_2)$ is an evolution operator given in terms of the interactions by the following formula

$$U(t_2,t_1) = T\,\exp i\int_{t_1}^{t_2} dx^0 d^3\mathbf{x}\;\mathcal{L}_{int}(\phi_{in}(x))\;. \tag{14.68}$$

We can therefore rewrite the generating functional solely in terms of the interaction picture field ϕ_{in},

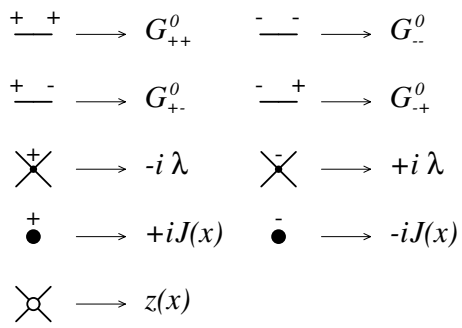
$$\begin{aligned}\mathcal{F}[z(\mathbf{x})] &= \big\langle 0_{in}\big|P\,\exp\int d^3\mathbf{x}\,\Big\{ i\int dx^0\;\mathcal{L}_{int}(\phi_{in}^+(x)) - \mathcal{L}_{int}(\phi_{in}^-(x)) \\ &\qquad + z(\mathbf{x})\,\mathcal{O}(\phi_{in}(t_f,\mathbf{x}))\Big\}\big|0_{in}\big\rangle\;,\end{aligned} \tag{14.69}$$

where P denotes the path ordering on the Schwinger-Keldysh time contour $\mathcal{C}$. We denote by ϕ_{in}^+ the (interaction picture) field that lives on the upper branch and by ϕ_{in}^- the field on the lower branch (the minus sign in front of the term $\mathcal{L}_{int}(\phi_{in}^-(x))$ comes from the fact that the lower branch is oriented from $+\infty$ to $-\infty$). The operator $\mathcal{O}(\phi_{in}(x))$ lives at the final time of this contour, and could either be viewed as made of fields of type $+$ or of type $-$ (the two choices lead to the same results).

[5] This is easily generalized to the case where the initial state is a coherent state instead of the vacuum.

Expression in the Schwinger-Keldysh formalism : Since the initial state is the vacuum, the generating functional defined in eq. (14.69) can be represented diagrammatically as the sum of all the vacuum-to-vacuum graphs (i.e. graphs without external legs) in the Schwinger-Keldysh formalism (see the section 1.14.5), extended by an extra vertex that corresponds to the insertions of the observable $\mathcal{O}$. Let us recall here that the Schwinger-Keldysh diagrammatic rules consist in having two types of interaction vertices ($+$ and $-$ depending on which branch of the contour the vertex lies on, the $-$ vertex being the opposite of the $+$ one) and four types of bare propagators (G^0_{++}, G^0_{--}, G^0_{+-} and G^0_{-+}) depending on the location of the endpoints on the contour. The additional vertex exists only on the final surface, at the time t_f. It is accompanied by a factor $z(\mathbf{x})$, and has as many legs as there are fields in $\mathcal{O}(\phi)$. There is only one kind of this vertex (we can decide to call it $+$ or $-$ without affecting anything). We recapitulate these Feynman rules in the figure 14.5. In the case of the vacuum initial state, we recall that the propagators have the

Figure 14.5: *Diagrammatic rules for the extended Schwinger-Keldysh formalism that gives the generating functional. The Feynman rules shown here for the self-interactions correspond to a $\lambda\phi^4/4!$ interaction term. In this illustration, we have assumed that the observable is quartic in the field when drawing the corresponding vertex (proportional to $z(\mathbf{x})$).*



following explicit expressions:

$$\begin{aligned}
&G^0_{-+}(x,y) = \int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\, e^{-ik\cdot(x-y)}\,, \quad G^0_{+-}(x,y) = \int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\, e^{ik\cdot(x-y)}\,,\\
&G^0_{++}(x,y) = \theta(x^0-y^0)\, G^0_{-+}(x,y) + \theta(y^0-x^0)\, G^0_{+-}(x,y)\,,\\
&G^0_{--}(x,y) = \theta(x^0-y^0)\, G^0_{+-}(x,y) + \theta(y^0-x^0)\, G^0_{-+}(x,y)\,.
\end{aligned} \tag{14.70}$$

Note that when we set $z \equiv 0$, these diagrammatic rules fall back to the pure Schwinger-Keldysh formalism, for which all the connected vacuum-to-vacuum graphs are zero. This implies that

$$\mathcal{F}[z\equiv 0] = 1\,, \tag{14.71}$$

in accordance with the fact that this should be $\langle 0_{\text{in}}|0_{\text{in}}\rangle = 1$.

Retarded-advanced representation : In order to clarify what approximations may be done in the strong field regime, it is useful to use a different basis of fields by introducing

$$\phi_2 \equiv \tfrac{1}{2}\left(\phi_+ + \phi_-\right)\,, \qquad \phi_1 \equiv \phi_+ - \phi_-\,. \tag{14.72}$$

The half-sum ϕ_2 in a sense captures the classical content (plus some quantum corrections), while the difference ϕ_1 is purely quantum (because it represents the different histories of the fields in the amplitude and in the complex conjugated amplitude). To see how the Feynman rules are modified in terms of these new fields, let us start from

$$\phi_\alpha = \sum_{\epsilon=\pm} \Omega_{\alpha\epsilon}\, \phi_\epsilon \qquad (\alpha = 1,2)\,, \tag{14.73}$$

where the matrix Ω reads:

$$\Omega_{\alpha\epsilon} \equiv \begin{pmatrix} 1 & -1 \\ 1/2 & 1/2 \end{pmatrix}\,. \tag{14.74}$$

In terms of this matrix, the new propagators are obtained as follows

$$G^0_{\alpha\beta} \equiv \sum_{\epsilon,\epsilon'=\pm} \Omega_{\alpha\epsilon}\Omega_{\beta\epsilon'} G^0_{\epsilon\epsilon'}\,. \tag{14.75}$$

Explicitly, these propagators read

$$\begin{aligned} G^0_{21} &= G^0_{++} - G^0_{+-}\,, \\ G^0_{12} &= G^0_{++} - G^0_{-+}\,, \\ G^0_{22} &= \tfrac{1}{2}\left[G^0_{+-} + G^0_{-+}\right]\,, \\ G^0_{11} &= 0\,. \end{aligned} \tag{14.76}$$

Note that G^0_{21} is the bare retarded propagator, while G^0_{12} is the bare advanced propagator. The vertices in the new formalism (here written for a quartic interaction) are given by

$$\Lambda_{\alpha\beta\gamma\delta} \equiv -i\lambda\left[\Omega^{-1}_{+\alpha}\Omega^{-1}_{+\beta}\Omega^{-1}_{+\gamma}\Omega^{-1}_{+\delta} - \Omega^{-1}_{-\alpha}\Omega^{-1}_{-\beta}\Omega^{-1}_{-\gamma}\Omega^{-1}_{-\delta}\right]\,, \tag{14.77}$$

where

$$\Omega^{-1}_{\epsilon\alpha} = \begin{pmatrix} 1/2 & 1 \\ -1/2 & 1 \end{pmatrix} \qquad [\Omega_{\alpha\epsilon}\Omega^{-1}_{\epsilon\beta} = \delta_{\alpha\beta}]\,. \tag{14.78}$$

More explicitly, we have :

$$\begin{aligned} &\Lambda_{1111} = \Lambda_{1122} = \Lambda_{2222} = 0 \\ &\Lambda_{1222} = -i\lambda\,, \quad \Lambda_{1112} = -i\lambda/4\,. \end{aligned} \tag{14.79}$$

(The vertices not listed explicitly here are obtained by permutations.) Finally, the rules for an external source in the retarded-advanced basis are :

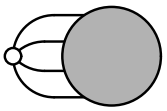
$$J_1 = J\,, \quad J_2 = 0\,. \tag{14.80}$$

Finally, note that the observable depends only on the field ϕ_2, i.e. $\mathcal{O} = \mathcal{O}(\phi_2)$. Indeed, the fields ϕ_+ and ϕ_- represent the field in the amplitude and in the conjugated amplitude. Their difference should vanish when a measurement is performed.

14.5.2 First derivative at tree level

First derivative of $\ln\mathcal{F}$: Differentiating the generating functional with respect to $z(\mathbf{x})$ amounts to exhibiting a vertex $\mathcal{O}$ at the point $\mathbf{x}$ at the final time (as opposed to weighting this vertex by $z(\mathbf{x})$ and integrating over $\mathbf{x}$). Furthermore, by considering the logarithm of the generating functional rather than $\mathcal{F}$ itself, we have only diagrams that are connected to the point $\mathbf{x}$, as shown in this representation:

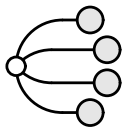
$$\frac{\delta \ln\mathcal{F}}{\delta z(\mathbf{x})} = \quad x \quad , \tag{14.81}$$



where the gray blob is a sum of graphs constructed with the Feynman rules of the figure 14.5, or their analogue in the retarded-advanced formulation. Therefore, these graphs still depend implicitly on z. Note that this blob does not have to be connected.

Tree level expression : Without further specifying the content of the blob, eq. (14.81) is valid to all orders, both in z and in g. At lowest order in g (tree level), a considerable simplification happens because the blob must be a product of disconnected subgraphs, one for each line attached to the vertex $\mathcal{O}(\phi(x))$:

$$\left.\frac{\delta \ln\mathcal{F}}{\delta z(\mathbf{x})}\right|_{\text{tree}} = \quad x \quad , \tag{14.82}$$



where now each of the light colored blob is a *connected tree* 1-point diagram. In the retarded-advanced formalism, there are two of these 1-point functions, that we will denote ϕ_1 and ϕ_2. At tree level, they can be defined recursively by the following pair of coupled integral equations:

$$\begin{aligned}
\phi_1(x) &= i\int_\Omega d^4y\; G^0_{12}(x,y)\,\frac{\partial \mathbf{L}_{\text{int}}(\phi_1,\phi_2)}{\partial\phi_2(y)} \\
&\qquad + \int_{t_f} d^3\mathbf{y}\; G^0_{12}(x,y)\; z(\mathbf{y})\; \mathcal{O}'(\phi_2(y))\;, \\
\phi_2(x) &= i\int_\Omega d^4y\; \left\{ G^0_{21}(x,y)\,\frac{\partial \mathbf{L}_{\text{int}}(\phi_1,\phi_2)}{\partial\phi_1(y)} + G^0_{22}(x,y)\,\frac{\partial \mathbf{L}_{\text{int}}(\phi_1,\phi_2)}{\partial\phi_2(y)} \right\} \\
&\qquad + \int_{t_f} d^3\mathbf{y}\; G^0_{22}(x,y)\; z(\mathbf{y})\; \mathcal{O}'(\phi_2(y))\;.
\end{aligned} \tag{14.83}$$

In these equations, $\mathcal{O}'$ is the derivative of the observable with respect to the field, Ω is the space-time domain comprised between the initial and final times, and we denote

$$\mathbf{L}_{\text{int}}(\phi_1,\phi_2) \equiv \mathcal{L}_{\text{int}}(\phi_2 + \tfrac{1}{2}\phi_1) - \mathcal{L}_{\text{int}}(\phi_2 - \tfrac{1}{2}\phi_1)\;. \tag{14.84}$$

For an interaction Lagrangian $-\frac{\lambda}{4!}\phi^4 + J\phi$, this difference reads

$$\mathbf{L}_{\text{int}}(\phi_1,\phi_2) = -\frac{\lambda}{6}\phi_2^3\phi_1 - \frac{\lambda}{4!}\phi_1^3\phi_2 + J\phi_1\;. \tag{14.85}$$

In terms of these fields, we have

$$\left.\frac{\delta \ln \mathcal{F}}{\delta z(\mathbf{x})}\right|_{\text{tree}} = \mathcal{O}(\phi_2(x)) \,, \tag{14.86}$$

i.e. simply the observable $\mathcal{O}$ evaluated on the field $\phi_2(x)$ (but this field depends on z to all orders, via the boundary terms in eqs. (14.83)).

Classical equations of motion : Using the fact that G^0_{12} and G^0_{21} are Green's functions of $\Box + m^2$, respectively obeying the following identities

$$(\Box_x + m^2)\, G^0_{12}(x,y) = -i\delta(x-y) \,, \quad (\Box_x + m^2)\, G^0_{21}(x,y) = +i\delta(x-y) \,, \tag{14.87}$$

while G^0_{22} vanishes when acted upon by this operator,

$$(\Box_x + m^2)\, G^0_{22}(x,y) = 0 \,, \tag{14.88}$$

we see that ϕ_1 and ϕ_2 obey the following classical field equations of motion:

$$\begin{aligned}(\Box_x + m^2)\, \phi_1(x) &= \frac{\partial \mathbf{L}_{\text{int}}(\phi_1,\phi_2)}{\partial \phi_2(x)} \,, \\ (\Box_x + m^2)\, \phi_2(x) &= \frac{\partial \mathbf{L}_{\text{int}}(\phi_1,\phi_2)}{\partial \phi_1(x)} \,. \end{aligned} \tag{14.89}$$

Note that here the point x is located in the "bulk" Ω; this is why the observable does not enter in these equations of motion. In fact, the observable enters only in the boundary conditions satisfied by these fields on the hypersurface at t_f. For later reference, let us also rewrite these equations of motion in the specific case of a scalar field theory with a $\lambda\phi^4/4!$ interaction term and an external source J:

$$\begin{aligned}&\left[\Box_x + m^2 + \tfrac{\lambda}{2}\phi_2^2\right] \phi_1 + \frac{\lambda}{4!}\,\phi_1^3 = 0 \,, \\ &(\Box_x + m^2)\, \phi_2 + \frac{\lambda}{6}\,\phi_2^3 + \frac{\lambda}{8}\,\phi_1^2\phi_2 = J \,. \end{aligned} \tag{14.90}$$

Boundary conditions : The equations of motion (14.89) are easier to handle than the integral equations (14.83), but they must be supplemented with boundary conditions in order to define uniquely the solutions. The standard procedure for deriving the boundary conditions is to consider the combination $G^0_{12}(x,y)\,(\Box_y + m^2)\,\phi_1(y)$, and let the operator $\Box_y + m^2$ act alternatively on the right and on the left,

$$\begin{aligned}G^0_{12}(x,y)\,(\overrightarrow{\Box}_y + m^2)\,\phi_1(y) &= G^0_{12}(x,y)\,\frac{\partial \mathbf{L}_{\text{int}}(\phi_1,\phi_2)}{\partial \phi_2(y)} \\ G^0_{12}(x,y)\,(\overleftarrow{\Box}_y + m^2)\,\phi_1(y) &= -i\delta(x-y)\,\phi_1(y) \,. \end{aligned} \tag{14.91}$$

By subtracting these equations and integrating over $y \in \Omega$, we obtain

$$\phi_1(x) = i\int_\Omega d^4y\; G^0_{12}(x,y)\,\frac{\partial \mathbf{L}_{\text{int}}(\phi_1,\phi_2)}{\partial \phi_2(y)} - i\int_\Omega d^4y\; G^0_{12}(x,y)\; \overleftrightarrow{\Box}_y\; \phi_1(y) \,. \tag{14.92}$$

The second term of the right hand side is a total derivative thanks to

$$A \overset{\leftrightarrow}{\Box} B = \partial_\mu \left[A \overset{\leftrightarrow}{\partial}{}^\mu B \right] . \tag{14.93}$$

Therefore, this term can be rewritten as a surface integral extended to the boundary of the domain Ω. With reasonable assumptions on the spatial localization of the source $J(x)$ that drives the field, we may disregard the contribution from the boundary at spatial infinity. The remaining boundaries are at the initial time t_i and final time t_f,

$$\phi_1(x) = i \int_\Omega d^4y \; G^0_{12}(x,y) \, \frac{\partial \mathbf{L}_{\rm int}(\phi_1,\phi_2)}{\partial \phi_2(y)} - i \int d^3\mathbf{y} \left[G^0_{12}(x,y) \overset{\leftrightarrow}{\partial}_{y^0} \phi_1(y) \right]^{t_f} . \tag{14.94}$$

Note that the boundary term vanishes at the initial time t_i, because G^0_{12} is the retarded propagator. Likewise, we obtain the following equation for ϕ_2:

$$\begin{aligned} \phi_2(x) \;\; = \;\; & i \int_\Omega d^4y \left\{ G^0_{21}(x,y) \, \frac{\partial \mathbf{L}_{\rm int}(\phi_1,\phi_2)}{\partial \phi_1(y)} + G^0_{22}(x,y) \, \frac{\partial \mathbf{L}_{\rm int}(\phi_1,\phi_2)}{\partial \phi_2(y)} \right\} \\ & - i \int d^3\mathbf{y} \left[G^0_{21}(x,y) \overset{\leftrightarrow}{\partial}_{y^0} \phi_2(y) + G^0_{22}(x,y) \overset{\leftrightarrow}{\partial}_{y^0} \phi_1(y) \right]^{t_f}_{t_i} . \end{aligned} \tag{14.95}$$

The boundary conditions at t_i and t_f are obtained by comparing eqs. (14.83) and (14.94-14.95). At the final time t_f, the boundary condition is

$$\phi_1(t_f,\mathbf{x}) = 0 \quad , \qquad \partial_0 \phi_1(t_f,\mathbf{x}) = i\, z(\mathbf{x})\, \mathcal{O}'(\phi_2(t_f,\mathbf{x})) . \tag{14.96}$$

At the initial time t_i, we must have

$$\int_{y^0=t_i} d^3\mathbf{y} \left[G^0_{21}(x,y) \overset{\leftrightarrow}{\partial}_{y^0} \phi_2(y) + G^0_{22}(x,y) \overset{\leftrightarrow}{\partial}_{y^0} \phi_1(y) \right] = 0 . \tag{14.97}$$

Some simple manipulations lead to the following equivalent form

$$\begin{aligned} & \int_{y^0=t_i} d^3\mathbf{y} \; G^0_{-+}(x,y) \overset{\leftrightarrow}{\partial}_{y^0} \left(\phi_2(y) + \tfrac{1}{2}\phi_1(y) \right) \\ = \quad & \int_{y^0=t_i} d^3\mathbf{y} \; G^0_{+-}(x,y) \overset{\leftrightarrow}{\partial}_{y^0} \left(\phi_2(y) - \tfrac{1}{2}\phi_1(y) \right) = 0 . \end{aligned} \tag{14.98}$$

From the explicit form of the propagators G^0_{+-} and G^0_{-+} (see eqs. (14.70)), we see that, at the initial time, the combination $\phi_2 + \frac{1}{2}\phi_1$ has no positive frequency components, and the combination $\phi_2 - \frac{1}{2}\phi_1$ has no negative frequency components. An equivalent way to state this boundary condition is in terms of the Fourier coefficients of the fields $\phi_{1,2}$. Let us decompose them at the time t_i as follows,

$$\phi_{1,2}(t_i,\mathbf{x}) \equiv \int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}} \left\{ \widetilde{\boldsymbol{\phi}}^{(+)}_{1,2}(\mathbf{k})\, e^{-ik\cdot x} + \widetilde{\boldsymbol{\phi}}^{(-)}_{1,2}(\mathbf{k})\, e^{+ik\cdot x} \right\} . \tag{14.99}$$

In terms of the coefficients introduced in this decomposition, the boundary conditions at the initial time read:

$$\widetilde{\boldsymbol{\phi}}^{(+)}_2(\mathbf{k}) = -\frac{1}{2}\, \widetilde{\boldsymbol{\phi}}^{(+)}_1(\mathbf{k}) , \quad \widetilde{\boldsymbol{\phi}}^{(-)}_2(\mathbf{k}) = \frac{1}{2}\, \widetilde{\boldsymbol{\phi}}^{(-)}_1(\mathbf{k}) . \tag{14.100}$$

14.5.3 Correlations in the quasi-classical regime

quasi-classical approximation : It is in principle possible to solve order by order in the function $x(\mathbf{x})$ the equations of motion (14.89) (or (14.90) for a quartic interaction term) with the boundary conditions (14.96) and (14.100). At order 0 in z, one easily recovers the result of eq. (14.34) for the 1-point function, which states that the expectation value of an observable at leading order is given by the solution Φ of the classical field equation of motion,

$$(\Box_x + m^2)\,\Phi(x) + U'(\Phi(x)) = j(x)\ , \tag{14.101}$$

with a boundary condition at the initial time that depends on the coherent state in which the system is initialized ($\Phi_{\rm ini} \equiv 0$ when the initial state is the vacuum). However, this expansion becomes increasingly cumbersome beyond this simple result. Instead of pursuing this very complicated expansion in powers of z, we present an approximation that allows for an all-orders solution of eqs. (14.89), (14.96) and (14.100). Here, we give only a very sketchy motivation for this approximation, and a lengthier discussion of its validity will be provided later in this section (after we have derived expressions for the fields ϕ_1 and ϕ_2).

Let us first recall that the fields ϕ_+ and ϕ_- represent, respectively, the space-time evolution of the field in amplitudes and in conjugate amplitudes. The fact that they are distinct leads to interferences when squaring amplitudes, a quantum effect controlled by $\hbar$. Consequently, we may expect the difference $\phi_1 \equiv \phi_+ - \phi_-$ to be small compared to $\phi_\pm$ themselves, i.e.

$$\phi_1 \ll \phi_2\ . \tag{14.102}$$

In this situation, that we will call the *quasi-classical approximation*, we can approximate the equations of motion (14.89) by keeping only the lowest order in ϕ_1. This amounts to keeping only the terms linear in ϕ_1 in eq. (14.84) (in the case of a ϕ^4 theory, it means dropping the $\phi_1^3\phi_2$ term in eq. (14.85)). In the approximation, they read

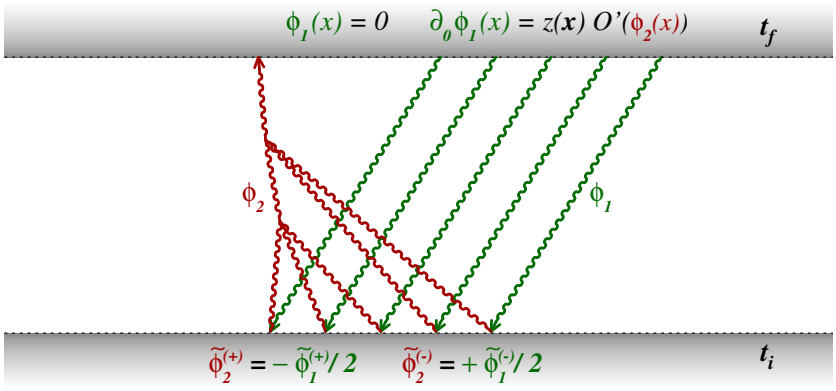
$$\begin{aligned}
&\left[\Box + m^2 - \mathcal{L}''_{\rm int}(\phi_2)\right]\phi_1 = 0\ ,\\
&(\Box + m^2)\,\phi_2 - \mathcal{L}'_{\rm int}(\phi_2) = 0\ ,
\end{aligned} \tag{14.103}$$

while the boundary conditions are still given by (14.96) and (14.98). The problem one must now solve is illustrated in the figure 14.6. The field ϕ_1 obeys a linear equation of motion (dressed by the field ϕ_2, although this aspect is not visible in the figure), with an advanced boundary condition that depends on ϕ_2. In parallel, the field ϕ_2 obeys the classical field equation of motion, with a retarded boundary condition that depends on ϕ_1. As we shall show, this tightly constrained problem admits a formal solution, valid to all orders in the function z, in the form of an implicit functional equation for the first derivative of $\ln \mathcal{F}[z]$.

Formal solution : In order to solve the equation of motion for ϕ_1, let us introduce *mode functions* $a_{\pm k}(x)$, defined as follows

$$\begin{aligned}
&\left[\Box_x + m^2 - \mathcal{L}''_{\rm int}(\phi_2(x))\right] a_{\pm k}(x) = 0\\
&\lim_{x^0 \to t_i} a_{\pm k}(x) = e^{\mp i k\cdot x}\ .
\end{aligned} \tag{14.104}$$

Figure 14.6: *Relationship between the fields ϕ_1 and ϕ_2 in the quasi-classical approximation.*



In other words, they form a basis of the linear space of solutions of the equation obeyed by ϕ_1, and therefore we may express ϕ_1 as a linear superposition of the mode functions. From unitarity, one may prove that the mode functions obey the following identity

$$\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}} \begin{pmatrix} a_{+\mathbf{k}}(x)\dot{a}_{-\mathbf{k}}(y) - a_{-\mathbf{k}}(x)\dot{a}_{+\mathbf{k}}(y) & a_{-\mathbf{k}}(x)a_{+\mathbf{k}}(y) - a_{+\mathbf{k}}(x)a_{-\mathbf{k}}(y) \\ \dot{a}_{+\mathbf{k}}(x)\dot{a}_{-\mathbf{k}}(y) - \dot{a}_{-\mathbf{k}}(x)\dot{a}_{+\mathbf{k}}(y) & \dot{a}_{-\mathbf{k}}(x)a_{+\mathbf{k}}(y) - \dot{a}_{+\mathbf{k}}(x)a_{-\mathbf{k}}(y) \end{pmatrix} \underset{x^0=y^0}{=} i\,\delta(\mathbf{x}-\mathbf{y}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} , \tag{14.105}$$

Thanks to these identities, it is easy to check that the field ϕ_1 that obeys the required equation of motion and boundary conditions is given by

$$\phi_1(x) = \int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\, d^3\mathbf{u} \left\{ a_{-\mathbf{k}}(x)a_{+\mathbf{k}}(t_f,\mathbf{u}) - a_{+\mathbf{k}}(x)a_{-\mathbf{k}}(t_f,\mathbf{u}) \right\} z(\mathbf{u})\; \mathcal{O}'(\phi_2(t_f,\mathbf{u})) \,. \tag{14.106}$$

The above equation formally defines $\phi_1(x)$ in the bulk, $x \in \Omega$, in terms of the field ϕ_2 at the final time. Besides the explicit factor $z(\mathbf{u})$, the right hand side contains also an implicit z dependence (to all orders in z) in the field $\phi_2(t_f,\mathbf{u})$ and in the mode functions $a_{\pm\mathbf{k}}$ (since they evolve on top of the background ϕ_2).

Then, using the boundary condition at the initial time, we obtain the following expression for the field ϕ_2 at t_i,

$$\begin{aligned} \phi_2(t_i,\mathbf{y}) \;=\; & \frac{1}{2}\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}} \int d^3\mathbf{u} \left\{ e^{+ik\cdot y}\, a_{+\mathbf{k}}(t_f,\mathbf{u}) \right. \\ & \left. + e^{-ik\cdot y}\, a_{-\mathbf{k}}(t_f,\mathbf{u}) \right\} z(\mathbf{u})\; \mathcal{O}'(\phi_2(t_f,\mathbf{u})) \,. \end{aligned} \tag{14.107}$$

This can be expressed in a more convenient way with eq. (14.51). In terms of the operators $\mathbb{T}_{\pm\mathbf{k}}$, we may rewrite ϕ_2 at the initial time as follows:

$$\phi_2(t_i,\mathbf{y}) = \frac{1}{2}\int \frac{d^3\mathbf{k}}{(2\pi)^3 2E_\mathbf{k}} \int d^3\mathbf{u}\; z(\mathbf{u})\; \mathcal{O}(\phi_2(t_f,\mathbf{u}))\left\{ \overleftarrow{\mathbb{T}}_{+\mathbf{k}}\; e^{+ik\cdot y} + \overleftarrow{\mathbb{T}}_{-\mathbf{k}}\; e^{-ik\cdot y}\right\}, \tag{14.108}$$

where the arrows indicate on which side the $\mathbb{T}_{\pm\mathbf{k}}$ operators act. This expression gives the initial condition for the first of eqs. (14.103), in the form of a linear superposition of plane waves $\exp(\pm ik\cdot y)$. The next step is to note that the field $\phi_2(x)$ that satisfies this equation of motion, and has the initial condition $\phi_2(t_i,\mathbf{y})$ is formally given by

$$\phi_2(x) = \exp\left\{\int d^3\mathbf{y}\,\left[\phi_2(t_i,\mathbf{y})\frac{\delta}{\delta\Phi_{ini}(t_i,\mathbf{y})} + (\partial^0\phi_2(t_i,\mathbf{y}))\frac{\delta}{\delta(\partial^0\Phi_{ini}(t_i,\mathbf{y}))}\right]\right\}\Phi(x)\,. \tag{14.109}$$

This formula follows from the fact that the derivative with respect to the initial field is the generator for shifts of the initial condition of Φ; its exponential is therefore the corresponding translation operator. The same formula applies also to any function of the field, e.g. $\mathcal{O}(\phi_2)$. Substituting $\phi_2(t_i,\mathbf{y})$ by eq. (14.108) inside the exponential, this leads to

$$\begin{aligned}\mathcal{O}\big(\phi_2(x)\big) &= \exp\left\{\frac{1}{2}\int\frac{d^3\mathbf{k}}{(2\pi)^3 2E_\mathbf{k}}\int d^3\mathbf{u}\; z(\mathbf{u})\,\mathcal{O}(\phi_2(t_f,\mathbf{u}))\right.\\ &\qquad\left.\times\left[\overleftarrow{\mathbb{T}}_{+\mathbf{k}}\overrightarrow{\mathbb{T}}_{-\mathbf{k}} + \overleftarrow{\mathbb{T}}_{-\mathbf{k}}\overrightarrow{\mathbb{T}}_{+\mathbf{k}}\right]\right\}\mathcal{O}\big(\Phi(x)\big)\Big|_{\Phi_{ini}\equiv 0}\\ &= \exp\left\{\int d^3\mathbf{u}\; z(\mathbf{u})\,\mathcal{O}(\phi_2(t_f,\mathbf{u}))\;\otimes\right\}\mathcal{O}\big(\Phi(x)\big)\Big|_{\Phi_{ini}\equiv 0}\,,\end{aligned} \tag{14.110}$$

where the $\otimes$ operation is defined by

$$A\otimes B \equiv \frac{1}{2}\int\frac{d^3\mathbf{k}}{(2\pi)^3 2E_\mathbf{k}}\; A\left[\overleftarrow{\mathbb{T}}_{+\mathbf{k}}\overrightarrow{\mathbb{T}}_{-\mathbf{k}} + \overleftarrow{\mathbb{T}}_{-\mathbf{k}}\overrightarrow{\mathbb{T}}_{+\mathbf{k}}\right] B\,. \tag{14.111}$$

Setting $x^0 = t_f$ and denoting

$$\mathcal{D}[\mathbf{x}_1;z] \equiv \mathcal{O}(\phi_2(t_f,\mathbf{x}_1)) \tag{14.112}$$

the first derivative of $\ln\mathcal{F}$, we see that it obeys the following recursive formula

$$\mathcal{D}[\mathbf{x}_1;z] = \exp\left\{\int d^3\mathbf{u}\; z(\mathbf{u})\,\mathcal{D}[\mathbf{u};z]\;\otimes\right\}\mathcal{O}\big(\Phi(t_f,\mathbf{x}_1)\big)\Big|_{\Phi_{ini}\equiv 0}\,. \tag{14.113}$$

Realization of the quasi-classical approximation : Let us now return on the condition $\phi_1 \ll \phi_2$, that was used in the derivation of eq. (14.113), in order to see a posteriori when it is satisfied. To that effect, we can use eq. (14.106) for ϕ_1. For ϕ_2, the initial condition at t_i is given by

eq. (14.108). For the sake of this discussion, it is sufficient to use a linearized solution for ϕ_2 in the bulk, that reads

$$\phi_2(x)\Big|_{\text{lin}} = \frac{1}{2}\int\frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\int d^3\mathbf{u}\,\Big\{a_{-\mathbf{k}}(x)a_{+\mathbf{k}}(t_f,\mathbf{u})+a_{+\mathbf{k}}(x)a_{-\mathbf{k}}(t_f,\mathbf{u})\Big\} \times z(\mathbf{u})\,\mathcal{O}'(\phi_2(t_f,\mathbf{u}))\,, \tag{14.114}$$

First of all, a comparison between eqs. (14.106) and (14.114) indicates that ϕ_1 and ϕ_2 have the same order in the coupling constant g, since they are made of the same building blocks (the only difference is the sign between the two terms of the integrand, and an irrelevant overall factor $\frac{1}{2}$).

However, a hierarchy between ϕ_1 and ϕ_2 arises dynamically when the classical solutions of the field equation of motion (14.101) are unstable. Such instabilities are fairly generic in several quantum field theories; in particular the scalar field theory with a ϕ^4 coupling that we are using as example is known to have a parametric resonance. Since the mode functions $a_{\pm\mathbf{k}}$ are linearized perturbations on top of the classical field ϕ_2, an instability of the classical solution ϕ_2 is equivalent to the fact that some of the mode functions grow exponentially with time, as $\exp(\mu(x^0-t_i))$ (where μ is the Lyapunov exponent). Thus, since eq. (14.114) is bilinear in the mode functions, we expect that

$$\phi_2(x)\Big|_{\text{lin}} \sim e^{\mu(x^0+t_f-2t_i)}\,. \tag{14.115}$$

Estimating the magnitude of ϕ_1 requires more care. Indeed, from eqs. (14.105), antisymmetric combinations of the mode functions at equal times remain of order 1 even if individual mode functions grow exponentially with time. Thus, at the final time, we have

$$\phi_1(t_f,\mathbf{x})\sim 1 \quad\text{and}\quad \frac{\phi_2(t_f,\mathbf{x})}{\phi_1(t_f,\mathbf{x})}\sim e^{2\mu(t_f-t_i)}\gg 1\,, \tag{14.116}$$

for sufficiently large t_f-t_i.

In order to estimate the ratio ϕ_2/ϕ_1 at intermediate times, one may use the following reasoning. The antisymmetric combination of mode functions that enters in eq. (14.106) is the advanced propagator G_A in the background ϕ_2. This advanced propagator may also be expressed in terms of a different set of mode functions $b_{\pm\mathbf{k}}$ defined to be plane waves at the final time t_f,

$$\begin{aligned}&\Big[\Box_x+m^2-\mathcal{L}''_{\text{int}}(\phi_2(x))\Big]\,b_{\pm\mathbf{k}}(x)=0\\ &\lim_{x^0\to t_f} b_{\pm\mathbf{k}}(x)=e^{\mp ik\cdot x}\,.\end{aligned} \tag{14.117}$$

In terms of these alternate mode functions, we also have

$$\phi_1(x)=\int\frac{d^3\mathbf{k}}{(2\pi)^3 2E_{\mathbf{k}}}\int d^3\mathbf{u}\,\Big\{b_{-\mathbf{k}}(x)b_{+\mathbf{k}}(t_f,\mathbf{u})-b_{+\mathbf{k}}(x)b_{-\mathbf{k}}(t_f,\mathbf{u})\Big\}\,z(\mathbf{u})\,\mathcal{O}'(\phi_2(t_f,\mathbf{u}))\,, \tag{14.118}$$

In the presence of instabilities, these backward evolving mode functions grow when x^0 decreases away from t_f, as $\exp(\mu(t_f-x^0))$ (in this sketchy argument, the Lyapunov exponent μ is assumed here to be the same for the forward and backward mode functions). This implies

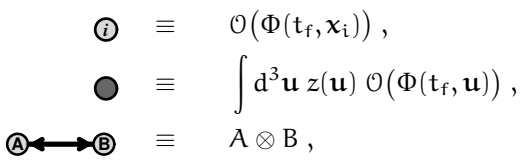
$$\phi_1(x)\sim e^{\mu(t_f-x^0)}\,, \tag{14.119}$$

and the following magnitude for the ratio ϕ_2/ϕ_1 at intermediate times

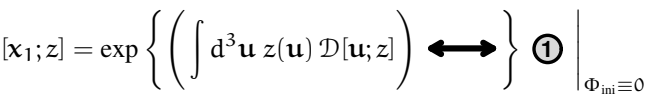
$$\frac{\phi_2(x)}{\phi_1(x)} \sim \frac{e^{\mu(x^0+t_f-2t_i)}}{e^{\mu(t_f-x^0)}} \sim e^{2\mu(x^0-t_i)} \,. \tag{14.120}$$

Thus, with instabilities and non-zero Lyapunov exponents, the quasi-classical approximation is generically satisfied thanks to the exponential growth of perturbations over the background.

Expansion of eq. (14.113) in powers of z : Although eq. (14.113) cannot be solved explicitly, it is fairly easy to obtain a diagrammatic representation of its solution. For this, let us introduce the following graphical notations:



in terms of which the functional equation obeyed by $\mathcal{D}[\mathbf{x}_1;z]$ reads:



At the order 0 in z, we just need to set $z \equiv 0$ inside the exponential, to obtain

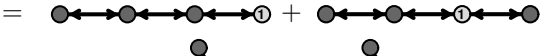
$$\mathcal{D}^{(0)}[\mathbf{x}_1;z] = \textcircled{1} \,. \tag{14.121}$$

Then, we proceed recursively. We insert the 0-th order result in the exponential, and expand to order 1 in z, leading to the following result at order 1:

$$\mathcal{D}^{(1)}[\mathbf{x}_1;z] = \tag{14.122}$$

The next two iterations give:

$$\mathcal{D}^{(2)}[\mathbf{x}_1;z] = + \frac{1}{2!} \,, \tag{14.123}$$

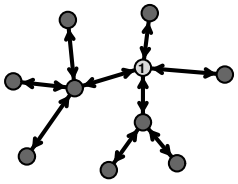


and

$$\mathcal{D}^{(3)}[\mathbf{x}_1;z] = + + \frac{1}{2!} + \frac{1}{3!} \,. \tag{14.124}$$

These examples generalize to all orders in z: the functional $\mathcal{D}[\mathbf{x}_1;z]$ can be represented as the sum of all the rooted trees (the root being the node carrying the fixed point $\mathbf{x}_1$) weighted by the corresponding symmetry factor $1/\mathcal{S}(T)$:

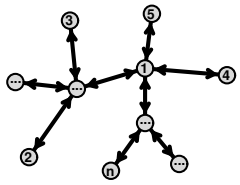
$$\frac{\delta \ln \mathcal{F}[z]}{\delta z(z_1)} = \mathcal{D}[\mathbf{x}_1;z] = \sum_{\substack{\text{rooted}\\ \text{trees } T}} \frac{1}{\mathcal{S}(T)} \tag{14.125}$$



.

Correlation functions : The n-point correlation function is obtained by differentiating this expression $n-1$ times, with respect to $z(\mathbf{x}_2), \cdots, z(\mathbf{x}_n)$, and by setting $z \equiv 0$ afterwards. This selects all the trees with n distinct labeled nodes[6] (including the node at $\mathbf{x}_1$). Moreover, since derivatives commute, these successive differentiations eliminate the symmetry factors, leading to

$$\left.\frac{\delta \ln \mathcal{F}[z]}{\delta z(\mathbf{z}_1)\cdots\delta z(\mathbf{x}_n)}\right|_{z\equiv 0} = \mathcal{C}_{\{1\cdots n\}} = \sum_{\substack{\text{trees with } n \\ \text{labeled nodes}}}$$



. (14.126)

The number of trees contributing to this sum is equal to n^{n-2} (*Cayley's formula*). Eq. (14.126) tells us that, at tree level in the quasi-classical regime, all the n-point correlation functions are entirely determined by the functional dependence of the solution of the classical field equation of motion with respect to its initial condition. Moreover, this formula provides a way to construct explicitly the correlation functions in terms of functional derivatives with respect to the initial field.

In the quasi-classical approximation, the final state correlations are entirely due to quantum fluctuations in the initial state, that are encoded in the function $G^0_{22}(x,y)$. If the initial state is the vacuum, it reads

$$G^0_{22}(x,y) = \int \frac{d^3\mathbf{k}}{(2\pi)^2 2E_{\mathbf{k}}}\, e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \,. \tag{14.127}$$

The support of this function is dominated by distances $|\mathbf{x}-\mathbf{y}|$ smaller than the Compton wavelength m^{-1}. Thus, in the tree representation of eq. (14.126), a link between the points $\mathbf{x}_i$ and $\mathbf{x}_j$ is nonzero provided that the past light-cones of summits x_i and x_j overlap at the initial time (or at least approach each other within distances $\lesssim m^{-1}$), as illustrated in the figure 14.7. A more thorough analysis would indicate that eq. (14.126) is exact at tree level for the 1-point and 2-point correlations, but is incomplete (even at tree-level) beyond 2 points. The corrections to this formula are nevertheless suppressed if the condition $\phi_1 \ll \phi_2$ holds.

[6] Thus, permuting nodes in general yields a different tree.

Figure 14.7: *Causal structure of the 3-point correlation function in the quasi-classical regime.*