Physics 253b Lecture Notes, spring 2013.

Contents

1 Poles and unitarity

1.1 What are quantum fields?

In this course, we will mostly be studying quantum field theories in $3 + 1$ spacetime dimensions. Our convention for the Lorentzian signature is $(-, +, +, +)$. Let $\mathcal{H}$ be the Hilbert space of all quantum states in the theory. Let $|0\rangle$ be the ground state, or vacuum. The Hamiltonian is assumed to be translation and Lorentz invariant, and so is the vacuum state. Excited states, of nonzero energy, should have a particle interpretation, either as a single particle or multi-particle states. These particles may or may not be excitations of an elementary particle field in the Lagrangian formalism. A special class of states are single particle states. The energy and momentum of one particle states obey a certain dispersion relation, which we now characterize.

Unless otherwise noted, we will work in the Heisenberg picture. Consider a complex scalar operator $\Phi(x)$, which may or may not be an elementary field. In a weakly coupled theory of complex scalar particles, for instance, one may take $\Phi(x)$ to be the elementary scalar field, but this need not be the case.

The vacuum expectation value of $\Phi(x)\Phi^\dagger(y)$, where x and y are two different points in spacetime, may be expressed as

$$\langle 0|\Phi(x)\Phi^\dagger(y)|0\rangle = \sum_n \langle 0|\Phi(x)|n\rangle\langle n|\Phi^\dagger(y)|0\rangle, \tag{1.1}$$

where $|n\rangle$ is a complete set of orthonormal basis of the Hilbert space $\mathcal{H}$. n generally denotes both discrete and continuous labels (in the latter case the sum will be replaced by the appropriate integral). We may choose the state $|n\rangle$ to have a definite four-momentum p_n^μ, and write[1]

$$\begin{aligned} \langle 0|\Phi(x)|n\rangle &= e^{ip_n\cdot x}\langle 0|\Phi(0)|n\rangle, \\ \langle n|\Phi^\dagger(y)|0\rangle &= e^{-ip_n\cdot y}\langle n|\Phi^\dagger(0)|0\rangle, \end{aligned} \tag{1.2}$$

[1] Here we used the translational symmetry which implies $\Phi(x) = e^{-iP\cdot x}\Phi(0)e^{iP\cdot x}$, where P_μ is the four-momentum operator.

and so

$$\begin{aligned}\langle 0|\Phi(x)\Phi^\dagger(y)|0\rangle &= \sum_n e^{ip_n\cdot(x-y)}|\langle 0|\Phi(0)|n\rangle|^2 \\ &= \int d^4p\, e^{ip\cdot(x-y)}\sum_n \delta^4(p-p_n)|\langle 0|\Phi(0)|n\rangle|^2.\end{aligned} \tag{1.3}$$

Since Φ is a scalar operator, it follows from Lorentz invariance[2] that the last sum is a scalar function of p^μ. It can therefore be expressed as a function of $p^2(\leq 0)$, multiplied by $\theta(p^0)$ as we require all states to have non-negative energy (θ is the step function that takes value 1 for positive argument and 0 for negative argument). That is,

$$\sum_n \delta^4(p-p_n)|\langle 0|\Phi(0)|n\rangle|^2 = \frac{\theta(p^0)}{(2\pi)^3}\rho(-p^2), \tag{1.4}$$

where ρ is a so-called spectral function, with the property $\rho(-p^2) = 0$ for all $p^2 > 0$ (as space-like four-momenta are forbidden on physical states). Clearly, $\rho(-p^2)$ is real and positive for $p^2 \leq 0$.

We can trivially rewrite the above two-point function as

$$\langle 0|\Phi(x)\Phi^\dagger(y)|0\rangle = \int_0^\infty d\mu^2 \rho(\mu^2)\Delta_+(x-y;\mu^2), \tag{1.5}$$

where

$$\Delta_+(x-y;\mu^2) = \frac{1}{(2\pi)^3}\int d^4p\, e^{ip\cdot(x-y)}\theta(p^0)\delta(p^2+\mu^2). \tag{1.6}$$

Similarly, with $\Phi(x)$ and $\Phi^\dagger(y)$ exchanged, we can write an analogous formula

$$\langle 0|\Phi^\dagger(y)\Phi(x)|0\rangle = \int_0^\infty d\mu^2 \bar\rho(\mu^2)\Delta_+(y-x;\mu^2), \tag{1.7}$$

where $\bar\rho(\mu^2)$ is defined by

$$\sum_n \delta^4(p-p_n)|\langle 0|\Phi^\dagger(0)|n\rangle|^2 = \frac{\theta(p^0)}{(2\pi)^3}\bar\rho(-p^2). \tag{1.8}$$

[2] Under a Lorentz transformation $x^\mu \to \Lambda^\mu{}_\nu x^\nu$, the field $\Phi(0)$ transforms into $U(\Lambda)^{-1}\Phi(0)U(\Lambda)$, where $U(\Lambda)$ is the unitary operator acting on the Hilbert space representation Lorentz transformation Λ. $U(\Lambda)$ takes $|n\rangle$ of four-momentum p_n^μ to another state $|\Lambda(n)\rangle$ of momentum $\Lambda^\mu{}_\nu p_n^\nu$. If you have studied Dirac equation and the Lorentz transformation of spinor wave functions in the context of relativistic quantum mechanics (which is confusing and logically incomplete), you might be confused by how there the Lorentz transformation wasn't unitary. In relativistic QFT, a Lorentz transformation is represented by a unitary operator acting on the Hilbert space. The field operators themselves, however, may not transform in a unitary representation under Lorentz boosts.

Let us examine the vacuum expectation value of the commutator of $\Phi(x)$ with $\Phi^\dagger(y)$,

$$\langle 0|[\Phi(x), \Phi^\dagger(y)]|0\rangle = \int_0^\infty d\mu^2 \left[\rho(\mu^2)\Delta_+(x-y;\mu^2) - \bar{\rho}(\mu^2)\Delta_+(y-x;\mu^2)\right]. \quad (1.9)$$

When x and y are space-like separated, i.e. $(x-y)^2 > 0$, the LHS is zero by the assumption of causality. Since $\Delta_+(x;\mu^2)$ is Lorentz invariant, for space-like x we may set $x^0 = 0$ by a Lorentz transformation, and then it is clear that $\Delta_+(x;\mu^2)$ is an even function of x when x is space-like. In fact, for space-like x, $\Delta_+(x;\mu^2)$ can be written as

$$\Delta_+(x;\mu^2) = \frac{\mu}{4\pi^2|x|}K_1(\mu|x|), \quad (1.10)$$

where $|x| = \sqrt{x^2}$, and K_1 is a Bessel function of the second kind. $\Delta_+(x;\mu^2)$ being a nontrivial and even function for spacelike x implies that, by causality, $\rho(\mu^2) = \bar{\rho}(\mu^2)$. This is a part of the CPT theorem.

The propagator for Φ, on the other hand, is the vacuum expectation value of the time ordered product of $\Phi(x)$ with $\Phi^\dagger(y)$. By the above manipulation, it is easy to see that

$$\langle 0|T\left[\Phi(x)\Phi^\dagger(y)\right]|0\rangle = -i\int_0^\infty d\mu^2 \rho(\mu^2)\Delta_F(x-y;\mu^2), \quad (1.11)$$

where $\Delta_F(x;\mu^2)$, given by

$$-i\Delta_F(x;\mu^2) = \theta(x^0)\Delta_+(x;\mu^2) + \theta(-x^0)\Delta_+(-x;\mu^2), \quad (1.12)$$

is identical to the Feynman propagator for a *free* scalar particle of mass μ. It has the momentum space representation

$$\int d^4x e^{-ip\cdot x}\Delta_F(x;\mu^2) = \frac{1}{p^2+\mu^2-i\epsilon}. \quad (1.13)$$

So, we can write

$$\begin{aligned}\Delta'(p) &\equiv i\int d^4x e^{-ip\cdot(x-y)}\langle 0|T\left[\Phi(x)\Phi^\dagger(y)\right]|0\rangle \\ &= \int_0^\infty d\mu^2 \frac{\rho(\mu^2)}{p^2+\mu^2-i\epsilon}.\end{aligned} \quad (1.14)$$

This seemingly simple result has some nontrivial implications. For instance, the positivity of $\rho(\mu^2)$ implies that $\Delta'(p)$ cannot vanish faster than $1/p^2$ at large momentum (i.e. large $|p^2|$). Naively, one may attempt to construct super-renormalizable interacting field theories by introducing say a four-derivative kinetic term, so that the propagator goes like $1/p^4$ at large momenta, and many UV divergences would be cured. But we see here that this is not possible in a unitary and Lorentz invariant theory; in fact, unitarity would be violated in such theories with higher derivative kinetic term.

In an interacting quantum field theory, $\Phi(0)|0\rangle$ could be a complicated state, generally containing multi-particle excitations. Let us now assume that there exists a one-particle state of momentum k^μ and mass m ($k^2 = -m^2$), $|\vec{k}\rangle$, which has nonzero overlap with $\Phi(0)|0\rangle$. The one-particle states are normalized so that

$$\langle\vec{k}|\vec{k}'\rangle = \delta^3(\vec{k} - \vec{k}'). \tag{1.15}$$

This normalization is not invariant under Lorentz boosts. For instance, the projection operator onto one-particle states can be written

$$\int d^3\vec{k}|\vec{k}\rangle\langle\vec{k}| = \int d^4k(2k^0)\delta(k^2 + m^2)|k\rangle\langle k| \tag{1.16}$$

It then follows from Lorentz invariance that

$$\langle 0|\Phi(0)|\vec{k}\rangle = N\left[\frac{1}{(2\pi)^3}\frac{1}{2\sqrt{\vec{k}^2 + m^2}}\right]^{\frac{1}{2}}, \tag{1.17}$$

where N is a constant. For simplicity, let us for now suppose that all other excited states are multi-particle states. We then find

$$\rho(\mu^2) = Z\delta(\mu^2 - m^2) + \sigma(\mu^2), \tag{1.18}$$

where $Z = |N|^2$, $Z\delta(\mu^2 - m^2)$ is the contribution from one-particle intermediate states $|n\rangle$, and $\sigma(\mu^2) \geq 0$ is the contribution from multi-particle states. The propagator $\Delta'(p)$ has a pole at $p^2 = -m^2$, with residue Z.

If the scalar field operator $\Phi(x)$ is canonically normalized, so that $\partial_t\Phi(\vec{x}, t)$ is the canonical momentum conjugate to $\Phi^\dagger$, then we have the equal time commutator

$$\left[\Phi^\dagger(\vec{x}, t), \frac{\partial\Phi(\vec{y}, t)}{\partial t}\right] = i\delta^3(\vec{x} - \vec{y}). \tag{1.19}$$

Plugging this into the vacuum two point function, and using

$$\left.\frac{\partial}{\partial x^0}\Delta_+(x; \mu^2)\right|_{x^0=0} = -\frac{i}{2}\delta^3(\vec{x}), \tag{1.20}$$

we find

$$\int_0^\infty \rho(\mu^2)d\mu^2 = 1. \tag{1.21}$$

It then follows from the presentation (1.18) that

$$1 = Z + \int_0^\infty \sigma(\mu^2)d\mu^2. \tag{1.22}$$

We see that the Z factor, which plays the role of field ("wave function") renormalization constant, is less than or equal to 1. It is equal to 1 only if Φ is a free field, so that $\Phi(x)|0\rangle$ contains no multi-particle states. When $Z < 1$, $\Phi(x)|0\rangle$ is a mixture of single particle and multi-particle states.

Our discussion here can be generalized straightforwardly to spinor and vector fields. Last semester, you have seen that the wave function renormalization of the photon field in QED at one-loop is given by

$$Z_3 = 1 - \frac{\alpha}{3\pi} \ln \frac{\Lambda^2}{\mu^2}, \tag{1.23}$$

where μ is an IR cutoff at the order of the electron mass, and $\Lambda \gg \mu$ is a UV cutoff, while the electron wave function renormalization is given at one loop by

$$Z_2 = 1 - \frac{\alpha}{4\pi} \ln \frac{\Lambda^2}{\mu'^2}. \tag{1.24}$$

In both cases, we see clearly that $Z < 1$.

1.2 The S-matrix

The most important physical observable in a quantum field theory is the S-matrix. The S-matrix is defined as the overlap between "in" and "out" asymptotic states, in the Heisenberg picture. Formally, the S-matrix element of ℓ outgoing particles and $n - \ell$ incoming particles is

$$S(\vec{p}_1, \cdots, \vec{p}_\ell; \vec{p}_{\ell+1}, \cdots, \vec{p}_n) = {}^{out}\langle \vec{p}_1, \cdots, \vec{p}_\ell | \vec{p}_{\ell+1}, \cdots, \vec{p}_n \rangle^{in}. \tag{1.25}$$

An "in" state, in the Schrödinger picture, is a state that looks like far separated wave packets of non-interacting particles in the far past. An "out" state in the Schrödinger picture is one that looks like far separated wave packets of free particle in the far future. This description is valid provided that the particles are weakly interacting at long distances. It is our assumption that all states in the Hilbert space can be represented as linear combinations of in states, or alternatively, as linear combinations of out states. The S-matrix is the linear transformation that relates these two bases of the Hilbert space. In talking about S-matrix elements, we typically take the limit where the wave packet of each particle in the in and out states are sharply peaked at some momentum, and therefore becomes a plane wave in position space.

A fundamental result in quantum field theory is the LSZ reduction formula, which allows us to express the S-matrix elements to the residues of time-ordered vacuum correlation functions of field operators, when the momenta associated with these field operators are put on the mass-shell of single particle states.

Let us begin write the momentum space correlation function

$$G(p_1, p_2, \cdots, p_n) = \int d^4x_1 \cdots d^4x_n e^{-i\sum p_i\cdot x_i}\langle 0|T\left[\Phi(x_1)\cdots\Phi(x_n)\right]|0\rangle, \tag{1.26}$$

where the momenta p_i are generally off shell. Now group $p_1, \cdots, p_\ell$ into the set of outgoing momenta, with $p_i^0 > 0$, and group $p_{\ell+1}, \cdots, p_n$ into incoming momenta, with $p_i^0 < 0$.

We are going to be interested in the poles of the correlation function $G(p_1, \cdots, p_n)$. The poles in p_i^0 cannot come from the integration of $e^{-ip_i\cdot x_i}$ times a function of x_i over a finite range of x_i. Rather, they come from integration over an infinite range of x_i^0, either in the far past or in the far future. For this purpose, we shall split off the contribution from where $\Phi(x_1)$ is inserted at a later time than all other field operators, and write

$$G(p_1, \cdots, p_n) = \int d^4x_1 \cdots d^4x_n \theta(x_1^0 - \max\{x_2^0, \cdots, x_n^0\}) e^{-i\sum p_i\cdot x_i}\langle 0|\Phi(x_1)T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle + \cdots\cdots, \tag{1.27}$$

where the omitted terms involve the vacuum expectation value of the string of operators with $\Phi(x_1)$ inserted to the right of one of the $\Phi(x_i)$'s, $i = 2, \cdots, n$. Now, we make use of the Fourier representation of the step function,

$$\theta(t) = -\frac{1}{2\pi i}\int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega t}}{\omega + i\epsilon}, \tag{1.28}$$

and write

$$\begin{aligned} &G(p_1, \cdots, p_n) \\ &= \int d^4x_1 \cdots d^4x_n e^{-i\sum p_i\cdot x_i}\frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega(x_1^0 - \max\{x_2^0, \cdots, x_n^0\})}}{\omega + i\epsilon}\langle 0|\Phi(x_1)T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle + \cdots, \end{aligned} \tag{1.29}$$

Now we insert a complete basis, $\sum_n |n\rangle\langle n|$, to the right of $\Phi(x_1)$, and focus on the one-particle states that is to be annihilated by $\Phi(x_1)$,

$$
\begin{aligned}
&G(p_1, \cdots, p_n) \\
&= \int d^4x_1 \cdots d^4x_n e^{-i\sum p_i\cdot x_i} \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega(x_1^0-\max\{x_2^0,\cdots,x_n^0\})}}{\omega+i\epsilon} \int d^3\vec{k}\langle 0|\Phi(x_1)|\vec{k}\rangle\langle\vec{k}|T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle \\
&\quad + (\text{multi}-\text{particle states annihilated by } \Phi(x_1)) + (\text{other ordering}), \\
&= \int d^4x_1 \cdots d^4x_n e^{-i\sum p_i\cdot x_i} \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega(x_1^0-\max\{x_2^0,\cdots,x_n^0\})}}{\omega+i\epsilon} \\
&\quad \times \int d^3\vec{k} e^{ik\cdot x_1} \frac{N}{(2\pi)^{\frac{3}{2}}\sqrt{2k^0}}\langle\vec{k}|T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle \\
&\quad + (\text{multi}-\text{particle states annihilated by } \Phi(x_1)) + (\text{other ordering}),
\end{aligned}
\tag{1.30}
$$

where inside the $\vec{k}$-integral, k^μ is understood to be the on-shell four-momentum, with $k^0 = \omega_{\vec{k}} \equiv \sqrt{\vec{k}^2+m^2}$. Integrating over x_1, and then integrating out $\vec{k}$ using the spatial momentum delta function, we have

$$
\begin{aligned}
G(p_1, \cdots, p_n) &= \int d^4x_2 \cdots d^4x_n e^{-i\sum_{i=2}^n p_i\cdot x_i} \frac{i}{2\pi}\int_{-\infty}^{\infty} d\omega \frac{e^{-i\omega(-\max\{x_2^0,\cdots,x_n^0\})}}{\omega+i\epsilon}(2\pi)^4\delta(\omega+\omega_{\vec{p}_1}-p_1^0) \\
&\quad \times \frac{N}{(2\pi)^{\frac{3}{2}}\sqrt{2\omega_{\vec{p}_1}}}\langle\vec{p}_1|T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle + \cdots \\
&= \int d^4x_2 \cdots d^4x_n e^{-i\sum_{i=2}^n p_i\cdot x_i} \frac{e^{i(p_1^0-\omega_{\vec{p}_1})\max\{x_2^0,\cdots,x_n^0\}}}{p_1^0-\omega_{\vec{p}_1}+i\epsilon}\frac{iN(2\pi)^{\frac{3}{2}}}{\sqrt{2\omega_{\vec{p}_1}}}\langle\vec{p}_1|T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle + \cdots,
\end{aligned}
\tag{1.31}
$$

The result has a pole at $p_1^0 = \omega_{\vec{p}_1} = \sqrt{\vec{p}_1^2+m^2}$, as expected when the external particle momentum p_1 goes on shell. The omitted terms either involve multi-particle states annihilated by $\Phi(x_1)$ (which do not contribute to the single particle pole), or states annihilated by $\Phi(x_1)$ together with other field operators inserted at later times (which only have poles when the sum of p_1 together with other external momenta goes on shell), generically do not have poles at $p_1^0 = \omega_{\vec{p}_1}$. In the limit $p_1^0 \to \omega_{\vec{p}_1}$, therefore,

$$
G(p_1, \cdots, p_n) \to \frac{-iN(2\pi)^{\frac{3}{2}}\sqrt{2\omega_{\vec{p}_1}}}{p_1^2+m^2-i\epsilon}\int d^4x_2 \cdots d^4x_n e^{-i\sum_{i=2}^n p_i\cdot x_i}\langle\vec{p}_1|T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle.
\tag{1.32}
$$

Similarly, the integration over x_1^0 in the far past gives rise to a pole at $p_1^0 \to -\omega_{\vec{p}_1}$,

$$
G(p_1, \cdots, p_n) \to \frac{-iN(2\pi)^{\frac{3}{2}}\sqrt{2\omega_{\vec{p}_1}}}{p_1^2+m^2-i\epsilon}\int d^4x_2 \cdots d^4x_n e^{-i\sum_{i=2}^n p_i\cdot x_i}\langle 0|T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|-\vec{p}_1\rangle.
\tag{1.33}
$$

We'd like to carry on the above procedure for $\Phi(x_2)$, $\Phi(x_3)$ etc. For example, the integration over x_2^0 at late times contributes to a pole at $p_2^0 \to \omega_{\vec{p}_2}$. To see this, it is most clear to consider finite size wave packets instead of plane waves. Namely, consider

$$\begin{aligned}&\int \frac{d^3\vec{p}_1}{(2\pi)^3}\widetilde{f}_1(\vec{p}_1)G(p_1,\cdots,p_n)\\ &\longrightarrow \int \frac{d^3\vec{p}_1}{(2\pi)^3}\widetilde{f}_1(\vec{p}_1)\frac{-iN(2\pi)^{\frac{3}{2}}\sqrt{2\omega_{\vec{p}_1}}}{-(p_1^0)^2+\omega_{\vec{p}_1}^2-i\epsilon}\int d^4x_2\cdots d^4x_n e^{-i\sum_{i=2}^n p_i\cdot x_i}\langle \vec{p}_1|T\left[\Phi(x_2)\cdots\Phi(x_n)\right]|0\rangle,\end{aligned} \tag{1.34}$$

for a momentum space 1-particle wave function $\widetilde{f}_1(\vec{p}_1)$ peaking around some spatial momentum $\vec{k}_1$, and p_1^0 close to $\omega_{\vec{k}_1}$. The LHS is a time-ordered correlation function involving

$$\int \frac{d^3\vec{p}_1}{(2\pi)^3}\widetilde{f}_1(\vec{p}_1)\int d^4x_1\, e^{ip_1^0x_1^0-i\vec{p}_1\cdot\vec{x}_1}\Phi(x_1)=\int d^4x_1 f_1(x_1)\Phi(x_1). \tag{1.35}$$

$f_1(x_1)$ is a position space wave packet that moves with momentum $\sim \vec{k}_1$. Now perform the same operation on x_2, we have

$$\begin{aligned}&\int \frac{d^3\vec{p}_1}{(2\pi)^3}\widetilde{f}_1(\vec{p}_1)\int \frac{d^3\vec{p}_2}{(2\pi)^3}\widetilde{f}_2(\vec{p}_2)G(p_1,p_2,\cdots,p_n)\\ &\longrightarrow \prod_{i=1}^2\int \frac{d^3\vec{p}_i}{(2\pi)^3}\widetilde{f}_i(\vec{p}_i)\frac{-iN(2\pi)^{\frac{3}{2}}\sqrt{2\omega_{\vec{p}_i}}}{-(p_i^0)^2+\omega_{\vec{p}_i}^2-i\epsilon}\int d^4x_3\cdots d^4x_n\left(e^{-i\sum_{i=3}^n p_i\cdot x_i}\right){}^{out}\langle \vec{p}_1,\vec{p}_2|T\left[\Phi(x_3)\cdots\Phi(x_n)\right]|0\rangle.\end{aligned} \tag{1.36}$$

Here we assumed that at late times, i.e. $x_1^0, x_2^0 \to +\infty$, the wave packets $f(x_1)$ and $f(x_2)$ are far separated, so that $\Phi(x_1)$ and $\Phi(x_2)$ do not interfere. This is known as cluster decomposition. Acting on $|0\rangle$, they (or rather, their conjugates $\Phi^\dagger(x_1)$ and $\Phi^\dagger(x_2)$) create a state that looks like two particles far from one another in the far future. This is the definition of the "out" state $|p_1,p_2\rangle^{out}$.

Repeating this argument for all outgoing and incoming momenta yields the LSZ reduction formula, namely in the limit $p_i^0 \to \pm\omega_{\vec{p}_i} = \pm\sqrt{\vec{p}_i^2+m^2}$, with the appropriate positive or negative sign of p_i^0 for outgoing or incoming particles, the residue of the time ordered vacuum correlation function is related to the S-matrix element via

$$G(p_1,\cdots,p_n)\to\left[\prod_{j=1}^n\frac{-i}{p_j^2+m^2-i\epsilon}N(2\pi)^{\frac{3}{2}}\sqrt{2\omega_{\vec{p}_j}}\right]S(\vec{p}_1,\cdots,\vec{p}_\ell;-\vec{p}_{\ell+1},\cdots,-\vec{p}_n). \tag{1.37}$$

The power of LSZ reduction formula lies in the fact that it does not depend on the precise choice of the field operator $\Phi(x)$, as long as we put the external momenta on-shell, and insert the appropriate field renormalization constant N. There can be many different choices of field operators, all of which create a 1-particle state along with other stuff, but the renormalized on-shell amplitude must all be the same one and is given by the S-matrix element which is unambiguously defined.

It is easy to generalize the above analysis to the case where the sum of a subset of momenta goes on shell. Consider

$$q = p_1 + \cdots + p_\ell = -p_{\ell+1} - \cdots - p_n, \tag{1.38}$$

and take the limit $q^2 \to -m^2$. G has a pole in q^2,

$$G \to (2\pi)^4 \delta^4(p_1 + \cdots + p_n) \frac{-i(2\pi)^3 \sqrt{\vec{q}^2 + m^2}}{q^2 + m^2 - i\epsilon} M(0|p_1, \cdots, p_\ell|q) M(q|p_{\ell+1}, \cdots, p_n|0), \tag{1.39}$$

where $M(0|p_1, \cdots, p_\ell|q)$ for instance is amplitude between the vacuum $\langle 0|$ and the one-particle state $|q\rangle$,

$$(2\pi)^4 \delta^4(p_1 + \cdots + p_\ell - q) M(0|p_1, \cdots, p_\ell|q) = \int \prod_{i=1}^{\ell} d^4 x_i e^{-ip_i \cdot x_i} \langle 0| T\left[\Phi(x_1) \cdots \Phi(x_\ell)\right] |q\rangle. \tag{1.40}$$

$M(q|p_{\ell+1}, \cdots, p_n|0)$ is a similar transition amplitude. Note that the pole structure in G when an intermediate momentum goes on shell occurs whenever there is a one-particle state with some mass m, regardless of whether this particle is a fundamental or composite particle in the theory. An example is in QCD, where $\Phi(x)$ could be taken to be the quark-anti-quark bilinear operator, which creates among other things, a one-particle state of the pion. The QCD is a strongly coupled theory at long distances, and the asymptotic states consist of hadrons rather than quarks and gluons. But the above argument is entirely general and non-perturbative. It shows that there is a pole in the hadron scattering amplitude when some intermediate momentum square approaches $-m^2$ of the pion; the contribution from the exchange of the pion is governed by the propagator $-i/(q^2 + m^2 - i\epsilon)$ near the pole, regardless of the nature of the composition of pion.

1.3 Constraints from unitarity

Let us discuss some general constraints on scattering amplitudes from the unitarity of the S-matrix. The S-matrix takes the form

$$S = \mathbb{I} + iT, \tag{1.41}$$

or in terms its matrix elements,

$$S_{\beta\alpha} = \delta_{\beta\alpha} + i(2\pi)^4 \delta^4(p_\beta - p_\alpha) M_{\beta\alpha}, \tag{1.42}$$

where $\langle \beta | iT | \alpha \rangle = iT_{\beta\alpha} = i(2\pi)^4 \delta^4(p_\beta - p_\alpha) M_{\beta\alpha}$ is the transition amplitude. The unitarity condition $SS^\dagger = S^\dagger S = \mathbb{I}$ is equivalent to

$$-i(T - T^\dagger) = T^\dagger T, \tag{1.43}$$

or

$$-i(M_{\gamma\alpha} - M^*_{\alpha\gamma}) = \int d\beta (2\pi)^4 \delta^4(p_\beta - p_\alpha) M_{\beta\alpha} M^*_{\beta\gamma}. \tag{1.44}$$

It is useful to consider the case $\gamma = \alpha$, from which we obtain

$$2\text{Im} M_{\alpha\alpha} = \int d\beta (2\pi)^4 \delta^4(p_\beta - p_\alpha) |M_{\beta\alpha}|^2 \tag{1.45}$$

The LHS is the forward scattering amplitude from the state α to itself. The RHS is related to the total transition rate from the state α to other states. The transition probability from α to a different state β is

$$P_{\alpha\to\beta} = |S_{\beta\alpha}|^2 = VT(2\pi)^4 \delta^4(p_\beta - p_\alpha) |M_{\beta\alpha}|^2 \tag{1.46}$$

where $VT = (2\pi)^4 \delta(0)$ is the spacetime volume. The total transition rate per unit volume is

$$\Gamma_\alpha = \int d\beta \frac{dP_{\alpha\to\beta}}{VT} = \int d\beta (2\pi)^4 \delta^4(p_\beta - p_\alpha) |M_{\beta\alpha}|^2. \tag{1.47}$$

So we derived the result

$$2\text{Im} M_{\alpha\alpha} = \Gamma_\alpha. \tag{1.48}$$

The unitarity of the S-matrix also implies that $TT^\dagger = T^\dagger T$, or in terms of transition rates,

$$\int d\beta \frac{d\Gamma(\alpha \to \beta)}{d\beta} = \int d\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha}. \tag{1.49}$$

Consider some statistical ensemble, in which $P_\alpha d\alpha$ is the probability of finding the system in the state α. The rate of change of P_α is given by the difference between the rate of all other states β turning in α and the the rate of α turning into some other state β,

$$\frac{dP_\alpha}{dt} = \int d\beta P_\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha} - P_\alpha \int d\beta \frac{d\Gamma(\alpha \to \beta)}{d\beta}. \tag{1.50}$$

The entropy of the system,

$$S = -\int d\alpha P_\alpha \ln P_\alpha, \tag{1.51}$$

changes by

$$\begin{aligned} \frac{dS}{dt} &= -\int d\alpha \frac{dP_\alpha}{dt}(\ln P_\alpha + 1) \\ &= -\int d\alpha d\beta (\ln P_\alpha + 1) \left[P_\beta \frac{d\Gamma(\beta \to \alpha)}{d\alpha} - P_\alpha \frac{d\Gamma(\alpha \to \beta)}{d\beta} \right] \\ &= \int d\alpha d\beta \, P_\beta \left(\ln \frac{P_\beta}{P_\alpha} \right) \frac{d\Gamma(\beta \to \alpha)}{d\alpha} \\ &\geq \int d\alpha d\beta \, (P_\beta - P_\alpha) \frac{d\Gamma(\beta \to \alpha)}{d\alpha}. \end{aligned} \tag{1.52}$$

In going from the second to the third line, we exchanged the integration variables α with β in the second term of the second line. The inequality in the last step holds for any non-negative P_α, P_β. Now we use the relation (1.49) and immediately see that the RHS of (1.52) vanishes. This establishes the Boltzmann H-theorem $dS/dt \geq 0$.

Suppose α is a two-particle asymptotic state, $|\vec{k}_1, \vec{k}_2\rangle$, and the transition amplitude describes the scattering of the two particles. We often normalize the one-particle states with the convention $\langle \vec{k}|\vec{k}'\rangle = \delta^3(\vec{k} - \vec{k}')$, and so $\langle \vec{k}|\vec{k}\rangle = \frac{V}{(2\pi)^3}$. The two-particle state then has the norm $\frac{V^2}{(2\pi)^6}$ in our convention. Let v be the relativity velocity between the two particles, and σ be the total cross section. The transition rate per unit volume is then related by $\Gamma_\alpha = \frac{V^2}{(2\pi)^6}\frac{1}{V^2}v\sigma = \frac{v\sigma}{(2\pi)^6}$.

Consider the case of $2 \to 2$ scattering $\alpha \to \beta$, where β is also a two-body state. Recall that the differential cross section $d\sigma(\alpha \to \beta)$ is related to the matrix element $M_{\beta\alpha}$ by

$$\begin{aligned} d\sigma(\alpha \to \beta) &= \frac{(2\pi)^6}{v}(2\pi)^4\delta^4(p_\beta - p_\alpha)|M_{\beta\alpha}|^2 d\beta \\ &= \frac{(2\pi)^6}{v}(2\pi)^4\delta^3(\vec{p}_1' + \vec{p}_2' - \vec{p}_\alpha)\delta(E_1' + E_2' - E_\alpha)|M_{\beta\alpha}|^2 d^3\vec{p}_1' d^3\vec{p}_2' \\ &= \frac{(2\pi)^6}{v}(2\pi)^4\frac{|\vec{p}_1'|^2}{\frac{dE_1'}{d|\vec{p}_1'|} + \frac{dE_2'}{d|\vec{p}_1'|}}|M_{\beta\alpha}|^2 d\Omega. \end{aligned} \tag{1.53}$$

In the center of mass frame, $|\vec{p}_1| = |\vec{p}_2| = k$, $|\vec{p}_1'| = |\vec{p}_2'| = k'$. We have

$$d\sigma(\alpha \to \beta) = \frac{(2\pi)^6}{v}(2\pi)^4\frac{k'E_1'E_2'}{E_\alpha}|M_{\beta\alpha}|^2 d\Omega = (2\pi)^{10}\frac{k'E_1'E_2'E_1E_2}{kE_\alpha^2}|M_{\beta\alpha}|^2 d\Omega. \tag{1.54}$$

Conventionally, the scattering amplitude $f(\alpha \to \beta)$ is defined with the normalization

$$f(\alpha \to \beta) = \frac{(2\pi)^5}{E_\alpha}\sqrt{\frac{k'E_1'E_2'E_1E_2}{k}}M_{\beta\alpha}, \tag{1.55}$$

so that the differential cross section is just the square of the amplitude,

$$\frac{d\sigma(\alpha \to \beta)}{d\Omega} = |f(\alpha \to \beta)|^2. \tag{1.56}$$

In particular, the forward scattering amplitude is

$$f(\alpha \to \alpha) = (2\pi)^5\frac{E_1E_2}{E_\alpha}M_{\alpha\alpha} = (2\pi)^5\frac{k}{v}M_{\alpha\alpha}. \tag{1.57}$$

We then have

$$\text{Im}f(\alpha \to \alpha) = \frac{k}{4\pi}\sigma. \tag{1.58}$$

This result is known as the optical theorem.

Suppose the $2 \to 2$ amplitude squared $|f|^2$ (which is part of the contribution to the total cross section σ) is close to that of the forward amplitude within a solid angle $\Delta\Omega$, say $|f|^2 \geq \frac{1}{2}|f(\alpha \to \alpha)|^2$. Then the optical theorem implies the inequality

$$\sigma \geq \frac{1}{2}|f(\alpha \to \alpha)|^2 \Delta\Omega \geq \frac{1}{2}|\text{Im} f(\alpha \to \alpha)|^2 \Delta\Omega = \frac{k^2}{32\pi^2}\sigma^2 \Delta\Omega. \tag{1.59}$$

It follows that there is an upper bound on $\Delta\Omega$,

$$\Delta\Omega \leq \frac{32\pi^2}{k^2\sigma}. \tag{1.60}$$

At high energies, σ typically grows or approach a constant, in which case $\Delta\Omega$ goes to zero like $1/k^2$, and the scattering amplitude is sharply peaked in the forward direction.

For analyzing angular distribution of cross section and scattering amplitudes, it is convenient to work in the angular momentum basis $|\ell m\rangle$, where ℓ is the total angular momentum, and m the z-component of the angular momentum. For simplicity we will ignore spin for now. Consider a two (non-identical) particle asymptotic state $|\vec{p}_1, \vec{p}_2\rangle$. We may change to the basis $|\vec{p}E\ell m\rangle$, where $\vec{p}$ and E are the total momentum and energy, and ℓ, m label the orbital angular momentum. In below we will mostly work in the center of mass frame, where $\vec{p} = 0$. The overlap is given by

$$\langle \vec{p}_1, \vec{p}_2 | 0 E \ell m \rangle = \delta^3(\vec{p}_1 + \vec{p}_2)\delta(\omega_{\vec{p}_1} + \omega_{\vec{p}_2} - E)\sqrt{\frac{E}{|\vec{p}_1|E_1E_2}} Y_\ell^m(\hat{p}_1). \tag{1.61}$$

Here $Y_\ell^m(\hat{p}_1)$ are the spherical harmonics, as a function of the direction of $\vec{p}_1$. The state $|\vec{p}E\ell m\rangle$ is normalized so that

$$\begin{aligned}\langle \vec{p}E\ell m | 0 E'\ell' m' \rangle &= \delta^3(\vec{p})\delta(E - E') \int d^3\vec{p}_1 d^3\vec{p}_2 \delta^3(\vec{p}_1 + \vec{p}_2)\delta(\omega_{\vec{p}_1} + \omega_{\vec{p}_2} - E)\frac{E}{|\vec{p}_1|E_1E_2} Y_\ell^{m*}(\hat{p}_1) Y_{\ell'}^{m'}(\hat{p}_1) \\ &= \delta^3(\vec{p})\delta(E - E') \int |\vec{p}_1|^2 d|\vec{p}_1| d\hat{p}_1 \delta(E_1 + E_2 - E)\frac{E}{|\vec{p}_1|E_1E_2} Y_\ell^{m*}(\hat{p}_1) Y_{\ell'}^{m'}(\hat{p}_1) \\ &= \delta^3(\vec{p})\delta(E - E')\delta_{\ell\ell'}\delta_{mm'}.\end{aligned} \tag{1.62}$$

If the two particles are identical, the integration $\int d^3\vec{p}_1 d^3\vec{p}_2$ is only over half of the space, and so for the normalization (1.62) to hold for the angular momentum states we need an extra factor of $\sqrt{2}$ on the RHS of (1.61).

Now the $2 \to 2$ S-matrix elements can be written as

$$S_{E'\vec{p}'\ell'm', E\vec{p}\ell m} = \delta^3(\vec{p})\delta(E - E') S_{\ell'm', \ell m}(E, \vec{p}), \tag{1.63}$$

with

$$S_{\ell' m',\ell m}(E,\vec{p}) = \delta_{\ell\ell'}\delta_{mm'} + i(2\pi)^4 M_{\ell' m',\ell m}(E,\vec{p}). \tag{1.64}$$

Note that the identity part of the S-matrix is now the identity matrix on the discrete partial wave basis. In the center of mass frame, i.e. $\vec{p}=0$, rotational invariance implies

$$S_{\ell' m',\ell m}(E,0) = \delta_{\ell\ell'}\delta_{mm'} S_\ell(E). \tag{1.65}$$

The scattering amplitude can be expressed in terms of the partial wave matrix elements as

$$\begin{aligned} f(\vec{k},-\vec{k}\longrightarrow\vec{k}',-\vec{k}') &= \frac{(2\pi)^5}{E_{tot}}\sqrt{\frac{k' E_1' E_2' E_1 E_2}{k}} M_{\vec{k}',-\vec{k}';\vec{k},-\vec{k}} \\ &= \frac{(2\pi)^5}{k}\sum_{\ell,m;\ell',m'} Y_{\ell'}^{m'*}(\hat{k}')Y_\ell^m(\hat{k})M_{\ell' m',\ell m}(E_{tot}) \\ &= \frac{2\pi i}{k}\sum_{\ell,m} Y_\ell^{m*}(\hat{k}')Y_\ell^m(\hat{k})(1-S_\ell(E)) \\ &= \frac{i}{2k}\sum_{\ell=0}^{\infty}(2\ell+1)P_\ell(\hat{k}\cdot\hat{k}')(1-S_\ell(E)). \end{aligned} \tag{1.66}$$

Here $P_\ell(x)$ is the Legendre polynomial of degree ℓ. The total two-body channel cross section (including both elastic and inelastic cross sections) is

$$\sigma_{2\to 2}(E) = \frac{\pi}{k^2}\sum_{\ell=0}^{\infty}(2\ell+1)|1-S_\ell(E)|^2. \tag{1.67}$$

Let us compare this to $4\pi/k$ times the imaginary part of the forward scattering amplitude,

$$\frac{4\pi}{k}\mathrm{Im} f(\vec{k},-\vec{k}\longrightarrow\vec{k},-\vec{k}) = \frac{2\pi}{k^2}\sum_{\ell=0}^{\infty}(2\ell+1)\mathrm{Re}(1-S_\ell(E)). \tag{1.68}$$

The optical theorem equates this to the total cross section $\sigma_{tot}(E)$, which is at least as big as the two- body channel cross section. Indeed,

$$2\mathrm{Re}(1-S_\ell(E)) = |1-S_\ell(E)|^2 + (1-|S_\ell(E)|^2). \tag{1.69}$$

$|S_\ell(E)|^2 < 1$ if there is a nonzero cross section for producing extra particles. If the two-body channel is the only channel at angular momentum ℓ, then $S_\ell(E) = e^{i\delta_\ell}$ for some phase δ_ℓ. Generally, it follows from $|S_\ell(E)|^2 \le 1$ that the partial wave cross section at each ℓ is bounded by $1/k^2$ times a constant. Thus we learned that unitarity implies that cross sections cannot be too large. But we have not put a bound on the *total* cross section so far, which we now discuss.

We will try to bound the total cross section by bounding the forward scattering amplitude. For this purpose, it will suffice to consider $2 \to 2$ scattering, with the same particles in the final state as in the initial state. It follows from the partial wave decomposition of the scattering amplitude, clearly, that

$$|f(\vec{k}' = \vec{k})| \leq \frac{1}{k}\sum_{\ell=0}^{L}(2\ell+1) + \frac{1}{2k}\sum_{\ell=L+1}^{\infty}(2\ell+1)|a_\ell(E)|, \tag{1.70}$$

where $ia_\ell(E) = S_\ell(E) - 1$, and L is so far an arbitrary number, which will be chosen suitably later. It will turn out that $a_\ell(E)$ decreases exponentially, for sufficiently large ℓ, say $\ell > K$ for some wave number K, which in turn puts a bound on the forward scattering amplitude.

Let the four-momenta of the initial particles be p_1, p_2 and the final particles p_3, p_4. The Lorentz invariant Mandelstam variables are defined as

$$\begin{aligned} s &= -(p_1+p_2)^2 = -(p_3+p_4)^2, \\ t &= -(p_1-p_3)^2 = -(p_2-p_4)^2, \\ u &= -(p_1-p_4)^2 = -(p_2-p_3)^2. \end{aligned} \tag{1.71}$$

For simplicity we will assume all masses are equal to m, and so $s + t + u = 4m^2$. In terms of the incident momenta $\vec{p}_1 = \vec{p}, \vec{p}_2 = -\vec{p}$, and the scattering angle θ, we have

$$\begin{aligned} s &= 4(\vec{p}^2 + m^2), \\ t &= -2\vec{p}^2(1-\cos\theta), \\ \cos\theta &= 1 + \frac{2t}{s-4m^2}. \end{aligned} \tag{1.72}$$

The scattering amplitude $f(p_1, p_2 \to p_3, p_4)$ will now be simply denoted $\mathcal{A}(\cos\theta)$. J. Schwinger once said, "One of the most remarkable discoveries of elementary particle physics has been that of the existence of the complex plane." We will analytically continue $\mathcal{A}(z)$ to the complex z-plane, and the analytic properties of $\mathcal{A}(z)$ will be crucial in determining the behavior of the amplitude in various limits.

The partial wave amplitudes a_ℓ can be written in terms of $\mathcal{A}(\cos\theta)$ as

$$a_\ell(s) = k\int_{-1}^{1}\mathcal{A}(x)P_\ell(x)dx. \tag{1.73}$$

We will make the assumption that the amplitude is bounded by some polynomial in s, t, u. At large z, then, $\mathcal{A}(z)$ should grow no faster than some power of z. $\mathcal{A}(z)$ may have singularities on the complex z-plane; these singularities are always associated with intermediate physical states. For instance, when either s or t or u is equal to

$4m^2$, the intermediate channel is at the threshold of allowing an on-shell two-particle state. Generally, the amplitude has a singularity (branch point) at these points. On the z-plane, these singularities occur at

$$z = \infty, \quad \frac{s+4m^2}{s-4m^2}, \quad \text{and} \quad -\frac{s+4m^2}{s-4m^2}. \tag{1.74}$$

Typically, there are branch cuts extending from $w = \frac{s+4m^2}{s-4m^2}$ to $+\infty$, and from $-w = -\frac{s+4m^2}{s-4m^2}$ to $-\infty$, on the z-plane. We will see that these assumptions together will imply that $a_\ell(s)$ decreases exponentially in ℓ at sufficiently large ℓ.

By our assumption, $\mathcal{A}(z) = z^{K+1}h(z)$, where $h(z)$ goes to zero at large z and K is some integer constant. We may now write

$$\begin{aligned} \frac{1}{k}a_\ell(s) &= \int_{-1}^{1} dx\, P_\ell(x) x^{K+1} h(x) \\ &= \int_{-1}^{1} dx\, P_\ell(x) x^{K+1} \oint_C \frac{dz}{2\pi i} \frac{h(z)}{z-x} \end{aligned} \tag{1.75}$$

The contour C, which goes around x and avoids the branch cuts of $\mathcal{A}(z)$ or $h(z)$, can be deformed to the sum of two contours, C_1 that goes clockwise around the branch cut $(-\infty, -w]$, and C_2 that goes clockwise around $[w, +\infty)$. After doing so, as the integral over z and over x both converges, we may exchange their order, and write

$$\frac{1}{k}a_\ell(s) = \oint_{C_1+C_2} \frac{dz}{2\pi i} h(z) \int_{-1}^{1} dx\, \frac{x^{K+1}}{z-x} P_\ell(x) \tag{1.76}$$

Now let us consider large wave numbers, $\ell > K$, and make use of the fact that the Legendre polynomial $P_\ell(x)$ is orthogonal to all polynomials of degree $< \ell$ on the interval $[-1, 1]$, which implies the identity

$$\int_{-1}^{1} dx\, \frac{x^{K+1}}{z-x} P_\ell(x) = z^{K+1} \int_{-1}^{1} dx \frac{P_\ell(x)}{z-x} = -2z^{K+1} Q_\ell(z). \tag{1.77}$$

$Q_\ell(z)$ is the Legendre function of the second kind. Note that $Q_\ell(z)$ is a priori well defined for z away from $[-1, 1]$, and is real on the real z-axis outside $[-1, 1]$. It has logarithmic singularity at $z = \pm 1$, but this does not affect the convergence of the integral over z along C_1, C_2. Now we can write

$$\frac{1}{k}a_\ell(s) = -2 \oint_{C_1+C_2} \frac{dz}{2\pi i} \mathcal{A}(z) Q_\ell(z), \tag{1.78}$$

where the integral on the RHS picks up the contribution from the jump of $\mathcal{A}(z)$ across the two branch cuts, $(-\infty, -w]$ and $[w, +\infty)$. The result thus takes the form

$$\frac{1}{k}a_\ell(s) = \frac{i}{\pi} \int_w^\infty dx \rho(x) Q_\ell(x), \tag{1.79}$$

where $|\rho(x)| < B|x|^K s^{A-\frac{1}{2}}$ for some constant B and A. Using the integral representation of the Legendre polynomial, we can derive an integral representation of $Q_\ell(x)$, namely for $x > 1$,

$$Q_\ell(x) = \int_{x+\sqrt{x^2-1}}^{\infty} \frac{d\zeta}{\zeta^{\ell+1}\sqrt{1-2\zeta x+\zeta^2}}. \tag{1.80}$$

In fact, apart from normalization factors, this differs from the integral representation of $P_\ell(x)$ only by a different choice of contour. A simple estimate of the integral results in the bound

$$\begin{aligned}
\frac{1}{k}|a_\ell(s)| &< \frac{B}{\pi} s^{A-\frac{1}{2}} \int_w^\infty dx\, x^K \int_{x+\sqrt{x^2-1}}^{\infty} \frac{d\zeta}{\zeta^{\ell+1}\sqrt{1-2\zeta x+\zeta^2}} \\
&< C' s^{A-\frac{1}{2}} \int_w^\infty dx \frac{x^K}{(x+\sqrt{x^2-1})^{\ell+1}} \\
&< C' \frac{s^{A-\frac{1}{2}}}{\ell - K} \left(\frac{\sqrt{s}-2m}{\sqrt{s}+2m}\right)^{\ell-K},
\end{aligned} \tag{1.81}$$

where C' is a constant (proportional to B). In above we have used $w + \sqrt{w^2-1} = \frac{\sqrt{s}+2m}{\sqrt{s}-2m}$. For fixed large s, $a_\ell(s)$ decreases exponentially in ℓ ($> K$). On the other hand, unitarity by itself constrains $|a_\ell(s)| < 2$. Only for sufficiently large ℓ we need to invoke the bound (1.81). In fact, the bound is efficiently implemented if we use (1.81) only for $\ell > L$, and use $|a_\ell(s)| < 2$ for $\ell \leq L$, where L is chosen to be

$$L \sim \frac{A}{4m}\sqrt{s}\left(\ln s + const\right). \tag{1.82}$$

A straightforward calculation of (1.70) then gives

$$|f(\vec{k}' = \vec{k})| < \frac{C}{k} s (\ln s)^2 \tag{1.83}$$

at large s. By optical theorem, finally, we find the bound

$$\sigma_{tot} < C(\ln s)^2 \tag{1.84}$$

for some constant C, at large s. This is known as Froissart bound.

2 Infrared divergences

Scattering amplitudes in QED generally suffers from the problem that loops of "soft", i.e. very low energy, photons attached on external propagators lead to large, in fact, divergent, radiative corrections. On the other hand, there are also infrared divergences associated with emission of real soft photons. The emitted photons with energy below some threshold cannot be observed in experiments, and scattering with the emission of such photons contribute to the same observed amplitude as those without the soft photon emission. We will see that this effect precisely cancel the infrared divergences in radiative corrections, which is crucial for the consistency of the theory.

2.1 Soft photon emission

Consider a general amplitude $\alpha \to \beta$ in scalar QED, where the scalar field ϕ has charge e and mass m. Suppose one of the outgoing scalar lines has momentum p (p is on-shell, $p^2 = -m^2$). Call this amplitude $\mathcal{A}(p, \cdots)$. Now consider the same amplitude with an extra photon of momentum q emitted from this scalar line. The new amplitude is of the form

$$\mathcal{A}(p+q, \cdots) \frac{-i}{(p+q)^2 + m^2 - i\epsilon} ie(2p^\mu + q^\mu) \tag{2.1}$$

where μ is the index of the photon field, to be contracted with a polarization vector. Strictly speaking, $\mathcal{A}(p+q, \cdots)$ is an off-shell amplitude, but in the limit of interest $q \to 0$ it becomes on-shell. The factor multiplying $\mathcal{A}(p+q, \cdots)$ becomes

$$\frac{-i}{(p+q)^2 + m^2 - i\epsilon} ie(2p^\mu + q^\mu) \;\to\; \frac{e\, p^\mu}{p \cdot q - i\epsilon}. \tag{2.2}$$

This factor diverges as $q \to 0$.

Let us examine the analogous amplitude in spinor QED, where the electron ψ has charge e and mass m. Consider an amplitude of the form

$$\bar{u}_\sigma(p) \mathcal{A}(p, \cdots), \tag{2.3}$$

where p is the momentum of an outgoing electron. $u_\sigma(p)$ is the spinor polarization, and σ is a helicity label. Then there is a similar amplitude with an extra soft photon emitted from the electron line,

$$\bar{u}_\sigma(p)(-e\gamma^\mu) \frac{-i}{i(\not{p} + \not{q}) + m - i\epsilon} \mathcal{A}(p+q, \cdots). \tag{2.4}$$

The polarization spinors obey the completeness relation

$$\sum_{\sigma'} u_{\sigma'}(p) \bar{u}_{\sigma'}(p) = \frac{-i\not{p} + m}{2p^0}, \tag{2.5}$$

as well as the orthogonality relation

$$\bar{u}_\sigma(p) \gamma^\mu u_{\sigma'}(p) = -i\delta_{\sigma\sigma'} \frac{p^\mu}{p^0}. \tag{2.6}$$

Using these, we find that in the $q \to 0$ limit,

$$\begin{aligned} \bar{u}_\sigma(p)(-e\gamma^\mu)\frac{-i}{i(\not{p}+\not{q})+m-i\epsilon} \;\to\; & ie\sum_{\sigma'}\bar{u}_\sigma(p)\gamma^\mu u_{\sigma'}(p)\frac{p^0}{p\cdot q-i\epsilon}\bar{u}_{\sigma'}(p) \\ & = \frac{ep^\mu}{p\cdot q-i\epsilon}\bar{u}_\sigma(p). \end{aligned} \tag{2.7}$$

This is precisely the same factor as (2.2). The same factor shows up when a soft photon is emitted from an external particle line of any spin. Furthermore, in the $q \to 0$ limit, the momenta of the electron lines on both sides of the soft photon vertex are on-shell, and the vertex with such values of momenta is not renormalized by the Ward identity. So the soft photon emission factor (2.2) in fact holds to all order in perturbation theory.

A soft photon emitted from an ingoing electron line with momentum p comes with the same factor, except that p is replaced by $-p$ in (2.2). The total amplitude with a soft photon emission, $M^\mu_{\beta\alpha}$, is given by summing over soft photon emission amplitudes from each external line. It is related to the amplitude $M_{\beta\alpha}$ without the soft photon by

$$M^\mu_{\beta\alpha}(q) \to M_{\beta\alpha}\sum_n \frac{e_n p_n^\mu}{\eta_n p_n\cdot q-i\epsilon}, \tag{2.8}$$

where n runs through all external lines, e_n is the charge of the corresponding particle, $\eta_n = +1$ (-1) for outgoing (incoming) lines.

Next, let us consider the emission of two soft photons, of momenta $q_1, q_2 \to 0$. If they are emitted from different external lines, the amplitude is multiplied by two factors like the one in (2.8). If the two soft photons are emitted from the same external line of momentum p, the amplitude is multiplied by

$$\frac{ep^{\mu_1}}{\eta p\cdot q_1-i\epsilon}\cdot\frac{ep^{\mu_2}}{\eta p\cdot(q_2+q_1)-i\epsilon}+(1\leftrightarrow 2) = \frac{ep^{\mu_1}}{\eta p\cdot q_1-i\epsilon}\cdot\frac{ep^{\mu_2}}{\eta p\cdot q_2-i\epsilon}, \tag{2.9}$$

which is once again simply the product of two factors as the one in (2.8). It is easy to generalize this to the amplitude with N soft photon emissions, which is

$$M_{\beta\alpha}\prod_{r=1}^N\left(\sum_n \frac{e_n p_n^{\mu_r}}{\eta_n p_n\cdot q_r-i\epsilon}\right). \tag{2.10}$$

Taking into account polarization vectors $\epsilon^\mu(q_r,h_r)$ for the r-th emitted soft photon, where $h_r = \pm 1$ is the helicity, we should contract the amplitude with

$$\prod_{r=1}^N \frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{2|\vec{q}_r|}}\epsilon^*_{\mu_r}(q_r,h_r) \tag{2.11}$$

Here the polarization vector is normalized so that the orthogonality and completeness relations take the form

$$\begin{aligned}
&\epsilon(q,h)\cdot\epsilon^*(q,h') = \delta_{hh'}, \\
&\sum_{h=\pm1}\epsilon_\mu(q,h)\epsilon^*_\nu(q,h) = \eta_{\mu\nu} + q_\mu c_\nu + q_\nu c_\mu, \qquad c^\mu = \frac{1}{2|\vec{q}|^2}(q^0,-\vec{q}).
\end{aligned} \tag{2.12}$$

The transition rate $\Gamma_{\beta\alpha}$ with N soft photons in the phase space $\prod_r d^3\vec{q}_r$ is therefore enhanced by the factor

$$\prod_{r=1}^{N}\frac{d^3\vec{q}_r}{(2\pi)^3 2|\vec{q}|}\sum_{n,m}\frac{\eta_n\eta_m e_n e_m}{(p_n\cdot q_r)(p_m\cdot q_r)}\sum_{h_r=\pm1}(p_n\cdot\epsilon(q_r,h_r))(p_m\cdot\epsilon^*(q_r,h_r)) \tag{2.13}$$

Now applying the completeness relation, the terms proportional to c_μ in (2.12) are unphysical and must drop out in order to preserve Lorentz invariance. This is ensured by charge conservation $\sum_n \eta_n e_n = 0$. The result is that we have the enhancement factor

$$\prod_{r=1}^{N}\frac{d^3\vec{q}_r}{(2\pi)^3 2|\vec{q}|}\sum_{n,m}\frac{\eta_n\eta_m e_n e_m}{(p_n\cdot q_r)(p_m\cdot q_r)}(p_n\cdot p_m) \tag{2.14}$$

The momentum integration over $\vec{q}_r$ can be separated into the integration over energy ω_r, and the angular integral, which is computed as

$$\begin{aligned}
&\frac{p_n\cdot p_m}{2(2\pi)^3}\int d^2\hat{q}\frac{1}{(p_n^0-\vec{p}_n\cdot\hat{q})(p_m^0-\vec{p}_m\cdot\hat{q})} \\
&=\frac{p_n\cdot p_m}{2(2\pi)^3}\int_0^1 dx\int d^2\hat{q}\frac{1}{[p_n^0 x + p_m^0(1-x) - (\vec{p}_n x + \vec{p}_m(1-x))\cdot\hat{q})]^2} \\
&=\frac{p_n\cdot p_m}{2(2\pi)^2}\int_0^1 dx\int_0^\pi d\theta\frac{\sin\theta}{[p_n^0 x + p_m^0(1-x) - |\vec{p}_n x + \vec{p}_m(1-x)|\cos\theta]^2} \\
&=\frac{p_n\cdot p_m}{2(2\pi)^2}\int_0^1 dx\frac{2}{m_n^2x^2 + m_m^2(1-x)^2 - 2p_n\cdot p_m x(1-x)} \\
&= -\frac{1}{8\pi^2\beta_{nm}}\ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right).
\end{aligned} \tag{2.15}$$

where

$$\beta_{nm} \equiv \sqrt{1-\frac{m_n^2 m_m^2}{(p_n\cdot p_m)^2}}. \tag{2.16}$$

Here m_n is the mass of the particle on the n-th external line. We then obtain the integrated N soft photon emission factor

$$\frac{A^N}{N!}\int\prod_{r=1}^{N}\frac{d\omega_r}{\omega_r}, \tag{2.17}$$

where

$$A = -\frac{1}{8\pi^2} \sum_{n,m} \frac{e_n e_m \eta_n \eta_m}{\beta_{nm}} \ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right). \tag{2.18}$$

The symmetry factor $1/N!$ is included due to permutations on the N photons. One can check that A is in fact always positive. The integration over ω_r diverges logarithmically in the infrared. We may introduce an IR cutoff μ and take $\mu \to 0$ in the end. In experiments, typically, photons below some energy E_T are not observed, and the corresponding transition rate indistinguished from the transition without soft photon emission. The details depends on the precise experiment. Roughly, E_T will be the upper bound of the integration over ω_r. Summing over all possible soft photon emissions of energies below E_T and above the IR cutoff μ then exponentiates (2.17) and gives the enhancement factor

$$\sim \left(\frac{E_T}{\mu}\right)^A. \tag{2.19}$$

2.2 Virtual soft photons

Now let us consider the loop corrections due to virtual soft photons emitted or absorbed by external lines. We will restrict the spatial momenta of the soft photons to be below a cutoff Λ, which is taken to be small compared to the momenta of external lines. Note that Λ here is *not* to be confused with the UV cutoff one uses to regularize the theory. The rest of the amplitude is computed by integrating over momenta *above* Λ. To regularize the IR divergence, we must introduce another IR cutoff μ, and take $\mu \to 0$ in the end. The same IR regulator μ was used to integrate over the real soft photon emissions. Be aware that this separation of momentum integration at scale Λ and regulator μ is not Lorentz invariant.

For instance, the exchange of a virtual soft photon between two external lines of momenta p_1 and p_2 contributes a factor

$$\int_{\mu<|\vec{q}|<\Lambda} \frac{d^4q}{(2\pi)^4} \frac{-i}{q^2 - i\epsilon} \cdot \frac{ep_1^\mu}{p_1 \cdot q - i\epsilon} \cdot \frac{ep_{2\mu}}{-p_2 \cdot q - i\epsilon}. \tag{2.20}$$

The contribution from the exchange of N soft photons can be computed by treating each exchange independently, by virtue of the previous section. Taking into account a symmetry factor $1/N!$ due to the permutation of N virtual soft photon propagators, we have the factor

$$\frac{1}{N!2^N} \left[\sum_{n,m \in \text{ext. lines}} e_n e_m J_{nm} \right]^N \tag{2.21}$$

multiplying the amplitude, where

$$J_{n,m} = -ip_n \cdot p_m \int_{\mu<|\vec{q}|<\Lambda} \frac{d^4q}{(2\pi)^4} \frac{1}{(q^2 - i\epsilon)(\eta_n p_n \cdot q - i\epsilon)(-\eta_m p_m \cdot q - i\epsilon)}. \tag{2.22}$$

Note that we have also included terms with $n = m$, with a factor of $\frac{1}{2}$. This may seem wrong because there are no such contributions to the amputated Feynman diagram. But in fact these terms precisely account for the IR divergent contribution to the electron field renormalization. To understand this, observe that $J_{n,n}$ computes the IR divergent 1-loop correction to the electron-photon vertex with zero photon momentum, in the limit where the external electron lines go on-shell. There is an analogous IR divergent contribution to the electron self-energy, denoted by δZ_2, that is given by $-J_{n,n}$. The corresponding field renormalization factors, $(Z_2^{\frac{1}{2}})^2 = Z_2$, precisely cancel the vertex correction when the external momenta are put on-shell. By LSZ reduction formula, in computing the S-matrix elements, each external electron line comes with a factor $Z_2^{\frac{1}{2}}$.[3] The IR divergences in these $Z_2^{\frac{1}{2}}$ factors are precisely accounted for by the $n = m$ terms in (2.22).

Now the sum over arbitrary number of virtual soft photons gives the exponential factor that multiplies the rest of the amplitude

$$M^{\mu}_{\beta\alpha} = M^{\Lambda}_{\beta\alpha} e^{\frac{1}{2}\sum_{n,m} e_n e_m J_{nm}}. \tag{2.23}$$

Let us calculate J_{nm}. The integration over q^0 from $-\infty$ to ∞ picks up the contribution from residues at a subset of the four poles:

$$q^0 = |\vec{q}| - i\epsilon, \quad -|\vec{q}| + i\epsilon, \quad \frac{\vec{p}_n \cdot \vec{q}}{p_n^0} - i\eta_n\epsilon, \quad \frac{\vec{p}_m \cdot \vec{q}}{p_m^0} + i\eta_m\epsilon. \tag{2.24}$$

If $\eta_n = -\eta_m = \pm 1$, we can close the contour on either the upper half or lower half q^0-plane and pick up the contribution from a single residue at $q^2 = \mp(|\vec{q}| - i\epsilon)$. If $\eta_n = \eta_m = \pm 1$, we pick up two residues by closing the contour on either upper or lower half plane. In the former case, we have

$$\begin{aligned} J_{nm} &= p_n \cdot p_m \int_{\mu<|\vec{q}|<\Lambda} \frac{d^3q}{(2\pi)^3} \frac{1}{2|\vec{q}|(\eta_n p_n \cdot q)(-\eta_m p_m \cdot q)} \bigg|_{q^0 = |\vec{q}| - i\epsilon} \\ &= \frac{p_n \cdot p_m}{2} \int_{\mu<|\vec{q}|<\Lambda} \frac{d^3q}{(2\pi)^3} \frac{1}{|\vec{q}|^3 (p_n^0 - \vec{p}_n \cdot \hat{q})(p_m^0 - \vec{p}_m \cdot \hat{q})}. \end{aligned} \tag{2.25}$$

[3] This is in the convention where we work with the unrenormalized electron fields ψ. In terms of the renormalized electron fields ψ_n, each vertex comes with a factor Z_2 and each internal electron propagator comes with a factor Z_2^{-1}, and there are no Z_2 factors associated with external lines. The net result is, of course, equivalent to attaching a $Z^{\frac{1}{2}}$ factor to each external leg in the diagram expressed in terms of unrenormalized fields ψ.

In the latter case, we have

$$\begin{aligned} J_{nm} &= p_n \cdot p_m \int_{\mu<|\vec{q}|<\Lambda} \frac{d^3q}{(2\pi)^3} \left[\frac{1}{2|\vec{q}|(p_n\cdot q)(-p_m\cdot q)}\Bigg|_{q^0=|\vec{q}|-i\epsilon} + \frac{1}{(q^2-i\epsilon)(-p_n^0)(-p_m\cdot q-i\epsilon)}\Bigg|_{q^0=\frac{\vec{p}_n\cdot\vec{q}}{p_n^0}-i\epsilon} \right] \\ &= -\frac{p_n\cdot p_m}{2} \int_{\mu<|\vec{q}|<\Lambda} \frac{d^3q}{(2\pi)^3} \frac{1}{|\vec{q}|^3(p_n^0-\vec{p}_n\cdot\hat{q})(p_m^0-\vec{p}_m\cdot\hat{q})} \\ &\quad + \frac{p_n\cdot p_m}{p_n^0 p_m^0} \int_{\mu<|\vec{q}|<\Lambda} \frac{d^3q}{(2\pi)^3} \frac{1}{|\vec{q}|^3(1-(\frac{\vec{p}_n\cdot\hat{q}}{p_n^0})^2)(\frac{\vec{p}_m\cdot\hat{q}}{p_m^0}-\frac{\vec{p}_n\cdot\hat{q}}{p_n^0}+i\epsilon)}. \end{aligned} \tag{2.26}$$

The $i\epsilon$ in the last term is important to make the integral well defined. This integral goes to minus itself under complex conjugation combined with the change of variable $\vec{q} \to -\vec{q}$, and hence the result is purely imaginary. This imaginary term contributes only a phase to the full amplitude, which is logarithmically divergent in the $\mu \to 0$ limit. We are interested in the transition rate $\Gamma_{\beta\alpha}$, and the soft photon contribution to $\Gamma_{\beta\alpha}$ involves only the real part of J_{nm},

$$\begin{aligned} \mathrm{Re} J_{nm} &= -\eta_n\eta_m \frac{p_n\cdot p_m}{2} \int_{\mu<|\vec{q}|<\Lambda} \frac{d^3q}{(2\pi)^3} \frac{1}{|\vec{q}|^3(p_n^0-\vec{p}_n\cdot\hat{q})(p_m^0-\vec{p}_m\cdot\hat{q})} \\ &= \frac{\eta_n\eta_m}{8\pi^2\beta_{nm}} \ln\left(\frac{1+\beta_{nm}}{1-\beta_{nm}}\right) \ln\frac{\Lambda}{\mu}. \end{aligned} \tag{2.27}$$

The effect on the transition rate due to soft photons in the energy range (μ, Λ) is therefore

$$\Gamma^{\mu}_{\beta\alpha} = \left(\frac{\mu}{\Lambda}\right)^A \Gamma^{\Lambda}_{\beta\alpha}. \tag{2.28}$$

Since A is positive, the effect of virtual soft photons is to multiply the amplitude or transition rate by a factor that goes to zero as the IR cutoff μ is sent to zero. But we see that it precisely cancel s the logarithmic divergence in the enhancement factor due to unobservable soft photon emissions, making the observed transition rate finite in the $\mu \to 0$ limit.

3 Nonabelian gauge theory

3.1 Constraints from unitarity and Lorentz invariance

Recall that in QED, or any quantum field theory with a massless vector boson A_μ, the polarization vector $\epsilon_\mu(p, h)$ does not transform as a vector under Lorentz transformation. For instance, take

$$p^\mu = (E, 0, 0, E), \qquad \epsilon_\mu(p, \pm) = \frac{1}{\sqrt{2}}(0, 1, \pm i, 0). \tag{3.1}$$

We may choose $\epsilon_0(p,\pm)$ to be purely spatial. Then clearly $\epsilon_\mu(p,\pm)$ cannot transform covariantly under Lorentz boosts in x, y directions. Instead, under a general Lonretz transformation Λ, we have

$$\epsilon_\mu(p,\pm) \to e^{\pm i\theta(p,\Lambda)}\left[(\Lambda\epsilon)_\mu + \alpha_\pm(p,\Lambda)(\Lambda p)_\mu\right], \tag{3.2}$$

Demanding the theory to be Lorentz invariant then forces us to introduce the $U(1)$ gauge symmetry/redundancy under which ϵ_μ and $\epsilon_\mu + \alpha p_\mu$ are identified. In particular, in any (on-shell) scattering amplitude, when the polarization of an external photon ϵ_μ is replaced by its momentum vector p_μ, the amplitude must vanish. We have seen that the amplitude with a soft photon emission (of polarization ϵ and momentum $q \to 0$) takes the form

$$M_{\beta\alpha} \sum_n \frac{e_n p_n \cdot \epsilon^*}{\eta_n p_n \cdot q - i\epsilon} \tag{3.3}$$

Setting $\epsilon = q$, the vanishing of the amplitude amounts to

$$\sum_n \eta_n e_n = 0, \tag{3.4}$$

which is the conservation of the total charge with respect to the photon.

Now we would like to construct a unitary, Lorentz invariant theory of several different species of massless vector bosons. We will label the type of the vector field by the indices $a, b, c, \cdots$, and label the type of charged "matter" field by $i, j, k, \cdots$ (all of which are assumed to have the same mass, for simplicity). Let us assume minimal coupling, but now generally allowing the matter field to change type upon emission or absorption of a vector field. For instance, with Dirac fermions ψ_i, we have couplings of the form

$$i(T^a)_{ij} A_{a\mu} \bar{\psi}^i \gamma^\mu \psi^j. \tag{3.5}$$

Consider the example of Compton scattering with an ingoing vector boson of type a, momentum q_1, polarization ϵ_1, and an outgoing vector boson of type b, momentum q_2 and polarization ϵ_2. The ingoing matter field is taken to be type i and outgoing type j. The amplitude in the soft limit $q_i \to 0$ takes the form

$$-\sum_k (T^b)_{jk}(T^a)_{ki} \frac{(p\cdot\epsilon_2^*)(p\cdot\epsilon_1)}{p\cdot q_2 - i\epsilon} - \sum_k (T^b)_{ik}(T^a)_{kj} \frac{(p\cdot\epsilon_1)(p\cdot\epsilon_2^*)}{-p\cdot q_1 - i\epsilon}. \tag{3.6}$$

Here k labels the type of the intermediate matter field, and is summed over all matter fields. Note that the on-shell condition implies $(p+q_1-q_2)^2 = p^2$, and hence $p\cdot q_1 \approx p\cdot q_2$ in the soft limit. If either ϵ_i is taken to be q_i, say $\epsilon_2^\mu = q_2^\mu$, then the soft amplitude reduces to

$$-\left[T^b, T^a\right]_{ji} p\cdot\epsilon_1. \tag{3.7}$$

Lorentz invariance requires this amplitude to be canceled by the contribution from additional diagrams. This is possible if we introduce a self coupling of the vector bosons, of the form

$$-f^{abc}A_a^\mu A_b^\nu \partial_\mu A_{c\nu}. \tag{3.8}$$

We assume that f^{abc} is completely antisymmetric in a, b, c. The corresponding cubic vertex of three vector fields, labelled by their type, Lorentz index, and momentum, (a, μ, k_1), (b, ν, k_2), and (c, ρ, k_3), is then

$$f^{abc}\left(\eta^{\mu\nu}k_{12}^\rho + \eta^{\nu\rho}k_{23}^\mu + \eta^{\rho\mu}k_{31}^\nu\right), \tag{3.9}$$

where $k_{ij} \equiv k_i - k_j$. There is now an additional diagram with the exchange of a gauge boson of type c that contributes to Compton scattering, giving in the soft limit

$$\begin{aligned} & if^{bac}(T^c)_{ji}\frac{(\epsilon_1\cdot\epsilon_2^*)(p\cdot(q_1+q_2)) - ((2q_1-q_2)\cdot\epsilon_2^*)(p\cdot\epsilon_1) - ((2q_2-q_1)\cdot\epsilon_1)(p\cdot\epsilon_2^*)}{-2q_1\cdot q_2 - i\epsilon} \\ &= if^{bac}(T^c)_{ji}\frac{\frac{1}{2}(\epsilon_1\cdot\epsilon_2^*)(p\cdot(q_1+q_2)) - (q_1\cdot\epsilon_2^*)(p\cdot\epsilon_1) - (q_2\cdot\epsilon_1)(p\cdot\epsilon_2^*)}{-q_1\cdot q_2 - i\epsilon} \end{aligned} \tag{3.10}$$

Now taking $\epsilon_2^\mu = q_2^\mu$ and using $p\cdot q_1 \approx p\cdot q_2$, we obtain

$$if^{bac}(T^c)_{ji}p\cdot\epsilon_1, \tag{3.11}$$

which cancels (3.7) if the T^a's obey

$$\left[T^a, T^b\right] = if^{abc}T^c. \tag{3.12}$$

In other words, the matrices T^a form a closed algebra under the Lie bracket (commutator), i.e. they generate a Lie algebra. The amplitudes must now be invariant under a bigger set of gauge equivalences, namely for each vector boson external line of type a and momentum p^μ, the amplitude should be invariant under

$$\epsilon_\mu^a \to \epsilon_\mu^a + \zeta^a p_\mu. \tag{3.13}$$

This will turn out to be the *linearized* gauge transformation in a theory of multiple types of massless vector bosons.

3.2 Simple Lie groups and nonabelian gauge invariance

We have seen in the previous subsection that Lorentz invariance of scattering amplitudes requires the Lagrangian of a set of interacting massless vector boson to take the form

$$\mathcal{L} = -\frac{1}{4}(\partial_\mu A_{a\nu} - \partial_\nu A_{a\mu})(\partial^\mu A_a^\nu - \partial^\nu A_a^\mu) - \frac{1}{2}f^{abc}A_\mu^a A_\nu^b(\partial^\mu A_a^\nu - \partial^\nu A_a^\mu) + \cdots \tag{3.14}$$

where $\cdots$ stands for higher order terms in $A_{a\mu}$. We will see momentarily that it suffices to stop at quartic order. We also saw that this Lagrangian should be gauge invariant under

$$\delta A_{a\mu} = \partial_\mu \zeta_a + \cdots \tag{3.15}$$

The simplest completion of these at the nonlinear level is such that the full gauge transformation takes the form

$$\delta A_{a\mu} = \partial_\mu \zeta_a + f^{abc} A_{b\mu} \zeta_c. \tag{3.16}$$

The gauge field Lagrangian can then be constructed out of the *nonabelian* field strength

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc} A_{b\mu} A_{c\nu}. \tag{3.17}$$

Under an infinitesimal gauge variation, $F_{\mu\nu}$ transforms by

$$\begin{aligned}\delta F^a_{\mu\nu} &= \partial_\mu \delta A^a_\nu - \partial_\nu \delta A^a_\mu + f^{abc} \delta A_{b\mu} A_{c\nu} + f^{abc} A_{b\mu} \delta A_{c\nu} \\ &= \partial_\mu (f^{abc} A_{b\nu} \zeta_c) - \partial_\nu (f^{abc} A_{b\mu} \zeta_c) + f^{abc} (\partial_\mu \zeta_b + f^{bde} A_{d\mu} \zeta_e) A_{c\nu} + f^{abc} A_{b\mu} (\partial_\nu \zeta_c + f^{cde} A_{d\nu} \zeta_e) \\ &= f^{abc} F_{b\mu\nu} \zeta_c.\end{aligned} \tag{3.18}$$

In the last step, we used the Jacobi identity

$$f^{abd} f^{bce} + f^{acb} f^{bde} = f^{abe} f^{bcd}, \tag{3.19}$$

which follows from (3.12). The gauge invariant Lagrangian that involves up to two-derivative terms takes the form

$$\mathcal{L} = -\frac{1}{4g^2} F_{a\mu\nu} F^{a\mu\nu}. \tag{3.20}$$

This is known as the Yang-Mills Lagrangian. The parameter g we have introduced here is a coupling constant. We could of course redefine the gauge fields $A_{a\mu}$ by

$$A_{a\mu} = g A'_{a\mu}, \tag{3.21}$$

so that the kinetic term for $A'_{a\mu}$ is canonically normalized; then a factor of g appears in front of the $A'^2 \partial A'$ cubic coupling and a factor of g^2 appears in front of the quartic term $(A')^4$. It is useful to represent the entire set of gauge fields into a single Lie algebra valued vector field,

$$A_\mu = \sum_a A_{a\mu} T^a, \tag{3.22}$$

where T^a are the matrices with entry $(T^a)_{ij}$, or equivalently thought of abstractly as generators of some Lie algebra $\mathbf{g}$. With the trace convention $\mathrm{Tr}(T^a T^b) = \delta^{ab}$, we may write the nonabelian field strength and Lagrangian as

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - i[A_\mu, A_\nu], \tag{3.23}$$

and

$$\mathcal{L} = -\frac{1}{4g^2}\mathrm{Tr}F_{\mu\nu}F^{\mu\nu}. \tag{3.24}$$

It is also convenient to write the gauge field as a 1-form,

$$A \equiv A_\mu dx^\mu, \tag{3.25}$$

and the field strength 2-form is given by

$$F \equiv \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu = dA - iA \wedge A. \tag{3.26}$$

The gauge transformations on the forms are

$$\begin{aligned}\delta A &= d\zeta - i[A,\zeta],\\ \delta F &= d(d\zeta - i[A,\zeta]) - i(d\zeta - i[A,\zeta]) \wedge A - iA \wedge (d\zeta - i[A,\zeta])\\ &= -i[dA,\zeta] + \zeta A \wedge A - A \wedge A\zeta = -i[F,\zeta].\end{aligned} \tag{3.27}$$

There is another possible Lorentz invariant term involving up to two derives on the gauge fields,

$$-\frac{1}{2}\theta\epsilon^{\mu\nu\rho\sigma}F_{a\mu\nu}F_{a\rho\sigma}, \tag{3.28}$$

where θ is a constant. This term is parity odd, and is in fact (locally) a total derivative. It does not affect the equation of motion. However, the integral of this term may evaluate to a nonzero number on nontrivial global configurations of the gauge fields. We will return to this point later. For now, we will set θ to zero.

The commutation relations of the generators T^a,

$$[T^a, T^b] = if^{abc}T^c, \tag{3.29}$$

defines a Lie algebra $\mathbf{g}$. To description the gauge fields themselves, it suffices to specify f^{abc} only, together with a notion of trace $\mathrm{Tr}(T^aT^b)$ (which we have normalized to δ^{ab}), without referring to any particular set of matrices. The index a ranges takes value $a = 1, 2, \cdots, d = \dim \mathbf{g}$. f^{abc} are subject to the Jacobi identity

$$[[T^a, T^b], T^c] + [[T^b, T^c], T^a] + [[T^c, T^a], T^b] = 0, \tag{3.30}$$

or equivalently,

$$f^{abd}f^{dce} + f^{bcd}f^{dae} + f^{cad}f^{dbe} = 0. \tag{3.31}$$

Here the index e is understood to be summed over (contracted). Note that with our trace normalization on the generators,

$$if^{bca} = \mathrm{Tr}(T^a[T^b, T^c]) = \mathrm{Tr}(T^b[T^c, T^a]) = if^{cab}, \tag{3.32}$$

and so f^{abc} is cyclically symmetric, hence completely anti-symmetric, as claimed earlier. Be cautious that if we are to change to an arbitrary basis of generators in which $\mathrm{Tr}(T^a T^b)$ is not diagonal, the structure constants would no longer be completely anti-symmetric.

To minimally couple n matter fields ψ_i to $A_{a\mu}$, we will need to consider a particular *representation* of $\mathbf{g}$, namely a linear map from $\mathbf{g}$ to the set of $n \times n$ matrices such that the commutation relation (3.29) is realized by the $n \times n$ matrices. The gauge fields themselves, $A_{a\mu}$, is then said to be in the *adjoint representation*, that is, the generators T^a act by $d \times d$ matrices

$$(T^a)^{\mathbf{ad}}_{bc} \equiv -i f_{bca}. \tag{3.33}$$

It is easy to verify that

$$[(T^a)^{\mathbf{ad}}, (T^b)^{\mathbf{ad}}] = i f^{abc} (T^c)^{\mathbf{ad}}, \tag{3.34}$$

which is equivalent to the Jacobi identity. Generally, for matter fields ψ_i in a representation R of $\mathbf{g}$, we may define the gauge covariant derivative

$$D_\mu \psi_j = \partial_\mu \psi_j - i A^a_\mu (T^a)_{jk} \psi_k. \tag{3.35}$$

With the assignment of gauge transformation

$$\delta \psi_i = i \zeta^a (T^a)_{ij} \psi_j = i(\zeta\psi)_i, \tag{3.36}$$

$D_\mu \psi$ then transforms covariantly, namely

$$\delta(D_\mu \psi) = D_\mu(i\zeta\psi) - i(D_\mu \zeta)\psi = i\zeta D_\mu \psi. \tag{3.37}$$

This property allows us to construct gauge invariant Lagrangian of matter fields easily using $D_\mu \psi$.

Let us discuss a few general features of the possible Lie algebras $\mathbf{g}$ and their representations. We have been a bit sloppy in introducing the trace $\mathrm{Tr}(T^a T^b)$, which must obey cyclicity, in claiming that it can be normalized to δ^{ab}. This is not always possible, but is in fact always possible whenever such gauge field kinetic term is positive definite for real fields A^a_μ. The existence of such a trace operation is equivalent to the statement that $\mathbf{g}$ is a direct sum of *compact*, *simple* Lie algebras, and abelian i.e. $U(1)$ subalgebras.

A Lie algebra $\mathbf{g}$ is called simple if there does not exist any nontrivial subalgebra $\mathbf{h}$ such that $[\mathbf{g}, \mathbf{h}] \subset \mathbf{h}$ (such an $\mathbf{h}$ is called an invariant subalgebra, or ideal). A Lie algebra can be exponentiated to give a group manifold, namely

$$g = e^{i\zeta^a T^a}, \tag{3.38}$$

either regarded as a formal power series in T^a, or as the actual exponential of the matrix constructed out of $(T^a)_{ij}$ in some representation of $\mathbf{g}$, comprise all elements of a group G (called Lie group). The Lie algebra $\mathbf{g}$ is called compact if the corresponding group G is a compact manifold.

Some basic examples of simple Lie groups are $SU(N)$ and $SO(N)$, the special unitary group and the special orthogonal group. The corresponding Lie algebra $su(N)$ or $so(N)$ may be realized as the space of Hermitian traceless matrices or real anti-symmetric matrices, with the Lie bracket defined as the standard matrix commutator. Note that unitary group $U(N) \simeq (SU(N) \times U(1))/\mathbb{Z}_N$ is not a simple Lie group. Another family of simple Lie groups are the symplectic group $Sp(N)$.[4] All simple Lie groups are classified. At the level of Lie algebra, the possible simple compact Lie algebras are $su(N)$, $so(N)$, $sp(N)$, and those of the exceptional Lie groups E_6, E_7, E_8, F_4, G_2. We will not discuss these exceptional cases for now. The gauge algebra $\mathbf{g}$ of a general nonabelian gauge theory (which has a positive definite kinetic term) must then be a direct sum of these simple Lie algebras and $u(1)$ algebras.

So far we have described the infinitesimal gauge transformations of the non-Abelian gauge fields $A_{a\mu}$ and matter fields ψ_i. By composing infinitesimal gauge transformations, one obtains the finite form of the gauge transformation, parameterized by a Lie *group* valued gauge function

$$g(x) = e^{i\zeta^a(x)T^a}. \tag{3.40}$$

Under the transformation by $g(x)$,

$$\begin{aligned} A_\mu(x) \to A^g_\mu(x) &= ig(x)\partial_\mu g(x)^{-1} + g(x)A_\mu(x)g(x)^{-1} \\ &= ig(x)(\partial_\mu - iA_\mu(x))g(x)^{-1}, \\ \psi(x) \to \psi^g(x) &= (g(x))_R\psi(x), \end{aligned} \tag{3.41}$$

where $(g(x))_R$ is the matrix corresponding to the group element $g(x)$ in the representation R of the matter field ψ. Typically, the matter fields should transform in a *unitary*

[4]In the context of nonabelian gauge theories we mostly deal with the *compact* form of $Sp(N)$. It may be realized as the group of quaternionic unitary matrices, and its Lie algebra that of quaternionic skew-Hermitian matrices. Alternatively, it is also the group of $U(2N)$ matrices that preserve the skew-symmetric bilinear form

$$\Omega = \begin{pmatrix} 0 & -I_N \\ I_N & 0 \end{pmatrix}. \tag{3.39}$$

Namely, it is the group of $2N \times 2N$ unitary matrices A that obey $A^T\Omega A = \Omega$. For this reason it is also denoted $USp(2N)$.

$Sp(N) = USp(N)$ is not to be confused with the *split* form of the symplectic group, $Sp(2N, \mathbb{R})$, which is the group of $2N$ dimensional real linear transformations that preserve Ω, i.e. the group of real $2N \times 2N$ matrices A that obey $A^T\Omega A = \Omega$. This group is *not* compact. For example, in the case $N = 1$, we have $Sp(2, \mathbb{R}) \simeq SL(2, \mathbb{R})$, while $Sp(1) \simeq SU(2)$.

representation of the gauge group G, i.e. $(g(x))_R^\dagger (g(x))_R = (g(x))_R (g(x))_R^\dagger = \mathbb{I}$, so that terms such as $-\bar{\psi}\gamma^\mu D_\mu \psi$ in the Lagrangian is gauge invariant.

3.3 Canonical quantization

Having written down the classical Lagrangian of nonabelian gauge fields, we now consider the quantization of this theory. We shall begin with our most familiar approach, namely canonical quantization. This approach has the advantage of being manifestly unitarity, but the disadvantage of being not manifestly Lorentz invariant.

For simplicity, we will consider the pure Yang-Mills theory here; the generalization to include matter fields transforming in some representation of the gauge group will be straightforward. Naively one starts by treating the fields $A_{a\mu}$ as canonical variables and write the canonical momenta $\Pi_{a\mu} = \partial\mathcal{L}/\partial(\partial_0 A^{a\mu})$. One immediately sees that $\Pi_{a0} = 0$. This is a primary constraint. There is also a secondary (Gauss-law) constraint from the equation of motion with respect to A_{a0}. These are first class constraints, in that the Poisson bracket of Π_{a0} with the equation of motion obtained by varying A_{a0} vanish, and cannot be quantized using the Dirac bracket. To proceed, we need to fix a gauge. A convenient gauge for quantizing nonabelian gauge theory is the axial gauge:

$$A_{a3} = 0. \tag{3.42}$$

While A_{ai}, $i = 1, 2$ are independent fields, i.e. canonical variables, A_{a0} is not and is determined by its equation of motion which does not involve time derivatives on A_{a0}. Let us see this:

$$-\partial_\mu \frac{\partial\mathcal{L}}{\partial(\partial_\mu A_{a0})} + \frac{\partial\mathcal{L}}{\partial A_{a0}} = \partial_\mu {F_a}^{\mu 0} + f_{cab} A_{b\mu} {F_c}^{\mu 0} \tag{3.43}$$

Using

$$\Pi_{ai} = \frac{\partial\mathcal{L}}{\partial(\partial_0 A_{ai})} = -{F_a}^{0i}, \tag{3.44}$$

and

$${F_a}^{30} = \partial_3 A_a^0 \tag{3.45}$$

in axial gauge, we can rewrite the equation of motion for A_a^0 as

$$-\partial_3^2 A_a^0 = \partial_i \Pi_a^i + f_{cab} A_{bi} \Pi_c^i. \tag{3.46}$$

Here i is understood to be summed over $1, 2$. Modulo zero frequency modes in x^3 direction, we can solve A_a^0 in terms of A_{ai} and Π_a^i.

In terms of the canonical variables A_{ai} and canonical momenta Π_{ai}, we can now write the Hamiltonian density in axial gauge,

$$\begin{aligned}\mathcal{H} &= \Pi_{ai}\partial_0 A_{ai} - \mathcal{L} \\ &= \Pi_{ai}(F_{a0i} + \partial_i A_{a0} - f_{abc}A_{b0}A_{ci}) - \left(\frac{1}{2}F_{a0i}F_{a0i} + \frac{1}{2}F_{a03}F_{a03} - \frac{1}{2}F_{aij}F_{aij} - \frac{1}{2}F_{ai3}F_{ai3}\right) \\ &= \Pi_{ai}(\partial_i A_{a0} - f_{abc}A_{b0}A_{ci}) + \frac{1}{2}\Pi_{ai}\Pi_{ai} + \frac{1}{2}F_{aij}F_{aij} + \frac{1}{2}\partial_3 A_{ai}\partial_3 A_{ai} - \frac{1}{2}\partial_3 A_{a0}\partial_3 A_{a0}.\end{aligned} \tag{3.47}$$

Here A_{a0} is understood to be solved using (3.46). After solving for A_{a0} which is linear in the canonical momenta Π_{ai}, we see that the Hamiltonian is quadratic in Π_{ai}.

One may now proceed by formulating the functional integral by treating A_{ai} and Π_{ai} as independent variables,

$$\int \mathcal{D}A_{ai}\mathcal{D}\Pi_{ai} \exp\left[i\int d^4x \left(\Pi_{ai}\partial_0 A_{ai} - \mathcal{H}[A_{ai}, \Pi_{ai}]\right)\right] \tag{3.48}$$

While a priori our quantization procedure requires solving for A_{a0} from (3.46) in $\mathcal{H}$, we note that $\mathcal{H}$ is up to quadratic in A_{a0}, and if we pretend that A_{a0} is an independent variable, the integration over A_{a0} is a Gaussian functional integral which gives a constant, field independent functional determinant. The result amounts to setting A_{a0} to the stationary point value of $\mathcal{H}$ with respect to A_{a0}, and this is precisely the equation (3.46). In other words, the path integral can be equivalently written as

$$\int \mathcal{D}A_{a0}\mathcal{D}A_{ai}\mathcal{D}\Pi_{ai} \exp\left[i\int d^4x \left(\Pi_{ai}\partial_0 A_{ai} - \mathcal{H}[A_{a0}, A_{ai}, \Pi_{ai}]\right)\right]. \tag{3.49}$$

We can now integrate out Π_{ai} (which is now a Gaussian with trivial determinant) and recover the Lagrangian form of the path integral,

$$\int \mathcal{D}A_{a\mu}\, \delta(A_{a3}) e^{i\int d^4x \mathcal{L}}. \tag{3.50}$$

where we formally restored the integration over A_{a3} by enforcing the axial gauge via a delta functional. The Lagrangian $\mathcal{L}$ is now back to the standard one, namely the covariant Yang-Mills Lagrangian.

This now looks like a straightforward path integral formulation with the gauge condition enforced by a delta functional, but it is not manifestly Lorentz invariant (in this case it is not rotationally invariant). One may attempt to simply replace the axial gauge condition by a covariant gauge condition. As we will see, to do this consistently requires inserting an additional functional Jacobian factor into the path integral, which leads to Faddeev-Popov "ghosts". The price we will pay for preserving manifest Lorentz invariance is that unitarity will not be manifest in the covariant quantization. The latter will be dealt with using the formalism of BRST quantization.

3.4 Faddeev-Popov ansatz and ghosts

Let us denote by ϕ collectively all fields in a gauge theory. The naive, formal path integral

$$\int \mathcal{D}\phi \, e^{iS} \tag{3.51}$$

is ill defined due to the overcounting of gauge equivalent field configurations. Here we assume that the path integral measure $\mathcal{D}\phi$ together with e^{iS} is gauge invariant. Typically in non-anomalous gauge theories this is the case for the measure and the classical action separately. To make the path integral well defined we must introduce some sort of gauge fixing. For instance, let $f[\phi]$ be a gauge non-invariant functional of ϕ (which is itself a function of spacetime coordinates), and consider the gauge condition $f[\phi] = 0$. This may be introduced in the path integral by inserting a delta functional $\delta(f[\phi])$. To fix the gauge completely, we need the functional Jacobian of the gauge variation of $f[\phi]$ to be non-degenerate. To ensure that the resulting functional integral is gauge invariant, we need to insert the Jacobian factor as well. Let ζ be the (finite) gauge parameter, and ϕ^ζ be the gauge transformation of ϕ. The Jacobian factor is

$$\mathcal{J}[\phi, \zeta] = \det \frac{\delta f[\phi^\zeta]}{\delta \zeta}. \tag{3.52}$$

Formally, we may insert

$$1 = \int D\zeta \, \delta(f[\phi^\zeta]) \det \frac{\delta f[\phi^\zeta]}{\delta \zeta} \tag{3.53}$$

into (3.51), and write the path integral as

$$\begin{aligned}
&\int D\phi D\zeta \, \delta(f[\phi^\zeta]) \det \frac{\delta f[\phi^\zeta]}{\delta \zeta} \, e^{iS[\phi]} \\
&= \int D\zeta D\phi^{\zeta'} \, \delta(f[\phi^{\zeta\circ\zeta'}]) \det \frac{\delta f[\phi^{\zeta\circ\zeta'}]}{\delta \zeta} \, e^{iS[\phi^{\zeta'}]} \\
&= \int D\zeta D\phi^{\zeta'} \, \delta(f[\phi^{\zeta\circ\zeta'}]) \det \frac{\delta f[\phi^{\zeta\circ\zeta'}]}{\delta(\zeta\circ\zeta')} \det \frac{\delta(\zeta\circ\zeta')}{\delta\zeta} \, e^{iS[\phi^{\zeta'}]},
\end{aligned} \tag{3.54}$$

where in the second line we change the functional integration variable ϕ to the transformed field by an arbitrary gauge parameter ζ'. $\zeta \circ \zeta'$ is the composition of the two gauge transformations. Now choose ζ' to be the inverse transformation of ζ, and use the gauge invariance of the measure together with e^{iS}, we obtain

$$\int D\zeta D\phi \det \frac{\delta f[\phi^{\zeta''}]}{\delta \zeta''}\bigg|_{\zeta''=0} \left(\det \frac{\delta(\zeta''\circ\zeta)}{\delta\zeta''}\bigg|_{\zeta''=0} \right)^{-1} \delta(f[\phi]) \, e^{iS[\phi]} \tag{3.55}$$

Now the integration over ζ with the factor $\left(\det \frac{\delta(\zeta''\circ\zeta)}{\delta\zeta''}\Big|_{\zeta''=0}\right)^{-1}$ is formally independent of fields and can be dropped from the path integral. We are then left with

$$\int D\phi \det \frac{\delta f[\phi^\zeta]}{\delta\zeta}\Big|_{\zeta=0} \delta(f[\phi])\, e^{iS[\phi]}. \tag{3.56}$$

The Jacobian factor

$$\Delta_{FP}[\phi] = \det \frac{\delta f[\phi^\zeta]}{\delta\zeta}\Big|_{\zeta=0} \tag{3.57}$$

is known as the Faddeev-Popov determinant. Now the path integral is free from the overcounting of gauge equivalent configurations, and is independent of the choice of gauge fixing condition $f[\phi]$ by virtue of the above formal argument (this can also be demonstrated directly starting from (3.56)). Note that we could have replaced the delta functional by any non-degenerate functional (say a Gaussian) of $f[\phi]$, i.e. some $F[f[\phi]]$, and the same argument still goes through.

Note that we did not encounter Δ_{FP} in Abelian gauge theories such as QED because with the standard choices of gauge conditions that are linear in the gauge field, the Faddeev-Popov determinant is field independent. It is also field independent in the case of axial gauge for non-Abelian gauge theories, as described in the previous section, but will be a nontrivial functional with Lorentz covariant gauge fixing conditions in non-Abelian gauge theories.

A simple Lorentz covariant gauge choice for nonabelian gauge theory is

$$f_a[A] = \partial_\mu A^\mu_a, \tag{3.58}$$

and a Gaussian form of $F[f[\phi]]$,

$$F[f_a] = e^{-\frac{i}{2\xi}\int d^4x f_a f_a} = e^{-\frac{i}{2\xi}\int d^4x (\partial_\mu A^\mu_a)^2}. \tag{3.59}$$

The gauge fixed Lagrangian is

$$\mathcal{L} = \mathcal{L}_{YM} - \frac{1}{2\xi}(\partial_\mu A^\mu_a)^2. \tag{3.60}$$

The gauge fixing term is quadratic in the gauge fields and hence only modifies the kinetic term. The modified propagator is the same as in the Abelian gauge theory, given in momentum space by

$$G_{a\mu,b\nu}(p) = \frac{-i\delta_{ab}}{p^2 - i\epsilon}\left[\eta_{\mu\nu} + (\xi - 1)\frac{p_\mu p_\nu}{p^2}\right]. \tag{3.61}$$

Due to the gauge transformation

$$\delta A^\mu_a = \partial^\mu \zeta_a + f_{abc} A^\mu_b \zeta_c, \tag{3.62}$$

the Faddeev-Popov determinant is

$$\Delta_{FP} = \det(\partial_\mu D^\mu_A), \tag{3.63}$$

where $D^\mu_A = \partial^\mu - i[A^\mu, \cdot]$ is the covariant derivative with respect to the gauge field A_μ.

The Faddeev-Popov determinant can be rewritten in a more familiar way as a fermionic (Grassmannian) Gaussian functional integral,

$$\Delta_{FP} = \det \frac{\delta f^\alpha[\phi^\zeta]}{\delta \zeta^\beta} = \int \prod_\alpha d\bar{\eta}_\alpha \prod_\beta d\eta^\beta \exp\left(i\bar{\eta}_\alpha \left. \frac{\delta f^\alpha[\phi^\zeta]}{\delta \zeta^\beta} \right|_{\zeta=0} \eta^\beta \right). \tag{3.64}$$

Here α, β are indices labeling the gauge fixing condition and the gauge parameter at all spacetime points (appropriately discretized) or in momenta space, all momenta. The continuum version of the Grassmannian integral on the RHS is

$$\int D\bar{\eta} D\eta \, e^{iS_{gh}[\phi,\eta,\bar{\eta}]}, \tag{3.65}$$

where the "ghost" action S_{gh} is obtained by linearizing $f[\phi^\zeta]$ with respect to the gauge parameter ζ, replace the infinitesimal gauge parameter ζ_a by η_a, and then multiply with $\bar{\eta}$ (which carries the same indices as $f[\phi]$) and integrate over the spacetime. η and $\bar{\eta}$ are Grassmannian scalar fields, and are referred to as Faddeev-Popov ghosts. They do not appear in asymptotic states, and their role is entirely played through coupling to the gauge fields (and matter fields if they enter the gauge fixing condition), and enter loop diagrams.

In particular, the ghost action associated with the Lorentz covariant gauge fixing function $\partial_\mu A^\mu_a$ is

$$\begin{aligned} S_{gh} &= \int d^4x \, \bar{\eta}_a \partial_\mu (\partial^\mu \eta_a + f_{abc} A^\mu_b \eta_c) \\ &= -\int d^4x \, (\partial_\mu \bar{\eta}_a \partial^\mu \eta_a + f_{abc} \partial_\mu \bar{\eta}_a A^\mu_b \eta_c) \, . \end{aligned} \tag{3.66}$$

We see that $\eta, \bar{\eta}$ has the same propagator as massless scalar fields (up to the overall sign which is a matter of convention), but couple to $A_{a\mu}$ through a cubic vertex only (unlike the minimally coupled scalar). Furthermore, since $\eta, \bar{\eta}$ are Grassmannian ghost fields, each ghost loop comes with a minus sign, as in the Feynman rules for fermions.

In our earlier discussion of unitarity, it was important that all particles appearing in loops can also be in the asymptotic states, in order for the optical theorem to hold, for instance. The appearance of the ghosts in the loops but not external states may seem to conflict with unitarity. However, Feynman rules in a Lorentz covariant gauge also instructs us to sum over intermediate states created by all components of $A_{a\mu}$, while $A_{a\mu}$ has only two physical degrees of freedom. The role of the ghosts is precisely to cancel the unphysical components of the gauge fields and restore unitarity in scattering amplitudes.

3.5 BRST symmetry

The quantum gauge theory is properly defined using the gauge fixing procedure and the introducing of Faddeev-Popov ghosts. Gauge invariance, however, now becomes obscure. Preserving gauge invariance at the quantum level is crucial to the consistency (unitarity and Lorentz invariance) of a gauge theory. Further, the well defined correlation functions are those of gauge invariant operators. Why we need to restrict ourselves to gauge invariant operators is not entirely obvious in the gauge fixed form of the path integral. These questions are addressed by the the BRST symmetry. Firstly, one realizes that the gauged fixed action, with FP ghosts, though no longer has the local gauge symmetry, has a new *global fermionic* nilpotent symmetry, called the BRST symmetry. The correlation functions or scattering amplitudes will be independent of the choice of the gauge fixing condition if the operators or asymptotic states are invariant under the BRST symmetry.

The gauge fixed path integral can be written in the form

$$\int D\phi D\eta D\bar{\eta} DB_\alpha e^{i(S_0[\phi]+S_{GF}[\phi,B_\alpha]+S_{gh}[\phi,\eta,\bar{\eta}])} \tag{3.67}$$

where we introduced a new field B_α to rewrite the gauge fixing term in the form

$$S_{GF} = \widetilde{F}[B_\alpha] + \int d^4x B_\alpha f^\alpha[\phi]. \tag{3.68}$$

For example, the gauge condition $f^\alpha[\phi] = 0$ would be imposed if $\widetilde{F}[B]$ is taken to be zero, in which case B_α becomes a Lagrangian multiplier field. Recall that the ghost action is

$$S_{gh} = \int d^4x \bar{\eta}_\alpha \eta^a \delta_a f^\alpha[\phi]. \tag{3.69}$$

where δ_a is the gauge variation with respect to ζ^a, at $\zeta = 0$.

The gauged fixed action is invariant under a symmetry variation δ_B (BRST transformation),

$$\begin{aligned} \delta_B\phi &= \epsilon\eta^a\delta_a\phi, \\ \delta_B\bar{\eta}_\alpha &= -\epsilon B_\alpha, \\ \delta_B\eta^a &= -\frac{1}{2}\epsilon\, f^a{}_{bc}\eta^b\eta^c, \\ \delta_B B_\alpha &= 0. \end{aligned} \tag{3.70}$$

Here ϵ is a constant *Grassmannian* parameter. $\delta_a\phi$ is the gauge variation of ϕ with respect to ζ^a. $f^a{}_{bc}$ is generally the structure constant of gauge transformations, namely

$$[\delta_b, \delta_c] = f^a{}_{bc}\delta_a. \tag{3.71}$$

In the case of Yang-Mills theory this coincides with the structure constant of the gauge group. A priori, in the most general gauge theory, it is possible for the structure "constants" $f^a{}_{bc}$ to be field dependent. The BRST quantization we describe for now applies only to the case where $f^a{}_{bc}$ are field independent; this suffices for nonabelian gauge theories.

Clearly, $S_0[\phi]$ is invariant under δ_B since $\delta_B\phi$ is proportional to the gauge variation of ϕ, and so is the term $\widetilde{F}[B_a]$ in S_{GF}. Instead of directly verifying that $S_{GF}+S_{gh}$ is BRST invariant, let us express it in the form

$$\epsilon(S_{GF}+S_{gh}-\widetilde{F}[B_a]) = -\delta_B\int d^4x\bar{\eta}_\alpha f^\alpha[\phi]. \tag{3.72}$$

Note that ϵ anti-commutes with $\eta,\bar{\eta}$, which is necessary in verify this. Let us write

$$\delta_B = i\epsilon Q_B, \tag{3.73}$$

where Q_B is a fermionic operation, which upon canonical quantization becomes the "BRST charge". The BRST symmetry is nilpotent, in the sense that $Q_B^2=0$.

Let us verify this. We have

$$\begin{aligned} Q_B^2\phi &= Q_B(-i\eta^a\delta_a\phi) \\ &= \left(\frac{1}{2}f^a{}_{bc}\eta^b\eta^c\right)\delta_a\phi + \eta^a\eta^b\delta_b\delta_a\phi \\ &= \frac{1}{2}f^a{}_{bc}\eta^b\eta^c\delta_a\phi - \frac{1}{2}\eta^a\eta^b[\delta_a,\delta_b]\phi = 0, \end{aligned} \tag{3.74}$$

and

$$\begin{aligned} Q_B^2\eta^a &= \frac{i}{2}f^a{}_{bc}((Q_B\eta^b)\eta^c - \eta^b Q_B\eta^c) \\ &= \frac{1}{2}f^a{}_{bc}\left(-\frac{1}{2}f^b{}_{de}\eta^d\eta^e\eta^c + \frac{1}{2}f^c{}_{de}\eta^b\eta^d\eta^e\right) \\ &= \frac{1}{2}f^a{}_{b[c}f^b{}_{de]}\eta^c\eta^d\eta^e = 0, \end{aligned} \tag{3.75}$$

where we used Jacobi identity in the last step. It then follows from $Q_B^2=0$ that $S_{GF}+S_{gh}$ is δ_B-invariant, thus verifying that Q_B is a symmetry of the total gauged fixed action.

A key consistency requirement of the gauge fixing procedure is that the physical amplitudes must be independent of the choice of gauge condition $f^\alpha[\phi]=0$. Under a variation of the gauge condition $f^\alpha\to f^\alpha+\delta' f^\alpha$, the transition amplitude between two physical states $\langle\alpha|\beta\rangle$ should be invariant. In other words,

$$\delta'\langle\alpha|\beta\rangle = i\langle\alpha|\delta'(S_{GF}+S_{gh})|\beta\rangle = \langle\alpha|\left[Q_B,\int d^4x\bar{\eta}_\alpha\delta' f^\alpha[\phi]\right]|\beta\rangle = 0. \tag{3.76}$$

In writing the above $S_{GF} + S_{gh}$ is understood to be the operator corresponding to inserting the gauge dependent part of the action into the path integral, and the BRST variation of $\int d^4x \bar{\eta}\delta' f[\phi]$ is written in the operator language as its commutator with the BRST charge Q_B. For the above variation to vanish for arbitrary $\delta' f[\phi]$, we need

$$\langle \alpha | Q_B = Q_B | \beta \rangle = 0. \tag{3.77}$$

Note that Q_B is Hermitian, $Q_B = Q_B^\dagger$. So physical states must be annihilated by the BRST charge, or "Q_B-closed". On the other hand, states of the form $Q_B|\chi\rangle$ ("Q_B-exact") are orthogonal to all physical states by the nilpotent property $Q_B^2 = 0$. Physical states that differ by a Q_B-exact state should be identified. The inequivalent physical states are therefore given by Q_B-*cohomology*, namely

$$\mathcal{H}_{phys} = \frac{\mathrm{Ker}\, Q_B}{\mathrm{Im}\, Q_B}. \tag{3.78}$$

Let us see how this works explicitly in the Lorentz covariant gauge with $S_{GF} = -\frac{1}{2\xi}\int(\partial_\mu A_a^\mu)^2$. In analyzing asymptotic states we can ignore interactions (this assumes that the particles are weakly interacting at long distances, which as we will see is not true in the pure Yang-Mills theory; we will ignore this issue for now), and the problem essentially reduces to that of an abelian gauge theory. The BRST transformation of the fields in an abelian gauge theory are

$$\begin{aligned} &\delta_B A_\mu = \epsilon \partial_\mu \eta, \\ &\delta_B \bar{\eta} = -\epsilon B = \frac{1}{\xi}\epsilon \partial_\mu A^\mu, \\ &\delta_B \eta = 0, \end{aligned} \tag{3.79}$$

where we have used the equation of motion for B. Let $a^\mu(p)$ be the annihilation operators in the mode expansion of A_μ, $b(p)$ the annihilation operators of $\bar{\eta}$, and $c(p)$ the annihilation operators of η. The BRST charge has the commutation relations

$$[Q_B, a_\mu(p)] = p_\mu c(p), \qquad \{Q_B, b(p)\} = \frac{1}{\xi} p_\mu a^\mu(p), \qquad \{Q_B, c(p)\} = 0, \tag{3.80}$$

and similarly with the creation operators. Suppose $|\psi\rangle$ is a physical state, i.e. it is annihilated by Q_B. The state with an additional photon, $e^\mu a_\mu^\dagger(p)|\psi\rangle$, is Q_B-closed if $e \cdot p = 0$, i.e. the polarization must be transverse. Further, $Q_B b^\dagger(p)|\psi\rangle = \frac{1}{\xi} p^\mu a_\mu^\dagger(p)|\psi\rangle$ is BRST exact, and hence we identify $e_\mu \sim e_\mu + \alpha p^\mu$. Now we have recovered the physical states of a single photon from the BRST cohomology. Note that the ghosts states $b^\dagger(p)|\psi\rangle$ and $c^\dagger(p)|\psi\rangle$ have negative norm, but the former is not Q_B-closed and the latter is Q_B-exact, and neither contributes to the BRST cohomology. One can prove that the norm is positive definite on the BRST cohomology, hence unitarity is maintained.

4 Scattering amplitudes in spinor helicity formalism

4.1 Color ordering

Let us examine scattering amplitudes of gluons in Yang-Mills theory, at tree level. The gluon in and out states are labelled by their momentum p^μ, polarization vector ε^μ, and color index a. We will consider $SU(N)$ gauge theory, and so $a = 1, 2, \cdots, N^2 - 1$. Recall that our gauge generators T^a are normalized so that

$$\mathrm{Tr}(T^a T^b) = \delta^{ab}, \quad [T^a, T^b] = i f^{abc} T^c, \quad \mathrm{Tr}([T^a, T^b]T^c) = i f^{abc}. \tag{4.1}$$

A general tree amplitude is computed by the sum of a number of Feynman diagrams, that involves various contractions of structure constants f^{abc} associated with each cubic vertex and $f^{abe} f^{cde}$ associated with each quartic vertex. These contractions a priori may look complicated. However, note the following simplification. Suppose a is the color index of an internal propagator. Then we will have color index contraction of the form $f^{abc} f^{acd} \cdots$. Generally, these may be expressed as

$$\mathrm{Tr}(\cdots T^a \cdots)\mathrm{Tr}(\cdots T^a \cdots) = \mathrm{Tr}(T^a M_1)\mathrm{Tr}(T^a M_2), \tag{4.2}$$

where the a index is summed over. Now one can make use of the identity

$$(T^a)_i{}^j (T^a)_k{}^\ell = \delta_i^\ell \delta_k^j - \frac{1}{N}\delta_i^j \delta_k^\ell, \tag{4.3}$$

and write (4.2) as

$$(M_1)_j{}^i (M_2)_\ell{}^k \left(\delta_i^\ell \delta_k^j - \frac{1}{N}\delta_i^j \delta_k^\ell\right) = \mathrm{Tr}(M_1 M_2) - \frac{1}{N}\mathrm{Tr}(M_1)\mathrm{Tr}(M_2). \tag{4.4}$$

For our purpose a slight further simplification can be made if we replace the $SU(N)$ gauge group by $U(N)$. The $U(N)$ contains a diagonal $U(1)$ that commutes with the $SU(N)$ generators. Correspondingly, the "photon" that corresponds to the $U(1)$ decouples from the gluons in the $SU(N)$, and thus the scattering amplitudes of gluons in the $U(N)$ pure Yang-Mills theory is identical to that of $SU(N)$ pure Yang-Mills theory. Now the $U(N)$ generators can be labeled by T^a with $a = 1, 2, \cdots, N^2$, that obey

$$(T^a)_i{}^j (T^a)_k{}^\ell = \delta_i^\ell \delta_k^j. \tag{4.5}$$

The $U(1)$ generator is $(T^{U(1)})_i{}^j = \frac{1}{\sqrt{N}}\delta_i^j$. Now we have

$$\sum_{a=1}^{N^2} \mathrm{Tr}(T^a M_1)\mathrm{Tr}(T^a M_2) = \mathrm{Tr}(M_1 M_2). \tag{4.6}$$

Now if we write all the color factors from the vertices using

$$f^{abc} = -i\left[\mathrm{Tr}(T^aT^bT^c) - \mathrm{Tr}(T^aT^cT^b)\right], \tag{4.7}$$

we can try to eliminate the summation over color indices in internal propagators using (4.6). This works as long as the pair of color indices on an internal line always appear in two separate traces, which is always the case for tree level diagrams, but not for loop diagrams. We will only discuss tree level diagrams for now.

By the above procedure, we can always reduce the summation over product of the structure constants into the form that is a sum over single-trace expressions, like

$$\mathrm{Tr}(T^{a_1}T^{a_2}\cdots T^{a_n}), \tag{4.8}$$

where $a_1, a_2, \cdots, a_n$ are the color indices of the n external gluons, in all possible ordering. In other words, the general tree level scattering amplitudes of n gluons labeled by color, momentum, and helicity

$$a_i,\ p_i^\mu,\ \lambda_i,\ i = 1, 2, \cdots, n, \tag{4.9}$$

always takes the form

$$\mathcal{A}^{tree}(a_i, p_i, \lambda_i) = \frac{g^{n-2}}{n}\sum_{\sigma\in S_n}\mathrm{Tr}(T^{a_{\sigma(1)}}\cdots T^{a_{\sigma(n)}})\mathcal{A}^{co}(\sigma(1), \sigma(2), \cdots, \sigma(n)). \tag{4.10}$$

Here S_n is the permutation group on n elements, $\{1, 2, \cdots, n\}$. σ is a permutation that takes i to $\sigma(i)$, $i = 1, 2, \cdots, n$. We included a factor $1/n$ to compensate for the fact that $\mathrm{Tr}(T^{a_{\sigma(1)}}\cdots T^{a_{\sigma(n)}})$ is invariant under cyclic permutation on the $T^{a_{\sigma(i)}}$'s. We have also separated out the factors of Yang-Mills coupling g.

$\mathcal{A}^{co}(1, 2, \cdots, n)$ is called the (tree-level) *color-ordered* amplitude. It depends on the momenta p_i and helicities $\lambda_i = \pm 1$; the color factor has been stripped off. Diagrammatically, what this means is that $\mathcal{A}^{co}$ is computed by summing over all *planar* diagrams with given cyclic ordering of external gluon lines $1, 2, \cdots, n$, and with each cubic vertex, instead of assigning the structure constant f^{abc}, we replace it by simply $-i$. This prescription can be summarized by the color-ordered Feynman rules, as follows. The cubic vertex of three gluons with momentum and vector index (k_1, μ), (k_2, ν), (k_3, ν) is now

$$-i(\eta^{\mu\nu}k_{12}^\rho + \eta^{\nu\rho}k_{23}^\mu + \eta^{\rho\mu}k_{31}^\nu), \tag{4.11}$$

and the color-ordered quartic vertex for four gluons with vector indices μ, ν, ρ, σ cyclically is given by

$$i\left(2\eta_{\mu\rho}\eta_{\nu\sigma} - \eta_{\mu\nu}\eta_{\rho\sigma} - \eta_{\mu\sigma}\eta_{\nu\rho}\right). \tag{4.12}$$

Now let us consider the simplest nontrivial example: the tree-level scattering of four gluons. We will compute the color ordered amplitude, $\mathcal{A}^{co}(1,2,3,4)$, for the four gluons with momenta p_i^μ and polarization vector ε_i^μ, $i = 1,2,3,4$. There are three diagrams that contribute: the s-channel, u-channel, and the diagram involving a quartic vertex. Note that the t-channel diagram is absent since we are restricting to planar diagrams.

We will work in the $\xi = 1$ Feynman gauge, so that the gluon propagator is simply

$$\frac{-i\eta_{\mu\nu}}{p^2 - i\epsilon}. \tag{4.13}$$

The s-channel diagram contributes to the amplitude

$$\begin{aligned} &i\varepsilon_1^\mu \varepsilon_2^\nu \varepsilon_3^\rho \varepsilon_4^\sigma \left[\eta_{\mu\nu}(k_1 - k_2)_\alpha + \eta_{\nu\alpha}(k_1 + 2k_2)_\mu + \eta_{\alpha\mu}(-2k_1 - k_2)_\nu\right] \\ &\times \frac{1}{(k_1+k_2)^2}\left[\eta_{\rho\sigma}(k_3 - k_4)^\alpha + \delta^\alpha_\sigma(k_3 + 2k_4)_\rho + \delta^\alpha_\rho(-2k_3 - k_4)_\sigma\right]. \end{aligned} \tag{4.14}$$

The u-channel diagram gives the same amplitude with a cyclic permutation $(1,2,3,4) \to (2,3,4,1)$. The diagram with quartic vertex gives

$$i\varepsilon_1^\mu \varepsilon_2^\nu \varepsilon_3^\rho \varepsilon_4^\sigma \left(2\eta_{\mu\rho}\eta_{\nu\sigma} - \eta_{\mu\nu}\eta_{\rho\sigma} - \eta_{\mu\sigma}\eta_{\nu\rho}\right) = i(2\varepsilon_1\cdot\varepsilon_3\varepsilon_2\cdot\varepsilon_4 - \varepsilon_1\cdot\varepsilon_2\varepsilon_3\cdot\varepsilon_4 - \varepsilon_1\cdot\varepsilon_4\varepsilon_2\cdot\varepsilon_3). \tag{4.15}$$

The color-ordered tree level amplitude is the sum of these three contributions. It looks a bit messy. The complexity of the computation of tree-level n-gluon amplitude by summing over Feynman diagrams increases quickly with n: even though we only need to sum over color-ordered diagrams, the number of such diagrams still grows like $n!$.

4.2 Spinor helicity formalism

We are now going to rewrite the gluon scattering amplitudes in terms of a different set of variables than the momenta and polarization vectors, in such a way that the on-shell condition of the external gluon states are implemented automatically. This will lead to dramatic simplification of the form of the gluon scattering amplitudes. Let us consider an *on-shell* gluon with momentum p^μ and polarization vector ε^μ. They obey

$$p^2 = p\cdot\varepsilon = 0. \tag{4.16}$$

Let $\sigma_0 = I_2$ be the 2×2 identity matrix, $\sigma_{1,2,3}$ the standard Pauli matrices. These will be denoted collectively as σ_μ. In other words, we have

$$\sigma_\mu = (I_2, \vec{\sigma}), \quad \sigma^\mu = (-I_2, \vec{\sigma}) \equiv \overline{\sigma}_\mu. \tag{4.17}$$

Consider the 2×2 matrix

$$p^\mu\sigma_\mu = \begin{pmatrix} p^0 + p^3 & p_1 - ip_2 \\ p_1 + ip_2 & p^0 - p^3 \end{pmatrix}. \tag{4.18}$$

We have for null p^μ,

$$\det(p^\mu \sigma_\mu) = -p^2 = 0. \tag{4.19}$$

We will write the components of the matrix σ^μ as $\sigma^\mu_{\alpha\dot\beta}$, where $\alpha = 1, 2$, and $\dot\beta = \dot 1, \dot 2$. What we learn here is that the matrix $p_{\alpha\dot\beta} \equiv p^\mu \sigma^\mu_{\alpha\dot\beta}$ is degenerate, and thus can be written as a column vector times a row vector,

$$p_{\alpha\dot\beta} = \lambda_\alpha \widetilde\lambda_{\dot\beta}. \tag{4.20}$$

For real p^μ, $p_{\alpha\dot\beta}$ is evidently Hermitian, and we can take $\widetilde\lambda_{\dot\alpha}$ to be the complex conjugate of λ_α. So we will write

$$p_{\alpha\dot\beta} = \lambda_\alpha \overline\lambda_{\dot\beta}. \tag{4.21}$$

Still, the decomposition (4.21) is not unique. λ and $\overline\lambda$ are defined up to a phase,

$$\lambda \to e^{i\phi}\lambda, \quad \overline\lambda \to e^{-i\phi}\overline\lambda. \tag{4.22}$$

Let us introduce some spinorial conventions. Note that

$$\overline\sigma_\mu = -\sigma_2 \sigma_\mu^T \sigma_2, \tag{4.23}$$

or in components,

$$(\overline\sigma_\mu)_{\dot\beta\alpha} = \epsilon^{\dot\beta\dot\gamma}(\sigma_\mu)_{\gamma\dot\gamma}\epsilon^{\gamma\alpha} = -(\sigma_\mu)^{\alpha\dot\beta}. \tag{4.24}$$

Here we have raised the spinor indices with ϵ tensor. Our convention for raising indices is

$$\epsilon^{\alpha\beta}u_\beta = u^\alpha, \quad \epsilon^{\dot\alpha\dot\beta}v_{\dot\beta} = v^{\dot\alpha}, \tag{4.25}$$

and for lowering indices,

$$u^\alpha \epsilon_{\alpha\beta} = u_\beta, \quad v^{\dot\alpha}\epsilon_{\dot\alpha\dot\beta} = v_{\dot\beta}. \tag{4.26}$$

These are consistent with the convention

$$\epsilon^{12} = \epsilon_{12} = \epsilon^{\dot 1\dot 2} = \epsilon_{\dot 1\dot 2} = 1. \tag{4.27}$$

Essentially, we always contract upper left with lower right indices. We have the formula

$$(\sigma_\mu)_{\alpha\dot\beta}(\sigma_\nu)^{\alpha\dot\beta} = -\mathrm{Tr}(\sigma_\mu\overline\sigma_\nu) = -2\eta_{\mu\nu}, \tag{4.28}$$

and

$$(\sigma^\mu)_{\alpha\dot\beta}(\sigma_\mu)_{\gamma\dot\delta} = -2\epsilon_{\alpha\gamma}\epsilon_{\dot\beta\dot\delta}. \tag{4.29}$$

Given any q_μ, k_ν, we have

$$q_{\alpha\dot\beta} = q_\mu \sigma^\mu_{\alpha\dot\beta}, \quad q^\mu = -\frac{1}{2}q_{\alpha\dot\beta}(\sigma^\mu)^{\alpha\dot\beta}, \quad q\cdot k = -\frac{1}{2}q_{\alpha\dot\beta}k^{\alpha\dot\beta} = -\frac{1}{2}\epsilon^{\alpha\gamma}\epsilon^{\dot\beta\dot\delta}q_{\alpha\dot\beta}k_{\gamma\dot\delta}. \tag{4.30}$$

Let us try to construct a polarization vector ε^μ using λ_α. Essentially all we need is that ε^μ is orthogonal to p^μ. We can take

$$\varepsilon_{\alpha\dot\beta} \equiv \varepsilon_\mu \sigma^\mu_{\alpha\dot\beta} \propto \lambda_\alpha \overline{u}_{\dot\beta}. \tag{4.31}$$

Such an ε^μ is orthogonal to p^μ, because $\varepsilon \cdot p = -\frac{1}{2}\varepsilon_{\alpha\dot\beta}p^{\alpha\dot\beta}$ is proportional to $\lambda_\alpha\lambda^\alpha = \epsilon^{\alpha\gamma}\lambda_\alpha\lambda_\gamma = 0$. To be completely explicit, let us take

$$p^\mu = (E, 0, 0, E). \tag{4.32}$$

Then we have

$$\lambda = e^{i\phi}\sqrt{2E}\begin{pmatrix}1\\0\end{pmatrix}. \tag{4.33}$$

The matrix $\varepsilon_{\alpha\dot\beta}$ is of the form

$$\begin{pmatrix}* & *\\0 & 0\end{pmatrix}. \tag{4.34}$$

If we choose ε^μ to be purely spatial, i.e. $\varepsilon^0 = 0$, then we must have

$$\varepsilon \propto (0, 1, i, 0) \tag{4.35}$$

This is a polarization vector of positive helicity. A negative helicity polarization vector can be constructed from

$$\varepsilon_{\alpha\dot\beta} \propto u_\alpha \overline{\lambda}_{\dot\beta}. \tag{4.36}$$

Now we will introduce some more notations. For a pair of chiral spinors $\lambda_\alpha, \mu_\alpha$, we will write their scalar product

$$\epsilon^{\alpha\beta}\lambda_\alpha\mu_\beta = \langle\lambda\mu\rangle = -\langle\mu\lambda\rangle. \tag{4.37}$$

For a pair of anti-chiral spinors $\widetilde{\lambda}_{\dot\alpha}, \widetilde{\mu}_{\dot\alpha}$, we write their scalar product as

$$\epsilon^{\dot\alpha\dot\beta}\widetilde{\lambda}_{\dot\alpha}\widetilde{\mu}_{\dot\beta} = [\widetilde{\lambda}\widetilde{\mu}] = -[\widetilde{\mu}\widetilde{\lambda}]. \tag{4.38}$$

Be aware that different sign conventions have been used in the literature, due to the different conventions for the signature and for the ϵ tensor.

Now if we write two null vectors p^μ, q^μ as

$$p_{\alpha\dot\beta} = \lambda_\alpha\widetilde{\lambda}_{\dot\beta}, \qquad q_{\alpha\dot\beta} = \mu_\alpha\widetilde{\mu}_{\dot\beta}, \tag{4.39}$$

we have

$$p \cdot q = -\frac{1}{2}\langle\lambda\mu\rangle[\widetilde{\lambda}\widetilde{\mu}] = \frac{1}{2}\langle\lambda\mu\rangle[\widetilde{\mu}\widetilde{\lambda}]. \tag{4.40}$$

The normalized polarization vectors of positive and negative helicities can now be written in terms of $\lambda_\alpha, \overline{\lambda}_{\dot\alpha}$, and a pair of arbitrary reference spinors $\mu_\alpha, \overline{\mu}_{\dot\alpha}$, as

$$\varepsilon^+_{\alpha\dot\beta} = -\sqrt{2}\frac{\lambda_\alpha \overline{\mu}_{\dot\beta}}{[\overline{\lambda}\,\overline{\mu}]}, \qquad \varepsilon^-_{\alpha\dot\beta} = \sqrt{2}\frac{\mu_\alpha \overline{\lambda}_{\dot\beta}}{\langle \mu\lambda\rangle}. \tag{4.41}$$

With this convention, ε^+_μ and ε^-_μ are the complex conjugate of one another, and obey $\varepsilon^+ \cdot \varepsilon^- = 1$, while $\varepsilon^\pm \cdot \varepsilon^\pm = 0$.

4.3 Tree-level amplitudes in spinor helicity formalism

We know from gauge invariance that the gluon scattering amplitude should depend on the gluon polarization only through the helicities. In particular, the choice of the reference spinors $\mu_i, \overline{\mu}_i$ for each $\varepsilon^\pm_i$ (polarization of the i-th gluon) should not affect the amplitude, once the momenta are put on-shell.

Let us now revisit the color-ordered four gluon tree level scattering amplitude. Suppose all helicities are $-$. Such color-ordered amplitude will be denoted

$$\mathcal{A}(1^-, 2^-, 3^-, 4^-). \tag{4.42}$$

We shall make appropriate choices of reference spinors for $\varepsilon^-_1, \cdots, \varepsilon^-_4$, to simplify the amplitude as much as possible. In particular, we can choose μ_i (and similarly $\overline{\mu}_i$) to be one of the λ_j's that come from the external momentum of a different leg, $j \neq i$. Say we choose $\mu_1 = \mu_2 = \mu_3 = \lambda_4$, and $\mu_4 = \lambda_1$, and similarly for the barred spinors. We then have

$$\varepsilon^-_1 \cdot \varepsilon^-_2 \propto \langle 44\rangle[12] = 0, \tag{4.43}$$

and similarly $\varepsilon^-_1 \cdot \varepsilon^-_3 = \varepsilon^-_2 \cdot \varepsilon^-_3 = 0$. We also have $\varepsilon_{1,2,3} \cdot k_4 = 0$, and $\varepsilon_4 \cdot k_1 = 0$. The diagram involving the quartic vertex simply vanishes in this case. The s-channel diagram reduces to

$$\begin{aligned} & i\left[(\varepsilon^-_2)_\mu \varepsilon^-_1 \cdot (k_1 + 2k_2) + (\varepsilon^-_1)_\mu \varepsilon^-_2 \cdot (-2k_1 - k_2)\right] \\ & \times \frac{1}{(k_1+k_2)^2}\left[(\varepsilon^-_3 \cdot \varepsilon^-_4)(k_3 - k_4)^\mu + (\varepsilon^-_3)^\mu \varepsilon^-_4 \cdot (-2k_3 - k_4)\right] \\ & = \frac{2i}{(k_1+k_2)^2}(\varepsilon^-_3 \cdot \varepsilon^-_4)\left[(\varepsilon^-_1 \cdot k_2)(\varepsilon^-_2 \cdot k_3) - (\varepsilon^-_1 \cdot k_3)(\varepsilon^-_2 \cdot k_1)\right]. \end{aligned} \tag{4.44}$$

Now using momentum conservation,

$$\begin{aligned} \varepsilon^-_1 \cdot k_2 &= -\varepsilon^-_1 \cdot (k_1 + k_3 + k_4) = -\varepsilon^-_1 \cdot k_3, \\ \varepsilon^-_2 \cdot k_1 &= -\varepsilon^-_2 \cdot (k_2 + k_3 + k_4) = -\varepsilon^-_2 \cdot k_3, \end{aligned} \tag{4.45}$$

we see that (4.44) is zero. Similarly, the u-channel diagrams vanishes as well. So in fact the amplitude with all four gluons of negative helicities is zero:

$$\mathcal{A}(1^-, 2^-, 3^-, 4^-) = 0. \tag{4.46}$$

Beware that we have taken our convention to be such that all four gluons are ingoing. When interpreted as a $2 \to 2$ scattering amplitude, say $1, 2 \to 3, 4$, we should flip the sign on the helicity of the out states. So what we found is that the tree level amplitude of $1^-, 2^- \to 3^+, 4^+$ vanishes.

In a similar manner, one can show that

$$\mathcal{A}(1^+, 2^-, 3^-, 4^-) = 0 \tag{4.47}$$

as well. In fact any 4-gluon tree amplitude with one + and three −'s, or one − and three +'s vanish identically. The only non-vanishing amplitudes are the ones that involve two +'s and two −'s.

Let us now compute such a color ordered amplitude, $\mathcal{A}(1^+, 2^+, 3^-, 4^-)$. We will take $\mu_1 = \mu_2 = \lambda_3$, and $\mu_3 = \mu_4 = \lambda_1$. We have

$$\varepsilon_1^+ \cdot \varepsilon_2^+ = \varepsilon_3^- \cdot \varepsilon_4^- = 0. \tag{4.48}$$

There is also

$$\varepsilon_1^+ \cdot \varepsilon_3^- \propto \langle 11 \rangle [33] = 0 = \varepsilon_1^+ \cdot \varepsilon_4^+, \tag{4.49}$$

and $\varepsilon_2^+ \cdot \varepsilon_3^- = 0$. Once again the diagram with the quartic vertex vanishes by itself. The only scalar product of polarization vectors that is non-vanishing is $\varepsilon_2^+ \cdot \varepsilon_4^-$. The s-channel diagram reduces to

$$\frac{i}{(k_1+k_2)^2} (\varepsilon_2^+ \cdot \varepsilon_4^-)\, \varepsilon_1^+ \cdot (k_1 + 2k_2)\, \varepsilon_3^- \cdot (k_3 + 2k_4) = \frac{2i}{k_1 \cdot k_2} (\varepsilon_2^+ \cdot \varepsilon_4^-)(\varepsilon_1^+ \cdot k_2)(\varepsilon_3^- \cdot k_4). \tag{4.50}$$

The u-channel diagram reduces to

$$\frac{i}{(k_2+k_3)^2} (\varepsilon_2^+ \cdot \varepsilon_4^-)\, \varepsilon_1^+ \cdot (-2k_4 - k_1)\, \varepsilon_3^- \cdot (-2k_2 - k_3) = \frac{2i}{k_2 \cdot k_3} (\varepsilon_2^+ \cdot \varepsilon_4^-)(\varepsilon_1^+ \cdot k_4)(\varepsilon_3^- \cdot k_2). \tag{4.51}$$

Since $\varepsilon_1^+ \cdot k_3 = 0$, by momentum conservation we have $\varepsilon_1^+ \cdot k_4 = -\varepsilon_1^+ \cdot k_2$. Similarly, $\varepsilon_3^- \cdot k_2 = -\varepsilon_3^- \cdot k_4$. The s and u channel diagrams combine into

$$-2i \frac{k_2 \cdot k_4}{(k_1 \cdot k_2)(k_2 \cdot k_3)} (\varepsilon_2^+ \cdot \varepsilon_4^-)(\varepsilon_1^+ \cdot k_2)(\varepsilon_3^- \cdot k_4). \tag{4.52}$$

In spinor helicity notation, we can write

$$\varepsilon_2^+ \cdot \varepsilon_4^- = \frac{\langle 21 \rangle [34]}{[23] \langle 14 \rangle}, \qquad \varepsilon_1^+ \cdot k_2 = \frac{1}{\sqrt{2}} \frac{\langle 12 \rangle [32]}{[13]}, \qquad \varepsilon_3^- \cdot k_4 = -\frac{1}{\sqrt{2}} \frac{\langle 14 \rangle [34]}{\langle 13 \rangle}. \tag{4.53}$$

The amplitude is then

$$2i\frac{\langle 24\rangle[42]}{\langle 12\rangle[12]\langle 23\rangle[23]}\frac{\langle 21\rangle[34]}{[23]\langle 14\rangle}\frac{\langle 12\rangle[32]}{[13]}\frac{\langle 14\rangle[34]}{\langle 13\rangle}=-2i\frac{[34]^2\langle 12\rangle}{[12]\langle 23\rangle[23]}. \tag{4.54}$$

In the above step we used momentum conservation, which implies

$$\langle 13\rangle[13]=\langle 24\rangle[24], \quad etc. \tag{4.55}$$

(4.54) is a remarkably simple result! We can simplify it by one more step, using the following relation:

$$\begin{aligned}&[41]\langle 12\rangle=\epsilon^{\dot\alpha\dot\beta}(\overline\lambda_4)_{\dot\alpha}(\overline\lambda_1)_{\dot\beta}\epsilon^{\alpha\beta}(\lambda_1)_\alpha(\lambda_2)_\beta=(\overline\lambda_4)_{\dot\alpha}p_1^{\beta\dot\alpha}(\lambda_2)_\beta\equiv\langle 2|1|4]\\&-\langle 2|2|4]-\langle 2|3|4]-\langle 2|4|4]=-\langle 2|3|4]=-\langle 23\rangle[34].\end{aligned} \tag{4.56}$$

So we have

$$\begin{aligned}\mathcal{A}(1^+,2^+,3^-,4^-)&=-2i\frac{[34]^2\langle 12\rangle}{[12]\langle 23\rangle[23]}\cdot\frac{[41]}{[41]}\\&=2i\frac{[34]^3\langle 23\rangle}{[12]\langle 23\rangle[23][41]}=2i\frac{[34]^3}{[12][23][41]}=2i\frac{[34]^4}{[12][23][34][41]}.\end{aligned} \tag{4.57}$$

There are many different ways to write this amplitude in spinor helicity notation. For instance, we have

$$\begin{aligned}&\frac{[34]^4}{[12][23][34][41]}=-\frac{[34]^2\langle 12\rangle\langle 41\rangle}{[12]\langle 23\rangle[23]\langle 41\rangle}=\frac{[34]\langle 12\rangle[32]\langle 21\rangle}{[12]\langle 23\rangle[23]\langle 41\rangle}=\frac{[34]\langle 12\rangle^2}{[12]\langle 23\rangle\langle 41\rangle}\\&=\frac{[12]\langle 12\rangle^3}{[12]\langle 23\rangle\langle 34\rangle\langle 41\rangle}=\frac{\langle 12\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 34\rangle\langle 41\rangle}.\end{aligned} \tag{4.58}$$

It can be shown that the color-ordered n-gluon amplitude ($n>3$) vanishes if the helicities are $(+,+,\cdots,+)$ or $(-,-,\cdots,-)$, as well as when the helicities are $(-,+,+,\cdots,+)$ and $(+,-,-,\cdots,-)$. The non-vanishing amplitudes must involve at least two $+$'s and at least two $-$'s. The amplitudes $\mathcal{A}(+,+,-,-,\cdots,-)$ are call "maximal helicity violating" amplitudes, or MHV amplitudes. Likewise, $\mathcal{A}(-,-,+,+,\cdots,+)$ is also maximal helicity violating, and is referred to as $\overline{\text{MHV}}$ amplitudes. The MHV amplitudes turn out to take an extremely simple form in spinor helicity formalism. We asset the general result without proof here:

$$\begin{aligned}\mathcal{A}(1^+,2^+,3^-,4^-,\cdots,n^-)&=2^{\frac{n}{2}-1}i\frac{\langle 12\rangle^4}{\langle 12\rangle\langle 23\rangle\cdots\langle n-1\ n\rangle\langle n1\rangle},\\\mathcal{A}(1^-,2^-,3^+,4^+,\cdots,n^+)&=2^{\frac{n}{2}-1}i\frac{[12]^4}{[12][23]\cdots[n-1\ n][n1]}.\end{aligned} \tag{4.59}$$

This is known as Parke-Taylor formula. Note that this result reduces to the four-gluon amplitude computed above. Note that the phase of the amplitude changes if we rotate the phase of the polarization vectors ε_i^+ by multiplying by $e^{2i\phi_i}$ and ε_i^- by $e^{-2i\phi_i}$. By our construction of $\varepsilon^\pm$ using reference spinors, such phase rotation can be achieved if we multiply λ_i by $e^{i\phi_i}$ and multiply $\overline{\lambda}_i$ by $e^{-i\phi_i}$. It follows that the amplitude that involves the i-th gluon of + helicity, when expressed in terms of λ_i and $\overline{\lambda}_i$, must have total degree +2, if we count each power of λ_i as degree +1 and $\overline{\lambda}_i$ as degree −1. For gluon with − helicity the degree must be the opposite. This is indeed consistent with Parke-Taylor formula.

4.4 Recursive relations

We have seen that gluon tree level scattering amplitude simplify dramatically when expressed in the spinor helicity formalism. So far we still derived the amplitude (four-gluon MHV) from explicit evaluation of Feynman diagrams, which is evidently very inefficiently. In fact, the general tree-level amplitudes can be computed much more efficiently using the analytic structure of the S-matrix elements.

Let us consider a general tree-level scattering amplitude, as a function of n external momenta,

$$\mathcal{A}(p_1, p_2, \cdots, p_n). \tag{4.60}$$

Here we omitted writing the helicity dependence. The amplitude is a priori defined only when all p_i's are on-shell, and obey momentum conservation $p_1+p_2+\cdots+p_n = 0$. Evidently, a tree-level amplitude is a rational function of the momenta. It makes sense, mathematically, to analytically continue $\mathcal{A}$ to complex momenta p_i, provided that the on-shell condition $p_i^2 = 0$ and momentum conservation are maintained. For external gluons, the polarization vectors must still obey transversality condition; this can be done if we fix the helicity and construct $\varepsilon^\pm$ using a fixed reference spinor, while analytically continuing $(p_i)_{\alpha\dot\beta} = (\lambda_i)_\alpha(\widetilde{\lambda}_i)_{\dot\beta}$.

A particularly interesting analytic continuation, or complex deformation, of the momenta, is the following one. Take a pair of the particles i, j. We make the replacement

$$p_i^\mu \to \widehat{p}_i^\mu = p_i^\mu + zq^\mu, \qquad p_j^\mu \to \widehat{p}_j^\mu = p_j^\mu - zq^\mu. \tag{4.61}$$

Momentum conservation is clearly satisfied. To maintain the on-shell condition $p_i^2 = p_j^2 = 0$ for any z, we need

$$q^2 = p_i \cdot q = p_j \cdot q = 0. \tag{4.62}$$

For generic null p_i, p_j, there are in fact two solutions to this equation up to rescaling. This is easily seen in spinor helicity formalism,

$$(p_i)_{\alpha\dot\beta} = (\lambda_i)_\alpha(\widetilde{\lambda}_i)_{\dot\beta}, \qquad (p_j)_{\alpha\dot\beta} = (\lambda_j)_\alpha(\widetilde{\lambda}_j)_{\dot\beta}, \tag{4.63}$$

and we can take

$$q_{\alpha\dot{\beta}} = (\lambda_i)_\alpha(\widetilde{\lambda}_j)_{\dot{\beta}} \quad \text{or} \quad (\lambda_j)_\alpha(\widetilde{\lambda}_i)_{\dot{\beta}}, \tag{4.64}$$

either of which obeys (4.62).

Let us denote by $\mathcal{A}(z)$ the tree-level amplitude after this analytic continuation,

$$\mathcal{A}(z) \equiv \mathcal{A}(p_1, \cdots, \widehat{p}_i, \cdots, \widehat{p}_j, \cdots, p_n). \tag{4.65}$$

Evidently, $\mathcal{A}(z)$ is a rational function in z, and generically has only simple poles in z. The denominator in each propagator is either independent of z, or linear in z. The poles of $\mathcal{A}(z)$ occurs when the sum of a subset of momenta that includes $\widehat{p}_i$ goes on-shell (here we are dealing with pure Yang-Mills theory and the only intermediate states are gluons). Namely, if we divide $\{1, 2, \cdots, n\}$ into two groups L, R, with $i \in L$, $j \in R$, then there is a pole at the value of z where

$$(P_L + zq)^2 = P_L^2 + 2zq \cdot P_L = 0, \quad P_L = \sum_{k \in L} p_k = -P_R = -\sum_{k \in R} p_k. \tag{4.66}$$

The residues of the amplitude at the poles of the internal propagators are expressed in terms of products of S-matrix elements with fewer external particles. The entire function $\mathcal{A}(z)$ is determined by its residues at all of its poles, together with its behavior at $z \to \infty$. In particular, if $\mathcal{A}(z)$ vanishes in the limit $z \to \infty$, it will be determined by its poles entirely, and we can recover $\mathcal{A}(0)$ which is the physical amplitude of interest.

Does $\mathcal{A}(z)$ vanish at infinity? If we are doing scalar ϕ^4 theory, then the answer is no, generally, since there are diagrams that are independent of z. In the case of gluon scattering, we are not only shifting the momenta, but the polarization vectors as well, in order to maintain the helicities. With one of the two choices of q (up to rescaling in q), the shift on the spinor helicity variables are

$$\begin{aligned} \lambda_i(z) &= \lambda_i, \quad \widetilde{\lambda}_i(z) = \widetilde{\lambda}_i + z\widetilde{\lambda}_j, \\ \lambda_j(z) &= \lambda_j - z\lambda_i, \quad \widetilde{\lambda}_j(z) = \widetilde{\lambda}_j. \end{aligned} \tag{4.67}$$

Now if the i-th gluon has $+$ helicity, and the j-th gluon has $-$ helicity, then the polarization vectors

$$(\varepsilon_i^+)_{\alpha\dot{\beta}} = -\sqrt{2}\frac{\lambda_{i\alpha}\widetilde{\mu}_{i\dot{\beta}}}{[\widetilde{\lambda}_i\widetilde{\mu}_i]}, \quad (\varepsilon_j^-)_{\alpha\dot{\beta}} = \sqrt{2}\frac{\mu_{j\alpha}\widetilde{\lambda}_{j\dot{\beta}}}{\langle\mu_j\lambda_j\rangle} \tag{4.68}$$

both go like $1/z$ at large z. In a tree level diagram, the only z dependence, apart from the polarization vectors, comes from the cubic vertices and propagators along a path going from the i-th external gluon to the j-th external gluon. Each cubic vertex goes like z and each internal propagator along this path goes like $1/z$. In total we have a factor of z from the vertices and internal propagators, which combines with the polarization vectors ε_i^+ and ε_j^- to give $1/z$. Thus, in this case, each diagram that contributes to $\mathcal{A}(z)$ goes like $1/z$, and $\mathcal{A}(z)$ indeed vanishes at $z = \infty$.

It turns out that when the i-th and j-th gluons have both + helicity or both − helicity, $\mathcal{A}(z)$ also vanishes at $z = \infty$. We won't prove it here. When the i-th gluon has − helicity and the j-th gluon has + helicity, then with the shift (4.67) $\mathcal{A}(z)$ does not vanish at infinity.

If we make a choice of shift z such that $\mathcal{A}(z)$ vanishes at infinity - as we see above this can always be done with a pair of i, j of appropriate helicities, then we can use the residue theorem

$$0 = \oint \frac{dz}{2\pi i} \frac{\mathcal{A}(z)}{z} = \mathcal{A}(0) + \sum_{\text{poles } z_I} \text{Res}_{z \to z_I} \frac{\mathcal{A}(z)}{z}. \tag{4.69}$$

We have seen that the residues come from dividing the external gluons into two sets L, R, with $i \in L$, $j \in R$. The pole is at

$$z_* = -\frac{P_L^2}{2q \cdot P_L} = \frac{P_R^2}{2q \cdot P_R}. \tag{4.70}$$

The residue of $\mathcal{A}(z)/z$ at z_*, which follows from our general analysis of the pole structure of S-matrix elements due to intermediate one-particle states, is given by

$$\sum_{h=\pm} \mathcal{A}(L,h) \frac{-i}{2q \cdot P_L} \mathcal{A}(R,-h) \frac{1}{z_*} = -\sum_{h=\pm} \mathcal{A}(L,h) \frac{-i}{P_L^2} \mathcal{A}(R,-h). \tag{4.71}$$

Here $\mathcal{A}(L,h)$ is the amplitude of the L gluons with an additional gluon of helicity h and momentum $-P_L - z_* q$, and $\mathcal{A}(L,-h)$ is the amplitude of the R gluons with an additional gluon of the opposite helicity $-h$ and momentum $P_L + z_* q = -P_R + z_* q$. Summing over all residues gives the result

$$\mathcal{A} \equiv \mathcal{A}(0) = \sum_{L,R} \sum_{h=\pm} \mathcal{A}(L,h) \frac{-i}{P_L^2} \mathcal{A}(R,-h). \tag{4.72}$$

This recusive relation, known as the BCFW relation, is a powerful formula that allows for efficient computation of tree level amplitudes of gluons in pure Yang-Mills theory. It admits many generalizations, e.g. to scattering amplitudes of gravitons. To some extent this approach can be extended to loop amplitudes as well, using the so called unitarity cut method. We won't discuss them further here.

5 Effective action

5.1 1PI effective action

The one-particle irreducible (1PI) effective action is a convenient way of organizing quantum corrections in a general quantum field theory, and is particularly useful in understanding gauge invariance and renormalization in gauge theories. The idea is that the full quantum correlation functions or scattering amplitudes can be obtained from a *tree level* (i.e. classical) calculation using an effective action S_{eff}. A general construction of such an effective action is achieved as follows. Denote collectively by $\phi(x)$ all fields in a quantum field theory, and $S[\phi]$ the classical action. Let us begin with the generating functional (that computes the vacuum amplitude with sources)

$$Z[J] = \int \mathcal{D}\phi \, \exp\left[iS[\phi] + i \int d^4x \phi(x) J(x) \right]. \tag{5.1}$$

Although we have only written the expression when ϕ is a scalar field, the generating function can be straightforwardly generated to the cases where ϕ carries Lorentz indices as well as for fermionic fields (in which case J would correspondingly be anti-commuting, i.e. Grassmannian sources). Diagrammatically, $Z[J]$ is given by the sum of all Feynman diagrams with a source $J(x)$ attached to each external ϕ-line. This sum includes both connected and disconnected diagrams. The generating functional for all *connected* diagrams, which we denote by $iW[J]$, is related to $Z[J]$ by

$$Z[J] = \exp\left(iW[J]\right). \tag{5.2}$$

The quantum effective action is then defined as the Legendre transform of $W[J]$,

$$\Gamma[\phi] \equiv \left[- \int d^4x \phi(x) J(x) + W[J] \right] \Bigg|_{\frac{\delta W[J]}{\delta J} = \phi} \tag{5.3}$$

$\Gamma[\phi]$ is in fact the "1PI effective action", in the sense that the n-point vertices derived from $\Gamma[\phi]$ sum up one-particle irreducible diagrams (i.e. diagrams that cannot be disconnected by cutting one particle line) with n external ϕ-lines. The simplest way to see this is to note the inverse Legendre transform

$$W[J] = \left[\Gamma[\phi] + \int d^4x \phi(x) J(x) \right]_{\frac{\delta \Gamma[\phi]}{\delta \phi} + J = 0} \tag{5.4}$$

may be computed from the path integral with action $\Gamma[\phi]$ in the $\hbar \to 0$ limit,

$$iW[J] = \lim_{\hbar \to 0} \hbar \ln \int \mathcal{D}\phi \exp\left\{ \frac{i}{\hbar} \left[\Gamma[\phi] + \int d^4x \phi(x) J(x) \right] \right\}. \tag{5.5}$$

The RHS side simply sums up all *connected tree level* diagrams computed from the effective action $\Gamma[\phi]$ while throwing away all loop corrections that are suppressed by powers of $\hbar$.

Equivalently, we may write

$$\begin{aligned} e^{i\Gamma[\phi_0]} &= \int \mathcal{D}\phi \, e^{iS[\phi]+i\int d^4x(\phi-\phi_0)J}\Big|_{\langle\phi\rangle_J=\phi_0} \\ &= \int \mathcal{D}\hat{\phi} \, e^{iS[\phi_0+\hat{\phi}]+i\int d^4x\hat{\phi}J}\Big|_{\langle\hat{\phi}\rangle_J=0} \end{aligned} \tag{5.6}$$

On the RHS, it is understood that the source J is *determined* by ϕ_0 via the constraint $\langle\phi\rangle_J = \phi_0$ or $\langle\hat{\phi}\rangle_J = 0$.

In other words, we may compute $\Gamma[\phi_0]$ by the following rule. First expand the classical action around a background field ϕ_0, and add a linear term in the fluctuation field $\hat{\phi}$ to cancel the expectation value of $\hat{\phi}$ (tadpoles). The shift of the action by $\int \hat{\phi}J$ with J such that $\langle\hat{\phi}\rangle_J = 0$ is precisely to cancel all diagrams that can be disconnected by cutting an internal line, leaving only the contribution from 1PI diagrams. In other words, we can write

$$e^{i\Gamma[\phi_0]} = \int_{\rm 1PI} \mathcal{D}\hat{\phi} \, e^{iS[\phi_0+\hat{\phi}]}. \tag{5.7}$$

Let ϕ_* be a stationary point of the effective action $\Gamma[\phi]$, i.e.

$$\frac{\delta\Gamma[\phi]}{\delta\phi}\Big|_{\phi=\phi_*} = 0. \tag{5.8}$$

On the other hand, we have

$$\phi_*(x) = \frac{\delta W[J]}{\delta J(x)}\Big|_{J=0} = -\frac{i}{Z[0]}\frac{\delta}{\delta J(x)}Z[J]\Big|_{J=0} = \langle\phi(x)\rangle. \tag{5.9}$$

So we see that the stationary point of the effective action is the quantum expectation value of ϕ.

5.2 Effective potential

Let us now consider the quantum effective action in the example of ϕ^4 theory. The classical action is

$$S[\phi] = -\int d^4x \left[\frac{1}{2}(\partial_\mu\phi)^2 + \frac{1}{2}m^2\phi^2 + \frac{1}{4!}g\phi^4\right]. \tag{5.10}$$

The full effective action $\Gamma[\phi_0]$ computed by summing over 1PI diagrams using the action $S[\phi_0 + \phi]$ is rather complicated. We will consider the simpler problem of computing $\Gamma[\phi_0]$ for constant ϕ_0. In this case, we may write

$$\Gamma[\phi_0] = -V(\phi_0) \int d^4x = -V(\phi_0)\mathcal{V}_4, \tag{5.11}$$

where $\mathcal{V}_4$ is the (infinite) volume of the spacetime. $V(\phi_0)$ is called the effective potential.

We now compute $V(\phi_0)$ in perturbation theory. The zeroth order contribution, i.e. the zero-loop term, is simply the classical potential $V^{(0)}(\phi_0) = \frac{1}{2}m^2\phi_0^2 + \frac{1}{4!}g\phi_0^4$. At one-loop and higher, we shall compute 1PI diagrams using the action

$$S[\phi_0 + \phi] = S[\phi_0] + (\text{linear term}) - \int d^4x \left[\frac{1}{2}(\partial_\mu\phi)^2 + \frac{1}{2}\mu^2(\phi_0)\phi^2\right] - \int d^4x \left(\frac{1}{6}g\phi_0\phi^3 + \frac{1}{24}g\phi^4\right), \tag{5.12}$$

where

$$\mu^2(\phi_0) = m^2 + \frac{1}{2}g\phi_0^2. \tag{5.13}$$

The 1-loop effective action is given by the logarithm of the functional determinant of the kinetic operator on ϕ,

$$\begin{aligned} i\Gamma^{1-\text{loop}}[\phi_0] &= \ln\left[\det(-\Box + \mu^2(\phi_0) - i\epsilon)\right]^{-\frac{1}{2}} \\ &= -\frac{1}{2}\mathrm{Tr}\ln(-\Box + \mu^2(\phi_0) - i\epsilon). \end{aligned} \tag{5.14}$$

The trace here is taken over the space of the classical field $\phi(x)$, i.e. the space of a real function in x. It can be computed as

$$\begin{aligned} \mathrm{Tr}\ln(-\Box + \mu^2(\phi_0) - i\epsilon) &= \int d^4x \langle x| \ln(-\Box + \mu^2(\phi_0) - i\epsilon)|x\rangle \\ &= \int d^4x \int \frac{d^4p}{(2\pi)^4} \langle x|p\rangle\langle p| \ln(-\Box + \mu^2(\phi_0) - i\epsilon)|x\rangle \\ &= \mathcal{V}_4 \int \frac{d^4p}{(2\pi)^4} \ln(p^2 + \mu^2(\phi_0) - i\epsilon). \end{aligned} \tag{5.15}$$

From this we derive the 1-loop effective potential,

$$V^{1-\text{loop}}(\phi_0) = \frac{-i}{2} \int \frac{d^4p}{(2\pi)^4} \ln(p^2 + \mu^2(\phi_0) - i\epsilon). \tag{5.16}$$

The UV divergence here simply reflects the need for renormalization at one-loop. We may proceed by taking the triple derivative in μ^2, so that the p-integral converges,

$$\left(\frac{\partial}{\partial\mu^2}\right)^3 V^{1-\text{loop}} = -i \int \frac{d^4p}{(2\pi)^4} \frac{1}{(p^2 + \mu^2(\phi_0) - i\epsilon)^3}. \tag{5.17}$$

The $i\epsilon$ prescription is such that we can perform the counterclockwise Wick rotation $p^0 = ip_4$ while avoiding running into poles, giving the result

$$\left(\frac{\partial}{\partial \mu^2}\right)^3 V^{1-\text{loop}} = \int \frac{d^4 p_E}{(2\pi)^4} \frac{1}{(p_E^2 + \mu^2(\phi_0))^3} = \frac{1}{32\pi^2 \mu^2(\phi_0)}. \tag{5.18}$$

Integrate three times in μ^2, we obtain the 1-loop effective potential

$$V^{1-\text{loop}} = \frac{\mu^4 \ln \mu^2}{64\pi^2} + A + B\mu^2 + C\mu^4. \tag{5.19}$$

While A, B, C are not determined this way and are in fact divergent, they can be absorbed into the zero point energy, mass squared m^2, and the quartic coupling g by shifting the bare values. In terms of the renormalized mass m_R and coupling g_R, we have up to one-loop contributions and the zero-point energy,

$$V(\phi_0) = \frac{1}{2} m_R^2 \phi_0^2 + \frac{1}{4!} g_R \phi_0^4 + \frac{\mu^4(\phi_0) \ln \mu^2(\phi_0)}{64\pi^2}. \tag{5.20}$$

While the field dependent mass $\mu^2(\phi_0)$ was originally expressed in terms of the bare mass and coupling, at this order it can be replaced by the same expression with renormalized mass and coupling,

$$\mu^2(\phi_0) = m_R^2 + \frac{1}{2} g_R \phi_0^2. \tag{5.21}$$

This is needed to make the one-loop effective potential finite. The replacement $m \to m_R, g \to g_R$ in $\mu^2(\phi_0)$ amounts to computing one-loop 1PI diagrams using the propagator with the one-loop corrected mass and coupling, effectively summing up an infinite set of diagrams.

We have seen that the expectation value of the field ϕ can be obtained by extremizing the effective potential. We can also give an interpretation to the effective potential away from its extrema. Generally, there are many states $|\Omega\rangle$ in which $\phi(\vec{x}, t)$ has a given expectation value at time t. Let $|\Omega_{\phi_0}\rangle$ be the state that minimizes the energy (density) expectation value subject to the constraint that the expectation value of $\phi(\vec{x}, t)$ is $\phi_0(\vec{x})$. This is achieved by minimizing $\langle\Omega|H|\Omega\rangle$ with respect to a state $|\Omega\rangle$ in the Hilbert space, subject to the constraints

$$\langle\Omega|\Omega\rangle = 1, \quad \langle\Omega|\phi(x)|\Omega\rangle = \phi_0(x). \tag{5.22}$$

Introducing Lagrangian multipliers, this requires extremizing

$$\langle\Omega|H|\Omega\rangle - \alpha\left(\langle\Omega|\Omega\rangle - 1\right) - \int d^3\vec{x}\beta(\vec{x})\left(\langle\Omega|\phi(\vec{x}, t)|\Omega\rangle - \phi_0(\vec{x})\right) \tag{5.23}$$

with respect to α, $\beta(\vec{x})$, and minimizing it with respect to unconstrained $|\Omega\rangle$. We obtain in particular

$$H|\Omega\rangle = \alpha|\Omega\rangle + \int d^3\vec{x}\beta(\vec{x})\phi(\vec{x},t)|\Omega\rangle. \tag{5.24}$$

This equation can be read as one that determines $|\Omega\rangle$ given α and $\beta(\vec{x})$. α and $\beta(\vec{x})$ are then adjusted so that (5.22) is satisfied (at a given time t). We can now interpret $\beta(\vec{x})$ as an external current, and

$$H - \int d^3\vec{x}\beta(\vec{x})\phi(\vec{x},t) \tag{5.25}$$

as the Hamiltonian in the presence of the current $\beta(\vec{x})$, derived from the modified action

$$\int d^4x\big[\mathcal{L} + \beta(\vec{x})\phi(\vec{x},t)\big]. \tag{5.26}$$

α is the ground state energy in the presence of the current $\beta(\vec{x})$. This follows from the assumption that $\langle\Omega|H|\Omega\rangle$ is minimized by our $|\Omega\rangle$. This ground state energy α can also be computed from the transition amplitude

$$\langle 0, out|0, in\rangle_\beta = e^{-i\alpha T} = e^{iW[\beta]}, \tag{5.27}$$

where the current $\beta(\vec{x})$ is turned on adiabatically, and turned off adiabatically after a long time T. In this adiabatic process the vacuum changes into the new ground state with respect to the current modified Hamiltonian. The result is nothing but our generating function $e^{iW[\beta]}$. Now we can derive the minimal energy

$$\begin{aligned}\langle\Omega|H|\Omega\rangle &= \alpha\langle\Omega|\Omega\rangle + \int d^3\vec{x}\beta(\vec{x})\langle\Omega|\phi(\vec{x},t)|\Omega\rangle \\ &= \alpha + \int d^3\vec{x}\beta(\vec{x})\phi_0(\vec{x}) \\ &= \frac{1}{T}\left[-W[\beta] + \int d^4x\beta(\vec{x})\phi_0(\vec{x})\right] \\ &= -\frac{1}{T}\Gamma[\phi_0] = \mathcal{V}_3 V(\phi_0).\end{aligned} \tag{5.28}$$

This result shows that the effective potential $V(\phi_0)$ has the interpretation of the *minimal* energy subject to the constraint $\langle\phi(x)\rangle = \phi_0$.

Note that the second order derivative of $V(\phi_0)$ in ϕ_0 is the inverse of the two-point function of ϕ in the matrix sense, evaluated at zero momentum. The latter is positive definite in the matrix sense. This implies that $V(\phi_0)$ is a convex function. Suppose we have a scalar field theory with a quartic coupling, and negative m^2, in its classical action. The potential is still bounded from below, but is not convex. What we just saw is that the quantum effective potential must be convex, seemingly contradicting (5.20). In fact, in deriving (5.20) we have assumed a stable vacuum with vanishing expectation value of $\phi(x)$, which is not the case when $m^2 < 0$. In this case, there are two vacuum states $|\Omega_1\rangle$ and $|\Omega_2\rangle$, in which the expectation value of $\phi(x)$ (to leading order in perturbation theory) is at the one of the two minima of the classical potential,

$$\langle\phi(x)\rangle = \phi_\pm \equiv \pm\sqrt{\frac{6|m^2|}{g}}. \tag{5.29}$$

If we restrict $\langle\phi(x)\rangle$ to lie between ϕ_- and ϕ_+, we can minimize the energy with a state of the form

$$|\Omega\rangle = c_1|\Omega_1\rangle + c_2|\Omega_2\rangle, \qquad \langle\phi\rangle_\Omega = |c_1|^2\phi_- + |c_2|^2\phi_+ \in (\phi_-, \phi_+). \tag{5.30}$$

Thus the true effective potential $V(\phi_0)$ is a *constant* in between ϕ_- and ϕ_+! We see that perturbation theory fails to capture this effective potential (in fact, the one-loop calculation would give a complex effective potential in between ϕ_- and ϕ_+).

5.3 Renormalization of nonabelian gauge theory

We will now deal with the renormalization of nonabelian gauge theory by studying the 1PI effective action

$$\Gamma[A_\mu, \psi, \eta, \bar\eta] \tag{5.31}$$

where ψ denotes matter fields (transforming in some representation R of the gauge group G), and $\eta, \bar\eta$ are the Fadeev-Popov ghosts. Γ is computed by shifting $A_\mu \to A'_\mu + A_\mu$, etc., and summing up 1PI diagrams of the quantum fluctuations $A'_\mu, \psi', \cdots$, in obtaining a functional of the background fields $A_\mu, \psi, \cdots$. To perform this calculation we need to choose a gauge (or rather, a gauge fixing term). A particularly convenient choice for computing the effective action is the background field gauge, where instead of working with the gauge fixing function $f_a[A] = \partial_\mu A'^\mu_a$, we take

$$f_a[A] = \overline{D}_\mu A'^\mu_a = \partial_\mu A'^\mu_a + f_{abc}A_{b\mu}A'^\mu_c. \tag{5.32}$$

Here $\overline{D}_\mu$ is the covariant derivative with respect to the background gauge field A_μ, and A'^μ_a is now thought of as a field in the adjoint representation. The gauge fixing term $-\frac{1}{2\xi}f_af_a$ is now invariant under the transformation

$$\begin{aligned} \delta A^\mu_a &= \overline{D}^\mu\epsilon_a = \partial^\mu\epsilon_a + f_{abc}A^\mu_b\epsilon_c, \\ \delta A'^\mu_a &= f_{abc}A'^\mu_b\epsilon_c. \end{aligned} \tag{5.33}$$

We can extend this formal transformation to the matter fields ψ' and the background matter field ψ in the same form as gauge transformations, and to the ghost fields $\eta, \bar\eta$

as if they are in the adjoint representation as well (note that the ghosts by definition do *not* transform under the original gauge transformation). In fact, the ghost action is now

$$S_{gh} = \int d^4x \mathrm{Tr}\Big[(\bar{\eta}+\bar{\eta}')\overline{D}_\mu\big(\overline{D}^\mu(\eta+\eta') - i[A'^\mu, \eta+\eta']\big)\Big] \tag{5.34}$$

and is invariant under our new transformation. It is not hard to see that the full background-gauge fixed action is invariant under this transformation. Consequently, after integrating out $A'_\mu, \psi', \cdots$, the 1PI effective action $\Gamma[A_\mu, \psi, \eta, \bar{\eta}]$ is invariant under the remaining "gauge" transformation

$$\delta A_\mu = \overline{D}_\mu \epsilon, \quad \delta\psi = i\epsilon_a t_R^a \psi, \quad \delta\eta = i[\epsilon, \eta], \quad \delta\bar{\eta} = i[\epsilon, \bar{\eta}]. \tag{5.35}$$

This gauge invariance of the effective action restricts possible ultraviolate divergent terms, namely the latter must be invariant under (5.35). By dimension analysis, divergent terms in the effective action can only involve operators made out of the fields and their derivatives of total mass dimension ≤ 4. The divergent part of the effective action must therefore take the form

$$\Gamma_\infty = \int d^4x \left[-\frac{1}{4}L_A F_{a\mu\nu}F_a^{\mu\nu} - L_\psi \bar{\psi}\gamma^\mu \overline{D}_\mu \psi - \delta m \bar{\psi}\psi - L_\omega(\overline{D}_\mu \bar{\eta}_a)(\overline{D}^\mu \eta_a)\right]. \tag{5.36}$$

The coefficients $L_A, L_\psi, \delta m, L_\omega$ are logarithmically divergent (on dimensional grounds, δm may also contain power divergences, but such divergences are absent in dimensional regularization scheme). As we have seen in the ϕ^4 theory example, these divergent contribution to the effective action can be absorbed in the renormalized coupling of the classical action. Note in particular that the renormalization of the structure constant f_{abc} can be absorbed simply into a renormalization of the Yang-Mills coupling g_{YM}.

Now we will study the gauge coupling renormalization by computing the one-loop 1PI effective action. Due to the gauge invariance of the effective action computed in background field gauge, there are various equivalent ways of extract the gauge coupling renormalization. For instance, we may compute the renormalization of the gauge kinetic term

$$\mathrm{Tr}(\partial_\mu A_\nu - \partial_\nu A_\mu)^2, \tag{5.37}$$

by computing the quadratic term in the effective action in the gauge fields, and extract the two-derivative term in the zero momentum limit, or we could compute the quartic gauge field coupling with no derivatives,

$$\mathrm{Tr}\,[A_\mu, A_\nu]^2\,. \tag{5.38}$$

We will do the latter calculation here since it can be performed by setting the background gauge fields to constant, while leaving the former calculation (which is more standard) as an exercise.

To proceed, we take the background fields $A_{a\mu}$ to be constant valued (and so $F_{a\mu\nu} = f_{abc}A_{b\mu}A_{c\nu}$) and the matter fields ψ as well as the ghosts $\eta_a, \bar{\eta}_a$ to zero. The full Lagrangian in background field gauge is now

$$\begin{aligned}\mathcal{L} = &-\frac{1}{4}\left(F_{a\mu\nu} + \overline{D}_\mu A'_{a\nu} - \overline{D}_\nu A'_{a\mu} + f_{abc}A'_{b\mu}A'_{c\nu}\right)^2 - \bar{\psi}'(\gamma^\mu\overline{D}_\mu - i\gamma^\mu A'_{a\mu}t^a_R + m)\psi' \\ &- \frac{1}{2\xi}(\overline{D}_\mu A'^\mu_a)^2 - (\overline{D}_\mu\bar{\eta}'_a)(\overline{D}^\mu\eta'_a + f_{abc}A'^\mu_b\eta'_c)\end{aligned} \tag{5.39}$$

The computation of the one-loop effective action makes use of only the part of the Lagrangian $\mathcal{L}$ that is quadratic in the fluctuation fields,

$$\begin{aligned}\mathcal{L}^{(2)} &= -\frac{1}{4}\left(\overline{D}_\mu A'_{a\nu} - \overline{D}_\nu A'_{a\mu}\right)^2 - \frac{1}{2}f_{abc}F^{\mu\nu}_a A'_{b\mu}A'_{c\nu} - \frac{1}{2\xi}(\overline{D}_\mu A'^\mu_a)^2 \\ &\quad - \bar{\psi}'(\gamma^\mu\overline{D}_\mu + m)\psi' - (\overline{D}_\mu\bar{\eta}'_a)(\overline{D}^\mu\eta'_a) \\ &= \mathcal{L}^{(2)}_{A'} + \mathcal{L}^{(2)}_{\psi'} + \mathcal{L}^{(2)}_{gh}.\end{aligned} \tag{5.40}$$

In the last step, we have separated the quadratic part of the action in three terms, involving A', ψ', and $(\eta', \bar{\eta}')$ respectively. The one-loop contribution to the effective action is simply given by the three functional determinants of the kinetic operators $\mathbb{D}_A$, $\mathbb{D}_\psi$, $\mathbb{D}_\eta$ on A', ψ', and $(\eta', \bar{\eta}')$,

$$e^{i\Gamma^{1-\text{loop}}[A]} = \frac{\det\mathbb{D}_\psi \det\mathbb{D}_\eta}{\sqrt{\det\mathbb{D}_A}}, \tag{5.41}$$

where

$$\begin{aligned}[\mathbb{D}_A]_{a\mu,b\nu} &= \left[-\overline{D}_\rho\overline{D}^\rho\eta_{\mu\nu} + \overline{D}_\nu\overline{D}_\mu - \frac{1}{\xi}\overline{D}_\mu\overline{D}_\nu\right]\delta_{ab} + f_{cab}F_{c\mu\nu}, \\ \mathbb{D}_\psi &= -\gamma^\mu\overline{D}_\mu - m, \\ [\mathbb{D}_\eta]_{ab} &= \overline{D}_\mu\overline{D}^\mu\delta_{ab}.\end{aligned} \tag{5.42}$$

Equivalently,

$$\begin{aligned}i\Gamma^{1-\text{loop}}[A] &= \text{Tr}\ln\mathbb{D}_\psi + \text{Tr}\ln\mathbb{D}_\eta - \frac{1}{2}\text{Tr}\ln\mathbb{D}_A \\ &= \int d^4x \int \frac{d^4p}{(2\pi)^4}\left[\text{Tr}\ln\mathbb{M}_\psi(p) + \text{Tr}\ln\mathbb{M}_\eta(p) - \frac{1}{2}\text{Tr}\ln\mathbb{M}_A(p)\right]\end{aligned} \tag{5.43}$$

where we have represented the kinetic operator in momentum space as

$$\langle q|\mathbb{D}|p\rangle = (2\pi)^4\delta^4(p-q)\mathbb{M}(p). \tag{5.44}$$

Here $\mathbb{M}(p)$ is a matrix that acts on the internal indices of the fields only. We will now work with the gauge choice $\xi = 1$. With constant background fields, they are particularly simple to write. We have

$$
\begin{aligned}
&[\mathbb{M}_A(p)]_{a\mu,b\nu} = (p_\rho\delta_{ac} - if_{adc}A_{d\rho})\,(p^\rho\delta_{cb} - if_{ceb}A_e^\rho)\,\eta_{\mu\nu} + (p_\nu\delta_{ac} - if_{adc}A_{d\nu})(p_\mu\delta_{cb} - if_{ceb}A_{e\mu}) \\
&\qquad - (p_\mu\delta_{ac} - if_{adc}A_{d\mu})(p_\nu\delta_{cb} - if_{ceb}A_{e\nu}) + f_{cab}F_{c\mu\nu}, \\
&\mathbb{M}_\psi(p) = i\gamma^\mu\,(p_\mu - A_{a\mu}t_R^a) + m, \\
&[\mathbb{M}_\eta(p)]_{ab} = (p_\mu\delta_{ac} - if_{adc}A_{d\mu})\,(p^\mu\delta_{cb} - if_{ceb}A_e^\mu)\,.
\end{aligned}
\tag{5.45}
$$

The remaining calculation reduces to an algebraic problem. $\mathrm{Tr}\ln\mathbb{M}(p)$ can be expanded in $A_{a\mu}$ and involves all orders terms in $A_{a\mu}$. Here we are interested in A^4 term only. The problem of computing A^4 term in $\mathrm{Tr}\ln\mathbb{M}(p)$ is equivalent to computing the integrand of the one-loop Feynman diagram with four external A lines at zero momentum, and momentum p running in the loop. Let us begin with the A-loop, i.e. the computation of $-\frac{1}{2}\mathrm{Tr}\ln\mathbb{M}_A(p)$. We can write

$$
\mathbb{M}_A(p) = \mathbb{M}_A^{(0)}(p) + \mathbb{M}_A^{(1)}(p) + \mathbb{M}_A^{(2)}(p), \tag{5.46}
$$

where $\mathbb{M}_A^{(n)}(p)$ is the term of order n in the background field A,

$$
\begin{aligned}
&\mathbb{M}_A^{(0)}(p) = p^2 - i\epsilon, \\
&\mathbb{M}_A^{(1)}(p) = -2p^\mu\,[A_\mu]^{\mathrm{adj}}\,, \qquad \text{i.e.} \quad \left[\mathbb{M}_A^{(1)}(p)\right]_{a\mu,b\nu} = -2if_{acb}A_{c\rho}p^\rho\eta_{\mu\nu}, \\
&\left[\mathbb{M}_A^{(2)}(p)\right]_{\mu\nu} = [\eta_{\mu\nu}A_\rho A^\rho - A_\nu A_\mu + A_\mu A_\nu + iF_{\mu\nu}]^{\mathrm{adj}}\,, \\
&\text{i.e.} \quad \left[\mathbb{M}_A^{(2)}(p)\right]_{a\mu,b\nu} = -f_{adc}f_{ceb}\,(A_{d\rho}A_e^\rho\eta_{\mu\nu} + A_{d\nu}A_{e\mu} - A_{d\mu}A_{e\nu}) + f_{cab}F_{c\mu\nu}.
\end{aligned}
\tag{5.47}
$$

Note a simplification due to the choice of $\xi = 1$: the linear term in A, $\mathbb{M}_A^{(1)}(p)$ is particularly simple and is proportional to $\eta_{\mu\nu}$; e.g. terms proportional to $p_\mu A_\nu$ are canceled.

Now expanding

$$
\begin{aligned}
\mathrm{Tr}\ln\mathbb{M}_A(p) &= \mathrm{Tr}\ln\mathbb{M}_A^{(0)}(p) + \mathrm{Tr}\left[\left(\mathbb{M}_A^{(0)}(p)\right)^{-1}\left(\mathbb{M}_A^{(1)}(p) + \mathbb{M}_A^{(2)}(p)\right)\right] \\
&\quad - \frac{1}{2}\mathrm{Tr}\left[\left(\mathbb{M}_A^{(0)}(p)\right)^{-1}\left(\mathbb{M}_A^{(1)}(p) + \mathbb{M}_A^{(2)}(p)\right)\left(\mathbb{M}_A^{(0)}(p)\right)^{-1}\left(\mathbb{M}_A^{(1)}(p) + \mathbb{M}_A^{(2)}(p)\right)\right] + \cdots\,,
\end{aligned}
\tag{5.48}
$$

we can extract the A^4 term

$$\begin{aligned}
\Big[\mathrm{Tr}\ln\mathbb{M}_A(p)\Big]_{A^4} &= -\frac{1}{2}\mathrm{Tr}\left[\left(\left(\mathbb{M}_A^{(0)}\right)^{-1}\mathbb{M}_A^{(2)}\right)^2\right] + \mathrm{Tr}\left[\left(\left(\mathbb{M}_A^{(0)}\right)^{-1}\mathbb{M}_A^{(1)}\right)^2\left(\mathbb{M}_A^{(0)}\right)^{-1}\mathbb{M}_A^{(2)}\right] \\
&\quad - \frac{1}{4}\mathrm{Tr}\left[\left(\left(\mathbb{M}_A^{(0)}\right)^{-1}\mathbb{M}_A^{(1)}\right)^4\right] \\
&= -\frac{1}{2}\mathrm{Tr}\left[(\eta_{\mu\nu}A^2 + [A_\mu, A_\nu] + iF_{\mu\nu})(\eta^{\mu\nu}A^2 + [A^\nu, A^\mu] + iF^{\nu\mu})\right]\frac{1}{(p^2 - i\epsilon)^2} \\
&\quad + \mathrm{Tr}\left[(-2p^\mu A_\mu)^2(4A^2)\right]\frac{1}{(p^2 - i\epsilon)^3} - \mathrm{Tr}\left[(-2p^\mu A_\mu)^4\right]\frac{1}{(p^2 - i\epsilon)^4} \\
&= -\frac{1}{2}\mathrm{Tr}\left[4(A^2)^2 + 4F_{\mu\nu}F^{\mu\nu}\right]\frac{1}{(p^2 - i\epsilon)^2} + 16\mathrm{Tr}\left[A_\mu A_\nu A^2\right]\frac{p^\mu p^\nu}{(p^2 - i\epsilon)^3} - 16\mathrm{Tr}\left[A_\mu A_\nu A_\rho A_\sigma\right]\frac{p^\mu p^\nu p^\rho p^\sigma}{(p^2 - i\epsilon)^4}.
\end{aligned} \tag{5.49}$$

After the second and third equality above, Tr stands for the trace over the color index only; the trace over Lorentz indices has already been carried out.

The momentum integration can be turned into an integration over Euclidean momentum by counterclockwise Wick rotation $p^0 \to ip^4$,

$$\int \frac{d^4p}{(2\pi)^4}\frac{1}{(p^2 - i\epsilon)^2} \to i\int d^4p_E \frac{1}{(p_E^2)^2}. \tag{5.50}$$

This is clearly logarithmically divergent, both in the UV and in the IR. We will defer the physical interpretations, and for now simply regularize by introducing a UV cutoff Λ and an IR cutoff μ, so that

$$\int_\mu^\Lambda \frac{d^4p_E}{(2\pi)^4}\frac{1}{(p_E^2)^2} = \frac{2\pi^2}{(2\pi)^4}\int_\mu^\Lambda \frac{dp}{p} = \frac{1}{8\pi^2}\ln\frac{\Lambda}{\mu}. \tag{5.51}$$

The momentum integrals with $p^\mu p^\nu$ and $p^\mu p^\nu p^\rho p^\sigma$ in the numerator can be performed easily using the Euclidean rotational symmetry. We can replace them by expressions proportional to invariant tensors,

$$\begin{aligned}
p^\mu p^\nu &\to \frac{1}{4}p^2\eta^{\mu\nu}, \\
p^\mu p^\nu p^\rho p^\sigma &\to \frac{1}{24}p^4\left(\delta^{\mu\nu}\delta^{\rho\sigma} + \delta^{\mu\rho}\delta^{\nu\sigma} + \delta^{\mu\sigma}\delta^{\rho\nu}\right).
\end{aligned} \tag{5.52}$$

The normalization of the RHS is fixed by comparing its trace with that of the LHS.

Thus we obtain

$$
\begin{aligned}
&\int \frac{d^4p}{(2\pi)^4}\Big[\mathrm{Tr}\ln\mathbb{M}_A(p)\Big]_{A^4}\\
&= \mathrm{Tr}\left[-2(A^2)^2 - 2F_{\mu\nu}F^{\mu\nu} + 4(A^2)^2 - \frac{2}{3}\left(2(A^2)^2 + A_\mu A_\nu A^\mu A^\nu\right)\right]\int\frac{d^4p}{(2\pi)^4}\frac{1}{(p^2-i\epsilon)^2}\\
&= \mathrm{Tr}\left[-2F_{\mu\nu}F^{\mu\nu} + \frac{2}{3}\left(A_\mu A^\mu A_\nu A^\nu - A_\mu A_\nu A^\mu A^\nu\right)\right]\int\frac{d^4p}{(2\pi)^4}\frac{1}{(p^2-i\epsilon)^2}\\
&= -\frac{5}{3}\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})\int\frac{d^4p}{(2\pi)^4}\frac{1}{(p^2-i\epsilon)^2}\\
&= -\frac{5}{3}\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})\frac{i}{8\pi^2}\ln\frac{\Lambda}{\mu}.
\end{aligned}
\tag{5.53}
$$

Note that the trace here is taken in the adjoint representation of G. Explicitly in components, we have

$$
\mathrm{Tr}_{\mathrm{adj}}(F_{\mu\nu}F^{\mu\nu}) = -F_{a\mu\nu}F_b^{\mu\nu}f_{cad}f_{dbc}. \tag{5.54}
$$

A slightly simpler calculation gives the the analogous quantition from the ghost loop,

$$
\int\frac{d^4p}{(2\pi)^4}\Big[\mathrm{Tr}\ln\mathbb{M}_{gh}(p)\Big]_{A^4} = \frac{1}{12}\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu})\frac{i}{8\pi^2}\ln\frac{\Lambda}{\mu}. \tag{5.55}
$$

For the purpose of understanding the high energy behavior of the nonabelian gauge theory, we can set the mass of the matter fields to zero. For massless fermion matter ψ in some representation R, we have

$$
\begin{aligned}
&\int\frac{d^4p}{(2\pi)^4}\Big[\mathrm{Tr}\ln\mathbb{M}_\psi(p)\Big]_{A^4} = -\frac{1}{4}\int\frac{d^4p}{(2\pi)^4}\mathrm{Tr}\left[\left(\left(\mathbb{M}_\psi^{(0)}(p)\right)^{-1}\mathbb{M}_\psi^{(1)}(p)\right)^4\right]\\
&= -\frac{1}{4}\mathrm{Tr}_R(T^aT^bT^cT^d)A_{a\mu}A_{b\nu}A_{c\rho}A_{d\sigma}\int\frac{d^4p}{(2\pi)^4}\frac{\mathrm{Tr}\left(\not{p}\gamma^\mu\not{p}\gamma^\nu\not{p}\gamma^\rho\not{p}\gamma^\sigma\right)}{(p^2-i\epsilon)^4}
\end{aligned}
\tag{5.56}
$$

The computation of the trace of the gamma matrices is left as an exercise. The result is

$$
\int\frac{d^4p}{(2\pi)^4}\Big[\mathrm{Tr}\ln\mathbb{M}_\psi(p)\Big]_{A^4} = -\frac{1}{3}\mathrm{Tr}_R(F_{\mu\nu}F^{\mu\nu})\frac{i}{8\pi^2}\ln\frac{\Lambda}{\mu}. \tag{5.57}
$$

Putting these together, including the sign for the fermion loop, and the factor of 1/2 for the gauge field determinant, we find the zero-derivative quartic term in the one-loop effective action to be

$$
\begin{aligned}
\left[\Gamma^{1-\mathrm{loop}}\right]_{A^4} &= \frac{1}{8\pi^2}\ln\frac{\Lambda}{\mu}\int d^4x\left[\left(\frac{1}{2}\cdot\frac{5}{3}+\frac{1}{12}\right)\mathrm{Tr}(F_{\mu\nu}F^{\mu\nu}) - \frac{1}{3}\mathrm{Tr}_R(F_{\mu\nu}F^{\mu\nu})\right]\\
&= \frac{1}{8\pi^2}\ln\frac{\Lambda}{\mu}\int d^4x\,F_{a\mu\nu}F_b^{\mu\nu}\left[\frac{11}{12}f_{acd}f_{bcd} - \frac{1}{3}\mathrm{Tr}_R(t^at^b)\right].
\end{aligned}
\tag{5.58}
$$

Here we have absorb the Yang-Mills coupling constant g into the structure constants f_{abc}. Generally, we have

$$f_{acd}f_{bcd} = g^2 C(\mathrm{adj})\delta_{ab}, \qquad \mathrm{Tr}_R(t^a t^b) = g^2 C(R)\delta_{ab}. \tag{5.59}$$

Here $C(R)$ is a constant that depends on the representation R. In fact it is the quadratic Casmir times $\dim R/\dim G$. For $G = SU(N)$, let $\mathbf{f}$ denote the N dimensional fundamental (defining) representation (in which the generators act as $N \times N$ traceless Hermitian matrices), $\bar{\mathbf{f}}$ the conjugate, anti-fundamental representation, and adj be the $N^2 - 1$ dimensional adjoint representation. We have

$$C(\mathrm{adj}) = N, \qquad C(\mathbf{f}) = C(\bar{\mathbf{f}}) = \frac{1}{2}. \tag{5.60}$$

Suppose there are N_f fermions in the fundamental (or anti-fundamental representation). The one-loop A^4 term in the effective action for the $SU(N)$ Yang-Mills theory coupled to N_f massless Dirac fermions is

$$\left[\Gamma^{1-\mathrm{loop}}\right]_{A^4} = \frac{g^2}{8\pi^2}\ln\frac{\Lambda}{\mu}\int d^4x\, F_{a\mu\nu}F_a^{\mu\nu}\left(\frac{11}{12}N - \frac{1}{6}N_f\right). \tag{5.61}$$

This corrects the A^4 term in the classical action, $-\frac{1}{4}F_{a\mu\nu}F_a^{\mu\nu}$ to

$$-\frac{1}{4}(1 + L_A)F_{a\mu\nu}F_a^{\mu\nu}. \tag{5.62}$$

We can rescale $A_{a\mu}$ by $(1 + L_A)^{-\frac{1}{2}}$ so that the kinetic term is canonically normalized, and then we see that the one-loop renormalized coupling is

$$\begin{aligned} g_R &= g\,(1 + L_A)^{-\frac{1}{2}} \\ &= g\left[1 + \frac{g^2}{4\pi^2}\left(\frac{11}{12}N - \frac{1}{6}N_f\right)\ln\frac{\Lambda}{\mu} + \mathcal{O}(g^4)\right]. \end{aligned} \tag{5.63}$$

This is an important result. When $N > \frac{2}{11}N_f$ (which is the case for QCD), the renormalized coupling is *greater* than the bare coupling, and increases as the cutoff scale Λ increases. Conversely, fixing the renormalized coupling g_R, the bare coupling g decreases as Λ increases. So far our definition of the bare coupling g is tied to the choice of regularization scheme. If we work in dimensional regularization, the $\ln\Lambda$ will be replaced by $1/(4-d)$, although the dependence on μ remains the same. The physical meaning of this "running" coupling will be clarified next.

6 Renormalization group

6.1 Evolution of the renormalized coupling constant

Our treatment of renormalization using perturbation theory so far suffers from an embarrassment: the renormalized coupling typically involves logarithms of the cutoff scale and the low energy physical scale (at which the renormalization point is defined). As the renormalization scale μ moves to a much larger or smaller value, the correction at each loop order becomes large. When the loop corrections to the renormalized coupling at the "old" energy scale becomes of the same order as the coupling itself, perturbation theory appears to breaks down, *even when the effective coupling stays small in the entire range of variation of the energy scale.*

The issue is that we have chosen to work with the bare coupling or the renormalized coupling defined at some fixed renormalization scale, whereas the effective coupling may change logarithmically with the energy scale of interactions. If the effective coupling g_E at some energy scale E is small, we should be able to calculate interactions at a *nearby* scale E' reliably, say corrections coming with a factor of $g_E \ln(E'/E)$, when $\ln(E'/E)$ is small or of order 1. As E' moves away from E, one may anticipate that the effective coupling at E' is still weak even if $g_E \ln(E'/E)$ is of order 1. But the traditional perturbation theory already breaks down at this point. However, if we vary the scale E' a little at a time, calculate the effective coupling at the new scale in terms of the one at the nearby old scale, then at each step perturbation theory is valid. This would allow us to extrapolate between two far separated scales E and E' as long as the true effective coupling stays small in the entire energy range from E to E'. This procedure amounts to a resummation of certain terms in the perturbation series to all order, and thus could be called an "improved perturbation theory".

Let us begin with a simple and concrete example, the ϕ^4 theory, with bare Lagrangian

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu \phi)^2 - \frac{1}{2}m^2\phi^2 - \frac{1}{4!}g\phi^4. \tag{6.1}$$

We will regularize the theory with a UV cutoff Λ, and define the renormalized coupling g_μ in terms of the *off-shell* four-point amplitude at a renormalization point $s = t = u = -\mu^2$, where s, t, u are the Mandelstam variables. Up to one-loop order, g_μ is given in terms of the bare coupling g by

$$\begin{aligned} g_\mu &= \mathcal{A}(s = t = u = -\mu^2) \\ &= g - \frac{3g^2}{32\pi^2}\int_0^1 dx \left[\ln \frac{\Lambda^2}{m^2 + \mu^2 x(1-x)} - 1\right] + \mathcal{O}(g^3). \end{aligned} \tag{6.2}$$

We may also express g_μ in terms of some other renormalization coupling say g_R defined by the off-shell amplitude at zero momentum, $s = t = u = 0$. The relation between g_μ and g_R is

$$g_\mu = g_R - \frac{3g_R^2}{32\pi^2} \int_0^1 dx \, \ln \frac{m^2}{m^2 + \mu^2 x(1-x)} + \mathcal{O}(g_R^3). \tag{6.3}$$

The one-loop correction becomes comparable to g_R itself when $g_R \ln(\mu/m)$ is of order 1, even if $g_R \ll 1$, and it would seem that we must take into account all order contributions already. But the coupling g_μ can still be weak at this point.

The efficient way to deal with this is to compute $g_{\mu'}$ in terms of the renormalized coupling g_μ at a nearby scale μ, and compose the relation between couplings at successive nearby scales in order to extrapolate the coupling at a far separated scale. From the one-loop result we have

$$g_{\mu'} = g_\mu - \frac{3g_\mu^2}{32\pi^2} \int_0^1 dx \, \ln \frac{m^2 + \mu^2 x(1-x)}{m^2 + \mu'^2 x(1-x)} + \mathcal{O}(g_\mu^3). \tag{6.4}$$

Taking μ' to be infinitesimally away from μ, we derive a differential equation

$$\mu \frac{dg_\mu}{d\mu} = \frac{3g_\mu^2}{16\pi^2} \int_0^1 dx \, \frac{\mu^2 x(1-x)}{m^2 + \mu^2 x(1-x)} + \mathcal{O}(g_\mu^3). \tag{6.5}$$

Suppose $\mu \gg m$, then this equation becomes simply

$$\mu \frac{dg_\mu}{d\mu} = \frac{3g_\mu^2}{16\pi^2} + \mathcal{O}(g_\mu^3). \tag{6.6}$$

Integrating it gives

$$\frac{1}{g_\mu} - \frac{1}{g'_\mu} = -\frac{3}{16\pi^2} \ln \frac{\mu}{\mu'}, \tag{6.7}$$

or

$$g_\mu = \frac{g_{\mu'}}{1 - \frac{3}{16\pi^2} g_{\mu'} \ln(\mu/\mu')}. \tag{6.8}$$

Now this relation between g_μ and $g_{\mu'}$ is valid for $g_{\mu'} \ln(\mu/\mu')$ of order 1, a significant improvement upon the ordinary perturbation theory at one-loop. For example, it predicts that even if we start with a vey small coupling $g_{\mu'}$ at scale μ', before we reach a large scale μ at which $g_{\mu'} \ln(\mu/\mu') = 16\pi^2/3$, the effective coupling becomes strong and perturbation theory breaks down. This is known as the Landau pole of ϕ^4 theory.

We could also keep nonzero m, and integrate (6.5) to

$$\frac{1}{g_\mu} - \frac{1}{g'_\mu} = -\frac{3}{32\pi^2} \int_0^1 dx \, \ln \frac{m^2 + \mu^2 x(1-x)}{m^2 + \mu'^2 x(1-x)}. \tag{6.9}$$

For instance, we can now derive an improved relation between g_μ and $g_R = g_{\mu'=0}$,

$$g_\mu = g_R \left[1 - \frac{3}{16\pi^2} g_R \int_0^1 dx \ln \left(1 + x(1-x) \frac{\mu^2}{m^2} \right) \right]^{-1}. \tag{6.10}$$

The differential equation (6.5) for the coupling constant as a function of the renormalization scale is known as the "renormalization group equation", first formulated by Gell-Mann and Low. Our renormalized coupling is defined essentially using the 1PI effective action (importantly, fields of all momenta are integrated out in the loops), and the resulting renormalization group (RG) equation is sometimes referred to as the 1PI RG equation. These are closely related to, but are not the same as, the "Wilsonian renormalization group", which we will discuss later.

Now let us consider the problem of computing off-shell correlation functions of a (generally composite) operator, denoted by $\mathcal{O}$. For example, in the scalar ϕ^4 theory, we can consider $\mathcal{O}(x) = \phi^n(x)$. Correlation functions of such operators generally receive quantum corrections involving logarithms of the energy scale, and in particular either diverges or goes to zero when expressed in terms of fixed bare couplings as the UV cutoff Λ is taken to infinity. We would like to consider a *renormalized* operator $\mathcal{O}_\mu$ at some energy scale μ, related to $\mathcal{O}$ by some μ-dependent factor,

$$\mathcal{O}_\mu(x) = N_\mu^{\mathcal{O}} \mathcal{O}(x). \tag{6.11}$$

Here $N_\mu^{\mathcal{O}}$ is chosen so that correlation functions of $\mathcal{O}_\mu(x)$ behave "nicely" (e.g. taking some fixed finite value) at a renormalization point defined by the scale μ. A priori, this prescription does not fix $N_\mu^{\mathcal{O}}$ in an unambiguous way, unless we specify the value of a particular correlation function of $\mathcal{O}_\mu$ at the renormalization point μ. In practice, the only important part of $N_\mu^{\mathcal{O}}$ is due to large logarithm in the energy scale. Once a scheme of specifying $N_\mu^{\mathcal{O}}$ is chosen, by comparing $N_\mu^{\mathcal{O}}$ at nearby scales μ and μ', one can derive an equation of the form

$$\mu \frac{d}{d\mu} N_\mu^{\mathcal{O}} = -\gamma^{\mathcal{O}}(g_\mu, m/\mu) N_\mu^{\mathcal{O}}. \tag{6.12}$$

Here $\gamma^{\mathcal{O}}$ is computed in terms of the renormalized coupling g_μ, and possibly depends on some mass parameter m of the theory. In high energy physics one is often interested in the limit $\mu/m \gg 1$, where the mass can be ignored. $\gamma^{\mathcal{O}}(g_\mu, m/\mu)$ will take some finite value as μ is varied, and can be calculated reliably in perturbation theory as long as g_μ is weak. Integrating the equation (6.12) then gives

$$N_E^{\mathcal{O}} = \exp\left[-\int^E \gamma^{\mathcal{O}}(g_\mu, m/\mu) \frac{d\mu}{\mu} \right] \tag{6.13}$$

If $\gamma^{\mathcal{O}}$ is close to being a constant in the energy range of interest, then we have

$$N_E^{\mathcal{O}} \sim \left(\frac{E}{\mu_0} \right)^{-\gamma^{\mathcal{O}}} \tag{6.14}$$

The exponent $\gamma^{\mathcal{O}}$ captures the effective scaling dimension of $\mathcal{O}$ when inserted into a correlation function.

Let us consider a concrete example, $\mathcal{O} = \phi^2$ in the ϕ^4 theory studied earlier. Inserting $\mathcal{O}$ into any correlation function, there are a series of loop corrections that can be absorbed by the renormalization of the $\mathcal{O}$ vertex. Let $\mathcal{O}(p)$ be the Fourier transform of $\mathcal{O}(x)$ into momentum space,

$$\mathcal{O}(p) = \int d^4x \, e^{-ip\cdot x} \mathcal{O}(x). \tag{6.15}$$

The correlator $\langle \mathcal{O}(p) \cdots \rangle$ then receives a one-loop correction factor

$$F(p) = 1 + \frac{1}{2}(-ig) \int \frac{d^4q}{(2\pi)^4} \frac{-i}{q^2 + m^2 - i\epsilon} \cdot \frac{-i}{(p-q)^2 + m^2 - i\epsilon}. \tag{6.16}$$

We are a little lucky in this case in that the one-loop correction is captured by a factor that depends only on the total inflowing momentum p, and not separately on the two momenta of the ϕ-lines coming out of the vertex $\mathcal{O}$. Generally we would have to specify the renormalization point more carefully in defining $F(p)$. The calculation of $F(p)$ is a familiar one in this case,

$$F(p) = 1 - \frac{g}{32\pi^2} \int_0^1 dx \left[\ln \frac{\Lambda^2}{m^2 + p^2 x(1-x)} - 1 \right]. \tag{6.17}$$

To cancel such (potentially large) logarithmic correction, we can choose

$$N_\mu^{\phi^2} = F(\mu)^{-1}, \tag{6.18}$$

where $F(\mu)$ is defined as $F(p)$ at $p^2 = \mu^2$. Replacing g by the renormalized coupling g_μ, and considering an infinitesimal change of μ on $N_\mu^{\phi^2}$ then gives the scaling exponent

$$\begin{aligned} \gamma^{\phi^2}(g_\mu, m/\mu) &= \frac{g_\mu}{16\pi^2} \int_0^1 dx \frac{\mu^2 x(1-x)}{m^2 + \mu^2 x(1-x)} + \mathcal{O}(g_\mu^2) \\ &\approx \frac{g_\mu}{16\pi^2} + \mathcal{O}(g_\mu^2), \quad \text{for} \ \ \mu \gg m. \end{aligned} \tag{6.19}$$

Let us consider another example, the Gell-Mann-Low equations for the gauge field renormalization and the electric charge in QED. We can define a renormalized $U(1)$ gauge field $A_\rho^{(\mu)}(x) = N_\mu^A A_\rho(x)$ by requiring its two-point function or Feynman propagator to be the canonical value at the renormalization point $p^2 = \mu^2$. The *renormalized* propagator for A_ρ in momentum space is

$$\langle A_\rho(p) A_\sigma(q) \rangle = (2\pi)^4 \delta^4(p+q) \left[\frac{\eta_{\rho\sigma}}{q^2 - \Pi(q^2) - i\epsilon} + q_\rho q_\sigma(\cdots) \right], \tag{6.20}$$

where $\Pi(q^2)$ is the photon self-energy. Our renormalization condition then requires

$$N_\mu^A = \left[1 - \mu^{-2}\Pi(\mu^2)\right]^{\frac{1}{2}}. \tag{6.21}$$

Recall that Z_3 was defined to be $[1 - \Pi'(0)]^{-1}$, so that residue of the propagator at $p^2 = 0$ is canonically normalized. Now we have chosen a different way to renormalize the photon field, adapted to an arbitrary renormalization scale μ. Using the one-loop photon self-energy, and varying N_μ^A in μ, we can derive

$$\gamma^A(e_\mu, m/\mu) = \frac{e_\mu^2}{2\pi^2}\int_0^1 dx \frac{\mu^2 x^2(1-x)^2}{m^2 + \mu^2 x(1-x)} + \mathcal{O}(e_\mu^4), \tag{6.22}$$

where m is the mass of the electron, and e_μ is the renormalized electric charge at scale μ. A priori, the definition of e_μ depends on the renormalization condition, which we have some freedom in choosing. This ambiguity would be reflected only in the $\mathcal{O}(e^4)$ term in γ^A. In QED, there is canonical normalization condition on the electric charge, namely we can define e_μ to be such that $e_\mu A_\rho^{(\mu)}$ to be independent of μ. It then follows that the "beta function" for the electric charge is

$$\beta(e_\mu) = \mu \frac{d}{d\mu} e_\mu = \gamma^A(e_\mu, m/\mu)\, e_\mu. \tag{6.23}$$

The ordinary fine structure constant $\alpha = e_R^2/4\pi$ is defined using the renormalized electric charge at $\mu = 0$, namely $e_R = e_{\mu=0}$. To compute renormalized electric charge e_μ at some general energy scale μ, we need to integrate the equation (6.23), which yields

$$\begin{aligned} \frac{1}{e_R^2} - \frac{1}{e_\mu^2} &= \frac{1}{\pi^2}\int_0^1 dx \int_0^\mu \frac{d\mu'}{\mu'} \frac{\mu'^2 x^2 (1-x)^2}{m^2 + \mu'^2 x(1-x)} + \mathcal{O}(e_\mu^2) \\ &= \frac{1}{2\pi^2}\int_0^1 dx\, x(1-x) \ln\left[1 + \frac{\mu^2}{m^2} x(1-x)\right] + \mathcal{O}(e_\mu^2). \end{aligned} \tag{6.24}$$

We know that e_R is small, and we are eventually interested in the regime $\mu \gg m$. For this purpose, it is not necessary to do the exact integral above, and we can simply patch up the solutions in two limits: (1) $e_R^2 \ln(\mu/m) \ll 1$, and (2) $\mu \gg m$, by matching them in the overlapping regime, that is when $\mu \gg m$ *and* $e_R^2 \ln(\mu/m) \ll 1$. The solution in the region (1) is

$$e_\mu = e_R + \frac{e_R^3}{12\pi^2}\left[\ln\frac{\mu}{m} - \frac{5}{6} + \mathcal{O}(\frac{m^2}{\mu^2})\right] + \mathcal{O}(e_R^5), \tag{6.25}$$

and the solution in the region (2) is

$$e_\mu = \left[\frac{1}{e_{\mu'}^2} - \frac{1}{6\pi^2}\ln\frac{\mu}{\mu'}\right]^{-\frac{1}{2}}. \tag{6.26}$$

Matching these two gives the answer at $\mu \gg m$ in terms of $e_R = e_{\mu=0}$,

$$e_\mu = e_R \left[1 - \frac{e_R^2}{6\pi^2}\left(\ln\frac{\mu}{m} - \frac{5}{6}\right)\right]^{-\frac{1}{2}}. \tag{6.27}$$

Given $\alpha = 1/137.036$ at $\mu = 0$, we may compute say at the scale of Z-boson mass $m_Z \approx 91$GeV, the running α_{m_Z} determined by Gell-Mann-Low equation takes the value 1/134.6. In reality the running of α_μ up to $\mu = m_Z$ receives contributions to its beta function from other charged particles besides the electron. The actual value of α_μ at $\mu = m_Z$ is 1/128.87.

6.2 Asymptotic behaviors

Let us recap our logic in the derivation of the renormalization group equation in the case of a single coupling constant $g(\mu)$. We started by defining the quantum field theory with a bare coupling g_B and a UV cutoff Λ, the latter is to be taken to infinity. We next compute the renormalized coupling $g(\mu)$, typically defined in terms of some off-shell amplitude at a renormalization point μ, to some loop order in perturbation theory. We then move to a slightly different scale μ', and re-express $g(\mu')$, previously computed using g_B, now in terms of $g(\mu)$ at the nearby scale μ. From this we derive a differential equation for $g(\mu)$, which is the Gell-Mann-Low equation.

This strategy can be restated in the following way. The bare coupling g_B is expressed in terms of $g(\mu)$, perturbatively, as (assuming all mass parameters $m \ll \mu$ and can be ignored)

$$g_B = g(\mu) + b\, g^2(\mu)\ln\frac{\Lambda}{\mu} + \mathcal{O}(g^3(\mu)) \tag{6.28}$$

We then vary μ, holding g_B fixed, and derive the equation for $g(\mu)$,

$$\mu\frac{d}{d\mu}g(\mu) = \beta(g(\mu)) = bg^2(\mu) + \mathcal{O}(g^3(\mu)). \tag{6.29}$$

Here we have assumed that the cutoff dependence cancel when $g(\mu)$ is expressed in terms of $g(\mu')$, which is the case when the theory is renormalizable with a single dimensionless coupling, such as ϕ^4 theory. For instance, going to the next order, we may have a log divergence and a $\log^2$ divergence in the two-loop diagram,

$$g_B = g(\mu) + b_1\, g^2(\mu)\ln\frac{\Lambda}{\mu} + c_1 g^3(\mu)\left(\ln\frac{\Lambda}{\mu}\right)^2 + b_2 g^3(\mu)\ln\frac{\Lambda}{\mu} + \mathcal{O}(g^4(\mu)) \tag{6.30}$$

The equation obtained by varying μ is

$$\mu\frac{d}{d\mu}g(\mu) = b_1 g^2(\mu) - 2b_1 g(\mu)\mu\frac{d}{d\mu}g(\mu)\ln\frac{\Lambda}{\mu} + 2c_1 g^3(\mu)\ln\frac{\Lambda}{\mu} + b_2 g^3(\mu) + \mathcal{O}(g^4(\mu))$$
$$= b_1 g_1^2(\mu) + 2(-b_1^2 + c_1)g^3(\mu)\ln\frac{\Lambda}{\mu} + b_2 g^3(\mu) + \mathcal{O}(g^4(\mu)) \tag{6.31}$$

For the Λ dependence to drop out, we must have $c_1 = b_1^2$. This is indeed what happens in ϕ^4 theory.

The idea of holding the bare coupling fixed while letting $g(\mu)$ vary with μ, is reformulated by Callan and Symanzik in terms of the off-shell Green's function. The Green's function of bare fields $\phi(x)$,

$$G_B(x_1, \cdots, x_n) = \langle 0|T\phi(x_1)\cdots\phi(x_n)|0\rangle, \tag{6.32}$$

in terms of the bare coupling g_B, is of course independent of the renormalization scale μ, by definition. When expressed in terms of the renormalized coupling $g(\mu)$, the dependence on the bare coupling, or counter term, also introduces explicit μ-dependence. Further, taking into account the field renormalization $\phi_\mu(x) = N_\mu\phi(x)$, we can write the Green's function of renormalized field in the form

$$G(x_1, \cdots, x_n; \mu, g(\mu)) = \langle 0|T\phi_\mu(x_1)\cdots\phi_\mu(x_n)|0\rangle, \tag{6.33}$$

The invariance of the bare Green function under the change of μ is then formally expressed as

$$\left[\mu\frac{\partial}{\partial\mu} + \frac{dg(\mu)}{d\mu}\frac{\partial}{\partial g} - n\mu\frac{dN_\mu}{d\mu}\right]G(x_1, \cdots, x_n; \mu, g) = 0, \tag{6.34}$$

or

$$\left[\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g} + n\gamma^\phi\right]G(x_1, \cdots, x_n; \mu, g) = 0. \tag{6.35}$$

This is called the Callan-Symanzik equation.

We have assumed the coupling $g(\mu)$ to be dimensionless so far. The formalism is straightforwardly generalized to several dimensionless couplings, $g^\ell(\mu)$. It can further be generalized to include dimensionful couplings as well. Suppose g_B^ℓ are dimensionful bare couplings, with mass dimension Δ_ℓ (multiplying an operator of dimension $4 - \Delta_\ell$ in the Lagrangian). We may define a corresponding dimensionless coupling $g^\ell(\mu)$, by taking appropriate power of μ times an amplitude at the renormalization point. The bare coupling is generally expressed in terms of $g^\ell(\mu)$'s by

$$g_B^\ell = \mu^{\Delta_\ell}\left[g^\ell(\mu) + \sum_{m,n} b_{mn}^\ell\, g^m(\mu)g^n(\mu)\ln\frac{\Lambda}{\mu} + \mathcal{O}(g^3(\mu))\right] \tag{6.36}$$

Varying μ while holding g_B^ℓ fixed, one obtains RG equation of the form

$$\mu \frac{d}{d\mu} g^\ell(\mu) = \beta^\ell(g(\mu)) = -\Delta_\ell g^\ell + \sum_{m,n} b^\ell_{mn} g^m g^n + \mathcal{O}(g^3). \tag{6.37}$$

The g_B^ℓ's can also include the bare mass m_B, in particular, and the corresponding $g^\ell(\mu)$ is then a renormalized dimensionless "mass" parameter, which may be defined using the renormalized two-point function at momentum scale μ. The precise definition of such a renormalized mass is not always physically meaningful. Nonetheless, it is useful to express physical amplitudes in terms of them.

The RG equations of the form (6.37) applies to renormalizable as well as non-renormalizable theories. The only difference is that in a renormalizable theory g^ℓ are a finite set of couplings, related to dimensionless or positive dimension bare couplings, whereas in a non-renormalizable theory one must consider all possible couplings allowed by the symmetries of the theory. If Δ_ℓ is negative, the renormalized dimensionless coupling $g^\ell(\mu)$ grows with μ, and perturbative expansion in g^ℓ breaks down at a sufficiently high energy scale.

Let us now discuss the possible high energy asymptotic behavior of the RG flow. To begin with, consider theories of a single (classically) dimensionless coupling g. There are several possibilities.

(1) Singularity at finite μ (Landau pole). Suppose $\beta(g)$ is positive for small positive g, then $g(\mu)$ increases with μ. If $\beta(g)$ grows sufficiently fast with increasing g, $g(\mu)$ could go to infinity at a finite energy scale μ_∞. This happens if

$$\ln \frac{\mu_\infty}{\mu} = \int_{g(\mu)}^\infty \frac{dg}{\beta(g)} < \infty. \tag{6.38}$$

In ϕ^4 theory, the one-loop beta function for the ϕ^4 coupling is $\beta(g) = \frac{3}{16\pi^2} g^2$. Naively plugging this in the the RG equation, the renormalized coupling blows up at the energy scale

$$\mu_\infty = \mu \, e^{\frac{16\pi^2}{3 g_\mu}}. \tag{6.39}$$

Similarly behavior occurs in QED using the one loop beta function. Of course, when the coupling g grows to be of order 1, the perturbative expansion breaks down and we can no longer use the one-loop beta function, so (6.39) cannot be taken seriously. Nonetheless, the positive beta function clearly indicates the breakdown of perturbation theory at a finite energy scale.

In the case of QED in the real world, new physics occurs at energies way below the scale of the Landau pole, and the UV behavior is entirely altered by the new physics.

(2) $g(\mu)$ could grow to infinity as μ goes to infinity, if $\beta(g)$ grows slowly enough with g. Once again, perturbation theory breaks down at some point, and it is not possible to distinguish this case from possibility (1) within the framework of perturbation theory.

(3) UV fixed point at finite coupling. Suppose $\beta(g)$ is positive for $0 < g < g_*$ but goes to zero at g_* and becomes negative for $g > g_*$. Then $g(\mu)$ will increase with μ for $0 < g < g_*$, decrease with μ for $g > g_*$, and in either case approach an RG fixed point $g = g_*$ as μ goes to infinity. When the coupling g is close to g_*, $\beta(g)$ can be approximated linearly in the coupling,

$$\beta(g) \approx \alpha(g_* - g), \qquad \alpha > 0. \tag{6.40}$$

In a neighborhood of g_*, we can solve the RG equation,

$$\mu\frac{dg}{d\mu} = \alpha(g_* - g) \quad \Rightarrow \quad g = g_* - C\mu^{-\alpha}. \tag{6.41}$$

Typically, the scaling exponent $\gamma^{\mathcal{O}}(g)$ for an operator $\mathcal{O}$ approaches a finite value $\gamma^{\mathcal{O}}(g_*) \equiv \gamma_*$ at $g = g_*$. Correlation functions of this operator at energy scale μ scales with the factor

$$N_\mu^{-1} \approx \exp\left[\int^\mu \gamma^{\mathcal{O}}(g_{\mu'})\frac{d\mu'}{\mu'}\right] \propto \mu^{\gamma_*} \tag{6.42}$$

in addition to the power of μ expected from the ordinary mass dimension of $\mathcal{O}$. γ_* is called the anomalous dimension of $\mathcal{O}$.

(4) Asymptotic freedom. Suppose $\beta(g)$ is negative at least for sufficiently small positive g. Then starting at some small value of g, $g(\mu)$ decreases with μ, and asymptotes to zero as $\mu \to \infty$. Thus the quantum field theory becomes *free* in the high energy limit. This is called asymptotic freedom.

Suppose for instance $\beta(g) \approx -b\,g^2$ at weak coupling. The RG equation is solved at weak coupling by

$$\frac{1}{g(\mu)} - \frac{1}{g(\mu_0)} = b\ln\frac{\mu}{\mu_0}, \tag{6.43}$$

or

$$g(\mu) = \frac{g(\mu_0)}{1 + b\,g(\mu_0)\ln\frac{\mu}{\mu_0}}. \tag{6.44}$$

While the coupling becomes weak at large μ, it becomes strong at low energies. At $\mu_\infty = \mu_0 e^{-\frac{1}{bg(\mu_0)}}$, (6.44) diverges, indicating breakdown of perturbation theory before the energy scale gets below μ_∞. Note that this scale is exponentially suppressed by the inverse of the coupling at μ_0, and is thus exponentially separated from the scale μ_0 at which the theory is weakly coupled.

Asymptotically free theories are the only case in the above four possibilities in which perturbation theory remains valid at arbitrarily high energies (in the case of UV fixed point, for the beta function to vanish it is necessary that contributions from different orders in perturbation theory cancel, and perturbation theory necessarily breaks down, unless there are other small expansion parameters). While the theory is completely well defined to arbitrarily high energies, the coupling constant $g(\mu)$ is not a true dimensionless parameter of the theory. Rather, the interaction strength of the theory is parameterized by the ratio between the energy scale μ of physical processes and the scale μ_∞ at which the theory becomes strongly coupled. In other words, the theory is defined by an energy scale μ_∞, rather than a coupling constant g. This is called "dimensional transmutation".

We have computed the renormalized effective coupling in the 1PI effective action of nonabelian gauge theory coupled to massless Dirac fermions at zero momentum. The result is IR divergent, and we have put in an IR cutoff μ. The logarithmic divergence in the UV cutoff Λ in the relation between the effective coupling and the bare coupling is universal however, and persists when we consider renormalized coupling at some nonzero finite energy scale μ. The relation between the bare coupling g_B and the renormalized coupling $g(\mu)$ is then

$$g_B = g(\mu) - \frac{g^3(\mu)}{4\pi^2}\left(\frac{11}{12}N - \frac{1}{6}N_f\right)\ln\frac{\Lambda}{\mu} + \cdots \tag{6.45}$$

From this we derive the one-loop beta function

$$\beta(g) = -\frac{g^3}{4\pi^2}\left(\frac{11}{12}N - \frac{1}{6}N_f\right) + \mathcal{O}(g^5) \tag{6.46}$$

When $N > \frac{2}{11}N_f$, the one-loop beta function is negative. As long as $g(\mu)$ is sufficiently weak at some energy scale μ, it will become smaller as μ increases, and the theory is asymptotically free. This is what happens in QCD of the real world, where $N = 3$, and there are 6 flavors of quarks. (6.46) was derived in the limit where the energy scale is much bigger than masses of the quarks. For instance, if we study scattering at energy scale lower than the top quark mass 174GeV but greater than the remaining 5 flavors of quarks (also assuming that $g(\mu)$ is weak at this scale), (6.46) would be approximately valid the number of flavors N_f taken to be 5. If there are some yet undiscovered heavy quark flavors, the beta function will change at energy scales higher than those heavy quarks. The $SU(3)$ gauge theory is asymptotically free as long as the total number of flavors is no more than 16.

On the other hand, QCD becomes strongly coupled at a low energy scale $\mu_\infty \sim$ 1GeV, and we can no longer speak of the asymptotic states as quarks and gluons separated at long distances, as the interaction strength becomes large at large distances.

It is believed (and observed) that the asymptotic states of QCD are color neutral particles (i.e. transforming in the singlet of $SU(3)$ gauge group), such as mesons and baryons. We will return to this point later.

Another possibility is that the two-loop (and higher loop) contribution to the beta function comes with the opposite sign, namely positive, which may alter the behavior of the renormalized coupling at low energies. This could happen when the number of (massless) flavors N_f is in a certain range below $\frac{11}{2}N$, and the one-loop and two-loop contributions to the beta function cancel at a coupling g_* of order 1. The theory with coupling g_* is a fixed point of the RG flow, called the Banks-Zaks fixed point.

Now let us discuss the case of multiple dimensionless (or dimensionful) couplings. Suppose the RG equation has a fixed point at $g^\ell = g^\ell_*$, namely

$$\beta^\ell(g_*) = 0 \tag{6.47}$$

for all ℓ. Near this fixed point, the beta functions behave as

$$\beta^\ell(g) \approx \sum_k M^\ell{}_k (g^k(\mu) - g^k_*). \tag{6.48}$$

The matrix $M^\ell{}_k$ may have positive and negative eigenvalues. As μ goes to infinity, $g^\ell(\mu)$ is attracted to g^ℓ_* along the directions of negative eigenvalues of M, and repelled from g^ℓ_* along the directions of positive eigenvalues. Suppose n_- is the number of negative eigenvalues of M. Then there is a hypersurface of dimension n_- in the space of coupling constants, such that starting at any point on this surface, $g^\ell(\mu)$ asymptotes to the fixed point g^ℓ_* at high energies. Even in non-renormalizable theories, where we need to taken into account infinitely many coupling constants, at a fixed point typically n_- is finite. A non-renormalizable theory that lies on such a surface in the space of couplings is called *asymptotically safe*. We will see examples of such quantum field theories in the next section.

6.3 Critical exponents

So far we have been mostly considering the asymptotic behavior of quantum field theories at high energies, which is particularly interesting in particle physics. Let us briefly discuss the low energy limit. If there are mass parameters in the theory, at very low energies the mass will dominate and the massive fields decouple from the low energy physics. Interesting low energy limit arises only when there are massless degrees of freedom. In the example of massless ϕ^4 theory in four dimensions, the beta function is positive and the renormalized coupling $g(\mu)$ goes to zero as $\mu \to 0$. In this case, while there are still massless degrees of freedom in the low energy limit, they are simply described by the free field ϕ.

The more interesting case is when there is an RG fixed point g_*, namely $\beta(g_*) = 0$. As described in the previous section, if the matrix $M^\ell{}_k$ has some negative eigenvalues, along the directions corresponding to these eigenvectors of $M^\ell{}_k$, going to low energies, the couplings run away from the fixed point g_*. The coupling parameters along these directions are called *relevant*.

On the other hand, the coupling parameters along the directions in the space couplings that corresponding to positive, or zero eigenvalues of $M^\ell{}_k$, are call *irrelevant* or *marginal*. Turning on an irrelevant coupling does not affect the low energy behavior of the theory, since $g(\mu)$ flows to g_* in the limit $\mu \to 0$.

Such RG fixed points generally occur in the effective field theories describing condensed matter systems at the point of a second-order phase transition. There are typically finitely many relevant coupling parameters of the system; for instance, they could be controlled by temperature, pressure, magnetic field, etc. Many examples of second order phase transition occur by adjust a single parameter, say the temperature T, to a critical temperature T_c. This indicates that there is 1 relevant coupling, corresponding to a negative-eigenvalue direction V^ℓ of the matrix $M^\ell{}_k$. Suppose the eigenvalue is $-\gamma$,

$$M^\ell{}_k V^k = -\gamma V^\ell. \tag{6.49}$$

Then an approximate solution of the RG equation near the fixed point along the relevant direction is

$$g^\ell(\mu) \approx (T - T_c) V^\ell \mu^{-\gamma} \tag{6.50}$$

where we have identified $T - T_c$ as the parameter that controls the deviation from the critical point. Exactly how $T - T_c$ enters this relation comes from the microscopic theory, and cannot be derived within the effective field theory. The exponent $-\gamma$, determined by the beta function near the critical point, on the other hand can be computed in the effective field theory. Correlation functions at energy scale μ, or distance scale $L = 1/\mu$, should be functions of the coupling $g^\ell(\mu)$, or equivalently functions of

$$(T - T_c) L^\gamma = \left(\frac{L}{\xi}\right)^\gamma. \tag{6.51}$$

Here we defined

$$\xi = (T - T_c)^{-1/\gamma}, \tag{6.52}$$

which is the characteristic correlation length. It diverges as T approaches T_c. $\nu = 1/\gamma$ is called the critical exponent, which can be measured in experiments.

Consider a d dimensional system near a second order phase transition, whose massless degrees of freedom are described by a single scalar field ϕ. In a ferromagnet, ϕ could be the local magnetization. We will assume the symmetry $\phi \to -\phi$, and so the only allowed couplings consistent with the symmetry are the coefficients g_{2n} of ϕ^{2n}, $n = 1, 2, \cdots$. Classically, ϕ has mass dimension $(d-2)/2$, and ϕ^{2n} has dimension $n(d-2)$. The coupling constants g_{2n} has dimension $d - n(d-2)$. In $d = 3$, there are two relevant coupling constants near the free theory point, g_2 (the mass) and g_4 (the ϕ^4 coupling). The RG flow starting from the free point by turning on g_4 has a fixed point in the IR. This is a strong coupling fixed point. Wilson and Fisher studied this fixed point by considering the same theory in $d = 4-\epsilon$ dimensions, where ϵ is regarded as a small expansion parameter, then then extrapolate the perturbative result in ϵ to $\epsilon = 1$. This method is called the "ϵ-expansion".

For $d = 4-\epsilon$, the bare coupling g_4^B has classical mass dimension ϵ. The renormalize coupling $g_4(\mu)$ can be defined as $\mu^{-\epsilon}$ times the four-point amplitude at $s = t = u = -\mu^2$. At one-loop, it is computed in terms of the bare coupling via

$$\begin{aligned} g_4(\mu) &= \mu^{-\epsilon}\left[g_4^B - \frac{3}{2}(g_4^B)^2 \int \frac{d^dq}{(2\pi)^d}\frac{1}{q^2(p-q)^2}\right] \\ &= \mu^{-\epsilon}\left[g_4^B - \frac{3}{2}(g_4^B)^2 \int_0^1 dx \int \frac{d^dq}{(2\pi)^d}\frac{1}{[q^2+\mu^2 x(1-x)]^2}\right] \\ &= \mu^{-\epsilon}\left[g_4^B - \frac{3}{2}(g_4^B)^2 (4\pi)^{-2+\frac{\epsilon}{2}}\frac{\Gamma(\frac{\epsilon}{2})\Gamma(1-\frac{\epsilon}{2})^2}{\Gamma(2-\epsilon)}\mu^{-\epsilon}\right], \end{aligned} \tag{6.53}$$

from which we derive

$$\mu\frac{d}{d\mu}g_4(\mu) = -\epsilon g_4(\mu) + \frac{3}{16\pi^2}g_4^2(\mu) + \cdots \tag{6.54}$$

where $\cdots$ are terms suppressed by powers of either ϵ or g_4. Note that up to this order, the beta function is the same as the one obtained from regularization by a UV cutoff Λ in 4 dimensions, after taking into account the classical scaling dimension of g_4, namely $-\epsilon$ in $d = 4-\epsilon$ dimensions. The higher loop beta function is generally regularization scheme dependent, due to the ambiguity in nonlinear reparameterization of the fields and couplings.

The renormalized coupling $g_2(\mu)$ may be defined by rewriting the mass term as $m_B^2\phi^2 = \mu^2 g_2(\mu)(N_\mu^{\phi^2}\phi^2)$, since the renormalization factor $N_\mu^{\phi^2}$ is chosen so that certain correlation functions of $\mathcal{O}_\mu^{\phi^2} = N_\mu^{\phi^2}\phi^2$ do not depend on μ (more precisely, correlation functions of $\mathcal{O}_\mu^{\phi^2}$ do not contain large logarithms from the renormalization of ϕ^2). $g_2(\mu)$ is then proportional to $\mu^{-2}(N_\mu^{\phi^2})^{-1}$, and thus obeys the RG equation (using the result of $N_\mu^{\phi^2}$ derived earlier)

$$\mu\frac{d}{d\mu}g_2(\mu) = -2g_2(\mu) + \frac{1}{16\pi^2}g_4(\mu)g_2(\mu) + \cdots \tag{6.55}$$

Up to this order, there is an RG fixed point at

$$(g_2)_* = 0, \qquad (g_4)_* = \frac{16\pi^2}{3}\epsilon. \tag{6.56}$$

The matrix $M^\ell{}_k$ for the two couplings $(g_2(\mu), g_4(\mu))$ is

$$M = \begin{pmatrix} -2 + \frac{1}{16\pi^2}(g_4)_* & \frac{1}{16\pi^2}(g_2)_* \\ 0 & -\epsilon + \frac{3}{8\pi^2}(g_4)_* \end{pmatrix} = \begin{pmatrix} -2 + \frac{\epsilon}{3} & 0 \\ 0 & \epsilon \end{pmatrix} + \mathcal{O}(\epsilon^2). \tag{6.57}$$

So at the RG fixed point, there is only one relevant coupling, namely $g_2(\mu)$, corresponding to the negative eigenvalue of the matrix M. Note that $g_4(\mu)$ which was relevant at the free point, becomes irrelevant at this nontrivial fixed point. The critical scaling dimension of g_2 is $-\gamma = -2 + \frac{\epsilon}{3}$, and the critical exponent is

$$\nu = \frac{1}{\gamma} = \frac{1}{2} + \frac{\epsilon}{12} + \mathcal{O}(\epsilon^2). \tag{6.58}$$

If we naively extrapolate the first order result in ϵ-expansion to $\epsilon = 1$, we obtain the critical exponential $\nu = \frac{7}{12} \approx 0.58$. (Alternatively, if we use the value $\gamma = \frac{5}{3}$ by keeping up to order $\mathcal{O}(\epsilon)$ contribution to γ, we obtain $\nu = \frac{1}{\gamma} = 0.6$.) This is surprisingly close to the experimentally measured value $\nu = 0.63 \pm 0.04$. It is left as an exercise to compute the order ϵ^2 correction to the critical exponent. The result is

$$\nu = \frac{1}{2} + \frac{\epsilon}{12} + \frac{7\epsilon^2}{162} + \mathcal{O}(\epsilon^3). \tag{6.59}$$

Extrapolating this to $\epsilon = 1$ gives $\nu \approx 0.627$.

6.4 A large N example

The Wilson-Fisher fixed point of scalar ϕ^4 theory discussed in the previous section has a generalization to the theory of N scalar fields. For reasons that will become clear shortly, the fixed point is in fact easier to understand when N is large. More precisely, let us begin with the (Euclidean) free field theory of N massless scalars ϕ_i, $i = 1, 2, \cdots, N$, in 3 dimensions. The Euclidean action is written as

$$S_E = \int d^3x \, \frac{1}{2}(\partial_\mu \phi_i)^2. \tag{6.60}$$

Classically, ϕ_i has mass dimension 1/2. There are two relevant operators that are invariant under the $O(N)$ symmetry that rotate the ϕ_i's. The operator $\mathcal{O}_1 = (\phi_i\phi_i)$ of dimension 1, and the operator $\mathcal{O}_2 = (\phi_i\phi_i)^2$ of dimension 2. If we deform the Lagrangian by the operator $\mathcal{O}_1$, we are simply adding a mass term for ϕ_i, and the theory becomes that of N free massive scalars. The more interested deformation is by the operator $\mathcal{O}_2$. Let us consider the deformed theory,

$$S_E = \int d^3x \left[\frac{1}{2}(\partial_\mu \phi_i)^2 + \frac{g}{2}(\phi_i \phi_i)^2 \right]. \tag{6.61}$$

In order for this theory to flow to a nontrivial fixed point under the RG flow to the infrared, we may need to add an appropriate bare mass term $\frac{1}{2}m_0^2\phi_i\phi_i$. The precise value of m_0 is regularization scheme dependent. Its value should be adjusted so that the physical mass of ϕ_i, which for instance can be determined from the two-point function $\langle \phi_i(x)\phi_j(y)\rangle$, is zero.

It is clear that any sort of fixed points would occur at strong coupling g. In the previous subsection we saw that this can be treated (in the $N = 1$ case) by ϵ-expansion. Here we will consider a different expansion: we will work still in three dimensions, but take N to be large, and consider the perturbative expansion in $1/N$. It is not immediately obvious that such an expansion is well defined. To see this, let us rewrite the path integral in an equivalent way by introducing auxiliary field $\sigma(x)$,

$$S_E = \int d^3x \left[\frac{1}{2}(\partial_\mu \phi_i)^2 + \frac{1}{2}\sigma\phi_i\phi_i - \frac{1}{8g}\sigma^2 \right]. \tag{6.62}$$

The functional integral over $\sigma(x)$ is a Gaussian integral, and after integrating out σ we get back to (6.61). You may be bothered by the wrong sign in front of σ^2. This is not essential, as we could redefine σ to $i\sigma$ (in other words, the functional integration "contour" should be along imaginary $\sigma(x)$).

Now the action in Gaussian in ϕ_i, and we can integrate out ϕ_i and obtain a functional determinant. If we are interested in correlation functions of ϕ_i, we can always add a source term $\int d^3x J_i(x)\phi_i(x)$, and perform the Gaussian integral with this linear term in ϕ_i included. For now we are interested in finding the RG fixed point, and so we won't bother with the source term.

After integrating out ϕ_i, the path integral becomes

$$\int D\phi_i\, D\sigma\, e^{-S_E} = \int D\sigma\, \exp\left[-\frac{N}{2}\mathrm{Tr}\ln(-\Box + \sigma(x)) + \int d^3x \frac{1}{8g}\sigma^2 \right]. \tag{6.63}$$

Writing $\lambda = gN$, the effective action in σ is

$$S_{eff}[\sigma] = \frac{N}{2}\mathrm{Tr}\ln(-\Box + \sigma(x)) - N\int d^3x \frac{\sigma^2}{8\lambda}. \tag{6.64}$$

Now we see that $1/N$ plays the role of $\hbar$. If λ is held fixed, then the perturbative expansion in $\hbar$ is equivalent to the expansion in $1/N$. At leading order in the large

N limit, we can calculate the path integral using the saddle point approximation, i.e. we substitute $S_{eff}[\sigma]$ by its value at a stationary point $\sigma(x)$. Presumably, such a stationary solution $\sigma(x)$ is translationally invariant, namely, $\sigma(x)$ takes a constant value, $\sigma(x) \equiv \sigma$. For constant σ, we need to extremize the action

$$\frac{N}{2}\int d^3x\left[\int \frac{d^3p}{(2\pi)^3}\ln(p^2+\sigma) - \frac{\sigma^2}{4\lambda}\right]. \tag{6.65}$$

Taking its derivative with respect to σ, we obtain the saddle point equation

$$\int \frac{d^3p}{(2\pi)^3}\frac{1}{p^2+\sigma} = \frac{\sigma}{2\lambda}. \tag{6.66}$$

The integral on the LHS appears to be UV divergent. This is the point where we need to choose a regularization scheme. Note that if we started with a bare mass m_0, the saddle point equation would be

$$\int \frac{d^3p}{(2\pi)^3}\frac{1}{p^2+\sigma+m_0^2} = \frac{\sigma}{2\lambda}. \tag{6.67}$$

We could then redefine $\overline{\sigma} = \sigma + m_0^2$, and write the above equation as

$$\int \frac{d^3p}{(2\pi)^3}\frac{1}{p^2+\overline{\sigma}} = \frac{\overline{\sigma}-m_0^2}{2\lambda}. \tag{6.68}$$

Here we see that m_0^2 can be chosen to cancel the power divergence $\sim \Lambda$, if we are to introduce a momentum cutoff Λ to regularize the p-integral. If we work with dimensional regularization, then the power divergence is absent and m_0^2 is not needed. In any case, we can subtract off the UV divergence by writing the equation as

$$\frac{\overline{\sigma}}{2\lambda} = \int \frac{d^3p}{(2\pi)^3}\left[\frac{1}{p^2+\overline{\sigma}} - \frac{1}{p^2}\right] = -\frac{\overline{\sigma}^{\frac{1}{2}}}{4\pi}, \tag{6.69}$$

This equation is obeyed for $\overline{\sigma} = 0$. Note that $\overline{\sigma}$ plays the role of mass squared for ϕ_i. Indeed if we are to compute the Green function of $\phi_i(x)$, we will find that the self energy of ϕ_i is equal to $\overline{\sigma}$ to leading order in the $1/N$ expansion. At the critical point, ϕ_i should be massless, and λ is sent to infinity.

Knowing that the saddle point value of $\overline{\sigma}$ is zero, we can then proceed and expand S_{eff} in $\overline{\sigma}$. We will drop the bar on $\overline{\sigma}$ from now, and simply write σ.

$$S_{eff} = -\frac{N}{2}\sum_{n=2}^{\infty}\frac{1}{n}\mathrm{Tr}\left[\frac{1}{\Box}\sigma(x)\right]^n \tag{6.70}$$

Note that in the above formula, $1/\Box$ is an operator that is multiplied by $\sigma(x)$ on the right; it doesn't act on $\sigma(x)$ only. In terms of the Fourier transformed field,

$$\sigma(x) = \int \frac{d^3p}{(2\pi)^3}e^{ip\cdot x}\widetilde{\sigma}(p), \tag{6.71}$$

we can write, say, the quadratic term in the effective action for σ as

$$\begin{aligned}
&-\frac{N}{4}\int\frac{d^3p}{(2\pi)^3}\frac{d^3q}{(2\pi)^3}\frac{1}{p^2}\langle p|\sigma(x)|q\rangle\frac{1}{q^2}\langle q|\sigma(x)|p\rangle\\
&=-\frac{N}{4}\int\frac{d^3p}{(2\pi)^3}\frac{d^3q}{(2\pi)^3}\frac{1}{p^2}\widetilde{\sigma}(p-q)\frac{1}{q^2}\widetilde{\sigma}(q-p)\\
&=-\frac{N}{4}\int\frac{d^3p}{(2\pi)^3}\widetilde{\sigma}(p)\widetilde{\sigma}(-p)\int\frac{d^3q}{(2\pi)^3}\frac{1}{q^2(p+q)^2}\\
&=-\frac{N}{32}\int\frac{d^3p}{(2\pi)^3}\widetilde{\sigma}(p)\widetilde{\sigma}(-p)\frac{1}{|p|}.
\end{aligned} \tag{6.72}$$

From this we see that the two point function of $\widetilde{\sigma}(p)$ is

$$\langle\widetilde{\sigma}(p)\widetilde{\sigma}(q)\rangle=-(2\pi)^3\delta(p+q)\cdot\frac{16}{N}|p|, \tag{6.73}$$

to leading order in the $1/N$ expansion. Fourier transforming back to position space, we have

$$\langle\sigma(x)\sigma(y)\rangle=-\int\frac{d^3p}{(2\pi)^3}e^{ip\cdot(x-y)}\frac{16|p|}{N}=\frac{16}{\pi^2N}\frac{1}{|x-y|^4}. \tag{6.74}$$

Here the p integral is evaluated using dimensional regularization. We see that $\sigma(x)$ behaves like a scalar operator of dimension 2. This statement is true only to leading order in the $1/N$ expansion. There are indeed nontrivial $1/N$ corrections to the exact scaling dimension of $\sigma(x)$.

Note that Gaussian integral over σ would have set (if we work in dimensional regularization in which m_0 is zero).

$$\sigma=2g\phi_i\phi_i. \tag{6.75}$$

So the scalar operator σ *is* the operator $\mathcal{O}_1=\phi_i\phi_i$, up to a constant rescaling. What we have learned here is that while $\mathcal{O}_1$ has scaling dimension 1 in the UV theory (free field theory of N massless scalars), it has scaling dimension 2 (plus $1/N$ corrections) in the IR fixed point theory.

6.5 An integrable example

In two space-time dimensions, there are examples of RG flows of interacting quantum field theories that are exactly solvable ("integrable", in a sense that we will describe below). One of the simplest examples is a completely massive deformation of a strongly interacting scaling invariant quantum field theory known as "3-state Potts model". The massive deformed theory is called the "scaling Potts model" (SPM). The UV theory can be obtained as the fine tuned critical point of the theory of a single massless scalar with ϕ^8 potential. We won't discuss this description here. We will be interested in an RG flow from the UV theory, the 3-state Potts model, to an IR theory which consists of interacting massive particles.

The entire RG trajectory of SPM is described by a strongly interacting massive quantum field theory. We won't bother with a Lagrangian description since we are not going to do perturbation theory. Instead, we will try to describe the *exact* S-matrix of this theory, and use it to learn something about the UV fixed point (the IR "fixed point" is empty because the theory is completely massive). This is possible thanks to a miraculous integrability property of the model: essentially, the theory has infinitely many conserved charges, which implies a particularly simple structure of its S-matrix elements.

The S-matrix of SPM is "factorized", in the following sense. The spectrum of single particle states will consist of only two types of particles: a particle A (which may be created by a complex scalar field), and its "anti-particle" $\overline{A}$. The scattering of n particles will only produce n-particle final states. In other words, the total particle number, including both A and $\overline{A}$, is preserved in any scattering process. Furthermore, the *set* of momenta of the asymptotic states, $p_1, p_2, \cdots, p_N$, are *unchanged* in a scattering process. The particle type may or may not change. What this means is that the $n \to n$ S-matrix can be written as an $2^n \times 2^n$ unitarity matrix $S_{IJ}(p_1, p_2, \cdots, p_n)$, where I, J label the types of the n particles in the out and in state, respectively.

We assert without proof a key property of such factorized S-matrix: the n-body S-matrix factorizes into the product of 2-body S-matrices. Imagine the n-body scattering occurs through $n(n-1)/2$ successive 2-body scatterings; the n-body S-matrix element is given by the product of the 2-body S-matrix elements, summed over the types of the intermediate particles. The resulting n-body S-matrix being independent of the choice of order of 2-body scatterings puts a nontrivial constraint on the 2-body S-matrix. This constraining relation is known as Yang-Baxter equation.

The S-matrix of SPM turns out to be an even simpler type: it is *totally elastic*, meaning that even the particle type cannot change in a scattering process. So the 2-body scattering can only take the form $A(p_1)A(p_2) \to A(p_1)A(p_2)$, $A(p_1)\overline{A}(p_2) \to A(p_1)\overline{A}(p_2)$, or $\overline{A}(p_1)\overline{A}(p_2) \to \overline{A}(p_1)\overline{A}(p_2)$. In other words, there is no reflection. The only nontrivial part of the S-matrix lies in the transmission phase. The 2-body S-matrix of SPM are given by:

$$\begin{aligned} S_{AA}(p_1,p_2) = S_{\overline{A}\overline{A}}(p_1,p_2) &= \frac{\sinh(\frac{\theta_{12}}{2}+\frac{\pi i}{3})}{\sinh(\frac{\theta_{12}}{2}-\frac{\pi i}{3})}, \\ S_{A\overline{A}}(p_1,p_2) = S_{\overline{A}A}(p_1,p_2) &= -\frac{\sinh(\frac{\theta_{12}}{2}+\frac{\pi i}{6})}{\sinh(\frac{\theta_{12}}{2}-\frac{\pi i}{6})}, \end{aligned} \tag{6.76}$$

where

$$\theta_{12} = \theta_1 - \theta_2, \quad (p_i^0, p_i^1) = (m\cosh\theta_i, m\sinh\theta_i). \tag{6.77}$$

θ_i is the rapidity of the on-shell momentum p_i in 1+1 dimensions. Lorentz boosts in 1+1 dimensions simply shifts θ_i by a constant, and leaves θ_{12} invariant. Therefore, the S-matrix elements above are consistent with Lorentz invariance.

They are also consistent with "crossing symmetry". Suppose we write, say $S_{AA}(p_1,p_2)$ as $S_{AA\overline{A}\overline{A}}(p_1,p_2,-p_1,-p_2)$, where all four momenta are taken to be ingoing, and the outgoing particle A of momentum p_i is replaced by the ingoing anti-particle of momentum $-p_i$. Crossing symmetry is the statement that the S-matrix is an analytic function of the momenta in cyclic ordering (except for poles and branch cuts that have interpretation via intermediate on-shell states), regardless of which ones come from ingoing particles and which ones are outgoing particles. For instance, $S_{AA}(p_1,p_2) = S_{AA\overline{A}\overline{A}}(p_1,p_2,-p_1,-p_2)$ should be the same as $S_{A\overline{A}\overline{A}A}(p_2,-p_1,-p_2,p_1) = S_{A\overline{A}}(p_2,-p_1)$. Associated with the momenta p_2, $-p_1$ are the rapidities θ_2 and $\theta_1-\pi i$. So cross symmetry implies the relation

$$S_{AA}(\theta_{12}) = S_{A\overline{A}}(\pi i - \theta_{12}). \tag{6.78}$$

This is indeed obeyed by (6.76). In fact, via crossing symmetry, any one particular 2-body S-matrix element above determines the rest.

The 2-body S-matrix (6.76) is also unitary:

$$S(p_1,p_2) = (S(p_2,p_1))^\dagger, \qquad S(p_1,p_2)S(p_2,p_1) = \mathbf{1}. \tag{6.79}$$

We see that there is a pole in $S_{AA}(\theta_{12})$ at

$$\theta_{12} = \frac{2\pi i}{3}. \tag{6.80}$$

In other words, the pole occurs when $\cosh\theta_{12} = -\frac{1}{2}$. This corresponds to invariant mass squared

$$-(p_1+p_2)^2 = 2m^2(1+\cosh\theta_{12}) = m^2. \tag{6.81}$$

This pole in fact corresponds to the intermediate state that is the anti-particle $\overline{A}$ (which has mass m).

We have not derived any of these statements about the S-matrix of the SPM model. For now we are just going to take this set of consistent S-matrix elements as what defines the SPM model, and try to understand the behavior of the theory in the UV. One question we could ask about the UV theory is how many degrees of freedom it has. One way to measure the number of degrees of freedom of a field theory is to compute its thermal free energy (say in a large box, or interval, in the case of 1+1 dimensions). In particular, we will try to compute the free energy of SPM at finite temperature, and see how it behaves in the high temperature limit.

To begin with let us assumes that there is one type of particle, of mass m, with factorized purely elastic S-matrix. The 2-body S-matrix is given by

$$S(p_1, p_2) = S(\theta_{12}). \tag{6.82}$$

It obeys unitarity $S(\theta)S(-\theta) = 1$ and crossing symmetry $S(\theta) = S(\pi i - \theta)$.

Now consider the system such particles in a 1-dimensional box of length L, with periodic boundary condition. In the limit $L \gg 1/m$, we can describe the energy eigenstates in terms of the scattering states. An N particle state may be described by the N-particle wave function

$$\Psi(x_1, x_2, \cdots, x_N). \tag{6.83}$$

Let us assume these are identical bosons, and so Ψ is symmetric in $x_1, \cdots, x_N$. A priori, the energy eigenstates may not have a definite particle number. What we know (or assumed) about the factorized S-matrix implies that, at least when $x_1, \cdots, x_N$ are far separated, the scattering state of N particles should be a good approximation of an energy eigenstate.

Suppose $x_1 < x_2 < \cdots < x_N$. The scattering state wave function with the N particles then takes the plane wave form

$$\Psi(x_1, x_2, \cdots, x_N) = \sum_{\sigma \in S_N} C_\sigma(p_1, \cdots, p_N)\, e^{ip_{\sigma(1)}x_1 + ip_{\sigma(2)}x_2 + \cdots ip_{\sigma(N)}x_N} \tag{6.84}$$

where $x_1, \cdots, x_N$ are far separated. The coefficients $C_\sigma(p_1, \cdots, p_N)$ are precisely the S-matrix element for the scattering (of N or fewer particles) that corresponds to the permutation σ on the momenta $p_1, \cdots, p_N$.

Since the x_i's live on a box with periodic boundary condition (i.e. a circle), we must have

$$\Psi(x_1, x_2, \cdots, x_N) = \Psi(x_2, x_3, \cdots, x_N, x_1 + L). \tag{6.85}$$

It follows that

$$e^{ip_1 L} \prod_{j=2}^{N} S(p_1, p_j) = 1. \tag{6.86}$$

An analogous equation applies to each p_i, namely,

$$e^{ip_iL}\prod_{j=1,\ j\neq i}^{N}S(p_i,p_j)=1,\qquad i=1,2,\cdots,N. \tag{6.87}$$

This is called the Bethe ansatz equation (BAE). Writing $S(\theta)=\exp(i\phi(\theta))$, and taking the logarithm on both sides of the BAE, we may write it as

$$mL\sinh\theta_i+\sum_{j\neq i}\phi(\theta_i-\theta_j)=2\pi n_i,\qquad n_i\in\mathbb{Z}. \tag{6.88}$$

The integers n_i now label energy eigenstates of the system in the periodic box. The total energy is simply the sum of each particle in the asymptotic state,

$$H=\sum_{i=1}^{N}m\cosh\theta_i. \tag{6.89}$$

What are the allowed values of n_i? Essentially, all possible assignments of n_i are allowed, except that we have to be careful when a pair of n_i's coincide. When the latter occurs, the BAE may or may not have a solution depending on whether $\phi(0)$ is equal to zero. In fact, by unitarity, $S(0)^2=1$, and $S(0)$ must be $+1$ or -1. In the scaling Potts model, $S(0)=-1$ and $\phi(0)=\pi$. In this case, the n_i's must all be distinct. So even though the particles are bosons, the way they occupy the integer "levels" n_i are fermion-like.

Now we are going to take a limit of large number of particles N. The scattering state can be characterized by the rapidity density function $\rho_1(\theta)$, i.e. there are $\rho_1(\theta)d\theta$ particles of rapidity between θ and $\theta+d\theta$. The levels n_i can be thought of as a function $n(\theta)$ of θ. Define the level density function

$$\rho(\theta)=\frac{dn(\theta)}{d\theta}. \tag{6.90}$$

The energy of the state is expressed in terms of $\rho_1(\theta)$,

$$H=m\int\cosh\theta\rho_1(\theta)d\theta, \tag{6.91}$$

where the BAE expresses ρ in terms of ρ_1,

$$mL\cosh\theta+\int\frac{d\phi(\theta-\theta')}{d\theta}\rho_1(\theta')d\theta'=2\pi\rho(\theta). \tag{6.92}$$

Next we ask how many states there are with given $\rho(\theta)$ and $\rho_1(\theta)$. The number of "available" levels corresponding to rapidities between θ and $\theta+\Delta\theta$ is $\Delta N\simeq\rho(\theta)\Delta\theta$.

The actual number of occupied levels is $\Delta n \simeq \rho_1(\theta)\Delta\theta$. Since no two particles can occupy the same level, as already discussed, the number of possible occupations for the particles of rapidity in the range θ to $\theta + \Delta\theta$ is

$$\frac{(\Delta N)!}{(\Delta n)!(\Delta N - \Delta n)!} \tag{6.93}$$

It contributes to the entropy S of the system by

$$\begin{aligned} \ln \frac{(\Delta N)!}{(\Delta n)!(\Delta N - \Delta n)!} &\simeq (\Delta N)\ln(\Delta N) - (\Delta n)\ln(\Delta n) - (\Delta N - \Delta n)\ln(\Delta N - \Delta n) \\ &\simeq d\theta \left[\rho \ln \rho - \rho_1 \ln \rho_1 - (\rho - \rho_1)\ln(\rho - \rho_1)\right]. \end{aligned} \tag{6.94}$$

The thermal free energy of the system at temperature T is therefore obtained by minimizing

$$\mathcal{F}[\rho_1] = H - TS = m \int \cosh\theta \rho_1(\theta) d\theta - T \int d\theta \left[\rho \ln \rho - \rho_1 \ln \rho_1 - (\rho - \rho_1)\ln(\rho - \rho_1)\right] \tag{6.95}$$

in ρ_1, with ρ understood as a functional of ρ_1 via the BAE. The variational equation is

$$\frac{m}{T}\cosh\theta + \ln \frac{\rho_1(\theta)}{\rho(\theta) - \rho_1(\theta)} - \int d\theta' \frac{\delta\rho(\theta')}{\delta\rho_1(\theta)} \ln \frac{\rho(\theta')}{\rho(\theta') - \rho_1(\theta')} = 0. \tag{6.96}$$

Using

$$\frac{\delta\rho(\theta')}{\delta\rho_1(\theta)} = \frac{1}{2\pi}\phi'(\theta' - \theta) = \frac{1}{2\pi}\phi'(\theta - \theta'), \tag{6.97}$$

where the second equality follows from unitarity, we have

$$\frac{m}{T}\cosh\theta + \ln \frac{\rho_1(\theta)}{\rho(\theta) - \rho_1(\theta)} - \int \frac{d\theta'}{2\pi} \frac{d\phi(\theta - \theta')}{d\theta} \ln \frac{\rho(\theta')}{\rho(\theta') - \rho_1(\theta')} = 0. \tag{6.98}$$

It is conventional to define a pseudoenergy ε via

$$\rho_1 = \frac{\rho}{e^{\varepsilon} + 1}, \tag{6.99}$$

so that (6.98) can be written as

$$\frac{m}{T}\cosh\theta - \varepsilon(\theta) - \int \frac{d\theta'}{2\pi} \frac{d\phi(\theta - \theta')}{d\theta} \ln\left[1 + e^{-\varepsilon(\theta')}\right] = 0. \tag{6.100}$$

This is called the thermodynamic Bethe ansatz equation (TBA). Once we solve $\varepsilon(\theta)$

from the TBA equation, the free energy of the system is given by

$$\begin{aligned}\mathcal{F} &= \int d\theta\rho_1\left[m\cosh\theta + T\ln\frac{\rho_1}{\rho-\rho_1}\right] - T\int d\theta\rho\ln\frac{\rho}{\rho-\rho_1}\\ &= T\int d\theta\rho_1(\theta)\int\frac{d\theta'}{2\pi}\phi'(\theta-\theta')\ln\frac{\rho(\theta')}{\rho(\theta')-\rho_1(\theta')} - T\int d\theta\rho\ln\frac{\rho}{\rho-\rho_1}\\ &= T\int\frac{d\theta'}{2\pi}(2\pi\rho(\theta') - mL\cosh\theta')\ln\frac{\rho(\theta')}{\rho(\theta')-\rho_1(\theta')} - T\int d\theta\rho\ln\frac{\rho}{\rho-\rho_1}\\ &= -TLm\int\frac{d\theta}{2\pi}\cosh\theta\ln\left[1+e^{-\varepsilon(\theta)}\right].\end{aligned} \tag{6.101}$$

Starting from the factorized 2-body S-matrix, we know the scattering phase $\phi(\theta)$ as a function of the rapidity θ. From the TBA equation, we can solve for the pseudoenergy $\varepsilon(\theta)$. Once we know $\varepsilon(\theta)$, it is straightforward to compute the free energy at finite temperature.

So far we have formulate the TBA equation in the case of one type of particle. When the S-matrix is purely elastic, the TBA equation is straightforwardly generalized to the case of several types of particles, of mass m_a, where a is an index that labels the type of the particle. All we need to do is to promote $\phi(\theta)$ to a matrix worth of phases $\phi_{ab}(\theta)$, which is $-i$ times the logarithm of the 2-body scattering amplitude $S_{ab}(\theta)$. Note that the subscripts ab labels the two incoming particles, which are the same as the two outgoing particles. There will be a pseudoenergy $\varepsilon_a(\theta)$ associated with the a-th type of particles. In the TBA equation, the convolution between $\phi'(\theta)$ and $\ln(1+e^{-\varepsilon})$ is now replaced by a matrix times a column vector. Namely,

$$\frac{m_a}{T}\cosh\theta - \varepsilon_a(\theta) - \int\frac{d\theta'}{2\pi}\phi'_{ab}(\theta-\theta')\ln(1+e^{-\varepsilon_b(\theta')}) = 0. \tag{6.102}$$

The formula for the free energy then involves a sum over the contribution from each type of particles.

Let us now apply the TBA equation to the SPM model. Since it is the derivative of the phase $\phi_{ab}(\theta) = -i\ln S_{ab}(\theta)$ that enters the TBA equation, we will use the formulae

$$\begin{aligned}\phi'_{AA}(\theta) = \phi'_{\overline{AA}}(\theta) &= -\frac{\sqrt{3}}{2\cosh\theta+1},\\ \phi'_{A\overline{A}}(\theta) = \phi'_{\overline{A}A}(\theta) &= -\frac{\sqrt{3}}{2\cosh\theta-1}.\end{aligned} \tag{6.103}$$

Let us define

$$G_a(\theta) = \ln(1+e^{-\varepsilon_a(\theta)}). \tag{6.104}$$

We then have

$$\begin{aligned}&\frac{m}{T}\cosh\theta - \varepsilon_a - \phi'_{ab} * G_b = 0,\\&\mathcal{F} = -TLm\int\frac{d\theta}{2\pi}\cosh\theta\sum_a G_a(\theta).\end{aligned}\tag{6.105}$$

Here the convolution product $*$ is defined as

$$f * g(\theta) = \int\frac{d\theta'}{2\pi}f(\theta')g(\theta-\theta').\tag{6.106}$$

For the SPM model, there is a symmetric solution with $\varepsilon_A = \varepsilon_{\overline{A}} \equiv \varepsilon$, that obeys

$$\frac{m}{T}\cosh\theta - \varepsilon - (\phi'_{AA} + \phi'_{A\overline{A}}) * G = 0,\tag{6.107}$$

with

$$\phi'_{AA} + \phi'_{A\overline{A}} = -2\sqrt{3}\frac{\sinh(2\theta)}{\sinh(3\theta)},\tag{6.108}$$

and then

$$\mathcal{F} = -2TLm\int\frac{d\theta}{2\pi}\cosh\theta G(\theta).\tag{6.109}$$

The TBA equation clearly looks like a horrendously complicated non-linear integral equation. Here we won't need the full solution to this equation. We are interested in the UV limit, i.e. $T\to\infty$. For a finite range of θ, the TBA equation reduces to

$$\varepsilon - 2\sqrt{3}\frac{\sinh(2\theta)}{\sinh(3\theta)} * G = 0.\tag{6.110}$$

This is solved by a contant $\varepsilon(\theta) = \varepsilon_0$ (and $G(\theta) = G_0$),

$$\begin{aligned}&\varepsilon_0 - 2\sqrt{3}\int\frac{d\theta}{2\pi}\frac{\sinh(2\theta)}{\sinh(3\theta)}\ln(1+e^{-\varepsilon_0}) = \varepsilon_0 - \ln(1+e^{-\varepsilon_0}) = 0\\&\Rightarrow\ \varepsilon_0 = \ln\frac{1+\sqrt{5}}{2} = G_0.\end{aligned}\tag{6.111}$$

The approximation $\varepsilon(\theta)\approx\varepsilon_0$ break down when θ is sufficiently large. When $\theta\gg 1$, the TBA equation becomes

$$\frac{m}{2T}e^{\theta} - \varepsilon + 2\sqrt{3}\frac{\sinh(2\theta)}{\sinh(3\theta)} * G = 0.\tag{6.112}$$

The contribution from this range of rapidity to the free energy is approximated by

$$-TLm\int_{\theta_0}^{\infty}\frac{d\theta}{2\pi}e^{\theta}G(\theta).\tag{6.113}$$

for some $\theta_0 \gg 1$. Now a bit of magic: taking the derivative of (6.112) with respect to θ, we have

$$\frac{m}{2T}e^{\theta} = \frac{d\varepsilon}{d\theta} - 2\sqrt{3}\frac{\sinh(2\theta)}{\sinh(3\theta)} * \frac{dG}{d\theta} = 0, \tag{6.114}$$

where we used an obvious property of the derivative of the convolution product. We can replace the e^{θ} inside the integral of (6.113) using (6.114), and obtain

$$\begin{aligned}
&-\frac{T^2L}{\pi}\int_{\theta_0}^{\infty} G(\theta)\left[\frac{d\varepsilon}{d\theta} - 2\sqrt{3}\frac{\sinh(2\theta)}{\sinh(3\theta)} * \frac{dG}{d\theta}\right]d\theta \\
&= -\frac{T^2L}{\pi}\left[\int_{\theta_0}^{\infty} G(\theta)\frac{d\varepsilon}{d\theta}d\theta - 2\sqrt{3}\int_{-\infty}^{\infty}\frac{d\theta'}{2\pi}\frac{\sinh(2\theta')}{\sinh(3\theta')}\int_{\theta_0}^{\infty} d\theta G(\theta)\frac{dG(\theta-\theta')}{d\theta}\right] \\
&\approx -\frac{T^2L}{\pi}\left[\int_{\varepsilon_0}^{\infty}\ln(1+e^{-\varepsilon})d\varepsilon + 2\sqrt{3}\int_{-\infty}^{\infty}\frac{d\theta'}{2\pi}\frac{\sinh(2\theta')}{\sinh(3\theta')}\frac{G_0^2}{2}\right] \\
&\approx -\frac{T^2L}{\pi}\left[\int_{\varepsilon_0}^{\infty}\ln(1+e^{-\varepsilon})d\varepsilon + \frac{\varepsilon_0^2}{2}\right].
\end{aligned} \tag{6.115}$$

We are then left with the contribution to the free energy

$$-\frac{T^2L}{\pi}\left[\int_{\varepsilon_0}^{\infty}\ln(1+e^{-\varepsilon})d\varepsilon - \int_{\varepsilon_0}^{\infty} d\varepsilon\frac{\ln(1+e^{-\varepsilon})}{e^{\varepsilon}+1}\right] = -\frac{\pi T^2 L}{15}. \tag{6.116}$$

The total free energy in the high temperature limit is twice of (6.116), from the integral over large positive θ and large negative θ (the contribution from the integral over θ of order 1 does not scale like T^2),

$$\mathcal{F} \approx -\frac{2\pi}{15}T^2L. \tag{6.117}$$

Remarkably, we find that the high temperature limit of the free theory of the SPM is independent of the mass parameter m. Generally, if we have a quantum field theory with no mass parameter, say in $1+1$ dimensions, the free energy at finite temperature, being extensive, must be proportional to L. By dimension analysis, then, it must have the temperature dependence T^2. The (negative) coefficient of T^2L is a number that characterizes the number of degrees of freedom of the theory. Generally, we can write the free energy of such a theory as $\mathcal{F} = -\pi c T^2 L/6$, where the constant c is call the central charge. We will see a different definition of the central charge later. A free massless scalar field in $1+1$ dimensions contributes $c = 1$ to the free energy[5], while a free massless real fermion in $1+1$ dimensions contributes $c = 1/2$. We see that the high temperature limit of the SPM has $c = 4/5$. This is a strongly interacting,

[5]There is an IR divergence in the free energy of a massless scalar field in $1+1$ dimensions, which may be regularized by adding a small mass. This IR divergent contribution is due to the constant mode of the scalar, and is not extensive.

scale invariant quantum field theory, known as one of the "minimal models" in two space-time dimensions. What we have learned here is that the scaling Potts model describes the renormalization group flow from the $c = 4/5$ minimal model in the UV to a completely massive theory in the IR.

6.6 Operator product expansion

The product of two local field operators $\mathcal{O}_i(x)$ and $\mathcal{O}_j(y)$ may be expanded on a linear basis of local operators, in the form

$$\mathcal{O}_i(x)\mathcal{O}_j(y) \to \sum_k F_{ij}^k(x-y)\mathcal{O}_k(y), \tag{6.118}$$

when x approaches y. This is call the operator product expansion (OPE). Let Δ_i be the mass scaling dimension of $\mathcal{O}_i$, then $F_{ij}^k(x-y)$ should scale like $(x-y)^{\Delta_k-\Delta_i-\Delta_j}$. In particular, higher dimensional operators come with less singular coefficients in the OPE as $x \to y$.

Let us begin with the theory of a *free* massless scalar field ϕ in four Euclidean dimensions. We can write

$$\begin{aligned} \phi(x)\phi(y) &= \frac{1}{4\pi(x-y)^2} + \sum_{k=0}^{\infty} \frac{1}{k!}(x-y)^{\mu_1}\cdots(x-y)^{\mu_k}\mathcal{O}_{\mu_1\cdots\mu_k}(y), \\ \mathcal{O}_{\mu_1\cdots\mu_k}(y) &= \lim_{z\to y}\partial_{z^{\mu_1}}\cdots\partial_{z^{\mu_k}}\left[\phi(z)\phi(y) - \frac{1}{4\pi(z-y)^2}\right]. \end{aligned} \tag{6.119}$$

In defining the local operator $\mathcal{O}_{\mu_1\cdots\mu_k}(y)$, we have subtracted the singularity $\sim (z-y)^{-2}$ that arises from the Wick contraction between $\phi(z)$ and $\phi(y)$ in a correlation function, so that the limit $z \to y$ gives to a well defined operator (i.e. one with well defined correlation functions). In Lorentzian signature, we can consider the time ordered OPE,

$$T\phi(x)\phi(y) = \Delta_F(x-y) + \sum_{k=0}^{\infty}\frac{1}{k!}(x-y)^{\mu_1}\cdots(x-y)^{\mu_k}\left[\partial_{z^{\mu_1}}\cdots\partial_{z^{\mu_k}} : \phi(z)\phi(y) : \right]\Big|_{z=y}. \tag{6.120}$$

The set of operators appearing in this OPE are identity operator and the normal ordered product of ϕ with its derivatives.

In momentum space, we can write the general time ordered OPE in the form

$$\int d^4x e^{-ik\cdot x} T\mathcal{O}_i(x)\mathcal{O}_j(0) \to \sum_\ell U_{ij}^\ell(k)\mathcal{O}_\ell(0) \tag{6.121}$$

in the limit of large momentum k. Note that the free OPE considered above, apart from the term proportional to identity operator, only gives rise to contact terms in $U_{ij}^\ell(k)$ and no contribution at large k, in momentum space.

Now let us consider the $\phi(x)\phi(0)$ OPE in ϕ^4 theory, with the Lagrangian

$$\mathcal{L} = -\frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m\phi^2 - \frac{1}{24}g\phi^4. \tag{6.122}$$

In the large k limit, we expect the timed ordered OPE of the renormalized field $\phi_R = Z^{-\frac{1}{2}}\phi$ to take the form

$$\int d^4x e^{-ik\cdot x} T\left[\phi_R(x)\phi_R(0)\right]_C \to U_{\phi^2}(k)(\phi^2(0))_R, \tag{6.123}$$

where $(\phi^2(0))_R = N^{\phi^2}\phi^2(0)$ is the lowest dimension operator that appear in the OPE and hence dominates at large k, N^{ϕ^2} is a renormalization constant that makes the correlation functions of $(\phi^2)_R$ finite. The subscript C stands for the connected part, i.e. dropping the term proportional to identity operator in the OPE.

The OPE coefficient $U_{\phi^2}(k)$ can be computed by taking the correlation function of both sides of (6.123) with $\widetilde{\phi}(-p)\widetilde{\phi}(-q)$. At tree level, the correlation function of the LHS of (6.123) with $\widetilde{\phi}(-p)\widetilde{\phi}(-q)$ gives

$$\frac{-ig}{(k^2+m^2)((p+q-k)^2+m^2)(p^2+m^2)(q^2+m^2)} \tag{6.124}$$

while the correlation function involving the RHS of (6.123) gives

$$U_{\phi^2}(k)(2\pi)^4\frac{-i}{p^2+m^2}\frac{-i}{q^2+m^2} \tag{6.125}$$

Comparing the two in the $k\to\infty$ limit gives

$$U_{\phi^2}(k) = \frac{ig}{(2\pi)^4}\frac{1}{(k^2)^2} + \mathcal{O}(g^2). \tag{6.126}$$

Carrying this computation to order g^2 requires the definition of $(\phi^2)_R$ at a renormalization point μ. This is left as an exercise.

The OPE coefficients $U_{ij}^\ell(k)$ generally depend on the renormalization scale μ, which enters through the definition of the renormalized operators $\mathcal{O}_i = N^{\mathcal{O}_i}\mathcal{O}_i^{bare}$, as well as the running coupling constant g that enters the computation of $U_{ij}^\ell(k)$ through correlation functions. Here we are assuming that the basis of operators $\mathcal{O}_i$ diagonalizes renormalization group flow, otherwise $N^{\mathcal{O}_i}$ will be replaced by a matrix acting on the set of operators that mix under RG. The OPE coefficients $U_{ij}^\ell(k;g_\mu,\mu)$ then obeys the RG equation

$$\mu\frac{d}{d\mu}U_{ij}^\ell = (\gamma_\ell-\gamma_i-\gamma_j)U_{ij}^\ell + \beta(g)\frac{\partial}{\partial g}U_{ij}^\ell. \tag{6.127}$$

This is closely related to Callan-Symanzik equation. Here we will work in the convention where $N^{\mathcal{O}_i}$ contains a factor μ^{-d_i}, where d_i is the classical dimension of $\mathcal{O}_i^{bare}$, so that the renormalized operator is dimensionless. U^ℓ_{ij} then has dimension -4 due to the Fourier transform to momentum space. If we write $k^\nu = \kappa n^\nu$, where n^ν is a fixed four-vector, we can write $U^\ell_{ij}(\kappa n; g_\mu, \mu) = \kappa^{-4} U^\ell_{ij}(n; g_\mu, \mu/\kappa)$. Integrating the RG equation gives

$$U^\ell_{ij}(\kappa n; g_\mu, \mu) = \kappa^{-4} \exp\left[\int_\mu^\kappa \frac{d\mu'}{\mu'} \left(\gamma_i(g_{\mu'}) + \gamma_j(g_{\mu'}) - \gamma_\ell(g_{\mu'})\right)\right] U^\ell_{ij}(n; g_\kappa, 1). \quad (6.128)$$

If g_μ approaches a fixed point g_* as $\mu \to \infty$, we have in the large κ limit

$$U^\ell_{ij}(\kappa n; g_\mu, \mu) \to \kappa^{\gamma_i(g_*)+\gamma_j(g_*)-\gamma_\ell(g_*)-4} \mathcal{C}^\ell_{ij}(n; g_\kappa), \quad (6.129)$$

i.e. the OPE coefficients scale with the momentum as dictated by the critical dimension of the operators $\mathcal{O}_i$ etc., up to possible logarithmic dependence of g_κ on κ at large κ.

In an asymptotically free theory, typically near the UV fixed point $g_* = 0$, the scaling exponent γ_i behaves as

$$\gamma_i(g) \to d_i + g^2 c_i + \mathcal{O}(g^4), \quad (6.130)$$

and the RG equation for the coupling g takes the form

$$\mu \frac{d}{d\mu} g_\mu^2 = -\frac{b}{8\pi^2} g_\mu^4 + \mathcal{O}(g_\mu^6). \quad (6.131)$$

It follows that in the $\kappa \to \infty$ limit,

$$\begin{aligned} g_\kappa^2 &\to \frac{8\pi^2}{b \ln \kappa}, \\ \int_\mu^\kappa \frac{d\mu'}{\mu'} \gamma_i(g_{\mu'}) &\to d_i \ln \frac{\kappa}{\mu} - c_i \frac{8\pi^2}{b} \ln g_\kappa^2 + const, \end{aligned} \quad (6.132)$$

and so

$$U^\ell_{ij}(\kappa n; g_\mu, \mu) \to \kappa^{d_i+d_j-d_\ell-4} (\ln \kappa)^{\frac{8\pi^2}{b}(c_i+c_j-c_\ell)} \mathcal{C}^\ell_{ij}, \quad (6.133)$$

where $\mathcal{C}^\ell_{ij}$ is a constant.

Generally, a convenient basis of operators $\mathcal{O}_i$ one chooses to work with may not diagonalize the RG flow. When this happens, say in asymptotically free theories, c_i is replaced by a matrix c_{ij}, which is block diagonal where each block involves operators of the same classical scaling dimension. (6.133) is then replaced by

$$U^\ell_{ij}(\kappa n; g_\mu, \mu) \to \kappa^{d_i+d_j-d_\ell-4} \left[(\ln \kappa)^{\frac{8\pi^2}{b} c}\right]_{ii'} \left[(\ln \kappa)^{\frac{8\pi^2}{b} c}\right]_{jj'} \left[(\ln \kappa)^{-\frac{8\pi^2}{b} c}\right]_{\ell'\ell} \mathcal{C}^{\ell'}_{i'j'}. \quad (6.134)$$

In the next section, the method of OPE and in particular the asymptotic behavior of OPE coefficients will be applied to the study of deep inelastic scattering of electron with hadron.

6.7 Deep inelastic scattering

Consider the scattering of an electron of four-momentum k off a nucleon N of momentum p and mass m_N, producing an electron of momentum k' and a general hadron state H. Let $q = k - k'$ be the momentum transfer. The scattering amplitude is given by

$$\mathcal{A}_{\sigma',\sigma}(k',k;q) = \bar{u}_{\sigma'}(k')(-e\gamma_\mu)u_\sigma(k)\frac{-i}{q^2 - i\epsilon}\langle H|eJ^\mu(0)|N\rangle \tag{6.135}$$

Define the Lorentz invariant quantity ν which is the energy loss in the nucleon rest frame,

$$\nu \equiv -\frac{q\cdot p}{m_N} = E_e - E'_e. \tag{6.136}$$

In the rest frame of the nucleon, we have

$$|\vec{k}'|\frac{d|\vec{k}'|}{d\nu} = \frac{k'\cdot p}{m_N^2}\frac{d(k'\cdot p)}{d\nu} = \frac{k'\cdot p}{m_N}. \tag{6.137}$$

The differential cross section summed over the final hadron state, in the rest frame of the nucleon, is given in terms of the amplitude by

$$\begin{aligned}\frac{d^2\sigma}{d\Omega d\nu} &= \frac{1}{v}|\vec{k}'|^2\frac{d|\vec{k}'|}{d\nu}\sum_H \delta^4(p_H - p - q)\sum_{\sigma,\sigma'}|\mathcal{A}_{\sigma',\sigma}(k',k;q)|^2 \\ &= -\frac{|\vec{k}'|(k\cdot p)(k'\cdot p)}{m_N^2|\vec{k}|}\sum_H \delta^4(p_H - p - q)\sum_{\sigma,\sigma'}|\mathcal{A}_{\sigma',\sigma}(k',k;q)|^2\end{aligned} \tag{6.138}$$

We have

$$\begin{aligned}\sum_{\sigma,\sigma'}|\mathcal{A}_{\sigma',\sigma}(k',k;q)|^2 &= \frac{e^4}{(q^2)^2}\mathrm{Tr}\left[\gamma_\mu\frac{-i\not{k}+m_e}{2k^0}\gamma_\nu\frac{-i\not{k}'+m_e}{2(k')^0}\right]\langle H|J^\mu(0)|N\rangle\langle H|J^\nu(0)|N\rangle^* \\ &= \frac{e^4}{(q^2)^2}\frac{-k_\mu k'_\nu - k_\nu k'_\mu + (k\cdot k' + m_e^2)\eta_{\mu\nu}}{k^0(k')^0}\langle H|J^\mu(0)|N\rangle\langle H|J^\nu(0)|N\rangle^* \\ &= \frac{e^4}{(q^2)^2}m_N^2\frac{-2k_\mu k_\nu + (k\cdot k' + m_e^2)\eta_{\mu\nu}}{(k\cdot p)(k'\cdot p)}\langle H|J^\mu(0)|N\rangle\langle H|J^\nu(0)|N\rangle^*,\end{aligned} \tag{6.139}$$

where in the last step we used that conservation of the current J^μ, which implies that the matrix element $\langle H|q_\mu J^\mu(0)|N\rangle$ vanishes, and used the fact that we are working in the rest frame of the nucleon.

In the high energy regime where the mass of the electron can be ignored, we have

$$\begin{aligned}\cos\theta &= \frac{\vec{k}\cdot\vec{k}'}{|\vec{k}||\vec{k}'|} \approx 1 - \frac{q^2}{2E_eE'_e}, \\ \sin^2(\theta/2) &\approx \frac{q^2}{4E_eE'_e},\end{aligned} \tag{6.140}$$

and so

$$\frac{d^2\sigma}{d\Omega d\nu} = \frac{e^4}{4q^2 E_e^2 \sin^2(\theta/2)} \left(2k_\mu k_\nu + \frac{1}{2} q^2 \eta_{\mu\nu} \right) \sum_H \delta^4(p_H - p - q) \langle H | J^\mu(0) | N \rangle \langle H | J^\nu(0) | N \rangle^* \tag{6.141}$$

It is useful to define the structure function

$$\begin{aligned} W^{\mu\nu}(q,p) &\equiv \frac{p_N^0}{2m_N} \sum_{\sigma_N} \sum_H \delta^4(p_H - p - q) \langle H | J^\mu(0) | N \rangle \langle H | J^\nu(0) | N \rangle^* \\ &= \frac{p_N^0}{2m_N (2\pi)^4} \sum_{\sigma_N} \int d^4x e^{-iq\cdot x} \langle N | J^\nu(x) J^\mu(0) | N \rangle, \end{aligned} \tag{6.142}$$

where the average is taken over the spin of the nucleon N, and a factor p_N^0/m_N is included to compensate for the Lorentz non-invariant normalization of the nucleon state $|N\rangle$, so that $W^{\mu\nu}$ transforms as a Lorentz tensor. Current conservation $q_\mu W^{\mu\nu} = q_\nu W^{\mu\nu} = 0$ then implies that $W^{\mu\nu}$ is a linear combination of two tensor structures,

$$W^{\mu\nu}(q,p) = -\left(\frac{q^\mu q^\nu}{q^2} - \eta^{\mu\nu} \right) W_1(\nu, q^2) + \frac{1}{m_N^2} \left(p^\mu - \frac{p\cdot q}{q^2} q^\mu \right) \left(p^\nu - \frac{p\cdot q}{q^2} q^\nu \right) W_2(\nu, q^2), \tag{6.143}$$

where W_1, W_2 are functions of the only two Lorentz invariant scalar variables made out of p, q, namely ν the energy loss in nucleon rest frame, and q^2. It follows from the definition of $W^{\mu\nu}$ that W_1, W_2 are real positive functions, and are non-vanishing only when $-p_H^2 = -q^2 + 2m_N\nu + m_N^2$ is above the hadron threshold, namely $\nu > q^2/(2m_N)$ if N is the lightest hadron among those of the same quantum number.

The differential cross section, averaged over the nucleon spin state, is now written as

$$\frac{d^2\sigma}{d\Omega d\nu} = \left(\frac{d\sigma}{d\Omega} \right)_{MOTT} \left(2W_1 \tan^2(\theta/2) + W_2 \right), \tag{6.144}$$

where

$$\left(\frac{d\sigma}{d\Omega} \right)_{MOTT} = \frac{e^4}{4E_e^2} \frac{\cos^2(\theta/2)}{\sin^4(\theta/2)} \tag{6.145}$$

is the differential cross section of electron scattering off a point spinless particle, in the high energy limit where the electron mass can be ignored.

The high energy behavior of the electron scattering cross section then depends on the high energy behavior of $W_{1,2}$, which is encoded in the singular terms of the OPE $J^\nu(x)J^\mu(0)$. A particularly interesting regime, the "deep inelastic scattering", is when q^2, ν are both large, with $\omega = 2m_N\nu/q^2$ fixed.

To analyze the structure function $W^{\mu\nu}(q,p)$, it is useful to consider the matrix element of the time ordered two point function of currents,

$$T^{\mu\nu}(q,p) = \frac{p_N^0}{2m_N(2\pi)^4}\sum_{\sigma_N}\int d^4x e^{-iq\cdot x}\langle N|T\left[J^\nu(x)J^\mu(0)\right]|N\rangle. \tag{6.146}$$

Lorentz invariance combined with current conservation $q_\mu T^{\mu\nu} = q_\nu T^{\mu\nu} = 0$ implies that $T^{\mu\nu}(q,p)$ can be expressed in terms of scalar functions $T_{1,2}(q,p)$ exactly the same way as in (6.143). Note that in rewriting $q_\nu T^{\mu\nu}$ in terms of the matrix element of $\partial_\nu\left\{T[J^\nu(x)J^\mu(0)]\right\}$, to show the latter vanishes we need to use causality which implies that the equal time commutator between $J^\nu(\vec{x},0)$ and $J^\mu(0)$ vanishes for nonzero $\vec{x}$.

We can relate $T^{\mu\nu}$ to $W^{\mu\nu}$ as follows. Let n^μ be a timelike vector with $n^0 > 0$. Since spacelike separated currents commute, we can replace the time ordering by ordering with respect to $-n\cdot x$, and so

$$\begin{aligned}
T^{\mu\nu}(q,p) &= \frac{p_N^0}{2m_N(2\pi)^4}\sum_{\sigma_N}\int d^4x e^{-iq\cdot x}\langle N|\left[\theta(-n\cdot x)J^\nu(x)J^\mu(0)+\theta(n\cdot x)J^\mu(0)J^\nu(x)\right]|N\rangle \\
&= \frac{i}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega}{\omega+i\epsilon}\,\frac{p_N^0}{2m_N(2\pi)^4}\sum_{\sigma_N}\int d^4x\langle N|\left[e^{-iq\cdot x+i\omega n\cdot x}J^\nu(x)J^\mu(0)+e^{iq\cdot x+i\omega n\cdot x}J^\mu(x)J^\nu(0)\right]|N\rangle \\
&= \frac{i}{2\pi}\int_{-\infty}^{\infty}\frac{d\omega}{\omega+i\epsilon}\left[W^{\mu\nu}(q-\omega n,p)+W^{\nu\mu}(-q-\omega n,p)\right] \\
&= \frac{1}{2}W^{\mu\nu}(q,p)+\frac{1}{2}W^{\nu\mu}(-q,p)-\frac{1}{2\pi i}⨍\frac{d\omega}{\omega}\left[W^{\mu\nu}(q+\omega n,p)-W^{\nu\mu}(-q-\omega n,p)\right]
\end{aligned} \tag{6.147}$$

Note that

$$-\frac{(q+\omega n)\cdot p}{m_N} = \nu-\omega\frac{n\cdot p}{m_N}, \qquad (q+\omega n)^2 = q^2+2\omega q\cdot n+\omega^2n^2. \tag{6.148}$$

In the regime of interest, q^μ is spacelike, and we can choose n^μ such that $n\cdot q = 0$, $n\cdot p = -m_N$. We can furthermore take the limit n^μ becoming lightlike, $n^2\to 0$, as $W^{\mu\nu}(q+\omega n,p)$ behaves regularly in this limit. In the principle part integral on the RHS of (6.147), a priori, additional Lorentz tensor structures arise. But the coefficients of these structures must vanish by Lorentz invariance of $T^{\mu\nu}(q,p)$ (that is, the tensor structure is independent of n^μ) and current conservation. We then derive the relation on the scalar functions

$$\begin{aligned}
T_r(\nu,q^2) &= \frac{1}{2}W_r(\nu,q^2)+\frac{1}{2}W_r(-\nu,q^2)-\frac{1}{2\pi i}⨍\frac{d\omega}{\omega}\left[W_r(\nu+\omega,q^2)-W_r(-\nu-\omega,q^2)\right] \\
&= \frac{1}{2}W_r(\nu,q^2)+\frac{1}{2}W_r(-\nu,q^2)-\frac{1}{2\pi i}⨍_{\frac{q^2}{2m_N}}^{\infty}d\omega\,W_r(\omega,q^2)\left(\frac{1}{\omega-\nu}+\frac{1}{\omega+\nu}\right).
\end{aligned} \tag{6.149}$$

where $r = 1, 2$. In the last step we used the property that $W_r(\nu, q^2)$ vanishes when $\nu < q^2/(2m_N)$. For $\nu > q^2/(2m_N)$, we can also write W_r in terms of the real part of T_r,

$$W_r(\nu, q^2) = 2\,\mathrm{Re}T_r(\nu, q^2). \tag{6.150}$$

We now examine the $q^2 \to \infty$ limit of $T^{\mu\nu}(q,p)$ from the current-current OPE. The operators that appear in $\int d^4x\, e^{-iq\cdot x} T J^\nu(x) J^\mu(0)$ OPE are Lorentz tensors contracted with $q^{\mu_1} \cdots q^{\mu_{s'}}$. We can therefore restrict our attention to symmetric traceless tensor operators

$$\mathcal{O}_{s,i}^{\mu_1\cdots\mu_s}, \tag{6.151}$$

where s is the spin of the operator, and i labels different operators of the same spin. Suppose this operator has mass dimension $d(s,i)$. Then classically, we expect the OPE coefficient to scale with q like

$$|q|^{3+3-d(s,i)-4} = |q|^{2-d(s,i)}. \tag{6.152}$$

It follows from Lorentz invariance that the spin averaged expectation value of $\mathcal{O}_{s,i}^{\mu_1\cdots\mu_s}$ in the nucleon state takes the form

$$\frac{1}{2}\sum_{\sigma_N} \langle N|\mathcal{O}_{s,i}^{\mu_1\cdots\mu_s}|N\rangle = \frac{m_N}{p^0}\left(p^{\mu_1}\cdots p^{\mu_N} - \text{traces}\right)\mathcal{C}_{s,i}, \tag{6.153}$$

where $\mathcal{C}_{s,i}$ is a constant. In its contribution to $T_1(\nu, q^2)$, $p^{\mu_1}\cdots p^{\mu_s}$ must be contracted with s factors of q's to give ν^s. In the contribution to $T_2(\nu, q^2)$, two factors of p^μ are already contained in the Lorentz tensor prefactor, and hence $s-2$ q's are contracted to give ν^{s-2}. Thus the contribution from the operator $\mathcal{O}_{s,i}$ in the OPE to $T_{1,2}(\nu, q^2)$ have the scaling

$$\begin{aligned} T_1(\nu, q^2) &\sim \nu^s (q^2)^{1-\frac{d(s,i)+s}{2}}, \\ T_2(\nu, q^2) &\sim \nu^{s-2} (q^2)^{2-\frac{d(s,i)+s}{2}}. \end{aligned} \tag{6.154}$$

In the deep elastic limit, $\omega = 2m_N\nu/q^2$ is held fixed, we have

$$\begin{aligned} T_1(\nu, q^2), W_1(\nu, q^2) &\sim \omega^s (q^2)^{-\frac{d(s,i)-s-2}{2}}, \\ \nu T_2(\nu, q^2), \nu W_2(\nu, q^2) &\sim \omega^{s-1} (q^2)^{-\frac{d(s,i)-s-2}{2}}. \end{aligned} \tag{6.155}$$

In particular, deep elastic scattering is dominated by operators $\mathcal{O}_{s,i}$ of the smallest value of "twist"

$$\tau(s,i) = d(s,i) - s. \tag{6.156}$$

In QCD, the gauge invariant operators has twist ≥ 2. For instance, the flavor current which can be expressed as quark-anti-quark bilinear $\bar\psi\gamma_\mu\psi$ has dimension $d = 3$ and spin 1, and hence twist 2. The complete independent set of twist-2 operators that are singlets with respect to the electromagnetic $U(1)$ are

$$\begin{aligned} \mathcal{O}_{s,f}^{\mu_1\cdots\mu_s} &= i^s\bar{\psi}_f\gamma^{(\mu_1}\overleftrightarrow{D}^{\mu_2}\cdots\overleftrightarrow{D}^{\mu_s)}\psi_f - \text{traces}, \\ \mathcal{O}_{s,g}^{\mu_1\cdots\mu_s} &= i^s F_{a\nu}{}^{(\mu_1}\overleftrightarrow{D}^{\mu_2}\cdots\overleftrightarrow{D}^{\mu_{s-1}}F_b{}^{\mu_s)\nu} - \text{traces}, \end{aligned} \tag{6.157}$$

where f labels the quark flavors. All other twist-2 operators can be written as linear combinations of these and their derivatives, up to equation of motion. Note that we did not include flavor changing operators here. The contribution from the twist-2 operators to the structure function $W_r(\nu, q^2)$ says that $W_1(\nu, q^2)$ and $\nu W_2(\nu, q^2)$ are finite functions of $\omega = 2m_N\nu/q^2$ in the $q^2 \to \infty$ limit with ω fixed. This is known as Bjorken scaling.

So far we have only made use of the classical scaling dimension of $\mathcal{O}_{s,i}$, and ignored the q dependence due to renormalization factors in the OPE coefficients. We have seen previously that the scaling exponent matrix γ_{ij} on the operators $\mathcal{O}_{s,i}$ corrects the q-dependence of the OPE coefficients. In asymptotically free theories such as QCD, this gives rise to logarithmic correction to Bjorken scaling. Note that the conserved current J^μ does not acquire quantum corrections to its scaling exponent. Use the one-loop RG corrected OPE coefficients derived earlier, we find the scaling behavior of $T_r(\nu, q^2)$,

$$\begin{aligned} T_1(\nu, q^2) &\to \sum_{s,i,j}\omega^s\mathcal{A}_{s,i}\left[\left(g_q^2/g_\mu^2\right)^{\frac{8\pi^2}{b}c(s)}\right]_{ij}\mathcal{C}_{s,j}, \\ \nu T_2(\nu, q^2) &\to \sum_{s,i,j}\omega^{s-1}\mathcal{B}_{s,i}\left[\left(g_q^2/g_\mu^2\right)^{\frac{8\pi^2}{b}c(s)}\right]_{ij}\mathcal{C}_{s,j}, \end{aligned} \tag{6.158}$$

where μ is a fixed scale, g_q and g_μ are the renormalized gauge couplings at scale $|q|$ and μ respectively. $c(s)_{ij}$ is the one-loop coefficient (i.e. coefficient of g^2) in the scaling exponent matrix $\gamma_{ij}(g)$ on spin-s twist-2 operators. The index i, j here runs through $1, 2, \cdots, N_f + 1$, or f, g for $\mathcal{O}_{s,f}$ and $\mathcal{O}_{s,g}$. Indeed at one-loop order the renormalized operators $\mathcal{O}_{s,f}$ and $\mathcal{O}_{s,g}$ are mixed under the RG flow. This computation is left as an exercise. The results are

$$\begin{aligned} c(s)_{ff'} &= \frac{C'}{8\pi^2}\left[1 - \frac{2}{s(s+1)} + 4\sum_{t=2}^{s}\frac{1}{t}\right]\delta_{ff'}, \\ c(s)_{gf} &= \frac{C'}{8\pi^2}\left[\frac{1}{s+1} + \frac{2}{s(s-1)}\right], \\ c(s)_{fg} &= \frac{C(f)}{\pi^2}\left[\frac{1}{s+2} + \frac{2}{s(s+1)(s+2)}\right], \\ c(s)_{gg} &= \frac{1}{2\pi^2}\left\{C(adj)\left[\frac{1}{12} - \frac{1}{s(s-1)} - \frac{1}{(s+1)(s+2)} + \sum_{t=2}^{s}\frac{1}{t}\right] + \frac{N_f}{3}C(f)\right\}. \end{aligned} \tag{6.159}$$

Here $C(adj)$ and $C(f)$ are as defined in the computation of one-loop beta function of QCD, C' is the Casimir invariant, defined by $t_a t_a = g^2 C' \mathbb{I}$. For $SU(N)$, we have

$$C(adj) = N, \quad C(f) = \frac{1}{2}, \quad C' = \frac{N^2-1}{2N}. \tag{6.160}$$

Note that $\mathcal{O}_{s,g}$ exists only for $s \geq 2$, and $c(1)_{ff'}$ vanishes. Note further that $c(2)_{gg}$ vanishes in pure Yang-Mills theory, i.e. when $N_f = 0$. This is because in pure Yang-Mills theory $\mathcal{O}^{\mu\nu}_{2,g}$ is the stress energy tensor, which does not receive quantum corrections to its scaling exponent.

We will illustrate here one example of the computation of the mixing matrix elements, $c(s)_{gf}$. This is particularly simple to calculate as there is only one 1-loop diagram that contributes to the mixing of $\mathcal{O}_{s,f}$ with $\mathcal{O}_{s,g}$, namely the one with two gluon lines coming out of the vertex $\mathcal{O}_{s,f}$ (at zero momentum), both attached to a quark line. We will consider the renormalization of

$$\mathcal{O}_{s,f}(n) \equiv \mathcal{O}^{\mu_1\cdots\mu_s}_{s,f} n_{\mu_1}\cdots n_{\mu_s} = i^s F_{a\nu\mu} n^\mu (n\cdot \overleftrightarrow{D})^{s-2} F_{b\rho}{}^\nu n^\rho, \tag{6.161}$$

where n_μ is a null vector. The corresponding vertex with momentum k flowing in on one gluon line and out on the other gluon line is

$$\delta_{ab}(n\cdot k)^{s-2}\left[(n\cdot k)^2\eta_{\mu\nu} + k^2 n_\mu n_\nu - (n\cdot k)(k_\mu n_\nu + k_\nu n_\mu)\right] \tag{6.162}$$

Let the momentum of the external quark lines be p and $-p$. The quarks will be taken to be massless. In $\xi = 1$ gauge, the diagram is computed as

$$\begin{aligned}
&\int \frac{d^4k}{(2\pi)^4}\gamma^\mu \frac{-i}{\not p + \not k - i\epsilon}\gamma^\nu \frac{(-i)^2}{(k^2-i\epsilon)^2}(n\cdot k)^{s-2}\left[(n\cdot k)^2\eta_{\mu\nu} + k^2 n_\mu n_\nu - (n\cdot k)(k_\mu n_\nu + k_\nu n_\mu)\right]\\
&= i\int \frac{d^4k}{(2\pi)^4}(n\cdot k)^{s-2}\frac{(n\cdot k)^2\gamma^\mu(\not p+\not k)\gamma_\mu + k^2\not n(\not p+\not k)\not n - (n\cdot k)\not n(\not p+\not k)\not k - (n\cdot k)\not k(\not p+\not k)\not n}{(k^2-i\epsilon)^2((p+k)^2-i\epsilon)}\\
&= i\int \frac{d^4k}{(2\pi)^4}(n\cdot k)^{s-2}\frac{-2(n\cdot k)^2\not k + k^2(n\cdot(2p+k))\not n - (n\cdot k)(p\cdot k)\not n - (n\cdot k)(p\cdot n)\not k}{(k^2-i\epsilon)^2((p+k)^2-i\epsilon)}
\end{aligned} \tag{6.163}$$

Using Feynman parameters, the relevant integrals are computed as

$$\begin{aligned}
&\int \frac{d^4k}{(2\pi)^4}(n\cdot k)^{s-2}\frac{n\cdot(2p+k)}{(k^2-i\epsilon)((p+k)^2-i\epsilon)}\\
&= \int_0^1 dx \int \frac{d^4k}{(2\pi)^4}(n\cdot(k-xp))^{s-2}\frac{n\cdot((2-x)p+k)}{[k^2+p^2x(1-x)-i\epsilon]^2}\\
&= (-)^s(n\cdot p)^{s-1}\int_0^1 dx\, x^{s-2}(2-x)\int \frac{d^4k}{(2\pi)^4}\frac{1}{[k^2+p^2x(1-x)-i\epsilon]^2}\\
&= (-)^s(n\cdot p)^{s-1}\left(\frac{2}{s-1}-\frac{1}{s}\right)\frac{i}{8\pi^2}\ln\Lambda + \text{finite},
\end{aligned} \tag{6.164}$$

$$
\begin{aligned}
&\int \frac{d^4k}{(2\pi)^4}\frac{(n\cdot k)^{s-1}(p\cdot k)}{(k^2-i\epsilon)^2((p+k)^2-i\epsilon)} \\
&= 2\int_0^1 dx\,(1-x)\int\frac{d^4k}{(2\pi)^4}\frac{(n\cdot(k-xp))^{s-1}(p\cdot(k-xp))}{\left[k^2+p^2x(1-x)-i\epsilon\right]^3} \\
&= 2(s-1)\int_0^1 dx\,(1-x)\int\frac{d^4k}{(2\pi)^4}\frac{(-xn\cdot p)^{s-2}(n\cdot k)(p\cdot k)}{\left[k^2+p^2x(1-x)-i\epsilon\right]^3}+\text{finite} \\
&= (-)^s\frac{s-1}{2}(n\cdot p)^{s-1}\int_0^1 dx\,x^{s-2}(1-x)\frac{i}{8\pi^2}\ln\Lambda+\text{finite} \\
&= (-)^s(n\cdot p)^{s-1}\frac{1}{2s}\frac{i}{8\pi^2}\ln\Lambda+\text{finite},
\end{aligned}
\tag{6.165}
$$

and

$$
\begin{aligned}
&\int \frac{d^4k}{(2\pi)^4}\frac{(n\cdot k)^{s}\not{k}}{(k^2-i\epsilon)^2((p+k)^2-i\epsilon)} \\
&= 2\int_0^1 dx\,(1-x)\int\frac{d^4k}{(2\pi)^4}\frac{(n\cdot(k-xp))^{s}(\not{k}-x\not{p})}{\left[k^2+p^2x(1-x)-i\epsilon\right]^3} \\
&= 2s\int_0^1 dx\,(1-x)\int\frac{d^4k}{(2\pi)^4}\frac{(-xn\cdot p)^{s-1}(n\cdot k)\not{k}}{\left[k^2+p^2x(1-x)-i\epsilon\right]^3}+\text{finite} \\
&= (-)^{s-1}\frac{s}{2}\not{n}(n\cdot p)^{s-1}\int_0^1 dx\,x^{s-1}(1-x)\frac{i}{8\pi^2}\ln\Lambda+\text{finite} \\
&= (-)^{s-1}(n\cdot p)^{s-1}\frac{1}{2(s+1)}\frac{i}{8\pi^2}\ln\Lambda+\text{finite}.
\end{aligned}
\tag{6.166}
$$

Putting these together, we obtain

$$
\begin{aligned}
&(-)^s\not{n}(n\cdot p)^{s-1}\left[-2\cdot\frac{-1}{2(s+1)}+\left(\frac{2}{s-1}-\frac{1}{s}\right)-\frac{1}{2s}-\frac{1}{2s}\right]\frac{i}{8\pi^2}\ln\Lambda+\text{finite} \\
&=(-)^s\not{n}(n\cdot p)^{s-1}\left[\frac{1}{s+1}+\frac{2}{s(s-1)}\right]\frac{i}{8\pi^2}\ln\Lambda+\text{finite}.
\end{aligned}
\tag{6.167}
$$

There is also the gauge group factor $t^at^a=g^2C'$. Comparing this with the vertex of the operator $\mathcal{O}_{s,f}(n)=i^s\bar{\psi}_f\not{n}(n\cdot\overleftrightarrow{D})^{s-1}\psi_f$, namely

$$
(-)^s\not{n}(n\cdot p)^{s-1}, \tag{6.168}
$$

we read off the one-loop mixing matrix element

$$
c(s)_{gf}=\frac{C'}{8\pi^2}\left[\frac{1}{s+1}+\frac{2}{s(s-1)}\right]. \tag{6.169}
$$

6.8 Wilsonian renormalization group

Gell-Mann and Low's formulation of the renormalization group equation, though practically straightforward, is conceptually rather contrived. The definition of the renormalized couplings, in terms of the 1PI amplitudes or effective action at a renormalization point, is somewhat ad hoc. In renormalizable theories, the fact that the RG equation for the coupling constants is free of UV divergences as the cutoff is taken to infinity relies on the statement that divergences in subdiagrams are canceled by counter terms of the corresponding vertices, even in the presence of overlapping divergences (Bogoliukov-Parasiuk-Hepp-Zimmermann).

There is a conceptually simpler and more natural version of the renormalization group, proposed by Kenneth Wilson. The idea is to take the cutoff Λ seriously, making it finite rather than taking it to infinity, and define an effective theory below the scale Λ, in which the functional integral that computes correlation functions only involves integration over Fourier modes of the fields with momentum p for $|p| < \Lambda$.

Since the momentum cutoff is most naturally introduced in Euclidean signature, we will be working with Euclidean quantum field theory in this section, related to the Lorentzian one by Wick rotation. Consider a quantum field theory of a single scalar field ϕ in d dimensions. We begin with the Euclidean action $S_\Lambda[\phi]$ of the form

$$S_\Lambda[\phi] = \int d^d x \left[\frac{1}{2}(\partial_\mu \phi)^2 + \frac{1}{2} m^2 \phi^2 + \frac{1}{4!} g \phi^4 + \cdots \right] \tag{6.170}$$

while the path integral is defined by integrating only over Fourier modes of ϕ with momentum below Λ,

$$Z = \int [\mathcal{D}\phi]_{|p|<\Lambda} \, e^{-S_\Lambda[\phi]} \tag{6.171}$$

So far we have chosen to work with a sharp cutoff. This is not essential. One may also replace it with a smooth cutoff, by replacing the propagator

$$\frac{1}{p^2+m^2} \to \frac{K(p^2/\Lambda^2)}{p^2+m^2}, \tag{6.172}$$

where $K(x)$ is some function that takes the value 1 for $x < 1$ and vanishes sufficiently fast as $x \to \infty$.

We may separate the path integral into an integration over modes below a lower cutoff Λ', and the integration over momenta in the shell $\Lambda' < |p| < \Lambda$. Namely,

$$Z = \int [\mathcal{D}\phi]_{|p|<\Lambda'} [\mathcal{D}\phi]_{\Lambda' \leq |p| < \Lambda} \, e^{-S_\Lambda[\phi]} \equiv \int [\mathcal{D}\phi]_{|p|<\Lambda'} \, e^{-S_{\Lambda'}[\phi]} \tag{6.173}$$

The "Wilsonian effective action" $S_{\Lambda'}[\phi]$ is obtained by functional integration over all momenta in the shell. It is generally a very complicated looking action, involving arbitrarily high order couplings with arbitrarily high number of derives. To gain some control, it is useful to integrate out a infinitesimally thin momentum shell at a time. For this purpose, we must start with the most general action $S_\Lambda[\phi]$ with arbitrarily high order terms as well.

More generally, using the momentum cutoff function $K(p^2, \Lambda^2)$, we may write the generating functional as

$$Z[J] = \int \mathcal{D}\phi \, \exp\left\{ -\int \frac{d^dp}{(2\pi)^d} \left[\frac{1}{2}\phi(p)\phi(-p)(p^2+m^2)K^{-1}(\frac{p^2}{\Lambda^2}) + J(p)\phi(-p) \right] - L_\Lambda[\phi] \right\}$$
$$\equiv \int \mathcal{D}\phi \, \exp\left\{ -S_\Lambda[\phi] - \int \frac{d^dp}{(2\pi)^d} J(p)\phi(-p) \right\}, \tag{6.174}$$

where in the first line we have separated the interaction part of the Wilsonian effective action, L_Λ. We have defined $S_\Lambda[\phi]$ as a functional of all momentum modes of ϕ (rather than only modes with momentum below the cutoff Λ), and the path integral is over all modes of ϕ, while the momentum cutoff is implemented through K^{-1} in the kinetic term of ϕ.

We assume that $m \ll \Lambda$, and that the source $J(p)$ is nonzero only in the momentum range where $K(p^2/\Lambda^2)$ is equal to 1.

Moving the cutoff Λ to Λ' can now be stated as changing Λ and $L_\Lambda[\phi]$ simultaneously in such a way that $Z[J]$ stays invariant. Namely, we demand

$$0 = \Lambda\frac{d}{d\Lambda} Z[J]$$
$$= \int \mathcal{D}\phi \left\{ -\frac{1}{2}\int \frac{d^dp}{(2\pi)^d}\phi(p)\phi(-p)(p^2+m^2)\Lambda\frac{\partial K^{-1}(p^2/\Lambda^2)}{\partial\Lambda} - \Lambda\frac{d}{d\Lambda}L_\Lambda[\phi] \right\} e^{-S_\Lambda[\phi]-\int J(p)\phi(-p)}. \tag{6.175}$$

This can be achieved if $L_\Lambda[\phi]$ is adjusted as follows:

$$\Lambda\frac{d}{d\Lambda}L_\Lambda[\phi] = \frac{1}{2}\int d^dp \frac{(2\pi)^d}{p^2+m^2}\Lambda\frac{\partial K(p^2/\Lambda^2)}{\partial\Lambda}\left[\frac{\delta L_\Lambda}{\delta\phi(p)}\frac{\delta L_\Lambda}{\delta\phi(-p)} - \frac{\delta^2 L_\Lambda}{\delta\phi(p)\delta\phi(-p)} \right] + (\text{field} - \text{independent}). \tag{6.176}$$

To see this, note that the RHS of (6.175) with $L_\Lambda[\phi]$ given by (6.176) can be rewritten as a functional total derivative,

$$\int d^dp\,\Lambda\frac{\partial K(p^2/\Lambda^2)}{\partial\Lambda}\int\mathcal{D}\phi\frac{\delta}{\delta\phi(p)}\left[\phi(p)K^{-1}(p^2/\Lambda^2) + \frac{1}{2}\frac{(2\pi)^d}{p^2+m^2}\frac{\delta}{\delta\phi(-p)} \right] e^{-S_\Lambda[\phi]-\int J(p)\phi(-p)}$$
$$= \int d^dp\,\Lambda\frac{\partial K(p^2/\Lambda^2)}{\partial\Lambda}\int\mathcal{D}\phi\frac{\delta}{\delta\phi(p)}\left[\frac{1}{2}\phi(p)K^{-1}(p^2/\Lambda^2) - \frac{1}{2}\frac{(2\pi)^d}{p^2+m^2}\frac{\delta L_\Lambda}{\delta\phi(-p)} \right] e^{-S_\Lambda[\phi]-\int J(p)\phi(-p)}$$
$$= \int d^dp\,\Lambda\frac{\partial K(p^2/\Lambda^2)}{\partial\Lambda}\int\mathcal{D}\phi\left[\frac{1}{2}\delta^d(0)K^{-1}(p^2/\Lambda^2) - \frac{1}{2}\frac{(2\pi)^d}{p^2+m^2}\frac{\delta^2 L_\Lambda}{\delta\phi(p)\delta\phi(-p)} \right.$$
$$- \frac{1}{2}\phi(p)\phi(-p)(p^2+m^2)K^{-2}(p^2/\Lambda^2) - \frac{1}{2}\phi(p)K^{-1}(p^2/\Lambda^2)\frac{\delta L_\Lambda}{\delta\phi(p)}$$
$$\left. + \frac{1}{2}\frac{\delta L_\Lambda}{\delta\phi(-p)}\phi(-p)K^{-1}(p^2/\Lambda^2) + \frac{1}{2}\frac{(2\pi)^d}{p^2+m^2}\frac{\delta L_\Lambda}{\delta\phi(p)}\frac{\delta L_\Lambda}{\delta\phi(-p)} \right] e^{-S_\Lambda[\phi]-\int J(p)\phi(-p)} \tag{6.177}$$

where the field independent shift of $L_\Lambda[\phi]$ is chosen to cancel the term proportional to $\delta^d(0) = (2\pi)^{-d} V_d$. From now on, we will ignore this field-independent shift, and work with the equation

$$\Lambda \frac{d}{d\Lambda} L_\Lambda[\phi] = \frac{1}{2} \int d^d p \frac{(2\pi)^d}{p^2 + m^2} \Lambda \frac{\partial K(p^2/\Lambda^2)}{\partial \Lambda} \left[\frac{\delta L_\Lambda}{\delta \phi(p)} \frac{\delta L_\Lambda}{\delta \phi(-p)} - \frac{\delta^2 L_\Lambda}{\delta \phi(p) \delta \phi(-p)} \right]. \quad (6.178)$$

Strictly speaking, under this transformation, the generating functional shifts by a constant factor, but physical correlation functions are not affected. The form (6.178) of the Wilsonian renormalization group equation was formulated by Polchinski: it describes the evolution of an infinite set of couplings in the Wilsonian effective action, of positive, zero, or negative mass dimensions, under the change of the cutoff scale Λ. The RHS of (6.178) has a very simple and clear interpretation: the first term is

$$-\frac{1}{2} \int d^d p \frac{(2\pi)^d}{p^2 + m^2} \Lambda \frac{\partial K(p^2/\Lambda^2)}{\partial \Lambda} \frac{\delta L_\Lambda}{\delta \phi(p)} \frac{\delta L_\Lambda}{\delta \phi(-p)} \quad (6.179)$$

with an opposite sign: it contracts two vertices in $L_\Lambda[\phi]$ using the ϕ propagator on the momentum shell $|p| \sim \Lambda$, as $\Lambda \partial K / \partial \Lambda$ is supported near $|p| = \Lambda$. The second term is

$$\frac{1}{2} \int d^d p \frac{(2\pi)^d}{p^2 + m^2} \Lambda \frac{\partial K(p^2/\Lambda^2)}{\partial \Lambda} \frac{\delta^2 L_\Lambda}{\delta \phi(p) \delta \phi(-p)} \quad (6.180)$$

with the opposite sign; it represents the contraction of two legs of a single vertex in $L_\Lambda[\phi]$ using the propagator on the momentum shell. These two operations take into account completely the effect of integrating out fields on the momentum shell, and generate the new effective couplings at a slightly lower scale $\Lambda - \delta\Lambda$.

Now we can state the meaning of renormalizability in terms of L_Λ. The relevant and marginal operators that preserve rotational (Lorentz) invariance and $\phi \to -\phi$, perturbatively in $d = 4$ dimensions, are

$$\phi^2, \quad (\partial_\mu \phi)^2, \quad \text{and } \phi^4. \quad (6.181)$$

Denote the coefficients of these operators in the "bare" interaction action L_{Λ_0} at a high scale Λ_0 by λ_1^B, λ_2^B, and λ_3^B. Assume that L_{Λ_0} has no other couplings. Flowing down to scale Λ, L_Λ contains the (renormalized) couplings λ_1, λ_2, λ_3, as well infinitely many irrelevant couplings. To be precise,

$$\begin{aligned} L_\Lambda[\phi] = \int \frac{d^4 p}{(2\pi)^4} L^{(2)}(p) \phi(p) \phi(-p) + \int \frac{d^4 p}{(2\pi)^4} \frac{d^4 q}{(2\pi)^4} \frac{d^4 k}{(2\pi)^4} \\ \times L^{(4)}(p, q, k) \phi(p) \phi(q) \phi(k) \phi(-p - q - k) + \cdots \end{aligned} \quad (6.182)$$

and λ_i are defined as

$$\lambda_1 = L^{(2)}(0), \qquad \lambda_2 = \frac{\partial}{\partial p^2} L^2(p) \Big|_{p=0}, \qquad \lambda_3 = L^{(4)}(0, 0, 0). \quad (6.183)$$

Now *define* $\lambda_i^B(g,\Lambda,\Lambda_0)$ to be the values of the bare couplings at scale Λ_0, *such that* after flowing down to scale Λ, $\lambda_1=\lambda_2=0$, and $\lambda_3=\frac{1}{4!}g$. The full interacting part of the Wilsonian effective action at the scale Λ, L_Λ, obtained by starting at Λ_0 with only the couplings $\lambda_i^B(g,\Lambda,\Lambda_0)$ and flowing down to Λ, now coincides with the ϕ^4 action with coupling g, plus infinitely many irrelevant operators. The L_Λ constructed this way depends on g and Λ_0; let us denote it by $L_\Lambda(g,\Lambda_0)$. *Renormalizability* is the statement that the limit

$$\lim_{\Lambda_0\to\infty} L_\Lambda(g,\Lambda_0) \tag{6.184}$$

exists. Note that it does *not* imply that the limit is just the ϕ^4 action: this is not at all the case, as we started with a pure ϕ^4 action at Λ_0 and after the RG flow to Λ all sorts of irrelevant terms are generated. Renormalizability, in this perspective, simply means that all these irrelevant couplings are determined by g,Λ, and is insensitive to Λ_0 as Λ_0 is taken to infinity.

Furthermore, one can show that the irrelevant terms in $L_\Lambda(g,\Lambda_0)$, in the large Λ_0 limit, deviates from those terms in $L_\Lambda(g,\infty)$ by $(\Lambda/\Lambda_0)^2$ times some power of $\ln(\Lambda/\Lambda_0)$.

As a toy example, let us make a naive truncation of the Wilsonian RG equations for the scalar field theory to a set of equations for two couplings only, $g_4\phi^4$ and $g_6\phi^6$. While we are no longer talking about the actual Wilsonian RG equations, the flow pattern of g_4 and g_6 in this toy model captures the essence of renormalizability. Define the dimensionless couplings $\lambda_4=g_4$, $\lambda_6=\Lambda^2 g_6$. The "RG" equation takes the form

$$\begin{aligned}\Lambda\frac{d\lambda_4}{d\Lambda}&=\beta_4(\lambda_4,\lambda_6),\\ \Lambda\frac{d\lambda_6}{d\Lambda}&=2\lambda_6+\beta_6(\lambda_4,\lambda_6),\end{aligned} \tag{6.185}$$

where we have separated the classically scaling dimension of λ_6, so that $\beta_6(\lambda_4,\lambda_6)$ is of quadratic and higher order in λ_4,λ_6.

The ordinary treatment of renormalization in a *renormalizable* theory amounts to the following: choose $\lambda_4=\lambda_4^0$, and $\lambda_6=0$ at a very high cutoff scale Λ_0, such that $\lambda_4(\Lambda_R)=\lambda_4^R$ at a lower renormalization point Λ_R. Note that $\lambda_6(\Lambda_R)$ is *not* zero, but given by the solution of the flow equation (6.185) with the initial condition $(\lambda_4,\lambda_6)=(\lambda_4^0,0)$ at Λ_0. However, $\lambda_6(\Lambda_R)$ approaches a fixed finite value as Λ_0 is taken to infinity, λ_4^0 adjusted such that $\lambda_4(\Lambda_R)$ stays fixed. Moreover, this limiting value of $\lambda_6(\Lambda_R)$ is determined by λ_4^R.

To see this is the case, let us consider a small deviation from a given RG trajectory $(\overline{\lambda_4},\overline{\lambda_6})$,

$$\begin{aligned}\lambda_4(\Lambda)&=\overline{\lambda_4}(\Lambda)+\varepsilon_4(\Lambda),\\ \lambda_6(\Lambda)&=\overline{\lambda_6}(\Lambda)+\varepsilon_6(\Lambda).\end{aligned} \tag{6.186}$$

Linearizing the equations in $\varepsilon_4, \varepsilon_6$ gives

$$\begin{aligned}\Lambda\frac{d\varepsilon_4}{d\Lambda} &= \overline{\frac{\partial\beta_4}{\partial\lambda_4}}\varepsilon_4 + \overline{\frac{\partial\beta_4}{\partial\lambda_6}}\varepsilon_6, \\ \Lambda\frac{d\varepsilon_6}{d\Lambda} &= \overline{\frac{\partial\beta_6}{\partial\lambda_4}}\varepsilon_4 + \left(2 + \overline{\frac{\partial\beta_6}{\partial\lambda_6}}\right)\varepsilon_6,\end{aligned} \tag{6.187}$$

We want to consider the nearby RG trajectories that obey

$$\varepsilon_4(\Lambda_R) = 0, \tag{6.188}$$

let Λ_0 vary, keep $\lambda_6(\Lambda_0) = 0$, and see how $\lambda_6(\Lambda_R)$ depends on Λ_0. It is convenient to consider variable

$$\xi_6(\Lambda) \equiv \varepsilon_6(\Lambda) - \frac{d\overline{\lambda_6}}{d\Lambda}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-1}\varepsilon_4(\Lambda). \tag{6.189}$$

A straightforward calculation gives the RG evolution of ξ_6,

$$\begin{aligned}\Lambda\frac{d\xi_6}{d\Lambda} &= \Lambda\frac{d\varepsilon_6}{d\Lambda} - \frac{d\overline{\lambda_6}}{d\Lambda}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-1}\Lambda\frac{d\varepsilon_4}{d\Lambda} - \Lambda\frac{d^2\overline{\lambda_6}}{d\Lambda^2}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-1}\varepsilon_4 + \Lambda\frac{d^2\overline{\lambda_4}}{d\Lambda^2}\frac{d\overline{\lambda_6}}{d\Lambda}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-2}\varepsilon_4 \\ &= \overline{\frac{\partial\beta_6}{\partial\lambda_4}}\varepsilon_4 + \left(2 + \overline{\frac{\partial\beta_6}{\partial\lambda_6}}\right)\varepsilon_6 - \frac{d\overline{\lambda_6}}{d\Lambda}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-1}\left(\overline{\frac{\partial\beta_4}{\partial\lambda_4}}\varepsilon_4 + \overline{\frac{\partial\beta_4}{\partial\lambda_6}}\varepsilon_6\right) \\ &\quad - \Lambda\frac{d^2\overline{\lambda_6}}{d\Lambda^2}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-1}\varepsilon_4 + \Lambda\frac{d^2\overline{\lambda_4}}{d\Lambda^2}\frac{d\overline{\lambda_6}}{d\Lambda}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-2}\varepsilon_4 \\ &= \left[2 + \overline{\frac{\partial\beta_6}{\partial\lambda_6}} - \frac{d\overline{\lambda_6}}{d\Lambda}\left(\frac{d\overline{\lambda_4}}{d\Lambda}\right)^{-1}\overline{\frac{\partial\beta_4}{\partial\lambda_6}}\right]\xi_6 \\ &= \left[2 + \overline{\frac{\partial\beta_6}{\partial\lambda_6}} + \overline{\frac{\partial\beta_4}{\partial\lambda_4}} - \Lambda\frac{d\ln\overline{\beta_4}}{d\Lambda}\right]\xi_6.\end{aligned} \tag{6.190}$$

In deriving the second equality above, we used (6.187). In the equality, we used (6.185). Integrating this equation gives

$$\xi_6(\Lambda) = \xi_6(\Lambda_0)\left(\frac{\Lambda}{\Lambda_0}\right)^2\frac{\overline{\beta_4}(\Lambda_0)}{\overline{\beta_4}(\Lambda)}\exp\left[-\int_\Lambda^{\Lambda_0}\frac{d\Lambda'}{\Lambda'}\left(\overline{\frac{\partial\beta_6}{\partial\lambda_6}}(\Lambda') + \overline{\frac{\partial\beta_4}{\partial\lambda_4}}(\Lambda')\right)\right]. \tag{6.191}$$

As long as we stay in weak coupling regime, $\overline{\frac{\partial\beta_6}{\partial\lambda_6}}, \overline{\frac{\partial\beta_4}{\partial\lambda_4}}$ are small, and $\overline{\beta_4}(\Lambda)$ does not grow too fast with Λ (i.e. the anomalous dimension of λ_4 is small), (6.191) is dominated by $(\Lambda/\Lambda_0)^2$. Consequently, $\xi_6(\Lambda_R)$ stays finite as Λ_0 is taken to infinity.

This argument shows that renormalizability of ϕ^4 follows very simply in Wilsonian renormalization group, and no argument based on analyzing divergences in Feynman diagrams is necessary.

The Wilsonian RG may be formulated also for theories with fermions and gauge fields. The naive momentum cutoff spoils gauge invariance of the effective action. This can be dealt with by an improved momentum cutoff scheme, where one uses Schwinger parameterization of the propagator, and introduce a cutoff function on the Schwinger parameter, namely

$$\frac{1}{p^2+m^2} = \int_0^\infty dt\, e^{-t(p^2+m^2)} \to \int_0^\infty dt\, f_\Lambda(t)\, e^{-t(p^2+m^2)}. \tag{6.192}$$

$f_\Lambda(t)$ is a function that takes the value 1 for $t > \Lambda^{-2}$ and goes to zero sufficiently fast near $t = 0$. In a gauge theory, one can simply replace p^2, or $-\partial^\mu \partial_\mu$ in position space, by the corresponding gauge covariant derivative.

6.9 Irreversibility of RG flows and the c theorem in two dimensions

Wilson's formulation of the renormalization group flow naturally goes from a high energy scale to a lower energy scale by integrating out momentum shells. Intuitively, one expects that the physical degrees of freedom are reduced in this coarse graining procedure, and the flow from one RG fixed point to another (both of which could be defined at all scales since they are RG fixed points) cannot be reversed as long as the renormalization scale keeps decreasing. This idea can be made precise: in two dimension it is called the c-theorem, and in four dimensions it is called the a-theorem. We will discuss the two dimensional c-theorem in this section, and visit the four dimensional a-theorem after discussing spontaneous symmetry breaking and effective field theories.

Let us consider a two dimensional Euclidean quantum field theory, that is related to a Lorentzian unitarity theory by analytic continuation. The stress energy tensor of the theory is denoted by $T_{ij}(x)$, $i, j = 1, 2$. It is convenient to define the complex coordinates $z = x_1 + ix_2$, $\bar{z} = x_1 - ix_2$, and write the components of the stress energy tensor in the complex coordinates,

$$\begin{aligned} T_{zz} &= \frac{\partial x^i}{\partial z}\frac{\partial x^j}{\partial z} T_{ij} = \frac{1}{2}T_{11} - \frac{1}{2}T_{22} - iT_{12}, \\ T_{\bar{z}\bar{z}} &= \frac{\partial x^i}{\partial \bar{z}}\frac{\partial x^j}{\partial \bar{z}} T_{ij} = \frac{1}{2}T_{11} - \frac{1}{2}T_{22} + iT_{12}, \\ T_{z\bar{z}} = T_{\bar{z}z} &= \frac{\partial x^i}{\partial z}\frac{\partial x^j}{\partial \bar{z}} T_{ij} = \frac{1}{2}T_{11} + \frac{1}{2}T_{22}. \end{aligned} \tag{6.193}$$

Instead of writing $T_{ij}(x)$, we will now write $T_{zz}(z, \bar{z})$ etc., where both z and $\bar{z}$ dependence are exhibited, emphasizing that T_{zz} is a general function of z and $\bar{z}$ (not necessarily holomorphic, for instance). Now consider the Euclidean two-point functions

$$\begin{aligned} F(|z|^2) &= z^4 \langle T_{zz}(z,\bar{z}) T_{zz}(0) \rangle, \\ G(|z|^2) &= 4|z|^2 z^2 \langle T_{zz}(z,\bar{z}) T_{z\bar{z}}(0) \rangle, \\ H(|z|^2) &= 16|z|^4 \langle T_{z\bar{z}}(z,\bar{z}) T_{z\bar{z}}(0) \rangle. \end{aligned} \tag{6.194}$$

We have used rotational invariance to write them as functions of $|z|^2 = x^2$. Note that current conservation, $\partial_i T_{ij} = 0$, is expressed in complex coordinates as

$$\partial_{\bar{z}} T_{zz} + \partial_z T_{z\bar{z}} = 0. \tag{6.195}$$

It is straightforward to check that current conservation implies

$$\begin{aligned} &4F' + G' - 3G = 0, \\ &4G' - 4G + H' - 2H = 0. \end{aligned} \tag{6.196}$$

Here $F' \equiv dF(|z|^2)/d(|z|^2)$, etc. Zamolodchikov defined the C function

$$C(|z|^2) = 2F - G - \frac{3}{8} H. \tag{6.197}$$

It has the property

$$\frac{dC}{d|z|^2} = -\frac{3}{4} H. \tag{6.198}$$

H may also be written as

$$H(x^2) = 4x^4 \langle T_{ii}(x) T_{jj}(0) \rangle \tag{6.199}$$

The trace of the stress energy tensor is a real scalar operator. Its Euclidean two-point function is positive definite, unless $T_{ii}(x)$ itself vanishes. This positivity condition is a general property of two-point functions in unitarity quantum field theory. One way to see this is to start from the Lorentzian signature, and consider the vacuum two-point function of a real operator $\mathcal{O}$,

$$\begin{aligned} \langle 0|\mathcal{O}(t)\mathcal{O}(0)|0\rangle &= \sum_n \langle 0|\mathcal{O}(t)|n\rangle\langle n|\mathcal{O}(0)|0\rangle \\ &= \sum_n e^{-iE_n t} \langle 0|\mathcal{O}(0)|n\rangle\langle n|\mathcal{O}(0)|0\rangle \end{aligned} \tag{6.200}$$

Wick rotating $t \to -i\tau$, for $\tau > 0$, gives the Euclidean two point function for $\mathcal{O}_E(\tau) = \mathcal{O}(-i\tau)$,

$$\langle \mathcal{O}_E(\tau)\mathcal{O}_E(0)\rangle_E = \sum_n e^{-E_n \tau} \langle 0|\mathcal{O}(0)|n\rangle\langle n|\mathcal{O}(0)|0\rangle, \tag{6.201}$$

which is clearly positive definite.

The positivity of H then implies that $C(|z|^2)$ decreases monotonously with $|z|^2$. We may take $|z|$ to be the length scale of the renormalization point. At an RG fixed point, $C(|z|^2)$ is invariant under rescaling of $|z|$, i.e. it is a constant. The value of C at an RG fixed point is called the central charge, commonly denoted by c. We also see that at the RG fixed point, the trace of the stress energy tensor vanishes as an operator, i.e. $T_{z\bar{z}} = 0$. Such a stress energy tensor gives rise to a larger set of global symmetries than Poincaré symmetry, namely, conformal symmetry in two dimensions. The statement that the central charge decreases along the RG flow (to lower energies) is part of the c theorem. If we perturb the theory by turning on a set of coupling constants λ_i multiplying operators $\mathcal{O}_i$, the C function will be a function of the λ_i's. One can show that the RG flow of λ_i is the *gradient flow* of the potential function C on the space of couplings, with a metric G_{ij} given in terms of the two-point function of $\mathcal{O}_i$ with $\mathcal{O}_j$ (G_{ij} is called the Zamolodchikov metric).

7 Spontaneous symmetry breaking and effective field theory

The vacuum states of a quantum field theory may not preserve all internal symmetries of the theory. This is called spontaneous symmetry breaking. When this happens, there are always degeneracy vacuum states, related by the "broken" symmetry. The possibility of degenerate vacuum states usually can only occur in quantum field theories in infinite space. In finite space, even if the classical potential of the field theory has degenerate minima, there are generically nonzero tunneling amplitudes between these classical minima, which breaks the degeneracy of the ground state. In infinity space, such tunneling amplitude goes to zero.

When there are degenerate vacua in a quantum field theory, the physical Hilbert space splits into superselection sectors, and the matrix elements of local operators vanish identically between vacuum states corresponding to the different superselection sectors. This can be seen as follows. Let $|u\rangle, |v\rangle, |w\rangle, \cdots$ be degenerate vacuum states. They are assumed to be translation invariant. The matrix element of the product of two local operators $A(\vec{x}), B(\vec{y})$ at equal time between two vacuum states can be written as

$$\langle u|A(\vec{x})B(\vec{y})|v\rangle = \sum_w \langle u|A(0)|w\rangle\langle w|B(0)|v\rangle + \int d^3\vec{p} \sum_n \langle u|A(0)|n,\vec{p}\rangle\langle n,\vec{p}|B(0)|v\rangle e^{-ip\cdot(x-y)}. \tag{7.1}$$

The first term on the RHS is a sum of degenerate vacuum states $|w\rangle$. The second term is a sum over all intermediate excited states of total momentum $\vec{p}$, and an integration over

$\vec{p}$. Assuming the $\vec{p}$ dependence is sufficiently smooth in the matrix element involving the excited states, in the limit $|\vec{x}-\vec{y}| \to \infty$, the integral over $\vec{p}$ goes to zero. We conclude that

$$\langle u|A(\vec{x})B(\vec{y})|v\rangle \longrightarrow \sum_w \langle u|A(0)|w\rangle\langle w|B(0)|v\rangle, \qquad |\vec{x}-\vec{y}| \to \infty. \tag{7.2}$$

The same argument also says

$$\langle u|B(\vec{y})A(\vec{x})|v\rangle \longrightarrow \sum_w \langle u|B(0)|w\rangle\langle w|A(0)|v\rangle, \qquad |\vec{x}-\vec{y}| \to \infty. \tag{7.3}$$

On the other hand, causality requires the equal time commutator between $A(\vec{x})$ and $B(\vec{y})$ to vanish when $\vec{x} \neq \vec{y}$. It follows that the transition matrix between vacuum states $\langle u|A(0)|v\rangle$ and $\langle u|B(0)|v\rangle$ must commute. It is then possible to find a basis for vacuum states, still denote by $|u\rangle, |v\rangle$, etc., that simultaneously diagonalizes all local operators, i.e.

$$\langle u|A(0)|v\rangle \propto \delta_{uv}, \tag{7.4}$$

for all local operators $A(x)$.

With $|u\rangle$ being one of the vacuum states in this basis, we have

$$\langle u|A(\vec{x})B(\vec{y})|u\rangle \longrightarrow \langle u|A(\vec{x})|u\rangle\langle u|B(\vec{x})|u\rangle, \qquad |\vec{x}-\vec{y}| \to \infty. \tag{7.5}$$

This allows for cluster decomposition and is needed to ensure that the S-matrix is well defined. Cluster decomposition would fail had we chosen the vacuum state to be a general linear combination of the basis $|u\rangle$ states.

7.1 Goldstone boson

We begin with the investigation of spontaneously broken continuous symmetry. Let $\phi_n(x)$ be a set of real scalar fields in the theory with a global symmetry whose action in the infinitesimal form is *linear*:

$$\phi_n(x) \to \phi_n(x) + i\epsilon \sum_m t_{nm}\phi_m(x). \tag{7.6}$$

In terms of the generating functional $Z[J]$, this symmetry transformation is a change of path integration variable which leaves $Z[J]$ invariant:

$$\begin{aligned} 0 = \frac{\delta_\epsilon Z[J]}{Z[J]} &= \frac{1}{Z[J]}\int \mathcal{D}\phi e^{iS[\phi]+i\int d^4x\phi_n(x)J_n(x)}(-\epsilon)\int d^4x \sum_{n,m} t_{nm}\phi_m(x)J_n(x) \\ &= -\epsilon\sum_{n,m} t_{nm}\int d^4x\langle\phi_m(x)\rangle_J J_n(x). \end{aligned} \tag{7.7}$$

Here $\langle \cdots \rangle_J$ stands for the expectation value in the presence of the source $J(x)$. Using

$$J_n(x) = -\left.\frac{\delta\Gamma[\phi]}{\delta\phi_n(x)}\right|_{\phi=\langle\phi\rangle_J}, \tag{7.8}$$

where $\Gamma[\phi]$ is the 1PI effective action, we can express (7.7) as

$$\sum_{n,m}\int d^4x \frac{\delta\Gamma[\phi]}{\delta\phi_n(x)} t_{nm}\phi_m(x) = 0. \tag{7.9}$$

In other words, the effective action is also invariant under the symmetry transformation (7.6). This is sometimes called Slavnov-Taylor identity. Note that we have assumed that the symmetry transformation is linear in deriving (7.9).

When ϕ_n are taken to be constant valued, we have expressed the effective action in terms of the effective potential $V(\phi)$, via

$$\Gamma[\phi] = -\mathcal{V}\, V(\phi), \tag{7.10}$$

where $\mathcal{V}$ is the spacetime volume. Specializing (7.9) to this case, we have

$$\sum_{n,m} \frac{\partial V(\phi)}{\partial \phi_n} t_{nm}\phi_m = 0. \tag{7.11}$$

Taking another derivative with respect to ϕ_ℓ, we obtain

$$\sum_n \frac{\partial V}{\partial\phi_n} t_{n\ell} + \sum_{n,m}\frac{\partial^2 V}{\partial\phi_n\partial\phi_\ell} t_{nm}\phi_m = 0. \tag{7.12}$$

Since the vacuum expectation value $\overline{\phi}_n = \langle\phi_n\rangle_{vac}$ minimizes $V(\phi)$, we derive

$$\sum_{n,m}\left.\frac{\partial^2 V}{\partial\phi_n\partial\phi_\ell}\right|_{\phi=\overline{\phi}} t_{nm}\overline{\phi}_m = 0. \tag{7.13}$$

In other words, the effective mass matrix $\partial^2 V/\partial\phi_n\partial\phi_\ell|_{\phi=\overline{\phi}}$ has a zero eigenvalue associated with each symmetry transformation that acts nontrivially on the vacuum (so that $\sum_m t_{nm}\overline{\phi}_m$ is nonzero).

More precisely, the Hessian of the effective potential at its minimum is the inverse of the quantum corrected ϕ propagator at zero momentum,

$$\left.\frac{\partial^2 V}{\partial\phi_n\partial\phi_\ell}\right|_{\phi=\overline{\phi}} = \Delta^{-1}_{n\ell}(0). \tag{7.14}$$

(7.13) means that the propagator $\Delta_{n\ell}(q)$ has a pole at $q^2 = 0$ associated with each nonzero eigenvector of $\Delta^{-1}(0)$, $\sum_m t_{nm}\overline{\phi}_m$. We see that associated with each spontaneously broken symmetry, there is a *massless* scalar field. This massless scalar particle is called the Nambu-Goldstone boson. Our argument based on effective action shows that this statement is exact in the quantum theory, and the masslessness is protected from renormalization.

A classic example of spontaneously broken symmetry is scalar fields interacting through a "Mexican hat" potential. Consider a theory of N real scalar fields ϕ_n, $n = 1, \cdots, N$, with Lagrangian

$$\mathcal{L} = -\frac{1}{2}\sum_n (\partial_\mu \phi_n)^2 - \frac{1}{2}m^2 \sum_n \phi_n\phi_n - \frac{g}{4}\left(\sum_n \phi_n\phi_n\right)^2 . \tag{7.15}$$

When $m^2 < 0$, the classical potential has its minima at $\phi_n = \overline{\phi}_n$, with

$$\sum_n \overline{\phi}_n\overline{\phi}_n = -\frac{m^2}{g}. \tag{7.16}$$

The mass matrix is

$$\begin{aligned} M^2_{nm} &= \left.\frac{\partial^2 V(\phi)}{\partial\phi_n\partial\phi_m}\right|_{\phi=\bar{\phi}} \\ &= m^2\delta_{nm} + g\delta_{nm}\sum_\ell \overline{\phi}_\ell\overline{\phi}_\ell + 2g\overline{\phi}_n\overline{\phi}_m \\ &= 2g\overline{\phi}_n\overline{\phi}_m. \end{aligned} \tag{7.17}$$

It has one eigenvalue of nonzero eigenvalue, namely ϕ_n itself, and the remaining $N-1$ eigenvalues are zero. If $\overline{\phi}$ acquires a vacuum expectation value obeying (7.16), the $O(N)$ symmetry that rotates all ϕ_n's is broken to $O(N-1)$. The number of broken symmetries is the difference between the dimension of the group $O(N)$ and $O(N-1)$,

$$\dim O(N) - \dim O(N-1) = \frac{N(N-1)}{2} - \frac{(N-1)(N-2)}{2} = N - 1. \tag{7.18}$$

This is indeed the number of massless fields in the vacuum labelled by $\overline{\phi}_n$ obeying (7.16).

A word of caution: we have argued before that the quantum effective potential is convex, and in fact the Mexican hat potential should be corrected in such a way that the true effective potential is constant in the region bounded by the minima of the classical potential. This is simply the statement that one can take linear combinations of the vacua in which $\overline{\phi}_n$ obey (7.16), and in such a state the expectation value of ϕ_n

lies in the domain bounded by (7.16) while the energy density stays the same. But we will only consider the basis of vacua obeying (7.16) which diagonalizes the matrix elements of local operators and in which cluster decomposition holds.

Now we will give a different proof of Goldstone's theorem, which does not make use of the effective action, but the current operator corresponding to the global symmetry. The method of this proof will be useful in studying the effective field theory of Nambu-Goldstone bosons later.

The symmetry (7.6) gives rise to a conserved Noether current $J_\mu(x)$, with

$$\partial_\mu J^\mu(x) = 0. \tag{7.19}$$

Acting on the Hilbert space, the symmetry generator is represented by the charge operator,

$$Q = \int d^3\vec{x} J^0(\vec{x}, t) \tag{7.20}$$

which is independent of time t by the conservation relation (7.19). (7.6) is now expressed in operator language as

$$[Q, \phi_n(x)] = -\sum_m t_{nm}\phi_m(x). \tag{7.21}$$

This is an operator relation, and does not involve a choice of vacuum state. It holds even when the symmetry is broken by the vacuum.

Suppose $|0\rangle$ is a vacuum state, in which the symmetry associated with J^μ is spontaneously broken. This is same as the statement that $Q|0\rangle \neq 0$. We will now show that $J^\mu(x)|0\rangle$ contains the one-particle state of a massless Nambu-Goldstone boson. The idea is to consider vacuum matrix elements of $J^\mu(x)$ with $\phi_n(y)$, as in our earlier discussion of Källen-Lehmann spectral representation. We can write

$$\begin{aligned}\langle 0| \left[J^\mu(x), \phi_n(y)\right] |0\rangle = \sum_N \big[& e^{ip_N\cdot(x-y)}\langle 0|J^\mu(0)|N\rangle\langle N|\phi_n(0)|0\rangle \\ & -e^{ip_N\cdot(y-x)}\langle 0|\phi_n(0)|N\rangle\langle N|J^\mu(0)|0\rangle\big] ,\end{aligned} \tag{7.22}$$

where $|N\rangle$ is a general excited state, with four-momentum p_N obeying $p_N^2 < 0$ and $p_N^0 > 0$. Since J^μ transforms as a Lorentz vector, we may express the summation of matrix elements in terms of a spectral function,

$$\sum_N \langle 0|J^\mu(0)|N\rangle\langle N|\phi_n(0)|0\rangle\delta^4(p - p_N) = \frac{i}{(2\pi)^3}p^\mu\rho_n(-p^2)\theta(p^0), \tag{7.23}$$

and

$$\sum_N \langle 0|\phi_n(0)|N\rangle\langle N|J^\mu(0)|0\rangle\delta^4(p - p_N) = \frac{i}{(2\pi)^3}p^\mu\tilde{\rho}_n(-p^2)\theta(p^0). \tag{7.24}$$

Using

$$\frac{i}{(2\pi)^3}\int d^4p\, e^{ip\cdot x}p_\mu\theta(p^0)\delta(p^2+\mu^2)=\frac{\partial}{\partial x^\mu}\Delta_+(x;\mu^2), \tag{7.25}$$

we can write (7.22) as

$$\langle 0|\left[J^\mu(x),\phi_n(y)\right]|0\rangle=\frac{\partial}{\partial x_\mu}\int d\mu^2\left[\rho_n(\mu^2)\Delta_+(x-y;\mu^2)+\tilde{\rho}_n(\mu^2)\Delta_+(y-x;\mu^2)\right]. \tag{7.26}$$

As before, when $x-y$ is spacelike, $\Delta_+(x-y;\mu^2)$ is even in $(x-y)$, and we can combine the two terms on the RHS. On the other hand, the commutator between spacelike separated $J^\mu(x)$ and $\phi_n(y)$ must vanish, from which we conclude

$$\tilde{\rho}_n(\mu^2)=-\rho_n(\mu^2). \tag{7.27}$$

So for general (non-spacelike separated) x,y, we have

$$\langle 0|\left[J^\mu(x),\phi_n(y)\right]|0\rangle=\frac{\partial}{\partial x_\mu}\int d\mu^2\rho_n(\mu^2)\left[\Delta_+(x-y;\mu^2)-\Delta_+(y-x;\mu^2)\right]. \tag{7.28}$$

When J^μ and ϕ_n are Hermitian operators, it also follows that $\rho_n(\mu^2)$ is real.

Now applying current conservation gives

$$\begin{aligned}0&=\frac{\partial}{\partial x^\mu}\langle 0|\left[J^\mu(x),\phi_n(y)\right]|0\rangle\\&=\int d\mu^2\rho_n(\mu^2)\,\Box_x\left[\Delta_+(x-y;\mu^2)-\Delta_+(y-x;\mu^2)\right]\\&=\int d\mu^2\mu^2\rho_n(\mu^2)\left[\Delta_+(x-y;\mu^2)-\Delta_+(y-x;\mu^2)\right].\end{aligned} \tag{7.29}$$

From this we conclude that

$$\mu^2\rho_n(\mu^2)=0, \tag{7.30}$$

and so $\rho_n(\mu^2)$ is supported only at $\mu^2=0$. On the other hand, we can relate (7.28) to the commutator between Q and ϕ_n: consider the equal time commutator

$$\begin{aligned}\langle 0|\left[J^0(\vec{x},t),\phi_n(\vec{y},t)\right]|0\rangle&=-\int d\mu^2\rho_n(\mu^2)\,\frac{\partial}{\partial t'}\left[\Delta_+(\vec{x}-\vec{y},t';\mu^2)+\Delta_+(\vec{y}-\vec{x},t';\mu^2)\right]\Big|_{t'=0}\\&=i\delta^3(\vec{x}-\vec{y})\int d\mu^2\rho_n(\mu^2).\end{aligned} \tag{7.31}$$

Integrating over $\vec{x}$ then gives

$$-\sum_m t_{nm}\overline{\phi}_m=\langle 0|\left[Q,\phi_n(\vec{y},t)\right]|0\rangle=i\int d\mu^2\rho_n(\mu^2), \tag{7.32}$$

where $\overline{\phi}_m = \langle 0|\phi_m|0\rangle$. If the symmetry generated by Q is spontaneously broken in that it acts nontrivially on the vacuum expectation value of $\phi_n(x)$, then $\rho_n(\mu^2)$ cannot be identically zero. In fact, we see that

$$\rho_n(\mu^2) = i\delta(\mu^2)\sum_m t_{nm}\overline{\phi}_m. \tag{7.33}$$

The existence of a contribution to $\rho_n(\mu^2)$ at $\mu^2 = 0$ form some state $|N\rangle$ means that there is a massless particle, otherwise $\rho_n(\mu^2)$ would vanish for values of μ^2 up to the mass squared of the lowest mass particle. Furthermore, since $\phi_n(x)$ is a scalar field, $\langle N|\phi_n(0)|0\rangle$ is non-vanishing only if N has zero helicity $(\vec{J}\cdot\hat{p})$. When $|N\rangle$ is a one-particle state, having nonzero overlap with $\phi_n|0\rangle$ means that this particle is a scalar. This massless scalar is the Nambu-Goldstone boson associated with the broken symmetry Q.

Let $|B,\vec{p}_B\rangle$ be the one-particle state of a Goldstone boson with momentum $\vec{p}_B$ (and energy $p_B^0 = |\vec{p}_B|$). The one-particle momentum eigenstates are delta function normalized as before. Lorentz invariance then determines the overlap between $|B,\vec{p}_B\rangle$ and $J^\mu(x)|0\rangle$ to be of the form

$$\langle 0|J^\mu(x)|B,\vec{p}_B\rangle = i\frac{Fp_B^\mu e^{ip_B\cdot x}}{(2\pi)^{\frac{3}{2}}\sqrt{2p_B^0}}, \tag{7.34}$$

where F is a constant (there are no Lorentz invariant variables it can depend on). Note that since $p_B^2 = 0$, the above formula is consistent with current conservation $\partial_\mu J^\mu(x) = 0$.

Now consider a scattering process in which an asymptotic state $|\alpha\rangle$ turns into the state $|\beta\rangle$ and emits a Goldstone boson with four-momentum q^μ. Since $J^\mu(x)$ is a field operator that creates a one Goldstone boson plus other states, the transition matrix element

$$\langle\beta|J^\mu(x)|\alpha\rangle \tag{7.35}$$

has an LSZ pole at $q^2 = 0$, where $q^\mu = p_\alpha^\mu - p_\beta^\mu$. The LSZ reduction formula then says in the limit $q^2 \to 0$,

$$\langle\beta|J^\mu(x)|\alpha\rangle \to \frac{iFq^\mu e^{iq\cdot x}}{q^2}M_{\beta,B|\alpha} \tag{7.36}$$

where $M_{\beta,B|\alpha}$ is related to the S-matrix element for the transition from α to β with the emission of a Goldstone boson of momentum q^μ by

$$i(2\pi)^4\delta^4(p_\alpha - p_\beta - q)\frac{M_{\beta,B|\alpha}}{(2\pi)^{\frac{3}{2}}\sqrt{2q^0}} = S(\beta,B;\alpha). \tag{7.37}$$

The factor $\frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{2q^0}}$ is usually attached with the emission of a boson in relating the S-matrix element to the Green's function, and here we have separated it from the definition of $M_{\beta,B|\alpha}$. For instance, in perturbation theory, if we are to treat the Goldstone boson as an elementary field, $M_{\beta,B|\alpha}$ would be computed using the vertex of emitting a Goldstone boson, but without attaching any extra factor to the external Goldstone boson line.

Generally, $J^\mu(x)|0\rangle$ (or rather, in our notation here, $\langle 0|J^\mu(x)$) also contains other, multi-particle, states. The full transition matrix element of J^μ between $|\alpha\rangle$ and $|\beta\rangle$ takes the form

$$\langle\beta|J^\mu(x)|\alpha\rangle = e^{iq\cdot x}N^\mu_{\beta\alpha} + \frac{iFq^\mu e^{iq\cdot x}}{q^2}M_{\beta,B|\alpha}, \tag{7.38}$$

where $N^\mu_{\beta\alpha}$ is the contribution from other than a Goldstone boson intermediate state, that is annihilated by $J^\mu(x)$. The conservation of the current implies

$$M_{\beta,B|\alpha} = \frac{i}{F}q_\mu N^\mu_{\beta\alpha}. \tag{7.39}$$

In the limit of a soft Goldstone boson emission, namely $q\to 0$, $M_{\beta,B|\alpha}$ is nonzero only if $N^\mu_{\beta\alpha}$ has a pole at $q=0$. Generically, such contribution come only from $J^\mu(x)$ turning on one-particle initial or final state into another one-particle state; in other words, J^μ is attached to an external line. For example, J^μ may act on an incoming particle of momentum p and mass m and turns it into a particle of momentum $p+q$ and the same mass. Such a contribution would involve an internal propagator of the particle with momentum $p+q$, that is

$$\frac{1}{(p+q)^2+m^2-i\epsilon} = \frac{1}{2p\cdot q+q^2-i\epsilon} \longrightarrow \frac{1}{2p\cdot q}, \tag{7.40}$$

which has a pole at $q=0$. From this we learn that the emission amplitude of a soft Goldstone boson can be determined in terms of the transition amplitude of J^μ between on-shell one-particle states.

We may define a set of renormalized Goldstone boson fields $\pi_a(x)$, so that their matrix element between the vacuum $|0\rangle$ and a one-Goldstone-boson state $|B_a,\vec{p}_B\rangle$ (here a labels the broken symmetry generator) is

$$\langle B_a,\vec{p}_B|\pi_b(x)|0\rangle = \delta_{ab}\frac{e^{-ip_B\cdot x}}{(2\pi)^{\frac{3}{2}}\sqrt{2p^0_B}}. \tag{7.41}$$

This of course does not specify $\pi_a(x)$ entirely; $\pi_a(x)$ may also create multi-particle states, but such ambiguity just amounts to possible nonlinear redefinitions of the field. The fundamental fields ϕ_n generally also creates Goldstone bosons:

$$\langle B_a,\vec{p}_B|\phi_n(x)|0\rangle = Z_{an}\frac{e^{-ip_B\cdot x}}{(2\pi)^{\frac{3}{2}}\sqrt{2p^0_B}}. \tag{7.42}$$

We may then write

$$\phi_n(x) = \sum_a Z_{an}\pi_a(x) + \cdots , \tag{7.43}$$

where $\cdots$ stands for fields operators that create other than Goldstone bosons. Z_{an} is tied to the matrix elements of the current J^μ,

$$\langle 0|J_a^\mu(x)|B_b, \vec{p}_B\rangle = iF_{ab}\frac{p_B^\mu e^{ip_B\cdot x}}{(2\pi)^{\frac{3}{2}}\sqrt{2p_B^0}}. \tag{7.44}$$

This can be seen from restricting (7.23) to the limit $p^2 \to 0$,

$$\begin{aligned}
&\sum_b \int d^3\vec{p}_B \langle 0|J_a^\mu(0)|B_b, \vec{p}_B\rangle\langle B_b, \vec{p}_B|\phi_n(0)|0\rangle\delta^4(p-p_B) = \frac{i}{(2\pi)^3}p^\mu\rho_{an}(-p^2)\theta(p^0) \\
&= \frac{i}{(2\pi)^3}\sum_b F_{ab}Z_{bn}\int \frac{d^3\vec{p}_B}{2p_B^0}p_B^\mu\delta^4(p-p_B) = \frac{i}{(2\pi)^3}\sum_b F_{ab}Z_{bn}\frac{p^\mu}{2|\vec{p}|}\delta(p^0-|\vec{p}|).
\end{aligned} \tag{7.45}$$

Combining this with

$$\rho_{an}(-p^2) = i\delta(-p^2)\sum_m t^a_{nm}\overline{\phi}_m, \tag{7.46}$$

we obtain the relation

$$\sum_b F_{ab}Z_{bn} = i\sum_m t^a_{nm}\overline{\phi}_m. \tag{7.47}$$

When a label a set of independent broken global symmetries, F_{ab} is invertible, and so we can write $Z_{an} = \sum_b F^{-1}_{ab}it^b_{nm}\overline{\phi}_m$, and

$$\phi_n(x) = \sum_{a,b} F^{-1}_{ab}it^b_{nm}\overline{\phi}_m\pi_a(x) + \cdots . \tag{7.48}$$

The scattering amplitude of N Goldstone boson can then be extract from the LSZ poles of the Green function of N $\phi_n(x)$'s. In the zero momentum limit, this is computed from the N-th derivative of the effective potential

$$\left.\frac{\partial^N V(\phi)}{\partial\phi_{n_1}\cdots\partial\phi_{nN}}\right|_{\phi=\overline{\phi}} \tag{7.49}$$

contracted with the factors $F^{-1}_{a_ib_i}it^{b_i}_{n_im_i}\overline{\phi}_{m_i}$. But this is zero by (7.11). Hence we see that, generally, the amplitude of Goldstone bosons at *zero* momentum vanishes. This implies that the effective field theory describing Goldstone bosons can only involve derivative interactions, as we will explore later.

7.2 Pion

While the statement of Goldstone's theorem may seem rather obvious in the example of scalars interacting through a Mexican hat potential, the general theory of spontaneously symmetry breaking is powerful in that the dynamics of Goldstone bosons at low energies can be understood based on symmetries alone, even when the nature of the symmetry breaking is complicated. A nontrivial example is the spontaneously breaking of the "chiral symmetry" in QCD. While the chiral symmetry is a global symmetry of the QCD Lagrangian with *massless* quarks, it is believed to be broken spontaneously by the vacuum due to strong coupling dynamics. The Goldstone bosons associated with the broken chiral symmetry is the pion, which would be exactly massless if the quarks were massless. When the quarks acquire a small mass, the chiral symmetry will be an approximate symmetry, and the pions in this case are pseudo-Goldstone bosons.

To begin with let us consider QCD with two flavors, u and d quarks. The masses of u and d are small and we will ignore them for the moment. The Lagrangian is

$$\mathcal{L} = -\bar{u}\not{D}u - \bar{d}\not{D}d + \cdots , \tag{7.50}$$

where $\not{D} = \gamma^\mu D_\mu = \gamma^\mu(\partial_\mu - igA_\mu^a T^a)$. T^a are the generators of $SU(3)$ gauge group. It should not be confused with the global flavor symmetry (isospin) we are about to discuss. The omitted terms in the Lagrangian do not involve u and d quarks. Then we have an $SU(2)\times SU(2)$ *global* symmetry of the Lagrangian, under which u, d transform as

$$\begin{pmatrix} u \\ d \end{pmatrix} \to \exp\left(i\vec{\theta}^V \cdot \vec{t} + i\gamma_5\vec{\theta}^A \cdot \vec{t}\right)\begin{pmatrix} u \\ d \end{pmatrix}, \tag{7.51}$$

where $\vec{\theta}^V, \vec{\theta}^A$ are three dimensional real vectors, and $\vec{t} = \frac{1}{2}\vec{\sigma}$ are the $SU(2)$ isospin generators, acting on (u, d) as a doublet. γ_5 on the other hand acts on the spinor indices of u and d. The symmetry is seen more clearly if we write ψ_i, $i = 1, 2$ for u, d, and define the left and right handed parts

$$\psi_{iL} = \frac{1+\gamma_5}{2}\psi_i, \qquad \psi_{iR} = \frac{1-\gamma_5}{2}\psi_i. \tag{7.52}$$

The Lagrangian is then written

$$\mathcal{L} = -\bar{\psi}^i_L\not{D}\psi_{iL} - \bar{\psi}^i_R\not{D}\psi_{iR} + \cdots . \tag{7.53}$$

where we wrote the flavor index for the conjugate quark fields as an upper index. We then have the symmetry $SU(2)_L \times SU(2)_R$, under which

$$\psi_{iL} \to (U_L)_i{}^j\psi_{jL}, \qquad \psi_{iR} \to (U_R)_i{}^j\psi_{jR}. \tag{7.54}$$

Here U_L and U_R are $SU(2)$ matrices, related to θ^V, θ^A by

$$U_{L,R} = \exp\left[i(\vec{\theta}^V \pm \vec{\theta}^A) \cdot \vec{t}\right]. \tag{7.55}$$

Note that a mass term would take the form $m_{ij}\bar{\psi}_{iL}\psi_{jR}$, which breaks the axial or chiral $SU(2)$ symmetry (parameterized by $\vec{\theta}^A$) even if the masses of the two flavors were identical. With small masses for u, d, the $SU(2)_L \times SU(2)_R$ is an approximate global symmetry.

A word of caution is that chiral symmetry, which acts differently on left handed versus right handed fermions, is potentially subject to *anomalies*. Namely, it could be a symmetry of the classical Lagrangian, but not a symmetry of the path integral measure and not a symmetry of the quantum theory. We will see later that the chiral symmetry of the massless QCD is anomaly free, though it becomes anomalous when the coupling of the quarks to the $U(1)$ electromagnetic field is introduced!

Even though the chiral symmetry is a true symmetry of massless QCD, it is believed to be *spontaneously* broken by the vacuum, and only the diagonal $SU(2) \subset SU(2)_L \times SU(2)_R$ (parameterized by $\vec{\theta}^V$ survives). The mechanism for the spontaneous chiral symmetry breaking involves strong coupling dynamics, because QCD with two flavors is strongly coupled at low energies, and would be difficult to understood using perturbation theory. Nonetheless, a great deal of information can be extracted, and predictions can be made, by merely assuming this spontaneous breaking.

To proceed let us write down the currents for the diagonal $SU(2)$ and the axial $SU(2)$,

$$\vec{J}^V_\mu = i\bar{\psi}\gamma_\mu\vec{t}\psi, \qquad \vec{J}^A_\mu = i\bar{\psi}\gamma_\mu\gamma_5\vec{t}\psi, \tag{7.56}$$

which are conserved. Here we have suppressed the flavor index of $\psi = (u, d)$, and it is understood that $\vec{t}$ acts on the flavor indices. The currents of $SU(2)_L$ and $SU(2)_R$ are $\frac{1}{2}(\vec{J}^V_\mu \pm \vec{J}^A_\mu)$. The conserved charges are

$$\vec{Q}^V = \int d^3x \vec{J}^{V0}, \qquad \vec{Q}^A = \int d^3x \vec{J}^{A0}. \tag{7.57}$$

The currents are normalized so that the corresponding conserved charges obey the commutation relations of $SU(2) \times SU(2)$, expressed in this basis of generators as

$$\begin{aligned} [Q^V_i, Q^V_j] &= i\epsilon_{ijk}Q^V_k, \\ [Q^V_i, Q^A_j] &= i\epsilon_{ijk}Q^A_k, \\ [Q^A_i, Q^A_j] &= i\epsilon_{ijk}Q^V_k. \end{aligned} \tag{7.58}$$

Note that these commutation relations unambiguously fix the normalization of the currents.

The statement of chiral symmetry breaking is that $\vec{Q}^A|0\rangle \neq 0$. In fact, if $\vec{Q}^A$ were a symmetry of the vacuum, then associated with each hadron state $|h\rangle$ there would be a degenerate hadron state $\vec{Q}^A|h\rangle$ with the opposite parity (since $\vec{Q}^A$ is parity odd) but all other quantum numbers equal to those of $|h\rangle$. This has not been observed (approximately, since the chiral symmetry is an approximate symmetry due to quark masses), indicating that $\vec{Q}^A$ must be spontaneously broken. If $\vec{Q}^A$ is spontaneously broken, then as we have seen, $\vec{J}^A_\mu$ creates an $SU(2)$ triplet of Goldstone bosons (and multi-particle states). These Goldstone bosons are the pions. Of course, pions in the real world are not massless; this is because the chiral symmetry is only approximate due to quark masses, and the pions are really pseudo-Goldstone bosons. Nonetheless, the masses of pions being much smaller than those of baryons indicates that the chiral symmetry is a valid approximation.

Let us consider the matrix element of $\vec{J}^A_\mu$ between a one pion state $|\pi_j\rangle$ (j is the isospin triplet index) of four-momentum p^μ_π and the vacuum. We have seen that it takes the form

$$\langle 0|J^A_{i\mu}(x)|\pi_j\rangle = iF_\pi \delta_{ij} \frac{p^\mu_\pi e^{ip_\pi\cdot x}}{2(2\pi)^{\frac{3}{2}}\sqrt{2p^0_\pi}}. \tag{7.59}$$

Note that F_π has dimension (mass), because J^A has dimension (mass)^3 and $|\pi\rangle$ is normalized with dimension (mass)$^{-\frac{3}{2}}$. Since the normalization of the current $\vec{J}^A_\mu$ is fixed by demanding the corresponding conserved charges obey the standard commutation relations in the $SU(2)\times SU(2)$ algebra, the constant F_π is unambiguously defined. It can be measured experimentally, by comparing the rate of pion μ-decay $\pi^+ \to \mu^+ + \nu_\mu$ to the decay $\pi^+ \to \pi^0 + e^+ + \nu_e$. The former is due to the coupling of the axial $SU(2)$ current to leptons, and is proportional to F_π, while the latter involves the matrix element of the vector current (at very small momentum) between $|\pi^+\rangle$ and $|\pi^0\rangle$, which is determined by the isospin Clebsch-Gordan coefficients, and does not depend on F_π. We will see later that in the standard model the axial current couples the same way to different generations of leptons (electrons and muons and their corresponding neutrinos in this case). This makes it possible to compute F_π by comparing the width of the two different π^+ decay channels; it is left as an exercise. The value of F_π determined this way is $F_\pi \simeq 184\,\mathrm{MeV}$.

Consider the quark bilinears

$$\begin{aligned} \Phi^a{}_b &= \bar\psi^a_L \psi_{bR} = \bar\psi^a \frac{1-\gamma_5}{2}\psi_b, \\ \overline{\Phi}^a{}_b &= \bar\psi^a_R \psi_{bL} = \bar\psi^a \frac{1+\gamma_5}{2}\psi_b. \end{aligned} \tag{7.60}$$

$\Phi^a{}_b$ transform in the $(\mathbf{2},\mathbf{2})$ representation of $SU(2)_L\times SU(2)_R \simeq SO(4)$. They can be decomposed into two set of fields transforming in the 4-dimensional real representations of $SO(4)$, which we denote by Φ_I^+ and Φ_I^-, $I = 1, 2, 3, 4$. We have

$$\begin{aligned}
\Phi_i^+ &= i\bar{\psi}\gamma_5\sigma_i\psi, \quad i = 1, 2, 3, \qquad \Phi_4^+ = \frac{1}{2}\bar{\psi}\psi, \\
\Phi_i^- &= \bar{\psi}\sigma_i\psi, \quad i = 1, 2, 3, \qquad \Phi_4^- = -\frac{i}{2}\bar{\psi}\gamma_5\psi.
\end{aligned} \tag{7.61}$$

$\Phi_i^\pm$ transform in the triplet of the diagonal $SU(2)_V$, while $\Phi_4^\pm$ are singlets. The axial $SU(2)_A$ on the other hand transforms $\Phi_4^\pm$ into $\Phi_i^\pm$, $i = 1, 2, 3$.

Now let us take into account the mass of u, d quarks, by adding the interaction Hamiltonian density

$$\begin{aligned}
\mathcal{H}_1 &= m_u\bar{u}u + m_d\bar{d}d \\
&= (m_u + m_d)\Phi_4^+ + (m_u - m_d)\Phi_3^-.
\end{aligned} \tag{7.62}$$

More generally, consider

$$\mathcal{H}_1 = \sum_n u_n\Phi_n, \tag{7.63}$$

where Φ_n denote all the quark bilinears. This leads to a symmetry breaking effective potential,

$$V_1(\phi_n) = \langle\mathcal{H}_1\rangle_{\langle\Phi_n\rangle=\phi_n} = \sum_n u_n\phi_n, \tag{7.64}$$

where the expectation value on $\mathcal{H}_1$ is taken subject to the condition that the expectation value of Φ_n is ϕ_n. The new effective potential, $V(\phi) = V_0(\phi) + V_1(\phi)$, is minimized at a shifted vacuum expectation value $\overline{\phi} = \phi_0 + \phi_1$. We have

$$\left.\frac{\partial V(\phi)}{\partial\phi_n}\right|_{\phi=\phi_0+\phi_1} = 0. \tag{7.65}$$

Expanding this to first order in ϕ_1, we have

$$\sum_m \left.\frac{\partial^2 V(\phi)}{\partial\phi_n\partial\phi_m}\right|_{\phi=\phi_0} \phi_{1m} + u_n = 0. \tag{7.66}$$

Suppose the symmetry generator Q^a acts on ϕ_n as the matrix t^a_{nm}, the invariance of $V_0(\phi)$ under Q^a then implies

$$\sum_n u_n(t^a\phi_0)_n = 0. \tag{7.67}$$

Note that we find that the vacuum expectation value ϕ_0 is no longer free to be any minimum of $V_0(\phi)$, but is constrained by the symmetry breaking perturbation u_n to the Hamiltonian/effective potential. This is called the vacuum alignment condition.

Let $|B_a, \vec{p}_B\rangle$ be the state of a single pseudo-Goldstone bosons created by Φ_n, with

$$\langle B_a, \vec{p}_B|\Phi_n(x)|0\rangle = Z_{an}\frac{e^{-ip_B\cdot x}}{(2\pi)^{\frac{3}{2}}\sqrt{2p_B^0}} \tag{7.68}$$

as before. The mass matrix of the pseudo-Goldstone bosons is then given by

$$
\begin{aligned}
M_{ab}^2 &= \sum_{m,n} Z_{an} Z_{bm} \left. \frac{\partial^2 V(\phi)}{\partial \phi_n \partial \phi_m} \right|_{\phi=\phi_0+\phi_1} \\
&= F_{ac}^{-1} F_{bd}^{-1} (it^c \phi_0)_n (it^d \phi_0)_m \left. \frac{\partial^2 V(\phi)}{\partial \phi_n \partial \phi_m} \right|_{\phi=\phi_0+\phi_1} .
\end{aligned}
\tag{7.69}
$$

To first order in the perturbation V_1, this is

$$
\begin{aligned}
M_{ab}^2 &= F_{ac}^{-1} F_{bd}^{-1} (it^c \phi_0)_n (it^d \phi_0)_m \left[\left. \frac{\partial^3 V_0(\phi)}{\partial \phi_n \partial \phi_m \partial \phi_\ell} \right|_{\phi=\phi_0} \phi_{1\ell} + \left. \frac{\partial^2 V_1(\phi)}{\partial \phi_n \partial \phi_m} \right|_{\phi=\phi_0} \right] \\
&= -F_{ac}^{-1} F_{bd}^{-1} \left[(t^d t^d \phi_0)_n \left. \frac{\partial V_1(\phi)}{\partial \phi_n} \right|_{\phi=\phi_0} + (t^c \phi_0)_n (t^d \phi_0)_m \left. \frac{\partial^2 V_1(\phi)}{\partial \phi_n \partial \phi_m} \right|_{\phi=\phi_0} \right] .
\end{aligned}
\tag{7.70}
$$

In the last step one makes use of the invariance of $V_0(\phi)$ under Q_a, and the stationary condition of $V(\phi)$ at $\phi = \phi_0 + \phi_1$, expanded to first order in ϕ_1. With our linear form of $V_1(\phi)$, we find

$$
M_{ab}^2 = -\sum_{c,d} F_{ac}^{-1} F_{bd}^{-1} \sum_n u_n (t^c t^d \phi_0)_n . \tag{7.71}
$$

In terms of expectation value of operators, it can be written in the form

$$
M_{ab}^2 = -\sum_{c,d} F_{ac}^{-1} F_{bd}^{-1} \langle 0 | \left[Q^c, \left[Q^d, \mathcal{H}_1 \right] \right] | 0 \rangle . \tag{7.72}
$$

Applying this to the chiral $SU(2)$ symmetry breaking, we have

$$
\begin{aligned}
\left[Q_a^A, \left[Q_b^A, \Phi_4^+ \right] \right] &= -\delta_{ab} \Phi_4^+ , \\
\left[Q_a^A, \left[Q_b^A, \Phi_3^- \right] \right] &= -\delta_{b3} \Phi_a^- ,
\end{aligned}
\tag{7.73}
$$

$F_{ab} = \frac{1}{2} F_\pi \delta_{ab}$, and the vacuum alignment condition

$$
\begin{aligned}
&\langle \Phi_{1,2}^\pm \rangle = 0, \\
&(m_u + m_d) \langle \Phi_3^+ \rangle + (m_d - m_u) \langle \Phi_4^- \rangle = 0.
\end{aligned}
\tag{7.74}
$$

The unbroken diagonal $SU(2)$, generated by $\vec{Q}^V$, then implies $\Phi_I^- = 0$ for all I and Φ_4^+ is the only non-vanishing component of Φ_I^+. These imply

$$
\begin{aligned}
M_{ab}^2 &= \frac{4}{F_\pi^2} \left[(m_u + m_d) \delta_{ab} \langle \Phi_4^+ \rangle + (m_u - m_d) \delta_{b3} \langle \Phi_a^- \rangle \right] \\
&= \delta_{ab} \frac{4(m_u + m_d)}{F_\pi^2} \langle \Phi_4^+ \rangle .
\end{aligned}
\tag{7.75}
$$

This formula says that to leading order in the u, d quark masses, the three pions $\pi^0, \pi^\pm$ have the same mass $m_\pi = 2F_\pi^{-1}\sqrt{(m_u + m_d)\langle\Phi_4^+\rangle}$, that depends on the sum of quark masses $m_u + m_d$. As long as m_u, m_d are both small (while their ratio need not be small), chiral symmetry breaking is a good approximation. The observed splitting in the masses of π^0 and $\pi^\pm$ is mainly due to the coupling to electromagnetism.

The non-chiral flavor symmetry, on the other hand, is not spontaneously broken. This can be understood by noting that the non-chiral symmetry is preserved when a mass parameter is turned on for all quarks, and that the path integral for massive fermions coupled to gauge fields is invariant under the flavor symmetry. In this argument, the massiveness of the fermions are needed to show that no singularities occur in the fermion functional determinant nor correlators. Spontaneous symmetry breaking for QCD like theories can only occur for symmetries that are violated by fermion mass terms.

7.3 Chiral perturbation theory

The scattering of Goldstone bosons at low energies is essentially determined by the symmetry breaking pattern entirely. An effective field theory can be formulated in terms of the massless Goldstone bosons only. Let us illustrate this with the example of mesons in QCD with N massless flavors. Here N is not to be confused with the rank of the nonabelian gauge group. The Lagrangian is

$$\mathcal{L} = -\bar{\psi}_L^i \not{D}\psi_{iL} - \bar{\psi}_R^i \not{D}\psi_{iR} + \cdots . \tag{7.76}$$

where the flavor index i is summed over $i = 1, 2, \cdots, N$. The theory has a global flavor symmetry $SU(N)_L \times SU(N)_R$, parameterized by unitary matrices U_L, U_R,

$$\psi_{iL} \to (U_L)_i{}^j \psi_{jL}, \qquad \psi_{iR} \to (U_R)_i{}^j \psi_{jR}. \tag{7.77}$$

There are also $U(1)_L \times U(1)_R$ global symmetry, whose axial part is not spontaneously broken, but rather, anomalous; we will discuss its role later.

When N is not too large compared to the number of colors, the theory is strongly coupled at low energies, and the chiral $SU(N)$ symmetry is believed to be spontaneously broken. In particular, we expect a nonzero vacuum expectation value of the quark-anti-quark bilinear

$$\langle \bar{\psi}_L^i \psi_{jR} \rangle = M_j^i. \tag{7.78}$$

Under $SU(N)_L \times SU(N)_R$, M_j^i is rotated by

$$M_j^i \to (U_R)_j{}^\ell M_\ell^k (U_L^\dagger)_k{}^i. \tag{7.79}$$

We can use the $SU(N)_L \times SU(N)_R$ symmetry to put the matrix $M^i{}_j$ in a diagonal form, and then it must then be proportional to the identity matrix in order for the diagonal $SU(N)$ to be preserved, namely $M^i_j = M\delta^i_j$.

We may write the quark fields in the form

$$\psi_{iL}(x) = (U^{-\frac{1}{2}})_i{}^j(x)\widetilde{\psi}_{jL}(x), \quad \psi_{iR}(x) = (U^{\frac{1}{2}})_i{}^j(x)\widetilde{\psi}_{jR}(x), \tag{7.80}$$

where $U(x)$ is a unitary matrix, and $\widetilde{\psi}_i(x)$ are free of Goldstone modes. The transformation of the quarks under $SU(N)_L \times SU(N)_R$,

$$\psi'_L(x) = U_L\psi_L(x), \quad \psi'_R(x) = U_R\psi_R(x), \tag{7.81}$$

is now represented as

$$\begin{aligned} &U_L U(x)^{-\frac{1}{2}} = (U'(x))^{-\frac{1}{2}}V(x), \quad U_R U(x)^{\frac{1}{2}} = (U'(x))^{\frac{1}{2}}V(x), \\ &\widetilde{\psi}'_{L,R}(x) = V(x)\widetilde{\psi}_{L,R}(x). \end{aligned} \tag{7.82}$$

Clearly, under the chiral symmetry $U_R = U_L^{-1}$, $\widetilde{\psi}_{L,R}(x)$ are invariant. We also see that $U(x)$ itself transforms under $SU(N)_L \times SU(N)_R$ as

$$U'(x) = U_R U(x) U_L^{-1}. \tag{7.83}$$

Mathematically, what we just did is to parameterize the coset

$$\frac{SU(N)_L \times SU(N)_R}{SU(N)} \tag{7.84}$$

(where the quotient is by the diagonal $SU(N)$) by the $SU(N)$ matrix U. This is possible because $(SU(N) \times SU(N))/SU(N)$ has the topology of $SU(N)$ itself.

Now we can construct the effective theory of the Goldstone bosons by writing down the most general $SU(N)_L \times SU(N)_R$ invariant Lagrangian of the field $U(x)$ (and its conjugate $U^\dagger(x)$), in a derivative expansion. It is clear that any such nontrivial term must involve derivatives of $U(x)$, i.e. there is no $SU(N)_L \times SU(N)_R$-invariant potential for $U(x)$. Up to two derivatives, the only invariant Lagrangian takes the form

$$\mathcal{L} = -\frac{1}{16}F^2\mathrm{Tr}(\partial_\mu U \partial^\mu U^\dagger), \tag{7.85}$$

where F is a constant. With four derivatives, there are three independent terms,

$$\left[\mathrm{Tr}(\partial_\mu U \partial^\mu U^\dagger)\right]^2, \quad \mathrm{Tr}(\partial_\mu U \partial_\nu U^\dagger)\mathrm{Tr}(\partial^\mu U \partial^\nu U^\dagger), \quad \mathrm{Tr}(\partial_\mu U \partial^\mu U^\dagger \partial_\nu U \partial^\nu U^\dagger), \tag{7.86}$$

and so forth.

Let us apply this to the $N = 2$ case, i.e. the effective field theory of the pion. We can parameterize the $SU(2)$ matrix valued field $U(x)$ as

$$U(x) = \frac{1 - \vec{\zeta}^2 + 2i\vec{\zeta} \cdot \vec{\sigma}}{1 + \vec{\zeta}^2}. \tag{7.87}$$

Using

$$U^{\frac{1}{2}}(x) = \frac{1 + i\vec{\zeta} \cdot \sigma}{\sqrt{1 + \vec{\zeta}^2}}, \tag{7.88}$$

we can calculate

$$\begin{aligned} U^{-\frac{1}{2}}(\partial_\mu U)U^{-\frac{1}{2}} &= \partial_\mu \vec{\zeta} \cdot \left[U^{-\frac{1}{2}} \frac{-2\vec{\zeta} + 2i\vec{\sigma}}{1 + \vec{\zeta}^2} U^{-\frac{1}{2}} - \frac{2\vec{\zeta}}{1 + \vec{\zeta}^2} \right] \\ &= 2i\vec{\sigma} \cdot \frac{\partial_\mu \vec{\zeta}}{1 + \vec{\zeta}^2}, \end{aligned} \tag{7.89}$$

and so

$$\begin{aligned} \mathcal{L} &= \frac{1}{16} F^2 \mathrm{Tr}(\partial_\mu U\, U^{-1} \partial^\mu U\, U^{-1}) \\ &= -\frac{1}{2} F^2 \frac{\partial_\mu \vec{\zeta} \cdot \partial^\mu \vec{\zeta}}{(1 + \vec{\zeta}^2)^2}. \end{aligned} \tag{7.90}$$

This is the nonlinear sigma model with its target space being the group manifold $SU(2) \simeq S^3$; the target space metric is written in stereograpphic coordinates as

$$ds^2 = \frac{d\vec{\zeta}^2}{(1 + \vec{\zeta}^2)^2}. \tag{7.91}$$

Under infinitesimal $SU(2)_L \times SU(2)_R$ transformations,

$$U \to U + \frac{i}{2} \vec{\theta}_R \cdot \vec{\sigma} U - \frac{i}{2} U \vec{\theta}_L \cdot \vec{\sigma}, \tag{7.92}$$

or in terms of $\vec{\zeta}$,

$$\delta \vec{\zeta} = -\vec{\theta}^V \times \vec{\zeta} + \frac{1}{2} \vec{\theta}^A (1 - \vec{\zeta}^2) + \vec{\zeta}(\vec{\theta}^A \cdot \vec{\zeta}), \tag{7.93}$$

where $\vec{\theta}^{V,A} = \frac{1}{2}(\vec{\theta}_R \pm \vec{\theta}_L)$. Note that the chiral $SU(2)$ is nonlinearly realized on $\vec{\zeta}$. On the other hand, the "covariant derivative"

$$\vec{D}_\mu = \frac{\partial_\mu \vec{\zeta}}{1 + \vec{\zeta}^2} \tag{7.94}$$

transforms linearly under the chiral symmetry, though in a field dependent way,

$$\begin{aligned} \delta \vec{D}_\mu &= -\vec{\theta}^V \times \vec{D}_\mu - \vec{\theta}^A (\vec{\zeta} \cdot \vec{D}_\mu) + \vec{\zeta}(\vec{\theta}^A \cdot \vec{D}_\mu) \\ &= -\vec{\theta}^V \times \vec{D}_\mu - (\vec{\zeta} \times \vec{\theta}^A) \times \vec{D}_\mu. \end{aligned} \tag{7.95}$$

The effective Lagrangian with higher order terms can be written in terms of $\vec{D}_\mu$ as

$$\mathcal{L} = -\frac{F^2}{2}\vec{D}_\mu \cdot \vec{D}^\mu - \frac{c_4}{4}(\vec{D}_\mu \cdot \vec{D}^\mu)^2 - \frac{c_4'}{4}(\vec{D}_\mu \cdot \vec{D}_\nu)(\vec{D}^\mu \cdot \vec{D}^\nu) + \cdots \tag{7.96}$$

This theory is non-renormalizable, in that counter terms for $c_4, c_4', \cdots$ are needed. Nonetheless, we can calculate scattering amplitudes in terms of renormalized $c_4, c_4', \cdots$. The coupling constant $1/F$ has dimension (mass)$^{-1}$, and thus the effective coupling at energy scale Q is Q/F. Let us examine the $2 \to 2$ scattering of pions. To lowest order in perturbation theory, the first term in (7.96) suffices. Define the canonically normalized field

$$\vec{\pi} = F\vec{\zeta}, \tag{7.97}$$

we have

$$\begin{aligned}\mathcal{L} &= -\frac{1}{2}\frac{(\partial_\mu\vec{\pi})^2}{(1+\vec{\pi}^2/F^2)^2} - \frac{c_4}{4F^4}\frac{((\partial_\mu\vec{\pi})^2)^2}{(1+\vec{\pi}^2/F^2)^4} - \frac{c_4'}{4F^4}\frac{(\partial_\mu\vec{\pi}\cdot\partial_\nu\vec{\pi})(\partial^\mu\vec{\pi}\cdot\partial^\nu\vec{\pi})}{(1+\vec{\pi}^2/F^2)^4} + \cdots \\ &= -\frac{1}{2}(\partial_\mu\vec{\pi})^2 + \frac{1}{F^2}\vec{\pi}^2(\partial_\mu\vec{\pi})^2 - \frac{c_4}{4F^4}((\partial_\mu\vec{\pi})^2)^2 - \frac{c_4'}{4F^4}(\partial_\mu\vec{\pi}\cdot\partial_\nu\vec{\pi})(\partial^\mu\vec{\pi}\cdot\partial^\nu\vec{\pi}) + \cdots\end{aligned} \tag{7.98}$$

The tree level S-matrix element is

$$S_{2\to 2} = i(2\pi)^4\delta^4(p_1+p_2-p_3-p_4)\prod_{i=1}^{4}\frac{1}{(2\pi)^{\frac{3}{2}}\sqrt{2E_i}} \times \mathcal{A}_{cd,ab}, \tag{7.99}$$

with the amplitude $\mathcal{A}$ given by

$$\mathcal{A}^{(1)}_{cd,ab} = \frac{4}{F^2}\Big[\delta_{ab}\delta_{cd}(-p_1\cdot p_2 - p_3\cdot p_4) + \delta_{ac}\delta_{bd}(p_1\cdot p_3 + p_2\cdot p_4) + \delta_{ad}\delta_{bc}(p_1\cdot p_4 + p_2\cdot p_3)\Big]. \tag{7.100}$$

To the next order, namely Q^4/F^4, we must include one loop contributions in $1/F$, as well as tree level vertices of c_4, c_4'. The amplitude at this order has the structure

$$\mathcal{A}^{(2)}_{cd,ab} = \frac{1}{F^4}\left[\delta_{ab}\delta_{cd}\mathcal{M}(s,t,u) + \delta_{ac}\delta_{dd}\mathcal{M}(t,s,u) + \delta_{ad}\delta_{bc}\mathcal{M}(u,t,s)\right], \tag{7.101}$$

where

$$\begin{aligned}
\mathcal{M}(s,t,u) = 16 \int \frac{d^4k_E}{(2\pi)^4} \Bigg[&\frac{3}{2} \frac{(-p_1 \cdot p_2 - k \cdot (p_1+p_2-k))\,(-p_3 \cdot p_4 - k \cdot (p_1+p_2-k))}{k^2(p_1+p_2-k)^2} \\
&+ 2\frac{(p_1 \cdot k + p_2 \cdot (p_1+p_2-k))\,(-p_3 \cdot p_4 - k \cdot (p_1+p_2-k))}{k^2(p_1+p_2-k)^2} \\
&+ \frac{(p_1 \cdot k - p_3 \cdot (p_1-p_3-k))\,(-p_2 \cdot k + p_4 \cdot (p_1-p_3-k))}{k^2(p_1-p_3-k)^2} \\
&+ \frac{(p_1 \cdot k - p_4 \cdot (p_1-p_4-k))\,(-p_2 \cdot k + p_3 \cdot (p_1-p_4-k))}{k^2(p_1-p_4-k)^2} \Bigg] \\
&- 2c_4(p_1 \cdot p_2)(p_3 \cdot p_4) - 2c_4' \left[(p_1 \cdot p_3)(p_2 \cdot p_4) + (p_1 \cdot p_4)(p_2 \cdot p_3)\right] \\
= \frac{s^2}{2\pi^2} \ln \frac{\Lambda^2}{-s} &+ \frac{u^2 - s^2 + 3t^2}{12\pi^2} \ln \frac{\Lambda^2}{-t} + \frac{t^2 - s^2 + 3u^2}{12\pi^2} \ln \frac{\Lambda^2}{-u} - \frac{c_4}{2} s^2 - \frac{c_4'}{2}(t^2+u^2).
\end{aligned} \tag{7.102}$$

Here we have thrown away power divergences. It is understood that a regularization scheme that eliminates power divergences is used to define the perturbation theory. Now the logarithmic divergence can be absorbed into the bare couplings c_4, c_4'. We may define renormalized couplings c_{4R}, c_{4R}' at energy scale μ by

$$\begin{aligned}
c_{4R} &= c_4 - \frac{2}{3\pi^2} \ln \frac{\Lambda^2}{\mu^2}, \\
c_{4R}' &= c_4' - \frac{4}{3\pi^2} \ln \frac{\Lambda^2}{\mu^2},
\end{aligned} \tag{7.103}$$

and write

$$\mathcal{M}(s,t,u) = \frac{s^2}{2\pi^2} \ln \frac{\mu^2}{-s} + \frac{u^2 - s^2 + 3t^2}{12\pi^2} \ln \frac{\mu^2}{-t} + \frac{t^2 - s^2 + 3u^2}{12\pi^2} \ln \frac{\mu^2}{-u} - \frac{c_{4R}}{2} s^2 - \frac{c_{4R}'}{2}(t^2+u^2). \tag{7.104}$$

We see that the log dependence on s, t, u in the order Q^4 amplitude is determined by the coupling $1/F$ entirely, and does not involve the higher order couplings c_{4R}, c_{4R}'. Despite that the effective chiral Lagrangian is non-renormalizable, we see that it is predictive at low energies.

The chiral current in the effective theory is given by

$$\vec{\theta}^A \cdot \vec{J}^A = -\frac{\partial \mathcal{L}}{\partial \partial^\mu \vec{\zeta}} \cdot \delta_A \vec{\zeta}, \tag{7.105}$$

and so

$$\begin{aligned}
\vec{J}^A_\mu &= -\frac{1}{2}(1 - \vec{\zeta}^2) \frac{\partial \mathcal{L}}{\partial \partial^\mu \vec{\zeta}} - \vec{\zeta}\,\vec{\zeta} \cdot \frac{\partial \mathcal{L}}{\partial \partial^\mu \vec{\zeta}} \\
&= F^2 \left[\frac{1}{2} \partial_\mu \vec{\zeta} \frac{1 - \vec{\zeta}^2}{1 + \vec{\zeta}^2} + \frac{\vec{\zeta}(\vec{\zeta} \cdot \partial_\mu \vec{\zeta})}{1 + \vec{\zeta}^2} \right] + \cdots \\
&= \frac{F}{2} \partial_\mu \vec{\pi} + \cdots
\end{aligned} \tag{7.106}$$

Comparing this to $\langle \pi_i, p_\pi | \vec{J}^{Aj}_\mu(x)|0\rangle$, we see that F is identified with F_π defined earlier. Note that higher order contributions to $\langle \pi_i, p_\pi | \vec{J}^{Aj}_\mu(x)|0\rangle$ come with powers of p^2/F^2, and do not contribute to the one-particle pole at zero momentum.

7.4 The WZW term

In writing down the effective field theory of the Goldstone bosons, it is important to include all possible terms allowed by symmetries. In the chiral Lagrangian, while we have written all terms in the Lagrangian *density* that is invariant under $SU(N)_L \times SU(N)_R$, there is in fact another term, called the Wess-Zumino-Witten term, that leads to an invariant effective action but not Lagrangian density; namely, its variation under $SU(N)_L \times SU(N)_R$ is a spacetime total derivative.

The WZW term can be introduced as follows. Let us consider the Wick rotated Euclidean theory, and take the four dimensional Euclidean spacetime to be a compact manifold M, that is the boundary of a five-dimensional manifold B. For instance, starting with $\mathbb{R}^4$, we may assume that the fields approach some constant value at infinity, so that they can be thought of as defined on the compactified space S^4, which is topologically $\mathbb{R}^4$ plus the point at infinity. B will be taken to be the five dimensional ball in this case. The WZW term will not involve the spacetime metric, and so there is no essentially difference whether this term is written in terms of fields on S^4 or $\mathbb{R}^4$. We now extend the $SU(N)$ matrix valued field $U(x)$ to a field $U(y)$ on B, where y denotes the five dimensional coordinates on B. A priori, there could be topological obstruction to this extension. If we take M to be S^4, the field $U(x)$ on M is a map from S^4 to the group manifold $SU(N)$. The existence of the extension $U(y)$ on B is equivalent to the statement that the map $U(x)$ can be continuously deformed to the map that takes the entire S^4 to a point in $SU(N)$, i.e. the map $U(x)$ represents a trivial element of the homotopy group $\pi_4(SU(N))$. With the exception of $N = 2$ case, where $\pi_4(SU(2)) = \mathbb{Z}_2$, $\pi_4(SU(N)) = \{1\}$ for all $N \geq 3$.[6] So for $N \geq 3$, we can always extend $U(x)$ to $U(y)$ on B, while for $N = 2$, there is generally a $\mathbb{Z}_2$-valued obstruction. We will consider the $N \geq 3$ below when there is no obstruction in defining $U(y)$.

Now consider the 5-form on B,

$$\omega = -\frac{i}{240\pi^2} \mathrm{Tr}\left((U^{-1}dU)^5\right), \tag{7.107}$$

where the power of $U^{-1}dU$ inside the trace is defined as wedge product on the forms

[6] This is a special case of "Bott periodicity", which implies that $\pi_k(SU(N)) = \{1\}$ for even k, and $\pi_k(SU(N)) = \mathbb{Z}$ for odd k, when $N \geq \frac{k+1}{2}$. Explicitly, the generator of $\pi_4(SU(2)) = \pi_4(S^3)$ can be constructed either by the reduced suspension of the Hopf map $S^3 \to S^2$, or by a rational map from the quaternions (after including the point at infinity gives S^4) to $Sp(1) \simeq SU(2)$.

and matrix product as $N \times N$ matrices. Under *any* infinitesimal variation of $U(y)$, ω varies by

$$\begin{aligned} \delta\omega &= -\frac{i}{48\pi^2}\mathrm{Tr}\left((U^{-1}dU)^4 \wedge (\delta U^{-1}dU + U^{-1}d\delta U)\right) \\ &= -\frac{i}{48\pi^2} d\left[\mathrm{Tr}\left((U^{-1}dU)^4\, U^{-1}\delta U\right)\right], \end{aligned} \tag{7.108}$$

which is a total derivative. So the integral of ω on B is independent of infinitesimal variations of the extension $U(y)$, given $U(x)$ on M. The following functional of $U(y)$

$$I_{WZW} = n\int_B \omega \tag{7.109}$$

is therefore well defined as a functional of $U(x)$, up to topologically distinct extensions of U to $U(y)$ on B. Given two extensions, $U(y)$ on B, and $U(y')$ on B', $\partial B = \partial B' = M$, let us examine the difference in the WZW terms

$$I_{WZW} - I'_{WZW} = n\int_{B\bigcup\overline{B'}} \omega, \tag{7.110}$$

where $\overline{B'}$ is the same manifold as B' but with opposite orientation. Now $B\bigcup\overline{B'} = Y$ is a closed five dimensional manifold, and $\int_Y \omega$ is invariant under continuous deformations of U. In the case when B and B' are both five dimensional balls, Y is the 5-sphere, and $\int_Y \omega$ is determined by the homotopy class of the map $U : S^5 \to SU(N)$. It follows from Bott periodicity that $\pi_5(SU(N)) = \mathbb{Z}$, for $N \geq 3$. Our normalization for ω is chosen so that $\int_Y \omega$ is an integer multiple of 2π. So if we choose n to be an integer, I_{WZW} is well defined up to a shift by an integer multiple of 2π. If we add I_{WZW} to the action, in the path integral e^{iS} is then entirely well defined as a functional of U.

Our treatment of the effective field theory of Goldstone bosons thus far does not determine the integer coefficient n. n can be computed using an "anomaly matching" argument. There are a variety of ways to formulate such an argument. Here I adopt the method used in Witten's 1983 paper.

Let Q be a generic chosen $su(N)$ generator, and consider the non-chiral, diagonal $U(1) \subset SU(N)$ symmetry generated by Q, under which U transforms by

$$\delta U = i\varepsilon[Q, U]. \tag{7.111}$$

The Noether current associated with the $U(1)$ symmetry generated by Q receives the following contribution from I_{WZW},

$$J^\mu = \frac{n}{48\pi^2}\epsilon^{\mu\nu\rho\sigma}\mathrm{Tr}\left[Q(\partial_\nu U U^{-1})(\partial_\rho U U^{-1})(\partial_\sigma U U^{-1}) + Q(U^{-1}\partial_\nu U)(U^{-1}\partial_\rho U)(U^{-1}\partial_\sigma U)\right], \tag{7.112}$$

Now suppose we gauge the $U(1)$ symmetry generated by Q, by coupling this current to a $U(1)$ gauge field A_μ. This $U(1)$ is not anomalous and the resulting theory should be gauge invariant. The effective action of U coupled to A_μ takes the form

$$\Gamma[U,A] = \Gamma[U] + \int d^4x A_\mu J^\mu + \cdots \tag{7.113}$$

Here $\Gamma[U]$ is the chiral Lagrangian with $\partial_\mu U$ replaced by $\partial_\mu U - i[A_\mu, U]$. The $\cdots$ are terms of higher order in A_μ. These terms are needed to ensure the gauge invariance of $\Gamma[U,A]$, because J^μ itself is not gauge invariant. One can show that by adding a quadratic term in A_μ, the action will be gauge invariant:

$$\Gamma[U,A] = \Gamma[U] + \int d^4x A_\mu J^\mu + \frac{in}{24\pi^2}\int d^4x \epsilon^{\mu\nu\rho\sigma}(\partial_\mu A_\nu)A_\rho \mathrm{Tr}\left(Q^2\{\partial_\sigma U, U^{-1}\} - QUQ\partial_\sigma U^{-1}\right). \tag{7.114}$$

Now writing $U = e^{i\zeta_a T^a}$, we can write the last term in the above action as

$$\begin{aligned} &-\frac{n}{8\pi^2}\int d^4x \epsilon^{\mu\nu\rho\sigma}(\partial_\mu A_\nu)A_\rho \mathrm{Tr}\left(Q^2\partial_\sigma\zeta\right) \\ &= -\frac{n}{8\pi^2}\int d^4x \epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}\mathrm{Tr}\left(Q^2\zeta\right). \end{aligned} \tag{7.115}$$

This agrees with the anomaly in the axial $SU(N)$ when the quarks are coupled to the $U(1)$ gauge field, provided that we identify n with the number of *colors* N_c. Note that the anomaly is absent in the $N=2$ case, consistent with the fact that we cannot write a WZW term for $SU(2)$ chiral Lagrangian.

In the approximation of QCD with three massless flavors (u, d, s quarks), in particular, the WZW term that governs Goldstone bosons of the breaking of $SU(3)_L \times SU(3)_R$ comes with the coefficient $n = N_c = 3$. The Goldstone bosons associated with chiral symmetry breaking are parity odd, and the WZW term is the only term involving odd number of Goldstone bosons allowed by parity symmetry in the chiral Lagrangian (due to the appearance of $\epsilon^{\mu\nu\rho\sigma}$). The effect of such a term is seen in the observed decay of ϕ meson (a vector meson that is a bound state of s and $\bar{s}$),

$$\phi \to K + \overline{K}, \quad \text{and} \quad \phi \to 3\pi. \tag{7.116}$$

The K mesons, together with η and the three pions form the octet of Goldstone bosons associated with chiral symmetry breaking of $SU(3)\times SU(3)$, all of which are parity odd. Transitions like $K + \overline{K} \to 3\pi$ would be forbidden by the effective theory of Goldstone boson had the WZW term not existed.

Generally, it can also be shown that "anomalous" terms like the WZW term written here are in correspondence with elements of the de Rham cohomology group $H^5(\,\cdot\,,\mathbb{R})$

of the coset parameterized by Goldstone bosons, namely $(SU(N)\times SU(N))/SU(N)$ in the present example. It turns out that the degree 5 de Rham cohomology of this coset (or the topologically equivalent $SU(N)$) has one generator, and thus the WZW term for the Goldstone bosons associated with $SU(N)\times SU(N)$ chiral symmetry breaking is unique.

8 Anomalies

In QCD with two massless flavors, we have said that the $SU(2)_L \times SU(2)_R$ global symmetry is spontaneously broken to the diagonal $SU(2)$, while the Nambu-Goldstone boson associated with the broken chiral $SU(2)$ symmetry are the pions. The classical Lagrangian also has $U(1)\times U(1)$ global symmetry. The chiral $U(1)$ symmetry is also not observed in the baryon spectrum, and yet there appears to be no Nambu-Goldstone boson (or pseudo-NG boson in the real world) associated with the breaking of this chiral $U(1)$ symmetry. The solution to this puzzle is that the chiral $U(1)$ symmetry is in fact not a symmetry of the quantum theory.

When a classical symmetry of the Lagrangian is not a symmetry of the quantum theory, the symmetry is "anomalous". This occurs when there does not exist a UV regularization scheme that is compatible with the classical symmetry of the Lagrangian. The most familiar example is the scaling symmetry of a classical scale invariant Lagrangian, such as the Yang-Mills theory or scalar ϕ^4 theory in four dimensions. Since the couplings in these theories have nonzero beta function, the quantum theory is not scale invariant. In this case, the UV regularization, say via a momentum cutoff, clearly violates the scaling symmetry.

In gauge theories, we have to be particularly careful, because the naive momentum cutoff is in conflict with gauge invariance. Two well known regularization methods that respects gauge invariance are Pauli-Villars regularization, where one introduces fictitious regulator fields associated with each matter field, of some large mass M and eventually take M to infinity, and dimensional regularization. In the case of chiral symmetry, both of these regularization methods could fail to preserve the symmetry: chiral symmetry requires fermions to be massless, and so the massive P-V regulator fermions would violate chiral symmetry; the chiral symmetry involves γ_5, which has no natural continuation into $4-\epsilon$ dimensions, and therefore dimensional regularization a priori would fail as well. Thus, the chiral symmetry in a gauge theory is always suspicious, and could be anomalous.

To begin with, let us consider a massless Dirac fermion ψ coupled to a $U(1)$ gauge field. For now we don't need to consider the dynamics of the gauge field, and it would suffice to treat it as a background field. The Lagrangian is

$$\mathcal{L} = -\bar{\psi}\not{D}\psi = -\bar{\psi}\gamma^\mu(\partial_\mu - iA_\mu)\psi. \tag{8.1}$$

The Noether current associated with the chiral $U(1)$ symmetry is

$$J^A_\mu = i\bar{\psi}\gamma_\mu\gamma_5\psi. \tag{8.2}$$

This current is conserved classically. If the symmetry is anomalous, say in the presence of a nontrivial background field strength $F_{\mu\nu}$, the quantum expectation value $\langle \partial^\mu J^A_\mu \rangle$ would be nonzero. To see this, let us adopt Pauli-Villars regularization, and introduce a commuting (as opposed to anti-commuting) regulator field ψ', of some large mass M, that couples to A_μ exactly the same way as ψ does. The total Lagrangian is now

$$\mathcal{L} = -\bar{\psi}\not{D}\psi - \bar{\psi}'(\not{D} + M)\psi'. \tag{8.3}$$

The regularized version of the chiral $U(1)$ current is now

$$J^A_\mu = i\bar{\psi}\gamma_\mu\gamma_5\psi + i\bar{\psi}'\gamma_\mu\gamma_5\psi'. \tag{8.4}$$

Using the equation of motion, we can write the divergence of J^A_μ as

$$\partial^\mu J^A_\mu = \partial^\mu\left(i\bar{\psi}'\gamma_\mu\gamma_5\psi'\right) = 2iM\bar{\psi}'\gamma_5\psi'. \tag{8.5}$$

Now let us compute the expectation value

$$\langle \partial^\mu J^A_\mu \rangle = 2iM\left\langle \bar{\psi}'\gamma_5\psi' \right\rangle \tag{8.6}$$

In perturbation theory, the first diagram that contributes to this 1-point function is the triangle diagram with ψ' in the loop and two external gauge field lines. Suppose that the background gauge field in momentum space is given by $A_\mu(p)$. We will consider the two external gauge field lines to have momentum p_1 and p_2 flowing in, with polarization vectors $\varepsilon_1^\mu = A^\mu(p_1)$ and $\varepsilon_2^\mu = A^\mu(p_2)$. There are two such triangle diagrams, related by exchanging $1 \leftrightarrow 2$. One of them is computed as (note that there is no minus sign from the fermion loop because ψ' is a commuting field)

$$2M\int \frac{d^4k}{(2\pi)^4}\frac{\mathrm{Tr}\left[\gamma_5(-i(\not{k}+\not{p}_1+\not{p}_2)+M)\not{\varepsilon}_2(-i(\not{k}+\not{p}_1)+M)\not{\varepsilon}_1(-i\not{k}+M)\right]}{((k+p_1+p_2)^2+M^2)((k+p_1)^2+M^2)(k^2+M^2)} \tag{8.7}$$

The trace in the numerator receives nonzero contribution only from γ_5 times four gamma matrices. In particular, we will have 2 powers of momentum in the numerator. (8.7) can be evaluated as

$$\begin{aligned}
&-2M^2\int \frac{d^4k}{(2\pi)^4}\frac{\mathrm{Tr}\left[\gamma_5\not{\varepsilon}_2(\not{k}+\not{p}_1)\not{\varepsilon}_1\not{k}\right] + \mathrm{Tr}\left[\gamma_5(\not{k}+\not{p}_1+\not{p}_2)\not{\varepsilon}_2\not{\varepsilon}_1\not{k}\right] + \mathrm{Tr}\left[\gamma_5(\not{k}+\not{p}_1+\not{p}_2)\not{\varepsilon}_2(\not{k}+\not{p}_1)\not{\varepsilon}_1\right]}{((k+p_1+p_2)^2+M^2)((k+p_1)^2+M^2)(k^2+M^2)}\\
&= -8iM^2\epsilon_{\mu\nu\rho\sigma}\int \frac{d^4k}{(2\pi)^4}\frac{\varepsilon_1^\mu k^\nu \varepsilon_2^\rho p_1^\sigma - \varepsilon_1^\mu k^\nu \varepsilon_2^\rho (p_1+p_2)^\sigma + \varepsilon_1^\mu (k+p_1)^\nu \varepsilon_2^\rho p_2^\sigma}{((k+p_1+p_2)^2+M^2)((k+p_1)^2+M^2)(k^2+M^2)}\\
&= -8iM^2\epsilon_{\mu\nu\rho\sigma}\varepsilon_1^\mu p_1^\nu \varepsilon_2^\rho p_2^\sigma\int \frac{d^4k}{(2\pi)^4}\frac{1}{((k+p_1+p_2)^2+M^2)((k+p_1)^2+M^2)(k^2+M^2)}.
\end{aligned} \tag{8.8}$$

We are interested in the $M \to \infty$ limit. We see that a finite contribution survives in this limit. It is given by

$$
\begin{aligned}
&-8iM^2\epsilon_{\mu\nu\rho\sigma}\varepsilon_1^\mu p_1^\nu \varepsilon_2^\rho p_2^\sigma \int \frac{d^4k}{(2\pi)^4}\frac{1}{(k^2+M^2)^3} = 8\epsilon_{\mu\nu\rho\sigma}\varepsilon_1^\mu p_1^\nu \varepsilon_2^\rho p_2^\sigma \int \frac{d^4k_E}{(2\pi)^4}\frac{M^2}{(k_E^2+M^2)^3} \\
&= \frac{1}{4\pi^2}\epsilon_{\mu\nu\rho\sigma}\varepsilon_1^\mu p_1^\nu \varepsilon_2^\rho p_2^\sigma .
\end{aligned}
\tag{8.9}
$$

Adding the (identical) contribution from the diagram with $1, 2$ exchanged, and expressing the answer in terms of the gauge field strength, we find

$$
\langle \partial^\mu J_\mu^A \rangle = \frac{1}{8\pi^2}\epsilon_{\mu\nu\rho\sigma}F^{\mu\nu}F^{\rho\sigma}. \tag{8.10}
$$

This is the chiral anomaly due to a massless Dirac fermion charged under a $U(1)$ (the electric charged has been normalized to 1, otherwise we have a factor e^2 on the RHS). The result is easily generalized to the non-Abelian case, where the axial current $J_{a\mu}^A$ associated with a symmetry generator t^a is given by

$$
\langle \partial^\mu J_{a\mu}^A \rangle = \frac{1}{16\pi^2}\mathrm{Tr}(t^a\{T^b, T^c\})\epsilon_{\mu\nu\rho\sigma}F_b^{\mu\nu}F_c^{\rho\sigma}, \tag{8.11}
$$

where T^b, T^c are the gauge generators, and the trace is taken over all fermions. So far we are consider Dirac fermions coupled to the gauge fields, and the anomaly is in the axial flavor symmetry, rather than gauge symmetry. The above trace is therefore proportional to $\mathrm{Tr}t^a$, and thus only the $U(1)$ part of the axial flavor symmetry can be anomalous.

Even though the above computation is based on the 1-loop triangle diagram only, the result is in fact exact. It is possible to prove this directly in perturbation theory. We will now, however, give an alternative derivation of the same result, which illustrates a universal feature of anomalies.

From the perspective of the path integral, the anomaly is reflected in the non-invariance of the path integral measure under the chiral symmetry transformation

$$
\psi \to e^{i\alpha\gamma_5}\psi. \tag{8.12}
$$

The path integral measure is written as

$$
\int DA_\mu D\psi D\bar{\psi}\cdots \tag{8.13}
$$

As before we will focus on the fermion integration measure, in a fixed background of A_μ, since A_μ is not affected by the chiral symmetry transformation. In order to deal with the fermion functional integral concretely, we will expand the Dirac field $\psi(x)$

on an orthonormal basis $\varphi_n(x)$, that consists of eigenfunctions of the kinetic operator (including the coupling to the classical background A_μ). φ_n obey

$$\gamma^\mu D_\mu \varphi_n(x) = \lambda_n \varphi_n(x), \qquad \bar{\varphi}_n(x)\gamma^\mu \overleftarrow{D}_\mu = -\lambda_n \bar{\varphi}_n(x). \tag{8.14}$$

They can be normalized according to

$$\int d^4x\, \bar{\varphi}_m(x)\varphi_n(x) = \delta_{mn}. \tag{8.15}$$

Strictly speaking, we need to put the spacetime in a box in order for the modes φ_n to be discrete; in infinite spacetime, φ_n's would be delta-function normalizable instead. This difference will not be important for us, and we will proceed by treating φ_n's as a discrete set of eigenfunctions.

Now expanding

$$\psi(x) = \sum_n a_n \varphi_n(x), \qquad \bar{\psi}(x) = \sum_n \bar{a}_n \bar{\varphi}_n(x), \tag{8.16}$$

the fermion functional integral measure can be written as

$$D\psi D\bar{\psi} = \prod_n da_n d\bar{a}_n. \tag{8.17}$$

Note that here $\varphi_n(x)$ is an ordinary function of x, while a_n are Grassmannian variables. The infinitesimal chiral symmetry rotation $\delta\psi = i\alpha\gamma_5\psi, \delta\bar{\psi} = i\alpha\bar{\psi}\gamma_5$ acts on the coefficients a_n by

$$\begin{aligned} \delta a_m &= i\alpha\, C_{mn} a_n, \qquad C_{mn} = \int d^4x\, \bar{\varphi}_m(x)\gamma_5\varphi_n(x), \\ \delta\bar{a}_m &= i\alpha\, \bar{a}_n C_{nm}. \end{aligned} \tag{8.18}$$

The functional integral measure varies by the Jacobian factor

$$(\det(\mathbb{I} + i\alpha C))^{-2} = \exp(-2i\alpha \mathrm{Tr} C), \tag{8.19}$$

where we have ignored higher order terms in α. The minus sign in the exponent -2 is due to the Grassmannian nature of da_n and $d\bar{a}_n$.

Note a particularly confusing point regarding the Euclidean version of the story. In the Euclidean signature, the Dirac operator $i\gamma^\mu D_\mu$ is Hermitian, and we would like to consider eigenfunctions $\varphi_n(x)$ that obey

$$i\gamma^\mu D_\mu \varphi_n(x) = \lambda_n \varphi_n(x), \qquad \varphi_n^\dagger i\gamma^\mu \overleftarrow{D}_\mu = \lambda_n \varphi_n^\dagger(x). \tag{8.20}$$

They are normalized via the standard inner product

$$\int d^4x\varphi_m^\dagger(x)\varphi_n(x) = \delta_{mn}. \tag{8.21}$$

In Euclidean signature, we must treat ψ and $\bar{\psi}$ as independent fields. In particular, the chiral symmetry transformation acts by

$$\delta\psi = i\alpha\gamma_5\psi, \qquad \delta\bar{\psi} = -i\alpha\bar{\psi}\gamma_5. \tag{8.22}$$

The reason for this is that while the complex conjugation takes a chiral spinor to an anti-chiral one in Lorentzian signature, the complex conjugation preserves chirality in Euclidean signature. With the transformation rule (8.22), and $C_{mn} = \int d^4x\varphi_m^\dagger\gamma_5\varphi_n$, we again recover the transformation

$$\delta a_m = i\alpha\, C_{mn}a_n, \qquad \delta\bar{a}_m = i\alpha\,\bar{a}_nC_{nm}. \tag{8.23}$$

The Jacobian factor of the functional measure takes the same form as in the Lorentzian case.

If we are to compute the divergence of the Noether current associated with the chiral symmetry, all we need to do is to promote α to a spacetime dependent function $\alpha(x)$, and the Jacobian factor becomes

$$\exp\left[-2i\int d^4x\,\alpha(x)\sum_n\bar{\varphi}_n(x)\gamma_5\varphi_n(x)\right]. \tag{8.24}$$

All we have to do now is to compute the quantity

$$\sum_n\bar{\varphi}_n(x)\gamma_5\varphi_n(x). \tag{8.25}$$

You may think that this is obviously zero due to the fact that γ_5 is traceless. But we can only evaluate it after having regularized the sum over n. A natural regularization is

$$\sum_n\bar{\varphi}_n(x)\gamma_5\varphi_n(x) \to \lim_{M\to\infty}\sum_n\bar{\varphi}_n(x)\gamma_5\varphi_n(x)e^{\lambda_n^2/M^2}. \tag{8.26}$$

The reason for the choice of + sign in the exponent is so that after continuation to Euclidean signature, λ_n becomes purely imaginary, and the sum will converge.

We can write the regularized sum in (8.26) as

$$\langle x|\mathrm{Tr}\gamma_5e^{(\gamma^\mu D_\mu)^2/M^2}|x\rangle. \tag{8.27}$$

Now, using

$$(\gamma^\mu D_\mu)^2 = D^\mu D_\mu + \frac{1}{2}\gamma^{\mu\nu}[D_\mu, D_\nu] = D^2 - \frac{i}{2}\gamma^{\mu\nu}F_{\mu\nu}, \tag{8.28}$$

(8.27) turns into

$$\langle x|\mathrm{Tr}\gamma_5 e^{(D^2-\frac{i}{2}\gamma^{\mu\nu}F_{\mu\nu})/M^2}|x\rangle = \langle x|\mathrm{Tr}\gamma_5 e^{(\Box-\frac{i}{2}\gamma^{\mu\nu}F_{\mu\nu}+\cdots)/M^2}|x\rangle. \tag{8.29}$$

Here we have expanded $D^2 = D^\mu D_\mu$ in powers of $F_{\mu\nu}$, with the leading term being simply $\Box$, while $\cdots$ contains higher order terms in F. Note that

$$\langle x|e^{\Box/M^2}|x\rangle = \int \frac{d^4k}{(2\pi)^4}e^{-k^2/M^2} = i\int \frac{d^4k_E}{(2\pi)^4}e^{-k_E^2/M^2} = \frac{iM^4}{16\pi^2}. \tag{8.30}$$

The only nontrivial contribution to (8.29) of order M^0 in the $M \to \infty$ limit is

$$\begin{aligned} &\lim_{M\to\infty}\langle x|\mathrm{Tr}\gamma_5 e^{\Box/M^2}\frac{1}{2}\left(-\frac{i}{2}\gamma^{\mu\nu}F_{\mu\nu}\right)^2|x\rangle \\ &= -\frac{1}{32\pi^2}\cdot\frac{1}{4}\mathrm{Tr}\gamma_5\gamma^{\mu\nu}\gamma^{\rho\sigma}F_{\mu\nu}(x)F_{\rho\sigma}(x) \\ &= -\frac{1}{32\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}. \end{aligned} \tag{8.31}$$

This means that under the chiral rotation $\alpha(x)$, the path integral changes by a phase factor

$$\exp\left[i\int d^4x\alpha(x)\frac{1}{16\pi^2}\epsilon^{\mu\nu\rho\sigma}F_{\mu\nu}F_{\rho\sigma}\right] \tag{8.32}$$

This reproduces the divergence of the chiral current obtained earlier (note that a factor 2 difference goes into the normalization convention for the Noether current.)

Yet another way to under the anomaly is the following. Let us separate the functional integral over ψ and $\bar{\psi}$ into the integration over left-handed and right-handed fermion modes,

$$D\psi_L D\bar{\psi}_L D\psi_R D\bar{\psi}_R, \tag{8.33}$$

and let us decompose $\psi_L(x)$ and $\psi_R(x)$ in terms of a basis of eigenfunctions of the Dirac operator $\gamma^\mu D_\mu$. Of course, since γ^μ takes a left-handed spinor to a right-handed one and vice versa, all we could do is to find a basis $\varphi_n^L(x)$ for $\psi_L(x)$, and $\varphi_n^R(x)$ for $\psi_R(x)$, with

$$\begin{aligned} \gamma^\mu D_\mu \varphi_n^L(x) &= c_n \varphi_n^R(x), \\ \gamma^\mu D_\mu \varphi_n^R(x) &= d_n \varphi_n^L(x), \end{aligned} \tag{8.34}$$

where c_n, d_n are constants (only the product $c_n d_n$ is independent of the normalization of ψ_n^L versus ψ_n^R.) Provided that c_n, d_n are nonzero, $\varphi_n^L(x)$ and $\varphi_n^R(x)$ always come in pairs. The Grassmannian integral over the corresponding left-handed and right-handed fermion modes is invariant under chiral symmetry $\psi \to e^{i\alpha\gamma_5}\psi$.

An exception to this is when c_n or d_n vanishes. In this case, φ_n^L and φ_n^R need not come in pairs. We refer to such solutions as *zero modes* of the Dirac operator. In the presence of some nontrivial background gauge field $A_\mu(x)$, the number of left-handed zero modes and the number of right-handed zero modes of the Dirac operator may not be equal. Their difference is called the *index* of the Dirac operator. When the index is nonzero, the Grassmannian integral over the zero modes is not invariant under the chiral symmetry, and the symmetry is anomalous. Our derivation of the chiral anomaly then gives a formula for the index,

$$\mathrm{index} \not{D} = \int d^4x \frac{1}{16\pi^2} \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}. \tag{8.35}$$

In the case of Dirac fermions transforming in some representation R of a non-Abelian gauge group, the index is

$$\mathrm{index} \not{D} = \int d^4x \frac{1}{16\pi^2} \mathrm{Tr}_R(t^a t^b) \epsilon^{\mu\nu\rho\sigma} F_{a\mu\nu} F_{b\rho\sigma}. \tag{8.36}$$

You might be surprised by such a formula, where the LHS is by definition an integer and the RHS is not obviously an integer (but it is). This is an example of an *index theorem*. We will see later that the RHS is proportional to the *instanton number* of the gauge field configuration.

We see here that the anomaly can be understood either in terms of the failure of UV regularization in preserving the classical symmetry, or as the violation of the classical symmetry by the path integral over the zero modes, which has to do with the IR physics.

9 Electroweak theory and the standard model

9.1 Spontaneously broken gauge symmetry

We have seen that the spontaneous breaking of global symmetry G to a subgroup H leads to massless Nambu-Goldstone bosons, whose low energy dynamics is described by the nonlinear sigma model on G/H. In particular, suppose a set of scalar fields $\phi_n(x)$ transform under the symmetry G by

$$\delta\phi_n(x) = i\epsilon^a \sum_m (t_a)_{nm} \phi_m(x), \tag{9.1}$$

and that they acquire vacuum expectation value

$$\langle \phi_n(x) \rangle = v_n, \tag{9.2}$$

then we may separate the Goldstone bosons from the non-Goldstone fields by writing

$$\begin{aligned} \phi_n(x) &= \sum_m \gamma_{nm}(x)\widetilde{\phi}_n(x), \\ \sum_{n,m} \widetilde{\phi}_n(x)(t_a)_{nm}v_m &= 0, \end{aligned} \tag{9.3}$$

where $\widetilde{\phi}_n(x)$ are now orthogonal to the Goldstone directions in field space. The unbroken group H is generated by $\widetilde{\epsilon}^a t_a$ that leave v_n invariant, and $\widetilde{\phi}_n(x)$ are defined up to transformation by $\widetilde{\epsilon}^a t_a$. This is seen as follows. If we work with real fields $\phi_n(x)$, then $(t_a)_{nm}$ are purely imaginary skew symmetric matrices. Under the variation of $\widetilde{\phi}_n$ by $\widetilde{\epsilon}^a t_a$, we have

$$\begin{aligned} \delta\left[\widetilde{\phi}_n(x)(t_a)_{nm}v_m\right] &= -i\widetilde{\epsilon}^b\widetilde{\phi}_\ell(x)(t_bt_a)_{\ell m}v_m \\ &= -i\widetilde{\phi}_\ell(x)(t_a)_{\ell n}\left[\widetilde{\epsilon}^b(t_b)_{nm}v_m\right] - f_{abc}\widetilde{\epsilon}^b\left[\widetilde{\phi}_\ell(x)(t_c)_{\ell m}v_m\right] = 0, \end{aligned} \tag{9.4}$$

and hence the condition that $\widetilde{\phi}_n$ are orthogonal to Goldstone modes is preserved. $\gamma_{nm}(x) \in G$ are thus defined up to multiplication by elements of H on the right, and are a set of fields that annihilate and create the Goldstone bosons.

Now suppose G is a *gauge* symmetry instead. In the weak gauge coupling limit, the gauge fields $A_{a\mu}$ decouple and the spontaneous breaking of G leads to Goldstone bosons. On the other hand, when the gauge coupling is nonzero, the gauge symmetry G is merely a redundancy in our description of the physics. What happens to the Goldstone bosons? The transformation from $\widetilde{\phi}(x)$ to $\phi(x)$ in (9.3) is now a symmetry for any function $\gamma_{nm}(x)$, and hence $\gamma_{nm}(x)$ drops out of the Lagrangian. There are no massless Goldstone bosons. It may seem puzzling that the number of propagating fields change discontinuously as the gauge coupling goes from zero to nonzero. A closer examination of the coupling of $\phi_n(x)$ to the gauge field reveals the answer. The Lagrangian takes the form

$$\mathcal{L} = -\frac{1}{2}\left(\partial_\mu\phi_n(x) - iA_{a\mu}t^a_{nm}\phi_m(x)\right)^2 + \cdots, \tag{9.5}$$

Still assuming the spontaneous breaking of G by the vacuum expectation values $\langle\phi_n(x)\rangle = v_n$, let us write

$$\phi_n(x) = v_n + \phi'_n(x). \tag{9.6}$$

Further, we can make a gauge transformation with $\gamma(x)$ so that $\phi_n(x)$ is orthogonal to $t^a_{nm}v_m$ for all a, and since $t^a_{nm}v_nv_m = 0$, we have $\phi'_n(x)t^a_{nm}v_m = 0$. This is a gauge condition, called the *unitarity gauge*. In this gauge, we can write the quadratic part of the Lagrangian in ϕ'_n and $A_{a\mu}$ that come from the minimal coupling of ϕ_n as

$$\begin{aligned}\mathcal{L} &= -\frac{1}{2}\left(\partial_\mu \phi'_n(x) - iA_{a\mu}t^a_{nm}v_m(x)\right)^2 + \cdots , \\ &= -\frac{1}{2}(\partial_\mu \phi'_n(x))^2 - \frac{1}{2}m^2_{ab}A_{a\mu}A_b{}^\nu + \cdots ,\end{aligned} \tag{9.7}$$

where m_{ab} is now a mass matrix for the vector gauge fields $A_{a\mu}$,

$$m^2_{ab} = -t^a_{nm}t^b_{n\ell}v_m v_\ell . \tag{9.8}$$

Each nonzero eigenvalue of m_{ab} is associated with a broken symmetry in G, parameterized by $t^a_{nm}v_m$. We see that the spontaneous breaking of *gauge* symmetry makes the corresponding vector field massive. The gauge fields associated with the unbroken part of the gauge group, H, generated by those $\widetilde{\epsilon}^a t_a$ that annihilate v_n, remains massless.

Let us examine the kinetic term of a single massive vector field A_μ,

$$-\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2}m^2 A_\mu A^\mu . \tag{9.9}$$

The equation of motion

$$\partial_\nu F^{\mu\nu} + m^2 A^\mu = 0 \tag{9.10}$$

leads to

$$\partial_\mu\partial_\nu F^{\mu\nu} + m^2\partial_\mu A^\mu = m^2\partial_\mu A^\mu = 0. \tag{9.11}$$

The constraint $\partial_\mu A^\mu = 0$ is formally identical to Lorentz gauge condition but is now a consequence of the equation of motion. Now that there is no gauge invariance, the massive vector field has three propagating degrees of freedom. The propagator for the massive gauge field A_μ in momentum space is

$$\frac{-i}{p^2 + m^2 - i\epsilon}\left(\eta_{\mu\nu} + \frac{p_\mu p_\nu}{m^2}\right). \tag{9.12}$$

Naively such a propagator spoils renormalizability by naive power counting. When the massive vector field is due to spontaneous symmetry breaking, we will see that it is possible to choose a gauge in which the vector field propagator goes like p^{-2} at large momenta and the power-counting test of renormalizability is obvious.

At weak coupling, what happened is that the would-be Goldstone boson degree of freedom is absorbed by the massive vector gauge field, and there is no physical discontinuity when the gauge coupling goes to zero. This can be seen very clearly by considering the amplitude due to exchange of a massive vector boson

$$-\langle\beta|J^\mu(p)|\alpha\rangle\langle\delta|J^\nu(-p)|\gamma\rangle\frac{-i}{p^2 + m^2 - i\epsilon}\left(\eta_{\mu\nu} + \frac{p_\mu p_\nu}{m^2}\right) \tag{9.13}$$

where J_μ is the current to which the vector boson couples. In our convention both the current J_μ and the mass m are proportional to the gauge coupling g, and in the $g \to 0$ limit, the above amplitude becomes

$$-\frac{p_\mu}{m}\langle\beta|J^\mu(p)|\alpha\rangle\,\frac{p_\nu}{m}\langle\delta|J^\nu(-p)|\gamma\rangle\frac{-i}{p^2-i\epsilon}. \tag{9.14}$$

Let us compare this to the matrix element of the current J^μ associated with a spontaneously broken global symmetry, in the zero momentum limit,

$$\langle\beta|J^\mu(p)|\alpha\rangle \to \frac{iFp^\mu}{p^2}M_{\beta,B|\alpha}, \tag{9.15}$$

where $M_{\beta,B|\alpha}$ is the transition amplitude from α to β with the emission of an extra Goldstone boson of momentum p^μ. The amplitude due to the exchange of a Goldstone boson is

$$M_{\beta,B(p)|\alpha}M_{\delta,B(-p)|\gamma}\frac{-i}{p^2-i\epsilon}. \tag{9.16}$$

In the Higgs mechanism, F is precisely the same as the vector boson mass m due to coupling to the would-be Goldstone mode, and (9.16) is precisely the $p \to 0$ limit of (9.14).

Now let us consider a gauge condition that makes renormalizability clear. Let us again write

$$\phi_n(x) = v_n + \phi'_n(x) \tag{9.17}$$

but without imposing unitary gauge condition. We instead consider a gauge fixing term

$$S_{GF} = -\frac{1}{2\xi}\int d^4x f_a f_a, \tag{9.18}$$

with the choice

$$f_a = \partial_\mu A^\mu_a - i\xi t^a_{nm}\phi'_n v_m. \tag{9.19}$$

This is so that the quadratic part of the action in ϕ' and A contains

$$\begin{aligned} S &= \int d^4x\left[-\frac{1}{2}\left(\partial_\mu\phi'_n(x) - iA_{a\mu}t^a_{nm}v_m(x)\right)^2 - \frac{1}{2\xi}(\partial_\mu A^\mu_a - i\xi t^a_{nm}\phi'_n v_m)^2 + \cdots\right] \\ &= \int d^4x\left[-\frac{1}{2}(\partial_\mu\phi'_n)^2 - \frac{1}{2}m^2_{ab}A_{a\mu}A_b{}^\mu - \frac{1}{2\xi}(\partial_\mu A^\mu_a)^2 + \frac{\xi}{2}(t^a_{nm}\phi'_n v_m)^2 + \cdots\right] \end{aligned} \tag{9.20}$$

and the mix terms between ϕ' and A cancel after integration by part. We see that a mass term is now generated for the would-be Goldstone modes of ϕ'_n. The mass matrix M^2_{nm} for ϕ'_n is shifted by

$$\Delta M^2_{nm} = -\xi\sum_a(t^a v)_n(t^a v)_m. \tag{9.21}$$

Suppose c^a is an eigenvector of m^2_{ab}, namely

$$m^2_{ab}c^b = -(t^a v)_n (t^b v)_n c^b = \mu^2 c^a, \tag{9.22}$$

then we note that

$$\Delta M^2_{nm}(c^a t^a v)_m = -\xi (t^b v)_n (t^b v)_m (t^a v)_m c^a = \xi \mu^2 (c^b t^b v)_n. \tag{9.23}$$

So the nonzero eigenvalues of ΔM^2_{nm} are precisely ξ times those of m^2_{ab}, and are in 1-1 correspondence with the broken gauge symmetries. Also, with the addition of the gauge fixing term, the transverse components of A_μ have mass matrix m^2_{ab}, whereas the kinetic term for the longitudinal components of A_μ is modified by ξ. The propagator for $A_{a\mu}$ is now

$$\begin{aligned}
&-i\left[(p^2+m^2)\eta_{\mu\nu} - p_\mu p_\nu\left(1-\frac{1}{\xi}\right)\right]^{-1}_{ab} \\
&= \left[\frac{-i}{p^2+m^2}\left(\eta_{\mu\nu} + \frac{p_\mu p_\nu}{m^2}\right) - \frac{-i}{p^2+\xi m^2}\frac{p_\mu p_\nu}{m^2}\right]_{ab} = \left[\frac{-i}{p^2+m^2}\left(\eta_{\mu\nu} - \frac{(1-\xi)p_\mu p_\nu}{p^2+\xi m^2}\right)\right]_{ab}.
\end{aligned} \tag{9.24}$$

We see that it decays like p^{-2} at large momentum, as claimed.

From the gauge variation of f_a,

$$\delta f_a = \Box \epsilon_a + f_{abc}\partial_\mu (A^\mu_b \epsilon_c) + \xi (t_a v)_n (\epsilon_b t_b \phi)_n, \tag{9.25}$$

we obtain the ghost Lagrangian

$$\mathcal{L}_{gh} = \bar\eta_a \big[\Box \eta_a + f_{abc}\partial_\mu (A^\mu_b \eta_c) + \xi (t_a v)_n (\eta_b t_b \phi)_n\big]. \tag{9.26}$$

In particular, the ghost kinetic term is

$$\bar\eta_a \left[\Box \eta_a + \xi (t_a v)_n (t_b v)_n\right] \eta_b = \bar\eta_a \left[\Box \eta_a - \xi m^2_{ab}\eta_b\right]. \tag{9.27}$$

So the ghosts acquire the mass matrix ξm^2_{ab} as well in the renormalizable ξ-gauge. In particular, the ghost mass matrix have the same set of eigenvalues as that of the would-be Goldstone modes of ϕ'_n. This allows for the diagrams with ghost exchange to cancel the unphysical, ξ-dependent, massive poles of the would-be Goldstone modes. Computation of Feynman diagrams are easiest in the $\xi = 1$ gauge, where the vector boson propagator becomes very simple. The unitarity gauge on the other hand is recovered in the $\xi \to \infty$ limit, where the ghosts become infinitely massive and decouple, and unitarity is manifest.

9.2 Electroweak theory and the standard model

Let us begin with the theory of leptons. Denote by e the electron field. The left and right handed parts are

$$e_L = \frac{1+\gamma_5}{2} e, \qquad e_R = \frac{1-\gamma_5}{2} e. \tag{9.28}$$

In the standard model, there is a purely left handed electron-neutrino ν_{eL}.

$$L = \begin{pmatrix} \nu_{eL} \\ e_L \end{pmatrix}$$

forms a doublet under an $SU(2)$ gauge symmetry. The corresponding $SU(2)$ gauge fields are denoted by A^i_μ, where $i = 1, 2, 3$ labels the gauge group generator. There is an additional $U(1)$ gauge field B_μ that couples to (ν_{eL}, e_L, e_R), such that the electromagnetic $U(1)$ gauge field is a linear combination of B_μ with A^3_μ. The Lagrangian of the $SU(2) \times U(1)$ gauge fields and the electronic leptons takes the form

$$\begin{aligned} \mathcal{L} = &-\frac{1}{4}(\partial_\mu A^i_\nu - \partial_\nu A^i_\mu + g\epsilon_{ijk}A^j_\mu A^k_\nu)^2 - \frac{1}{4}(\partial_\mu B_\nu - \partial_\nu B_\mu)^2 \\ &- \bar{L}\gamma^\mu \frac{1+\gamma_5}{2}(\partial_\mu - igA^i_\mu t^i_L - ig' y_L B_\mu)L - \bar{e}_R\gamma^\mu \frac{1-\gamma_5}{2}(\partial_\mu - ig' y_R B_\mu)e_R \end{aligned} \tag{9.29}$$

In above we wrote explicitly the chiral and anti-chiral projection operator to emphasize the left and right handedness of L and e_R respectively. These projections will be omitted from now. $t^i_L = \frac{1}{2}\sigma^i$ are the $SU(2)$ generators. The $U(1)$ B_μ charge assignments are $y_L = -\frac{1}{2}$, $y_R = -1$. The $SU(2) \times U(1)$ will be spontaneously broken to the electromagnetic $U(1)$, A_μ, which is coupled to the current

$$\bar{L}\gamma_\mu(t^3_L - \frac{1}{2})L + \bar{e}_R\gamma_\mu(-1)e_R. \tag{9.30}$$

Namely, it assigns charge 0 to ν_{eL} and -1 to e_L and e_R. The electromagnetic $U(1)$ charge Q_{em} is embedded in the $SU(2) \times U(1)$ as $Q_{em} = t^3 + Y$, where t^3 is the third generator of the $su(2)$ and Y is the hypercharge. Of course, the choice of t^3 as opposed to other $su(2)$ generators is a matter of convention; it is related to the choice of Higgs vev, as we will discuss below.

The simplest mechanism for the spontaneous breaking of $SU(2) \times U(1)$ is to introduce a pair of scalar fields

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}$$

namely the Higgs field, that transform in a doublet of the $SU(2)$. The vacuum expectation value of ϕ will also give mass to the electron via the Yukawa coupling

$$\mathcal{L}_Y = -\lambda_e \bar{L}\phi e_R + c.c. \tag{9.31}$$

ϕ is assigned $U(1)$ charge $y_\phi = \frac{1}{2}$ to make the Yukawa coupling gauge invariant. Consequently, the two components of ϕ, ϕ^+ and ϕ^0, have charge $+1$ and 0 under the electromagnetic $U(1)$, as the notation suggested.

The most general renormalizable Lagrangian that couples ϕ to the $SU(2) \times U(1)$ gauge fields is

$$\mathcal{L}_\phi = -\frac{1}{2}\left|(\partial_\mu - igA^i_\mu t_i - ig' B_\mu y_\phi)\phi\right|^2 - \frac{\mu^2}{2}\phi^\dagger\phi - \frac{\lambda}{4}(\phi^\dagger\phi)^2. \tag{9.32}$$

Note that slightly unconventional normalizable of the kinetic term here. The quartic coupling λ is positive, where μ^2 is negative for spontaneous symmetry breaking to occur at tree level. The tree approximation for the vev of ϕ is

$$|\langle\phi\rangle|^2 \equiv v^2 = \frac{-\mu^2}{\lambda}. \tag{9.33}$$

Up to an $SU(2)$ rotation we may take

$$\langle\phi^+\rangle = 0, \qquad \langle\phi^0\rangle = v > 0. \tag{9.34}$$

Further, as seen before, we may work in the *unitarity gauge* in which ϕ^+ is set to zero, and ϕ^0 is real. In this gauge, the Goldstone modes of ϕ are absorbed into the longitudinal modes of the now massive gauge bosons, whereas the remaining real ϕ^0 field is a massive Higgs field.

The gauge fields now acquire a mass term

$$\begin{aligned} -\frac{1}{2}\left|(gA^i_\mu t_i + g' B_\mu y_\phi)\langle\phi\rangle\right|^2 &= -\frac{v^2}{2}\left[\left|-\frac{g}{2}(A^1_\mu + iA^2_\mu)\right|^2 + \left(-\frac{g}{2}A^3_\mu + \frac{g'}{2}B_\mu\right)^2\right] \\ &= -\frac{g^2v^2}{4}(W_\mu)^\dagger W^\mu - \frac{(g^2+g'^2)v^2}{8}Z_\mu Z^\mu. \end{aligned} \tag{9.35}$$

Now we see that the couplings g and g' can be expressed in terms of the electric charge e (not to be confused with the electron field) as

$$g = -\frac{e}{\sin\theta}, \qquad g' = -\frac{e}{\cos\theta}, \tag{9.36}$$

while the fields of $W^\pm$, Z, and the photon are the following linear combinations of the $SU(2) \times U(1)$ gauge fields,

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}}(A^1_\mu \pm iA^2_\mu), \\ Z_\mu &= \cos\theta A^3_\mu - \sin\theta B_\mu, \\ A_\mu &= \sin\theta A^3_\mu + \cos\theta B_\mu. \end{aligned} \tag{9.37}$$

We also see that the masses of $W^\pm$ and Z are given by

$$m_W = \frac{vg}{2}, \qquad m_Z = \frac{v\sqrt{g^2 + g'^2}}{2} = \frac{m_W}{\cos\theta}. \tag{9.38}$$

at tree level. The vev of ϕ also gives rise to the electron mass,

$$m_e = \lambda_e v. \tag{9.39}$$

The theory of electroweak interactions described so far easily generalizes to the case of arbitrarily many lepton generations. In the standard model, there are 3 Yukawa couplings for the three generations of leptons, λ_e, λ_μ and λ_τ.

Before the experimental discovery of W and Z bosons, the value of v was determined by the muon decay width of $\mu^+ \to e^+ + \nu_e + \overline{\nu}_\mu$ (through W exchange), to be $v \simeq$ 247GeV. Note that since this process occurs at energy much lower than electroweak scale, the effective vertex that couple the four leptons from integrating out the W-boson is proportional to $g^2/m_W^2 = \frac{4}{v^2}$. To determine the weak angle θ, or g' (knowing v and the electric charge e), one needs to study processes involving the neutral current, i.e. Z boson exchange. Purely leptonic neutral current processes such as $\nu_\mu + e \to \nu_\mu + e$ scattering are difficult to study in experiments. Historically, the weak angle θ was first determined through the study of neutral current exchange involving hadrons.

Now let us describe the quark sector of the standard model. Each generation of quark can be written in terms of the left handed fermions

$$Q_L^i = \begin{pmatrix} u_L^i \\ d_L^i \end{pmatrix} \tag{9.40}$$

and the right handed fermions

$$u_R^i, \quad d_R^i. \tag{9.41}$$

$i = 1, 2, 3$ here labels the generation. The left handed quarks couple to the electroweak $SU(2) \times U(1)$ as a doublet under the $SU(2)$ and has $U(1)$ charge $y_Q = \frac{1}{6}$, while the right handed quarks transform in the singlet of $SU(2)$ and has $U(1)$ charge $y_U = \frac{2}{3}$, $y_D = -\frac{1}{3}$. The representations of all fermions ("*QudLe*") together with the Higgs field ϕ in the standard model under the $SU(3) \times SU(2) \times U(1)$ gauge group are summarized in the following chart:

	$SU(3)$	$SU(2)$	$U(1)_Y$
Q	**3**	**2**	$\frac{1}{6}$
u	**3**	**1**	$\frac{2}{3}$
d	**3**	**1**	$-\frac{1}{3}$
L	**1**	**2**	$-\frac{1}{2}$
e	**1**	**1**	-1
ϕ	**1**	**2**	$\frac{1}{2}$

After the $SU(2) \times U(1)_Y$ is broken to the electromagnetic $U(1)$ by the Higgs vev, the charge of the fermions under the electromagnetic $U(1)$ is given by the linear combination $t^3 + Y$, where $t^3 = \frac{1}{2}\sigma^3$ is the third $SU(2)$ generator.

The quarks acquire their masses by coupling to the Higgs field and its Hermitian conjugate (which also transform in the doublet of $SU(2)$) via Yukawa type coupling that mixes generations. The complete Yukawa coupling takes the form

$$\mathcal{L}_Y = -\lambda_i(\overline{L}^i_{La}\phi^a)e^i_R - \lambda^{ij}_d(\overline{Q}^i_{La}\phi^a)d^j_R - \lambda^{ij}_u\epsilon^{ab}(\overline{Q}^i_{La}\phi^\dagger_b)u^j_R + c.c. \tag{9.42}$$

Here we exhibited the $SU(2)$ doublet index a explicitly, and used the $SU(2)$ invariant tensor ϵ^{ab} to contract $\overline{Q}^i_L$ with $\phi^\dagger$. We could have introduced lepton Yukawa coupling that mixes different generations of leptons as well, but such a Yukawa coupling matrix can be put to a diagonal form by separate unitary rotations on L^i_L and E^j_R respectively. We can also perform unitary rotation on Q_L, u_R, d_R separately to eliminate some components of the quark Yukawa matrix, but cannot diagonalize λ^{ij}_d and λ^{ij}_u simultaneously, while preserving the $SU(2)$ doublet Q_L.

Now consider the quark mass term generated from the Yukawa coupling via Higgs vev

$$\langle\phi\rangle = \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{9.43}$$

The mass terms for $u_{L,R}$ and $d_{L,R}$ are

$$-\lambda^{ij}_d v\, \overline{d}^i_L d^j_R - \lambda^{ij}_u v\, \overline{u}^i_L u^j_R + c.c. \tag{9.44}$$

So far we have chosen to work in the basis such that u^i_L and d^i_L transform in the same weak $SU(2)$ doublet, so that the expression for the weak current does not mix generations. However, the quark fields defined this way are not in an eigenbasis of the matrix matrix. We can perform unitary rotation on u^i_L and d^i_L in Q_L *separately*, and diagonalize the mass matrix. Let us consider such a change of basis

$$u'_L = A^U_L u_L, \quad u'_R = A^U_R u_R, \quad d'_L = A^D_L d_L, \quad d'_R = A^D_R d_R, \tag{9.45}$$

where $A^{U,D}_{L,R}$ are four 3×3 unitary matrices. The mass matrix in terms of the primed quark fields are

$$M'_d = A^D_L \lambda_d (A^D_R)^\dagger v, \quad M'_u = A^U_L \lambda_u (A^U_R)^\dagger v. \tag{9.46}$$

For any matrices λ_d, λ_u, we can find the A matrices so that M'_d and M'_u are diagonal. The corresponding $u'_{L,R}$ and $d'_{L,R}$ are the fields of what we know as the three generations of quarks, with definite masses. The price to pay is that the expression for the weak current now mixes different generations of the primed quark fields. For example, the hadronic $SU(2)$ current that couples to W^+ can be written as

$$\begin{aligned} J_\mu^- &= \overline{u}_L^i \gamma^\mu d_L^i \\ &= \overline{u}'^i_L \gamma^\mu \left(A_L^U (A_L^D)^{-1} \right)_{ij} d'^j_L. \end{aligned} \tag{9.47}$$

In fact, we see that the only free parameter among the quark Yukawa couplings, besides the six mass parameters in M'_d and M'_u, are contained in the 3×3 *unitary* matrix

$$V_{ij} = \left(A_L^U (A_L^D)^{-1} \right)_{ij}. \tag{9.48}$$

V is known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix. So far we are still free to rotate the six phases of $u'^i_{L,R}$, $d'^i_{L,R}$ independently, without changing the mass matrix. Such a phase rotation does change the matrix V. If there were only two generations, and V a 2×2 unitary matrix, then it is always possible to perform such a phase rotation to make V a real unitary, i.e. orthogonal, matrix, of the form

$$\begin{pmatrix} \cos\theta_c & \sin\theta_c \\ -\sin\theta_c & \cos\theta_c \end{pmatrix} \tag{9.49}$$

θ_c is known as the Cabibbo angle. Including the third generation, V can no longer be put to an orthogonal matrix, though the experimentally measured value of its components that mix the third generation with the first two generations is very small (but nonzero). The significance of a CKM matrix that cannot be set to real is that the standard model necessarily breaks CP symmetry (charge conjugation combined with parity) invariance. Under CP symmetry, $q_{L,R}$ transform into $Bq^*_{L,R}$, where the matrix B obeys $B^\dagger \beta \gamma^\mu B = (\gamma^\mu \beta)^T$, $\beta = i\gamma^0$. It exchanges for the instance the couplings to W^+ and W^-,

$$iW_\mu^+ \overline{u}'^i_L \gamma^\mu V_{ij} d'^j_L - iW_\mu^- \overline{d}'^j_L \gamma^\mu V^*_{ij} u'^i_L. \tag{9.50}$$

The Lagrangian has CP invariance if V_{ij} is real in some basis. As a $U(3)$ matrix, V has 9 parameters, which may be expressed in terms of 3 angles of $O(3)$ and 6 phases; 5 of the phases can be removed by independent phase rotations on the six flavors of quarks, leaving one independent phase. This phase gives rise to CP violation.

We have been a bit sloppy in making the above *chiral* rotation on the quark flavors, since such chiral rotations generally changes the functional integral measure of the fermions by a phase, in the presence of background gauge fields. Closely related are the theta terms for the $SU(3) \times SU(2) \times U(1)$ gauge fields,

$$\sum_{I=1}^{3} \frac{\theta_I g_I^2}{64\pi^2} \epsilon^{\mu\nu\rho\sigma} F^{a(I)}_{\mu\nu} F^{a(I)}_{\rho\sigma}, \tag{9.51}$$

where $I = 1, 2, 3$ labels the three gauge group $U(1)$, $SU(2)$, $SU(3)$ respectively; g_I the corresponding Yang-Mills coupling. While these terms can be written as total derivatives involving the *vector potential*, for nonabelian gauge group there are gauge field configurations with vanishing *field strength* at infinity (and finite action) but nonzero integral of the above "theta term". The angle θ_i is defined periodically in the functional integral, due to the quantized value of the integral of $F\tilde{F}$. If we make a chiral rotation on the fermions, and if such a chiral rotation is anomalous in the presence of $F^{(I)}_{\mu\nu}$, then the change of phase of the path integral measure shifts θ_I in the effective action. Due to the chiral coupling of $SU(2)\times U(1)$ gauge fields to the leptons, it is possible to make a chiral rotation on the leptons that leaves the lepton mass matrix invariant, while shifting away θ_1, θ_2. In other words, θ_1 and θ_2 are not observable physical parameters of the standard model. θ_3, on the other hand, cannot be rotated away by a chiral rotation on the quarks that leaves the quark mass matrix invariant, since $SU(3)$ couples non-chirally to the quarks. As a consequence, θ_3 is an independent parameter of the theory. In reality, however, in the basis of quark fields with diagonalized mass matrix, there is no evidence for a nonzero θ_3 angle, as the latter would contribute to the neutron electric dipole moment which has been excluded experimentally to high accuracy. The vanishing or very small value of θ_3 cannot be explained within the standard model; this is known as the "strong CP problem".

To summarize, let us write down the full renormalizable Lagrangian of the standard model:

$$\mathcal{L} = \mathcal{L}_{YM} + \mathcal{L}_V + \mathcal{L}_F + \mathcal{L}_Y, \tag{9.52}$$

where $\mathcal{L}_{YM}$ is the Yang-Mills Lagrangian of $SU(3) \times SU(2) \times U(1)_Y$ gauge group, possibly with theta term. We have worked with the convention in which the gauge coupling is absorbed into the structure constants. $\mathcal{L}_V$ is the Lagrangian for the Higgs field, minimally coupled to $SU(2) \times U(1)_Y$, that transforms as a doublet of the $SU(2)$ and has $U(1)_Y$ charge $-\frac{1}{2}$, and with Mexican hat scalar potential. $\mathcal{L}_F$ stands for the Lagrangian of minimally coupled left and right handed fermions Q_L, u_R, d_R, L_L, e_R, whose representation content under $SU(3) \times SU(2) \times U(1)_Y$ are described as before. $\mathcal{L}_Y$ is the Yukawa coupling of the Higgs field to the massless fermions. The independent parameters are the three gauge couplings, the Higgs mass and quartic coupling, the 3 lepton Yukawa couplings (or mass parameters), the 6 quark mass parameters, the 4 independent entries of the CKM matrix, and the QCD theta angle θ_3, making a total of 19 parameters.

9.3 Anomaly cancelation

Chiral currents are a priori anomalous when the fermions are coupled to gauge fields. When the corresponding chiral symmetry is a global symmetry, the anomaly means that the symmetry does not exist in the quantum theory. This is the case for the $U(1)$ axial symmetry of QCD with massless quarks. If the chiral symmetry is gauged, however, anomaly would indicate inconsistency of the gauge theory. The fermion spectrum of the standard model is chiral, and the cancelation of anomaly for the $SU(3)\times SU(2)\times U(1)$ gauge currents is required for the consistency of the theory. This is not at all automatic: for instance, as we will see, the anomaly would not be canceled if the number of lepton generations is not the same as the number of quark generations, or if the number of colors is not equal to 3.

The divergence of an anomalous, non-vector, current $J^\mu = -i\bar{\psi}\tau\gamma^\mu\psi$, where $\tau = \tau_L\frac{1+\gamma_5}{2} + \tau_R\frac{1-\gamma_5}{2}$, due to the coupling of fermion ψ to background gauge field strength $F^a_{\mu\nu}$ is given by

$$\partial_\mu J^\mu = -\frac{1}{32\pi^2}\mathrm{tr}(\gamma_5\tau\{t_a, t_b\})\epsilon^{\mu\nu\rho\sigma}F^a_{\mu\nu}F^b_{\rho\sigma}. \tag{9.53}$$

where t^a are the matrices for the generators of the gauge group acting on the fermions, and the trace is taken over all fermions involved in the current J^μ. Note that in our convention the gauge coupling has been included in the generators t_a (and hence structure constants). This formula can be derived by computing the triangle diagram, which gives the contribution from the part of $F^a_{\mu\nu}$ that is linear in $A_{b\rho}$, and then the higher order terms in A can be completed simply by demanding gauge invariance.

Let us note that the formula (9.53) is most straightforwardly derived for the anomaly of a non-vector type current associated with a global symmetry, due to coupling to non-anomalous gauge fields $A_{a\mu}$. It is much more subtle to compute the anomaly of a gauge current. First of all, the current itself would not be gauge invariant, and we should write the gauge covariant derivative on the LHS. Furthermore, the RHS need not be a gauge invariant expression either, when the gauge fields are anomalous. In such situations, there is in fact an ambiguity in computing the triangle diagram, which allows one to "move" anomalies from one gauge current to another. In particular, if we replace J^μ by the gauge current J^μ_a itself, and τ by t^a (which involves γ_5), there is an extra factor 1/3 multiplying the term quadratic in $\partial_\mu A_{a\nu}$ on the RHS. We will not discuss this subtle computation here (see chapter 22 of Weinberg). It is nonetheless still the case that the anomaly is proportional to

$$\begin{aligned} d_{abc} &= \mathrm{tr}\left(t^L_a\{t^L_b, t^L_c\}\right) - \mathrm{tr}\left(t^R_a\{t^R_b, t^R_c\}\right) \\ &= \mathrm{tr}\left(\gamma_5 t_a\{t_b, t_c\}\right), \end{aligned} \tag{9.54}$$

where the trace is taken over all fermions that transform under t_a.

The consistency of the standard model requires the vanishing of d_{abc}, for a, b, c running through all generators of the gauge group $SU(3)\times SU(2)\times U(1)$. Since only the quarks transform under the $SU(3)$ gauge fields, and the coupling to $SU(3)$ is non-chiral, there are no anomaly coefficients of the type $SU(3)-SU(3)-SU(3)$ (we will write it as 3-3-3 for short). The anomaly coefficients that involve only one

$SU(2)$ or $SU(3)$ generator obviously also vanish, since the generators are traceless in the fundamental representation. This leaves us with the potentially nonzero anomaly coefficients, 2-2-2, 1-3-3, 1-2-2, 1-1-1, which we now compute.

The 2-2-2 anomaly coefficient is clearly zero since $\mathrm{tr}(\tau^a\tau^b\tau^c) \propto \epsilon^{abc}$ for the $SU(2)$ generators τ^a and so $\mathrm{tr}(\tau^a\{\tau^b, \tau^c\}) = 0$. In fact, this is a property for any *real* representation of a gauge group G, since the generators for the complex conjugate representation is given by $-(t_a)^T$, and so the coefficient d_{abc} for the conjugate representation has the opposite sign, and vanishes for real representations.

1-3-3: It suffices to compute the anomaly coefficient for one generation of quarks. We have

$$\mathrm{tr}(Y\{t^a, t^b\}) = \delta^{ab} \sum_q Y_q, \tag{9.55}$$

where q labels the quarks Q_L, u_R, d_R, Y_q the corresponding $U(1)_Y$ charged counted with + sign for left handed quarks and an extra − sign for right handed quarks.

$$\sum_q Y_q = 2 \times \frac{1}{6} - \left(\frac{2}{3} - \frac{1}{3}\right) = 0. \tag{9.56}$$

1-2-2: Only $SU(2)$ doublets, i.e. left handed leptons L_L and left handed quarks Q_L contribute. For one generation of leptons and quarks, we have

$$\begin{aligned} \mathrm{tr}(Y\{\tau^a, \tau^b\}) &= \delta^{a,b} \sum Y_L \\ &= \delta^{a,b} \left(-\frac{1}{2} + 3 \times \frac{1}{6}\right) = 0. \end{aligned} \tag{9.57}$$

The factor of 3 is due to the colors of the left handed quark Q_L.

1-1-1: All leptons and quarks contribute. The anomaly is proportional to

$$\mathrm{tr}(Y^3) = 2 \times \left(-\frac{1}{2}\right)^3 - (-1)^3 + 3 \times 2 \times \left(\frac{1}{6}\right)^3 - 3 \times \left(\frac{2}{3}\right)^3 - 3 \times \left(-\frac{1}{3}\right)^3 = 0. \tag{9.58}$$

Here the trace is again understood to be taken with the opposite sign for left handed and right handed fermions. Now we see that all gauge anomalies cancel.

There is one more possible anomaly, for the $U(1)_Y$ gauge current, due to the coupling to background gravitational fields. Since this is a gauge anomaly due to gravitational field, it is often referred to as the gauge-gravitational mixed anomaly (not to be confused with the purely gravitational anomaly, which only exists in $4n + 2$ spacetime dimensions). The derivation of this anomaly by computing the triangle diagram with two stress-energy tensor insertions is left as an exercise. The resulting divergence of the $U(1)_Y$ current is proportional to

$$\mathrm{tr}(\gamma_5 Y)\epsilon^{\mu\nu\rho\sigma}R_{\mu\nu\eta\tau}R_{\rho\sigma}{}^{\eta\tau}. \tag{9.59}$$

Once again, we can verify that this anomaly in the standard model cancel as well:

$$\mathrm{tr}(Y) = 2\times\left(-\frac{1}{2}\right) - (-1) + 3\times 2\times\left(\frac{1}{6}\right) - 3\times\left(\frac{2}{3}\right) - 3\times\left(-\frac{1}{3}\right) = 0. \tag{9.60}$$

9.4 Higher dimensional operators and physics beyond the standard model

Assuming the gauge group of $SU(3)\times SU(2)\times U(1)$ and the matter field content (QUDLE plus a Higgs doublet), we have written the most general renormalizable Lagrangian for the standard model. However, the standard model Lagrangian a priori describes only a low energy effective theory, and there could be irrelevant operators in the full Lagrangian of mass dimension greater than 4, that come with coefficients suppressed by inverse powers of some large mass scale Λ (much larger than the electroweak scale).

If such higher dimensional terms do appear, their effects in physical processes occurring at energies $E\ll\Lambda$ is negligible, except when the higher dimensional operator violates symmetries of the renormalizable Lagrangian. Indeed, there are two accidental $U(1)$ global symmetries of the standard model Lagrangian, the lepton number and baryon number conservation.[7]

The first gauge and Lorentz invariant operator that violate lepton number occurs at dimension 5. Such an operator can be constructed out of the $SU(2)$ singlet formed by the left handed lepton with the Higgs doublet, $(L_i\bar\phi) = L_i^a\bar\phi_a$, which is also neutral under $U(1)_Y$. Here a stands for the $SU(2)$ doublet index, and i labels the generation. We may then write the dimension 5 operator

$$f_{ij}(L_{i\alpha}\bar\phi)(L_{j\beta}\bar\phi)\epsilon^{\alpha\beta}. \tag{9.61}$$

Here we wrote L_i in Weyl spinor notation, and exhibited explicitly the chiral spinor index $\alpha(=1,2)$. In Dirac spinor notation, the above operator is written as

$$f_{ij}(\overline{L_i^{\mathbf{c}}}\bar\phi)(L_j\bar\phi), \tag{9.62}$$

[7] The baryon and lepton $U(1)$ global symmetries are in fact anomalous due to the chiral coupling of quarks and leptons to the $SU(2)\times U(1)$ gauge fields. Amplitudes that violate baryon and lepton number conservation due to anomalies, as we will see later, are mediately by instantons. Such amplitudes are exponentially suppressed with a small electroweak coupling constant, and are unlikely to be observed in experiments.

where $L_i^{\mathbf{c}}$ is the charge conjugate field of L_i, $L_i^{\mathbf{c}} = CL_i^*$, where C obeys $C^\dagger \beta \gamma^\mu C = (\beta\gamma^\mu)^*$. In components, $L_i = (\nu_i, \ell_{Li})$, $\bar{\phi} = (\phi^0, -\phi^+)$, we can write this term as

$$f_{ij}(\overline{\nu}_i^{\mathbf{c}}\phi^0 - \overline{\ell}_{Li}^{\mathbf{c}}\phi^+)(\nu_j\phi^0 - \ell_{Lj}\phi^+) \tag{9.63}$$

At energies much below the electroweak scale, this gives rise to a Majorana mass term for the neutrinos via the Higgs vev,

$$f_{ij}v^2\overline{\nu}_i^{\mathbf{c}}\nu_j. \tag{9.64}$$

We expect f_{ij} to be suppressed by $1/\Lambda$, where Λ is a large mass scale at which the perturbative standard model Lagrangian breaks down. The neutrino mass arising this way would be of order v^2/Λ. Since $v \sim 300$GeV, if the neutrino mass is of order 0.1eV (as the typical value indicated from neutrino oscillation experiments), it would imply $\Lambda \sim 10^{15}$GeV. We will see later that, remarkably, this is also roughly the energy scale of coupling unification!

Operators that lead to baryon number violation appear at dimension 6, involving three quark fields (so as to form an $SU(3)$ color singlet) and one lepton field. For the operator to be invariant under $U(1)_Y$, the following combinations are allowed: $QQQL$, $QQue$, $duQL$, $duue$. All of these operators violate both the baryon number and lepton number, but not the difference between baryon and lepton numbers. Their coefficients are expected to be suppressed by $1/\Lambda^2$. Such an operator leads to proton decay rate of order $10^{-3}m_p^5/\Lambda^4$. The current bound on the proton lifetime $\tau > 5 \times 10^{33}$ years then says $\Lambda > 4 \times 10^{15}$GeV, which is yet again consistent with the scale of coupling unification.

Let us now turn to radiative corrections in the standard model. There are three gauge couplings, g_3, g and g' for the $SU(3)$, $SU(2)$, and $U(1)_Y$ respectively. The renormalized couplings run under RG. We are interested in the behavior of these couplings at energies far above the electroweak scale (say the scale of Z-boson mass). For this purpose, in computing the one-loop beta function, the effect of spontaneous breaking of $SU(2) \times U(1)$ may be ignored, and all fermions treated as massless fields. We have previous derived the one-loop beta function of nonabelian gauge theory with fermion matter fields. The only modification we need here is to include the contribution from a scalar matter field (the Higgs). The computation is identical to the earlier calculation of the contribution from ghost loops, except for the overall minus sign due to opposite statistics, and the group theory factor $C(adj)$ for ghosts now replaced by the corresponding factor for the scalar matter. The general one-loop beta function for a gauge theory with gauge group G, massless Dirac fermions in representation R_f and massless scalars (the renormalization scale will be assumed to be much higher than the Higgs mass) in representation R_b is

$$\beta(g) = -\frac{g^3}{4\pi^2}\left[\frac{11}{12}C(adj) - \frac{1}{3}C(R_f) - \frac{1}{12}C(R_b)\right]. \tag{9.65}$$

Recall that $C(adj) = N$ for $SU(N)$ gauge group, and $C(\mathbf{f}) = C(\bar{\mathbf{f}}) = \frac{1}{2}$. For $U(1)$ gauge group, $C(adj) = 0$, and for a representation R of charge q, $C(R) = q^2$. Note that the contribution from a left-handed (or right-handed) Weyl fermion is half that of a Dirac fermion. Suppose there are n_g generations of quarks and leptons. We have

$$\begin{aligned}
\beta_3(g_3) &= -\frac{g_3^3}{4\pi^2}\left(\frac{11}{12}\times 3 - \frac{1}{3}\times\frac{1}{2}\times 2n_g\right) = -\frac{g_3^3}{4\pi^2}\left(\frac{11}{4} - \frac{n_g}{3}\right),\\
\beta_2(g) &= -\frac{g^3}{4\pi^2}\left(\frac{11}{12}\times 2 - \frac{1}{12}\times\frac{1}{2} - \frac{1}{3}\times\frac{1}{2}\times\frac{1}{2}\times(3+1)n_g\right) = -\frac{g_2^3}{4\pi^2}\left(\frac{43}{24} - \frac{n_g}{3}\right),\\
\beta_1(g') &= -\frac{g'^3}{4\pi^2}\left[-\frac{1}{12}\times 2\times(\frac{1}{2})^2 - \frac{1}{3}\times\frac{n_g}{2}\left(3\times 2\times(\frac{1}{6})^2 + 3\times(\frac{2}{3})^2 + 3\times(-\frac{1}{3})^2 + 2\times(-\frac{1}{2})^2 + (-1)^2\right)\right]\\
&= \frac{g'^3}{4\pi^2}\left(\frac{1}{24} + \frac{5n_g}{9}\right).
\end{aligned} \tag{9.66}$$

With $n_g = 3$, the beta function for the $SU(3)$ gauge coupling g_3 and $SU(2)$ gauge coupling g are negative, while the beta functions for $U(1)_Y$ gauge coupling g' is positive. As the renormalization scale increases,the three couplings tend to converge at a "unification scale", and one may suspect that the $SU(3) \times SU(2) \times U(1)_Y$ is the low energy gauge group that is the result of spontaneous breaking of a yet larger "grand unification" gauge group.

The simplest proposal for such a grand unification theory (GUT) postulates an $SU(5)$ gauge group, that is spontaneously broken to its maximal subgroup $SU(3) \times SU(2) \times U(1)$ at a certain mass scale M_{GUT}. The standard model gauge group generators can be embedded in $su(5)$ in the form

$$\begin{pmatrix} * & * & * & 0 & 0\\ * & * & * & 0 & 0\\ * & * & * & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & * & *\\ 0 & 0 & 0 & * & * \end{pmatrix}, \qquad \sqrt{\frac{3}{5}}\begin{pmatrix} \frac{1}{3} & 0 & 0 & 0 & 0\\ 0 & \frac{1}{3} & 0 & 0 & 0\\ 0 & 0 & \frac{1}{3} & 0 & 0\\ 0 & 0 & 0 & -\frac{1}{2} & 0\\ 0 & 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix}, \tag{9.67}$$

where the 3×3 and 2×2 blocks are traceless, corresponding to generators of $su(3)$ and $su(2)$, and the last matrix is the $u(1)$ generator Y embedded in $su(5)$, with the normalization convention $\mathrm{tr}(Y^2) = \frac{1}{2}$. Due to this normalization, in the $SU(5)$ GUT scenario, we expect the standard model gauge couplings g_3, g and $\sqrt{\frac{5}{3}}g'$ to become the same, namely the $SU(5)$ gauge coupling, at mass scale M_{GUT}. Based on one-loop beta function computed in the standard model, the three couplings indeed become the same order at $M_{GUT} \sim 10^{15}$GeV though they don't meet exactly.[8]

The question then is whether the standard model matter fields can be fit into representations of $SU(5)$. In the most naive scenario, there would have to be more Higgs fields whose vevs break the $SU(5)$ to the standard model gauge group. This could be achieved if there are Higgs fields transforming in the **24** dimensional adjoint representation of $SU(5)$, whose vev is proportional to the Y generator written above. One then assumes that the electroweak Higgs doublet is part of another Higgs multiplet, transforming in the **5** of $SU(5)$. The spontaneous symmetry breaking in this scenario is such that first $SU(5) \to SU(3) \times SU(2) \times U(1)$ by the **24** Higgs with vev at scale M_{GUT}, and then $SU(2) \times U(1)$ is broken to the electromagnetic $U(1)$ by the **5** Higgs at electroweak scale.

The case for unification is much more compelling in the fermion sector. First let us note that the representations **5** and **10** of $SU(5)$ decomposes into irreducible representations of $SU(3) \times SU(2) \times U(1)$ as

$$\begin{aligned}
&\mathbf{5} \to (\mathbf{3}, \mathbf{1})_{1/3} \oplus (\mathbf{1}, \mathbf{2})_{-1/2}, \\
&\bar{\mathbf{5}} \to (\bar{\mathbf{3}}, \mathbf{1})_{-1/3} \oplus (\mathbf{1}, \mathbf{2})_{1/2}, \\
&\mathbf{10} = (\mathbf{5} \otimes \mathbf{5})_{asym} \to (\bar{\mathbf{3}}, \mathbf{1})_{2/3} \oplus (\mathbf{3}, \mathbf{2})_{-1/6} \oplus (\mathbf{1}, \mathbf{1})_{-1}, \\
&\overline{\mathbf{10}} \to (\mathbf{3}, \mathbf{1})_{-2/3} \oplus (\bar{\mathbf{3}}, \mathbf{2})_{1/6} \oplus (\mathbf{1}, \mathbf{1})_{1}.
\end{aligned} \tag{9.68}$$

We see that d_L^c (the charge conjugate of d_R) and L_L would fit nicely into a left handed fermion in the **5** of $SU(5)$, while u_L^c, Q_L, and e_L^c would fit precisely into a left handed fermion in the $\overline{\mathbf{10}}$.

The $SU(5)$ GUT, first proposed by Georgi and Glashow, is a tantalizing scenario for physics beyond the standard model. A more careful study of the low energy effective Lagrangian obtained after symmetry breaking reveals difficulties in satisfying phenomenological constraints. For instance, integrating out the color triplet Higgs in the **5** generates dimension 6 operators that violate baryon and lepton number and mediate proton decay. The coefficient of this operator is proportional to the inverse square of the triplet Higgs mass. To satisfy current experimental bounds on proton lifetime, the color triplet Higgs must have its mass at the GUT scale, while the remaining electroweak doublet Higgs has its mass at the electroweak scale. This is known as the "doublet-triplet splitting". There are other dimension 6 operators that mediates proton decay, generated by integrating out the massive vector boson corresponding to the broken part of $SU(5)$. The mass of the massive GUT vector bosons are essentially

[8] The unification of couplings is significantly improved in the supersymmetric version of the standard model, and the unification scale would be at slightly higher value, $M_{GUT} \sim 10^{16}$GeV.

determined by the scale of unification of couplings. The proton decay rate mediated by the massive GUT vector boson turns out to violate the current experimental bound by several orders of magnitude, which effectively rules out the simplest $SU(5)$ GUT scenario. There are various extended GUT theories, involving other gauge groups, e.g. $SO(10)$ and E_6. We will not discuss them here.

10 Non-perturbative aspects

In the Hamiltonian approach to quantum field theory, we have canonically quantized the elementary fields and their canonical momenta, and constructed the Hilbert space of states by applying the creation operators of fields on the vacuum. In perturbation theory, this is done by expanding the fields around an equilibrium configuration that solves the equation of motion, and quantize the fluctuations. It can happen that there are topologically distinct field configurations of finite energy that cannot be obtained from one another by continuous deformations. The states arising from such field configurations would be missed in the perturbation theory. An example is the magnetic monopole in certain gauge theories, which we will explore.

In the Lagrangian, or path integral formulation, perturbation theory is again defined by first choosing a stationary point of the action, and performing the functional integration over fluctuations around this field configuration, in Wicked rotated Euclidean spacetime. Generally, there can be (Euclidean time dependent) field configurations that locally minimizes the action, that differ from the one we expand around in the ordinary perturbation theory. For the action to be finite, such field configurations must have their field strength vanish at spatial and time infinity, and in particular localized in Euclidean time. These field configurations are called instantons. While in weakly coupled quantum field theory, the contribution due to instantons are very small compared to the perturbative configurations, there are tunelling processes that are forbidden in perturbation theory and only occur through instantons. We will also see that the presence of instantons are closely tied to the large order behavior of perturbation theory, and are often required for the consistency of the quantum field theory at the non-perturbative level.

10.1 Monopoles

The simplest kind of monopole is the Dirac monopole of a $U(1)$ gauge theory. We want to consider a solution to the equation of motion with magnetic field strength

$$F_{ij} = q_m \epsilon_{ijk} \frac{x^k}{4\pi|\vec{x}|^3}, \tag{10.1}$$

where q_m is the magnetic charge. The solution is singular at the origin $\vec{x} = 0$, and has infinite energy and hence not an acceptable field configuration in the free abelian gauge theory. However, if our $U(1)$ gauge theory is defined with a UV cutoff, with a suitable UV completion, then such a monopole could have finite energy. Note that there is no way to write the vector potential A_i (say in temporal gauge $A_0 = 0$) on the entire space away from the origin. This is not a problem, since A_μ is not gauge invariant by itself. It suffices to be able to write A_μ on simply connected patches of space(time), such that the vector potentials on different patches are related by gauge transformations. For instance, in radial coordinates (r, θ, ϕ), we can write the vector potential on the upper $(0 \le \theta \le \pi/2)$ and lower $(\pi/2 \le \theta \le \pi)$ half space respectively,

$$\vec{A}_N = \frac{q_m}{4\pi r} \frac{1 - \cos\theta}{\sin\theta} \hat{e}_\phi, \tag{10.2}$$

and

$$\vec{A}_S = -\frac{q_m}{4\pi r} \frac{1 + \cos\theta}{\sin\theta} \hat{e}_\phi. \tag{10.3}$$

Indeed, A_N and A_S differ by a gauge transformation along $\theta = \pi/2$,

$$\vec{A}_N - \vec{A}_S = \frac{q_m}{2\pi r} \frac{\hat{e}_\phi}{\sin\theta} = \nabla \left(\frac{q_m}{2\pi} \phi \right). \tag{10.4}$$

One may worry that this is not a well defined gauge transformation as the angular coordinate ϕ is not single valued. The finite form of the gauge transformation is

$$A_\mu \to A_\mu + ig\partial_\mu g^{-1}, \tag{10.5}$$

and if there are matter fields, say ψ with one unit of electric charge, then ψ transforms as

$$\psi(x) \to g(x)\psi(x). \tag{10.6}$$

The gauge function that relates the monopole gauge fields on the upper and lower half space is

$$g(x) = \exp\left(i \frac{q_m}{2\pi} \phi \right) \tag{10.7}$$

Requiring $g(x)$ to be single valued leads to the quantization condition on the magnetic charge,

$$q_m = 2\pi n, \qquad n \in \mathbb{Z}. \tag{10.8}$$

Here we have assumed that the electric charge is integer quantized. If the electric charges are multiples of q_e, we should then require $(g(x))^{q_e}$ to be single valued, which leads to the Dirac quantization condition

$$q_m q_e = 2\pi n, \qquad n \in \mathbb{Z}. \tag{10.9}$$

While the Dirac monopole solution by itself is singular, it can be embedded in UV completions of abelian gauge theory, where the $U(1)$ gauge group is embedded in a nonabelian gauge group that is spontaneously broken. The simplest such solutions were found by 't Hooft and by Polyakov, in $SU(2)$ Yang-Mills theory coupled to a scalar triplet Φ_a with a potential that breaks the $SU(2)$ spontaneously. Note that since the scalars are in the adjoint of $SU(2)$, their vev breaks the $SU(2)$ into a $U(1)$ subgroup that commutes with the scalar vev. This is unlike the Higgs doublet in the standard model. It is nonetheless common to refer to such a scalar field coupled to the nonabelian gauge field as a "Higgs field", and the Lagrangian "Yang-Mills-Higgs" Lagrangian.

The Lagrangian is

$$\mathcal{L} = -\frac{1}{4} F_{a\mu\nu} F_a^{\mu\nu} - D^\mu \Phi_a D_\mu \Phi_a + V(\Phi), \tag{10.10}$$

where

$$\begin{aligned} F_{a\mu\nu} &= \partial_\mu A_{a\nu} - \partial_\nu A_{a\mu} + g\epsilon_{abc} A_{b\mu} A_{c\nu}, \\ D_\mu \Phi_a &= \partial_\mu \Phi_a + g\epsilon_{abc} A_{b\mu} \Phi_c, \\ V(\Phi) &= \frac{\lambda}{4}(\Phi_a \Phi_a - v^2)^2. \end{aligned} \tag{10.11}$$

The precise form of the scalar potential is not important as long as it leads to the spontaneous symmetry breaking. We take it to be the quartic mexican hat potential to be concrete.

Classically, the vacuum expectation value of Φ^a is such that $\Phi_a \Phi_a = v^2$. The $SU(2)$ is broken to the $U(1)$ subgroup that commutes with $\langle \Phi_a \rangle t^a$. The perturbative spectrum consists of the $U(1)$ gauge field, massive "W-bosons" $W_\mu^\pm$ that acquire their mass gv through Higgs mechanism, and the component of Φ^a along the direction of the vev, with mass $\sqrt{2\lambda}v$. At energies much lower than gv or $\sqrt{2\lambda}v$, the theory is effectively that of a $U(1)$ gauge field. We can ask whether there are Dirac monopole like solutions of this $U(1)$ theory that embeds into the full $SU(2)$ Yang-Mills-Higgs theory as a smooth solution.

The answer is yes, and for a good reason. Loosely speaking, the unbroken $U(1)$ gauge field is the component of the nonabelian gauge field aligned with Φ. To be precise, this is true only in unitarity gauge, where Φ^a points along a fixed 3-vector in the $su(2)$. While this can be achieved by a gauge rotation locally, it cannot always be done globally. Firstly, to have finite total energy, Φ_a should approach a minimum of $V(\Phi)$ at spatial infinity, but need not approach the same value in all directions. The lowest energy configurations for Φ_a are parameterized by the sphere $S^2 : \Phi_a\Phi_a = v^2$ in the field target space. There are topologically distinct classes of allowed asymptotic boundary conditions for Φ at spatial infinity, namely, the map

$$\Phi : S^2_\infty \to S^2 \tag{10.12}$$

is classified by its winding number (labeling elements of the homotopy group $\pi_2(S^2)$). With such spatial dependent Φ_a, it is possible for A_μ to have nonzero magnetic flux at infinity and yet $A_{a\mu}$ are still well defined and non-singular in the entire space.

Let us examine more closely the behavior of Φ_a and $A_{a\mu}$ at spatial infinity. For the total energy to be finite, we also need $D_\mu\Phi_a$ to fall off at infinity faster than $r^{-3/2}$, where r is the radial distance. For

$$D_\mu\Phi_a = \partial_\mu\Phi_a + g\epsilon_{abc}A_{b\mu}\Phi_c \sim \mathcal{O}(r^{-3/2}), \tag{10.13}$$

the gauge field $A_{a\mu}$ should take the form

$$A_{a\mu} = -\frac{1}{gv^2}\epsilon_{abc}\Phi_b\partial_\mu\Phi_c + \frac{1}{v}\Phi_a a_\mu + \mathcal{O}(r^{-3/2}) \tag{10.14}$$

where a_μ is an arbitrary vector field. Note that a_μ is *not* the same as the vector potential of the unbroken $U(1)$. The latter is given by $\widehat{\Phi}_a A_{a\mu}$ in the *unitarity gauge*, where $\widehat{\Phi}_a = \Phi_a/|\Phi|$ is fixed to a constant unit vector. In this gauge, we may write the field strength of the unbroken $U(1)$ as

$$F_{\mu\nu} = \partial_\mu(\widehat{\Phi}_a A_{a\nu}) - \partial_\nu(\widehat{\Phi}_a A_{a\mu}). \tag{10.15}$$

Relaxing the gauge condition, this expression for $F_{\mu\nu}$ can be covariantized into

$$\begin{aligned} F_{\mu\nu} &= \partial_\mu(\widehat{\Phi}_a A_{a\nu}) - \partial_\nu(\widehat{\Phi}_a A_{a\mu}) - \frac{1}{g}\epsilon_{abc}\widehat{\Phi}_a\partial_\mu\widehat{\Phi}_b\partial_\nu\widehat{\Phi}_c \\ &= \widehat{\Phi}_a F^a_{\mu\nu} - \frac{1}{g}\epsilon_{abc}\widehat{\Phi}_a D_\mu\widehat{\Phi}_b D_\nu\widehat{\Phi}_c. \end{aligned} \tag{10.16}$$

For the field configuration (10.14) at large distance, the nonabelian field strength takes the form

$$\begin{aligned} &F_{a\mu\nu} = \partial_\mu A_{a\nu} - \partial_\nu A_{a\mu} + g\epsilon_{abc}A_{b\mu}A_{c\nu} \approx \frac{1}{v}\Phi_a F_{\mu\nu}, \\ &\text{where} \quad F_{\mu\nu} \approx -\frac{1}{gv^3}\epsilon_{abc}\Phi_a\partial_\mu\Phi_b\partial_\nu\Phi_c + \partial_\mu a_\nu - \partial_\nu a_\mu. \end{aligned} \tag{10.17}$$

The magnetic flux of the unbroken $U(1)$ field strength $F_{\mu\nu}$ on S^2_∞ is given precisely by the winding number of Φ, since a_μ, being well defined on the entire S^2_∞, does not have magnetic flux:

$$\int_{S^2_\infty} \frac{1}{2} F_{ij} dx^i \wedge dx^j = -\frac{1}{2gv^3} \int_{S^2_\infty} \epsilon_{abc} \Phi_a \partial_i \Phi_b \partial_j \Phi_c dx^i \wedge dx^j = -\frac{4\pi n}{g}, \tag{10.18}$$

where $n \in \mathbb{Z}$ is the winding number.

One could try to find the most general monopole solution carrying a given number of magnetic charge. It is useful to start with the lowest energy solutions of the given charge, and describe the rest in terms of fluctuations on the minimal energy solution. So instead of trying to solve the second order equation of motion, we will directly minimize the Hamiltonian. Write the nonabelian electric and magnetic fields

$$E_{ai} = F_{a0i}, \qquad B_{ai} = -\frac{1}{2}\epsilon_{ijk} F_{ajk}. \tag{10.19}$$

The minus sign in the convention for $\vec{B}_a$ is due to the sign in (10.18). The Hamiltonian is

$$H = \int d^3\vec{x} \left[\frac{1}{2} \left(\vec{E}_a^2 + \vec{B}_a^2 + (D_0\Phi_a)^2 + (\vec{D}\Phi_a)^2 \right) + V(\Phi) \right]. \tag{10.20}$$

We will consider static field configurations in temporal gauge $A_0 = 0$, with $\vec{E}_a = D_0\Phi_a = 0$. Next, we restrict ourselves to the sector with a given magnetic charge q_m, that is

$$q_m = \int_{S^2_\infty} \frac{1}{v} \Phi_a \vec{B}_a \cdot d\vec{\sigma} = \frac{1}{v} \int d^3\vec{x}\, \vec{B}_a \cdot \vec{D}\Phi_a \tag{10.21}$$

Without loss of generality, we will assume that q_m is positive. The energy of the configuration can then be written as

$$H = q_m v + \int d^3\vec{x} \left[\frac{1}{2} (\vec{B}_a - \vec{D}\Phi_a)^2 + V(\Phi) \right] \geq q_m v + \int d^3\vec{x}\, V(\Phi). \tag{10.22}$$

The inequality is saturated when

$$\vec{B}_a = \vec{D}\Phi_a. \tag{10.23}$$

This is known as the Bogomol'nyi equation. Our scalar potential $V(\Phi)$ is nonzero when evaluated on such a solution. In the small λ limit, we can ignore the contribution of $V(\Phi)$ to the energy of the monopole (V then only plays the role of allowing for spontaneous symmetry breaking); in this case, the minimal energy configuration is achieved by solving the Bogomol'nyi equation. Such magnetic monopole solutions are called BPS monopoles.

An explicit BPS monopole solution can be found using the ansatz

$$\begin{aligned} \Phi_a &= \frac{\hat{r}_a}{gr} H(vgr), \\ A_{ai} &= \epsilon_{aij} \frac{\hat{r}_j}{gr} (1 - K(vgr)). \end{aligned} \tag{10.24}$$

The nonabelian magnetic field is

$$\begin{aligned} B_{ai} &= -\epsilon_{ijk} \left(\partial_j A_{ak} + \frac{1}{2} g \epsilon_{abc} A_{bj} A_{ck} \right) \\ &= \delta_{ai} \frac{1 - K^2}{gr^2} + \left(\delta_{ai} - \frac{r_a r_i}{r^2} \right) (1 - K + r\partial_r) \frac{1 - K}{gr^2}. \end{aligned} \tag{10.25}$$

Note that this ansatz is neither rotationally invariant nor invariant under the global $SU(2)$ rotation, as the gauge index is mixed with the spatial index. It is, however, invariant under the diagonal subgroup of $SO(3)$ spatial rotation and $SU(2) \sim SO(3)$ global gauge rotation. The function $H(\rho)$ should have the asymptotic behavior $H(\rho) \to \rho$ as $\rho \to \infty$, so that Φ_a approaches the vacuum configuration at spatial infinity, while $H(\rho) \to 0$ as $\rho \to 0$ in order for Φ_a to be smooth at the origin. Similarly, $K(\rho) \to 1$ as $\rho \to 0$ for A_{ai} to be smooth at the origin, and $K(\rho) \to 0$ as $\rho \to \infty$, so that the magnetic charge is $4\pi/g$. The normalization of A_{ai} at spatial infinity is fixed in terms of the winding number of Φ, which is 1 in this case, by the Bogomol'nyi equation.

Let us now solve the equation

$$\begin{aligned} B_{ai} &= D_i \Phi_a = \partial_i \Phi_a + g \epsilon_{abc} A_{bi} \Phi_c \\ &= \delta_{ai} (1 + r\partial_r) \frac{H}{gr^2} - \left(\delta_{ai} - \frac{r_a r_i}{r^2} \right) \left[(1 + r\partial_r) \frac{H}{gr^2} - \frac{KH}{gr^2} \right]. \end{aligned} \tag{10.26}$$

The equation then reduces to

$$\begin{aligned} (1 + \rho\partial_\rho) \frac{H(\rho)}{\rho^2} &= \frac{1 - K(\rho)^2}{\rho^2}, \\ (1 - K + \rho\partial_\rho) \frac{1 - K(\rho)}{\rho^2} &= \frac{1 - K(\rho)^2 - K(\rho)H(\rho)}{\rho^2}. \end{aligned} \tag{10.27}$$

The solutions are

$$H(\rho) = \rho \coth \rho - 1, \qquad K(\rho) = \frac{\rho}{\sinh \rho}. \tag{10.28}$$

Now we have found one BPS solution carrying one unit of magnetic charge. But such BPS solutions are generally highly non-unique, though often all such solutions lie in a continuous family, and one solution can be deformed into another. Such a family in the one-monopole case (i.e. monopole with one unit of magnetic charge, as the one we found) is particularly simple to characterize. First of all we can shift the center of the monopole to an arbitrary position, which obviously still preserve the asymptotic behavior of the fields. Somewhat less trivially, we can also perform a "large" gauge transformation that does not reduce to identity at infinity; configurations related by large gauge transformations are distinct physical configurations. This is in contrast to "small" gauge transformations that reduce to identity at infinity, and configurations related by small gauge transformations are identified as the same physical configuration.

The only large gauge transformation of relevance here is that of the unbroken $U(1)$ by the vev of Φ at infinity. The corresponding gauge function can be chosen as

$$h = e^{\chi\Phi} = e^{\chi\Phi_a t^a}, \tag{10.29}$$

which commutes with Φ. The admissible large gauge transformation must preserve the falloff behavior of the fields at infinity; for $A_{a\mu}$ to still vanish at infinity requires that χ is a constant. The large gauge transformation by h leaves Φ invariant, but changes A_{ai} (and leaves $A_0 = 0$). Since χ is periodically valued, this gives rise to a circle worth of physically distinct one BPS monopole configurations. It turns out that these combined with the spatial translation are all deformations of the BPS solution with one unit of magnetic charge that preserve its energy. We say that the *moduli space* of one BPS monopole is $\mathcal{M}_1 \simeq \mathbb{R}^3 \times S^1$.

More generally, given say a BPS solution, we can consider a deformation $(\delta A_i, \delta\Phi)$ that leaves the BPS equation invariant, and preserves the Gauss law constraint (equation of motion with respect to A_0, and then set $A_0 = 0$ by temporal gauge condition; note that the Gauss law depends on time derivatives, and was automatically satisfied for static field configurations). That is, we demand

$$\begin{aligned} &-\epsilon_{ijk} D_j \delta A_k = D_i \delta\Phi - i[\delta A_i, \Phi], \\ &D_i \delta\dot{A}_i - i[\Phi, \delta\dot{\Phi}] = 0. \end{aligned} \tag{10.30}$$

Moving along the S^1 of the monopole moduli space corresponds to the solution

$$(\delta A_i, \delta\Phi) = (D_i(\chi\Phi), 0), \tag{10.31}$$

which is a large gauge transformation.

The low energy fluctuations of the monopole are described in terms of time dependent solutions to (10.30), corresponding to motion along the moduli space. Let us consider motion along the S^1, described by the deformation

$$(\delta A_i, \delta\Phi) = (D_i(\chi(t)\Phi), 0), \tag{10.32}$$

where $\chi(t)$ is some function of time t. While this deformation still solves the BPS equation, the time dependent A_i gives rise to a nonzero electric field and increases the energy by

$$\int d^3\vec{x}\frac{1}{2}\vec{E}_a^2 = \frac{1}{2}\dot{\chi}^2\int d^3\vec{x}(\vec{D}_i\Phi_a)^2 = \frac{1}{2}\dot{\chi}^2\int d^3\vec{x}\,\vec{B}_a\cdot\vec{D}_i\Phi_a = \frac{1}{2}\frac{4\pi v}{g}\dot{\chi}^2. \tag{10.33}$$

χ is an example of a collective coordinate of the monopole. Here we see that the low energy dynamics can be described in terms of an effective mechanics of the collective coordinate, with the Hamiltonian given by (10.33). More generally (say for the configuration of n monopoles), the minimal energy solutions of given charges are parameterized by a set of collective coordinates z^α, say

$$A_i = A_i(\vec{x}; z^\alpha), \quad \Phi = \Phi(\vec{x}; z^\alpha). \tag{10.34}$$

z^α are local coordinates on the moduli space $\mathcal{M}$ of solutions. The low energy dynamics of the monopoles can be described by making z^α to time dependent, namely to consider

$$A_i = A_i(\vec{x}; z^\alpha(t)), \quad \Phi = \Phi(\vec{x}; z^\alpha(t)). \tag{10.35}$$

The action $S[A_i, \Phi]$ evaluated on such field configurations takes the form

$$S = \frac{1}{2}\int dt\,\mathcal{G}_{\alpha\beta}(z)\dot{z}^\alpha\dot{z}^\beta \tag{10.36}$$

There are no potential term in z^α nor linear term in $\dot{z}^\alpha$ since z^α parameterize solutions to the equation of motion. We also assumed the action $S[A_i, \Phi]$ has up to two derivatives in the fields, and so involve only two time derivatives on $z^\alpha(t)$. This action is that of a non-relativistic particle moving on the moduli space $\mathcal{M}$ parameterized by z^α with metric $\mathcal{G}_{\alpha\beta}$. In the one monopole case, $\mathcal{G}_{\alpha\beta}$ is simply the flat metric on $\mathbb{R}^3\times S^1$. Quantization of this non-relativistic mechanics gives rise to states that carry n units of momentum along the S^1, where n is an integer. In other words, the wave function $e^{2\pi i n\chi/v}$ in this effective quantum mechanics describes a "dyon" state with one unit of monopole charge and n units of electric charge.

The 't Hooft-Polyakov monopoles we have describe requires the scalar field to be in the adjoint representation of the gauge group. They are not present for the $SU(2)\times U(1)$ Yang-Mills-Higgs theory in the standard model, though they would exist in various grand unification models, say the $SU(5)$ GUT where there are Higgs in the adjoint of $SU(5)$ with vev's that break the $SU(5)$ into $SU(3)\times SU(2)\times U(1)$. The mass of these monopoles is of order v/g, which is $1/g^2$ times that of the massive GUT vector bosons.

10.2 Yang-Mills instantons

In non-abelian gauge theory, there are gauge field configurations that contribute to the *Euclidean* path integral on which the θ term evaluates to nonzero numbers. These field configurations are not continuously connected to the trivial configuration $A_\mu = 0$. There are such solutions to the Euclidean equation of motion, which are local minima of the action functional; they look like a lump in both space and Euclidean time, and are hence called instantons.

Consider the functional (θ-term)

$$\mathcal{I} = \frac{1}{8}\int d^4x\, \epsilon^{\mu\nu\rho\sigma} F_{a\mu\nu} F_{a\rho\sigma}. \tag{10.37}$$

The integrand can be written in a total derivative form,

$$\frac{1}{4}\epsilon^{\mu\nu\rho\sigma} F^a_{\mu\nu} F^a_{\rho\sigma} = \epsilon^{\mu\nu\rho\sigma}\partial_\mu \left(A^a_\nu \partial_\rho A^a_\sigma + \frac{1}{3} f_{abc} A^a_\nu A^b_\rho A^c_\sigma \right), \tag{10.38}$$

but A_μ need not vanish by themselves at infinity. An admissible field configuration should have finite action, which requires the field strength to vanish at infinity. In other words, A_μ becomes a pure gauge at infinity,

$$A_\mu(x) \to i(g(x))^{-1}\partial_\mu g(x), \qquad |x| \to \infty, \tag{10.39}$$

where the gauge function $g(x)$ takes value in the gauge group G. The functional $\mathcal{I}$ then reduces to a boundary integral,

$$\begin{aligned} \mathcal{I} &= \frac{1}{2}\int_{S^3_\infty} \mathrm{Tr}\left(A \wedge dA - \frac{2i}{3} A \wedge A \wedge A \right) \\ &= \frac{1}{2}\int_{S^3_\infty} \mathrm{Tr}\left(-g^{-1}dg \wedge dg^{-1} \wedge dg - \frac{2}{3}(g^{-1}dg)^3 \right) \\ &= \frac{1}{6}\int_{S^3_\infty} \mathrm{Tr}\left[(g^{-1}dg)^3 \right]. \end{aligned} \tag{10.40}$$

We have seen a similar integral in the construction of the WZW term. This integral is invariant under arbitrary infinitesimal variation of $g(x)$, and is a topological invariant that is proportional to the winding number of the map $g(x) : S^3_\infty \to G$. We will determine the normalization through an explicit calculation below.

Now consider the *Euclidean* Yang-Mills action (related to Lorentzian action S via Wick rotation and $S = iS_E$),

$$S_E = \frac{1}{4}\int d^4x F^a_{\mu\nu} F^{a\mu\nu}. \tag{10.41}$$

We have the inequality

$$S_E \mp \mathcal{I} = \frac{1}{8}\int d^4x \left(F^a_{\mu\nu} \mp \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{a\rho\sigma}\right)^2 \geq 0, \tag{10.42}$$

and hence $S_E \geq |\mathcal{I}|$, where the equality holds if and only if the following equation is obeyed:

$$F^a_{\mu\nu} = \pm\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}F^{a\rho\sigma}. \tag{10.43}$$

This is called the self-dual (or anti-self-dual) Yang-Mills equation, or the instanton equation. A solution to this equation is a local minimum of the Euclidean action, with $S_E = |\mathcal{I}|$.

Let us consider the case of gauge group $G = SU(2)$. A simple nontrivial solution to the self-dual Yang-Mills equation is

$$A_\mu(x) = \frac{x^2}{x^2+R^2} i(g_1(\hat{x}))^{-1}\partial_\mu g_1(\hat{x}), \tag{10.44}$$

where R is an arbitrary length parameterizing the "size" of the instanton, $x^2 \equiv x^\mu x_\mu$, $x^\mu = (\vec{x}, x^4)$, $\hat{x} = x/|x|$, and $g_1(\hat{x})$ is the following map from the unit S^3 to $SU(2)$,

$$g_1(\hat{x}) = \hat{x}_4 + i\vec{\hat{x}}\cdot\vec{\sigma}. \tag{10.45}$$

$g_1(\hat{x})$ has winding number 1, and the corresponding solution $A_\mu(x)$ can be called a 1-instanton solution. The 1-instanton solution comes in a family, parameterized by the size R and the choice of x_4 direction (different choices are related by a constant $SU(2)$ gauge transformation). It is straightforward to calculation its action,

$$S_E = \mathcal{I} = 8\pi^2. \tag{10.46}$$

There are also families of n-instanton solutions, with $S_E = \mathcal{I} = 8\pi^2 n$, for all positive integer n, as well as n anti-instanton solutions with $S_E = -\mathcal{I} = 8\pi^2 n$. When the n instantons are far separated in Euclidean spacetime (or their size much smaller than their separation), the solution is approximately given by the superposition of the shifted 1-instanton solution. The exact multi-instanton solutions can be explicitly constructed (Atiyah-Drinfeld-Hitchin-Manin), though we will not discuss the general constructions here. The $SU(2)$ instanton solutions can be easily extended to the general nonabelian gauge group G by choosing an embedding of $SU(2)$ into G.

In writing the $SU(2)$ instanton solution, we have worked in the convention that the Yang-Mills coupling g multiplies the overall action and does not enter the structure constant. Taking into account the coupling g, the n-instanton (or anti-instanton) solution contributes to the path integral with a factor

$$\exp\left(-\frac{1}{4g^2}\int d^4x F^a_{\mu\nu}F^{a\mu\nu}\right) = \exp\left(-\frac{8\pi^2 n}{g^2}\right), \tag{10.47}$$

where $n \geq 1$. Such contributions are non-perturbative, in the sense that they have vanishing Taylor series in g around $g = 0$ and hence never appear in perturbation theory. This doesn't mean they are negligible, however. First of all, g runs under the renormalization group. For instance, in nonabelian gauge theory with a negative beta function (such as QCD), the one-loop running coupling at renormalization scale μ takes the form

$$g_\mu^2 = \frac{8\pi^2}{b \ln(\mu/\Lambda)}, \tag{10.48}$$

where Λ is the scale at which the theory is strongly coupled. One may expect that the contribution from an instanton of size R should be calculated with the renormalized coupling at scale $1/R$. Such a contribution comes with the factor

$$\exp\left(-\frac{8\pi^2 n}{g_{1/R}^2}\right) = (R\Lambda)^b. \tag{10.49}$$

When $R\Lambda$ is order 1 or greater, there is no sense in which the instanton contribution is suppressed; of course, we also cannot calculate the effect of "large" instanton in this naive manner.

Even when the theory stays weakly coupled, there can be amplitudes that receive contribution only through instantons, such as those that violate the conservation of charge associated with an anomalous symmetry. For instance, in QCD with massless quarks, the $U(1)$ axial symmetry is anomalous. The divergence of the $U(1)$ axial current is given by

$$\partial_\mu J_A^\mu = -\frac{1}{16\pi^2} \mathrm{tr}_R(t^a t^b) \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^b, \tag{10.50}$$

where the trace tr is taken over the representation of the fermion fields, namely the fundamental representation in the case of QCD. Under the axial $U(1)$ rotation,

$$\psi \to e^{i\alpha\gamma_5}\psi, \tag{10.51}$$

the path integral measure for the fermions picks up the phase factor

$$[d\bar{\psi}] \to [d\psi][d\bar{\psi}] \exp\left(-\frac{i\alpha}{16\pi^2} \mathrm{tr}_R(t^a t^b) \int d^4x \epsilon^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^b\right) = [d\psi][d\bar{\psi}] e^{4inC(R)\alpha}, \tag{10.52}$$

where n is the instanton number. Recall that $C(R)$ is defined as the normalization factor $\mathrm{tr}_R(t^a t^b) = C(R)\delta_{ab}$ (here we have left out the coupling in the definition of the gauge group generators t^a). For the fundamental representation, $C(\mathbf{f}) = \frac{1}{2}$. From this we see that the configuration of n instantons mediates an amplitude that violates the axial $U(1)$ quantum number by $4nC(R)$, or $2n$ in QCD with massless quarks. To actually computes such amplitudes, one must perform the functional integration over quantum fluctuations around the instanton solution, and also integrate over the moduli space of n instantons. We will carry out this computation for the 1-instanton contribution in the next section.

10.3 θ vacua and $U(1)_A$ violating amplitude

We will first compute the one-instanton contribution to the vacuum amplitude, or vacuum energy, of QCD. The result will give the dependence of the vacuum energy on the θ angle. We will consider a more general Lagrangian, that involves minimally coupled scalars as well as Dirac fermions. There can be mass terms for the fermions and scalar potential, though we will ignore them for the time being. It suffices to illustrate this computation for the $SU(2)$ gauge group, which will be assumed through this section. The Lagrangian is

$$\mathcal{L} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} - (D_\mu \Phi)^* D^\mu \Phi - \bar{\psi}\gamma^\mu D_\mu \psi, \tag{10.53}$$

with $F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\epsilon_{abc} A^b_\mu A^c_\nu$. We also need to introduce gauge fixing and ghost term. Expanding around a classical background $A_\mu = A^{cl}_\mu + A'_\mu$, we can work in the background field gauge, defined using the gauge condition $f_a = \overline{D}_\mu A'^\mu$ as before, where $\overline{D}_\mu$ is the covariant derivative with respect to the background gauge field A^{cl}_μ. The gauge fixing term and the ghost term are of the form

$$\mathcal{L}_{GF} = -\frac{1}{2}(\overline{D}_\mu A'^\mu)^2 + \bar{\eta}_a \overline{D}_\mu D^\mu \eta_a, \tag{10.54}$$

where $D_\mu = \overline{D}_\mu - ig[A'_\mu, \cdot]$ is understood. In perturbation theory around the trivial configuration $A_\mu = 0$, it is convenient to work with $A^{cl}_\mu = 0$. In perturbation theory around an instanton solution, on the other hand, it would be convenient to work in the background field gauge where A^{cl}_μ is the instanton solution itself. Previously, we have concluded that the path integral does not depend on the choice of f_a, as the functional determinant due to the integration over pure gauge modes of A_μ exactly cancel the Grassmannian functional Gaussian integral over the ghosts. While this cancelation holds for the gauge and ghost modes of nonzero eigenvalue with respect to their kinetic operators, it is subtle for the modes that are annihilated by the kinetic operator, i.e. *"zero modes"* (suitable finite volume regularization is generally needed). We will later determine the relative normalization of the instanton contribution after understanding the zero modes in the instanton background. For now, we will proceed to compute the functional integral at one-loop order around the one-instanton background in the background field gauge.

The one $SU(2)$ instanton solution centered at z^μ can be written explicitly in components as

$$A^{cl}_{a\mu} = \frac{2}{g} \frac{\eta_{a\mu\nu}(x-z)^\nu}{(x-z)^2 + R^2}, \tag{10.55}$$

where the symbol $\eta_{a\mu\nu}$ is anti-symmetric in μ, ν, defined as

$$\begin{aligned} &\eta_{aij} = \epsilon_{aij}, \quad i,j = 1,2,3, \\ &\eta_{ai4} = -\eta_{a4i} = \delta_{ai}. \end{aligned} \tag{10.56}$$

Note that the one anti-instanton solution would be obtained by replacing $\eta_{a\mu\nu}$ with $\bar{\eta}_{a\mu\nu}$, defined by $\bar{\eta}_{aij} = \eta_{aij}$, $\bar{\eta}_{ai4} = -\eta_{ai4}$.

The Euclidean action S_E is related to the Lorentzian one by $S = iS_E$. The one instanton solution has Euclidean action $S^{cl}_E = 8\pi^2/g^2$. In background field gauge, the action for the quantum fluctuations is

$$S_E = S^{cl}_E + \int d^4x \left[\frac{1}{2}(\overline{D}_\mu A'_{a\nu})^2 + g\epsilon_{abc} F^{\mu\nu}_a A'_{b\mu} A'_{c\nu} + (\overline{D}_\mu \Phi)^* (\overline{D}^\mu \Phi) + \bar{\psi}\gamma^\mu \overline{D}_\mu \psi + (\overline{D}_\mu \bar{\eta}_a)(\overline{D}^\mu \eta_a) + \cdots \right] \tag{10.57}$$

where we have exhibited only quadratic terms in the quantum fluctuations, which is all we need to the one-loop computation. Note that the term proportional to $(\overline{D}^\mu A'_\mu)^2$ from expanding the Yang-Mills Lagrangian is canceled by the gauge fixing term. As before, we will write $\mathbb{M}_A$, $\mathbb{M}_\Phi$, $\mathbb{M}_\psi$ and $\mathbb{M}_\eta$ for the kinetic operators on A', Φ, ψ, and $(\eta, \bar{\eta})$ respectively. The result of the one-loop Gaussian integral is then

$$e^{-8\pi^2/g^2} \int d(\text{zero modes}) \, \frac{\det' \mathbb{M}_\psi \det' \mathbb{M}_\eta}{\sqrt{\det' \mathbb{M}_A} \det' \mathbb{M}_\Phi}, \tag{10.58}$$

where we have separated out the integration over zero modes of A', Φ, ψ, η in the instanton background: these modes are annihilated by the respective kinetic operator $\mathbb{M}$; $\det'$ is the determinant taken over modes of nonzero eigenvalues only, defined using a suitable (gauge invariant) regulator.

Explicitly, the kinetic operators are

$$\begin{aligned} &\mathbb{M}_A A'_{a\mu} = -\overline{D}^2 A'_{a\mu} - 2g\epsilon_{abc} F_{b\mu\nu} A'_{c\nu}, \\ &\mathbb{M}_\psi = \gamma^\mu \overline{D}_\mu, \quad \mathbb{M}^2_\psi = \overline{D}^2 - \frac{i}{2}\gamma^{\mu\nu} F_{a\mu\nu} T^a, \\ &\mathbb{M}_\Phi = \mathbb{M}_\eta = -\overline{D}^2. \end{aligned} \tag{10.59}$$

Here T^a are the generators of the $SU(2)$ gauge group. Let us first calculate $\overline{D}^2$ in the background of an instanton with $z=0$ and $R=1$,

$$\begin{aligned}\overline{D}^2 &= (\partial_\mu - igT^a A^{cl}_{a\mu})(\partial^\mu - igT^b A^{cl,\mu}_b) \\ &= \Box - 2iT^a \eta_{a\mu\nu}\{\frac{x^\nu}{r^2+1}, \partial_\mu\} - 4T^a T^b \eta_{a\mu\nu}\eta_b{}^\mu{}_\rho \frac{x^\nu x^\rho}{(r^2+1)^2} \\ &= \Box + \frac{4}{r^2+1} T^a \eta_{a\mu\nu}(-ix^\nu \partial_\mu) - \frac{4r^2}{(r^2+1)^2} T^2,\end{aligned} \tag{10.60}$$

where $r^2 = x^\mu x_\mu$, $T^2 = T^a T^a$. Note that the $SO(4)$ rotation generator $-ix_{[\mu}\partial_{\nu]}$ can be split into the $SU(2)\times SU(2)$ generators

$$L^a_1 = -\frac{i}{2}\eta_{a\mu\nu} x^\mu \partial^\nu, \qquad L^a_2 = -\frac{i}{2}\overline{\eta}_{a\mu\nu} x^\mu \partial^\nu. \tag{10.61}$$

They obey

$$L^a_1 L^a_1 = L^a_2 L^a_2 \equiv L^2 = -\frac{1}{8}(x_\mu \partial_\nu - x_\nu \partial_\mu)^2. \tag{10.62}$$

Writing the Laplacian $\Box$ in terms of L^2 also, we have

$$\overline{D}^2 = \frac{1}{r^3}\partial_r\left(r^3 \frac{\partial}{\partial r}\right) - \frac{4L^2}{r^2} - \frac{8}{r^2+1} T^a L^a_1 - \frac{4r^2}{(r^2+1)^2} T^2. \tag{10.63}$$

This expression can be generalized to the kinetic operator acting on fields of nonzero spin,

$$\mathbb{M} = -\frac{1}{r^3}\partial_r\left(r^3 \frac{\partial}{\partial r}\right) + \frac{4L^2}{r^2} + \frac{8}{r^2+1} T^a L^a_1 + \frac{4r^2}{(r^2+1)^2} T^2 + \frac{16}{(r^2+1)^2} T^a S^a_1, \tag{10.64}$$

where S^a_1 together with S^a_2 are the $SU(2)\times SU(2)$ generators associated with the intrinsic spin of the field. The fermion ψ transforms in the representation $(\frac{1}{2},0)\oplus(0,\frac{1}{2})$ of the $SU(2)\times SU(2) \sim SO(4)$ rotation (Euclidean version of Lorentz) group, whereas A'_μ transforms in the representation $(\frac{1}{2},\frac{1}{2})$. Explicitly, acting on the spinor, S_1 and S_2 are

$$S^a_1 = -\frac{i}{8}\eta_{a\mu\nu}\gamma^{\mu\nu}, \quad S^a_2 = \frac{i}{8}\overline{\eta}_{a\mu\nu}\gamma^{\mu\nu}, \quad S^2_1 = \frac{3}{4}\frac{1-\gamma_5}{2}, \quad S^2_2 = \frac{3}{4}\frac{1+\gamma_5}{2}. \tag{10.65}$$

Acting on the vector, they are

$$(S^a_1)_{\mu\nu} = -\frac{i}{2}\eta_{a\mu\nu}, \quad (S^a_2)_{\mu\nu} = -\frac{i}{2}\overline{\eta}_{a\mu\nu}, \quad S^2_1 = S^2_2 = \frac{3}{4}. \tag{10.66}$$

$\mathbb{M}$ commutes with the following angular momentum operators

$$L^2, \qquad \vec{J}_1 \equiv \vec{L}_1 + \vec{S}_1 + \vec{T}, \qquad \vec{L}_2, \qquad \vec{S}_2. \tag{10.67}$$

Therefore, we can diagonalize $\mathbb{M}$ with modes of quantum numbers $(\ell, j_1, j_1^3, \ell_2^3, s_2^3)$,

$$\begin{aligned}
&L^2 = \ell(\ell+1), \quad \ell(=\ell_1=\ell_2) = 0, \frac{1}{2}, 1, \cdots, \\
&\vec{J}_1^2 = j_1(j_1+1), \quad |\ell - s_1 - t| \le j_1 \le \ell + s_1 + t, \quad j_1^3 = -j_1, -j_1+1, \cdots, j_1, \\
&\ell_2^3 = -\ell, \cdots, \ell, \\
&s_2^3 = -s_2, \cdots, s_2.
\end{aligned} \tag{10.68}$$

Here t is the spin of the representation of the field (on which $\mathbb{M}$ acts) with respect to the $SU(2)$ gauge group.

Let us compare this with the kinetic operator in the trivial background $A = 0$,

$$\mathbb{M}_0 = -\frac{1}{r^3}\partial_r\left(r^3\frac{\partial}{\partial r}\right) + \frac{4L^2}{r^2}. \tag{10.69}$$

We are interested in computing the ratio of determinants,

$$\frac{\det' \mathbb{M}}{\det' \mathbb{M}_0} \tag{10.70}$$

as a regularized product of nonzero eigenvalues. This is a bit complicated because $\vec{T}\cdot\vec{L}_1$ and $\vec{T}\cdot\vec{S}_1$ in the expression for $\mathbb{M}$ do not commute. For our purpose, s_1 is either 0 or $\frac{1}{2}$. The *nonzero* eigenvalues in the two cases are in fact the same set. This can be seen as follows.

Let us begin with an $s_1 = 0$ eigenmode Ψ obeying

$$\mathbb{M}\Psi = \lambda\Psi. \tag{10.71}$$

The s_2 quantum number does not enter the expression for $\mathbb{M}$. We could nonetheless think of Ψ as a right handed fermion in the representation $(0, \frac{1}{2})$, so that $\mathbb{M}$ acts as $-\mathbb{M}_\psi^2 = -(\gamma^\mu \overline{D}_\mu)^2$. Now consider

$$\Psi' = \gamma^\mu \overline{D}_\mu \psi. \tag{10.72}$$

Ψ' now is in the representation $(\frac{1}{2}, 0)$, and in particular has $s_1 = \frac{1}{2}$. Clearly it has the same eigenvalue λ with respect to $\mathbb{M}$, and is non-vanishing when $\lambda \neq 0$. This shows that the $s_1 = 0$ eigenvalues of $\mathbb{M}$ are in 1-1 correspondence with the $s_1 = \frac{1}{2}$ eigenvalues of $\mathbb{M}$. This argument works for any s_2, and therefore also relates the nonzero eigenmodes of the scalar and vector fields in the $SU(2)$ representation of spin t.

To compute the determinant for $\mathbb{M}$ acting on the scalar field, a further simplification occurs if we consider instead the operator

$$\begin{aligned}
\mathbb{V} &= \frac{1}{4}(r^2+1)\mathbb{M}(r^2+1) \\
&= -\frac{1}{4}(r^2+1)^2\left[\frac{1}{r^3}\partial_r\left(r^3\frac{\partial}{\partial r}\right) - \frac{4L^2}{r^2} - \frac{4}{r^2+1}(J_1^2 - L^2) + \frac{4r^2}{(r^2+1)^2}T^2\right].
\end{aligned} \tag{10.73}$$

Note the flipped sign of the T^2 term. Although the set of eigenvalues of $\mathbb{V}$ are not simply related to those of $\mathbb{M}$, the ratio of determinants

$$\frac{\det' \mathbb{V}}{\det' \mathbb{V}_0} \tag{10.74}$$

is formally the same as (10.70). This is not quite correct because we need to regularize the determinant, say using Pauli-Villars regularization. But the Pauli-Villars regularization for the $\mathbb{V}$ determinant with regulator mass M would amount to doing Pauli-Villars regularization for $\mathbb{M}$ but with a spatially dependent mass for the regulator field, $2M/(1+r^2)$. The correction factor due to this difference in the regulator is nontrivial and will be computed shortly.

To proceed, we shall carry out the following steps:

I. Compute the determinant over nonzero modes of $\mathbb{V}$ via Pauli-Villars regularization, for $\mathbb{V}$ acting on scalar fields.

II. Compute the correction factor due to the change of regulator mass from $M/(1+r^2)$ to M.

III. The zero mode integral must be treated separately, as they do not give rise to Gaussian integral. There are zero modes of A' and ψ in the instanton background. The zero modes of A' are associated with translation, dilaton, and global gauge rotation of the instanton solution.

10.3.1 The nonzero modes

The equation $\mathbb{V}\Psi = \lambda\Psi$ is of the type of hypergeometric equation. The eigenvalues are found to be

$$\lambda_n = (n+\ell+j_1+1-t)(n+\ell+j_1+2+t), \qquad n = 0, 1, 2, \cdots . \tag{10.75}$$

Note that since $j_1 \geq |\ell - t|$, there are no zero eigenvalues for the scalar field in any representation of the $SU(2)$ gauge group. $\mathbb{V}_0$ has a corresponding set of eigenvalues $\lambda_n^{(0)} = (n+2\ell+1)(n+2\ell+2)$. The logarithm of the determinant over nonzero eigenvalues is

$$\ln \frac{\det' \mathbb{V}}{\det' \mathbb{V}_0} = \sum_n \ln \frac{\lambda_n}{\lambda_n^{(0)}} \;\to\; \sum_n \sum_{i=0}^{N} e_i \ln \frac{\lambda_n + M_i^2}{\lambda_n^{(0)} + M_i^2} \tag{10.76}$$

where we now introduce N Pauli-Villars regulator fields with statistics relative to the original physical field labelled by $e_i = \pm 1$, and mass M_i, $i = 1, \cdots, N$, subject to the conditions

$$\sum_{i=1}^{N} e_i = -1, \quad \sum_{i=1}^{N} e_i M_i^k = 0, \quad k = 1, 2, \cdots, k_0, \quad \sum_{i=1}^{N} e_i \ln M_i \equiv -\ln \mu, \tag{10.77}$$

where k_0 is sufficiently large to ensure that all power divergences cancel (by taking N to be sufficiently large and taking M_i to infinity at the same time with appropriate ratios). If the $\ln \mu$ term survives in the end, it will give rise to a logarithmic dependence on the renormalization scale, reflecting the running under RG.

The summation is over $n(= 0, 1, 2, \cdots)$, $\ell(= 0, \frac{1}{2}, 1, \cdots)$, and j_1 subject to the constraint $|\ell - t| \leq j_1 \leq \ell + t$, and there is a degeneracy of $(2j_1 + 1)(2\ell + 1)$ for each (n, ℓ, j_1). Let us change the summation variables from n, ℓ, j_1 to,

$$w = n + j_1 + \ell - t, \quad p = j_1 + \ell - t, \quad q = j_1 - \ell + t, \tag{10.78}$$

so that the ranges are $w = 0, 1, 2, \cdots$, $0 \leq p \leq w$, $0 \leq q \leq 2t$. We can now write the sum

$$\begin{aligned}\sum \ln(\lambda_n + M^2) &= \sum_{w=0}^{\infty} \sum_{p=0}^{w} \sum_{q=0}^{2t} (2j_1 + 1)(2\ell + 1) \ln\left[(w + 1)(w + 2t + 2) + M^2\right] \\ &= \frac{2t + 1}{3} \sum_{w=0}^{\infty} (w + 1)(w + t + \frac{3}{2})(w + 2t + 2) \ln\left[(w + 1)(w + 2t + 2) + M^2\right].\end{aligned} \tag{10.79}$$

The ratio of the regularized determinants is given by

$$\ln \frac{\det' \mathbb{V}}{\det' \mathbb{V}_0} = \frac{2t + 1}{3} \sum_{i=0}^{N} e_i \left[f_{t+\frac{1}{2}}(M_i) - f_{\frac{1}{2}}(M_i)\right], \tag{10.80}$$

where the function $f_s(M)$ is

$$f_s(M) \equiv \sum_{r=s+1}^{\Lambda} r(r^2 - s^2) \ln(r^2 - s^2 + M^2). \tag{10.81}$$

Λ will be taken to infinity after performing the sum over i (Pauli-Villars regulators). In the limit of large Λ and large M, this sum can be performed using Euler-Maclaurin formula,

$$\sum_{r=s+1}^{\Lambda} h(r) = \int_s^{\Lambda} h(x)dx + \left[\frac{1}{2}h(x) + \frac{1}{12}h'(x) - \frac{1}{720}h'''(x) + \cdots\right]\Bigg|_s^{\Lambda}. \tag{10.82}$$

From this one can derive

$$\begin{aligned}f_s(M) = &-\frac{s^2}{2}M^2 - \frac{s^2}{6}\ln M^2 - s^2\left[(\Lambda^2 + \Lambda + \frac{1}{6})\ln\Lambda + \frac{\Lambda}{2} + \frac{1}{4}\right] + \frac{s^4}{4}(2\ln\Lambda + 1) \\ &+ \mathcal{O}(\frac{1}{M}) + (\text{terms independent of } s),\end{aligned} \tag{10.83}$$

where terms that vanish in the large Λ limit are dropped. The terms that are independent of s will cancel in (10.80).

The $i = 0$ contribution has $M = 0$,

$$f_s(0) = \sum_{r=s+1}^{\Lambda} r(r^2 - s^2) \ln(r^2 - s^2). \tag{10.84}$$

This sum cannot be approximated using Euler-Maclaurin formula.

Putting them together, taking the Pauli-Villars regulator masses to infinity subject to the constraints, and finally taking the $\Lambda \to \infty$ limit, one finds the following formula for the regularized determinant,

$$\ln \frac{\det' \mathbb{V}}{\det' \mathbb{V}_0} = \frac{t(t+1)(2t+1)}{3} \left[\frac{\ln M}{3} + \frac{1}{t(t+1)} \sum_{r=1}^{2t+1} r(2t+1-r)(r-t-\frac{1}{2}) \ln r - \frac{t(t+1)}{3} - \frac{1}{6} - 4\zeta'(-1) \right] \tag{10.85}$$

10.3.2 Adjusting regulator mass

Now let us consider the effect of changing the Pauli-Villars regulator mass (for the original kinetic operator $\mathbb{M}$) from M to $2M/(r^2+1)$. Consider a scalar regulator field Φ' of mass M. Under an infinitesimal variation $\delta M(x)$, the effective action is shifted by

$$\begin{aligned} \delta S &= \delta \operatorname{Tr} \ln(-\overline{D}^2 + M^2) \\ &= \operatorname{Tr} \left[\delta M^2 \frac{1}{-\overline{D}^2 + M^2} \right] \\ &= \int d^4x \, \delta M^2(x) \langle x | \frac{1}{-\overline{D}^2 + M^2} | x \rangle. \end{aligned} \tag{10.86}$$

In the presence of a background gauge field $A_\mu(x)$, the contribution that survives the $M \to \infty$ comes from the triangle diagram (similarly to the derivation of chiral anomaly) together with a bubble diagram that cancels the logarithmic divergence,

$$
\begin{aligned}
\delta\mathcal{L}(x) &= \delta M^2(x) 2g^2 \int \frac{d^4p d^4q}{(2\pi)^8} e^{i(p+q)\cdot x} \mathrm{tr} \widetilde{A}_\mu(p) \widetilde{A}_\nu(q) \\
&\times \int \frac{d^4k}{(2\pi)^4} \left[\frac{(2k-p)^\mu (2k+q)^\nu}{(k^2+M^2)((k+q)^2+M^2)((k-p)^2+M^2)} - \frac{\delta^{\mu\nu}}{((k+q)^2+M^2)((k-p)^2+M^2)} \right] \\
&= \delta M^2(x) 2g^2 \int \frac{d^4p d^4q}{(2\pi)^8} e^{i(p+q)\cdot x} \mathrm{tr} \widetilde{A}_\mu(p) \widetilde{A}_\nu(q) \cdot 2 \int_0^1 dx \int_0^{1-x} dy \int \frac{d^4k}{(2\pi)^4} \\
&\times \frac{(2k-(1-2x)p-2yq)^\mu (2k+2xp+(1-2y)q)^\nu - \delta^{\mu\nu}((k+xp-yq)^2+M^2)}{[k^2+xp^2+yq^2-(xp-yq)^2+M^2]^3} \\
&\to \frac{\delta M^2}{32\pi^2 M^2} 2g^2 \int \frac{d^4p d^4q}{(2\pi)^8} e^{i(p+q)\cdot x} \mathrm{tr} \widetilde{A}_\mu(p) \widetilde{A}_\nu(q) \\
&\times 2 \int_0^1 dx \int_0^{1-x} dy \left[(-(1-2x)p-2yq)^\mu (2xp+(1-2y)q)^\nu - \delta^{\mu\nu}(2(xp-yq)^2+M^2-(xp^2+yq^2)) \right] \\
&= -\frac{\delta M^2}{32\pi^2 M^2} 2g^2 \int \frac{d^4p d^4q}{(2\pi)^8} e^{i(p+q)\cdot x} \mathrm{tr} \widetilde{A}_\mu(p) \widetilde{A}_\nu(q) \frac{1}{3} \left[q^\mu p^\nu - \delta^{\mu\nu} p\cdot q + 3M^2 \delta^{\mu\nu} \right]
\end{aligned}
\tag{10.87}
$$

The term proportional to δM^2 will be canceled by the condition on Pauli-Villars regulator masses. We are left with

$$
\begin{aligned}
&-\frac{\delta M^2}{32\pi^2 M^2} \frac{2g^2}{3} \int \frac{d^4p d^4q}{(2\pi)^8} e^{i(p+q)\cdot x} \mathrm{tr} \widetilde{A}_\mu(p) \widetilde{A}_\nu(q) \left(q^\mu p^\nu - \delta^{\mu\nu} p\cdot q \right) \\
&\to -\frac{\delta M^2}{32\pi^2 M^2} \cdot \frac{g^2 t(t+1)(2t+1)}{18} F_{a\mu\nu} F_a^{\mu\nu}
\end{aligned}
\tag{10.88}
$$

where in the last step we restored the higher order terms in A by gauge invariance, and the trace is taken over the spin t representation of $SU(2)$. When δM is constant in space, this is nothing but the contribution from a scalar field to the beta function of the $SU(2)$ Yang-Mills coupling. We see here that even when δM is spatially dependent, the effect of adjusting the regulator mass is accounted by local counterterms. Integrating (10.88) in M gives

$$
\Delta S = -\frac{g^2}{32\pi^2} \frac{t(t+1)(2t+1)}{9} \int d^4x \ln \frac{M(x)}{M_0} F_{a\mu\nu} F_a^{\mu\nu}. \tag{10.89}
$$

With $M(x) = 2M/(r^2+1)$, and the instanton field strength

$$
F_{a\mu\nu} = -\frac{4}{g} \frac{\eta_{a\mu\nu}}{(r^2+1)^2}, \tag{10.90}
$$

we can compute the correction to the effective action,

$$
\begin{aligned}
\Delta S &= -\frac{g^2}{32\pi^2} \frac{t(t+1)(2t+1)}{9} \frac{16}{g^2} \int d^4x \frac{12}{(r^2+1)^4} \ln \frac{2M}{M_0(r^2+1)} \\
&= -\frac{t(t+1)(2t+1)}{9} \left(\ln \frac{2M}{M_0} - \frac{5}{6} \right).
\end{aligned}
\tag{10.91}
$$

Combining this with the regularized $\mathbb{V}$-determinant over nonzero modes, we obtain the fully Pauli-Villars regularized determinant over nonzero modes,

$$\ln \frac{\det' \mathbb{M}}{\det' \mathbb{M}_0} = \frac{t(t+1)(2t+1)}{3} \left[\frac{1}{3} \ln \frac{M_0}{2} + \frac{1}{t(t+1)} \sum_{r=1}^{2t+1} r(2t+1-r)(r-t-\frac{1}{2}) \ln r \right.$$
$$\left. - \frac{t(t+1)}{3} + \frac{1}{9} - 4\zeta'(-1) \right]. \tag{10.92}$$

Note that this result is derived for the instanton of size $R = 1$. In restoring the R-dependence, M_0 will be replaced by $M_0 R$.

10.3.3 The zero mode integral

The zero modes of $\mathbb{M}$ on fields of general spin are not related to modes of the scalar field, and must be analyzed separately for each spin. Zero modes of the vector field in the instanton background can be found either by directly solving the equation $\mathbb{M}\Psi = 0$, or by varying the instanton solution with respect to its collective coordinates, which include 4 translations (parameterized by the center position z^μ), 1 dilation (the size R), and 3 constant $SU(2)$ gauge rotations.

To determine the zero modes of the spin $\frac{1}{2}$ fermion we must solve the equation $\mathbb{M}\Psi = 0$, with $(s_1, s_2) = (\frac{1}{2}, 0)$ or $(0, \frac{1}{2})$. Since only S_1 enters the expression for $\mathbb{M}$, and we know that there are no zero modes for the scalar field, the fermion zero mode must have $s_1 = \frac{1}{2}$, $s_2 = 0$. We will consider the example of fermions in the fundamental representation of $SU(2)$, i.e. $t = \frac{1}{2}$. On modes with $\ell = 0$, $\mathbb{M}$ simplifies to

$$\mathbb{M} = -\frac{1}{r^3} \partial_r \left(r^3 \frac{\partial}{\partial r} \right) + \frac{3r^4}{(r^2+1)^2} + \frac{8(j_1(j_1+1) - \frac{3}{2})}{(r^2+1)^2}, \tag{10.93}$$

where $j_1 = 0, 1$. The zero mode has $j_1 = 0$, and is given by

$$\Psi(r) = \frac{1}{(r^2+1)^{\frac{3}{2}}}. \tag{10.94}$$

We see that for each flavor of massless fermion (quark) there is a *complex* zero mode, or two real zero modes, in the one-instanton background. The zero modes are purely left handed. The Grassmannian integral over the fermion zero modes would vanish, unless the operator insertions in the path integral contain the fermion zero modes. Non-vanishing amplitudes therefore violate the axial $U(1)$ quantum number precisely by the number of chiral zero modes (minus anti-chiral zero modes, which are absent in this case), namely $2N_f$ for N_f flavors and $SU(2)$ gauge group.

Let us now analyze the integration over zero modes of the $SU(2)$ gauge vector fields. They can be obtained by solving the equation $\mathbb{M}\Psi = 0$ with $s_1 = s_2 = \frac{1}{2}$, $t = 1$. The zero modes are:

$$\begin{aligned} j_1 &= \frac{1}{2}, \quad \ell = 0, \quad \Psi(r) = \frac{1}{(r^2+1)^2}, \qquad (2j_1+1)(2s_2+1) = 4 \text{ zero modes;} \\ j_1 &= 0, \quad \ell = \frac{1}{2}, \quad \Psi(r) = \frac{r}{(r^2+1)^2}, \qquad (2\ell+1)(2s_2+1) = 4 \text{ zero modes.} \end{aligned} \tag{10.95}$$

The first set of 4 zero modes are associated with translation. The second set of 4 zero modes split into a singlet and a triplet with respect to $\vec{J}_2$. The singlet is associated with dilation, whereas the triplet is associated with $SU(2)$ global gauge rotation, as we shall see.

The translational zero modes can be written as

$$\begin{aligned} (A_{(\nu)})_{a\mu} &= -\frac{\partial}{\partial z^\nu}\left[\frac{2}{g}\frac{\eta_{a\mu\nu}(x-z)^\nu}{(x-z)^2+R^2}\right] + D_\mu \Lambda_{(\nu)a} \\ &= \frac{4}{g}\frac{\eta_{a\mu\nu}}{(r^2+1)^2}, \end{aligned} \tag{10.96}$$

where $\Lambda_{(\nu)a}$ is some gauge parameter. The gauge transformation by Λ is needed because the zero modes are those of the kinetic operator with a particular gauge fixing term, namely that of the background field gauge in the one instanton background.

The functional integral measure for a general field $\phi(x)$ can be written in terms of the integral over the eigenmodes of the kinetic operator as follows. Decompose $\phi(x) = \sum_n c_n \phi_n(x)$, where $\phi_n(x)$ are an orthonormal basis of eigenmodes. We can then write

$$[D\phi] = \prod_n dc_n. \tag{10.97}$$

If $\phi_n(x)$ are not normalized, we have

$$[D\phi] = \prod_n dc_n \left[\int d^4x (\phi_n(x))^2\right]^{\frac{1}{2}}. \tag{10.98}$$

While the overall normalization of the measure is ambiguous and cancels in the computation of expectation values, in writing the Gaussian integral as the determinant we have implicitly adopted the normalization with an extra factor of $\pi^{-\frac{1}{2}}$. The zero mode integral therefore should be written as

$$\int dc_0 \left[\frac{1}{\pi}\int d^4x (\phi_0(x))^2\right]^{\frac{1}{2}}. \tag{10.99}$$

The integral over the translational zero modes of the instanton is therefore

$$\int \prod_{\nu} dz^{\nu} \left[\frac{1}{\pi} \int d^4x (A_{(\nu)})_{a\mu} (A_{(\nu)})_a^{\mu} \right]^{\frac{1}{2}} = \int d^4 z \left(\frac{8\pi}{g^2} \right)^2 \tag{10.100}$$

This is not all. We have adopted Pauli-Villars regulator with mass M_0 (more precisely, $M_0 = \prod_{i=1}^{N} M_i^{-e_i}$), and multiply the functional integral by that of the regulator fields. The zero modes of A'_{μ} also gives rise to modes of the corresponding regulator field, of eigenvalue M_0^2 rather than zero with respect to $\mathbb{M}$. Therefore, after including the regulator, the integral over translational zero modes is

$$M_0^4 \left(\frac{8\pi}{g^2} \right)^2 \int d^4 z. \tag{10.101}$$

In the absence of fermions, if we compute the vacuum amplitude, the integral over translational zero modes of the instanton just gives the spacetime volume. After dividing the result by the spacetime volume, we obtain a contribution to the vacuum energy density. If we compute an amplitude with operator insertions that depend on the position of the instanton, we then have the integral of some function of z over the entire spacetime, much like the integration of an interaction vertex in the ordinary Feynman rule.

The dilation zero mode is treated similarly,

$$(A_{(D)})_{a\mu} = \frac{\partial}{\partial R} \left[\frac{2}{g} \frac{\eta_{a\mu\nu} x^{\nu}}{x^2 + R^2} \right] \Bigg|_{R=1} = \frac{4}{g} \frac{\eta_{a\mu\nu} x^{\nu}}{(r^2+1)^2}. \tag{10.102}$$

We have been restricting ourselves to the $R = 1$ case, for the simplicity of notation; the R dependence will be restored easily by dimensional analysis in the end. After including the regulator, the integration measure over the dilation zero mode, i.e. integration over the size modulus of the instanton, at $R = 1$, is

$$\mu_0 \frac{4\sqrt{\pi}}{g} dR. \tag{10.103}$$

Now let us turn to the remaining 3 zero modes, related to the constant $SU(2)$ gauge rotation. These zero modes of the kinetic operator $\mathbb{M}$ are

$$(A_{(b)})_{a\mu} = 2\eta_{a\rho\mu} \eta_b{}^{\rho}{}_{\nu} \frac{x^{\nu}}{(r^2+1)^2}. \tag{10.104}$$

The normalization of (10.104) is arbitrary so far. The question is how to determine the normalization of the zero mode integral. We would like to do so by relating (10.104) to the constant gauge variations of the instanton solution itself. The latter for gauge parameter $(\Lambda_{(b)})_a = \delta_{ab}$ are given by

$$D_\mu(\Lambda_{(b)})_a = 2\epsilon_{acb}\eta_{c\mu\nu}\frac{x^\nu}{r^2+1}. \tag{10.105}$$

(10.105) is not a zero mode in the A^{cl}_μ-background field gauge (it would be a zero mode in the Lorentz covariant gauge i.e. background field gauge with trivial background $A_\mu = 0$). Both (10.104) and (10.105) are of a pure gauge form, though the corresponding gauge parameters are not square normalizable. Namely, the zero mode $A_{(b)}$ can be written as

$$(A_{(b)})_{a\mu} = D_\mu(\Omega_{(b)})_a, \qquad (\Omega_{(b)})_a = \eta_{a\rho\mu}\eta_b{}^\rho{}_\nu\frac{x^\mu x^\nu}{r^2+1}. \tag{10.106}$$

We now decompose $\Omega_{(b)}$ into a component proportional to the constant gauge rotation $\Lambda_{(b)}$, and an orthogonal component. Even though neither $\Omega_{(b)}$ nor $\Lambda_{(b)}$ are square normalizable, we can put the theory in a ball of some large but finite radius, make this orthogonal decomposition on the gauge parameters, and then take the ball to infinite size again. This procedure yields

$$(\Omega_{(b)})_a = (\Lambda_{(b)})_a + \left(\delta_{ab}\frac{r^2-1}{r^2+1} + \frac{\eta_{a\rho\mu}\eta_b{}^\rho{}_\nu x^\mu x^\nu - \delta_{ab}r^2}{r^2+1}\right). \tag{10.107}$$

We learn that $A_{(b)}$ is in fact the correctly normalized zero mode associated with the constant $SU(2)$ gauge rotation. The zero mode integral is then written as (including Pauli-Villars regulator)

$$\left[\frac{1}{\pi}\int d^4x (A_{(b)})_{a\mu}(A_{(b)})^{a\mu}\right]^{\frac{3}{2}} M_0^3\frac{1}{2}\int_{SU(2)}[d^3\vec{\xi}] = (4\pi)^{\frac{3}{2}}M_0^3\frac{1}{2}\int_{SU(2)}[d^3\vec{\xi}], \tag{10.108}$$

where the constant gauge parameter $\vec{\xi}$ is related to the $SU(2)$ element U by

$$U = e^{i\frac{g}{2}\vec{\xi}\cdot\vec{\sigma}}. \tag{10.109}$$

Note that $SU(2)$ elements that differ by multiplication by -1 give the same gauge transformation on the gauge fields. Our integration over $SU(2)$ here come from the zero mode integral of the gauge field, and hence we need to multiply the $SU(2)$ volume by $\frac{1}{2}$. The invariant measure on the $SU(2)$ gauge group therefore gives (recall our parameterization of the pion in the chiral Lagrangian)

$$\frac{1}{2}\int_{SU(2)}[d^3\vec{x}] = \frac{1}{2}\int\frac{d^3\vec{\xi}}{\left[1+(\frac{g}{4})^2\vec{\xi}^2\right]^3} = \frac{8\pi^2}{g^3}. \tag{10.110}$$

The result is that the integration over the three zero modes $A_{(b)}$ gives a factor

$$(4\pi)^{\frac{3}{2}}M_0^3\frac{8\pi^2}{g^3}. \tag{10.111}$$

10.3.4 Putting everything together

The zero mode integral together with the regularized determinant over nonzero modes, with the R-dependence restored from dimension analysis, now give the one-instanton contribution

$$\int 2^{14}\pi^6 g^{-8}\frac{d^4z dR}{R^5}\exp\left[-\frac{8\pi^2}{g_{M_0}^2}-b\ln(M_0R)-\alpha\right]=\int 2^{14}\pi^6 g^{-8}\frac{d^4z dR}{R^5}\exp\left(-\frac{8\pi^2}{g_{1/R}^2}-\alpha\right), \tag{10.112}$$

where b is precisely the one-loop beta function coefficient, so that the logarithmic correction turns g_{M_0} into $g_{1/R}$. The constant α collects the constant terms in (10.92) from the gauge field, ghost, and matter fields. Its explicit expression is not important for now. When there are fermion matter fields that couple to the $SU(2)$ gauge fields, as we have seen, there is also an integration over the fermion zero modes, which gives zero when there are no operator insertions in the path integral.

Let us first consider the theory of pure $SU(2)$ gauge fields, possibly coupled to scalar fields, but with no fermions. (10.112) gives the one-instanton contribution to the vacuum energy. The z^μ integral simply gives the spacetime volume factor. The integration over the size modulus R may seem problematic. When $b < -4$ (which is the case for pure $SU(2)$ Yang-Mills theory coupled to sufficiently few matter fields, for instance), there is no UV divergence from the integration over small R. This is because the instanton action computed using the renormalized coupling is large at high mass scale, due to asymptotic freedom.

There would be, however, an IR divergence from the integration over large size instantons. This IR divergence may be cutoff either by the strong coupling dynamics of the gauge theory at large distances, or if the gauge symmetry is spontaneously broken by the vev of the scalar (with an Mexican hat potential). In the former case, we expect the R integral to be cut off at $R \sim 1/\Lambda_0$, where Λ_0 is the scale at which the renormalized coupling diverges. The contribution from large instantons then goes beyond the control of perturbation theory, and we cannot calculate the one instanton contribution reliably. The latter case is calculable if the coupling stays small at the symmetry breaking scale. To see this, suppose the scalar field ("Higgs") has vev $\Phi^2 = v^2$ in the vacuum. We will consider the limit of small "Higgs" mass (i.e. small quartic coupling λ), so that the equation for Φ can be approximated by that of a massless scalar coupled to the gauge field,

$$\overline{D}_\mu \overline{D}^\mu \Phi = 0, \qquad \Phi^2 \to v^2 \ (r\to\infty). \tag{10.113}$$

The zero mode with the desired asymptotic value has $j_1 = 0$, $\ell = t$, with

$$|\Phi(x)| = v\left(\frac{r^2}{r^2+1}\right)^t. \tag{10.114}$$

The scalar kinetic term then contributes to the action (using Φ equation of motion)

$$\begin{aligned}
\int d^4x (D_\mu \Phi)^* (D^\mu \Phi) &= \int d^4x \partial_\mu (\Phi^* D^\mu \Phi) \\
&= \int_{S^3_\infty} \Phi^* D_r \Phi \\
&= \int_{S^3_\infty} \Phi^* \partial_r \Phi \\
&= 4\pi^2 t v^2.
\end{aligned} \tag{10.115}$$

This is the result for the $R = 1$ instanton. Restoring R dependence, we find that the "Higgs" kinetic term contributes $4\pi^2 t v^2 R^2$ to the action, giving a factor $\exp(-4\pi^2 t v^2 R^2)$ in the path integral which cuts off the large R divergence. Note that the instanton solution of the gauge field itself is not an exact classical solution any more, but the Higgs action balances the quantum corrections to the instanton action in the path integral.

Let us also include the θ term in the Euclidean action,

$$i\theta \frac{g^2}{64\pi^2} \int d^4x \epsilon^{\mu\nu\rho\sigma} F_{a\mu\nu} F_{a\rho\sigma} = in\theta, \tag{10.116}$$

where n is the instanton number. Since n is integer valued, θ is defined up to periodicity 2π. Note that a priori one might think that it is possible to weigh the contributions from sectors of different instanton number n with some arbitrary function of n, but in order for cluster decomposition to hold, this weighting factor must be multiplicative for far separated instantons and hence of the form $e^{-in\theta}$.

Now given the one instanton contribution to the vacuum amplitude of the form $KVTe^{-S_0}$, where $S_0 = 8\pi^2/g^2$ is the instanton action, and K the quantum correction with the integral over translational zero mode (giving the spacetime volume factor VT) excluded, we can sum over the contribution from arbitrarily many instantons and anti-instantons in the dilute gas approximation, where the instantons and anti-instantons are far apart so that the solution is given by the superposition,

$$\sum_{n,\overline{n}=0}^{\infty} \left(KVTe^{-S_0}\right)^{n+\overline{n}} \frac{e^{i(n-\overline{n})\theta}}{n!\overline{n}!} = \exp\left(2KVTe^{-S_0}\cos\theta\right). \tag{10.117}$$

The factor $1/n!$ is due to the fact that the n instantons are identical and the integral over the translational zero modes really gives $(VT)^n/n!$. From this we find the θ-dependence of the vacuum energy density

$$\frac{E(\theta)}{V} = -2Ke^{-S_0}\cos\theta. \tag{10.118}$$

Finally, let us include N_f Dirac fermions with $t = \frac{1}{2}$. As already mentioned, if the fermions are massless, the vacuum amplitude is zero due to the fermion zero mode integral. If the fermions have a very small mass m, then what would be the (chiral) zero mode is replaced by the mode of eigenvalue m with respect to $\mathbb{M}_\psi$. The Grassmannian integral over this lowest mode (combined with the contribution from the Pauli-Villars regulator) gives

$$\left(\frac{m}{M_0}\right)^{N_f} \tag{10.119}$$

A $(M_0 R)^{-N_f}$ factor combine with the determinant over nonzero modes to give the fermion contribution to the one loop beta function of the gauge coupling, leaving $R^{N_f} m^{N_f}$ multiplying rest of the instanton amplitude.

For massless fermions with $t = \frac{1}{2}$, there are 2 zero modes for each fermion field in the 1-instanton background, of the form

$$\psi_{i\alpha} = \frac{u_{i\alpha}}{(r^2+1)^{\frac{3}{2}}}, \tag{10.120}$$

where i is the $SU(2)$ doublet index and α the chiral Weyl spinor index. As we have seen, the zero mode has $\ell = 0$, and $\vec{J}_1 = \vec{S}_1 + \vec{T} = 0$, i.e. $u_{i\alpha}$ obeys

$$\left(-\frac{i}{8}\eta_{a\mu\nu}\gamma^{\mu\nu} + \frac{1}{2}\sigma_a\right) u = 0, \tag{10.121}$$

where σ_a are Pauli matrices acting on the $SU(2)$ index i and $\gamma^{\mu\nu}$ acts on the spinor index α. For convenience we may normalize $u_{i\alpha}$ by normalizing the integral of the square of the mode (10.120) to 1. Now the general amplitude involving fermions can be computed using the generating function with source $\mathcal{J}_{IJ}(x)$ (where $I, J = 1, \cdots, N_f$ are flavor indices; $\mathcal{J}_{IJ}$ acts on the Dirac spinor index as a matrix), defined by the functional integral with the action shifted by

$$S \to S + \int d^4x\, \bar{\psi}^{Ii}(x)\mathcal{J}_{IJ}(x)\psi^J{}_i(x), \tag{10.122}$$

and differentiate the result with respect to $\mathcal{J}_{IJ}$. The integral over the $2N_f$ fermion zero modes then gives the N_f dimensional determinant

$$M_0^{-N_f} \det{}_{IJ} \int d^4x \frac{\bar{u}^i \mathcal{J}_{IJ}(x) u_i}{(x^2+1)^3} \tag{10.123}$$

Since u_i is a left handed spinor, this is nonzero only when $\mathcal{J}_{IJ}(x)$ is a linear combination of $\frac{1+\gamma_5}{2}\gamma^\mu$. Under the axial $U(1)$ that assigns charge $+1$ to $\frac{1+\gamma_5}{2}\psi_i^J$ and -1 to $\frac{1-\gamma_5}{2}\psi_i^J$, such a $\mathcal{J}_{IJ}$ has charge -2. Combining with the rest of the quantum corrected one instanton contribution gives the lowest order amplitude that violates the axial $U(1)$ charge by $2N_f$.