RENORMALIZATION AND GAUGE INVARIANCE*

Gerard 't Hooft

Institute for Theoretical Physics
Utrecht University
and

Spinoza Institute
Postbox 80.195
3508 TD Utrecht, the Netherlands

e-mail: g.thooft@phys.uu.nl
internet: http://www.phys.uu.nl/~thooft/

Abstract

An overview is presented of the developments that led to our present insights in how renormalization works, from the initial skepticism to our present views of quantized field theory being an effective tool for understanding a vast domain of scales in the world of the elementary particles.

December 18, 2009

---

1 Introduction: Sweeping the infinities under the rug

When the dynamical laws of continuous systems, such as the equations for fields in a multi-dimensional world, are subject to the rules of Quantum Mechanics, one encounters divergent expressions that, at first sight, seem to invalidate the entire theory. Quantum corrections are never small, because they are described by divergent integrals. Physicists searched for alternative approaches. Perhaps one should talk about discrete particles instead of continuous fields? But then, consistent descriptions of relativistic particles requires that we allow particles to be created and annihilated. The contributions of virtual particles (particles that are annihilated very shortly after being created) also tend to divergent expressions.

It was soon suggested that the infinite parts of these effects are somehow invisible, while the finite, computable parts in some cases lead to effects that agree with experimental observations. These ideas were elaborated in a very careful manner, and apparently meaningful theories emerged. Indeed, in the case of purely electro-magnetic effects, the finite parts of the amplitudes could be calculated with impressive precision, and confrontation with experimental results demonstrated without doubt that these calculations indeed reflect the real world.

In spite of these successes, however, renormalization theory was greeted with considerable skepticism. Critics observed that "the infinities are just being swept under the rug". This obviously had to be wrong; all agreements with experimental observations, according to some, had to be accidental. Why should one believe in a theory where all one has is a perturbation expansion, which, even if one could do away with the infinities, would still not converge when summed? The search for a far superior theoretical framework was intensified.

In this contribution, however, I want to emphasize that such criticism was unjustified. In a very special class of *renormalizable* theories, it is possible, and it is correct, to remove the infinities, but one has to understand what is happening in these theories. Renormalization is a natural feature, and the fact that renormalization counter terms diverge in the ultra-violet is unavoidable.

2 The principle of renormalization

The fact that mass terms in the Lagrangian of a quantized field theory do not exactly correspond to the real masses of the physical particles it describes, and that the coupling constants do not exactly correspond to the scattering amplitudes, should not be surprising. The interactions among particles have the effect of modifying masses and coupling strengths; if we had a quantum field theory that is finite, for instance because the space-time points on which the fields are defined form a discrete lattice, then these mass and coupling "renormalizations" will be finite as well, and the emergence of such renormalizations should not be surprising at all. This means that, when masses and scattering amplitudes are measured, one still does not know exatly the *bare* mass- and coupling

parameters of the theory; what is known from experiment is the values of the masses and couplings of the real particles, which already include the renormalization effects. We call these the *renormalized* masses and coupling strengths.

In principle, observed phenomena only require finite field theories for their description anyway, since there is an energy limit to the collisions that can be studied, and if we choose a cutoff Λ, for instance by introducing a lattice with mesh size $a = 1/\Lambda$, our theories may well be accurate for all phenomena at energies $E \ll \Lambda$. In other words, limiting our experiments to an energy maximum E, implies momenta p with $|p| < E$, so the uncertainties Δx in space in general will obey $|\Delta x| > \hbar/E$.

The finite theory depends on finite values of its mass parameters and coupling strengths, but as soon as we modify our choice for the lattice length a, one finds that these bare parameters have to be modified as well. It so happens that the required choice for these bare parameters diverges in the limit $a \to 0$, or $\Lambda \to \infty$, but, if our theory is renormalizable, this divergence has no physical consequences at all.

One might suspect that this phenomenon is an artifact of the perturbation expansion, but in most cases where one can check, it persists beyond this expansion — or it might get worse! For many theories (such as Quantum Electrodynamics, QED), perturbation expansion is all that matters, in practice.

3   Calculations

For conceptual considerations, lattice field theories are very illustrative, but in practice they are rather cumbersome; there are many more sophisticated ways to introduce cut-off procedures. What happens when one does the calculations?

Essentially, lattice theories imply that there is a cut-off in momentum space. This cut-off, like the lattice itself, breaks rotational invariance. Since, for our present considerations, we are interested in more general features of such cut-offs, we could just as well replace it by a rotationally invariant one:

$$\int \mathrm{d}^3\vec{k}\,\mathrm{d}k_0 \quad \to \quad \int_{|\vec{k}|<\Lambda} \mathrm{d}^3\vec{k}\,\mathrm{d}k_0 \ , \tag{3.1}$$

or, better still, after a Wick rotation from Minkowski space-time to Euclidean space-time,

$$\int_{|k|<\Lambda} \mathrm{d}^4 k \ . \tag{3.2}$$

Still more convenient for practical calculations is the replacement of propagators by more convergent propagators:

$$\frac{1}{k^2 + m^2 - i\varepsilon} \quad \to \quad \frac{1}{(k^2 + m^2 - i\varepsilon)(1 + k^2/\Lambda^2 - i\varepsilon)^N} \ , \tag{3.3}$$

where N is a conveniently chosen power (1 or 2 usually suffices).

It is very important to check, upon completion of the calculations, that all these different cut-off procedures are *equivalent*, which means that, apart from small adjustments of the bare parameters of the theory, the physical predictions will be the same in all these schemes, in the limit where the cut-off is moved to infinity. This indeed could be checked for all cut-off procedures discussed here.
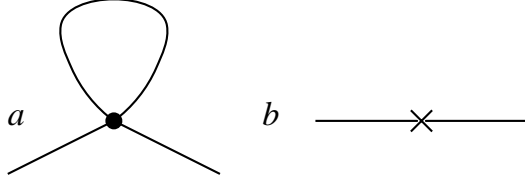

Figure 1: *a)* Diagram occurring in $\lambda\phi^4$ theory; *b)* Diagram for the counter term.

The bare parameters, $m_B(\Lambda)$, $g_B(\Lambda)$, etc., will in general diverge as $\Lambda \to \infty$. For instance take $\lambda\phi^4$ theory. The one-loop mass correction comes from the diagram of Fig. 1 *a*). In the limit $\Lambda^2 \gg m^2$, the integral is

$$
\begin{aligned}
\delta m^2 &= \frac{\lambda}{2(2\pi)^4} \int \frac{\mathrm{d}^4 k}{(k^2 + m^2 - i\varepsilon)(1 + k^2/\Lambda^2 - i\varepsilon)^2} = \\
&= \frac{\lambda}{32\pi^2} \left(\Lambda^2 + m^2 \left(1 - \log \frac{\Lambda^2}{m^2} \right) \right) \ .
\end{aligned}
\tag{3.4}
$$

The divergent part of the required counter term is

$$
\Delta m^2 = \frac{\lambda}{32\pi^2} \left(-\Lambda^2 + m^2 \log \Lambda^2 \right) \ ;
\tag{3.5}
$$

Consequently, we have to choose

$$
m_{\mathrm{Bare}}^2 = m^2 + \Delta m^2 \ .
\tag{3.6}
$$

Similarly, the coupling strength λ will require a counter term $\Delta\lambda$ due to the three diagrams of Fig. 2. Its value is

$$
\Delta\lambda = \frac{3\lambda^2}{32\pi^2} \left(\log(\Lambda^2/m^2) \right) \ .
\tag{3.7}
$$

Similar divergent counter terms are required for the higher loop corrections.

In Quantum Electrodynamics, one can replace the photon propagator analogously:

$$
\frac{\delta_{\mu\nu} - k_\mu k_\nu/(k^2 - i\varepsilon)}{k^2 - i\varepsilon} \to \frac{\delta_{\mu\nu} - k_\mu k_\nu/(k^2 - i\varepsilon)}{(k^2 - i\varepsilon)(1 + k^2/\Lambda^2 - i\varepsilon)^N} \ .
\tag{3.8}
$$

This is allowed in spite of an apparent breach of gauge invariance because the photon itself is electrically neutral. However, if one would do this for the electron propagator, gauge-invariance would be destroyed:

$$
\frac{m - i\gamma k}{(k^2 + m^2)(1 + k^2/\Lambda^2)} = \frac{m - i\gamma k}{(1 - m^2/\Lambda^2)} \left(\frac{1}{k^2 + m^2} - \frac{1}{k^2 + \Lambda^2} \right) \ .
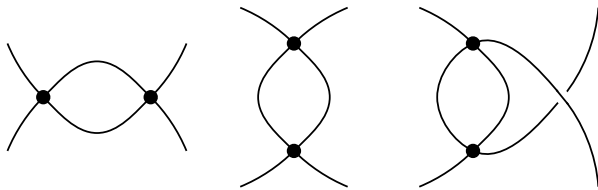\tag{3.9}
$$

Figure 2: The three diagrams contributing to $\Delta\lambda$ at the one-loop level.

This generates two fermions, ψ_1 and ψ_2; the second one, having negative metric because of the wrong sign, would couple to the photon via terms in the Lagrangian of the form (among others)

$$ie\overline{\psi}_1\gamma_\mu A_\mu \psi_2 \ . \tag{3.10}$$

This cannot be combined with the kinetic terms in the Lagrangian in a gauge-invariant way.

A gauge-invariant procedure to regularize diagrams with fermions was discovered by Pauli and Villars[1]. If a fermion occurs in a closed loop, we replace the entire loop,

$$-\mathrm{Tr}\ \big(P_m(k_{(1)})P_m(k_{(2)})\cdots P_m(k_{(\ell)})\big)\ ,\quad P_m(k)=\frac{m-i\gamma k}{k^2+m^2}\ , \tag{3.11}$$

by

$$-\sum_i e_i \mathrm{Tr}\ \big(P_{m_i}(k_{(1)})P_{m_i}(k_{(2)})\cdots P_{m_i}(k_{(\ell)})\big)\ , \tag{3.12}$$

where e_i are weights, and the index i runs from 0 to ℓ, such that

$$e_0 = 1\ ;\quad \sum_{i=0}^{\ell} e_i = 0\ ,\quad \sum_{i=0}^{\ell} e_i m_i^s = 0 \qquad (s=1,\cdots,\ell-1)\ ;$$
$$m_0 = m\ ;\quad m_{i>0} > \Lambda\ ;\quad \sum_{i=0}^{\ell} e_i\, m_i^s \log m_i\ =\ \text{finite}\ ; \tag{3.13}$$

This corresponds to a theory where several fermion species, with masses m_i and weights e_i are coupled in a gauge-invariant way to the photons. Negative e_i does not invalidate gauge-invariance. It *does* invalidate unitarity, but only at energy levels $E > \Lambda$, which is harmless in the limit $\Lambda \to \infty$. So, we have a gauge-invariant regulator.

4   Yang-Mills theory

None of these tricks work to regularize Yang-Mills theory in a gauge-invariant way. Even the lattice theory has problems because it does not respect chiral invariance. In the continuum, the Lagrangian is

$$\mathcal{L}\ =\ -\tfrac{1}{4}G_{\mu\nu}^a G_{\mu\nu}^a - \overline{\psi}(\gamma D + m)\psi\ ;$$

4

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + g\, f^{abc} A^b_\mu A^c_\nu \;;$$
$$D^{ij}_\mu = \delta^{ij}\partial_\mu + ig\, T^{ij}_a A^a_\mu \;; \quad [T_a,\, T_b] = -if^{abc}T_c \;. \tag{4.1}$$

How do we renormalize the Yang-Mills theory? One of the first simplifications that were tried was to add a mass term, $-\frac{1}{2}M^2(A^a_\mu)^2$ to the Lagrangian. This is not gauge-invariant, but so what? Gauge-invariance may be expected to be restored in the limit $M \to 0$. Also, renormalization counter terms refer to the situation where the energies go to infinity. Surely, these mass terms should become insignificant in the high-energy limit? This, as it turned out later, was a mistake. The mass terms correspond to new polarizations of the vector particle, which in the massless case should be removed by gauge-fixing. In the massive theory, however, gauge-fixing would clash with unitarity. Therefore, even in the high-energy limit of a massive theory, gauge fixing would be illegal; consequently, the gauge degrees of freedom, which would be physically observable for the massive theory, would fluctuate so much at high momentum that they render this theory non-renormalizable, in general.

5 Unitarity

To investigate the situation, one must first understand the restrictions that we wish to impose on our theories. These are: *Unitarity* and *causality*[1]. Since the perturbation expansion for a field theory can be reformulated in terms of Feynman diagrams, it is extremely useful to discuss these notions in terms of properties of these diagrams. A second reason for looking at these diagrams is that the origin of the divergences in the loop corrections is easiest to trace in these diagrams, so one can read off what modifications are necessary to regularize these expressions while still obeying the requirements.

Diagrams consist of *propagators* connected by *vertices*. The amplitudes associated to the diagrams can either be computed in coordinate space or in terms of their Fourier transforms, *i.e.*, momentum space. We use x, y, z, for coordinates and p, q, k, $\cdots$ for momentum space, all being 4-vectors. In coordinate space, the propagator for a scalar field with mass m is

$$\Delta^F(x) = -i \int \mathrm{d}^4k \frac{e^{ik\cdot x}}{k^2 + m^2 - i\varepsilon} \;; \tag{5.1}$$

One introduces the *on-shell* propagator $\Delta^{\pm}$ as follows:[2]

$$\Delta^{\pm}(x) = 2\pi \int \mathrm{d}^4k \; e^{ik\cdot x}\, \delta(k^2 + m^2)\, \theta(\pm k^0) \;, \tag{5.2}$$

and subsequently we fully employ the useful property

$$\Delta^F(x) = \theta(x^0)\Delta^+(x) + \theta(-x^0)\Delta^-(x) \;. \tag{5.3}$$

---

[1]Causality can be linked to so-called "dispersion relations" that must be obeyed by the S-matrix elements of the theory. They are closely related to the unitarity requirement and will not be further discussed here.
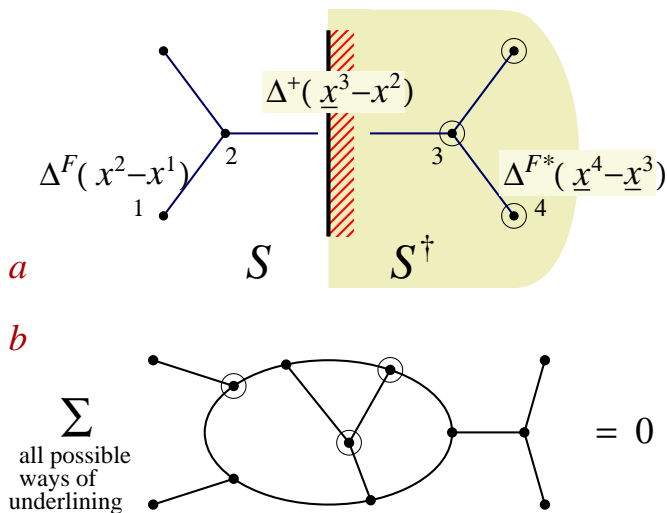
Figure 3: a) A diagram contributing to the unitarity condition. b) Veltman's largest time equation. Underlined vertices (indicated by circles) are the vertices that occur in the diagrams for $S^\dagger$. The sum is only over all diagrams that have the same topology, apart from the underlinings.

Unitarity corresponds to the relation $\sum_n S|n\rangle\langle n|S^\dagger = \mathbb{I}$. If we have the diagrams contributing to the scattering matrix S, then the diagrams for establishing this relation are sketched in Fig. 3a. Using Eq. (5.3), one arrives at the so-called *largest time equation*[2], which is a property of diagrams sketched in Fig. 3b). One uses this to show that all interaction parts cancel out when checking the unitarity relation

$$S\,S^\dagger = \mathbb{I} \,. \tag{5.4}$$

Unitarity then automatically follows for a scalar field theory generated by a real Lagrangian, provided regularization of the infinities is done by modifications of the prapagator that are such that Eqs. (5.3) are kept intact. The latter is indeed the case for the propagator modifications (3.3), (3.8), (3.9), and (3.12), but we do see that the new terms correspond to particles with masses Λ or m_i that might have (components with) *negative metric*: their contribution in $|n\rangle\langle n|$ is negative, so we have states with negative norm. The only way to ensure that unitarity survives is to choose Λ larger than the largest energies considered in S. In that case, simply energy conservation ensures the absence of these particles in the intermediate states $|n\rangle$. In the limit $\Lambda \to \infty$ then unitarity in terms of the physical particles only, is completely restored.

The propagators for particles with spin $\frac{1}{2}$ or 1 may be written as scalar propagators *with extra polynomials of k_μ in the numerator*. In that case, they also obey Eq. (5.3), because the factors k_μ turn into space-time differentials ∂_μ. The propagators $\Delta^\pm$ then have these polynomials on mass shell. Unitarity will then require that the residues of these polynomials on mass shell be non-negative. This is the case for the numerator $g_{\mu\nu} + k_\mu k_\nu/M^2$ in the case of spin 1. It also holds for the numerators $\gamma_4(m - i\gamma k)$ for spin

$\frac{1}{2}$ particles[2], while for their antiparticles the overall sign switches (due to the presence of γ_4), which explains the need for a minus sign for every closed fermion loop. Thus, we see how unitarity can be proven in theories with these propagators, assuming that regularization was carried out with heavy mass terms only. Clearly, unitarity requires the propagator

$$\frac{\delta_{\mu\nu} + k_\mu k_\nu/M^2}{k^2 + M^2 - i\varepsilon} \tag{5.5}$$

for massive vector particles.

The problem with the massive vector propagator (5.5) is that the $k_\mu k_\nu$ term diverges badly at high energies. In a renormalizable theory, we wish to replace it by

$$\frac{\delta_{\mu\nu} - k_\mu k_\nu/(k^2 - i\varepsilon)}{k^2 + M^2 - i\varepsilon} \ , \tag{5.6}$$

because it would be the same on mass shell, and off mass shell it would converge much better. However, here the second part generates new massless poles, *with the wrong sign!* An illegal massless particle with negative norm states would invalidate the unitarity relation (5.4).

6   Ghosts

Particles that can be added to the Feynman rules in order to replace the divergent $k_\mu k_\nu$ terms of the vector propagator, without modifying the scattering matrix, and that seem to spoil the unitarity relation in a renormalizable theory, are called "ghosts". They were first investigated by Feynman[3], Veltman[4] and DeWitt[5]. For the massive Yang-Mills theory, Feynman and Veltman found a single, non-oriented (real) scalar ghost with negative metric. It replaces the divergent $k_\mu k_\nu/M^2$ term in all one-loop diagrams, see Fig. 4. Unfortunately, they could not find any similar properties for diagrams with two or more overlapping loops. The relations that were found [4] indicated that renormalizability was lost. Today, we know why this is so. In the absence of the scalar Higgs particle, Massive Yang-Mills theory does not respect gauge invariance exactly. The gauge ghost particles decouple approximately, but not entirely in the ultraviolet limit, so that their uncontrolled fluctuations in that limit would cause integrals to diverge there, beyond repair. Having approximate gauge-invariance in the UV limit is not good enough!

On the contrary, a completely gauge-invariant Yang-Mills theory where the vector particles do have a mass *does* exist: it is the Brout-Englert[6]-Higgs[7] theory. Only that theory can serve as a renormalizable theory for vector particles.

This theory features two species of ghost particles. One is the so-called Faddeev-Popov[8]-DeWitt[5] ghost particle. Its Feynman rules are much like those of the ghost that had been found by Feynman and Veltman, including the minus sign for each closed

---

[2]We wrote the metric tensor $g_{\mu\nu}$ rather than the Euclidean $\delta_{\mu\nu}$ because unitarity is considered in Minkowski space-time, and because signs, as well as complex conjugation, are crucial in these arguments.

Figure 4: Relation for all one-loop insertions found by Feynman and Veltman. The blobs represent only tree diagrams, and the external lines must be physical particles on mass shell.

loop, except that it is described by a complex rather than a real field, and because of this, a closed FP loop receives a factor -1 rather than $-\frac{1}{2}$ as in Fig. 4. The other ghost particle is the transverse isospin polarization of the Higgs particle. It happens to feature the same quantum numbers as the FP particle, and it obeys exactly the same Feynman rules, except, here, we have a real particle, and it has a positive metric. Thus, we now find the explanation as to why Feynman and Veltman found a real-valued ghost with negative metric:

$$-1 + \tfrac{1}{2} = -\tfrac{1}{2} \; . \tag{6.1}$$

If now the Higgs mass is sent to infinity, no Higgs contribution is seen in the one-loop diagrams of the vector particles, and so, renormalizability happened to survive at the one-loop level.

7 Gauge-invariant regularization

We saw that the Brout-Englert-Higgs theory is formally renormalizable, since unitary Feynman rules are formally equivalent to renormalizable ones. This, however, should be verified explicitly as soon as cut-off procedures are used that actually break gauge-invariance at an intermediate stage. Will gauge-invariance be restored in the limit where the cut-off parameter Λ is sent to infinity, in spite of the gauge-dependence of the procedure? This is not obvious, unless one could demonstrate at least the existence of a

procedure that does respect gauge-invariance. The road towards finding such a procedure contains many dead alleys that all had to be explored.

First, it was noted that, at least the fermions in the theory can be regularized. We simply employ the same method as the one used by Pauli and Villars[1], see Eqs. (3.12) and (3.13). This implies that the internal fermion lines cannot cause any problems[3]. However, the Yang-Mills bosons also couple to themselves. How do we regularize a diagram containing loops of vector lines? Our next step was now as follows.[9] Observe that Yang-Mills theory is also gauge-invariant in n dimensions, where n can be any integer bigger (or smaller) than 4. Let us replace the coordinates x^μ by multi-dimensional vectors, $(x^{0,\cdots,3}, x^5, x^6, \cdots)$. Then the vector fields A_μ are also n-component vectors, $(A_\mu^a, A_5^a, A_6^a, \cdots)$, and when we sum over its components we have to remember that $\sum_\mu \delta_\mu^\mu \to n$. Consider now all diagrams with exactly one closed loop of vector particles. Let us take $n = 5$. It turned out that all properties of the diagrams that we need, such as the largest-time equation, Fig. 3 b) and the Slavnov-Taylor identities (see Section 8), continue to hold if we keep the 5th component of the momentum k running around in the loop, fixed[9]. We take the contributions where $k_5 = m_i$, at weight e_i, and treat these diagrams as regulator diagrams just as in the Pauli-Villars method, Eqs. (3.12) and (3.13).

To prove that this procedure *is* gauge-invariant is not trivial, but it turns out that nowhere one actually needs to integrate over the 5th component of the momentum. This is why we can keep k_5 fixed and use diagrams with different, large values of k_5 as regulators. However, only in one-loop diagrams (or diagrams with non-overlapping loops), one can keep one single value for k_5. In diagrams with overlapping loops, gauge-invariance does link internal lines with momentum k_5 to other diagrams where the fifth component is $2k_5$, $3k_5$, etc. Thus, for multi-loop diagrams, a more clever technique was called for. We tried to go to 6, 7, or more dimensions, but in vain[4].

The transition to a non-integral number of dimensions[10] was, in a sense, the last straw. If you have only ε extra dimensions, with ε very close to zero, then the amplitudes should hardly change. Yet, the expressions for integrals in divergent diagrams do change dramatically. The general formula for a typical integral in a one-loop diagram in n dimensions (for simplicity in Euclidean space) is

$$\int \mathrm{d}^n k \frac{(k^2)^p}{(k^2 + m^2)^q} = \pi^{n/2} m^{2p+n-2q} \frac{\Gamma(q - p - \frac{n}{2})\Gamma(p + \frac{n}{2})}{\Gamma(q)\Gamma(\frac{n}{2})} , \qquad (7.1)$$

where $\Gamma(x)$ is Euler's gamma function. This integral actually only converges when the entries of all gamma functions are positive. If $p \geq q + n/2$, the integral is divergent, hence meaningless. But the gamma function is only infinite if this quantity is exactly integral, see Fig. 5; any tiny deviation from that would make the expression finite. Could this not be used?

Indeed, this had been used. If the propagator is replaced by a slightly different power

---

[3]unless *chirality* is an issue, see the last paragraph of this section.

[4]Nevertheless, one author wrote a book where the use of finite extra dimensions was re-invented, but without valid proofs; we claim that this cannot be done.
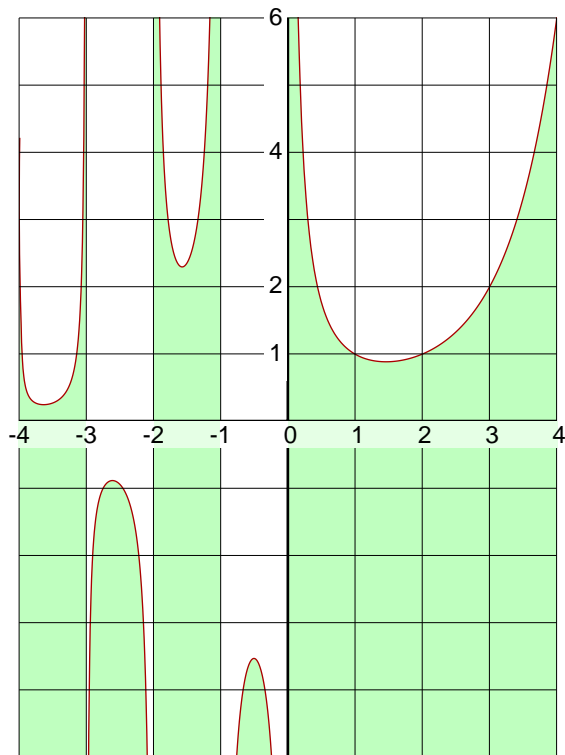
Figure 5: Euler's Γ-function.

of $k^2 + m^2$, say, $q = 1 + \varepsilon$, the expression would be finite. This, however, could be seen to violate gauge-invariance, an essential property of the theory. Strangely enough, the only modification of the theory that did not appear to violate any such basic principle, is a replacement of the number of dimensions, n, by a non-integer, such as $4 - \varepsilon$. In particular, the theory would retain gauge-invariance and unitarity, if all divergent integrals were evaluated using the gamma functions of Eq. (7.1). Now, when the limit $n \to 4$ is taken, the poles of the gamma functions are hit (see Fig. 5), reflecting the original divergence. All that needed to be done now is allow these poles to be cancelled by allowing the 'bare parameters' such as coupling strengths and masses to have poles of the form $(n-4)^{-L}$ that cancel these effects, where L is the number of overlapping loops. It is exactly the renormalizable theories where such procedures work. We now had a gauge-invariant regulator scheme that would work flawlessly for diagrams with any number of overlapping loops in them. It was called *dimensional regularization*, to be followed by *dimensional renormalization*.

One barrier still remained: in a theory with chiral symmetry, the completely anti-symmetric tensor $\varepsilon_{\mu\nu\kappa\lambda}$ would occur; for instance, this tensor appears in the definition of the Dirac matrix γ_5. It can only be used when the number of dimensions is exactly 4. Thus, chiral symmetry would be violated by dimensional renormalization. This actually was unavoidable; we knew that chiral symmetry in gauge theories is offset by an anomaly, called the 'Adler-Bell-Jackiw anomaly'.[11] Some theories do have chirality but the Adler-Bell-Jackiw anomaly happens to cancel out. To regularize such theories, special

10

tricks had to be devised. This was important to do; the Standard Model belongs to this class.

8 Proofs

To verify that the required counter terms could all be provided by simply absorbing the poles in the bare parameters of the theory required mathematical proofs. The most natural way to provide such a proof would be to associate the required mathematical properties with a *symmetry*. In the early days, however, no such symmetry could be identified. Symmetry groups are characterized by generators obeying commutation rules, but here, we were dealing with Faddeev Popov ghosts obeying *anti*-commutator rules. It did not cross our minds that symmetries might exist between fermionic and bosonic fields, and that these symmetries indeed are controlled by anticommutators. So, we derived identities relating different diagrams that only later would be recognized as expressions of Becchi-Rouet-Stora-Tyutin symmetry[12].
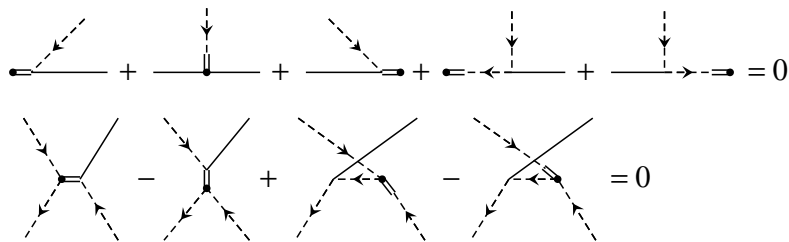

Figure 6: Top: identity between propagators and vertices; the double line refers to an auxiliary vertex as described in Ref [9]. Below: identity between ghost diagrams requiring the Jacobi identity (8.1).

This is why we used the only method known at the time: combinatorial relations between Feynman diagrams. Due to gauge-invariance, the vertices and the propagators are related. Inside a diagram, we could use identities such as the ones depicted in Fig. 6. Repeating this at all ghost vertices, one ends up at spots where ghosts couple to ghosts. There, the Jacobi identity

$$f_{aib}f_{bjk} + f_{bjb}f_{bki} + f_{akb}f_{bij} = 0 \; , \tag{8.1}$$

is needed, to show that the diagrams obey a special simple relation. Here, f_{abc} are the structure constants of the local gauge group (see Eq. (4.1)). The relation we found was generalized by Slavnov[13] and Taylor[14] for diagrams with external lines off mass shell. these relations are now called Slavnov-Taylor identities.

At the time, we needed only our own, on mass shell identities, to prove that the renormalized theory is gauge-invariant and unitary, and we avoided off-mass-shell expressions because they would require new loop diagrams to be renormalized, which would complicate matters. However, when regarding these identities as symptoms of a *symmetry*,

the off-shell expressions are useful. It is these identities that inspired Becchi, Rouet and Stora[12] to describe their symmetry.

9   BRST symmetry

Today, we no longer need to perform manipulations relating diagrams. We can take our refuge in symmetry considerations, which indeed are time saving and more fool proof. What BRST discovered was that, *after* fixing the gauge constraint, local gauge invariance is replaced by a global symmetry. The reason why it was not found earlier is that it is a *super*-symmetry, requiring anticommuting generators rather than commuting ones, since it is the fermionic ghost field that is transformed into the bosonic field variables and back. If we write the action as

$$
S \;=\; \int \mathrm{d}^4 x \mathcal{L}(x) \;, \quad \text{where} \tag{9.1}
$$

$$
\mathcal{L} \;=\; \mathcal{L}^{\mathrm{inv}} + \lambda^a(x) C^a(A,x) + \overline{\eta}^a(x) \frac{\partial C^a(x)}{\partial \Lambda^b(x')} \eta^b(x') \;, \tag{9.2}
$$

then the Lagrange multiplier field $\lambda^a(x)$ forces the gauge condition $C^a(x) = 0$, where C^a depends on the fields A in such a way that gauge transformations can transform this quantity to zero. The last term contains the ghost fields $\eta(x)$ and its conjugate $\overline{\eta}(x)$, which are required to interact as dictated by the way $C^a(x)$ depends on an infinitesimal gauge transformation generated by the gauge generator $\Lambda^b(x')$. BRST found that this lagrangian is invariant under the following (infinitesimal) super-transformation:

$$
\begin{aligned}
\delta A^a(x) &= \overline{\varepsilon}\, \frac{\partial A^a(x)}{\partial \Lambda^b(x')}\, \eta^b(x') \;; \\
\delta \eta^a(x) &= \tfrac{1}{2}\overline{\varepsilon}\, f^{abc}\eta^b(x)\eta^c(x) \;; \\
\delta \overline{\eta}^a(x) &= -\overline{\varepsilon}\, \lambda^a(x) \;; \\
\delta \lambda^a(x) &= 0 \;.
\end{aligned} \tag{9.3}
$$

Here, $\overline{\varepsilon}$ is the (anticommuting) generator of the infinitesimal super-transformation. To prove this invariance, the same Jacobi identity (8.1) is needed, and the relations among the diagrams that one derives from this invariance are exactly the ones we derived by our combinatorial methods. These relations are needed in order to prove that the required renormalization counter terms can all be attributed to mass- and coupling constant renormalizations, and/or field rescalings. The theory we obtained provides for a unitary scattering matrix.

10   Conclusion

The choice of Yang-Mills fields to describe the weak force was inspired by experimental observations. They forced us to investigate the most general field theories that still obey

the requirement of renormalizability. The original skepticism appeared to be justified at first, but gradually gave way to enthusiasm. Infinite counter terms should not scare us; when treated with caution, we not only obtain theories with a high predictive power, but we can also understand how to interpret them. There is no question of sweeping any of the difficulties under the rug. We do encounter the well-known Adler-Bell-Jackiw anomalies, which can only be admitted if they are arranged to cancel out for all currents to which gauge fields are coupled.

The original results and proofs appeared to be quite technical, and were difficult to grasp by many, but later much more elegant formalisms were invented, and today, gauge theories are routine. Today, quantized string theories appear to be technically complicated and difficult to interpret. Their fate in the near future could be similar to what happened to the gauge theories.

References

[1] W. Pauli and F. Villars, Rev. Mod. Phys. **21** (1949) 434.

[2] G. 't Hooft and M. Veltman, "DIAGRAMMAR", CERN Report 73/9 (1973), reprinted in "Particle Interactions at Very High Energies", Nato Adv. Study Inst. Series, Sect. B, vol. 4b, p. 17.

[3] R.P. Feynman, *Acta Phys. Polonica* **24** (1963) 697.

[4] M. Veltman, Physica **29** (1963) 186, id., *Nucl. Phys.* **B7** (1968) 637; J. Reiff and M. Veltman, Nucl. Phys. **B13** (1969) 545; M. Veltman, *Nucl. Phys.* **B21** (1970) 288.

[5] B.S. DeWitt, *Phys. Rev. Lett.* **12** (1964) 742.

[6] F. Englert and R. Brout, *Phys. Rev. Lett.* **13** (1964) 321.

[7] P.W. Higgs, *Phys. Lett.* **12** (1964) 132; *Phys. Rev.* Lett. **13** (1964) 508; *Phys. Rev.* **145** (1966) 1156.

[8] L.D. Faddeev and V.N. Popov, *Phys. Lett.* **25B** (1967) 29.

[9] G. 't Hooft, *Nucl. Phys.* **B33** (1971) 173.

[10] G. 't Hooft and M. Veltman, *Nucl. Phys.* **B44** (1972) 189; C.G. Bollini and J.J. Giambiagi, *Phys. Lett.* **40B** (1972) 566; J.F. Ashmore, Nouvo Cimento Letters **4** (1972) 289.

[11] S.L. Adler, *Phys. Rev.* **177** (1969) 2426; J.S. Bell and R. Jackiw, *Nuovo Cim.* **60A** (1969) 47; S.L. Adler and W.A. Bardeen, *Phys. Rev.* **182** (1969) 1517; W.A. Bardeen, *Phys. Rev.* **184** (1969) 1848.

[12] C. Becchi, A. Rouet and R. Stora, *Commun. Math. Phys.* **42** (1975) 127; id., *Ann. Phys.* (N.Y.) **98** (1976) 287; I.V. Tyutin, Lebedev Prepr. FIAN 39 (1975), unpubl.

[13] A. Slavnov, Theor. Math. Phys. **10** (1972) 153 (in Russian), Theor. Math. Phys. **10** (1972) 99 (Engl. Transl.)

[14] J.C. Taylor, *Nucl. Phys.* **B33** (1971) 436.