Aspects Of Perturbative Unitarity

Damiano Anselmi

Dipartimento di Fisica "Enrico Fermi", Università di Pisa,
Largo B. Pontecorvo 3, 56127 Pisa, Italy
and INFN, Sezione di Pisa,
Largo B. Pontecorvo 3, 56127 Pisa, Italy

damiano.anselmi@unipi.it

Abstract

We reconsider perturbative unitarity in quantum field theory and upgrade several arguments and results. The minimum assumptions that lead to the largest time equation, the cutting equations and the unitarity equation are identified. Using this knowledge and a special gauge, we give a new, simpler proof of perturbative unitarity in gauge theories and generalize it to quantum gravity, in four and higher dimensions. The special gauge interpolates between the Feynman gauge and the Coulomb gauge without double poles. When the Coulomb limit is approached, the unphysical particles drop out of the cuts and the cutting equations are consistently projected onto the physical subspace. The proof does not extend to nonlocal quantum field theories of gauge fields and gravity, whose unitarity remains uncertain.

1 Introduction

The problem of quantum gravity is the apparent incompatibility between unitarity and renormalizability. For example, the quantization of Einstein gravity gives a theory that is unitary, but not renormalizable [1, 2]. If the counterterms generated by renormalization are included, the theory becomes renormalizable with infinitely many independent couplings[1] and predictive at low energies (see, for example, [3]). It is also possible to build theories of quantum gravity that are renormalizable (with finitely many independent couplings), but not unitary. One way to achieve this goal is by including higher-derivative terms that make the propagators fall off more rapidly at high energies [4]. It is not known how to build a theory that is renormalizable and unitary at the same time.

Perturbative unitarity is thus a key issue in quantum field theory. In scalar and fermion theories it can be proved by means of the cutting equations [5, 6]. In gauge theories, additional aspects need to be addressed, such as the compensation between the Faddeev-Popov ghosts and the temporal and longitudinal components of the gauge fields. This compensation can be proved diagrammatically [7] by means of the Ward identities or more formally at the level of the Fock space [8]. For a variety of reasons, we believe that the last word has not been said on this topic and that an attempt to reorganize and generalize the proof is most welcome. First, a treatment of perturbative unitarity in quantum gravity is still missing. Second, the existing proofs in gauge theories are involved, which suggests that they are not optimized.

In this paper we offer a more economic approach and a new, exhaustive proof that works not only in Abelian and non-Abelian gauge theories, but also in quantum gravity and a variety of other nonrenormalizable local theories, in arbitrary dimensions d greater than 3.

First we prove a number of basic tools, such as the largest time equation and the cutting equations, paying attention to the minimum assumptions that they require. Then, we show that the unphysical degrees of freedom can be consistently dropped from the cuts. To achieve this goal, we identify a *special gauge* that leads to the unitarity equation in a straightforward way.

In gauge theories and gravity, several common gauges have inconvenient features. The Lorenz gauge, for example, gives propagators that have double poles, which prevents the derivation of the cutting equations. The Coulomb gauge has a nice feature, because it just propagates the physical degrees of freedom. However, it introduces unwanted singularities in the Feynman diagrams, which are under control only in QED.

The special gauge is a new gauge that has no double poles, interpolates between the Feynman gauge and the Coulomb gauge and satisfies all the assumptions that are required to derive the cutting equations. Moreover, when the gauge fields are given a mass to regulate their on shell

[1]For this reason, "nonrenormalizable" and "renormalizable with infinitely many independent couplings" are often used interchangeably.

infrared divergences and the Coulomb limit is approached, the threshold for the production of unphysical particles grows enough to drop them out from the cuts. After that, a few technical tricks allow us to complete the proof. The special gauge is unique in both Yang-Mills theory and gravity.

An even simpler proof of perturbative unitarity is available in QED, where it is possible to work directly in the Coulomb gauge.

We pay attention to details such as the regularization and the renormalization of the cutting equations, the presence of contact terms, the double poles of the gauge field propagators, the orders of the limits with which various parameters are removed, the infrared divergences and other singularities that disappear by summing up the cut diagrams. Some of these problems are not treated carefully (or are not even mentioned) in the existing literature.

Because of the key role played by the largest time equation, the proofs we give in this paper do not generalize to nonlocal quantum field theories of gauge fields and gravity, including those whose propagators have no poles on the complex plane besides the graviton one [9]. For this reason, the consistency of those theories remains unclear.

The paper is organized as follows. In section 2 we derive the cutting equations under the minimum assumptions. In section 3 we derive the unitarity equation. In section 4 we introduce the special gauge in Abelian and non-Abelian gauge theories. In section 5 we give the simplest proof of perturbative unitarity, using the Coulomb gauge in QED. In section 6 we use the special gauge to prove unitarity in all gauge theories. In section 7 we generalize the special gauge and the proof of unitarity to quantum gravity. Section 8 contains the conclusions.

2 The cutting equations

We investigate perturbative unitarity following the guidelines of refs. [6, 10, 11], which consist of proving, in the order:

— the largest time equation,

— the cutting equations,

— the pseudounitarity equation,

— the unitarity equation.

The pseudounitarity equation is a more general version of the unitarity equation, where the cuts may propagate both physical and unphysical particles. In gauge theories and gravity it is helpful to first derive the pseudounitarity equation and then prove that it implies the unitarity equation, by showing that the external legs and the cuts can be consistently projected onto the physical subspace.

In this section we reconsider the cutting equations and search for the minimum assumptions that are necessary to derive them. We assume invariance under translations and spatial rotations, but we do not assume Lorentz invariance. Indeed, we need results that can be applied to gauge choices that violate Lorentz invariance, such as the Coulomb gauge and the special gauge. We do not assume from the beginning that the theory is local. Nevertheless, along with the derivation it emerges that the vertices must be local.

2.1 Regularization

We use the dimensional regularization, or one of its variants [12], directly in Minkowski spacetime. Several assumptions and arguments of our derivations make no sense unless there are exactly one time component and one energy component, so we dimensionally continue the space coordinates, but do not continue the time coordinate.

Let d denote the physical spacetime dimension and $d-\varepsilon$ the continued one, where ε is a complex number. Split the continued spacetime $\mathbb{R}^{d-\varepsilon}$ into the product of the time line $\mathbb{R}$ times the continued space $\mathbb{R}^{d-1-\varepsilon}$. Denote the metric of flat spacetime with $\eta^{\mu\nu}=$diag$(1,-1,-1,\cdots,-1)$.

Typically, we work with theories whose vertices are local and whose gauge fixed propagators $\tilde{f}(E,\mathbf{p})$ are equal to ratios of polynomials $u(E,\mathbf{p})$ and $v(E,\mathbf{p})$ of the energy E and the space momentum $\mathbf{p}$,

$$\tilde{f}(E,\mathbf{p})=\frac{u(E,\mathbf{p})}{v(E,\mathbf{p})}, \tag{2.1}$$

with denominators $v(E,\mathbf{p})$ equal to products of polynomials $aE^2-b\mathbf{p}^2-m^2+i\epsilon$, where a and b are positive constants and m is real. The symbol ϵ is used to specify the contour prescription.

These theories are well regularized by the prescription of first integrating on the space momenta $\mathbf{p}$, then on the energies E. Indeed, after the integration on the space momenta the energy integrals behave as

$$\sim\int^{E=\pm\infty}\mathrm{d}E\,\frac{E^m}{(E^2)^{n\varepsilon/2}} \tag{2.2}$$

for large $|E|$, where n and m are nonnegative integers and $n\neq 0$. The analytic continuation in ε makes these integrals well defined.

Various manipulations can simplify the propagators and generate local integrands. Then the result is zero, because the dimensionally regularized $\mathbf{p}$ integral vanishes, as in

$$\int_{-\infty}^{+\infty}\frac{\mathrm{d}E}{2\pi}\int\frac{\mathrm{d}^{d-1-\varepsilon}\mathbf{p}}{(2\pi)^{d-1-\varepsilon}}E^m p_{i_1}\cdots p_{i_n}=0, \tag{2.3}$$

where, again, n and m are nonnegative integers.

Note that we may not be able to perform the usual contour integrations on the energy. Nonetheless, each step of the calculation is consistent. For example, in $d=4$, we have

$$\begin{aligned}\int_{-\infty}^{+\infty}\frac{\mathrm{d}E}{2\pi}\int\frac{\mathrm{d}^{3-\varepsilon}\mathbf{p}}{(2\pi)^{3-\varepsilon}}\frac{i}{E^2-\mathbf{p}^2-m^2+i\epsilon}&=-\frac{i\Gamma\left(\frac{\varepsilon-1}{2}\right)}{(4\pi)^{(3-\varepsilon)/2}}\int_{-\infty}^{+\infty}\frac{\mathrm{d}E}{2\pi}(m^2-E^2-i\epsilon)^{(1-\varepsilon)/2}\\&=\frac{\Gamma\left(\frac{\varepsilon}{2}-1\right)}{(4\pi)^{(2-\varepsilon)/2}}(m^2-i\epsilon)^{1-\frac{\varepsilon}{2}}.\end{aligned} \tag{2.4}$$

Interchanging the energy and momentum integrals does not make sense, in general, as in (2.3), but in specific cases it may be allowed, as in (2.4).

The propagators have the structure (2.1) in all the cases we consider, with two exceptions: the Coulomb gauge in QED and the mass terms introduced to regulate the (on shell) infrared divergences in non-Abelian gauge theories and nonrenormalizable theories. In both cases some denominators have $a = 0$, but the regularization can be proved to work well by means of *ad hoc* methods and/or appropriate truncations.

Equipped with this regularization technique, we are ready to begin our investigation. The algorithm to renormalize the divergences is described along the way.

2.2 The largest time equation

The largest time equation is implied by the following minimum assumptions:

(a) the vertices are localized in time;

(b) the propagators $f(x)$ in coordinate space can be decomposed as

$$f(x) = \theta(x^0)g_+(x) + \theta(-x^0)g_-(x) \tag{2.5}$$

in the sense of distributions.

For the moment, we do not make further assumptions about the distributions $g_\pm(x)$. Formula (2.5) and similar formulas written below are exact identities among distributions. In particular, they imply that $f(x)$ contains no contributions proportional to $\delta(x^0)$ or its derivatives.

By assumption (a), each vertex is associated with a definite time x^0, but need not be associated with a unique space coordinate $\mathbf{x}$. By translational invariance, a propagator is described by a time difference $x^0 - y^0$ and a space difference $\mathbf{x} - \mathbf{y}$, as usual.

Consider a raw Feynman diagram in coordinate space. By this we mean the plain product of the vertices and the propagators, with no integrations over the space and time coordinates. We denote the raw diagram by $F(x_1^0, \cdots, x_n^0)$, where x_i^0 are the locations of the vertices in time, while the dependences on the space coordinates are omitted.

Next, build variants F_M of the diagram F as follows. Mark any subset of vertices by putting hats on their times x_i^0. Multiply by an overall factor $(-1)^s$, where s is the number of marked vertices. Replace the propagators connecting two unmarked vertices, two marked vertices and a marked vertex with an unmarked one, respectively, as specified by the following scheme:

$$\begin{array}{ll}
x \longrightarrow y: & \theta(x^0 - y^0)g_+(x-y) + \theta(y^0 - x^0)g_-(x-y), \\
\hat{x} \longrightarrow \hat{y}: & \theta(x^0 - y^0)g_-(x-y) + \theta(y^0 - x^0)g_+(x-y), \\
\hat{x} \longrightarrow y: & g_+(x-y), \qquad\qquad x \longrightarrow \hat{y}: \quad g_-(x-y).
\end{array} \tag{2.6}$$

Finally, do not modify the values of the vertices. For the sake of generality we assume that the propagators are oriented. The orientation is specified by the arrows.

Now, assume that the vertices have distinct times. Then, we have the identity

$$\sum_{\text{markings } M} F_M(x_1^0, \cdots, \hat{x}_i^0, \cdots, \hat{x}_j^0, \cdots, x_n^0) = 0, \tag{2.7}$$

which is known as the largest time equation. The sum is over all the ways to mark the vertices, including the cases where the vertices are all marked and all unmarked.

Here is the proof of (2.7). Since the vertices have distinct times, one vertex must have the largest time. Denote that time by z^0. Pick any diagram $F_{\bar{M}}$ of the sum (2.7). The time z^0 may be marked or not in $F_{\bar{M}}$. If it is marked (unmarked), the sum (2.7) contains another diagram $F'_{\bar{M}}$ that is identical to $F_{\bar{M}}$ except for the fact that z^0 is unmarked (marked). The sum $F_{\bar{M}} + F'_{\bar{M}}$ vanishes, because the propagators between a point[2] $z = (z^0, \mathbf{z})$ and any other points x, y are, in the various cases,

$$z \longrightarrow y: \; g_+(z-y), \qquad \hat{z} \longrightarrow y: \; g_+(z-y), \qquad z \longrightarrow \hat{y}: \; g_-(z-y), \qquad \hat{z} \longrightarrow \hat{y}: \; g_-(z-y),$$
$$x \longrightarrow z: \; g_-(x-z), \qquad x \longrightarrow \hat{z}: \; g_-(x-z), \qquad \hat{x} \longrightarrow z: \; g_+(x-z), \qquad \hat{x} \longrightarrow \hat{z}: \; g_+(x-z).$$

In the end, the diagrams $F_{\bar{M}}$ and $F'_{\bar{M}}$ are equal except for an overall minus sign due to the marking/unmarking of z^0. This implies (2.7).

2.3 Contact terms

To derive the cutting equations, we must calculate the Fourier transforms of the largest time equations, which demands to integrate on the coordinates. However, in the derivation of (2.7) we have assumed that the vertices had different times. We want to make sure that this assumption can be dropped, because only in that case the result of the Fourier transform has a straightforward diagrammatic interpretation.

More precisely, we need to show that when we take any (one-sided) limits of coinciding times on the functions F_M of equation (2.7), we do not miss terms that give nontrivial contributions to the integrals on the coordinates.

Call two vertices nearest neighbors if they are connected by a propagator. Observe that, to prove (2.7), the point z of largest time just needs to be compared with its nearest neighbors. For this reason, equation (2.7) trivially extends to the case where there are vertices with coinciding times, as long as no pairs of them are made of nearest neighbors. Precisely, denote the vertices with coinciding times by w_i and call their time w^0. When w^0 is not the largest time, we can proceed exactly as above, which leads to (2.7). When w^0 is the largest time, we can pick any of the w_i as the vertex z and, again, proceed as above to obtain (2.7). Thus, the only situation that deserves attention is when some nearest neighbors have coinciding times. Nontrivial contributions to the integrals on the coordinates can only appear when contact terms are present.

[2]There maybe more than one point with time z^0, if the vertex is nonlocal in space.

Consider that the vertices may carry time derivatives. For example, in quantum gravity the Einstein-Hilbert action is corrected by terms built with the Riemann tensor and its derivatives, which may contain an arbitrary number of time derivatives acting on the metric tensor. By means of partial integrations, the derivatives can be moved to the propagators (2.6). Then, they may generate contact terms proportional to $\delta(x^0)$ or its derivatives:

$$\partial_0^n f(x) = \theta(x^0)\partial_0^n g_+(x) + \theta(-x^0)\partial_0^n g_-(x) + \sum_{k=1}^{n} \delta^{(k-1)}(x^0) \left[\lim_{x^0 \to 0^+} \partial_0^{n-k} g_+(x) - \lim_{x^0 \to 0^-} \partial_0^{n-k} g_-(x) \right]. \tag{2.8}$$

However, the largest time equation (2.7) is only sensitive to the first two terms that appear on the right-hand side of this equation, since the vertices must have distinct times.

In specific cases, such as when the vertices cannot provide enough time derivatives to create nontrivial contact terms, assumption (a) is sufficient for our purposes. However, in general it is necessary to replace it with the stronger assumption that

(a') the vertices are local

and further assume that

(c) the contact terms are local, i.e. the time derivatives of the propagators satisfy the property

$$\partial_0^n f(x) = \theta(x^0)\partial_0^n g_+(x) + \theta(-x^0)\partial_0^n g_-(x) + \text{local terms}; \tag{2.9}$$

(d) $g_\pm(x)$ and their derivatives $\partial_0^n g_\pm(x)$ have well-defined limits for $x^0 \to 0$.

When the propagators have the structure (2.1) property (c) follows as a consequence. Observe that a nontrivial contact term arises when the numerator contains a power of E greater than or equal to the maximum power of E appearing in the denominator. Let r denote the degree of $v(E, \mathbf{p})$ as a polynomial in E. Assumption (b) implies that the degree of $u(E, \mathbf{p})$ in E must be smaller than r. Write

$$\tilde{f}(E, \mathbf{p}) = \frac{u(E, \mathbf{p})}{E^r + w(E, \mathbf{p})},$$

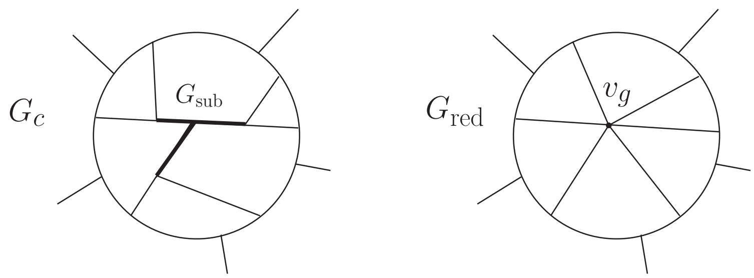
where $w(E, \mathbf{p})$ also has degree smaller than r. When $\tilde{f}(E, \mathbf{p})$ is multiplied by a sufficient power of E, the numerator may contain a power E^r that simplifies the power E^r of the denominator. Write $u(E, \mathbf{p}) = E^{r-1}u'(\mathbf{p}) + u''(E, \mathbf{p})$, where $u'(\mathbf{p})$ is a polynomial of $\mathbf{p}$ and $u''(E, \mathbf{p})$ is a polynomial of degree $r - 2$ in E. Then,

$$E\tilde{f}(E, \mathbf{p}) = \frac{Eu''(E, \mathbf{p}) - w(E, \mathbf{p})u'(\mathbf{p})}{E^r + w(E, \mathbf{p})} + u'(\mathbf{p}).$$

The ratio on the right-hand side does not contain contact terms, because the numerator contains at most $r - 1$ powers of E. Thus, (2.9) holds for $n = 1$. The argument can be easily iterated for $E^n \tilde{f}(E, \mathbf{p})$, $n > 1$, which proves (2.9) for every n.

It is easy to show that property (d) follows from (2.1), as long as (2.1) has only simple poles and $g_\pm$ are regular distributions.

We are ready to describe the procedure to deal with the contact terms. Consider a diagram G where some differentiated propagators carry contact terms. Separate them from the rest of $\partial_0^n f(x)$ as in formula (2.9) and write G as a sum of contributions G_{c}, such that each internal line of G_{c} is either a contact term or does not carry contact terms. The contact terms of G_{c} draw a subdiagram G_{sub} (which may be disconnected), as shown in the picture



It is easy to prove that if G_{sub} contains loops, it vanishes. Indeed, by assumptions (a') and (c), the vertices and the contact terms are both local, so each loop of contact terms is a linear combination of integrals (2.3) in momentum space. Thus, we can assume that G_{sub} is a tree subdiagram. Each connected component $G_{\text{sub}}^{\text{conn}}$ of G_{sub} is equal to a product of (derivatives of) delta functions $\delta(x_i^0 - x_j^0)$ times a new local vertex v_{g} that can be obtained by gluing the vertices of $G_{\text{sub}}^{\text{conn}}$ together. In turn, G_{c} is a product of (derivatives of) delta functions $\delta(x_i^0 - x_j^0)$ times a reduced diagram G_{red}, built with the ordinary vertices and the vertices v_{g}.

Now, G_{red} has no contact terms and thus satisfies the largest time equation (2.7). It is connected if the original diagram G is connected.

By property (d), a line that connects a marked vertex with an unmarked one is not interested by contact terms. By the same property, $\partial_0^n g_{\pm}(x)$ have well-defined limits for $x^0 \to 0$. Then, formula (2.6) shows that the contact terms carried by the lines connecting pairs of marked vertices are equal to minus the contact terms of $\partial_0^n f(x)$. We can easily show that, thanks to this fact, a minus sign is associated with each marked vertex of type $\hat{v}_{\text{g}}$, as expected. Indeed, $\hat{v}_{\text{g}}$ originates from the markings of all the vertices of $G_{\text{sub}}^{\text{conn}}$. Each such vertex provides a minus sign, but other minus signs come from the contact terms of $G_{\text{sub}}^{\text{conn}}$, because they are associated with pairs of marked vertices. Since $G_{\text{sub}}^{\text{conn}}$ is a tree diagram, the sum of the number of its vertices plus the number of its lines is odd, so $\hat{v}_{\text{g}}$ always carries a minus sign.

If we sum the G largest time equation (2.7), derived under the condition that all the nearest neighbors have distinct times, to the largest time equations satisfied by the diagrams G_{red}, multiplied by the appropriate products of (derivatives of) delta functions $\delta(x_i^0 - x_j^0)$, the right-hand side of (2.8) is fully reconstructed, for each propagator. Observe that assumption (a) plays an important role here, because it ensures that each diagram involves a finite number of time derivatives.

The conclusion is that if we add assumptions (a'), (c) and (d), the largest time equation (2.7) holds even if we drop the assumption that the vertices are located at distinct times. Then formula (2.7) can be interpreted as an identity of distributions and we can safely compute its Fourier transform.

The same conclusion holds when the vertices do not provide enough time derivatives to generate contact terms, in which case assumptions (a'), (c) and (d) need not be satisfied.

2.4 The cutting equations

Once the contact terms are dealt with as explained above, the Fourier transform of the largest time equation (2.7) is an analogous equation in momentum space, where the propagators and the vertices are replaced by their Fourier transforms. Denoting the Fourier transform of F_M with G_M, we get

$$\sum_{\text{markings } M} G_M(p_1, \cdots, p_n) = 0. \tag{2.10}$$

Now we simplify this identity by converting it into a set of cutting equations. The cutting equations are consequences of the assumptions made so far and the following additional one:

(e) the Fourier transforms $\tilde{g}_\pm(p)$ of $g_\pm(x)$ have the form

$$\tilde{g}_\pm(p) = \theta(\pm p^0) h_\pm(p). \tag{2.11}$$

For the moment, we make no further assumptions about the distributions $h_\pm(p)$. We interpret formulas (2.11) by saying that the energy flows from an unmarked vertex to a marked vertex, that is to say from the past to the future.

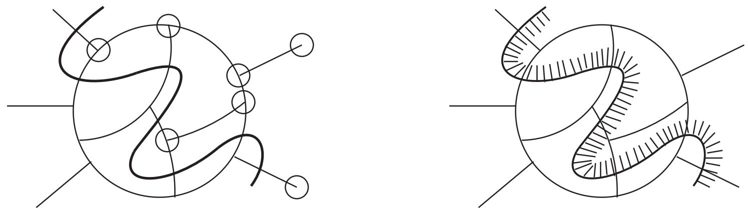
Consider a connected, amputated diagram of formula (2.10). Call the external legs whose energies flow into (out of) the diagram ingoing (outgoing). Mark the end points of the outgoing external legs and leave the end points of the incoming external legs unmarked.

We refer to the vertices and the end points of the external legs by simply calling them "points". Thus, the energy flows from an unmarked point to a marked point. Between two marked points or two unmarked points it can flow in both directions.

We want to show that every diagram G_M of (2.10) vanishes, unless it can be cut into two pieces, leaving the marked and unmarked points on opposite sides of the cut. If that is the case, we denote the diagram by G_C.

Consider a marked vertex. Its nearest points cannot be all unmarked, because then the orientations of the energy flows would imply the violation of energy conservation. Thus, at least one of its nearest neighbors is a marked point. Next, consider a connected subdiagram made of some marked vertices and the legs attached to them. Again, energy conservation implies that the nearest points of the subdiagram must include at least another marked point. Extending the subdiagram point by point, we find that each connected subdiagram of marked points must include the end point of an outgoing line. Similarly, a connected subdiagram of unmarked points must include the end point of an incoming line.

Because of this, the diagram is cut into two (not necessarily connected) subdiagrams. The cut crosses the propagators that connect a marked vertex to an unmarked vertex, as well as the external lines that connect a marked point to an unmarked point. For example, we have



(2.12)

In the left figure, the marked points are circled and the solid line denotes the cut. From now on, instead of marking the vertices, we just shadow the marked side of the cut, as shown in the right figure of (2.12). Normally, the incoming legs are drawn on the left-hand side and the outgoing legs are drawn on the right-hand side.

Since the external legs are amputated, the cutting of an external leg does not have any particular meaning besides the graphical one: all the marked points must lie on one side of the cut and all the unmarked points must lie on the other side.

We conclude that the Fourier transform (2.10) of the largest time equation (2.7) simplifies into the cutting equation

$$\sum_{\text{cuttings } C} G_C(p_1, \cdots, p_n) = 0. \tag{2.13}$$

We stress that equations (2.7), (2.10) and (2.13) do not assume that the external legs are on shell.

The sum of formula (2.13) contains two special contributions that it is convenient to single out. They are the contributions G and $\bar{G}$ of the diagrams where all the vertices are unmarked or marked, respectively. We have

$$G(p_1, \cdots, p_n) + \bar{G}(p_1, \cdots, p_n) = -\sum_{\text{proper cuttings } C} G_C(p_1, \cdots, p_n), \tag{2.14}$$

where the sum is restricted to the "proper" cuttings, which are those where at least one vertex is marked and at least one vertex is unmarked.

Everything we have said so far is valid at the regularized level. If the locality of counterterms holds, the diagrams built with the counterterms satisfy analogous properties. Combining the cutting equation of one diagram with the cutting equations satisfied by the diagrams that subtract its subdivergences and overall divergence, we obtain the renormalized cutting equation.

Note that in the renormalized cutting equation every side of the cut is appropriately renormalized. On the other hand, no counterterms are associated with subdiagrams containing the cut or part of it. The consistency of this fact is proved by the renormalized cutting equation itself. Indeed, after the inclusion of the counterterms the left-hand side of formula (2.14) is convergent, so the right-hand side must also be convergent.

So far, the assumptions we have made are more general than the usual ones. However, we anticipate that we cannot obtain the pseudounitarity equation unless we impose further restrictions.

2.5 Examples

Now we give some simple examples concentrating on scalar fields φ. Examples with fermions and gauge fields are given later on.

If we interpret the decomposition (2.5) as the usual T-ordered one, where

$$f(x) = \langle 0|T\varphi(x)\varphi(0)|0\rangle, \qquad g_+(x) = \langle 0|\varphi(x)\varphi(0)|0\rangle, \qquad g_-(x) = \langle 0|\varphi(0)\varphi(x)|0\rangle = g_+(-x),$$

and further assume Lorentz invariance, then we obtain the standard Källén-Lehman (KL) representation. Indeed, now $f(x)$ and $g_\pm(x)$ are Lorentz invariant and so are $\tilde{g}_\pm(p) = \theta(\pm p^0)h_\pm(p)$. However, the sign of p^0 depends on the reference frame, unless $p^2 > 0$. Thus, $h_\pm(p)$ vanish for $p^2 < 0$ and depend only on p^2 for $p^2 > 0$. Then, $g_-(x) = g_+(-x)$ implies that h_+ and h_- must be the same function, which we denote by $(2\pi)\rho(p^2)$. Inserting $\tilde{g}_\pm(p) = (2\pi)\theta(\pm p^0)\rho(p^2)$ inside (2.5) and working out the Fourier transform $\tilde{f}(p)$ of $f(x)$, we find the KL decomposition

$$\tilde{f}(p) = \int_0^{+\infty} \frac{i\rho(s)\mathrm{d}s}{p^2 - s + i\epsilon}. \tag{2.15}$$

We have used the identity

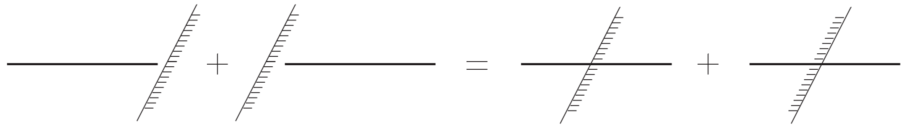
$$\theta(x^0) = \frac{i}{2\pi}\int_{-\infty}^{+\infty} \frac{\mathrm{e}^{-i\tau x^0}\mathrm{d}\tau}{\tau + i\epsilon}, \tag{2.16}$$

then changed the integration variable from τ to $\tau^2 - \mathbf{p}^2$ and used $\rho(p^2) = 0$ for $p^2 < 0$. Note that ρ is not assumed to be nonnegative.

In the case of ordinary (i.e. non-higher-derivative) free scalar fields, we have

$$\tilde{f}(p) = \frac{i}{p^2 - m^2 + i\epsilon}, \qquad \tilde{g}_\pm(p) = 2\pi\theta(\pm p^0)\rho(p^2), \qquad \rho(s) = \delta(s - m^2).$$

The simplest cutting equation is the one satisfied by the propagator:

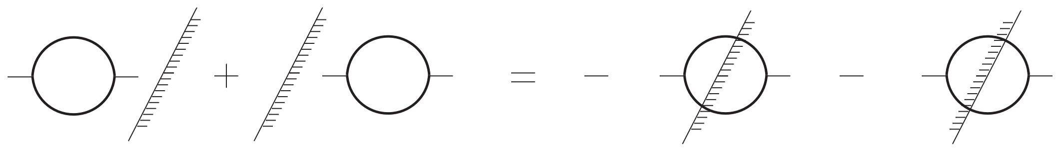


(2.17)

In deriving this equation, the end points must be imagined as vertices, so each shadowed end point gives a factor -1. This explains the signs of (2.17). In formulas, we have

$$\frac{i}{p^2 - m^2 + i\epsilon} + \frac{-i}{p^2 - m^2 - i\epsilon} = 2\pi\theta(p^0)\delta(p^2 - m^2) + 2\pi\theta(-p^0)\delta(p^2 - m^2). \qquad (2.18)$$

At one loop we have



(2.19)

which can be checked easily (see for example [11]).

Nonlocal quantum field theories do not satisfy the assumptions that lead to the largest time equation, unless their vertices are local in time. Then, however, either Lorentz invariance or gauge invariance is violated.

3 The pseudounitarity and unitarity equations

In this section we derive the pseudounitarity equation and explain when it implies the unitarity equation. As said, we must make additional assumptions, which eventually lead to a general Källén-Lehman spectral representation, even if we do not assume it from the start.

First, the shadowed regions of the cutting equations should correspond to the complex conjugate diagrams, that is to say we must assume that

(f) the action is Hermitian.

In particular, this implies that the shadowed propagator is the Hermitian conjugate of the unshadowed one and that the cut propagators (2.11) are Hermitian, i.e. $\tilde{g}_{\pm}^{\dagger}(p) = \tilde{g}_{\pm}(p)$. The minus sign associated with each marked vertex is then justified by the fact that the vertices are anti-Hermitian.

Second, the cut propagators must project onto the on shell states of the free field limit. This means that we must replace (e) by the more restrictive assumption that

(e') the Fourier transforms $\tilde{g}_{\pm}(p)$ of $g_{\pm}(x)$ have the form

$$\tilde{g}_{\pm}(p) = \pi \sum_{i=1}^{N} a_i^{\pm}(p)\delta(p^0 \mp \omega_i), \qquad (3.1)$$

where $\omega_i(\mathbf{p}^2)$ are positive functions of $\mathbf{p}$, while $a_i^{\pm}(p)$ are Hermitian matrices whose entries are functions of p.

Using the identity (2.16) it is easy to check that the Fourier transform of formula (2.5) is

$$\tilde{f}(p) = \frac{i}{2} \sum_{i=1}^{N} \frac{a_i^+(\omega_i, \mathbf{p})(p^0 + \omega_i) - a_i^-(-\omega_i, \mathbf{p})(p^0 - \omega_i)}{(p^0)^2 - \omega_i^2 + i\epsilon}. \tag{3.2}$$

With suitable assumptions on $a_i^\pm$ and ω_i, this formula matches (2.1).

At this point, we diagonalize the matrices $a_i^+(\omega_i, \mathbf{p})$ and $a_i^-(-\omega_i, \mathbf{p})$ and normalize their eigenvalues to 1, -1 and 0. Calling the diagonalizing matrices $u_i(\mathbf{p})$ and $v_i(\mathbf{p})$, we have

$$a_i^+(\omega_i, \mathbf{p}) = u_i(\mathbf{p}) H_i u_i^\dagger(\mathbf{p}), \qquad a_i^-(-\omega_i, -\mathbf{p}) = (-1)^{\sigma_i} v_i(\mathbf{p}) H_i' v_i^\dagger(\mathbf{p}),$$

where $\sigma_i = 0$, 1 for bosons and fermions, respectively, and H_i, H_i' are diagonal matrices with eigenvalues 1, 0 and -1. The matrices $u_i(\mathbf{p})$ and $v_i(\mathbf{p})$ collect the external particle and antiparticle states.

Writing the S matrix as $S = 1 + iT$, the cutting equations (2.14) can be collected into the pseudounitarity equation

$$-iT + iT^\dagger = THT^\dagger, \tag{3.3}$$

where H is the diagonal matrix having diagonal blocks H_i and H_i'.

If there exists a subspace V of states of the free field theory such that equation (3.3) holds with $H = 1$ when the external legs and the cut legs are projected onto V, then the pseudounitarity equation implies perturbative unitarity, which is expressed by the equation

$$-iT + iT^\dagger = TT^\dagger \tag{3.4}$$

in V.

Summarizing, the assumption that turns the pseudounitarity equation into the unitarity equation is that

(g) there exists a subspace V of states with $a_i^+(\omega_i, \mathbf{p}) > 0$, $(-1)^{\sigma_i} a_i^-(-\omega_i, -\mathbf{p}) > 0$, such that the cutting equations still hold after the external legs and the cut legs are projected onto V.

3.1 The Källén-Lehman spectral representation

We have found that, in general, the propagator must have the form (3.2), which means, in particular, that there can only be simple poles on the real axis, but no double poles and no poles away from the real axis. We can recast formula (3.2) in the form of the general Källén-Lehman representation

$$\tilde{f}(p) = i \int_0^{+\infty} \frac{\rho(s, \mathbf{p}) + p^0 \sigma(s, \mathbf{p})}{(p^0)^2 - s + i\epsilon} \mathrm{d}s, \tag{3.5}$$

where the densities σ and ρ are the Hermitian matrices given by

$$\rho(s,\mathbf{p}) = \sum_{i=1}^{N} \frac{\omega_i}{2} \left(a_i^+(\omega_i,\mathbf{p}) + a_i^-(-\omega_i,\mathbf{p}) \right) \delta(s-\omega_i^2),$$
$$\sigma(s,\mathbf{p}) = \sum_{i=1}^{N} \frac{1}{2} \left(a_i^+(\omega_i,\mathbf{p}) - a_i^-(-\omega_i,\mathbf{p}) \right) \delta(s-\omega_i^2).$$

It is easy to check formula (2.17) in this general case.

The representation (3.5) has a form similar to the one known from Lorentz violating theories [13]. When Lorentz symmetry holds, the densities $\rho(s,\mathbf{p})$ and $\sigma(s,\mathbf{p})$ vanish for $s<\mathbf{p}^2$, so after a translation the representation acquires a more common form, that is to say (2.15) with $\rho(s)$ replaced by $\rho(s+\mathbf{p}^2,\mathbf{p}) + p^0\sigma(s+\mathbf{p}^2,\mathbf{p})$. Lorentz invariance also implies that this sum has the form $\mathbf{p}\cdot\boldsymbol{\rho}'(s) + p^0\sigma'(s) + \rho''(s)$ and further relates the functions $\boldsymbol{\rho}'(s)$ and $\sigma'(s)$.

3.2 Examples

Most bosons have $a^-(-\omega,\mathbf{p}) = a^+(\omega,\mathbf{p})$, so the coefficient $\sigma(s,\mathbf{p})$ of p^0 in the numerator of (3.5) vanishes. This gives

$$\tilde{f}(p) = i\frac{\omega a^+(\omega,\mathbf{p})}{(p^0)^2-\omega^2+i\epsilon}. \tag{3.6}$$

Lorentz invariant scalars have $a^+(\omega,\mathbf{p}) = 1/\omega$, $\omega = \sqrt{\mathbf{p}^2+m^2}$.

Examples where the coefficient of p^0 does not vanish are the Chern-Simons gauge fields and the fermions. In particular, free Dirac fermions have $a^+(p) = a^-(p) = (p_\mu\gamma^\mu + m)\gamma^0/\omega$, which gives

$$\tilde{f}(p)\gamma^0 = i\frac{p_\mu\gamma^\mu + m}{p^2-m^2+i\epsilon}.$$

Interacting fermions in Lorentz invariant theories have

$$\tilde{f}(p)\gamma^0 = i\int_0^{+\infty} \frac{\rho(s)+p_\mu\gamma^\mu\sigma(s)}{p^2-s+i\epsilon}\mathrm{d}s.$$

Let us now consider gauge fields. If we choose a covariant gauge the pseudounitarity equation exists only when the propagators have the form

$$\tilde{f}_{\mu_1\cdots\mu_n,\nu_1\cdots\nu_n}(p) = i\frac{\mathcal{I}_{\mu_1\cdots\mu_n,\nu_1\cdots\nu_n}}{p^2+i\epsilon}, \tag{3.7}$$

where $\mathcal{I}_{\mu_1\cdots\mu_n,\nu_1\cdots\nu_n}$ is a constant tensor built with the metric $\eta_{\mu\nu}$. In other words, no covariant gauges besides the Feynman ones satisfy the assumptions. The common Lorenz gauge for vector fields, which gives the propagator

$$-\frac{i}{p^2}\left(\eta_{\mu\nu} - (1-\lambda)\frac{p_\mu p_\nu}{p^2}\right), \tag{3.8}$$

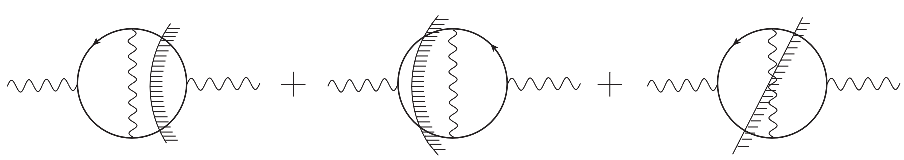
does not lead to the cutting equations (2.14) when the gauge-fixing parameter λ is different from 1, because of the double pole. It is possible to deform (3.8) by introducing fictitious masses that split the double pole into simple poles, but it is not easy to study the limit where the fictitious masses are removed in the cutting equations.

We see that, in the end, the conditions imposed by the very existence of the cutting equations and the requirement that they lead to the pseudounitarity and unitarity equations are very restrictive.

The largest time equation (2.7), the cutting equations (2.14), the pseudounitarity equation (3.3) and the unitarity equation (3.4) also hold when the external legs of the diagrams correspond to the insertions of local composite fields. Indeed, it is easy to check that the arguments that lead to those equations remain valid. More generally, the equations still hold when the external legs include both elementary fields and local composite fields.

3.3 Infrared divergences and other singularities

The uncut diagrams that appear on the left-hand side of the cutting equations (2.14) are regular off shell. However, the individual diagrams that appear on the right-hand side have cuts, which are necessarily on shell. In the presence of massless particles there can be infrared divergences. For example, consider the sum



(3.9)

in QED. The first two diagrams contain the infrared divergences of the one-loop radiative corrections to the vertex. However, the third diagram is also infrared divergent and compensates the divergences of the other two.

The cancellation of the infrared divergences on the right-hand side of equation (2.14) (when the external legs are off shell) is a well-known fact [14], so we do not need to spend more words on it. At the same time, for various arguments of the next sections we need to deal with cut diagrams that are individually infrared convergent. This can be achieved by inserting fictitious masses in the propagators. It is possible to do so without violating the assumptions we have made so far. However, the fictitious masses violate gauge invariance and it is necessary to remove them with care to successfully prove the perturbative unitarity of gauge theories.

Other singularities occur when self-energy subdiagrams are present. For example, the product of a cut propagator times an unshadowed propagator with the same momentum is equal to

$$\frac{i(2\pi)\theta(p^0)\delta(p^2-m^2)}{p^2-m^2+i\epsilon}=\frac{2\pi}{\epsilon}\theta(p^0)\delta(p^2-m^2), \tag{3.10}$$

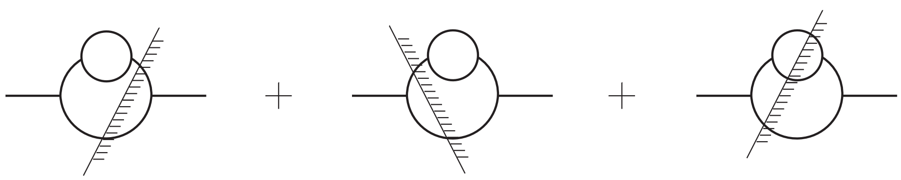
in the case of ordinary scalar fields. On the other hand, the product of an unshadowed propagator times a shadowed one with the same momentum is

$$\frac{i}{p^2 - m^2 + i\epsilon} \frac{-i}{p^2 - m^2 - i\epsilon} = \frac{1}{(p^2 - m^2 + i\epsilon)^2} + \frac{2\pi}{\epsilon}\delta(p^2 - m^2), \tag{3.11}$$

where we have used (2.18).

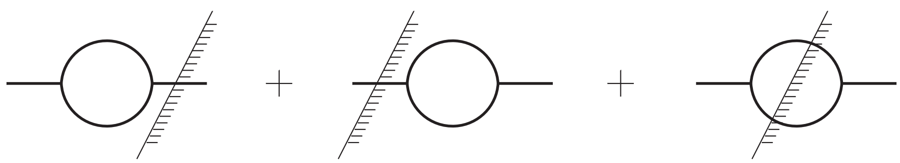
Again, the left-hand sides of the cutting equations are smooth, while the individual diagrams on the right-hand side may have singularities for $\epsilon \to 0$ that cancel out in the sum. The cancellation can be seen by keeping the width ϵ different from zero and taking the limit $\epsilon \to 0$ only at the very end.

For example, consider the bubble diagram where one propagator is replaced by the one-loop self-energy. The right-hand side of the cutting equation is equal to minus the sum



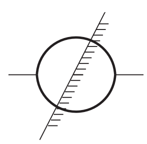
(3.12)

where we have assumed, for definiteness, that the energy flows in from the left. Using (2.19), (3.10) and (3.11), it is easy to check that the sum



(3.13)

(with propagators on the external legs) is equal to

 $\times \dfrac{1}{(p^2 - m^2 + i\epsilon)^2}$,

which is regular when $\epsilon \to 0$. Thus, (3.12) is also regular.

4 The special gauge

We have seen that the only covariant gauge that leads to the pseudounitarity equation is the Feynman gauge, which corresponds to formula (3.8) with $\lambda = 1$. However, the Feynman gauge has ghosts, that is to say the matrix H of formula (3.3) has negative entries.

The Feynman gauges do not make unitarity manifest. Actually, all the propagators (3.7) for $n > 0$ have ghosts. It is hard to prove that the ghosts compensate each other in the Feynman gauge, although not impossible [7, 8]. Here we prefer to follow a different strategy, which amounts to prove perturbative unitarity in gauge theories and gravity by working in a new, noncovariant gauge that satisfies all the requirements we have outlined and interpolates between the Feynman gauge and the Coulomb gauge.

We call the new gauge "special", because of its properties. In this section we build the special gauge in Abelian and non-Abelian gauge theories, while in section 7 we build it in quantum gravity. We work in arbitrary dimensions[3] d greater than 3. The gauge group indices are understood in the formulas written below.

Consider the gauge-fixed Lagrangian

$$\mathcal{L}_{\mathrm{YM}} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} - \frac{1}{2\lambda}\mathcal{G}^2(A) - \bar{C}\mathcal{G}(DC), \tag{4.1}$$

where $\mathcal{G}(A)$ is the gauge choice, which is assumed to be linear in A, while D denotes the covariant derivative. In $d > 4$ the theory is nonrenormalizable, so we should include infinitely many corrections of higher dimensions, which are optional in $d = 4$. We do not write them explicitly, because, for our purposes, it is sufficient to assume that they are perturbative, local and Hermitian. We also omit the matter contributions, which are not important for the moment.

Now, take

$$\mathcal{G}(A) = \zeta\partial_0 A_0 + \nabla\cdot\mathbf{A}, \tag{4.2}$$

where ζ is another gauge-fixing parameter. For $\zeta = 1, 0$ we have the Lorenz and Coulomb gauges, respectively.

The gauge field propagators read

$$\langle A^0(k)A^0(-k)\rangle_0 = -\frac{i(\lambda E^2 - \mathbf{k}^2)}{(\zeta E^2 - \mathbf{k}^2)^2}, \qquad \langle A^i(k)A^0(-k)\rangle_0 = \frac{i(\zeta - \lambda)k^i E}{(\zeta E^2 - \mathbf{k}^2)^2},$$
$$\langle A^i(k)A^j(-k)\rangle_0 = \frac{i\Pi^{ij}}{E^2 - \mathbf{k}^2} + \frac{i(\zeta^2 E^2 - \lambda\mathbf{k}^2)}{(\zeta E^2 - \mathbf{k}^2)^2}\frac{k^i k^j}{\mathbf{k}^2}, \tag{4.3}$$

where the $\langle\cdots\rangle_0$ denotes the free field limit of the average and

$$\Pi^{ij} = \delta^{ij} - \frac{k^i k^j}{\mathbf{k}^2} \tag{4.4}$$

is the projector onto the transversal components of the gauge field. The ghost propagator is

$$\langle C(k)\bar{C}(-k)\rangle_0 = \frac{i}{\zeta E^2 - \mathbf{k}^2}. \tag{4.5}$$

[3]The case $d = 3$, which we do not treat here in detail, can be studied by including the Chern-Simons term, to avoid the infrared problems due to the superrenormalizability of the gauge coupling.

The propagators just listed do not satisfy the assumptions required by the pseudounitarity equation for generic values of λ and ζ, because they have double poles. The special gauge is defined as the one with $\lambda = \zeta > 0$, where the double poles disappear. We obtain

$$\left\langle A^0(k)A^0(-k)\right\rangle_0 = \frac{-i}{\lambda E^2 - \mathbf{k}^2 + i\epsilon} = -\left\langle C(k)\bar{C}(-k)\right\rangle_0, \qquad \left\langle A^i(k)A^0(-k)\right\rangle_0 = 0,$$
$$\left\langle A^i(k)A^j(-k)\right\rangle_0 = \frac{i\Pi^{ij}}{E^2 - \mathbf{k}^2 + i\epsilon} + \frac{i\lambda}{\lambda E^2 - \mathbf{k}^2 + i\epsilon}\frac{k^i k^j}{\mathbf{k}^2}, \tag{4.6}$$

where we have inserted the contour prescriptions, which are now straightforward. Note that formulas (4.6) have good power counting behaviors. In particular, the denominators $\mathbf{k}^2$ cancel out in the sum. The KL spectral representation (3.5) is satisfied, although the densities ρ are not positive definite.

The limit $\lambda \to 0^+$ takes us to Coulomb gauge, actually the Landau limit of the Coulomb gauge. There, the assumptions we have made do not hold, because, for example, $\left\langle A^0(x)A^0(y)\right\rangle$ is proportional to $\delta(x^0 - y^0)$, which violates (2.5). In the next section we show that in QED there is a way to circumvent this difficulty and work directly at $\lambda = 0$. Instead, in non-Abelian gauge theories (and *a fortiori* gravity) it is necessary to work at $\lambda \neq 0$.

To deal with the infrared divergences, we need to introduce an infrared cutoff m_γ and remove it later. This can be done in various ways. We describe two methods that are good for our purposes, a simpler one and a more involved one. The simpler method works well in renormalizable theories, the more involved one is designed to work in nonrenormalizable theories.

4.1 Renormalizable theories

For the moment, we concentrate on $d = 4$ renormalizable Abelian and non-Abelian gauge theories, possibly coupled to matter. There, it is sufficient to replace the propagators (4.6) with

$$\left\langle A^0(k)A^0(-k)\right\rangle_0 = -\frac{i}{\lambda E^2 - \mathbf{k}^2 - m_\gamma^2 + i\epsilon} = -\left\langle C(k)\bar{C}(-k)\right\rangle_0, \qquad \left\langle A^i(k)A^0(-k)\right\rangle_0 = 0,$$
$$\left\langle A^i(k)A^j(-k)\right\rangle_0 = \frac{i\Pi^{ij}}{E^2 - \mathbf{k}^2 - m_\gamma^2 + i\epsilon} + \frac{i\lambda}{\lambda E^2 - \mathbf{k}^2 - m_\gamma^2 + i\epsilon}\frac{k^i k^j}{\mathbf{k}^2}. \tag{4.7}$$

The Lagrangian that leads to the propagators (4.7) is equal to (4.1) plus the mass terms

$$\mathcal{L}_{m_\gamma} = \frac{m_\gamma^2}{2}A_\mu A^\mu + \frac{m_\gamma^2}{2\lambda}(1-\lambda)(\boldsymbol{\nabla}\cdot\mathbf{A})\frac{1}{\Delta}(\boldsymbol{\nabla}\cdot\mathbf{A}) - m_\gamma^2\bar{C}C, \tag{4.8}$$

and is nonlocal in space.

We must show that the regularization defined in subsection 2.1 is well defined in the special gauge, because the denominator $\mathbf{k}^2$ of the projector $k^i k^j/\mathbf{k}^2$ does not have the form specified in formula (2.1). We must also pay attention to the contact terms, because the procedure of subsection 2.3 to deal with them does depend on (2.1).

For definiteness, we call an expression *regular* if it just involves denominators equal to products of polynomials $aE^2 - b\mathbf{k}^2 - m^2 + i\epsilon$ with $a > 0$ and $b > 0$. We call any other expression *irregular*. For example, the projector $k^i k^j/\mathbf{k}^2$ is irregular, although it has good infrared and ultraviolet behaviors.

Let us point out a few obvious facts. The propagators of the bosonic fields and those of the Faddeev-Popov ghosts decrease like $1/E^2$ at large energies. Instead, the fermionic propagators decrease like $1/E$. Call the vertices that carry at least three legs "proper". In $d = 4$ renormalizable Abelian and non-Abelian gauge theories coupled to matter the proper vertices with no fermionic legs contain at most one time derivative, while the proper vertices involving fermionic legs have no time derivatives. Finally, the irregular contributions to $\left\langle A^i(k)A^j(-k)\right\rangle_0$ read

$$\frac{i(1-\lambda)m_\gamma^2}{(\lambda E^2 - \mathbf{k}^2 - m_\gamma^2 + i\epsilon)(E^2 - \mathbf{k}^2 - m_\gamma^2 + i\epsilon)}\frac{k^i k^j}{\mathbf{k}^2} \tag{4.9}$$

and behave like $1/E^4$ at large energies.

We treat the quadratic counterterms as two-leg vertices. To minimize the number of time derivatives acting on the same propagator in coordinate space, the kinetic counterterms are assumed to have the forms $\sim (\partial_0\phi)^2$ for bosons and $\sim \bar{\psi}\partial_0\psi$ for fermions.

Now, consider a Feynman diagram. The power of the energy brought by the vertices of each loop is at most equal to the number of proper vertices with no fermionic legs, plus twice the number of bosonic quadratic counterterms, plus the number of fermionic quadratic counterterms. Then, if we ignore the tadpoles for a moment, every energy integral is convergent (before integrating on the space momenta) and the multiple energy integrals are overall convergent.

The tadpoles can be treated apart. The fermionic tadpole is straightforward, because it does not involve the projector $k^i k^j/\mathbf{k}^2$. The bosonic tadpole involves it by means of (4.9), which contributes to a convergent energy integral. Moreover, adopting the prescription of symmetric integration, the energy integrals are convergent in both types of tadpoles.

Thus, the regularization of subsection 2.1 is well defined. The integrals on the energy and those on the space momenta can be freely interchanged.

The same arguments prove that the irregular contributions (4.9) to the propagators cannot generate contact terms. Indeed, the vertices cannot provide enough E powers to compensate the E^4 appearing in the denominator of (4.9). Thus, the contact terms can only come from the regular terms and can be treated as explained in subsection 2.3.

The locality of counterterms is usually proved by differentiating the Feynman diagrams with respect to the external energies and space momenta, then showing that a sufficient number of such derivatives makes the integrals overall convergent [15]. This strategy works when the integrands are regular. Instead, the derivatives of $k^i k^j/\mathbf{k}^2$ with respect to the components of $\mathbf{k}$ just improve the ultraviolet behavior of the integral on $\mathbf{k}$, but do not improve the behavior of the integral on the energy E. Nevertheless, we have shown that all the integrals on the energies are convergent by themselves, so their ultraviolet behaviors do not need to be improved. For this reason, a sufficient number of derivatives with respect to the external energies and space momenta does make a diagram overall convergent. Once the diagram is equipped with the counterterms that subtract its subdivergences, the same operation makes the sum fully convergent. It is easy to check that all the regions of integration are properly subtracted. This proves the locality of counterterms. For similar reasons, it is straightforward to prove that the counterterms are polynomial in m_γ^2.

In the end, the renormalization in the special gauge is straightforward. The renormalized Lagrangian coincides with the one at $m_\gamma = 0$ plus the counterterms

$$\Delta\mathcal{L}_{m_\gamma} = \frac{m_\gamma^2}{2}\Delta Z_0 (A^0)^2 - \frac{m_\gamma^2}{2}\Delta Z_s (A^i)^2 - m_\gamma^2 \Delta Z_g \bar{C} C,$$

where ΔZ_0, ΔZ_s and ΔZ_g are divergent constants.

The other requirements of the previous sections are fulfilled, before and after renormalization. This ensures that the largest time equation, the cutting equations and the pseudounitarity equation hold in the special gauge for arbitrary $\lambda > 0$ in $d = 4$, if the theory is renormalizable by power counting.

4.2 Nonrenormalizable theories

The construction just given is sufficient for renormalizable gauge theories, such as the standard model in flat space. In view of the generalization to quantum gravity, we explain how to adapt the special gauge to nonrenormalizable gauge theories in arbitrary dimensions $d > 3$.

We introduce two fictitious masses, μ and m_γ, which play different roles. Define

$$P_{\lambda,\theta,\eta} \equiv \frac{1}{\lambda E^2 - \theta \mathbf{k}^2 - \eta\mu^2 - m_\gamma^2 + i\epsilon}$$

and replace the propagator $\left\langle A^i(k) A^j(-k) \right\rangle_0$ of (4.7) by

$$\left\langle A^i(k) A^j(-k) \right\rangle_0 = i P_{1,1,1}\delta^{ij} + i\left(Q_N(\lambda, r) - P_{1,1,1}\right)\frac{k^i k^j}{\mathbf{k}^2 + \mu^2}, \qquad (4.10)$$

where

$$Q_N(\lambda, r) \equiv \lambda \sum_{n=0}^{N} m_\gamma^{2n} (\lambda - 1)^n \prod_{q=0}^{n} P_{\lambda, r_q, r_q}, \qquad (4.11)$$

$r \equiv \{r_0, r_1, \ldots\}$ and r_q are positive constants such that $r_q \neq r_{q'}$ for $q \neq q'$, with $r_0 = 1$. For example, we can choose $r_q = q + 1$.

Before explaining where the idea for the replacement (4.10) comes from, we give its key properties, in connection with unitarity and renormalization.

It is easy to prove the identity

$$Q_N(\lambda, r) = \sum_{n=0}^{N} P_{\lambda,r_n,r_n} \sum_{q=n}^{N} \left(\frac{m_\gamma^2}{\mathbf{k}^2 + \mu^2} \right)^q c_{nq}(\lambda), \tag{4.12}$$

where $c_{nq}(\lambda)$ are polynomials of λ. We see that μ regulates the infrared divergences that would appear in the individual terms on the right-hand side of this formula at $\mathbf{k} = 0$.

The irregular term $ik^i k^j/(\mathbf{k}^2 + \mu^2)$ inside $\langle A^i(k)A^j(-k)\rangle_0$ is multiplied by the difference $Q_N(\lambda, r) - P_{1,1,1}$, which satisfies the property

$$\frac{ik^i k^j}{\mathbf{k}^2 + \mu^2} (Q_N(\lambda, r) - P_{1,1,1}) = -i \frac{k^i k^j}{\mathbf{k}^2 + \mu^2} \frac{(\lambda - 1)^{N+1} m_\gamma^{2N+2}}{E^2 - \mathbf{k}^2 - \mu^2 - m_\gamma^2 + i\epsilon} \prod_{n=0}^{N} P_{\lambda,r_n,r_n} + \text{ regular terms.} \tag{4.13}$$

To derive this formula, note that if we set $r_q = 1$ for every q and $N = \infty$, the function $Q_N(\lambda, r)$ resums into

$$Q_\infty(\lambda) \equiv \lambda \sum_{n=0}^{\infty} m_\gamma^{2n} (\lambda - 1)^n \prod_{q=0}^{n} P_{\lambda,1,1} = \frac{\lambda}{\lambda E^2 - \mathbf{k}^2 - \mu^2 - \lambda m_\gamma^2 + i\epsilon}. \tag{4.14}$$

Replacing $Q_N(\lambda, r)$ by $Q_\infty(\lambda)$ in (4.10), we obtain

$$\langle A^i(k)A^j(-k)\rangle_0 = iP_{1,1,1}\delta^{ij} + \frac{i(1 - \lambda)k^i k^j}{(E^2 - \mathbf{k}^2 - \mu^2 - m_\gamma^2 + i\epsilon)(\lambda E^2 - \mathbf{k}^2 - \mu^2 - \lambda m_\gamma^2 + i\epsilon)}. \tag{4.15}$$

Then, there are no irregular terms, the regularization of subsection 2.1 is well defined, the contact terms are under control by means of the procedure of subsection 2.3 and the locality of counterterms is obvious. The point is that the arguments of section 6 about unitarity do not work well with the choice (4.15), because (4.14) shows that the squared mass m_γ^2 gets multiplied by λ in some cuts, which invalidates the inequality (6.3) at $\mu = 0$.

Nonetheless, the resummation (4.14) gives us the inspiration for the replacement (4.10). Indeed, truncate the sum of (4.14) to $N < \infty$ and replace the coefficients of $\mathbf{k}^2 + \mu^2$ in the denominators with arbitrary numbers $r_0, r_1, \ldots$, so as to obtain (4.11). It is clear that these operations lead to the behavior (4.13). In particular, the variations of the $\mathbf{k}^2 + \mu^2$ coefficients just affect the regular terms. The role of those coefficients is to make sure that there are no double poles.

Now we use (4.13) to prove that the modification (4.10) has the properties we need, that is to say the regularization of subsection 2.1 is well defined, the contact terms are under control and the locality of counterterms holds. Such properties are obviously satisfied by the regular contributions to the Feynman diagrams, so we can concentrate on the contributions that involve irregular terms.

We recall that we are considering a nonrenormalizable theory, whose Lagrangian contains infinitely many vertices. It is helpful to expand the interaction Lagrangian in powers of the energy and focus on some finite truncation. If, at the same time, we truncate the loop expansion to a finite order, only a finite number of amplitudes, vertices and diagrams are involved in every calculation. So doing, we are able to prove perturbative unitarity within any finite truncation, which is enough to prove perturbative unitarity for the whole theory.

By formula (4.13), there exists an N such that all the irregular contributions to the Feynman diagrams are overall convergent within the truncation. At one loop, the integrals that contain irregular terms are convergent by themselves. At higher orders, they are convergent once the counterterms that subtract the subdivergences (associated with the regular contributions to the subdiagrams) are included. Thus, the regularization of subsection 2.1 is well defined. Since the irregular terms do not contribute to the renormalization of the theory, the locality of counterterms obviously holds. Moreover, formula (4.13) shows that for N large enough the vertices cannot provide enough powers of E to match the total E powers appearing in the denominators of the irregular terms. This means that the contact terms are local, within the truncation, because they can only be generated by the regular terms. This fact, together with the locality of the vertices, ensures that the procedure of subsection 2.3 to deal with the contact terms is still valid.

The construction also works in the case of renormalizable theories, where it is sufficient to choose $N > d/2 - 1$.

Formulas (4.12) and (4.10) show that the assumptions that lead to the pseudounitarity equation are satisfied at $\lambda > 0$, $m_\gamma \neq 0$, for arbitrary $N \geqslant 0$.

Some remarks are in order, about the recovery of gauge invariance and gauge independence when m_γ is sent back to zero. Using the Batalin-Vilkovisky formalism [16], gauge invariance is encoded into the antiparentheses $(S, S) = 2(S, S_{m_\gamma})$, where $S_{m_\gamma} = \int \mathcal{L}_{m_\gamma}$ collects the m_γ mass terms, while gauge independence is encoded in the expression

$$\frac{\partial S}{\partial \lambda} - (S, \Psi_\lambda) = \frac{\partial S_{m_\gamma}}{\partial \lambda},$$

where Ψ_λ is the λ derivative of the gauge fermion Ψ, which is the local functional that performs the gauge fixing (for a recent reference, with details and the notation, see [17]). The right-hand sides of both equations should vanish, at least when $\varepsilon = 0$, but they do not if $m_\gamma \neq 0$. Their effects on the generating functional Γ of the one-particle irreducible correlation functions are encoded into the averages $2\langle (S, S_{m_\gamma}) \rangle$ and $\langle \partial S_{m_\gamma} / \partial \lambda \rangle$, which contain insertions of new vertices besides those of the standard Feynman rules. We want to make sure that the Feynman diagrams that contain such insertions are also well regularized and satisfy the locality of counterterms, and check that their contact terms are still under control.

Write the free massive Lagrangian in compact notation as

$$\mathcal{L}_{\text{free}} + \mathcal{L}_{m_\gamma} = \frac{1}{2} \Phi^\alpha Q_{\alpha\beta} \Phi^\beta,$$

where Φ^α are all the fields (including the ghosts, the antighosts and the Lagrange multipliers for the gauge fixing). We have

$$\frac{\partial S_{m_\gamma}}{\partial\lambda}=\frac{1}{2}\int\Phi^\alpha\frac{\partial Q_{\alpha\beta}}{\partial\lambda}\Phi^\beta-\frac{\partial}{\partial\lambda}\int\mathcal{L}_{\text{free}}.$$

The last contribution is local. The other term gives $-(i/2)(\partial f^{\alpha\beta}/\partial\lambda)$, if we include the propagators $f^{\alpha\beta}$ attached to its legs. The irregular part can then be easily derived from formula (4.13). Again, if N is large enough this irregular insertion cannot generate contact terms and every subintegral that contains it is overall convergent.

Similarly,

$$2(S,S_{m_\gamma})=(S,\Phi^\alpha)Q_{\alpha\beta}\Phi^\beta-2(S,S_{\text{free}}).$$

The last contribution is local, while the other term becomes local, precisely equal to $i(S,\Phi^\alpha)$, once it is inserted in a Feynman diagram and the propagator attached to the right field Φ is included.

5 Proof of unitarity in QED

We are now ready to give the simplest proof of perturbative unitarity in gauge theories, which applies to QED (in $d=4$).

First we show that the Feynman diagrams can be calculated directly in the Coulomb gauge, which can be reached as the $\lambda\rightarrow 0$ limit of the special gauge defined in the previous section. The quadratic part of the Lagrangian is singular at $\lambda=0$, but the Feynman rules are regular. The propagators are

$$\left\langle A^0(k)A^0(-k)\right\rangle_0=\frac{i}{\mathbf{k}^2+m_\gamma^2}=-\left\langle C(k)\bar{C}(-k)\right\rangle_0,\qquad\left\langle A^i(k)A^j(-k)\right\rangle_0=\frac{i\Pi^{ij}}{E^2-\mathbf{k}^2-m_\gamma^2+i\epsilon},$$
$$\left\langle A^i(k)A^0(-k)\right\rangle_0=0,\qquad\left\langle\psi(p)\bar{\psi}(-p)\right\rangle_0=\frac{i(p_\mu\gamma^\mu+m)}{p^2-m^2+i\epsilon},\tag{5.1}$$

while the vertices are encoded in the interaction Lagrangian $\mathcal{L}_I=-eA_\mu\bar{\psi}\gamma^\mu\psi$.

We want to show that the correlation functions are also regular in the limit $\lambda\rightarrow 0$. The problem is that the propagators $\left\langle A^0A^0\right\rangle_0$ and $\left\langle C\bar{C}\right\rangle_0$ do not have the form (2.1) at $\lambda=0$. Moreover, the propagator of the longitudinal component $\mathbf{A}_\parallel\equiv(1/\sqrt{-\Delta})\boldsymbol{\nabla}\cdot\mathbf{A}$ of the photon naively disappears in the limit, because it is multiplied by λ. We must show that the contributions coming from this component can be consistently dropped from the Feynman diagrams. The ghosts can be ignored, because they decouple.

Each loop contains at least one different fermion propagator, which makes the energy integrals behave at worst like $\sim\int\mathrm{d}E/E$ for large energies E. Such integrals are convergent by symmetric integration. At $\lambda>0$ the symmetric integration is justified by the regularization of section 2. At $\lambda=0$ it must be assumed by default. Then, the Feynman diagrams are well regularized. Moreover, the energy integrals can be freely interchanged with the integrals on the space momenta. Finally, the extra λ factor carried by the $\mathbf{A}_{\parallel}$ propagator multiplies integrals that are regular for $\lambda \to 0$. Thus, every diagram that has internal $\mathbf{A}_{\parallel}$ lines disappears in the limit. The regular behavior for $\lambda \to 0$ implies that the counterterms of the Coulomb gauge are the $\lambda \to 0$ limit of those evaluated at $\lambda > 0$. The diagrams that contain external $\mathbf{A}_{\parallel}$ legs can be ignored, as well as their counterterms, because, by formula (5.1), the vertices that depend on $\mathbf{A}_{\parallel}$ do not contribute beyond the tree level.

Now we assume that λ vanishes and inquire about unitarity. The main difficulty is that the propagator $\langle A^0 A^0 \rangle_0$ does not satisfy the assumptions that lead to the pseudounitarity equation, because it does not have the form (3.5). We can solve this problem as follows. Since we do not need A^0 on the external legs of equation (2.14), we integrate it out.

Consider a generic (uncut) Feynman diagram with A^i, ψ and $\bar{\psi}$ on the external legs. Every A^0 internal leg must connect two vertices proportional to $A_0 \psi^\dagger \psi$. Focus on a subdiagram made by such vertices and the A^0 propagator that connects them. Replace this subdiagram with an effective four fermion vertex proportional to $\psi^\dagger \psi \langle A^0 A'^0 \rangle_0 \psi'^\dagger \psi'$. This operation is equivalent to replace the Feynman rules listed above with the ones where the propagators are just $\langle A^i A^j \rangle_0$ and $\langle \psi \bar{\psi} \rangle_0$, while the vertices are those encoded in the effective interaction Lagrangian

$$\mathcal{L}_{I\mathrm{eff}}(t, \mathbf{r}) = -\frac{m_\gamma e^2}{8\pi} \int \rho(t, \mathbf{r}) V(m_\gamma |\mathbf{r} - \mathbf{r}'|) \rho(t, \mathbf{r}') \mathrm{d}^{3-\varepsilon} \mathbf{r}' + e \mathbf{J}(t, \mathbf{r}) \cdot \mathbf{A}(t, \mathbf{r}) \tag{5.2}$$

(at the tree level), where $\rho = \psi^\dagger \psi$ and $J^i = \bar{\psi} \gamma^i \psi$, while the function $V(x)$ can be considered as the dimensional continuation of the Yukawa potential [indeed, $V(x) = \mathrm{e}^{-x}/x$ at $\varepsilon = 0$]. At the renormalized level, $\mathcal{L}_{I\mathrm{eff}}(t, \mathbf{r})$ keeps its form, except for renormalization constants in front of the two vertices and m_γ.

The new Feynman rules satisfy our assumptions. Observe that (5.2) is local in time and nonlocal in space, which means that assumption (a) is satisfied, but assumption (a') is not. This is not a problem, because assumption (a') is just required to handle the loops of contact terms, which cannot be generated here. Moreover, the propagators $\langle A^i A^j \rangle_0$ and $\langle \psi \bar{\psi} \rangle_0$ satisfy the general KL decomposition (3.5). Thus, the largest time equation (2.7) and the cutting equations (2.14) hold. Since the cut propagators have only (massive) physical states, this leads to the unitarity equation (3.4).

Next, we build the physical amplitudes. We still have m_γ, so the theory defined by the new Feynman rules is not gauge invariant. If the external fields are off shell, the left-hand side of the cutting equation does not have infrared divergences when m_γ tends to zero. Thus, the sum of the cut diagrams on the right-hand side is also smooth in the limit $m_\gamma \to 0$.

When we put the external legs on shell, other infrared divergences appear. They can be canceled by summing cutting equations associated with diagrams that have the same types of infrared divergences [14].

Finally, we must project the external legs onto gauge invariant states. The external photons $A^i(k)$ can be multiplied by the physical polarizations or the projectors Π^{ij} of (4.4). The cut photon legs are already multiplied by such projectors. The external electron legs and the cut electron legs do not need a special treatment in the Coulomb gauge, because an insertion of ψ is equivalent to the insertion of the gauge invariant operator

$$\psi' = \psi \exp\left(ie\frac{1}{\Delta}\boldsymbol{\nabla}\cdot\mathbf{A}\right).$$

The transverse form of the propagator $\left\langle A^i A^j \right\rangle_0$ appearing in formula (5.1) shows that the insertions of ψ' and ψ in the correlation functions give the same results.

The operations just described allow us to conclude that the S matrix of perturbative QED is unitary, as desired. Since the physical amplitudes are gauge independent, the conclusion extends from the Coulomb gauge to any other gauge.

6 Proof of unitarity in non-Abelian gauge theories

In this section we prove the perturbative unitarity of non-Abelian gauge theories, while in the next section we extend the proof to quantum gravity.

The arguments of the previous section do not generalize beyond QED. Indeed, in non-Abelian gauge theories and quantum gravity it is not consistent to work at $\lambda = 0$, where several Feynman diagrams become singular. For example, it is possible to build loops of circulating ghosts C, $\bar{C}$ and/or gauge fields A_0. According to formula (5.1), the integrands of such loops are polynomial in the energies, but not in the space momenta, so the integrals are ill defined.

If we could dimensionally continue the energies, the mentioned integrals would vanish. However, since we work in Minkowski spacetime, a continuation of the energy clashes against assumptions (2.5) and (2.11). One may think of using an *ad hoc* regularization just for the energy integrals, but the removal of that regulator is very problematic.

We avoid these difficulties by working in the special gauge. That is to say, we use the propagators (4.7), possibly with the modification (4.10), and keep $\lambda > 0$. We know from section 4 that the assumptions that lead to the pseudounitarity equation are satisfied, the cut propagators being $2\theta(E)$ or $2\theta(-E)$ times

$$\begin{aligned}
&\mathrm{Im}\left[i\left\langle A^0 A^0\right\rangle_0\right] = -\pi\delta(\lambda E^2 - \mathbf{k}^2 - m_\gamma^2) = -\mathrm{Im}\left[i\left\langle C\bar{C}\right\rangle_0\right], \qquad \mathrm{Im}\left[i\left\langle A^i A^0\right\rangle_0\right] = 0, \\
&\mathrm{Im}\left[i\left\langle A^i A^j\right\rangle_0\right] = \pi\delta(E^2 - \mathbf{k}^2 - \mu^2 - m_\gamma^2)\left(\delta^{ij} - \frac{k^i k^j}{\mathbf{k}^2+\mu^2}\right) \\
&\qquad + \pi\sum_{n=0}^{N}\delta(\lambda E^2 - r_n\mathbf{k}^2 - r_n\mu^2 - m_\gamma^2)\sum_{q=n}^{N}\left(\frac{m_\gamma^2}{\mathbf{k}^2+\mu^2}\right)^q c_{nq}(\lambda)\frac{k^i k^j}{\mathbf{k}^2+\mu^2}.
\end{aligned} \tag{6.1}$$

In addition to the physical degrees of freedom, we have unphysical ones, which are the ghosts C and $\bar{C}$, the temporal component A_0 of the gauge field and the longitudinal component $\mathbf{A}_\parallel$. To prove that the pseudounitarity equation turns into the unitarity equation, we must prove that the unphysical degrees of freedom do not contribute, when the external legs are all physical.

Consider the cutting equation (2.14) associated with some Feynman diagram and assume that the external legs are physical. Focus on a cut diagram G_C and observe that the incoming and outgoing legs may be crossed by the cut or not, as shown in figure (2.12). Denote the total energy of the uncut (cut) incoming and outgoing legs by E_i (E_i') and E_o (E_o'), respectively. Denote the total energy of the cut legs by E_c, which is also equal to the total incoming energy $E_i + E_i'$ and the total outgoing energy $E_o + E_o'$.

An internal cut leg may give contributions of three types, with dispersion relations

$$DR_1 : E^2 = \frac{1}{\lambda}(\mathbf{k}^2 + m_\gamma^2), \qquad DR_2 : E^2 = \mathbf{k}^2 + \mu^2 + m_\gamma^2, \qquad DR_3 : E^2 = \frac{1}{\lambda}(r_n\mathbf{k}^2 + r_n\mu^2 + m_\gamma^2). \tag{6.2}$$

Write G_C as a sum $\sum_I G_C^{(I)}$, where $G_C^{(I)}$ is such that each internal cut leg contributes by means of one dispersion relation. Decompose the energy E_c of $G_C^{(I)}$ as the energy $E_i' + E_o'$ of the external cut legs, plus the energy E_λ of the internal cut legs that contribute by means of DR_1 or DR_3, plus the energy E_2 of the internal cut legs that contribute by means of DR_2.

Formulas (6.2) imply $E_\lambda \geqslant m_\gamma/\sqrt{\lambda}$. Thus, whenever DR_1 or DR_3 contribute, the total energy $E_{\text{tot}} = E_i + E_i' = E_o + E_o' = E_c$ satisfies the inequality

$$E_{\text{tot}} \geqslant \frac{m_\gamma}{\sqrt{\lambda}}. \tag{6.3}$$

Assume that m_γ is fixed and nonvanishing. The total energy E_{tot} has a given, λ-independent value. Then, when λ is small enough, the condition (6.3) cannot be satisfied. This means that for any diagram with physical external legs, any configuration of external energies and space momenta and any nonvanishing m_γ, the dispersion relations DR_1 and DR_3 of (6.2) cannot contribute to the cuts of formula (2.14), if λ belongs to the interval $(0, m_\gamma^2/E_{\text{tot}}^2)$.

For such values of λ, we can ignore DR_1 and DR_3 and the cut propagators (6.1) effectively become

$$\begin{aligned}
\text{Im}\left[i\left\langle A^0A^0\right\rangle_0\right] &= \text{Im}\left[i\left\langle C\bar{C}\right\rangle_0\right] = \text{Im}\left[i\left\langle A^iA^0\right\rangle_0\right] = 0, \\
\text{Im}\left[i\left\langle A^iA^j\right\rangle_0\right] &= \pi\delta(E^2 - \mathbf{k}^2 - \mu^2 - m_\gamma^2)\left(\delta^{ij} - \frac{k^ik^j}{\mathbf{k}^2 + \mu^2}\right).
\end{aligned}$$

At this point, we can take the limit $\mu \to 0$, which gives

$$\text{Im}\left[i\left\langle A^0A^0\right\rangle_0\right] = \text{Im}\left[i\left\langle C\bar{C}\right\rangle_0\right] = \text{Im}\left[i\left\langle A^iA^0\right\rangle_0\right] = 0, \quad \text{Im}\left[i\left\langle A^iA^j\right\rangle_0\right] = \pi\delta(E^2 - \mathbf{k}^2 - m_\gamma^2)\Pi^{ij}. \tag{6.4}$$

The limit $\mu \to 0$ is regular, not only in the cut propagators with $\lambda \in (0, m_\gamma^2/E_{\text{tot}}^2)$, but also in the uncut ones (4.10) and in the behavior (4.13), because as long as $m_\gamma \neq 0$ there are no infrared divergences in the individual diagrams of the cutting equation. Note that we cannot take $\mu \to 0$ for generic values of λ, because the right-hand side of (6.1) would give infrared divergences at $\mathbf{k} = 0$.

Next, we take the limit $\epsilon \to 0$ on the cutting equation, after grouping the cut diagrams as explained in subsection 3.3, so that each group is regular in the limit.

Formula (6.4) shows that, in the end, the unphysical degrees of freedom disappear from the cuts. This is a good starting point, but not the end of the story, because we must eventually send the fictitious mass m_γ to zero, which makes the interval $(0, m_\gamma^2/E_{\text{tot}}^2)$ disappear. To overcome this difficulty, we use the following trick. Before sending m_γ to zero, we analytically continue the cutting equation (2.14) in λ from the interval $(0, m_\gamma^2/E_{\text{tot}}^2)$ on the real axis to the complex plane.

Consider an uncut Feynman diagram G. At $\mu = 0$, the propagators of the unphysical degrees of freedom have denominators $\lambda E^2 - r_n \mathbf{k}^2 - m_\gamma^2$, so their poles move away from the real axis during the analytic continuation from real λ to complex λ. The continuation is consistent if we deform the integration contours so that the poles are never crossed. The continued λ will be called either ζ or ζ^*, leading to the continued diagrams G_ζ and G_{ζ^*}, respectively.

To better understand what we are doing, write the analytically continued (renormalized) cutting equation (2.14) in the form

$$G_\zeta + G_{\zeta^*}^\dagger = -\sum_{\text{proper cuttings}} \tilde{G}_\zeta \mathcal{C} \tilde{G}'^\dagger_{\zeta^*}, \tag{6.5}$$

where $\tilde{G}_\zeta$ and $\tilde{G}'^\dagger_{\zeta^*}$ denote the analytic continuations of the amputated subdiagrams identified by the shadowed and unshadowed sides of the cuts and $\mathcal{C}$ collects the propagators of the cut lines. The integration on the momenta of the cut legs is understood.

The diagrams $G_{\zeta^*}^\dagger$ and $\tilde{G}'^\dagger_{\zeta^*}$ of the shadowed regions are not the conjugates of the diagrams G_ζ and $\tilde{G}'_\zeta$ of the unshadowed regions at $\zeta \neq \lambda$, but the conjugates of the diagrams G_{ζ^*} and $\tilde{G}'_{\zeta^*}$. Indeed, the continued equation (6.5) depends only on ζ and not on ζ^*.

Note that the propagators (4.7) do not satisfy the KL spectral representation (3.5) for complex λ. That is why we cannot obtain equation (6.5) as a standard cutting equation. We must start from (2.14), specialize λ to the interval $(0, m_\gamma^2/E_{\text{tot}}^2)$, send μ to zero and then analytically continue from λ to ζ. An obvious, but important fact is that the cut propagators of $\mathcal{C}$ are independent of ζ, because they are those of the physical degrees of freedom.

Now we study the properties of the analytic continuation from λ to ζ. The singularities of a Feynman diagram arise when the contours are pinched[4] and can be studied by means of

[4]Strictly speaking, we must first decompose the integral I into a sum $\sum_i I_i$ of integrals I_i, each of which admits a domain of convergence D_i in the sense of the dimensional regularization. Then, we can study I_i in D_i and close

the Landau equations [18]. In our case, the Landau equations are algebraic and imply that the singularities in the ζ plane are a finite number of branch points.

Consider all the diagrams G_ζ, G_{ζ^*}, $\tilde{G}_\zeta$ and $\tilde{G}_{\zeta^*}$ at once and the trajectories described by the locations of their singularities when m_γ is small and tends to zero. There exists a $\bar{m}_\gamma$ such that, for every nonvanishing m_γ smaller than $\bar{m}_\gamma$ the following three facts hold: (*i*) the trajectories are continuous; (*ii*) those which do not coincide do not intersect; and (*iii*) there exists a positive $\lambda_{\text{min}} \leqslant m_\gamma^2/E_{\text{tot}}^2$ such that no singularities occur in the interval $(0, \lambda_{\text{min}})$ on the real axis.

Define the domain U as the complex plane $\mathbb{C}$ minus the trajectories with $m_\gamma \in (0, \bar{m}_\gamma)$. Consider a simply connected open subset V of U (we can take a disk, for simplicity). For any $m_\gamma \in (0, \bar{m}_\gamma)$, it is possible to analytically continue the cutting equation (2.14) from $(0, \lambda_{\text{min}})$ to V. To achieve this goal, it is sufficient to identify a path γ that connects the interval $(0, \lambda_{\text{min}})$ to some $z \in V$ and does not cross any singularity. When m_γ decreases, the path γ can be continuously deformed to avoid the singularities. So doing, the continued equation (6.5) holds in V at $m_\gamma = 0$.

Now we come to the physical amplitudes. We use gauge independence, following the recent treatment of ref. [17], which covers all gauge theories, including the nonrenormalizable ones and the potentially anomalous ones (as long as they are nonanomalous at one loop, by explicit cancellation, and at higher orders, by the Adler-Bardeen theorem [19]). The Hermitian conjugates are never involved in the arguments about gauge independence, so the results of [17] also apply to a complex ζ.

Start again from $m_\gamma \in (0, \bar{m}_\gamma)$. Pick any correlation function, make the analytic continuation from λ to ζ and identify the domain V. At nonzero m_γ, the equations of gauge dependence are violated. It is easy to show that the violations, which are due to insertions proportional to m_γ^2 [see e.g (4.8)], disappear when m_γ tends to zero. Consider an (off shell) Feynman diagram that contains such insertions. The limit $m_\gamma \to 0$ can generate divergences, but those divergences are always beaten by the multiplying factors of m_γ^2. For example, a single insertion may generate a logarithmic divergence $\sim \ln m_\gamma^2$ in $d = 4$, a powerlike divergence $\sim 1/m_\gamma$ in $d = 3$ and no singularity in $d > 4$. Two insertions may give a powerlike divergence $\sim m_\gamma^{d-6}$ in $d < 6$, a logarithmic one $\sim \ln m_\gamma^2$ in $d = 6$ and no singularity in $d > 6$, and so on. In the end, the violations are at worst multiplied by $m_\gamma^{d-2} \ln m_\gamma^2$ and m_γ^{d-2}, or products and powers of these expressions, and so vanish for $m_\gamma \to 0$. These properties are true even when we include the modifications of formula (4.10).

Thus, the equations of gauge dependence are satisfied after the limit $m_\gamma \to 0$ in V. They ensure [17] that the ζ dependence of the generating functional $\Gamma(\Phi, K, \zeta)$ of the one-particle irreducible correlation functions can be absorbed into a canonical transformation F_ζ (in the Batalin-

the contours of integration on the energies at infinity. So doing, we see that there are no end point singularities, but just pinching singularities.

Vilkovisky sense [16]) of the fields Φ and the sources K coupled to the Φ transformations, plus redefinitions of the couplings and the other parameters of Γ.

Finally, we put the external legs on shell. The infrared divergences generated by this operation are of a few universal types. The same types occur in different diagrams, so it is possible to identify combinations of cutting equations that are free of such divergences. In the rest of the discussion, we focus on such combinations, which have a well-defined on shell limit.

As said, the ζ dependence is encoded into a canonical transformation, plus redefinitions of the parameters. A canonical transformation maps a correlation function, evaluated up to some order within the truncation, onto a sum of correlation functions, each of which satisfies its own cutting equations (6.5). What is important for us is that the canonical transformation is trivial on the legs that are on shell. Indeed, the nonlinear terms of the field redefinition correspond to insertions of local composite fields. Once the diagram is amputated, those insertions get multiplied by inverse propagators, which vanish on shell.

In the end, the combinations of amplitudes that have a well-defined on shell limit are also independent of ζ (possibly after redefining the couplings and the other parameters). In particular, they can be trivially continued back from $\zeta \in V$ to real positive values λ. Moreover, by our construction, they satisfy the unitarity equation (3.4). This concludes the proof.

6.1 Examples

The argument involving the analytic continuation from λ to ζ is rather new, at least to our knowledge, so we give some examples in four dimensions to better visualize what happens.

The simplest situation is the one-loop self-energy of the four-dimensional φ^3 theory with propagator $iP_{\lambda,1,0}$ and coupling g. The diagram has the usual value apart from a rescaling of the energy by a factor $1/\sqrt{\lambda}$. Using the Feynman parameters, without including the combinatorial factor, we get

$$\frac{ig^2\Gamma\left(\frac{\varepsilon}{2}\right)}{(4\pi)^{(4-\varepsilon)/2}\sqrt{\lambda}}\int_0^1 \mathrm{d}x \left[x(1-x)(\mathbf{k}^2-\lambda E^2)+m_\gamma^2-i\epsilon\right]^{-\varepsilon/2}.$$

The analytic continuation in λ is done from the interval $(0, m_\gamma^2/E^2)$, where the right-hand side of (6.5) vanishes, because no physical degrees of freedom are present. The left-hand side of formula (6.5) is then

$$\frac{ig^2}{(4\pi)^2\sqrt{\zeta}}\int_0^1 \mathrm{d}x \ln\frac{x(1-x)(\mathbf{k}^2-\zeta E^2)+m_\gamma^2+i\epsilon}{x(1-x)(\mathbf{k}^2-\zeta E^2)+m_\gamma^2-i\epsilon}, \tag{6.6}$$

where we have taken the limit $\varepsilon \to 0$, for simplicity. Thus, the cutting equations imply that formula (6.6) should just give zero.

To prove this fact, it is sufficient to study the analytic continuation of the function

$$\Theta(z) \equiv \frac{1}{2\pi i}\ln\frac{z+i\epsilon}{z-i\epsilon}. \tag{6.7}$$

Clearly, $\Theta(z)$ vanishes for z real and positive, so the analytic continuation of $\Theta(z)$ from the positive real axis gives zero everywhere. Instead, the analytic continuation from the negative real axis gives $\Theta = 1$ everywhere. In some sense, the function $\Theta(z)$ is the complex θ function.

In the end, formula (6.6) turns into

$$-\frac{g^2}{8\pi\sqrt{\zeta}}\int_0^1 \mathrm{d}x\,\Theta(a(x)-\zeta),$$

where

$$a(x)=\frac{\mathbf{k}^2}{E^2}+\frac{m_\gamma^2}{E^2x(1-x)}>\frac{m_\gamma^2}{E^2}.$$

The analytic continuation from the interval $(0,m^2/E^2)$ of the real axis gives zero everywhere, as expected.

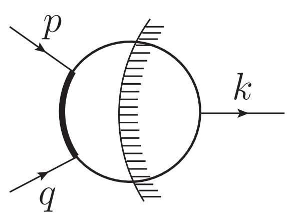
To show that the essential features just described survive when the λ is not just a simple rescaling of the energies, we consider a second example, which is a variant of the first one where the diagram remains the same and one internal leg keeps the propagator $iP_{\lambda,1,0}$, but the other internal leg gets the propagator $iP_{1,1,0}$. The right-hand side of (6.5) still vanishes for $\lambda < m_\gamma^2/E^2$.

For simplicity, we take a vanishing external space momentum. We find that formula (6.6) is replaced by

$$\frac{ig^2}{(4\pi)^2}\int_0^1 \frac{\mathrm{d}x}{\sqrt{x+\zeta(1-x)}}\ln\frac{(x+\zeta(1-x))\,m_\gamma^2-\zeta E^2x(1-x)+i\epsilon}{(x+\zeta(1-x))\,m_\gamma^2-\zeta E^2x(1-x)-i\epsilon}.$$

Again, the analytic continuation from $(0,m_\gamma^2/E^2)$ gives zero on the entire complex plane.

The third example is a one-loop three-point function, where two propagators are $iP_{1,1,0}$ and one is $iP_{\lambda,1,0}$, so that the right-hand side of (6.5) is nonzero. There is just one nontrivial cut diagram for $\lambda < m_\gamma^2/E_{\text{tot}}^2$, which is



where the arrows denote the energy flows and the solid line stands for the λ-dependent propagator. To evaluate this diagram, we assume that the external particle of momentum k is at rest and has mass M. We make no assumptions on the masses of the other external particles. Then we get

$$\frac{g^3\theta\,(M-2m_\gamma)}{16\pi M|\mathbf{p}|}\ln\frac{4m_\gamma^2+(M\sigma-2|\mathbf{p}|)^2-\zeta\left(M-2p^0\right)^2-i\epsilon}{4m_\gamma^2+(M\sigma+2|\mathbf{p}|)^2-\zeta\left(M-2p^0\right)^2-i\epsilon}$$

where $\sigma = \sqrt{1 - \frac{4m_\gamma^2}{M^2}}$. We see that the analytic continuation from λ to ζ is straightforward, as well as the limit $m_\gamma \to 0$.

These examples show that no big surprise occurs when we make the operations described in this section.

We emphasize that the proof we have given, being purely perturbative, does not deal with nonperturbative issues, such as confinement and chiral symmetry breaking, which affect the physical spectrum. It simply shows that the physical amplitudes that are built with the elementary fields satisfy the unitary cutting equation. The definition of asymptotic states is not necessary to make these statements meaningful. The true asymptotic states may be completely different from those suggested by the classical Lagrangian, as we know from QCD. We recall that, instead of dealing with the elementary fields, it is often convenient to study the correlation functions of gauge invariant composite fields and extract the S-matrix elements as residues of their poles [20].

7 Proof of unitarity in quantum gravity

In this section we generalize the proof of perturbative unitarity to quantum gravity (with vanishing cosmological constant).

First, we prove that quantum gravity also admits the special gauge, in arbitrary dimensions $d > 3$. The gauge-fixed Lagrangian is

$$\mathcal{L}_{\text{gf}} = -\frac{1}{2\kappa^{d-2}}\sqrt{|g|}R + \frac{1}{4\lambda_1\kappa^{d-2}}\mathcal{G}_0^2(g) - \frac{1}{4\lambda_2\kappa^{d-2}}\mathcal{G}_i^2(g) + \bar{C}_0\mathcal{G}_0(\overline{DC}) - \bar{C}_i\mathcal{G}_i(\overline{DC}), \tag{7.1}$$

where κ is a constant of dimension -1 in units of mass, C_μ and $\bar{C}_\mu$ are the ghosts and antighosts, respectively, the gauge-fixing functions $\mathcal{G}_0(g)$ and $\mathcal{G}_i(g)$ are linear in the metric tensor $g_{\mu\nu}$ and $\overline{DC}$ stands for $D_\mu C_\nu + D_\nu C_\mu$.

We start from the most general linear gauge-fixing functions with one derivative, which are

$$\mathcal{G}_0(g) = \alpha\partial_0 g_{ii} + \beta\partial_i g_{0i} + \chi\partial_0 g_{00}, \qquad \mathcal{G}_i(g) = \gamma\partial_0 g_{0i} + \delta\partial_i g_{00} + \xi\partial_i g_{jj} + \tau\partial_j g_{ij},$$

and determine the constants in front of the various terms as follows. First, we require that the propagators with an odd number of indices 0 vanish. Second, we simplify the double poles. Third, we eliminate the redundant constants and arrange the result in the most economic form. At the end, we find

$$\mathcal{G}_0(g) = \frac{\lambda}{2}\partial_0 g_{00} + \frac{1}{2}\partial_0 g_{ii} - \partial_i g_{0i}, \qquad \mathcal{G}_i(g) = -\lambda_1\partial_j g_{ij} + \frac{1}{2}\left(2\lambda_1 - 1\right)\partial_i g_{jj} + \lambda\partial_0 g_{0i} - \frac{\lambda}{2}\partial_i g_{00},$$

together with

$$\lambda_1 = \frac{\lambda(d-3)+d-1}{2(d-2)}, \qquad \lambda_2 = \lambda\lambda_1.$$

Only one gauge-fixing parameter, which we still call λ, does survive.

The ghost propagators are

$$\langle C^0\bar{C}^0\rangle_0 = -i\bar{P}_{\lambda,1}, \qquad \langle C^0\bar{C}^i\rangle_0 = \langle C^i\bar{C}^0\rangle_0 = 0, \qquad \langle C^i\bar{C}^j\rangle_0 = i\bar{P}_{\lambda,\lambda_1}\Pi^{ij} + i\bar{P}_{\lambda,1}\frac{k^ik^j}{\mathbf{k}^2}, \tag{7.2}$$

where $\bar{P}_{a,b} = 1/(aE^2 - b\mathbf{k}^2 + i\epsilon)$. The propagators of the fluctuations $h_{\mu\nu} = \kappa^{1-(d/2)}(g_{\mu\nu} - \eta_{\mu\nu})/2$ around flat space are

$$\begin{aligned}
\langle h_{00}h_{00}\rangle_0 &= \frac{d-3}{d-2}i\bar{P}_{\lambda,1}, \qquad \langle h_{00}h_{ij}\rangle_0 = \frac{\delta_{ij}}{d-2}i\bar{P}_{\lambda,1}, \\
\langle h_{0i}h_{0j}\rangle_0 &= -\frac{i\lambda_1}{2}\left(\bar{P}_{\lambda,\lambda_1}\Pi_{ij} + \bar{P}_{\lambda,1}\frac{k_ik_j}{\mathbf{k}^2}\right), \qquad \langle h_{00}h_{0i}\rangle_0 = \langle h_{0i}h_{jk}\rangle_0 = 0, \\
\langle h_{ij}h_{mn}\rangle_0 &= \frac{i\bar{P}_{1,1}}{2}\left(\Pi_{im}\Pi_{jn} + \Pi_{in}\Pi_{jm} - \frac{2}{d-2}\Pi_{ij}\Pi_{mn}\right) - \frac{\lambda}{\mathbf{k}^2}\frac{i\bar{P}_{\lambda,1}}{d-2}\left(\Pi_{ij}k_mk_n + k_ik_j\Pi_{mn}\right) \\
&\quad + \frac{\lambda i\bar{P}_{\lambda,\lambda_1}}{2\mathbf{k}^2}\left(\Pi_{im}k_jk_n + \Pi_{in}k_jk_m + \Pi_{jm}k_ik_n + \Pi_{jn}k_ik_m\right) + \lambda i\bar{P}_{\lambda,1}\frac{d-3}{d-2}\frac{k_ik_jk_mk_n}{(\mathbf{k}^2)^2}.
\end{aligned} \tag{7.3}$$

At $\lambda = 1$ the special gauge coincides with the de Donder one, which is the gravitational analogue of the Feynman gauge of Yang-Mills theories. In the limit $\lambda \to 0$ we get an analogue of the Coulomb gauge, different from the Prentki gauge.

To have control on the infrared divergences of the individual cut diagrams, we keep $\langle C^0\bar{C}^i\rangle_0 = \langle C^i\bar{C}^0\rangle_0 = \langle h_{00}h_{0i}\rangle_0 = \langle h_{0i}h_{jk}\rangle_0 = 0$ and replace the other propagators with

$$\begin{aligned}
\langle C^0\bar{C}^0\rangle_0 &= -iP_{\lambda,1,1}, \qquad \langle h_{00}h_{00}\rangle_0 = \frac{d-3}{d-2}iP_{\lambda,1,1}, \qquad \langle h_{00}h_{ij}\rangle_0 = \frac{\delta_{ij}}{d-2}iP_{\lambda,1,1}, \\
\langle h_{0i}h_{0j}\rangle_0 &= -\frac{i\lambda_1}{2\lambda}\left(Q_N(\lambda,s)\pi_{ij} + Q_N(\lambda,r)\omega_{ij}\right) = -\frac{\lambda_1}{2}\langle C^i\bar{C}^j\rangle_0, \\
\langle h_{ij}h_{mn}\rangle_0 &= \frac{iP_{1,1,1}}{2}\left(\pi_{im}\pi_{jn} + \pi_{in}\pi_{jm} - \frac{2}{d-2}\pi_{ij}\pi_{mn}\right) - \frac{iQ_N(\lambda,r)}{d-2}\left(\pi_{ij}\omega_{mn} + \omega_{ij}\pi_{mn}\right) \\
&\quad + \frac{i}{2}Q_N(\lambda,s)\left(\pi_{im}\omega_{jn} + \pi_{in}\omega_{jm} + \pi_{jm}\omega_{in} + \pi_{jn}\omega_{im}\right) + iQ_N(\lambda,r)\frac{d-3}{d-2}\omega_{ij}\omega_{mn},
\end{aligned}$$

where

$$\pi_{ij} = \delta_{ij} - \frac{k_ik_j}{\mathbf{k}^2 + \mu^2}, \qquad \omega_{ij} = \frac{k_ik_j}{\mathbf{k}^2 + \mu^2},$$

and the sequence $s = \{s_0, s_1, \ldots\}$ is related to $r = \{r_0, r_1, \ldots\}$ by the formula

$$s_n = \frac{\lambda(d-3) + r_n(d-1)}{2(d-2)}. \tag{7.4}$$

Moreover, we choose r_n such that $r_n \neq s_{n'}$ for every n and n', at λ small.

As in Yang-Mills theories, the propagators contain irregular terms. It can be shown that those terms satisfy a property analogous to (4.13). Precisely, they are equal to the product of projectors built with δ_{ij} and ω_{ij} times regular terms that factorize $N+1$ powers of m_γ^2. Note that the irregular denominators can now be as bad as $(\mathbf{k}^2+\mu^2)^2$. The relation (7.4) is crucial to cancel those, up to $\mathcal{O}(m_\gamma^{2N+2})$.

The powers m_γ^{2N+2} that factorize in front of the irregular terms lower the degree of divergence. If N is large enough, the irregular contributions to the Feynman diagrams are overall convergent within any given truncation. Thus, the locality of counterterms holds within the truncation, for the same reasons explained in the study of nonrenormalizable gauge theories. Moreover, for N sufficiently large the irregular parts of the propagators cannot generate contact terms. Then, contact terms can only come from the regular contributions and can be dealt with by means of the procedure explained in subsection 2.3.

The assumptions that are required to derive the pseudounitarity equation are satisfied at $m_\gamma \neq 0$, $\lambda > 0$. When λ is sufficiently small, the threshold for the production of the unphysical particles in the cuts is raised enough to get rid of all of them (once μ is sent to zero), for any given m_γ and any total energy of the incoming particles.

The operators projecting onto the physical degrees of freedom that propagate in the cuts can be read from formulas (7.2) and (7.3), which show that C^μ, $\bar{C}^\mu$, h_{00}, h_{0i}, h_{ii} and $\partial_j h_{ij}$ do not propagate for λ small. The external legs can be projected in a similar way. Precisely, it is sufficient to set the sources coupled to the external legs C^μ, $\bar{C}^\mu$, h_{00} and h_{0i} to zero and project the external legs h_{ij} by means of

$$\prod\nolimits_{ij,mn} \equiv \frac{1}{2}\left(\Pi_{im}\Pi_{jn} + \Pi_{in}\Pi_{jm} - \frac{2}{d-2}\Pi_{ij}\Pi_{mn}\right).$$

Every other argument of the previous section can be generalized straightforwardly.

8 Conclusions

In this paper we worked out a proof of perturbative unitarity that is more economical and general than the ones available in the literature and applies to renormalizable and nonrenormalizable gauge theories and quantum gravity, in arbitrary dimensions d greater than 3. With an eye on future generalizations, we searched for the minimum assumptions that lead to the various equations involved in the proof, which are the largest time equation, the cutting equations and the pseudounitarity equation. The minimum assumptions are actually very restrictive and imply a general Källén-Lehman spectral representation, even if it is not assumed from the start. The pseudounitarity equation turns into the unitarity equation when the incoming and outgoing particles, as well as the particles propagating in the cuts, can be projected onto a subspace of physical states.

We also filled some gaps that exist in the current literature, as in the treatment of contact terms.

The simplest proof of perturbative unitarity is available in QED by working directly in the Coulomb gauge. Unfortunately, a similar strategy cannot be pursued in non-Abelian gauge theories and gravity. In those cases, we identified a special gauge that fulfills all the assumptions and has several other virtues. It depends on a unique gauge-fixing parameter λ and interpolates between the Feynman gauge ($\lambda = 1$) and the Coulomb gauge ($\lambda = 0$). When the gauge fields are given fictitious masses to regulate the on shell infrared divergences, the threshold for the production of the unphysical particles in the cuts grows while λ becomes small. Eventually, it projects the unphysical particles away. Thus, there exists an interval of values of λ where the pseudounitarity equation turns into the unitarity equation. To recover gauge invariance, the fictitious masses must be removed. This can be achieved without jeopardizing the unitarity equation by making the analytic continuation in λ.

Various theories do not obey the assumptions that lead to the cutting equations. Examples are the local higher-derivative theories whose propagators have poles outside the real axis. Other examples are the nonlocal theories of gauge fields and gravity formulated in refs. [9]. Indeed, if the vertices are not localized in time, the largest time equation cannot be derived, because no "largest time" can be identified in the analogue of the raw diagram $F(x_1^0, \cdots, x_n^0)$. Moreover, if the vertices are nonlocal in space, the contact terms of subsection 2.3 cannot be treated as explained there. For these reasons, the consistency of the theories of refs. [9] remains an open problem, even if their propagators have no poles on the complex plane besides the graviton one.

These remarks also suggest an unforseen connection between unitarity and locality that is worth further investigation.

Acknowledgments

I am grateful to U. Aglietti, M. Bochicchio, R. Ferrari, M. Mintchev, G. Paffuti and M. Piva for helpful discussions.

References

[1] G. 't Hooft and M. Veltman, One-loop divergences in the theory of gravitation, Ann. Inst. Poincaré, 20 (1974) 69;

P. van Nieuwenhuizen, On the renormalization of quantum gravitation without matter, Ann. Phys. (NY) 104 (1977) 197;

M.H. Goroff and A. Sagnotti, The ultraviolet behavior of Einstein gravity, Nucl. Phys. B 266 (1986) 709;

A. van de Ven, Two loop quantum gravity, Nucl. Phys. B 378 (1992) 309.

[2] S. Weinberg, Ultraviolet divergences in quantum theories of gravitation, in *An Einstein Centenary Survey*, edited by S. Hawking and W. Israel, Cambridge University Press, Cambridge 1979, p. 790.

[3] D. Anselmi, Properties of the classical action of quantum gravity, J. High Energy Phys. 05 (2013) 028, 13A2 Renormalization.com and arXiv:1302.7100 [gr-qc].

[4] K.S. Stelle, Renormalization of higher derivative quantum gravity, Phys. Rev. D 16 (1977) 953;

E. S. Fradkin and A. A. Tseytlin, Renormalizable asymptotically free quantum theory of gravity, Nucl. Phys. B 201 (1982) 469.

[5] R.E. Cutkosky, Singularities and discontinuities of Feynman amplitudes, J. Math. Phys. (NY) 1 (1960) 429.

[6] M. Veltman, Unitarity and causality in a renormalizable field theory with unstable particles, Physica 29 (1963) 186.

[7] G. 't Hooft, Renormalization of massless Yang-Mills fields, Nucl. Phys. B 33 (1971) 173;

G. 't Hooft, Renormalizable Lagrangians for massive Yang-Mills fields, Nucl. Phys. B 35 (1971) 167.

[8] C. Becchi, A. Rouet and R. Stora, The Abelian Higgs Kibble model, unitarity of the S-operator, Phys. Lett. B 52 (1974) 344;

C. Becchi, A. Rouet and R. Stora, Renormalization of gauge theories, Ann. Phys. (NY) 98 (1976) 287;

G. Curci and R. Ferrari, An alternative approach to the proof of unitarity for gauge theories, Nuovo Cimento A 35 (1976) 273;

T. Kugo and I. Ojima, Local covariant operator formalism of non-Abelian gauge theories and quark confinement problem, Prog. Theor. Phys. Suppl. 66 (1979) 1;

C. Becchi, "Lectures on the renormalization of gauge theories," in Les Houches 1983, Proceedings, Relativity, Groups and Topology, II, p. 787;

for a recent investigation, see R. Ferrari and A. Quadri, Physical unitarity for massive non-Abelian gauge theories in the Landau gauge: Stueckelberg and Higgs, J. High Energy Phys. 11 (2004) 019 and hep-th/0408168.

[9] Yu.V. Kuz'min, Finite nonlocal gravity, Sov. J. Nucl. Phys. 50 (1989) 6 [Yad. Fiz. 50 (1989) 1630];

E.T. Tomboulis, Super-renormalizable gauge and gravitational theories, arXiv:hep-th/9702146;

L. Modesto, Super-renormalizable quantum gravity, Phys. Rev. D86 (2012) 044005 and arXiv:1107.2403 [hep-th];

T. Biswas, E. Gerwick, T. Koivisto and A. Mazumdar, Towards singularity and ghost free theories of gravity, Phys. Rev. Lett. 108 (2012) 031101 and arXiv:1110.5249 [gr-qc];

L. Modesto, Finite quantum gravity, arXiv:1305.6741 [hep-th];

T. Biswas, A. Conroy, A. S. Koshelev and A. Mazumdar, Generalized ghost-free quadratic curvature gravity, Classical Quantum Gravity 31 (2014) 015022 and arXiv:1308.2319 [hep-th].

L. Modesto and L. Rachwal, Super-renormalizable and finite gravitational theories, Nucl. Phys. B 889 (2014) 228 and arXiv:1407.8036 [hep-th].

[10] G. 't Hooft and M. Veltman, *Diagrammar*, report No. CERN-73-09, available at this link.

[11] M. Veltman*, Diagrammatica. The path to Feynman rules* (Cambridge University Press, New York, 1994).

[12] D. Anselmi, Weighted power counting and chiral dimensional regularization, Phys. Rev. D 89 (2014) 125024, 14A2 Renormalization.com and arXiv:1405.3110 [hep-th].

[13] D. Anselmi, Weighted scale invariant quantum field theories, J. High Energy Phys. 02 (2008) 051, 08A1 Renormalization.com and arXiv:0801.1216 [hep-th].

[14] F. Bloch and A. Nordsieck, Note on the radiation field of the electron, Phys. Rev. 52 (1937) 54;

T. Kinoshita, Mass singularities of Feynman amplitudes, J. Math. Phys. 3 (1962) 650;

T. D. Lee and M. Nauenberg, Degenerate systems and mass singularities, Phys. Rev. 133 (1964) B1549;

S. Weinberg, Infrared photons and gravitons, Phys. Rev. 140 (1965) B516.

[15] See, for example, J.C. Collins, *Renormalization* (Cambridge University Press, Cambridge, UK (1984), Sects. 5.2 and 5.8.

[16] I.A. Batalin and G.A. Vilkovisky, Gauge algebra and quantization, Phys. Lett. B 102 (1981) 27-31;

I.A. Batalin and G.A. Vilkovisky, Quantization of gauge theories with linearly dependent generators, Phys. Rev. D 28 (1983) 2567; 30 (1984) 508(E);

See also S. Weinberg, *The Quantum Theory of Fields*, vol. II, Cambridge University Press, Cambridge (1995).

[17] D. Anselmi, Ward identities and gauge independence in general chiral gauge theories, Phys. Rev. D 92 (2015) 025027, 15A1 Renormalization.com and arXiv:1501.06692 [hep-th].

[18] C. Itzykson and J.B Zuber, *Quantum Field Theory* (McGraw-Hill Inc., New York (1980), Sect. 6.3.

[19] S.L. Adler and W.A. Bardeen, Absence of higher order corrections in the anomalous axial vector divergence, Phys. Rev. 182 (1969) 1517;

for a review untill 2004, see S.L. Adler, Anomalies to all orders, in "*Fifty Years of Yang-Mills Theory*", edited by G. 't Hooft (World Scientific, Singapore, 2005), p. 187-228, and arXiv:hep-th/0405040;

D. Anselmi, Adler-Bardeen theorem and manifest anomaly cancellation to all orders in gauge theories, Eur. Phys. J. C74 (2014) 3083, 14A1 Renormalization.com and arXiv:1402.6453 [hep-th];

D. Anselmi, Adler-Bardeen theorem and cancellation of gauge anomalies to all orders in nonrenormalizable theories, Phys. Rev. D 91 (2015) 105016, 15A2 Renormalization.com and arXiv:1501.07014 [hep-th].

[20] S. Weinberg, *The Quantum Theory of Fields*, Vol. I (Cambridge University Press, Cambridge, 1995), Sect. 10.2, p. 428.