Cancellation of the Infrared Singularity through the Unitarity Relation

Nobuo YOSHIDA

Department of Physics, University of Tokyo, Tokyo 113

(Received January 16, 1981)

It is shown that the infrared singularities (both soft and collinear divergences) in usual field theories including QCD are cancelled among topologically equal “double cut diagrams” by appropriate summation over initial and/or final cuts. The basic tools of the analysis are the largest time equation and the power counting method in the space consisting of time plus three momentum components.

§ 1. Introduction

Although the origin of the infrared (IR) singularity in the perturbative calculation has been brought to light by many early works, there still remains an obscurity about the IR finite inclusive cross section. The general theorem by Lee and Nauenberg\textsuperscript{1,2)} says that we can eliminate the IR singularity order by order in the coupling constant by summing over all degenerate initial and final states in the cross section, but this theorem does not make clear among which diagrams the cancellation occurs. Until recently, so far as soft divergences are concerned, it had been supposed that the so-called Bloch-Nordsieck (BN) inclusive cross section is finite as a consequence of the cancellation of the IR singularities between virtual corrections and real emissions of soft particles,\textsuperscript{1,3)} at least in gauge theories.\textsuperscript{4)} However, by a direct calculation Doria et al.\textsuperscript{5)} showed that this cancellation does not hold in non-forward quark-quark scattering and the Drell-Yan process. Hence a simple and certain method for obtaining a finite cross section is still lacking.

The chief purpose of this paper is to show, as a possible answer to this problem, that the soft and collinear divergences in the cross section will disappear after appropriate summation over initial and/or final cuts for each set of topologically equal double cut diagrams.\textsuperscript{**} Here a “double cut diagram” means a vacuum-bubble-like diagram which is obtained by connecting external initial legs of a usual cut diagram.\textsuperscript{4,5)} Although in Ref. 4\textsuperscript{)} this diagram is treated in the four

\textsuperscript{*} The proof in the original article is somewhat unassuring for theories with complicated Hamiltonian, but the validity of the theorem is indubitable from the general consideration of the unitarity. See the following discussion.

\textsuperscript{**} For the case with collinear divergences in QCD this is true only when one works in the physical gauge.
momentum space, it is convenient to work in the $t \cdot p$ (time plus three momentum) space.

The basic ingredient of the proof is the fact that a time-ordered double cut diagram with the IR singularity can be always compensated by use of the "largest time equation". The reason for the cancellation is as follows: The IR singularities originate from the subdiagrams which do not suffer the damping when going to plus or minus infinity of time. Since such a singular subdiagram necessarily contains the latest or earliest time vertex which represents the transition between degenerate states, we can always apply the largest time equation to cancel out the singularity without mixing the distinguishable processes. Thus the cancellation of the IR singularity is a direct consequence of the unitarity relation and holds in any field theory in which massless particles do not interact via (bare or induced) super-renormalizable couplings.

The most interesting application is that to QED, on which we are to concentrate in this paper. Since the cancellation is only owing to the unitarity relation and holds in any colour channel (as can be seen later), we need not elaborate a theory which treats the colour singlet bound state of quarks and gluons.

The construction of this paper is as follows. First, in § 2 the IR singular subdiagrams in a time-ordered double cut diagram are determined. Section 3 is devoted to the proof of the cancellation of the IR singularity. The correctness of the procedure roughly stated above is assured by power counting. Section 4 contains the physical interpretation of this cancellation.

§ 2. Determination of IR singular subdiagrams

The reason why soft and collinear divergences appear can be most easily understood in the time-ordered perturbation theory. As an example let us consider one photon emission (with three momentum $k$) from an electron (with $p$). The energy difference in this transition is:

$$\Delta E = \sqrt{(p - k)^2 + m^2} + |k| - \sqrt{p^2 + m^2},$$

where $m$ is the mass of the electron. Usually $\Delta E$ is non-zero, and owing to the oscillation factor, $\exp(\pm i\Delta E t)$, the transition damps when $t \to \pm \infty$. The problem occurs when (i) $|k| = 0$, and (ii) $k \parallel p$ if $m = 0$. In these cases $\Delta E$ becomes zero, and the emission persists for all times. In such a situation the time integration for $t \to \pm \infty$ may yield a divergence unless some suppression mechanism works.

*) We ignore the case where the energy difference happens to be zero without soft nor collinear momenta, since such diagrams are suppressed by the phase space as is easily seen.
Thus the origin of the IR singularity can be captured as the singular configuration of a time-ordered diagram in which non-oscillating vertices are at infinity of time.

To discuss more precisely it is necessary to determine the IR singular subdiagrams. This is the subject of this section. This paper is chiefly devoted to QCD, although the following discussion can be applied, with slight modifications, to usual field theories which suffer the IR singularity such as massive or massless QED, the massless $\phi^4$ theory in the six dimension and the massless pseudoscalar theory interacting with fermions via Yukawa couplings. The exceptions are theories in which massless particles interact via (bare or induced) superrenormalizable couplings. The problem of these theories is that the effective vertices with less than $2d/(d-2)$ legs, where $d$ is the space-time dimension, are not suppressed in the IR region. In the following the characteristic feature in gauge theories is stated in the square bracket.

As the first step we separate the IR and non-IR regions in a double cut diagram. The criterion of the separation is quite arbitrary. For example, we may introduce the cutoff in the spatial or transversal momentum region. Since the IR singularities originate from the behavior in the large time scale, it is convenient to localize any subdiagram in the non-IR region through the partial integration. This procedure, which is afterwards referred to as "skeleton expansion", induces the effective vertices not included in the Lagrangian.

After the skeleton expansion the double cut diagram represents a product of two point functions with non-IR corrections, effective vertex functions, and delta functions of energy-momentum conservation,

$$\int \prod_i dk_i \prod \Gamma'(k_i) \delta\left(\sum_j k_j\right),$$

(Here $i$ denotes an internal line and $\alpha$ a vertex.) $D^\alpha'(k)$ is chosen among the Feynman propagator $D^F(k)$, the anti-Feynman propagator $D^{F}(k)$, the cut propagator with positive energy $D^{+}(k)$, and the cut propagator with negative energy $D^{-}(k)$, according to how the unitarity cut goes across the diagram.

Here, for purely technical reason the momentum assignment is kept unchanged for each topological set of double cut diagrams whichever line is cut. The observed momentum is introduced by limiting the integration region of three momenta on the hypersurface $C$.

It is worthwhile noting that we may use the "bare" expression for $D^\alpha$'s. This can be seen from the Källen-Lehmann representation,

$$D^\alpha(k^2) = \int ds D_{\text{bare}}(k^2 ; s) \rho(s),$$

where $\rho(s)$ is the spectral function in the skeleton expansion diagram. Since $\rho(s)$ does not include the corrections in the IR region, only the single particle...
state contributes to the IR singularity, so that we may replace $D^\lambda(k)$ by the bare expression.

In the above manipulation other possible degrees of freedom than momenta are ignored. Let me briefly comment on this point.

(i) We may sum over unphysical degrees of freedom such as longitudinal polarizations either in a cut or uncut propagator. The cancellation among unphysical contributions is assured by other mechanism$^9$ and we need not care about it.

(ii) Flavors or helicities are set to be definite for each topological set of double cut diagrams. This is allowed because they are not changed by the interaction in the IR regions.

(iii) Colour charges are IR sensitive in a sense that they vary through the soft or collinear interactions. Hence if we sum over colour charges in a certain line of a double cut diagram, we should sum even if the line is cut so that it represents the real state.

Following the arguments in the first paragraph of this section, let us Fourier transform the energy components into the time coordinates. (The spatial momenta are left as they are for two reasons: First in this representation the power counting can be carried out more easily, and second the integrand is a regular function, i.e., integrable in the finite region, so that the singularity originates only from the infinite volume of the integration region.) By this procedure Eq. (1) is transformed into the following form:

$$
\int \prod \frac{d^3k_i}{2\pi^3} \int \frac{dt_a}{\sqrt{t}} \prod \frac{D^\lambda(A_{t_i}, k_i)}{2\pi^3} \times \prod \Gamma_{t_a}(t_a, \sum_k k_a) \delta(\sum_k k_a), \tag{2}
$$

where $A_{t_i}$ represents the time interval between two end points of the propagator. Since the translational invariance along the time is assured, we may set the vertex function including the hard interaction at the origin of time. (Here the cluster decomposable interactions are ignored, although we can deal with them also.)

Now we proceed to determine which diagram really gives rise to the divergences. Although the IR singularities originate from non-oscillating subdiagrams, we know that such subdiagram does not always cause the divergence. In fact, in most cases the power suppression from the phase space prevents the integration from blowing up. Thus what is necessary is to compare the phase space volume and the degree of divergence. This can be done with the help of power counting method. As a whole this method is the $t$-$p$ space version of what was adopted by Sterman$^8$ or Ellis et al.$^{10}$ The setting of the subject is: If all momenta in some subdiagram become soft or collinear in a power of, say, $\lambda$, what
power behavior does the contribution of the diagram in some time ordering sector show?

In order to investigate the power behavior, the following power counting rules are adopted.

(i) In the typical term of a propagator, \( 1/\omega e^{+i\omega t} \) where \( \omega = |k| \) for massless particles and \( \sqrt{k^2 + m^2} \) for massive particles, only \( \omega \) is counted as a propagator factor.

(ii) The time dependent exponential factor, which is definite for each time ordered diagram, is counted as a vertex factor. To estimate the power behavior after time integration the following formula is useful,

\[
\int_0^\infty dt e^{\pm i\Delta E t} = \frac{1}{2} e^{\pm i\Delta E t} \left( \delta(\Delta E) \pm \frac{i}{\pi} P \frac{1}{\Delta E} \right).
\]

If \( \Delta E \to 0 \) like the non-oscillating vertex case, the r.h.s. behaves in \( O(\Delta E^-1) \). This factor gives the worst power behavior, and will be referred to as the "singular vertex factor". Some examples of vertices that do not suffer the damping due to the oscillation factor and yield the singular vertex factor are shown in Fig. 1. (Note that this factor is in \( O(k) \) in the soft divergence case, and in \( O(kc^2) \) in the collinear divergence case.)

(iii) The following factors are to be associated with each vertex and called collectively an ordinary vertex factor:

a) momentum factor due to the derivative couplings.

b) Dirac spinors from the numerator of fermion propagators,

\[
\gamma = \sum_k \left[ u(p, h) \bar{u}(p, h) \delta(p_\mu) - \bar{v}(-p, h) v(-p, h) \delta(-p_\mu) \right].
\]

c) polarization vectors from the transverse vector boson propagators,

\[
d_{\mu\nu}(p) = \sum_k \epsilon_{\mu\nu}(p, h) \epsilon(p, h).
\]

* If the IR singularity remains even after some subdiagram is shrinking to a point, we are to face with the accidental zero, \( \sum \langle \Delta E_n \rangle = 0 \), in the iterative time integration of the original diagram. This case is, however, included in another skeleton expansion diagram.
Now let us investigate the power behavior of cases with soft or collinear divergences, respectively.

(A) Soft Divergences

Let all the momenta of a subdiagram go to zero as $k = \lambda x$ and $\lambda \to 0$ with $x$ fixed.

The scaling behavior of each factor is as follows:

- a propagator factor: $1/\omega \sim O(\lambda^{-1})$,
- momentum integration: $d^3 k \sim O(\lambda^3)$,
- a delta function: $\delta(\Sigma k_i) \sim O(\lambda^{-3})$,
- a singular vertex factor
  - of a non-oscillating vertex: $O(\lambda^{-1})$,
  - of an oscillating vertex: $O(1)$,
- an ordinary vertex factor
  - of a three point vertex: $O(\lambda)$,
  - of other vertices: $O(1)$.

The $O(\lambda)$ behavior of the ordinary vertex factor of a three point vertex is owing to the derivatives or the spinor factor. (The vertex factor of the massless soft fermions like $\bar{u}(k)\gamma_\mu u(k)$ behaves in $O(\lambda)$.)

In gauge theories it is necessary to consider the instantaneous interactions if one works in some non-covariant gauge. Fortunately, however, the following result is unchanged. For instance, in the Coulomb gauge the soft Coulomb or ghost propagator behaves in $O(\lambda^{-2})$, but the singular vertex factor yield the $O(\lambda^{-1})$ behavior for the set of all equal time vertices, so that the total power behavior does not change.

Now we are to consider a soft subdiagram which contains $I$ internal lines and $N_n$ $n$-point effective vertices, and which attaches to the rest of the whole diagram via $H_n$ effective vertices with $n$ soft legs. Note that in this manipulation we may make any internal line of a double cut diagram soft.

The total power behavior of the subdiagram is:

$$(3-1)I - \sum_n 3N_n - \sum_n (N_n^{\text{non}} + H_n^{\text{non}}) + N_3$$

$$= (N_3 - N_3^{\text{non}}) + \sum_{n \geq 4} ((n-3)N_n - N_n^{\text{non}}) + \sum_{n \geq 1} (nH_n - H_n^{\text{non}}).$$

The suffix “non” means that the vertex is non-oscillating.

From this equation we can easily see that the degree of convergence becomes positive unless the subdiagram has only non-oscillating three or four point effective vertices and attaches to the rest of the diagram via non-oscillating vertices with one soft leg. This condition implies that the soft subdiagram should not attach to the “hard” interaction part. The important point is that all vertices should be three or four point couplings and do not suffer the damping due
Cancellation of the Infrared Singularity

(B) Collinear Divergences

In a similar way we can treat the case with collinear divergences. In this case the gauge choice becomes essential, and the most convenient gauge is the Coulomb gauge. Fortunately we may contract the Coulomb and ghost propagators to induce effective vertices, because they are instantaneous.] As the momentum $k = sp + \mathbf{h}$ goes parallel to $p$ in such a way that

$$k_T = \lambda x_T, \lambda \to 0 \text{ with } x_T \text{ fixed},$$

then each factor scales like these:

- a propagator factor: $1/\omega \sim O(1)$,
- momentum integration: $d^3 k = \pi |p| dx d\mathbf{k}_\perp \sim O(\lambda^2)$,
- a delta function: $\delta(\sum \mathbf{k}_\perp) \sim O(\lambda^{-2})$,
- a singular vertex factor
  - of a non-oscillating vertex: $O(\lambda^{-2})$,
  - of an oscillating vertex: $O(1)$,
- an ordinary vertex factor
  - of a three point vertex: $O(\lambda)$,
  - of other vertices: $O(1)$,
- momentum conservation: $\delta(p - \sum \mathbf{k}_i) \sim O(\lambda^{-2})$.

The last condition is necessary because $p$ must be determined from the observed momentum in some way or other. (If we do not observe and integrate the momentum, we obtain the finite result as is easily seen.) The $O(\lambda)$ behavior of the ordinary vertex factor of a three point vertex is due to the helicity conservation. [In gauge theories this behavior is assured by the transversality of the physical polarization. In another gauge we have to sum up the gauge invariant set in order to obtain the finite result.]

The total degree of convergence becomes:

$$2I - 2 \sum_n N_n - 2 \sum_n N_n^{(00)} + N_3 - 2$$

$$= 2(N_3 - N_3^{(00)}) + \sum_{n \geq 4} ((n - 2)N_n - 2N_n^{(00)}).$$

Here a subdiagram with only two legs, namely a two particle reducible diagram, is considered, because only such a subdiagram yields the divergence. The result is similar to the case with soft divergences. The divergent subdiagram should include only three or four point vertices, all of which must be non-oscillating.

In the above the soft and collinear divergences are treated separately. Then is it not dangerous when their overlapping appears? The answer is “no”, because even in this case we can count the power behavior, which becomes no
worse than the above result. This is because the above power counting is valid for any choice of soft or collinear subdiagrams. We can even change the scaling law for some set of internal lines.

§ 3. Cancellation through the unitarity relation

Now we are in a position to show that it is possible to cancel the singular configuration of a time-ordered double cut diagram. The weapon for this purpose is the largest time equation.2) (In Ref. 7 this equation is applied only to the latest time vertex, but here it is applied to the earliest time vertex also.)

First the proof of the largest time equation in the IR region is given. The basis of the largest time equation is the relation:

\[
D^F(t, k) = \theta(t) D^{+}(t, k) + \theta(-t) D^{-}(t, k),
\]

\[
D^{F} (t, k) = \theta(t) D^{-}(t, k) + \theta(-t) D^{+}(t, k).
\]

Let \( F(\cdots t_{a} \cdots; \{ k_{i} \}) \) stand for the integrand in Eq. (2), where the vertex function with the underlined time coordinate is in the complex conjugate region. It is worthwhile noting that the effective vertex function in the complex conjugate has the same absolute value as and the opposite sign to that in the Feynman amplitude. This feature is owing to the fact that in the skeleton expansion all legs and no internal loop momenta of a vertex function are assigned to the IR region so that any intermediate state cannot be real. (We ignore the ambiguity near the cutoff, because it does not matter for the IR problem.) Taking this feature into account together with the relations in Eq. (3) we easily see that the following “largest time equation in the IR region” holds:

\[
F(\cdots t_{a} \cdots; \{ k_{i} \}) + F(\cdots t_{a} \cdots; \{ k_{i} \}) = 0.
\]

The asterisk means that the value of the time is the largest (or the smallest). The unitarity cut for this vertex represents the final (or initial) state. The diagrammatical expression of this equation is shown in Fig. 2(a) and (b). (Note that we may add some extra cuts allowed from Eq. (3), as is shown in Fig. 2(c)). These extra cuts imply the existence of non-interacting particles.

The largest time equation is usually used to prove the local unitarity,\(^{11}\) and we may call it a unitarity relation. Here, however, it is used in order to cancel only the IR singular configuration.

The proof of the cancellation is very easy. From what was shown in the previous section, an IR singular subdiagram inevitably has the non-oscillating latest (or earliest) time vertex. The type of the allowed latest (earliest) time vertex is limited. [For QCD possible types of the latest time three point vertices yielding the divergences are shown in Fig. 3. Here diagrams that can be
obtained by taking the complex conjugate are omitted. The diagram (c2) is added formally; it results in the null contribution after the time integration from minus to plus infinity. From the largest time equation it follows that soft or collinear divergences are cancelled between (a1) and (a2), (b1) and (b2), and (c1) and (c2).] It is easy to see that any diagram contributing to the divergences has the counterpart of its own which is determined by the largest time equation. Furthermore we should note that the cancellation set is energetically degenerate; otherwise the latest (or earliest) time vertex suffers the damping due to the oscillation factor and does not yield a divergence.

Thus it has been proved that in a topologically equal diagram both soft and collinear divergences disappear after the summation over initial and/or final cuts. The cutting rule is determined by the unitarity relation as depicted in Fig. 2. [The "topological equality" assures in QCD that the cancellation holds in a definite colour channel, as desired.]

In Fig. 4 the lowest order examples of cancellation in massive quark-antiquark annihilation process are given. The IR singularity appears when both of the quark-quark-gluon vertices go to minus infinity of time. For its cancellation we should apply the largest time equation to the earliest time vertex. Thus, following the relations in Fig. 2, the cancellation holds between: (a) and (b), (c)
Fig. 4. The lowest order diagrams in $qq \rightarrow \gamma^*$ with the IR singularity. Here diagrams that can be obtained by taking the complex conjugate are omitted. In the upper half are depicted the time ordered double cut diagrams, and in the lower the corresponding cut diagrams. The notations are tabulated in Table I.

Fig. 5. Similar examples as in Fig. 4 with one soft emission.

and (d). If one admits a soft gluon emission by reversing the direction of energy flow of a gluon in Fig. 4(b) and (c), the counterpart of the cancellation becomes somewhat different as in Fig. 5.

It is worthwhile noting that when applying the largest time equation to the earliest time vertex, we should sum over “initial” cuts. If we sum over final cuts instead, the sum on the l.h.s. of Eq. (4) does not yield zero. This situation becomes more comprehensible if we decompose the propagators as follows:

$$D^{(F)}(t, k) = D_i(|t|) + \theta(t)D(|t|) + \theta(-t)D(|t|),$$
$$D^{(F)}_r(t, k) = D_i(|t|) - \theta(t)D(|t|) - \theta(-t)D(|t|),$$
$$D^{(F)}_l(t, k) = D_i(|t|) + \theta(t)D(|t|) - \theta(-t)D(|t|),$$
$$D^{(F)}_m(t, k) = D_i(|t|) - \theta(t)D(|t|) + \theta(-t)D(|t|),$$

where the symbols are tabulated in Table I. These relations are depicted diagrammatically in Fig. 6. The remainder come from the retarded propagators, $\theta(t)D(t, k)$. Therefore, in the BN inclusive cross section, i.e., the cross section including all soft emissions, divergences originating from minus infinity of
Cancellation of the Infrared Singularity

Table 1. Expression of various propagators.

<table>
<thead>
<tr>
<th>$p$ space</th>
<th>$t-p$ space</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D^{(1)}$</td>
<td>$i/(k_0^2 - \omega^2 + i\epsilon)$</td>
</tr>
<tr>
<td>$D^{(2)}$</td>
<td>$-i/(k_0^2 - \omega^2 - i\epsilon)$</td>
</tr>
<tr>
<td>$D^{(+)}$</td>
<td>$2\pi\theta(k_0)\delta(k_0^2 - \omega^2)$</td>
</tr>
<tr>
<td>$D^{(-)}$</td>
<td>$2\pi\theta(-k_0)\delta(k_0^2 - \omega^2)$</td>
</tr>
<tr>
<td>$D_1$</td>
<td>$\pi\delta(k_0^2 - \omega^2)$</td>
</tr>
<tr>
<td>$D$</td>
<td>$\pi\epsilon(k_0)\delta(k_0^2 - \omega^2)$</td>
</tr>
</tbody>
</table>

time are, if they exist, left uncancelled. In QED, the cancellation holds among topologically different diagrams, but in QCD the colour charge interrupts such a cancellation.\(^9,12\) This is why the so-called BN theorem breaks down in QCD.

§ 4. Physical interpretation

The discussion in the previous section shows that for the natural cancellation of the IR singularity, the summation over cuts is necessary. Then the question arises: Is this procedure allowed from the physical point of view? Since the cancellation holds among energetically degenerate states, the extra soft or collinear particles which this procedure necessitates are unobservable, so long as detectors are colour-blind. However, the problem remains about how to interpret the extra soft particles. We cannot interpret them as the “infrared cloud”, because they are not associated with a single massive particle field. Also it should be noted that they do not originate from interactions with the apparatus of the experiment. If this were true, the processes with different initial or final states can interfere, and hence it becomes impossible to draw a double cut diagram.

Instead we should interpret these extra particles as phantoms owing to the
defect of the interaction picture. As Kulish and Faddeev have discussed elegantly,\textsuperscript{2,13} the IR singularity originates from the bad separation of the Hamiltonian into the free and interaction part; the contribution from the latter is to be calculated perturbatively, though it persists at infinity of time and gives rise to the non-integrable yields.

In contrast with the fact that Kulish and Faddeev changed the asymptotic state itself, the calculational method in the previous section is, as it were, to remedy the defect by introducing the extra soft or collinear particles which cancel the interaction in the asymptotic region.

The relation between these two approaches becomes obvious if we separate the $S$ matrix in such a way as $S = Z_f S' Z_i \dagger$, where $Z_f$ and $Z_i \dagger$ denote the evolution operator in the final and initial asymptotic region. (The approximated form of $Z$ in QED is proposed in Ref. 3.) The inclusive cross section in the previous section takes the form,

$$\sigma = \text{Tr} Z_i \dagger \left( \sum_{\beta f} |\beta \rangle \langle \beta| Z_f S' \right),$$

In the asymptotic dynamics the $S$ matrix element is given by

$$S_{\tilde{f} f} = \langle \tilde{f} | Z_i Z \dagger \dagger S' Z_i | f \rangle = \langle \tilde{f} | Z_f S' Z_i \dagger | f \rangle.$$

The tilde denotes the asymptotic state à la Kulish and Faddeev. In both approaches the IR finiteness of $Z_i Z$, which can be proved by the largest time equation, plays an essential role. The “real” soft or collinear particles which appear in the inclusive cross section approach are absorbed by the definition of the asymptotic state.

References

1) F. Bloch and A. Nordsieck, Phys. Rev. 52 (1939), 54.