Ref.TH.3358-CERN

OPTIMIZATION AND THE ULTIMATE CONVERGENCE OF
QCD PERTURBATION THEORY

P.M. Stevenson
CERN -- Geneva

A B S T R A C T

I suggest that QCD perturbation theory can be convergent, and that "optimization" of the renormalization scheme choice is essential in achieving this. Arguing that higher orders probe shorter distances, I suggest that the effective expansion parameter (the "optimized" couplant) decreases at high orders, leading to an induced convergence. The mechanism is illustrated by a simple mathematical example. The point is that, even if the perturbation series is divergent in all fixed renormalization schemes, the sequence of "optimized" approximations may still converge. It is emphasized that the limit approached by perturbation theory, if any, will not be the exact result of the full theory. Allegations that QCD series are not Borel-summable are critically re-examined in this light.

Ref.TH.3358-CERN
12 July 1982

1. - INTRODUCTION

Does perturbation theory in realistic theories such as QED or QCD converge? The question is an old and intriguing one ¹⁾⁻¹⁸⁾ - and, of course, it will not be answered here. My aim is rather to re-examine the issues involved, and to suggest some new possibilities.

Why is the question interesting? One reason for considering it is simply to learn more about the nature of field theory. It may be that we shall only be successful in going beyond perturbation theory when we have a clearer conception of what perturbation theory is, and what it is trying to tell us. Another reason is a very practical one : the usual folklore is that in QED the perturbation series ceases to converge after about 137 terms, with the n^{th} term ultimately behaving something like $(\frac{1}{2}n)! \alpha^n$. This sort of $n!$ behaviour is supposed to be a generic feature of field theory ⁷⁾⁻¹¹⁾. One might worry, therefore, that QCD series will show a similar phenomenon - perhaps after only a few terms. In that case, even in the foreseeable future, there is a horrifying possibility that higher-order calculations might result in a worsening of the approximation.

The conventional wisdom that perturbation theory is divergent is based largely on experience with finite and super-renormalizable models. However, the realistic case of renormalizable theories is very different, because now the expansion parameter itself is renormalized. That is, the coupling constant is no longer a fixed parameter, but is something that depends on the renormalization scheme (RS) used. As a consequence, finite-order results in perturbation theory are ambiguous : they depend on the choice of RS. Hence, the predictions of perturbation theory, and its initial or "apparent" convergence rate, and its ultimate behaviour, all depend on how the RS is to be chosen.

In the author's view, there is only one basis for choosing the RS that is even remotely satisfactory ; namely, to use the renormalization-group invariance ¹⁹⁾ property of physical quantities. The argument is simple : since we know that the exact result must be RS independent, we have most reason to trust that approximant which is least sensitive to small changes in RS. This very general idea has been named the "principle of minimal sensitivity". I shall not expand on the ramifications of this argument here, as I have already done so at length in previous papers ^{20),21)}. I emphasize, though, that this point is of crucial importance in what follows.

The relevance to the convergence question is the following [cf. some tentative remarks in 20)]. The "optimum" RS identified by the principle of minimal sensitivity is not only different for different physical quantities, but it is also different in different orders of perturbation theory. This feature radically alters the convergence question. The convergence of perturbation theory is not the same as the convergence of the perturbation series. The high-order behaviour of the series in a fixed RS is no longer relevant when, as in "optimized" perturbation theory, one will be adjusting the RS at each new order. All the familiar mathematical lore concerning power series is simply not applicable. In fact, it is possible for the sequence of optimized approximations to converge, even if the perturbation series is divergent in all fixed RS's. In other words, renormalization group invariance ¹⁹⁾, as embodied in "optimization" ²⁰⁾, may act as a built-in summation device for an otherwise divergent perturbation series. Section 2 is devoted to an example which illustrates this point.

Changing the emphasis slightly, the example also answers a question that has been raised in the literature ²²⁾ : - how does "optimization" fare when applied to a divergent series? The answer is both positive and interesting, for it shows that a popular "truism" is false. It is often remarked that, since the RS ambiguity in n^{th} order is $O(\alpha^{n+1})$, the problem becomes less and less serious as one works to higher and higher orders. That is just not true when the series is potentially divergent. Proper treatment of the RS choice is more vital than ever in high orders.

"Optimization" applies only to physical quantities, since only such quantities are guaranteed to be RS invariant. This is not a drawback, since we are - or should be - only interested in the physical aspects of the theory. Green's functions, however much they may fascinate theorists, are only a means to an end. Questions about the convergence or analytic properties of Green's functions are inevitably scheme-dependent, and may not be interesting, or even meaningful. The same remark applies to anomalous dimensions and β functions. These are indispensable tools, but ultimately it is only the finished product that counts.

Convergence properties will presumably depend, to some extent, on which physical quantity one is considering. To talk abstractly of "the behaviour of perturbation theory in QCD" - as I shall do - is an over-simplification. It should be understood that I am considering only perturbatively accessible quantities. That is, I am speaking of situations in which low-order perturbation theory, while not necessarily accurate, is not obviously absurd. In particular, it is assumed that the quantity in question has an infinite-energy limit predictable by leading order renormalization group improved perturbation theory. For example, one may

think of the e^+e^- total cross-section, or the Q^2 dependence of structure function moments. Quantities which are inherently sensitive to the infra-red properties of the theory will not be covered by the arguments here.

Within the limitations just discussed, I offer the conjecture that "optimized" perturbation theory is convergent in asymptotically free theories. I shall accept the general belief that the perturbation series in a fixed RS is probably divergent with zero radius of convergence ; otherwise there would be no problem to discuss. However, I argue that, as in my example, convergence is induced by a gradual decrease of the "optimized" expansion parameter as the order increases. This may be anticipated on the basis of a rather simple physical argument.

Consider a physical quantity at some fixed values of the external momenta. In going to very high orders in perturbation theory one is mainly "dressing-up" the basic process. That is, one is looking ever closer at the polarization cloud around each particle, putting in explicitly more and more of the quantum fluctuations which endow each particle with an internal fine structure. In this sense, higher orders in perturbation theory are a theoretical probe of shorter and shorter distances. Let me temporarily use a QED example ; in zeroth order the electron has the normal Dirac moment of a pointlike fermion. An anomalous magnetic moment arises from higher-order effects precisely because these give the electron an internal structure at short distances. Proceeding to ever higher orders serves to describe this structure in ever finer detail. Similarly, in QCD, higher-order corrections to $\sigma(e^+e^- \rightarrow \text{hadrons})$ reflect the fact that quarks are not pointlike, but have a structure induced by QCD quantum fluctuations. I stress that this viewpoint is different from, and complementary to, the familiar idea that one can "experimentally" probe short distances by increasing the external momentum scale, while keeping the order of perturbation theory fixed. Here the size of the interaction region is fixed by the fixed external scale, but one is looking at ever smaller scale subprocesses within it. This is another aspect of the "partons-within-partons" picture.

The next step of the argument is simple. In asymptotically free theories, short distances are under perturbative control. Therefore, high orders in perturbation theory ought to be controllable - not convergent in the ordinary sense, perhaps, but controllable by renormalization-group means. Since high-order diagrams will be dominated by very small scale structure, the "typical" loop momentum will become large, and one expects this to be reflected in a growth of the optimum

renormalization scale $\bar{\mu}$. [Here I am employing the argument of the "momentum subtractionists" 23),24) as an intuitive interpretation of the workings of "optimization" 20),21),25). The extra realization here is that the "typical" loop momentum is a function not only of the external scale(s) Q , but also of the order of perturbation theory n . I am arguing that, intuitively, it should be an increasing function of n .] Thanks to asymptotic freedom, a growth in $\bar{\mu}$ corresponds to a shrinking of the effective couplant $\bar{a} (\equiv \bar{\alpha}_s/\pi)$ as one proceeds to ever higher orders. This decrease of the effective expansion parameter can compensate for the tendency of the coefficients to grow, leading to convergence, as in the example.

One problem with this hopeful scenario is that it appears to be at odds with much of the earlier literature, particularly with statements that "QCD is not Borel summable" 17),26). I argue in Section 3 that this contradiction is more apparent than real. It is vital not to confuse the question of whether perturbation theory converges with the question of whether the full theory is somehow recoverable from the all-orders perturbation series. There can be - and in QCD there are - terms which are not visible in perturbation theory, not even in "perturbation-theory-summed-to-all-orders". The existence of such terms in no way implies that the perturbation series must be horribly and uncontrollably divergent. It is hoped to clarify these issues in Section 3.

I now turn to an exposition of the mathematical example that provided the inspiration for this paper.

2. - OPTIMIZATION AND CONVERGENCE

A. - A toy model

The potential impact of "optimization" on perturbation theory can be illustrated in the following class of toy examples [inspired by Appendix A of 24)]. Given a series

$$\mathcal{R} = a_0 (1 + r_1(0)a_0 + r_2(0)a_0^2 + \dots), \quad (2.1)$$

let us make the substitution

$$a \equiv a(\tau) = a_0 / (1 + \tau a_0), \quad \tau \in \mathbb{R}, \quad (2.2)$$

and consider the resulting re-expansions

$$\mathcal{R} = \alpha (1 + r_1 \alpha + r_2 \alpha^2 + \dots). \quad (2.3)$$

The coefficients are easily determined to be

$$r_k \equiv r_k(\tau) = \sum_{j=0}^k \binom{k}{j} \tau^j r_{k-j}(0), \quad (2.4)$$

where

$$\binom{k}{j} \equiv \frac{k!}{j! (k-j)!} \quad (2.5)$$

are the binomial coefficients. This simple mathematical system mimics, to some extent, the RS dependence problem. Both the expansion parameter $\underline{a}$ and the coefficients r_k in Eq. (2.3) depend on the extraneous variable τ , which plays the rôle of the RS. Obviously, $\mathcal{R}$ itself does not depend on τ , and the τ dependences of $\underline{a}$ and the r_k must cancel in Eq. (2.3). However, this cancellation is spoiled if the series is truncated.

For the analogy with RS dependence to hold good it is important that all the expansion parameters should be on an equal footing. Equation (2.2) is the simplest example of a substitution which achieves this. The symmetry between any two expansion parameters is made manifest by the relation

$$1/\underline{a} - 1/\underline{a}' = \tau - \tau'. \quad (2.6)$$

By contrast, a substitution such as $\underline{a} = \underline{a}_0(1 + \tau \underline{a}_0)$ would not have been suitable, since it would have given $\underline{a}_0$ a unique status. Equation (2.2) is particularly appropriate since it directly recalls the (leading-order) QCD running coupling constant formula, if we identify τ as proportional to $\ln(\mu/\mu_0)$, where μ is the renormalization scale. Indeed, one has the " β function equation"

$$d\underline{a}/d\tau = -\underline{a}^2. \quad (2.7)$$

The analogy with renormalized perturbation theory is not perfect. Here the multiplicity of expansion parameters, the "unphysical" variable, the " β function", and the resulting ambiguity of finite order results, have all been introduced artificially. In renormalizable field theories these features arise naturally - indeed unavoidably. Moreover, only a single unphysical variable has been introduced here, whereas the real problem involves $(n-1)$ RS parameters at n^{th} order ²⁰. Nevertheless, this oversimplified model can be very instructive.

Differentiating (2.3) and using (2.7) leads to

$$\frac{d\mathcal{R}}{d\tau} = (-\alpha^2)(1 + 2r_1\alpha + 3r_2\alpha^2 + \dots) + (\dot{r}_1\alpha^2 + \dot{r}_2\alpha^3 + \dot{r}_3\alpha^4 + \dots). \quad (2.8)$$

The cancellations required for this to vanish lead to the "self-consistency conditions"

$$dr_k/d\tau \equiv \dot{r}_k = kr_{k-1}, \quad (2.9)$$

which determine the τ dependence of the coefficients [cf. 20)]. Integration of these equations leads, of course, to Eq. (2.4), previously obtained by direct substitution.

The n^{th} order approximant, defined as

$$\mathcal{R}^{(n)} \equiv \alpha(1 + r_1\alpha + \dots + r_{n-1}\alpha^{n-1}), \quad (2.10)$$

is now a known, well-defined function of τ , as soon as we know the values of $\underline{a}$ and $r_1, \dots, r_{n-1}$ at some fixed value of τ , such as $\tau = 0$. [It is not often necessary to have an explicit formula for $\mathcal{R}^{(n)}(\tau)$, but it is sometimes helpful. One may write $\mathcal{R}^{(n)}(\tau)$ as the rational function

$$\mathcal{R}^{(n)}(\tau) = (\alpha_0^{-1} + \tau)^{-n} \sum_{k=1}^n \tau^{n-k} \alpha_0^{-k} \binom{n}{k} \mathcal{R}^{(k)}(0). \quad (2.11)$$

This formula is most easily proved by induction : a direct proof requires a remarkable identity between binomial coefficients.]

The derivative of $\mathcal{R}^{(n)}$ is the residuum of the cancellations in Eq. (2.8) :

$$d\mathcal{R}^{(n)}/d\tau = -nr_{n-1}\alpha^{n+1}. \quad (2.12)$$

The principle of minimal sensitivity defines the "optimum" value of τ at n^{th} order, $\bar{\tau}_n$, to be the value at which $\mathcal{R}^{(n)}(\tau)$ is least sensitive to small variations in τ . Hence, one seeks to make the right-hand side of (2.12) vanish, which leads to the "optimization condition" *)

*) Superficially, this has the look of a "fastest apparent convergence" criterion. However, see the Appendix.

$$r_{n-1}(\tau = \bar{\tau}_n) = 0. \quad (2.13)$$

[If this equation has no solution - a case that never arises in field theory, but which will occur in the examples here - then one must think again. Although the slope $|\partial \mathcal{R}^{(n)} / \partial \tau|$ cannot be made zero, it can be minimized, and this is where $\mathcal{R}^{(n)}$ will be "least sensitive to small variations" in τ . One is then interested in solutions of $\partial^2 \mathcal{R}^{(n)} / \partial \tau^2 = 0$.] Having solved for $\bar{\tau}_n$, the "optimized" approximant $\mathcal{R}_{\text{opt}}^{(n)}$ is found by evaluating $\mathcal{R}^{(n)}(\tau)$ at the point $\tau = \bar{\tau}_n$.

Clearly, one can investigate many examples within this framework, simply by beginning with a different initial series in Eq. (2.1). I leave it as an exercise for the reader to show that if the initial series is geometric, then the optimized result in second, and all subsequent orders is exact. This is true irrespective of whether the initial series was within its radius of convergence or not [cf. 27)]. In the following subsections I study a more challenging example in detail.

B. - The alternating factorial series

Consider the series

$$\mathcal{R} = a_0 (1 - 1! a_0 + 2! a_0^2 - 3! a_0^3 + \dots). \quad (2.14)$$

This is a classic example of a divergent series. It has zero radius of convergence. In the transformed series (2.3), the coefficients are given by

$$r_k = (-1)^k k! \sum_{j=0}^k \frac{(-\tau)^j}{j!} = (-1)^k k! \{e^{-\tau}\}_{k+1}, \quad (2.15)$$

where the notation $\{f(\tau)\}_n$ means "the first n terms $(\tau^0, \dots, \tau^{n-1})$ of the power series expansion of $f(\tau)$ ". If τ is fixed then the coefficients will eventually behave as $(-1)^k k! e^{-\tau}$ for sufficiently large k , with $e^{-\tau}$ being a constant. Therefore, the series remains divergent for all fixed values of τ . However, if we "optimize" the choice of τ at each order then the resulting sequence of approximations is convergent. In the next subsection I outline the proof of (a slightly modified form of) this statement. First, I discuss the optimization procedure in more detail, and present a numerical example.

In this case the optimization condition, Eq. (2.13) becomes simply

$$\{e^{-\bar{\tau}_n}\}_n = 0. \quad (2.16)$$

For even values of n this equation has a single real root, which grows approximately linearly with n [see Eq. (2.21) below]. For odd n there is no real root. As mentioned earlier, in such cases one seeks instead to minimize the slope $|\partial \mathcal{R}^{(n)} / \partial \tau|$, since it cannot be made to vanish. This corresponds to the condition $\partial^2 \mathcal{R}^{(n)} / \partial \tau^2 = 0$, which leads to

$$(\alpha_0^{-1} + \bar{\tau}_n) \{e^{-\bar{\tau}_n}\}_{n-1} + (n+1) \{e^{-\bar{\tau}_n}\}_n = 0. \quad (2.17)$$

This equation has, apparently, two real roots, but the larger corresponds to a maximum of the slope, and hence is not relevant. The appropriate root lies just beyond the $\bar{\tau}$ of the previous even order.

Having solved for $\bar{\tau}_n$ one can immediately evaluate the optimized values of $\underline{a}$, r_k , and $\mathcal{R}^{(n)}$ from Eqs. (2.2), (2.15), (2.10).

Some numerical results for the case $a_0 = \frac{1}{4}$ are shown in the Table. The naïve results, corresponding to the partial sums of (2.14), converge initially, but soon develop violent fluctuations. The optimized results, by contrast, converge in a steady fashion. This comes about because $\bar{\tau}_n$ grows with n at just the right rate. The resulting decrease in the effective expansion parameter $\bar{a}_n$ counterbalances the potential growth of the coefficients. The radius of convergence of the original series is zero, but asymptotically one is expanding in powers of a vanishing parameter.

It is interesting to see how the terms of the series behave numerically at the "optimized" value of τ . A typical example is 8th order, which looks like

$$\begin{aligned} \mathcal{R}^{(8)}(\bar{\tau}_8) = & 0.148 (1 + 0.260 + 0.090 + 0.028 \\ & + 0.011 + 0.003 + 0.002 + 0). \end{aligned} \quad (2.18)$$

Thus, the "apparent convergence" of the series in the "optimized" scheme is perfectly satisfactory. The residual error that one would probably guess from this - say, $\lesssim 0.148 \times 0.001$ - is a reasonable indication of the actual error (which, in fact, is only about half this size).

It is also instructive to plot some graphs of $\mathcal{R}^{(n)}(\tau)$ against τ for various values of n . This has been done in Fig. 1, by utilizing Eq. (2.11). One can describe each curve in terms of three main regions : (i) an "overdamped" region at large τ in which $\mathcal{R}^{(n)}(\tau)$ is small, (ii) a plateau around $\tau = \bar{\tau}_n$, and (iii) a "diverging" region for $\tau \lesssim \bar{\tau}_n$ where the function tends rapidly to $+\infty$ or $-\infty$, according to whether n is odd or even. The regions (i) and (ii) merge smoothly, but the break between regions (ii) and (iii) is rather abrupt. Both boundaries move rapidly to the right as n increases. They do so at slightly different rates, so that the width of the plateau region between them slowly grows.

If we sit at some fixed, reasonably large τ , we see each of these regions in turn. Low order results are too small, but they grow steadily and begin to settle down to an almost constant value. Then, quite suddenly, we are overtaken by the diverging region, and the results start to oscillate increasingly violently from one order to the next. The view from the "co-moving frame" of the optimized τ is quite different. One sees the approximation converging smoothly, and becoming flat over an increasing range of τ . From this vantage point the approximation is not only becoming steadily more accurate, it is also successfully mimicking the τ independence property of the exact result.

This example provides further evidence in favour of the principle of minimal sensitivity. Moreover, it is evidence of a new and dramatic kind. Previous examples ^{20),27)} are analogous to the case of an initial series that is convergent. In such cases, any finite τ will eventually lie within the flat region. "Optimization" serves to speed up the convergence process, but one can always obtain an arbitrarily good approximation at any τ , provided one is prepared to calculate enough orders. Here, by contrast, "optimization" is essential for convergence. Although the flat region still grows with n , it moves faster than it grows - and it is essential to keep pace with it. Without optimization, brute-force calculation to very high orders is worse than useless - it is counterproductive !

C. - Convergence proof

I now outline a proof of the convergence of the "optimized" approximants in the example described above. The story involves several surprises. Firstly, one needs to know the behaviour of $\bar{\tau}_n$ as $n \rightarrow \infty$. For this purpose the integral representation

$$\{e^{-\tau}\}_n = e^{-\tau} \left[1 + \frac{(-1)^{n-1}}{\Gamma(n)} \int_0^{\tau} t^{n-1} e^t dt \right], \quad (2.19)$$

obtained by repeated integration by parts, is extremely useful. In the case $n = \text{even}$, the optimization condition, $\{e^{-\bar{\tau}_n}\}_n = 0$, reduces to

$$\int_0^{\bar{\tau}_n} t^{n-1} e^t dt = \Gamma(n). \quad (2.20)$$

One can now consider this as an abstract problem with n ranging over all real values. Taking the difference of Eq. (2.20) and the corresponding equation with $n \rightarrow n+1$; putting bounds on the remaining integral term; and using the asymptotic behaviour of the Γ function one can show that

$$\bar{\tau}_n \sim \chi n + \xi \ln n + O(1), \quad (2.21)$$

where $\chi \approx 0.278$ is the solution to

$$\ln \chi + \chi + 1 = 0, \quad (2.22)$$

and $\xi = \frac{1}{2}\chi/(1+\chi) \approx 0.109$. For odd n an analysis of Eq. (2.17) leads to the same result, except that the coefficient of $\ln n$ in (2.21) is altered. It will turn out, though, that the $\ln n$ term is not important.

The alternating factorial series, Eq. (2.14), is a classic example of a Borel-summable series. The sum of the series, in the sense of Borel, is given by the integral

$$\mathcal{R}_B \equiv \int_0^\infty du e^{-u/a_0}/(1+u). \quad (2.23)$$

Heuristically, the series is generated by expanding $1/(1+u)$ as $1 - u + u^2 - u^3 + \dots$, and then integrating term by term. Of course, the expansion is only valid for $|u| < 1$, whereas the integral involves $0 < u < \infty$. The difficulty is exposed by noting that truncations of (2.14) correspond to

$$\mathcal{R}^{(n)}(0) = \int_0^\infty du e^{-u/a_0} \frac{(1 - (-u)^n)}{(1+u)}, \quad (2.24)$$

in which the $(-u)^n$ term is far from negligible as $n \rightarrow \infty$, and is the cause of the wild oscillations. Recalling that $1/a_0 = 1/a - \tau$, one can write down a similar integral representation of the n^{th} order approximant at a general value of τ :

$$\mathcal{R}^{(n)}(\tau) = \int_0^\infty du e^{-u/a} \{e^{u\tau} (1+u)^{-1}\}_n. \quad (2.25)$$

Straightforward algebraic manipulations then lead to

$$\mathcal{R}^{(n)}(\tau) = \int_0^\infty du \frac{e^{-u/a}}{(1+u)} [\{e^{u\tau}\}_n - (-u)^n \{e^{-\tau}\}_n]. \quad (2.26)$$

The optimization condition chooses τ such that $\{e^{-\tau}\}_n = 0$, and so it eliminates the troublesome $(-u)^n$ term. (I restrict the discussion to the case $n = \text{even}$, for the moment.) If $\{e^{u\tau}\}_n$ could be replaced by $e^{u\tau}$ as $n \rightarrow \infty$, then one would have established that $\mathcal{R}_{\text{opt}}^{(n)} \rightarrow \mathcal{R}_B$.

However..., this replacement is not valid when τ is $\bar{\tau}_n$, which is linearly increasing with n . Each term in the difference $(e^{u\bar{\tau}_n} - \{e^{u\bar{\tau}_n}\}_n)$ generates only an $O(1/n)$ correction, but the sum of these corrections is finite. To see this one can use the representation (2.19) to rewrite (2.26), for $\tau = \bar{\tau}_n$, as

$$\mathcal{R}_{\text{opt}}^{(n)} = \mathcal{R}_B - \int_0^\infty du \frac{e^{-u/a_0}}{(1+u)} \frac{\gamma(n, u\bar{\tau}_n)}{\Gamma(n)}, \quad (2.27)$$

where $\gamma(n, z)$ is the incomplete Γ function

$$\gamma(n, z) \equiv \int_0^z v^{n-1} e^{-v} dv, \quad (2.28)$$

and $\Gamma(n)$ is, of course, $\gamma(n, \infty)$. One requires the large n behaviour of $\gamma(n, z(n))$ when z behaves as $u(\chi n + \xi \ln n + O(1))$. Applying Laplace's method [see 28), Chapter 3], one can show that

$$\lim_{n \rightarrow \infty} [\gamma(n, \chi u n) / \Gamma(n)] = \begin{cases} 0, & u\chi < 1 \\ 1, & u\chi > 1. \end{cases} \quad (2.29)$$

The same technique establishes that the $\xi \ln n$ term is negligible. (Incidentally, this indicates that the flat region grows at least as fast as $\ln n$.) Thus, the extra term in (2.27) receives contributions only for $u > 1/\chi \approx 3.591$, but its integrand has exactly the same form as that of $\mathcal{R}_B$.

Thus, the limit of the optimized approximants is not the same as the Borel sum. Instead,

$$\lim_{\substack{n \rightarrow \infty \\ n \text{ even}}} \mathcal{R}_{\text{opt}}^{(n)} = \int_0^{1/\chi} du e^{-u/a_0} / (1+u). \quad (2.30)$$

The difference from $\mathcal{R}_B$ lies only in the upper limit of the integral - something which does not affect the series expansion. The difference is exponentially small, $O(a_0 e^{-1/(\chi a_0)})$ for small a_0 . In the numerical example of the Table, the ultimate "discrepancy" is $< 3.1 \times 10^{-8}$.

The story takes another turn when we consider the case $n = \text{odd}$. The different sub-asymptotic behaviour of $\bar{\tau}_n$ makes no difference, but now the $\{e^{-\tau}\}_n$ term in (2.26), although highly suppressed, is non-vanishing at $\tau = \bar{\tau}_n$. Analysis of the corresponding contribution requires a generalization of Laplace's method [see 28), Chapter 9]. The outcome is a finite contribution

$$e^{-1/(\chi a_0)} / (1 + \chi)^2. \quad (2.31)$$

Thus, the "optimized" approximants are actually "bi-convergent" : the odd orders converge to a value slightly greater than the even orders ! This curious feature arises from the absence of a solution to the usual optimization condition in odd orders - a subtlety that does not occur in the field theory case. [One could banish the effect here by allowing complex solutions to $\{e^{-\tau_n}\}_n = 0$. Odd orders would then acquire a small imaginary part of ambiguous sign, but would converge to the same limit as even orders - barring some unforeseen subtlety.]

Although these differences from the Borel sum are amusing and curious, I do not believe they are anything to worry about, as far as field theory is concerned. They merely reflect the artificiality of this kind of example. The sum of a divergent series is a matter of definition ²⁹⁾, and which definition is most appropriate depends on the context in which it appears. Within the present, abstract mathematical context, the Borel sum is certainly a very natural and attractive definition. In comparison, the manipulations employed here seem rather artificial. The substitution $a = a_0/(1 + \tau a_0)$, which introduced the unphysical variable τ , was made purely by fiat : the only motive was the ulterior one of wanting to mimic the RS dependence problem. So, the "discrepancy" problem is not due to any shortcoming of the principle of minimal sensitivity itself, but is rather a reflection of the artificiality of the framework within which it operates here ^{*)}. None of these "artificiality" problems would arise in a field theory context. There the occurrence of unphysical variables is natural and inescapable, and "optimization" is necessary even to make sense of finite-order perturbation theory. If this process tends to a limit, then that is the most natural definition of the "sum" of perturbation theory. In comparison, Borel summation appears, in a field theory context, to be something of an artificial trick, without physical motivation.

*) This can be seen particularly clearly by noting that even-order approximants are always less than the Borel sum by a finite amount. "Optimization", by picking the maximum of the approximant, is incontrovertibly finding the closest approximation to the Borel sum, out of the possibilities open to it.

The unfortunate feature of this effect is that I fear the reader may confuse it with the point to be made in the next section. Here I started with a series expansion : there I will be talking about the function from which the series arises. This function may itself possess $e^{-1/a}$ type terms, which need have nothing to do with the series, or its "sum".

3. - SUMMABILITY AND RECOVERABILITY

A. - Contradiction with previous results?

It is now time for some discussion of the existing literature on high-order behaviour in field theory. I embark on this with some trepidation, as it takes me into realms of mathematical physics outside my radius of competence. However, some comment is surely required, for the following simple reason : the suggestion that QCD perturbation theory is convergent is seemingly at odds with most previous work ^{17),7)-11)}. The conventional wisdom has it that "QCD is not Borel-summable" [e.g., 26)]. This supposedly implies that not only are QCD series divergent, but they are so horribly divergent that not even such a powerful resummation method as Borel's can make sense of them. If this is so, then it would seem a forlorn hope that the series can be tamed by "optimization".

Of course, one could speculate about several technical let-outs from this apparent contradiction. Physical quantities might behave differently from the Green's functions studied in the literature. Furthermore, the "convergence-through-optimization" mechanism is not the same as Borel summation, and it might work even when Borel summation fails. These points may well be important, but I do not think they are the heart of the matter. The contrast between the present view and the conventional one goes deeper than that. For instance, conventionally the failure of Borel summation appears to be a consequence of asymptotic freedom, whereas here asymptotic freedom plays a vital, beneficial rôle in the suggested convergence mechanism.

I shall argue the following provocative thesis : - There is actually no hard evidence that QCD series are horribly divergent (or even that they are divergent at all) : any impression to the contrary stems from a confusion of two separate issues ; in particular, from a misleading use of the word "summable". I shall attempt to clarify these issues in this section. My central point is

the existence of terms invisible to perturbation theory ; something that is very well known. However, in this context, the full implications do not seem to have always been properly appreciated.

In brief, there are two kinds of questions that must be distinguished :
 (i) Does ("optimized") perturbation theory converge? (or, Can perturbation theory be summed?) and (ii) Is the full theory somehow recoverable from the all-orders perturbation series? The answer to the second question, in QCD (and probably in any realistic theory), must surely be : "No, - perturbation theory, even to all orders, is not the whole story ; and no amount of mathematical juggling - unless it involves new, physical input - will ever make it so". The existence of instantons and a θ parameter in QCD, as well as other work of 't Hooft¹⁷⁾, surely allows no other conclusion. However, contrary to popular impression, there is no inference that the perturbation series is uncontrollably divergent. It is still possible that perturbation theory converges to a limit, and, while this limit will not be the true result of the full theory, it may still be useful and meaningful.

The two issues raised above I shall refer to as the questions of "summability" and "recoverability" respectively. In the next subsection these terms are carefully defined. The subsequent subsections deal with the implications for the interpretation of existing analyses relating to the high-order behaviour of perturbation series.

B. - Mathematical background

Consider a well-defined function $f(a)$, all of whose derivatives exist when $a \rightarrow 0_+$, so that it possesses a formal perturbation series (Taylor expansion), which I denote by $S\{f(a)\}$.

$$S\{f(a)\} = \sum_{n=0}^{\infty} r_n a^n, \quad (3.1)$$

where

$$r_n \equiv \frac{1}{n!} \left. \frac{d^n f(a)}{da^n} \right|_{a \rightarrow 0_+}. \quad (3.2)$$

The function and the series are not to be confused. Indeed, the series is defined only as a collection of symbols, whose numerical meaning is not yet defined. One can, however, construct numerical quantities from $S\{f(a)\}$; in particular, one can define the partial sums $S_f^{(N)}(a)$ by truncating the series after N terms. The series is convergent or divergent according to whether the limit $\lim_{N \rightarrow \infty} S_f^{(N)}(a)$

exists or not. [It is implicit here that I am talking about "sufficiently small a ". That is convergent (divergent) will be synonymous with finite (zero) radius of convergence.]

In the convergent case the limit of the partial sums yields the "ordinary sum" of $S\{f(a)\}$, denoted here by $\text{Ord}[S\{f(a)\}]$. This may or may not be equal to $f(a)$. That is, even if the series is convergent one does not necessarily recover the original function. For example, consider

$$f(a) = 1/(1-a) + e^{-1/a}, \quad (3.3)$$

whose series expansion, $S\{f(a)\} = \sum a^n$, converges to $1/(1-a)$, which is not equal to $f(a)$. The important point is that functions will, in general, contain an "invisible" or "non-perturbative" part which is characterized by zero derivatives as $a \rightarrow 0_+$. One sees such terms neither in finite orders, nor in the sum-to-all-orders.

[The statement that a function is analytic at $a = 0$ implies both that its series converges and that the sum of the series is equal to the original function (these statements being true in some circular region around the origin). However, non-analyticity at $a = 0$ does not necessarily imply a divergent series - as can be seen from the example above. The original argument for the divergence of QED series, as presented by Dyson¹⁾ and frequently repeated elsewhere, is flawed in this respect : it falsely assumes that non-analyticity implies a divergent series : cf. 5), 15).]

In the divergent case one may still be able to ascribe a "sum" to the series, in a generalized sense. As emphasized by Hardy²⁹⁾, the "sum" of a divergent series is a matter of definition. There are many different methods for summing divergent series, each usually associated with the name of a famous mathematician. (I shall consider only regular methods, i.e., methods which always yield the ordinary sum when applied to a convergent series.) Under certain circumstances, different methods can be equivalent, but this is not true in general. Thus, one should speak of the "X sum", or the "sum-in-the-sense-of-X", where $X = \text{Borel, Euler, etc.}$, denotes the summation method. I write the X sum of the series as $X[S\{f(a)\}]$, or, more briefly, as $f_X(a)$. It is a function of a , but it is not necessarily the same as the original function $f(a)$.

The distinction to be made is between summability and recoverability *). The two concepts correspond, respectively, to the following two statements (each of which should in practice be qualified by a restriction to some domain of a.)

- 1) " $S\{f(a)\}$ is X summable". This is a property of the series ; namely, that the series can be ascribed a finite sum in the sense of X ; i.e., $X[S\{f(a)\}] \equiv f_X(a)$ exists.
- 2) " $f(a)$ is X recoverable". This is a property of the function ; namely, that when its formal series expansion is resummed by the method of X , one recovers the original function exactly ; i.e., $f_X(a) = f(a)$.

The relation between summability 1) and recoverability 2) may be summarized thus :

$$2 \Rightarrow 1, \quad \bar{1} \Rightarrow \bar{2}, \quad \text{but } 1 \not\Rightarrow 2, \quad \bar{2} \not\Rightarrow \bar{1}. \quad (3.4)$$

The positive statements are obvious ; recoverability is not possible without summability. The example of Eq. (3.3) is sufficient to establish the negative statements (for any regular summation method).

The central point may be encapsulated thus : just because a function $f(a)$ has a formal perturbation series expansion $\sum_n r_n a^n$, we are not entitled to write

$$f(a) = \sum_{n=0}^{\infty} r_n a^n. \quad (3.5)$$

wrong

It is distressing to find such a misidentification underlying many classic works on the high-order behaviour of perturbation theory. To write such an equation is to assume recoverability at the outset. That is a sweeping assumption in any context, and in QCD it is surely false.

*) A rare previous instance of a clear distinction between these two concepts occurs in a paper by Khuri 12), who refers to them as "weak" and "strong" summability, respectively. I prefer my terminology because it emphasizes that not only are these properties distinct, but also they are properties of different things. Nevertheless, I am much indebted to Khuri's paper for confirming and illuminating several of the conclusions to which I had been coming.

C. - Implicit and invisible singularities

The traditional approach to the question of the divergence of perturbation series proceeds roughly as follows. (It is convenient for the present to consider QED : I return to QCD in a moment.) When α , the fine-structure constant of QED, is imagined to be negative, one expects on "physical" grounds that the normal vacuum is unstable to tunnelling¹⁾. This implies that amplitudes will have a cut on the negative α axis, with a discontinuity proportional to $i e^{-C/|\alpha|}$ for small $|\alpha|$, with C being a computable constant^{*)}. From this knowledge one proceeds to extract, by a suitable contour integration, the high-order behaviour of the series. Crudely speaking, the result is that the coefficients grow something like $(n!)^{7)-11)$.

Such arguments may well give the right result in several cases, but they are not rigorous. As made clear by Khuri¹²⁾, the result depends entirely on the assumption that the $i e^{-1/|\alpha|}$ cut is the only singularity in the neighbourhood of the origin. This happens to be true for ϕ^4 in $(0+1)$ dimensions⁶⁾, but it has never been established for QED, or even for $\phi_{(3+1)}^4$.

This objection is closely connected with the possible presence of "invisible" terms, as can be seen from the following counter-example. Consider the function

$$f(\alpha) = \frac{(\ln \alpha) e^{1/\alpha}}{(1 + e^{1/\alpha^3})} + f_X(\alpha), \quad (3.6)$$

where I leave the series part - whose X sum is $f_X(\alpha)$ - unspecified for the present. The first term is invisible for $\alpha \rightarrow 0_+$ (since the large numerator is overwhelmed by the huge denominator) ; it does not contribute to the series expansion at any order. When α is small and negative, however, the invisible term yields a cut with a discontinuity proportional to $i e^{-1/|\alpha|}$. I can easily arrange that this behaviour is not significantly altered by the $f_X(\alpha)$ term. Thus, the function $f(\alpha)$ possesses a Dyson "vacuum-instability" cut. Nevertheless, the existence of this cut does not determine the high-order behaviour of the series. For example, I may choose $f_X(\alpha)$ to be the sum of a convergent series, if I wish.

*) Even this part of the argument is questionable. If $e^2/4\pi$ is negative, then e must be imaginary. The Hamiltonian is then not Hermitian, and hence unphysical, so it is not obvious that one can employ physical reasoning^{5),16)}. Furthermore, one is not merely generalizing the theory by allowing the parameters in the Lagrangian to take complex values. Since physical quantities are the square moduli of amplitudes, if e is complex then " α " generalizes to $|e|^2/4\pi, (e^2+e^{*2})/8\pi, \dots$, depending on the context. Consequently, it is not at all clear what it means to consider a continuation of the theory to non-positive α .

The point is this : the usual argument points to a singularity expected on "physical" grounds, and assumes that the singularity is reflected in the X sum of the series. This is not necessarily the case : the singularity may reside in an invisible term which leads an existence independent of the series.

In the above example there are other singularities arbitrarily close to the origin, besides the branch-point/essential-singularity at $\alpha = 0$. The zeros of the denominator $(1+e^{1/\alpha^3})$ give rise to poles at

$$\alpha = \frac{\pm i}{((2k+1)\pi)^{1/3}}, \quad \frac{\frac{1}{2}(\sqrt{3} \pm i)}{((2k+1)\pi)^{1/3}}, \quad (k=0,1,2,\dots), \quad (3.7)$$

which accumulate at the origin, as shown in Fig. 2. However, this is not a necessary feature of such a counter-example (i.e., even the assumption of no additional singularities is not strong enough to guarantee the conventional results ¹²⁾). Consider instead

$$f(\alpha) = (\ln \alpha) \exp\left(\frac{1}{\alpha}(1-e^{1/\alpha})\right) + f_x(\alpha). \quad (3.8)$$

Again, the first term is invisible for $\alpha \rightarrow 0_+$, and has a cut with a discontinuity proportional to $i e^{-1/|\alpha|}$ for $\alpha \rightarrow 0_-$. The origin is a very complicated essential singularity, but there are no other singularities, except at infinity. The series term $f_x(\alpha)$ remains at my disposal.

The lesson of these examples is this : unless we have a complete knowledge of the precise nature of the singularities at or near the origin, it is dangerous to draw conclusions for the high-order behaviour of the series. One must be able to distinguish implicit from invisible singularities. An implicit singularity is one present in the X sum of the series. Such a singularity implies a certain kind of divergent behaviour in the series. Conversely, a particular high-order behaviour of the series implies a certain type of singularity in its X sum. However, the function $f(\alpha)$ may have other singularities which are not implied by the series. These are invisible singularities present in invisible terms (which are necessarily essentially-singular at $\alpha = 0$).

A similar comment applies to the interpretation of some classic work of 't Hooft on the Borel summability of QCD ¹⁷⁾. 't Hooft investigates the analytic structure of QCD Green's functions by mapping the cuts representing physical meson (or baryon) thresholds in the energy plane on to the complex α plane. He shows that these branch points leave too narrow an angular region astride the positive

real axis for Watson's theorem to apply ; i.e., the conditions for Borel-recoverability are not met. However, this does not necessarily imply a non-Borel-summable series - or even a divergent one -, since the singularities may reside in an invisible term.

Watson's theorem [see, e.g., 26)] specifies the conditions under which a function is Borel recoverable. The relevant requirement - violated by QCD - is that the function should be analytic, for sufficiently small $|\alpha|$, in a sector $|\arg \alpha| < \theta_{\max}$, where $\theta_{\max}$ is greater than $\pi/2$. Clearly, the purpose of this stipulation is to rule out the possibility of "invisible" terms, such as $e^{-1/\alpha}$. Failure to meet this condition would seem to be an indication that invisible terms are present, rather than a sign that the series is too rapidly divergent. (The latter possibility is actually covered by the other requirement of the theorem.)

The suggestion that 't Hooft's singularities are invisible rather than implicit, is not unreasonable. The singularities correspond to the thresholds for producing bound states. ('t Hooft's argument assumes that QCD confines, and gives rise to the hadrons we observe.) Confinement and bound-state thresholds are certainly not seen in any finite order of perturbation theory. Whether they are "implicit" or "invisible" effects is not clear, but I believe one could make a good case for the latter possibility.

There is some analogy with the example (3.6), whose analytic structure (see Fig. 2) is not so dissimilar to 't Hooft's Green's functions : there are lines of singularities accumulating at the origin, with the analytic region astride the positive real axis being too narrow for Borel recoverability. Nevertheless, the behaviour of the series in this example is quite unconstrained. In spite of its horrifying analytic structure, $f(\alpha)$ can have a Borel-summable - or even a convergent - series.

D. - Borel and Borel-Laplace transforms

Another important approach to the question of high-order behaviour centres on the Borel transform. Given a power series

$$S\{f(\alpha)\} = \sum_{n=0}^{\infty} r_n \alpha^n, \quad (3.9)$$

its Borel transform is defined as

$$F_B(z) \equiv \sum_{n=0}^{\infty} \frac{r_n}{n!} z^n. \quad (3.10)$$

The Borel sum of the original series is defined as

$$B[S\{f(\alpha)\}] \equiv f_B(\alpha) \equiv \frac{1}{\alpha} \int_0^\infty e^{-z/\alpha} F_B(z) dz. \quad (3.11)$$

It is easy to see that, on reversing the order of summation and integration, one will reproduce the original series, in a formal case.

Clearly, for this method of summation to work, both Eqs. (3.10), (3.11) need to be well defined : the series (3.10) must converge - at least for a finite range of z , with the result to be analytically continued to all z -, and the integral (3.11) must exist. The latter condition requires that the integral converges at infinity, and that $F_B(z)$ has no (non-integrable) singularities on the positive real axis.

According to the literature, the singularities of $F_B(z)$ are determined by classical solutions of the Euclidean field equations¹⁷⁾. In QCD, because of instantons, these singularities were found to lie on the positive real axis, and it was concluded that QCD series are not Borel summable^{17),26)}. However, what has really been shown is that QCD is not Borel recoverable (which may, perhaps, have been the authors' main interest). There is no necessary implication that QCD series are not Borel summable.

One must be careful not to confuse the Borel transform $F_B(z)$ with the "Borel-Laplace" transform $F_{BL}(z)$, defined by

$$f(\alpha) \equiv \frac{1}{\alpha} \int_0^\infty e^{-z/\alpha} F_{BL}(z) dz. \quad (3.12)$$

Whereas $F_B(z)$ was defined through the series $S\{f(\alpha)\}$, here $F_{BL}(z)$ is defined directly in terms of the original function $f(\alpha)$, being essentially its inverse Laplace transform. A comparison of the last two equations shows that $F_{BL}(z) = F_B(z)$ only if $f(\alpha) = f_B(\alpha)$. Thus, in general the two transforms are not to be identified - just as the original function is not to be confused with its series.

In the context of field theory, it is the singularities of the Borel-Laplace transform which one finds by studying classical solutions of the field equations. These singularities are not necessarily present in the Borel transform. For example,

$$F_{BL}(z) = e^z + \delta(1-z) \quad (3.13)$$

has a singularity at $z = 1$. Its Laplace transform, in the sense of Eq. (3.12), is

$$f(a) = 1/(1-a) + a^{-1}e^{-1/a}, \quad (3.14)$$

and so the singularity in $F_{BL}(z)$ arises from an invisible term, and has nothing to do with the series' behaviour. The series expansion of $f(a)$, $\sum a^n$, yields $F_B = \sum z^n/n! = e^z$, which is free of singularities. [Another example of an invisible singularity might be $e^{-1/z}/(1-z)$.]

Thus, as in the previous subsection, one must distinguish between implicit and invisible singularities in the Borel-Laplace transform. Implicit singularities are those also present in the Borel transform, and they reflect the behaviour of the series. Invisible singularities are present in non-power-series terms, and they correspond to invisible terms in $f(a)$.

For QCD, the singularities that have been found in $F_{BL}(z)$ must surely be "invisible". They arise from instantons, which are $e^{-1/a}$ effects, invisible to perturbation theory. The corresponding terms in $F_{BL}(z)$ have the form of derivatives of delta functions¹⁷⁾, and so the situation is closely analogous to my example above.

The conclusion on instanton singularities is the same as for the α plane approach discussed in the previous subsection. These arguments establish the (Borel) non-recoverability of QCD, but they do not necessarily say anything about the high-order behaviour of QCD perturbation series.

4. - CONCLUSIONS

My firmest conclusion is that the behaviour of perturbation theory in high orders remains very much an open question. The principal aim here has been to draw attention to important aspects of the problem which have previously been neglected. More speculatively, I have argued for a particular scenario, which I now summarize.

Firstly, the purpose of Section 3 was to show that there is really no convincing evidence that QCD series are horribly divergent - or even that they are divergent at all. Furthermore, in view of the amazing cancellations that are an everyday miracle in gauge theories¹³⁾, one should not take too seriously

arguments based on diagram counting ²⁾. However, such arguments do indicate that one needs to be very optimistic to believe that perturbation series are convergent in the ordinary sense. I suggest instead that, though the series is presumably divergent, it can be controlled by appropriate use of the renormalization group : that is to say, "optimization" of the RS choice leads automatically to a convergent sequence of approximations. The example of Section 2 demonstrates that this is a genuine possibility, and the intuitive argument given in the Introduction suggests that the possibility is realized in QCD, thanks to its asymptotic freedom ^{*)}.

If the mechanism does work it has conceptual advantages over other possible methods based on, say, Borel summation ¹⁷⁾, or Padé approximants ³⁰⁾. Firstly, it is an automatic mechanism, in that "optimization" is already necessary to make sense out of finite-order perturbation theory ^{20),21)} : if it acts as a summation device then that is a bonus. Moreover, it is founded on a fundamental physical principle, namely renormalization-group invariance ¹⁹⁾. In comparison, Borel summation, Padé approximants, etc., seem like artificial, mathematical tricks, for which there is no specific, physical motivation.

Moreover, I question whether naïve use of devices and concepts from the mathematical theory of power series is really appropriate in renormalizable field theories. It is no longer possible to view perturbation theory as a power-series expansion. In Section 3 I was prepared, for the sake of argument, to go along with the usual view, but I repudiate it now. Firstly, there are two series involved - for the physical quantity in question, and for the β function. Secondly, the "expansion parameter" is not fixed : it is not the parameter in the Lagrangian, but something whose definition is open to choice - and the right choice may change from one order to the next. Conventionally, these complications are glossed over, and renormalization is viewed as a minor nuisance, merely requiring some technical finessing. The view here is quite the contrary : renormalization is of major conceptual importance to the question of high-order behaviour in realistic field theories.

Perhaps this may shed some light on the following question. As is well known, it is not legitimate to re-arrange the order of the terms of a power series which is not absolutely convergent. On the other hand, many popular and apparently successful techniques, such as leading-log summation, or the double-leading-log approximation, rely on just such a re-arrangement of the perturbation series. The questionable validity of such procedures has long been a cause of

*) These are not empty words. The prediction that the optimized couplant tends to shrink in higher orders may be supported or denied by explicit calculations, within the foreseeable future.

concern ^{2),3),31)}. Of course, a mere shift from one RS to another constitutes a re-ordering of the perturbation series. For example, an adjustment of the re-normalization scale μ causes a part of each higher-order term to be re-absorbed into the leading term. Clearly, one cannot outlaw such RS changes entirely. There is no God-given "right" RS. In the ordinary sense, then, there is no "right" arrangement of the terms of the series. One needs a more sophisticated concept of the "right arrangement". Any static arrangement of the terms - i.e., any fixed RS - is wrong, and presumably leads to divergent results. However, "optimization" defines a dynamic arrangement, in the sense that the RS changes from one order to the next. I suggest that this is the "right" ordering because it is based on enforcing, as far as possible, the physically necessary property of RS independence.

It is then conceivable that certain re-arrangements are allowable if they respect this "dynamic" property, so that the expansion parameter is not fixed but has an order dependence. As an example of what I have in mind, consider the ladder-diagram calculations of Ref. 32). Those calculations certainly work, in the sense that they reproduce the results of ordinary perturbation theory used together with the operator-product expansion. It is crucial, though, that the resummation involves not a fixed coupling constant, but one which varies down the ladder, reflecting the energy scale of each individual bremsstrahlung subprocess. That is the sort of thing I mean by "dynamic". I can only make this clearer in the form of a question. Peterman has shown that, essentially, "the RG invariance of the theory underwrites the legitimacy of the leading-log approximation" ³¹⁾ : can this be generalized? Given a theory for which "optimized" perturbation theory converges, what kinds of "dynamical" re-arrangements can be rigorously justified?

Finally, I turn to a question that has probably been bothering the reader. I have repeatedly emphasized that the "sum" of perturbation theory will not be the same as the true prediction of the full theory. That clearly begs the question : "What is the point of all this, if one is going to get to the wrong result?" Well, firstly, the error terms, being "invisible", must necessarily vanish faster than any power of the couplant $\underline{a}$ ($\equiv \alpha_s/\pi$) as $\underline{a} \rightarrow 0$. Thus, in QCD, they quickly become negligible once one is above the characteristic energy scale of the theory. One then has an excellent approximation to the right result in the $Q \gg \Lambda$ region. That is, the phenomenological successes of low-order perturbation theory are not a fluke : they represent the first stage in a nicely convergent, but incomplete (though only very slightly when $Q \gg \Lambda$), approximation scheme.

Furthermore, the notion of a "sum" of perturbation theory provides a starting point for an attempt to do better. It provides something to which one can sensibly add "non-perturbative" terms - with the aim of achieving better accuracy and/or relaxing the $Q \gg \Lambda$ restriction somewhat. The idea that one should supplement perturbation theory with "non-perturbative" or "higher-twist" terms is commonplace, and is already being vigorously pursued in various ways [e.g., Refs. 33)-35)]. However, it is difficult to see how this would make sense if the perturbative term, to which one is adding things, had no meaning. If perturbation theory actually degenerates into uncontrollable divergence, and has no natural sum, then any "perturbative-plus-non-perturbative" picture is fraught with ambiguity - for if the sum-of-perturbation-theory is not well-defined, then the definition of what is "non-perturbative" must be equally ambiguous. So, in conjecturing that there is a "natural sum" of perturbation theory, I am also implicitly saying that there is a natural definition of "non-perturbative". The "perturbative/non-perturbative" split-up I envisage is meaningful, in the sense that each term separately is renormalization-group invariant.

Perhaps I can make this more concrete : the terms I have called "invisible" would be referred to in other contexts as "dynamical higher-twist effects" ³³⁾ (as opposed to "trivial" kinematic higher-twist effects due to quark masses). This is so because Λ/Q can be written as $\exp(-\int^a(Q) da'/\beta(a')) \simeq e^{-1/ba(Q)}$, so that any power of Λ/Q is necessarily an "invisible" term ³⁶⁾. Thus, the "twist expansion" currently being pursued by practical theorists ³³⁾⁻³⁵⁾ is something like

$$f(a) = S_0(a) + e^{-1/a} S_1(a) + e^{-2/a} S_2(a) + \dots, \quad (4.1)$$

where each term $S_i(a)$ is a perturbation series. If $S_0(a)$ does not have a well-defined "sum" then the meaning of this kind of expansion is unclear, and may suffer from serious ambiguities. It would be worse still if the sum of $S_0(a)$ were the complete answer, since then (4.1) would represent double counting. Thus, the conclusion that perturbation theory has a natural sum, to which, however, other effects must be added, is just what one wants, phenomenologically.

[It is interesting to ask the questions of summability and recoverability with regard to (4.1). Does the right-hand side converge, or otherwise have a natural sum? If so, does one recover the physical answer? That is : is the expansion (4.1) complete, or are there still other effects?]

I hope after these remarks that the phenomenon of non-recoverability no longer seems strange or distasteful. The reader who remains unconvinced is urged to consider the large N^2 Gross-Neveu model, as described in the early sections of 36) [see also 27)]. Like QCD, this model is renormalizable and asymptotically free. Its perturbation theory is well-behaved, indeed convergent, but does not yield quite the right answer, because of dynamical mass generation which leads to "invisible", "higher-twist effects".

In conclusion, one needs to be realistic about the merits and failings of perturbation theory. Perturbation theory is an approximation scheme. It is not a definition of the theory, nor can it be made into one. It has major limitations which render it useless for all too many interesting questions. Nevertheless, within its own domain it is relatively practicable, and remarkably successful. The goal of this and my earlier work ²⁰⁾ has been to show that it also makes some kind of sense : it is not hopelessly arbitrary, and does not necessarily go beserk in high orders. However, to reach such a conclusion it is necessary to recognize the vital rôle of the renormalization group in "optimizing" the RS choice. One must discard the out-dated notion of perturbation theory as merely a power-series expansion.

ACKNOWLEDGEMENTS

It is a pleasure to thank A.J. Macfarlane A. Peterman, G.M. Shore, and J.W. van Holten for their helpful comments on the first draft of this work.

Order	Naive	$\bar{\tau}$	$\bar{a}$	Opt
1	0.25	-	-	-
2	0.188	1.00000	0.200	0.2000000
3	0.219	1.43845	0.184	0.2061121
4	0.195	1.59607	0.179	0.2054645
5	0.219	2.00000	0.167	0.2062757
6	0.189	2.18061	0.162	0.2061192
7	0.233	2.59875	0.152	0.2063073
8	0.156	2.75900	0.148	0.2062634
9	0.310	3.19736	0.139	0.2063235
10	- 0.036	3.33355	0.136	0.2063083
11	0.829	3.79230	0.128	0.2063320
12	- 1.550	3.90545	0.126	0.2063259
13	5.588	4.38352	0.119	0.2063367
14	-17.610	4.47541	0.118	0.2063340
15	63.581	4.97148	0.111	0.2063394

TABLE - Approximations to the alternating factorial series, Eq. (2.14), for $a_0 = 0.25$. The Borel sum of the series has the value 0.2063457.

APPENDIX

In this Appendix I comment briefly on the "fastest apparent convergence" (FAC) criterion in relation to the examples presented here. Historically, FAC criteria have often been used to fix the choice of unphysical variables, such as τ , in analogous situations. The basic idea is to force different orders of approximation to agree, so that the series - as far as it has been calculated - appears to converge as fast as possible. I have argued elsewhere that this idea is ill-conceived and unreliable ^{20),21)}. Recently, a popular view has been that, in the case of field theory, the principle of minimal sensitivity (PMS) and FAC criteria are essentially equivalent ³⁷⁾. Superficially, this view is supported by the following feature of the examples described here : - the PMS optimization condition $d\mathcal{R}^{(n)}/d\tau = 0$ turns out to be the same as requiring the last calculated coefficient to vanish ($r_{n-1}(\bar{\tau}_n) = 0$).

However, one should be cautious about attaching any deep significance to this fact. Firstly, it is a special consequence of the trivial " β function" in these examples, and so is an artificial, and not a realistic, feature. Secondly, there are at least two ways of interpreting the notion of "fastest apparent convergence" in the present context. [As in 20), I distinguish them as FAC and FAC'.] One can require approximants in adjacent orders to agree [$\mathcal{R}^{(n)} = \mathcal{R}^{(n-1)}$, i.e., $r_{n-1} = 0$: FAC] - which happens to coincide with PMS here. Alternatively, one may require the n^{th} order result to agree with the lowest order form [$\mathcal{R}^{(n)} = \underline{a}^{(n)}$: FAC']. The latter procedure in these examples would clearly be disastrous. The relative merits of FAC and FAC' here are exactly the reverse of what is found in another class of examples, described in Section IIB of 20). One can only conclude that neither method is reliable.

Moreover, in the field theory case, one has $(n-1)$ unphysical parameters at n^{th} order, and so one must impose $(n-1)$ conditions to fix the RS. The most obvious implementation of the FAC idea is then to require $r_1 = \dots = r_{n-1} = 0$. This procedure has been proposed by Grunberg ³⁸⁾. Since this method results in $\mathcal{R}^{(n)} = \underline{a}^{(n)}$, one can argue that it is more analogous to the disastrous FAC than to the successful FAC' ($\approx$ PMS) in these examples.

Finally, I remark that the suggested mechanism whereby "optimization" induces convergence of QCD perturbation theory cannot also operate for the Grunberg FAC method. The induced convergence results from a vanishing of the "optimized" expansion parameter, which - if it were also to occur in FAC - would lead to zero as the ultimate prediction for the physical quantity $\mathcal{R}$, simply because $\mathcal{R}^{(n)}$

and $\underline{a}^{(n)}$ are identified in FAC. On the other hand, if $a_{\text{FAC}}^{(n)}$ does not vanish as $n \rightarrow \infty$, then it is almost inconceivable that one could obtain convergence from a series which presumably has zero radius of convergence to begin with. Thus, if the "convergence-through-optimization" scenario has any validity, then PMS and FAC cannot remain numerically similar beyond low orders.

REFERENCES

- 1) F.J. Dyson - Phys.Rev. 85 (1952) 631.
- 2) C.A. Hurst - Proc.Roy.Soc.London 214A (1952) 44 ; Proc.Camb.Phil.Soc. 48 (1952) 625 ;
W. Thirring - Helv.Phys.Acta 26 (1953) 33 ;
A.J. Riddel, Jr., Phys.Rev. 91 (1953) 1243.
- 3) A. Peterman - Helv.Phys.Acta 26 (1953) 291.
- 4) A. Jaffe - Comm.Math.Phys. 1 (1965) 127.
- 5) B. Simon - Nuovo Cimento 59 (1969) 199 ; Ann.Phys. 58 (1970) 76.
- 6) C.M. Bender and T.T. Wu - Phys.Rev. 184 (1969) 1231 ; D7 (1973) 1620 ;
T. Banks, C.M. Bender and T.T. Wu - Phys.Rev. D8 (1973) 3346.
- 7) L.N. Lipatov - Sov.Phys. JETP 45 (1977) 216 ; JETP Letters 24 (1976) 157 ;
25 (1977) 104.
- 8) E. Brézin, J.C. Le Guillou and J. Zinn-Justin - Phys.Rev. D15 (1977) 1544 ;
1558 ;
A.J. McKane and D.J. Wallace - J.Phys. A11 (1978) 2285.
- 9) G. Parisi - Phys.Letters 66B (1977) 167.
- 10) C. Itzykson, G. Parisi and J.B. Zuber - Phys.Rev. D16 (1977) 996 ;
R. Balian, C. Itzykson, G. Parisi and J.B. Zuber - Phys.Rev. D17 (1978) 1041.
- 11) J. Zinn-Justin - Physics Reports 49 (1979) 205 ;
G. Parisi - Physics Reports 49 (1979) 216 ;
J. Zinn-Justin - Physics Reports 70 (1981) 109.
- 12) N.N. Khuri - Phys.Rev. D16 (1977) 1754.
- 13) P. Cvitanović - Nuclear Phys. B127 (1977) 176.
- 14) B. Lautrup - Phys.Letters 69B (1977) 109.
- 15) I.W. Herbst and B. Simon - Phys.Letters 78B (1978) 304.
- 16) A.S. Wightman - in The Whys of Subnuclear Physics (Erice, 1977), ed.
A. Zichichi (Plenum Press, New York, 1979).
- 17) G. 't Hooft, in "Deeper Pathways in High-Energy Physics" (Orbis Scientiae,
Coral Gables, 1977), ed. B. Kursunoglu, A. Perlmutter and W. Scott (Plenum
Press, New York, 1977) ; in "The Whys of Subnuclear Physics" (Erice 1977),
ed. A. Zichichi (Plenum Press, New York, 1977).
- 18) G. 't Hooft - Phys.Letters 109B (1982) 474.
- 19) E.C.G. Stueckelberg and A. Peterman - Helv.Phys.Acta 26 (1953) 449 ;
M. Gell-Mann and F. Low - Phys.Rev. 95 (1954) 1300 ;
N.N. Bogoliubov and D.V. Shirkov - "Introduction to the Theory of Quantized
Fields", (Wiley-Interscience, New York, 1959).
- 20) P.M. Stevenson - Phys.Rev. D23 (1981) 2916.

- 21) P.M. Stevenson - Nuclear Phys. B203 (1982) 472.
- 22) A. Peterman - CERN Preprint TH. 3202 (1981).
- 23) W. Celmaster and R.J. Gonsalves - Phys.Rev. D20 (1979) 1420 ; Phys.Rev. D21 (1980) 3112 ;
P. Pascual and R. Tarrach - Nuclear Phys. B174 (1980) 123 [E. B181 (1981) 546] ;
E. Braaten and J.P. Lveill - Phys.Rev. D24 (1981) 1369.
- 24) W. Celmaster and D. Sivers - Phys.Rev. D23 (1981) 227.
- 25) H.D. Politzer - Nuclear Phys. B194 (1982) 493.
- 26) W. Marciano and H. Pagels - Physics Reports 36 (1978) 137, Section 7.2.
- 27) P.M. Stevenson - Phys.Rev. D24 (1981) 1622.
- 28) F.W.J. Olver - "Asymptotics and Special Functions" (Academic Press, New York, 1974).
- 29) G.H. Hardy - "Divergent Series" (Oxford University Press, 1949).
- 30) G.A. Baker and J.L. Gammel (eds.), "The Pad Approximant in Theoretical Physics" (Academic Press, New York, 1970) ;
P.R. Graves-Morris (ed.), "Pad Approximants and their Applications" (Academic Press, New York, 1973).
- 31) A. Peterman - Phys.Letters 96B (1980) 137.
- 32) Yu.L. Dokshitser, D.I. D'Yakonov and S.I. Troyan - Physics Reports 58 (1980) ;
C.H. Llewellyn Smith - Schlading Lectures 1978, Acta Phys.Austriaca, Suppl. XIX (1978) 331 ;
A.B. Carter and C.H. Llewellyn Smith - Nuclear Phys. B162 (1980) 397.
- 33) H.D. Politzer - Nuclear Phys. B172 (1980) 349 ;
R.K. Ellis, W. Furmanski and R. Petronzio - CERN Preprints TH 3254 and 3301 (1982).
- 34) M.A. Shifman, A.I. Vainshtein and V.I. Zakharov - Nuclear Phys. B147 (1978) 385, 448, 519.
- 35) S.J. Brodsky and G.P. Lepage - Proc. of SLAC 1979 Summer Inst. on Particle Physics ;
E.L. Berger and S.J. Brodsky - Phys.Rev. D24 (1981) 2428.
- 36) D.J. Gross and A. Neveu - Phys.Rev. D10 (1974) 3235.
- 37) M.R. Pennington - Durham Preprint (Phys.Rev.D, to appear) ;
J. Kubo and S. Sakakibara - Dortmund Preprint (Phys.Rev.D, to appear) ;
G. Grunberg - Ecole Polytechnique Preprint A474.1281 (December 1981).
- 38) G. Grunberg - Phys.Letters 95B (1980) 70.

FIGURE CAPTIONS

- Figure 1 Approximants $\mathcal{R}^{(n)}(\tau)$, $n = 10, 11, 12, 14$, for the alternating factorial series example, shown as functions of the unphysical variable τ .
- Figure 2 Analytic structure of $f(\alpha)$ in Eq. (3.6), showing the accumulation of poles at the origin.

- Figure 1 -

- Figure 2 -