PHYSICAL REVIEW D VOLUME 23, NUMBER 12 15 JUNE 1981

Optimized perturbation theory

P. M. Stevenson
Physics Department, University of Wisconsin-Madison, Madison, Wisconsin 53706
(Received 21 July 1980; revised manuscript received 17 February 1981)

Conventional perturbation theory gives different results in different renormalization schemes, a problem which is especially serious in quantum chromodynamics (QCD). I propose a theoretical resolution of this ambiguity which uses the full renormalization-group invariance of the theory. The idea is that, in any kind of approximation scheme which does not respect the known invariances of the exact result, the "optimum" approximant is the one that is "most invariant," i.e., least sensitive to variations in the unphysical parameters. I discuss this principle in several examples, including the Halliday-Suranyi expansion for the anharmonic oscillator. Turning to massless field theories, I identify the unphysical variables which label a particular renormalization scheme as the renormalization point μ, and the β-function coefficients. I describe how perturbative approximations depend on these unphysical variables, and show how to find the stationary point which represents the "optimum" result. Certain renormalization-scheme invariants, in one-to-one correspondence with the perturbation-series coefficients, arise naturally in the analysis. An application to the e,μ magnetic moments in QED provides a partial test of these ideas, with encouraging results. I suggest possible further theoretical developments, and advocate the method as a sound basis for quantitative QCD phenomenology.

I. INTRODUCTION

Physical quantities in quantum chromodynamics (QCD), or indeed in any field theory, are independent of the particular scheme used to renormalize the theory. However, this exact invariance is respected only approximately in perturbation theory: Two nth-order results, calculated in two different renormalization schemes (RS's), will in general differ by a term of order $(n+1)$. The results of perturbative QCD calculations are therefore RS dependent.[1-8] This is a very serious problem. Of what use is a perturbative prediction if it depends entirely on an arbitrary choice of RS? It is essential to have an objective means of resolving the RS ambiguity if perturbative approximations are to have any real, quantitative meaning.

A resolution of the RS-ambiguity problem is *not* a matter of finding a "good expansion parameter for QCD." It is quite unjustified to assume that the same RS—and hence the same expansion parameter—must be used for every physical quantity. This false assumption is a holdover from QED, where it is traditional to use the canonical "on-shell" RS for everything. It is time to recognize that *the traditional* QED *procedure is wrong.* This has been implicit ever since the advent of the "running coupling constant."[9] One does not have to use the same renormalization point for different physical quantities: It can, and should, be adjusted to suit the energy scale of the process in question. Similarly, there is *no* consistency requirement that forbids one from using different RS's for different physical quantities. Therefore, one must answer the question of what is the best way to choose the RS *separately* for each individual physical quantity. Otherwise, one is not facing up to the issues raised by the RS ambiguity.

The aim of this paper is to show that the RS ambiguity can be resolved satisfactorily by purely theoretical means. The proposed method is simple, precise, and applicable at any order of perturbation theory. I see my analysis as a refinement and extension of the idea of "renormalization-group-improved perturbation theory"[9] and, to exaggerate slightly, as a new kind of perturbation theory, quite different from the traditional concept of a power-series expansion in a fixed parameter.

My starting point is the philosophy that the RS ambiguity is not so much a QCD problem as a problem in "approximation theory." The solution is therefore to be found, not in a study of Feynman diagrams, but rather by a careful reconsideration of the nature of perturbative approximations. The key idea is that one has *two* kinds of information about a physical quantity: (i) the first few terms of its perturbation expansion in some RS and (ii) the knowledge that the exact result is independent of RS (in other words, the invariance of physical quantities under the full renormalization group of Stueckelberg and Petermann[10]). I shall be concerned with the proper utilization of the second piece of information. The point is this: Since the true result is exactly RS independent, the best approximation is the one which is least sensitive to small changes in RS. This is a particular example of a quite general principle, applicable to any kind of approximation scheme involving "unphysical" parameters, i.e.,

parameters of which the true result is independent. I call this the "principle of minimal sensitivity" (PMS).

If an approximant depends on unphysical parameters, then their values should be chosen so as to minimize the sensitivity of the approximant to small variations in those parameters.

The motivation for this principle is the following. In the space of the unphysical parameters the exact result is a constant. Therefore the calculated result cannot possibly be a successful approximation where it is rapidly varying. The most reliable numerical result is likely to be where the calculation shows the correct qualitative behavior, i.e., where the approximate result is flattest. I discuss the justification of this principle in detail in Sec. II. As an example I show its success when applied to the Halliday-Suranyi expansion[11] for the anharmonic oscillator.

The application of this principle to field theory requires a system for labeling RS's so that the notion of a "small variation" in RS is meaningful. I address this problem in Sec. III and conclude that RS's can be labeled by their renormalization point[12] μ and by the values of their β-function coefficients. I also discuss the nature of perturbative approximations and how they depend on the RS parameters. In Sec. IV the formula for optimizing second- (i.e., next-to-leading) order results is derived. This is extended to third and higher orders in Sec. V. I show that for every (RS-dependent) coefficient in the perturbation series there is a corresponding RS invariant. These invariants play a central role in the analysis. Section VI describes some examples and special cases of the optimization formulas. In particular the existing QED calculations of electron and muon magnetic moments provide a test of the ideas developed here, with encouraging results. The conclusions are given in Sec. VII.

My notation is sometimes unorthodox and is described in detail in Appendix A, where I also give a careful discussion of the definition and meaning of the QCD Λ parameters. For convenience I employ a parameter $\tilde{\Lambda}$ defined slightly differently from the conventional Λ (see Appendix A). Another important piece of nonstandard notation is the use of (α_s/π) rather than α_s as the expansion parameter in the perturbation series. This is referred to as the "couplant" and denoted by the symbol a. Also, I use the expressions "first order," "second order," etc., as being synonymous with "leading order," "next-to-leading order," etc.

This paper addresses the problem of RS dependence in *massless* field theories with a *single* coupling constant. Although these two restrictions do simplify the problem considerably, I believe that the basic ideas can be generalized to any field theory. My main interest is with massless QCD (ignoring quark masses). Actually, in gauge theories such as QCD some extra considerations are involved: I discuss these in Appendix B and show that they do not change anything. In order to keep the presentation self-contained, only brief references are made in the text to the existing literature on the RS-dependence problem. I discuss the relation of this work to some other approaches to the RS ambiguity in Appendix C.

This paper is an improved and extended version of some unpublished work,[7] which in turn grew out of an earlier article[13] on the renormalization-group equations. The common theme of all this work is that the concept of "renormalization-group invariance"[10,14] is best understood and applied *without* invoking the apparatus of renormalization itself. The *idea* of renormalization is crucial in what follows, but its technicalities will be irrelevant.

II. THE PRINCIPLE OF MINIMAL SENSITIVITY

A. A lesson from the anharmonic oscillator

In this section I shall try to explain how and why the principle of minimal sensitivity works. I begin by describing the example which inspired the formulation of this principle. The problem is the calculation of the eigenvalues of the anharmonic-oscillator Hamiltonian

$$H = H_{\mathrm{SHO}}(m) + \frac{g}{4} x^4$$

with

$$H_{\mathrm{SHO}}(m) \equiv \tfrac{1}{2}(p^2 + m^2 x^2) . \tag{2.1}$$

A simple and ingenious expansion technique has recently been proposed by Halliday and Suranyi[11] (hereafter referred to as HS). Introducing an arbitrary frequency parameter Ω they write

$$H = H_0(\Omega) + H_{\mathrm{int}}(\Omega) , \tag{2.2}$$

where

$$H_0(\Omega) \equiv \frac{g}{\Omega^4} [H_{\mathrm{SHO}}(\Omega)]^2 , \tag{2.3}$$

$$H_{\mathrm{int}}(\Omega) \equiv H_{\mathrm{SHO}}(m) + \frac{g}{4} x^4 - \frac{g}{\Omega^4} [H_{\mathrm{SHO}}(\Omega)]^2 . \tag{2.4}$$

Using $H_0(\Omega)$ as the unperturbed Hamiltonian they then apply the usual Rayleigh-Schrödinger or Brillouin-Wigner perturbation theory. The calculations are most easily done by introducing creation and annihilation operators $a^\dagger$, a, of frequency Ω: Both $H_0(\Omega)$ and $H_{\mathrm{int}}(\Omega)$ are fourth-order polynomials in $a, a^\dagger$. The calculations are no more

difficult than the conventional perturbation expansion, based on $H_{\mathrm{SHO}}(m)$ as the unperturbed Hamiltonian. Unlike the conventional expansion, which diverges quite horribly,[15] the HS expansion has been proved to converge.[11] A peculiar feature is that the HS series is not an expansion in powers of any particular parameter of the theory.[11]

(The anharmonic oscillator, which can be regarded as a field theory in one dimension,[15] has long served as a testing ground for new ideas for solving field theories. Most such techniques have been horribly complicated, even for this very simple case. The HS expansion reduces the anharmonic-oscillator problem to the level of an undergraduate exercise. Moreover, the success of the method is very impressive, as will be seen. Consequently an extension of the technique to field theories would be a very exciting development. One stands to gain from a benevolent paradox: good, nonperturbative approximations can be obtained by a familiar technique—perturbation theory. This is, however, beyond the scope of the present paper.)

As both H_0 and H_{int} are Ω dependent, finite-order results in the HS scheme are also dependent on this arbitrary, unphysical parameter. I shall argue that the best choice of Ω is given by the PMS criterion. (This observation has recently been made, independently, by HS,[16] but, somewhat surprisingly, it was not noticed by them in Ref. 11, although they came very close to it. To quote from Ref. 11: "Order by order the perturbation theory ... depends on Ω ... . Clearly this fake dependence imposes strong renormalization-group-type constraints in our series, but we have been unable to use this." To fix Ω they instead adopted what I call a "fastest apparent convergence" criterion, requiring the sum of the corrections to the zeroth order term to vanish. This gives much poorer numerical results, as will be seen.)

It is convenient to replace Ω by a dimensionless parameter Z,

$$Z \equiv \Omega^3/g, \quad Z>0 \tag{2.5}$$

and to consider the Hamiltonian

$$H(Z,\eta) \equiv H_0(Z) + \eta H_{\mathrm{int}}(Z) \tag{2.6}$$

with H_0 and H_{int} as defined in Eqs. (2.3) and (2.4). The parameter η is introduced for purely pedagogical reasons: It has no physical significance. The Hamiltonian we are interested in is $H(Z,\eta=1) = H$, *which is independent of* Z. The Hamiltonian we can solve is $H_0 = H(Z,\eta=0)$. By Taylor-series expansion we can find an expression for the eigenvalue $E(Z,\eta)$, valid for sufficiently small values of η. The perturbative approximation consists of extrapolating this result all the way to $\eta=1$. Although we do not know what the true result for the eigenvalue E at $\eta=1$ actually is, we do know that it is a constant, independent of Z. Therefore we can judge the reliability of the extrapolation from $\eta=0$ by observing how flat the result is at $\eta=1$ (see Fig. 1). The result will not be flat everywhere, so one should choose Z to lie in the middle of the flattest portion of the curve. This is the PMS criterion.

Numerically this criterion is highly successful. Consider the ground-state energy of the quartic oscillator [i.e., $m=0$ in Eq. (2.1)]. This problem has been extensively studied numerically, so the approximate results can be compared with the exact result,[17]

$$E/g^{1/3} = 0.420805 \ldots . \tag{2.7}$$

For brevity I shall quote all results in units of $g^{1/3}$. The approximate results, up to third order in the HS expansion, have been calculated in Ref. 11. In zeroth order the result is

$$E^{(0)} = \tfrac{1}{4} Z^{-2/3}. \tag{2.8}$$

Clearly the PMS criterion gives no information on how to choose Z here. This is not surprising since $E^{(0)}$ is not really an approximation to E, but merely the eigenvalue of $H_0(\Omega)$, our chosen starting point. The first-order result does, however, give a genuine approximation,

$$E^{(1)} = Z^{-2/3}(\tfrac{3}{16} + \tfrac{1}{4} Z) \to 0.429, \tag{2.9}$$

where $E^{(1)}$ is defined as $E^{(0)}$ plus the first-order correction term. The single minimum of this function at $Z = \frac{3}{2}$ provides the optimum result, according to the PMS criterion. The result is good to 2% [although there would be no way of judging the reliability of (2.9) without knowing the true result]. In this case the PMS criterion picks out

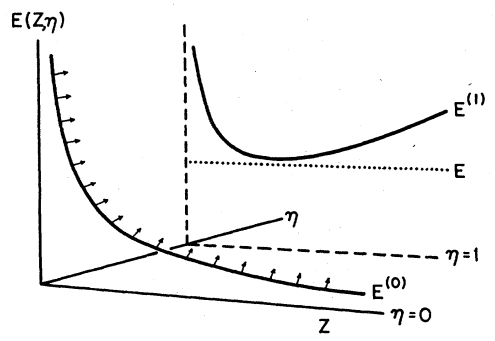

FIG. 1. The ground-state energy of the Hamiltonian $H(Z,\eta)$ of Eq. (2.6). The exact eigenvalue is known to be Z independent at $\eta=1$, and is $\frac{1}{4}Z^{-2/3}$ for $\eta=0$. The approximate eigenvalue $E^{(1)}$ is obtained by a Taylor-series expansion from $\eta=0$, but the extrapolation is reliable only in the region where $E^{(1)}$ is insensitive to Z.

the optimum result in the rigorous sense of minimizing the error, $|E^{(\text{approx})}-E|$. This is because first-order perturbation theory corresponds to a variational approximation

$$E^{(1)}=E^{(0)}+{}_\Omega\langle 0|H_{\text{int}}|0\rangle_\Omega={}_\Omega\langle 0|H|0\rangle_\Omega, \tag{2.10}$$

where $|0\rangle_\Omega$, the ground state of $H_0(\Omega)$, can be regarded as the trial wave function. The Rayleigh-Ritz principle guarantees that $E^{(1)}(Z)\geq E$, so the optimum $E^{(1)}$ is obtained by minimizing with respect to Z, in agreement with the PMS criterion.

To second order the result is

$$E^{(2)}=Z^{-2/3}\left(\tfrac{87}{640}+\tfrac{5}{16}Z-\tfrac{1}{48}Z^2\right)\rightarrow 0.426. \tag{2.11}$$

This function is flattest in the vicinity of its two turning points at $Z=(15\pm 9/\sqrt{5})/8$, $\simeq 1.3719$, 2.3781. The variation in the function between these two points is less than 0.002, so although the PMS criterion does not give a unique optimum value of Z, there is no real problem. In fact—without referring to the exact result—the size of this ambiguity, taken together with the change between first and second order, indicates that the error is of order of, say, 0.003. This is, in fact, quite a fair error estimate.

One should note that the function $E^{(2)}(Z)$ in Eq. (2.11) takes on all values from $+\infty$ to $-\infty$. At some point, therefore, it actually passes through the exact result. However, without knowing the exact result, there is no way of identifying this point. Thus it is still reasonable to claim that the PMS criterion indicates the "optimum" choice of Z, provided it is understood that the word optimum is used in the loose sense of "the best guess in the absence of further information." The strict sense of optimum as minimizing the modulus of the difference between the exact and approximate results is just not useful in the present context.

The third-order result is

$$E^{(3)}=Z^{-2/3}\left(\tfrac{573}{5120}+\tfrac{2253}{6400}Z-\tfrac{41}{960}Z^2+\tfrac{1}{288}Z^3\right)\rightarrow 0.421^{+0.003}_{-0.001}. \tag{2.12}$$

These results are illustrated in Fig. 2. (Note the tenfold expansion in scale between 0.4 and 0.45.) These curves are strong evidence for the correctness of the PMS criterion. It is quite remarkable how flat the curves become in the vicinity of the true result, given that, by the nature of the HS procedure, they must diverge at $Z=0$ and $Z=\infty$.

The numerical convergence from the first-order result, Eq. (2.9), towards the exact result may seem a little slow, but this cannot be helped. Indeed, the slow rate of improvement is mainly a reflection of the excellence of the first-order result. By contrast the fastest apparent convergence criterion employed by HS gives the results[11] $E\simeq 0.63$, 0.51, 0.45 in first, second, and third orders (see also Fig. 2) which are much poorer than even the first-order PMS result.

The HS expansion and the PMS criterion can also be successfully applied to the excited states of the anharmonic oscillator. For example, the first-order result for the kth energy level of the quartic oscillator is[11]

$$E_k^{(1)}(Z)=Z^{-2/3}\left[\tfrac{3}{8}(k+\tfrac{1}{2})^2+\tfrac{1}{2}Z(k+\tfrac{1}{2})+\tfrac{3}{32}\right]. \tag{2.13}$$

The PMS criterion leads to the optimum choice

$$Z=\frac{3}{8}\frac{[4(k+\frac{1}{2})^2+1]}{(k+\frac{1}{2})}. \tag{2.14}$$

The optimized approximation is then

$$E_k^{(1)}(\text{opt})=\tfrac{3}{4}(k+\tfrac{1}{2})\left(\frac{3}{8}\frac{[4(k+\frac{1}{2})^2+1]}{(k+\frac{1}{2})}\right)^{-1/3}. \tag{2.15}$$

A comparison with the exact eigenvalues is given in Table I. Note that the agreement is within 2% or better *over the entire spectrum*. Moreover, while the success of the PMS criterion for the ground state (and for the first excited state, because the parity under $x\rightarrow -x$ is a conserved quantum number) can be attributed to the Rayleigh-Ritz theorem, this is not true for the higher states. In fact, by examining the exact result, one finds that it lies just *above* the minimum of the curve (2.13) for $k\geq 2$. Thus the PMS criterion actually *maximizes* the error in a very local sense. However, and this is the point, the exact result lies only *just* above the minimum of $E_k^{(1)}(Z)$: The PMS criterion continues to work in a real, if undefinable, sense.

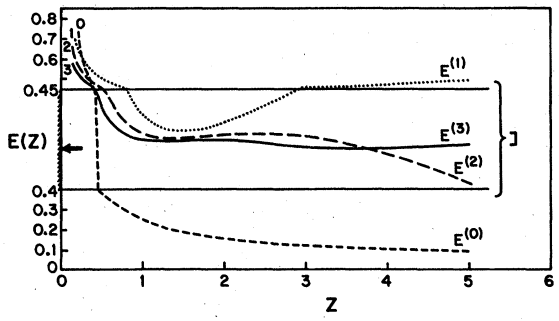

FIG. 2. Successive approximations to the ground state of the quartic oscillator in the HS expansion, as a function of the unphysical parameter $Z\equiv\Omega^3/g$. Note the tenfold expansion in the energy scale between 0.4 and 0.45. The exact result, indicated by the arrow, is independent of Z. The best approximations correspond to the flattest portions of the curves, as implied by the PMS criterion. By contrast, the fastest apparent convergence principle gives successive approximations corresponding to the intersection of the nth-order curve with the $E^{(0)}$curve.

Apart from noting the evident success of the PMS criterion, there are three important lessons to be drawn from this example. Firstly, the optimum value of Z is different for different energy levels [see Eq. (2.14)]. This is entirely natural: It corresponds to adjusting Ω so that $H_0(\Omega)$ gives a good fit to the full Hamiltonian in the relevant energy range. Secondly, even for a single energy level, the optimum value of Z is different in different orders [see Eqs. (2.9)–(2.12) and Fig. 2]. It becomes larger at higher orders because the ultimate convergence of the series is fastest at large Z.[11] The change is small, in the sense that the previous optimum Z gives a good approximation, although not the best. Finally, note that the operation of optimizing is not a linear one. If one adds or subtracts energy eigenvalues, then it matters whether one does the optimization before or afterwards. This may seem a little disturbing at first, but it seems fairly clear that it is most natural to adjust Z, and hence the energy parameter Ω, to each energy level separately, as was done above.

B. Truth without theorems?

Although the anharmonic-oscillator case discussed above provided the inspiration for this work, I emphasize that my arguments are *not* based on making analogies between the dynamics of the anharmonic oscillator and that of field theories. The argument is simply that the principle of minimal sensitivity embodies a general, abstract truth in the "theory of approximations."

Now, it is not possible to prove, or even formulate, a useful theorem to this effect. One is dealing with a situation in which the exact result is *unknown*. There is a strictly limited amount of information about the answer, and the problem is to make best use of that information to find the approximation which is most likely to be close to the exact result. It seems almost impossible to formulate this problem in precise mathematical terms (particularly since our information does not include a bound on the error). The situation is such that any useful, general statement can always be contradicted by a suitably pathological example.

For instance, any physicist seeing the series $f(x)=1+2x+x^2+\frac{5}{6}x^3+x^4+\cdots$ would assume that for $x=0.1$ the calculated terms give an answer to much better than 1% accuracy. However, a mathematician could point out that the right answer might be $f(x)=1/(1-x)+\sin x+e^{4\pi^2}x^5$, in which case the fourth-order result is monstrously inaccurate at $x=0.1$. The physicist would say that this is highly unlikely, but he could not explain what he meant by "unlikely."

TABLE I. The energy levels of the quartic oscillator. The approximate eigenvalues $E_k^{(1)}$ are obtained from first-order HS expansion using the PMS criterion [see Eq. (2.15)]. The exact results E_k are taken from Ref. 17, Table I, dividing by a factor of $4^{1/3}$. For $k=0,1$ the approximation is of the variational type so the error must be positive.

k	$E_k^{(1)}$	E_k	% error
0	0.429 27	0.420 805	2.01
1	1.526 9	1.507 901	1.26
2	2.951 4	2.958 796	−0.25
3	4.593 1	4.621 220	−0.61
10	19.754 8	19.944 208	−0.95
$k\to\infty$	$(0.858\,5)\times(k+\frac{1}{2})^{4/3}$	$(0.867\,145)\times(k+\frac{1}{2})^{4/3}$	−0.99

In the same way, when I talk of an optimum approximant as that "most likely" to represent the true result, I am necessarily using undefined terms. However, the very idea of perturbation theory is equally scandalous, mathematically speaking. Therefore I hope that the reader is open to persuasion that the PMS criterion represents a "pragmatic truth," even if that is a kind of truth which no respectable mathematician could recognize.

I can illustrate this pragmatic truth further in a simple class of examples. Consider the following problem: One wishes to calculate

$$F\equiv\int_A^B f(x)dx\,, \tag{2.16}$$

where the function $f(x)$ is not known exactly: $f(x)$ is known to be analytic on $[A,B]$ (where A,B are given, and $A<B$), and one has a means of calculating the Taylor expansions about $x=A$ and $x=B$. That is all the available information about $f(x)$.

By writing F as

$$F=\int_A^\xi f(x)dx+\int_\xi^B f(x)dx\,, \tag{2.17}$$

one can form approximations to F by replacing $f(x)$ by its Taylor series about A and B in the first and second terms, respectively. That is, the (n,m)th approximation to F is

$$F^{(n,m)}(\xi)\equiv\int_A^\xi f_A^{(n)}(x)dx+\int_\xi^B f_B^{(m)}(x)dx\,, \tag{2.18}$$

with

$$f_X^{(n)}(x)\equiv\sum_{i=0}^{n}\frac{1}{i!}(x-X)^i\left(\frac{d^if}{dx^i}\right)\Big|_{x=X} \tag{2.19}$$

(see Fig. 3).

The approximants $F^{(n,m)}(\xi)$ depend on the choice of ξ, even though F manifestly does not. Of

course, for any fixed ξ one can obtain an arbitrarily good approximation to F by calculating sufficient terms in the two Taylor series. However, that is not the issue. The problem is to get the best "value for effort": Given only a small finite number of terms in each series, which value of ξ is likely to be the best? This problem is a paradigm of the situation in which the principle of minimal sensitivity is supposed to apply. Within this framework I can make some general remarks about the PMS criterion, before proceeding to specific examples.

Firstly, in this context the PMS criterion states that the approximant $F^{(n,m)}(\xi)$ should be evaluated at the point $\xi = \bar{\xi}$ where it is flattest. Essentially this means that the optimum point $\bar{\xi}$ is characterized by

$$\left.\frac{dF^{(n,m)}}{d\xi}\right|_{\xi=\bar{\xi}} = 0. \tag{2.20}$$

Actually this is an oversimplification—although one which is adequate later on in the paper. By the nature of the approximation scheme one would only consider solutions for $\bar{\xi}$ in the range A to B. There may be no real solutions to Eq. (2.20) in this range, in which case one seeks the point at which $dF^{(n,m)}/d\xi$ is minimized, as the point where the approximant is least sensitive to small variations in ξ. (If all derivatives are monotonic, as frequently happens in the lowest order of many approximation schemes, then the approximant is irredeemably ambiguous.)

Equation (2.20) may have two or more solutions with $A < \bar{\xi} < B$. This is interpreted as the approximant undulating around, or near to, the exact result (cf. the anharmonic-oscillator example in second and higher orders). The size of the undulations indicates the scale on which the approximant successfully mimics the constant, exact result. Therefore, although the PMS criterion is then somewhat vague—indicating that ξ should be chosen somewhere in the middle of the flat region—the uncertainty is of the order of the *intrinsic* error of the approximation. [The situation discussed in this paragraph does not seem to occur in the field theory case (see Sec. VD).]

[In the integration example the geometrical interpretation of the PMS condition (2.20) is simple: $\bar{\xi}$ corresponds to the intersection(s) of the Taylor series $f_A^{(n)}(x)$ and $f_B^{(m)}(x)$. This is intuitively a good choice for ξ (see Figs. 3 and 4).]

Secondly, the approximation scheme discussed here clearly assumes that $f(x)$ is smooth and does not have a great deal of structure between A and B. If this is not true, the approximations will be intrinsically poor, particularly in low orders. In such circumstances the PMS criterion yields poor numerical results—but then, so would any other criterion, except by a pure fluke. Thus, it is not that the PMS criterion fails, but rather that the approximation itself is inadequate. The result for F is poor because of the inadequacy of our information about $f(x)$, not because we have failed to make best use of it. (The element of tautology in

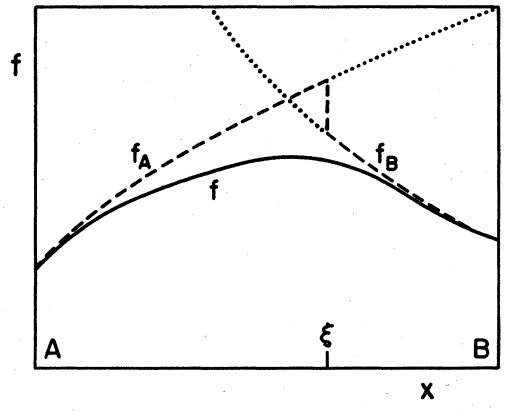

FIG. 3. The integration example discussed in Sec. IIB. The sketch shows a function $f(x)$ and its finite-order Taylor series f_A, f_B. The integral of f is approximated by the area under f_A from A to ξ, plus the area under f_B from ξ to B.

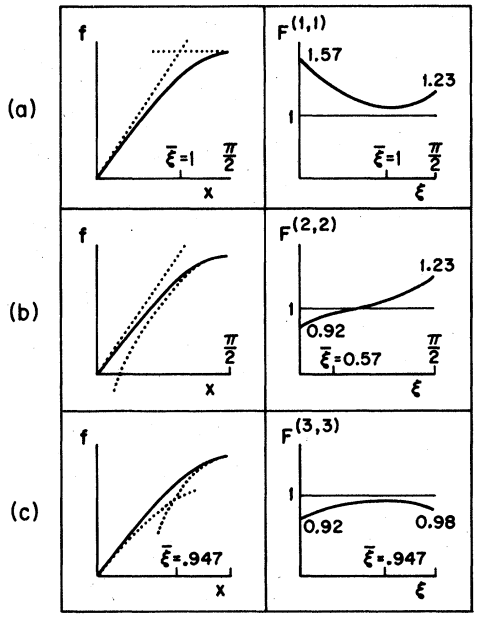

FIG. 4. The integration example for $f(x) = \sin x$, $0 < x < \pi/2$. The left-hand graphs show $\sin x$ and its Taylor-series approximations in first, second, and third orders. The right-hand graphs show the resulting approximations to $\int_0^{\pi/2} \sin x\, dx$, as a function of the break point ξ. The optimum choice $\bar{\xi}$ is indicated. (The graphs are sketches, and are not drawn to scale.)

this argument is inevitable, since without something such as the PMS criterion the approximation has no precise, numerical meaning.)

Thirdly, I can use the present problem to contrast the PMS criterion with the "fastest apparent convergence" (FAC) criterion—the idea that ξ should be chosen to make different orders of approximation agree. I argue against the FAC criterion for several reasons: (i) Higher-order approximants are, in some sense, better than lower-order approximants, so one should expect them to differ by some intrinsic amount representing that improvement. By forcing different orders to agree one runs the risk of the better (higher-order) approximant being "dragged down" by the poorer one. In contrast, the PMS criterion gives a meaning to each approximant, based on its own properties, whether or not it is one of a series of approximants. (ii) It is somewhat artificial to force the higher-order corrections to vanish, since there is no reason to suppose that the approximations "want" to behave in this way (whereas they do want to become flat). The FAC criterion appeals to a property which we would *like* the result to have—rapid convergence—whereas the PMS criterion appeals to a property which the exact result *does* have—exact insensitivity to variations in unphysical parameters, such as ξ. (iii) The results in different orders of approximation can quite often coincide for reasons largely, or even entirely, to do with the nature of the approximation scheme. The FAC criterion is then wholly misleading as to where the true result lies. This phenomenon is quite common in the integration example, and also occurs in the anharmonic oscillator case: In Fig. 2 the intersection of the higher-order results with the zeroth-order curve is strongly influenced by the fact that all the approximants diverge in a similar fashion as $Z \to 0$, which is purely an artifact of the approximation method.

Turning now to a specific example, I consider approximations to

$$F \equiv \int_0^{\pi/2} \sin x \, dx = 1 , \tag{2.21}$$

using Taylor expansions about $x = 0$ and $\frac{1}{2}\pi$. The zeroth-order result $F^{(0,0)}(\xi) = \pi/2 - \xi$ is, as usual, irredeemably ambiguous. In first-order one obtains

$$F^{(1,1)}(\xi) = \int_0^{\xi} x \, dx + \int_{\xi}^{\pi/2} 1 \, dx = \tfrac{1}{2}\xi^2 + (\tfrac{1}{2}\pi - \xi) , \tag{2.22}$$

which has a stationary point at $\xi = 1$, giving $F_{\text{opt}}^{(1,1)} = \frac{1}{2}(\pi - 1) \simeq 1.07$ [see Fig. 4(a)]. In this case the PMS result is optimal in the strict sense. The reason is obvious geometrically. A function which is convex in the region A to B lies below the tangents at A and B, so that $F \leqslant F^{(1,1)}(\xi)$. Therefore the optimum result is obtained by *minimizing* $F^{(1,1)}(\xi)$. A parallel argument applies if $f(x)$ is concave on (A, B), implying $F \geqslant F^{(1,1)}(\xi)$, when the optimum is obtained by *maximizing* $F^{(1,1)}(\xi)$. These statements are analogs of the Rayleigh-Ritz theorem, which applies to the ground state of the anharmonic oscillator. Such Rayleigh-Ritz inequalities seem to be common, and one should be alert to the possibility of proving them in specific, practical contexts.

Notice that the FAC criterion, which requires $F^{(0,0)}(\xi)$ and $F^{(1,1)}(\xi)$ to agree, would lead one to choose $\xi = 0$, yielding $F_{\text{FAC}}^{(1,1)} = \frac{1}{2}\pi \simeq 1.57$. This is positively the *worst* first-order result. The reason for the agreement between $F^{(0,0)}$ and $F^{(1,1)}$ at $\xi = 0$ is obvious geometrically, and is clearly an artifact of the approximation method, illustrating a point made above.

The second-order approximant

$$F^{(2,2)}(\xi) = \tfrac{1}{2}\xi^2 + (\tfrac{1}{2}\pi - \xi) - \tfrac{1}{6}(\tfrac{1}{2}\pi - \xi)^3 \tag{2.23}$$

has no stationary point—corresponding to the Taylor series not intersecting [see Fig. 4(b)]. The approximant $F^{(2,2)}(\xi)$ is least sensitive to small changes in ξ where its slope is minimum, which occurs at $\xi = \frac{1}{2}\pi - 1$. This yields $F_{\text{opt}}^{(2,2)} = \frac{1}{8}\pi^2 - \frac{1}{2}\pi + \frac{4}{3} \simeq 0.9962$. Proceeding to third order one obtains

$$F^{(3,3)}(\xi) = -\tfrac{1}{24}\xi^4 + \tfrac{1}{2}\xi^2 + (\tfrac{1}{2}\pi - \xi) - \tfrac{1}{6}(\tfrac{1}{2}\pi - \xi)^3 , \tag{2.24}$$

which has a maximum at $\xi \simeq 0.947$ [see Fig. 4(c)], giving $F_{\text{opt}}^{(3,3)} = 0.9982$.

FAC-based criteria are less successful. Beyond first order, there are two ways of interpreting the FAC idea. One can require approximants in *adjacent* orders to agree (FAC′), which gives $\xi = 0$ or $\frac{1}{2}\pi$ always, and is clearly disasterous. The alternative, choosing ξ by making $F^{(n,n)}(\xi)$ agree with *zeroth* order (FAC), is not as bad. (However, the relative merits of FAC and FAC′ in the anharmonic-oscillator case seem to be the reverse.) In second order one finds $F_{\text{FAC}}^{(2,2)} = 0.9965$, which is marginally better than the PMS result, 0.9962. However, the significance of this is made dubious by the retrogression at third order, which gives 0.9921 with this criterion—the error being four times that of the PMS result.

One could compute still higher orders, and consider approximants $F^{(n,m)}(\xi)$ for $n \neq m$, but this adds little to the discussion. I have studied several other specific examples, and find that the

features discussed above are not atypical. The reader is encouraged to investigate these points in his own examples.

One other instructive example is worth describing here:

$$F \equiv \int_0^\pi \sin^2 x\, dx = \tfrac{1}{2}\pi \simeq 1.57 . \tag{2.25}$$

This is a somewhat "pathological" case, in that $\sin^2 x$ does have considerable structure between 0 and π, and is inadequately represented by the low-order Taylor series about these end points. In fact the initial results $F^{(0,0)}$ and $F^{(1,1)}$ are both zero. Being ignorant of the exact result, one might think that all is well because (i) both results are perfectly flat in ξ, and (ii) they agree exactly. This false picture is dispelled by a second-order calculation, which gives

$$F^{(2,2)}(\xi) = \tfrac{1}{3}\xi^3 + \tfrac{1}{3}(\pi - \xi)^3 . \tag{2.26}$$

The PMS criterion yields $F^{(2,2)}_{\text{opt}} = \frac{1}{12}\pi^3 \simeq 2.58$, which, although poor, is the best available result. The large variation of $F^{(2,2)}(\xi)$ for $0 \leq \xi \leq \pi$, and the big change from first to second order, are danger signs that the second-order result is not reliable. Several more terms in the Taylor series are needed for an accurate approximation in this particular case.

The moral is that one can never be *sure* with approximations that all is going well: Problems could always arise in the next order. One can only (i) calculate as many terms as humanly possible, (ii) use that information, *and any other information*, as sensibly as possible, and (iii) hope for the best. The PMS criterion comes in at stage (ii), utilizing the known invariance of the exact result, but there is no way it can eliminate stages (i) and (iii).

"Noninvariant approximations"—ones involving supposedly arbitrary parameters which actually affect the approximate results—occur in many branches of theoretical physics. Various methods have been used to deal with the problem in specific contexts.[18] The special features of any particular problem must, of course, be considered, because they may provide extra information about the exact result. Nevertheless, I argue that the principle of minimal sensitivity is a basic, general tool for making sense of noninvariant approximations.

III. APPLYING THE PMS CRITERION IN FIELD THEORY

A. Setting up the problem

Having described the justification for the PMS criterion, I now consider how to apply it to the problem of RS dependence in field-theoretic perturbation theory. I shall talk in terms of QCD, but the discussion applies equally to any massless, renormalizable field theory with a single coupling constant and a perturbative β function which begins with an a^2 term (of either sign). Actually, there are some extra complications connected with the gauge nature of QCD which are ignored here. Appendix B deals with these subtleties.

For definiteness I shall often use $R \equiv \sigma(e^+e^- \to \text{hadrons})/\sigma(e^+e^- \to \mu^+\mu^-)$ as an example. (Really I use the "theorist's R", appropriate to a world without heavy-flavor thresholds.) R is a function of one physical variable Q, the total e^+e^- energy, although I usually suppress this dependence to simplify the notation. All of the following analysis is performed at fixed Q. For convenience I define

$$R \equiv \left(3\sum_i q_i^2\right)(1 + \mathcal{R}) \tag{3.1}$$

so that $\mathcal{R}$ is given by a perturbation expansion of the form

$$\mathcal{R} = a(1 + r_1 a + r_2 a^2 + \cdots) . \tag{3.2}$$

In fact, the analysis will apply to any physical quantity having a perturbation series of this form. Later I shall generalize the results to the case where the series begins $a^N(1 + r_1 a + \cdots)$ for arbitrary N.

In Eq. (3.2) the couplant $a \equiv \alpha_s/\pi$ is RS dependent. The coefficients $r_1, r_2, \ldots$ are also RS dependent. In the full series these RS dependences exactly cancel, because *physical quantities are independent of RS*. However, these cancellations occur *between* different orders. Thus when the series is truncated, the resulting approximant *does* depend on RS. This is just the kind of situation in which the PMS criterion applies: $\mathcal{R}$ itself is independent of RS, so the optimum ith-order approximant $\mathcal{R}^{(i)}$ is the one which is least sensitive to small changes in RS. That is, one should look for an optimum RS characterized by

$$\left.\frac{\partial \mathcal{R}^{(i)}}{\partial(\text{RS})}\right|_{\text{RS=Optimum RS}} = 0 . \tag{3.3}$$

The problem is, of course, to decide what is meant by $\partial/\partial(\text{RS})$.

Therefore, I first need to explain how to translate RS dependence into dependence on a certain number of RS parameters. This is perhaps best explained by stating the solution first, and then justifying it.

For my purposes RS's can be labeled by a countable infinity of parameters $(\tau, c_2, c_3, \ldots)$, defined as follows. In any kind of RS there is a parameter μ with dimensions of mass—the "renormalization point."[12] The dependence of the couplant $a \equiv \alpha_s/\pi$ on the renormalization point is given by the β-function equation

$$\mu\frac{\partial a}{\partial \mu} = \beta(a)\,, \tag{3.4}$$

$$\beta(a) \equiv -ba^2(1 + ca + c_2a^2 + c_3a^3 + \cdots)\,. \tag{3.5}$$

The boundary condition on Eq. (3.4) is not provided by the theory, but may be expressed in terms of a (μ-independent but RS-*dependent*) scale parameter $\tilde{\Lambda}$, which must be fitted to experiment. [$\tilde{\Lambda}$ is related to the conventional Λ parameter, as defined by Buras *et al.*,[4] by

$$\tilde{\Lambda} = \Lambda\left(\frac{2c}{b}\right)^{-c/b}. \tag{3.6}$$

I discuss the definition and meaning of $\tilde{\Lambda}$ in detail in Appendix A.]

The first RS parameter is τ,

$$\tau \equiv b\ln(\mu/\tilde{\Lambda})\,, \tag{3.7}$$

which represents the same "degree of freedom" in the choice of RS as the renormalization point μ, but proves to be a more convenient variable to use.

As is well known, the first two β-function coefficients, b and c in Eq. (3.5), are independent of the RS.[19] However, the higher-order coefficients $c_2, c_3, \ldots$ are RS dependent. Turning this statement around: Different RS's are characterized by different values of $c_2, c_3, \ldots$. These coefficients can therefore be used to label RS's, and they become the other RS parameters. The dependence of a on these parameters can be written as

$$\frac{\partial a}{\partial c_i} \equiv \beta_i(a)$$

$$= \frac{1}{(i-1)}\,a^{i+1}(1 + W_1^i a + W_2^i a^2 + \cdots)\,. \tag{3.8}$$

The functions $\beta_i(a)$ can be determined from the requirement that $\tau; c_2, c_3, \ldots$ should form a set of mutually independent variables with commuting partial derivatives, as demonstrated in Sec. III B. I prove in Sec. III D that these variables give a complete parametrization of RS's as far as physical quantities are concerned, and also that ith-order approximations depend only on $(i-1)$ of these RS parameters (for $i \geq 2$).

In Sec. III B (which may be omitted on a first reading) I discuss the motivation for using $c_2, c_3, \ldots$ as RS parameters, and determine the $\beta_i(a)$ functions. In Secs. III C and III D I define ith-order perturbative approximations, and describe how they depend on the RS parameters.

B. The parametrization of renormalization scheme dependence

As explained in Sec. III A, I require some method of labeling RS's by a set of parameters so that the notion of a small variation in RS is meaningful. In other words, I need an explicit realization of the full renormalization group of Stueckelberg and Petermann (Ref. 10 and see also Refs. 20 and 21). The solution to this problem has already been outlined above, but here I describe the derivation in more detail.

The problem of parametrizing RS dependence is simplified by three restrictions: (i) the restriction to physical quantities, (ii) the restriction to massless theories, and (iii) the restriction to perturbation theory. Obviously I shall only want to consider the optimization of perturbative approximations to *physical* quantities. (Moreover, it is only for physical quantities that one *can* apply the PMS criterion, since only these quantities are RS independent.) Consequently, I need only consider two RS's as different if they correspond to different definitions of the couplant a ($\equiv \alpha_s/\pi$), i.e., different definitions of the finite part of the renormalization constant Z in the relation $a = Za_{\text{bare}}$. Differences in the definitions of the other renormalization constants will be irrelevant since, although these affect the Green's functions, the differences will cancel order-by-order in expressions for physical quantities.

If mass terms are present in the theory, then different ways of renormalizing the mass could also affect the calculations of physical quantities. I avoid this complication by restricting myself to massless theories in this paper. Thus I shall use the phrase "different RS's" as synonymous with "different definitions of the renormalized couplant a."

The restriction to perturbation theory is important. It allows me to assume that anything and everything can be expressed as a power series in a (in whatever RS). In particular, the couplant a' in some general RS must be expressible as a power series in the couplant a of some "base" scheme, i.e.,

$$a' = a(1 + v_1a + v_2a^2 + \cdots)\,. \tag{3.9}$$

In a sense the coefficients $v_1, v_2, \ldots$ label the "primed" scheme, but this is not satisfactory because it singles out the arbitrary base scheme as special. The point is not that the a scheme corresponds to the special values $v_1 = 0$, $v_2 = 0, \ldots$, but rather that by regarding $v_1, v_2, \ldots$ as independent variables (in the sense that differentiation with respect to v_1 is implicitly performed with v_2, $v_3, \ldots$ held constant) one would be giving the a scheme a unique status.

However, it seems natural to assume that each coefficient $v_1, v_2, \ldots$ corresponds to a distinct degree of freedom in the specification of one RS in terms of another. This suggests that there exists

a hierarchy of parameters $v_1, v_2, \ldots$ such that $\partial a/\partial v_1 = O(a^2)$, $\partial a/\partial v_2 = O(a^3)$, etc., although one must abandon Eq. (3.9) as a means of defining them. I have already argued that the parameter v_1 should be identified with $\tau \equiv b \ln(\mu/\tilde{\Lambda})$, since this gives an $O(a^2)$ variation:

$$\frac{\partial a}{\partial \tau} = \frac{\beta(a)}{b} = -a^2(1 + ca + c_2 a^2 + \cdots). \tag{3.10}$$

By analogy, one can write down a similar expansion, with unknown coefficients, for the dependence of a on the other parameters, whatever they may be:

$$\frac{\partial a}{\partial v_i} = N_i a^{i+1}(1 + W_1^i a + W_2^i a^2 + \cdots). \tag{3.11}$$

The previous discussion suggests that one needs to consider the mutual independence of any candidates for RS parameters. Consequently, I next investigate the constraints imposed by requiring $(\tau; v_2, v_3, \ldots)$ to be independent variables. In particular, the partial derivatives

$$\frac{\partial}{\partial \tau} \equiv \frac{\partial}{\partial \tau}\bigg|_{v_i = \text{const}} \quad \text{and} \quad \frac{\partial}{\partial v_i} \equiv \frac{\partial}{\partial v_i}\bigg|_{\tau, v_j = \text{const},\ j \neq i} \tag{3.12}$$

must commute. It is easy to see, by differentiating (3.10) and (3.11) with respect to v_i and τ, respectively, that this requires that either the coefficients $N_i, W_1^i, W_2^i, \ldots$ are τ dependent, or the coefficients $c_2, c_3, \ldots$ are v_i dependent, or both. The first and last possibilities can be rejected immediately because they imply that the W's are μ dependent, and dimensionless quantities such as the W's can only depend on μ via $a(\mu)$, since there are no other massive variables in the problem. (This is just dimensional analysis, or, more correctly, "one-parameter ambiguous dimensional analysis" in the sense of Ref. 13.) But a dependent W's would contradict the requirement that $\partial a/\partial v_i$ must have an expansion in powers of a. Therefore the β-function coefficients $c_2, c_3, \ldots$ must be v_i dependent, i.e., they are RS dependent, as is well known. However, one recovers the usual result[19] that (like b) c is an invariant [see Eq. (3.13) below].

By evaluating $\partial^2 a/\partial\tau\partial v_i$ and $\partial^2 a/\partial v_i \partial\tau$ and equating coefficients of $a^3, a^4, \ldots, a^{i+2}, a^{i+3}, \ldots$ one obtains the following equations ($i = 2, 3, \ldots$):

$$\frac{\partial c}{\partial v_i} = \frac{\partial c_2}{\partial v_i} = \cdots = \frac{\partial c_{i-1}}{\partial v_i} = 0, \tag{3.13}$$

$$\frac{\partial c_i}{\partial v_i} = N_i(i-1), \tag{3.14}$$

$$\frac{\partial c_{i+j}}{\partial v_i} = N_i \sum_{r=0}^{j} (i + j - 1 - 2r) c_r W_{j-r}^i, \tag{3.15}$$

with $j = 1, 2, \ldots$ and $c_1 \equiv c$, $c_0 \equiv W_0^i \equiv 1$.

The most striking result is that c_i depends linearly on v_i. This suggests a very appealing solution, namely, to *identify* v_i *with* c_i. (Later results are in fact independent of this specific choice: see Sec. V.) This is a self-consistent solution, as I shall now show. The independence of the v_i from τ was used to derive Eqs. (3.13)–(3.15), but one also requires the mutual independence of the v_i, i.e., $\partial v_i/\partial v_j = 0$, $i \neq j$. This will be assured by Eqs. (3.13) and (3.15), provided the RHS of Eq. (3.15) is equal to zero, which leads to a set of linear equations that determine the W_j^i coefficients uniquely, i.e., identifying

$$v_i = c_i \Rightarrow N_i = 1/(i-1) \tag{3.16}$$

and

$$W_j^i = \frac{-1}{(i+j-1)} \sum_{r=1}^{j} (i + j - 1 - 2r) c_r W_{j-r}^i, \tag{3.17}$$

$j = 1, 2, \ldots$ with $c_1 \equiv c$, $W_0^i \equiv 1$.

The solution to these equations can be expressed as a determinant (see Appendix A). The variation $\partial a/\partial c_i$ therefore has a well-determined series expansion,

$$\frac{\partial a}{\partial c_i} = \frac{1}{(i-1)} a^{i+1}\left[1 - \frac{(i-2)}{i} ca + \left(\frac{(i-1)(i-2)}{i(i+1)} c^2 - \frac{(i-3)}{(i+1)} c_2\right) a^2 + \cdots\right]. \tag{3.18}$$

Notice that $i = 1$ in Eq. (3.18) reproduces the β function, Eq. (3.10), apart from the singular normalization. Thus one has the amusing situation that c_1 behaves in one sense like $\tau \equiv b \ln(\mu/\tilde{\Lambda})$, and in another sense like the invariant c (cf. the way x^0 in calculus formulas behaves like both $\ln x$ and 1). I will always use $c_1 = c$, as in Eqs. (3.15) and (3.17).

In fact the above analysis need not be phrased in terms of series expansions. Defining

$$\frac{\partial a}{\partial \tau} \equiv \hat{\beta}(a), \quad \frac{\partial a}{\partial c_i} \equiv \beta_i(a), \tag{3.19}$$

where $\hat{\beta}(a) \equiv \beta(a)/b$, the commutation of the partial derivatives requires

$$\beta_i'(a)\hat{\beta}(a) = \hat{\beta}'(a)\beta_i(a) - a^{i+2}, \tag{3.20}$$

where the prime indicates differentiation with respect to a, regarding the c_i as fixed. This leads to the expression

$$\beta_i(a) = -\hat{\beta}(a) \int_0^a \frac{x^{i+2} dx}{[\hat{\beta}(x)]^2} \tag{3.21}$$

for β_i in terms of $\hat{\beta}$. It is easily verified that a formal expansion of this equation leads to Eq. (3.18).

The problem posed at the beginning of this subsection is now solved: RS's can be labeled by a countable infinity of parameters $(\tau; c_2, c_3, \ldots)$. The dependence of a on these parameters is, in principle, fully specified by Eqs. (3.10) and (3.21). A simple argument, given in Sec. III D, shows that this parametrization of RS's is a complete parametrization for physical quantities, and that only $(i-1)$ parameters are relevant in ith order, for $i \geq 2$.

C. Definition of perturbative approximations

Before I can describe how to optimize perturbative approximations by the best choice of RS, I must define precisely what I mean by a perturbative approximation. Consider for the moment working in one particular renormalization scheme. In this scheme the couplant satisfies

$$\frac{\partial a}{\partial \tau} = \frac{\beta(a)}{b} = -a^2(1 + ca + c_2 a^2 + c_3 a^3 + \cdots), \quad (3.22)$$

and the physical quantity $\mathcal{R}$ has the expansion

$$\mathcal{R} = a(1 + r_1 a + r_2 a^2 + r_3 a^3 + \cdots). \quad (3.23)$$

The reason we have to resort to approximations is because we only know a finite number of the coefficients $c, c_2, c_3, \ldots$ and $r_1, r_2, r_3, \ldots$. Therefore it is most natural to define the ith-order perturbative approximation $\mathcal{R}^{(i)}$ by (i) truncating the series (3.23) at ith order, and (ii) replacing a by its ith-order approximant, defined as the solution to Eq. (3.22) with the RHS truncated at ith order.

Having adopted this definition of $\mathcal{R}^{(i)}$, it is important to recognize that no further analytic approximations are permitted. In particular I cannot now approximate the solution to

$$\frac{\partial a^{(i)}}{\partial \tau} = -(a^{(i)})^2[1 + ca^{(i)} + \cdots + c_{i-1}(a^{(i)})^{i-1}] \quad (3.24)$$

by, say, expanding $a^{(i)}$ in inverse powers of τ and keeping only i terms. To do so, i.e., to redefine $\mathcal{R}^{(i)}$ by reexpanding in powers of $1/\ln(\mu/\tilde{\Lambda})$ would be unnecessary, unmotivated, and *dangerous*—because it loses some of the information we have. The point is that $1/\ln(\mu/\Lambda)$ is a very arbitrary and artificial choice of expansion parameter, since the precise definition of Λ is a matter of convention (as witnessed by my use of a Λ parameter $\tilde{\Lambda}$ defined differently from the conventional Λ: see Appendix A, paragraph 3). The same point has been made by Celmaster and Sivers [see Ref. 6, especially following Eq. (2.34)].

My insistence on respecting Eq. (3.24) exactly means that transcendental equations will appear. In practice this means that one must resort to numerical methods. Although this is troublesome, the effort involved is much less than that already invested in calculating perturbative coefficients, so I do not regard this as a drawback. The only approximations that are forced on us are the truncation of the series (3.22) and (3.23). Any other approximations can only be justified by showing that they have negligible *numerical* effect on the final result.

D. Scheme dependence of perturbative approximations

Having defined the perturbative approximation $\mathcal{R}^{(i)}$ in a particular RS, I now consider how $\mathcal{R}^{(i)}$ varies when the RS is changed. In particular I will prove the statement that only $(i-1)$ RS parameters are relevant in ith order $(i \geq 2)$. From the definition of $a^{(i)}$, Eq. (3.24), it is manifestly independent of $c_i, c_{i+1}, \ldots$, and clearly can depend on RS *only* through the $(i-1)$ parameters $(\tau; c_2, \ldots, c_{i-1})$. Therefore $\mathcal{R}^{(i)}$, defined by

$$\mathcal{R}^{(i)} \equiv a^{(i)}[1 + r_1 a^{(i)} + \cdots + r_{i-1}(a^{(i)})^{i-1}], \quad (3.25)$$

will be a function of $\tau; c_2, \ldots, c_{i-1}$ too. It remains to show that $\mathcal{R}^{(i)}$ cannot depend on RS through any other parameters.

Consider perturbative approximations $\mathcal{R}^{(i)}$ and $\mathcal{R}^{(i)\prime}$ in two different RS's. Since[22]

$$\mathcal{R} = \mathcal{R}^{(i)} + O(a^{i+1}) \quad (3.26)$$

and

$$\mathcal{R} = \mathcal{R}^{(i)\prime} + O(a^{i+1}),$$

it follows that $\mathcal{R}^{(i)}$ and $\mathcal{R}^{(i)\prime}$ can differ only by a term which is, formally, $O(a^{i+1})$. This self-consistency condition, which I can write symbolically as

$$\frac{\partial \mathcal{R}^{(i)}}{\partial (\mathrm{RS})} = O(a^{i+1}), \quad (3.27)$$

has important consequences. It shows immediately that the coefficients r_j can depend on RS *only* through the parameters $\tau; c_2, \ldots, c_j$. Suppose this were not the case, i.e., suppose there existed some mysterious, extra parameter η which was part of the characterization of the RS, yet independent of $\tau; c_2, \ldots, c_{i-1}$. Then, by differentiating $\mathcal{R}^{(i)}$ holding $\tau; c_2, \ldots, c_{i-1}$, *and hence* $a^{(i)}$, constant, one would have

$$\frac{\partial \mathcal{R}^{(i)}}{\partial \eta} = \sum_{j=1}^{i-1} \frac{\partial r_j}{\partial \eta} (a^{(i)})^{j+1}, \quad (3.28)$$

which contradicts Eq. (3.27) unless

$$\frac{\partial r_j}{\partial \eta} = 0, \quad j = 1, \ldots, i-1. \quad (3.29)$$

Therefore $\mathcal{R}^{(i)}$ cannot depend on the mysterious η. (The proof generalizes trivially to the case where η is presumed to be a discrete parameter.) This shows conclusively that the parametrization of RS dependence by $\tau; c_2, \ldots, c_{i-1}$ is complete for ith-order approximations to physical quantities.

The self-consistency condition, Eq. (3.27), has even stronger consequences. It requires that there be a (partial) cancellation between the RS dependence of $a^{(i)}$ and that of the coefficients r_j, which is enough to fix the functional dependence of the r_j on the RS parameters $\tau; c_2, \ldots, c_j$. This will be demonstrated in the following sections.

IV. OPTIMIZING SECOND-ORDER RESULTS

A. The optimization formula

I now specialize to the case of second-order approximations. According to the previous arguments, there is only one relevant RS parameter, namely $\tau \equiv b \ln(\mu/\tilde{\Lambda})$. This is saying that the RS-dependence problem at second order is entirely equivalent to the problem of how best to choose the renormalization point within one particular kind of RS. I can verify this statement by an independent argument at the end of this subsection, but it may be helpful to discuss this point a little further here.

In e^+e^- annihilation, which I continue to use as an example, it is traditional to fix the renormalization point μ to be equal to the total e^+e^- energy Q, but there is no solid justification for this choice. The argument is only a dimensional argument—the couplant depends on one massive parameter μ, and the only massive variable in the problem is Q. However, the argument provides no reason for setting $\mu = Q$ rather than $\mu = \frac{1}{2}Q$, or $\mu = 4Q$, etc. Moreover, in other processes with more complicated kinematics, this ambiguity is even more evident. What is needed is an objective criterion for choosing μ, and this is what the PMS criterion supplies. I shall show that by resolving this renormalization-point ambiguity one automatically resolves the RS ambiguity in second order.

To begin with I choose some particular kind of RS (say, minimal subtraction[23] for definiteness), and consider the second-order approximant $\mathcal{R}^{(2)}$ as a function of the renormalization point μ. For convenience I use the variable $\tau \equiv b \ln(\mu/\tilde{\Lambda})$ instead of μ itself, and write $\mathcal{R}^{(2)}$ as

$$\mathcal{R}^{(2)}(\tau) = a^{(2)}(\tau)[1 + r_1(\tau)a^{(2)}(\tau)]. \tag{4.1}$$

The renormalization-point ambiguity arises because one has an infinite set of approximants $\{\mathcal{R}^{(2)}(\tau) \mid -\infty < \tau < \infty\}$ which, although formally equivalent, are numerically different. The PMS criterion resolves this ambiguity by selecting a unique member of this set, namely $\mathcal{R}^{(2)}_{\text{opt}} \equiv \mathcal{R}^{(2)}(\bar{\tau})$, where $\bar{\tau}$ is characterized by

$$\left.\frac{\partial \mathcal{R}^{(2)}}{\partial \tau}\right|_{\tau=\bar{\tau}} = 0. \tag{4.2}$$

Differentiating Eq. (4.1) and using

$$\frac{\partial a^{(2)}}{\partial \tau} = \frac{\beta^{(2)}(a^{(2)})}{b} = -(a^{(2)})^2(1 + ca^{(2)}), \tag{4.3}$$

one obtains (for the remainder of this section I write a as shorthand for $a^{(2)}$)

$$\frac{\partial \mathcal{R}^{(2)}}{\partial \tau} = -a^2(1 + ca)(1 + 2r_1 a) + a^2 \frac{\partial r_1}{\partial \tau}. \tag{4.4}$$

The order-a^2 terms must cancel for the formal self-consistency of perturbation theory, as discussed in Sec. IIID. This gives

$$\frac{\partial r_1}{\partial \tau} = 1, \tag{4.5}$$

corresponding to the well-known fact (easily verified by Feynman-diagram calculations) that under a change of renormalization point $\mu \to \mu'$, $r_1 \to r_1' = [r_1 + b \ln(\mu'/\mu)]$. Equation (4.5) shows that

$$\rho_1 \equiv \tau - r_1(\tau) \tag{4.6}$$

is a constant, independent of the unphysical variable τ.

The PMS criterion requires that the remaining terms of Eq. (4.4) should vanish at $\tau = \bar{\tau}$, i.e.,

$$2\bar{r}_1(1 + c\bar{a}) + c = 0, \tag{4.7}$$

where $\bar{r}_1 \equiv r_1(\bar{\tau})$, $\bar{a} \equiv a(\bar{\tau})$. Using this to eliminate $\bar{r}_1$ in $\mathcal{R}^{(2)}_{\text{opt}} \equiv \bar{a}(1 + \bar{r}_1\bar{a})$ yields

$$\mathcal{R}^{(2)}_{\text{opt}} = \bar{a}\frac{(1 + \frac{1}{2}c\bar{a})}{(1 + c\bar{a})}. \tag{4.8}$$

All that remains is to find $\bar{a}$. This is, in fact, given by Eq. (4.7) because r_1 is implicitly a function of a. Recall that r_1 is simply related to τ by Eq. (4.6), and that τ is given as a function of a by integrating the β-function equation, Eq. (4.3):

$$\tau = \hat{K}^{(2)}(a) \equiv b \int_{\infty}^{a} \frac{da'}{\beta^{(2)}(a')}, \tag{4.9}$$

where

$$\hat{K}^{(2)}(a) = -\int_{\infty}^{a} \frac{da'}{a'^2(1 + ca')} = \frac{1}{a} + c \ln\left(\frac{ca}{1 + ca}\right). \tag{4.10}$$

[The lower limit of integration in Eq. (4.9) reflects the particular definition of $\tilde{\Lambda}$ appearing in

$\tau \equiv b \ln(\mu/\tilde{\Lambda})$ (see Appendix A).] Substituting

$$r_1 = \tau - \rho_1 = \hat{K}^{(2)}(a) - \rho_1 \tag{4.11}$$

into Eq. (4.7) gives an implicit equation for $\bar{a}$:

$$\hat{K}^{(2)}(\bar{a}) + \frac{1}{2}\frac{c}{(1+c\bar{a})} = \rho_1 . \tag{4.12}$$

The optimum second-order result is found by solving this transcendental equation for $\bar{a}$ and then substituting the result into Eq. (4.8). [This is really no more difficult than the usual perturbation theory procedure, which also requires inverting a transcendental equation, namely $\tau = \hat{K}^{(2)}(a)$.]

I can now show that the final result for $\mathcal{R}^{(2)}_{\text{opt}}$ is independent of the particular kind of RS adopted initially. Consider the quantity ρ_1: Eq. (4.6) holds for any value of τ, so I can, in particular, evaluate it for $\mu = Q$ to obtain

$$\rho_1 = b \ln(Q/\tilde{\Lambda}) - r_1(\text{calc}) . \tag{4.13}$$

In the e^+e^- example r_1(calc) is the coefficient calculated in Refs. 2 and 32, which use $\mu = Q$. Now, r_1(calc) depends on which kind of RS is used in the calculation, *but so does* $\tilde{\Lambda}$. Moreover, it has been shown by Celmaster and Gonsalves[1] that the RS dependence of $\tilde{\Lambda}$ is determined exactly by a one-loop calculation and is precisely such that the RS dependences of $\tilde{\Lambda}$ and r_1(calc) cancel in Eq. (4.13) (see Appendix A, paragraph 4). *Thus* ρ_1 *is a RS invariant.* The same is true of c, so it is immediately obvious from Eqs. (4.12) and (4.8) that $\bar{a}$ and $\mathcal{R}^{(2)}_{\text{opt}}$ are independent of the RS.

The physics of the particular process under consideration enters the formulas for the optimized result only through ρ_1. So, from this point of view, the goal of a Feynman-diagram calculation is to determine the RS invariant ρ_1: The convention-dependent quantity r_1 is only an intermediate step. Given ρ_1, one can immediately obtain the optimum second-order approximant $\mathcal{R}^{(2)}_{\text{opt}}$ from Eqs. (4.12) and (4.8). The size of ρ_1 is an objective indicator of the reliability of perturbation theory. The larger ρ_1 is, the smaller $\bar{a}$ will be, and hence the more reliable one can expect low-order perturbative approximations to be.

The next subsection is devoted to some remarks aimed at clarifying the meaning of the preceding algebra. Section IV C then summarizes the second-order optimization formula for a general physical quantity of the form $\mathcal{R} = a^N(1 + r_1 a + \cdots)$.

B. Remarks

The following points are noted.

(1). An important step in the last subsection was the change of variables from τ to a. It is instructive, therefore, to look at $\mathcal{R}^{(2)}$ as a function of a:

$$\begin{aligned}\mathcal{R}^{(2)}(a) &= a[1 + r_1(a)a] \\ &= a\{1 + [\hat{K}^{(2)}(a) - \rho_1]a\} \\ &= a\{2 - \rho_1 a + ca \ln[ca/(1+ca)]\} .\end{aligned} \tag{4.14}$$

This makes plain a very important fact: *Perturbative approximations are not polynomials in the couplant.* Figure 5 shows, however, that a graph of $\mathcal{R}^{(2)}$ against a is roughly parabolic (for large ρ_1/c). The curve is fully determined once c and ρ_1 are known, and this is equivalent to knowledge of c and the value of $\mathcal{R}^{(2)}(a_0)$ at some arbitrary initial value of a_0. The PMS criterion selects the stationary point of the curve as giving the most reliable approximate result. The optimization formulas simply perform the necessary analytical geometry to find this result: Eq. (4.12) locates the position of the maximum, and Eq. (4.8) gives the value of the function $\mathcal{R}^{(2)}$ at that point.

(2). In view of the similarities with the examples of Sec. II, it is very likely that in some cases, although not in all, the exact result will actually lie just above the maximum of the curve. Moreover, it is tempting to speculate that one might be able to prove an analog of the Rayleigh-Ritz theorem, i.e., an inequality $\mathcal{R} \geq \mathcal{R}^{(2)}(a)$, all a, in certain simple cases. This would be an interesting point to investigate with functional integration techniques.

(3). Another similarity with the examples of Sec. II is that the *lowest*-order approximation gives a monotonic function, since the first-order approximant is just $\mathcal{R}^{(1)}(a) = a$. Thus leading-order results are irredeemably ambiguous—there is no objective means for choosing the correct renormalization point μ at which to evaluate the couplant. [This is, in fact, the same as the ambiguity pointed out by Baće,[24] although expressed in differ-

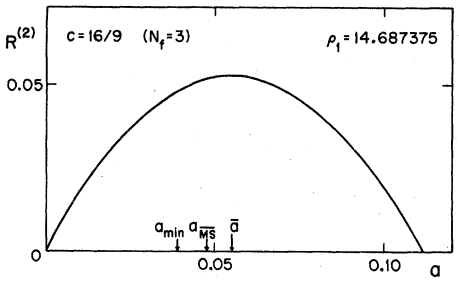

FIG. 5. The second-order approximant $\mathcal{R}^{(2)}$ as a function of the choice of couplant [see Eq. (4.14)]. This example is drawn from e^+e^- annihilation at an energy $Q = 100\tilde{\Lambda}_{\text{min}}$ (see Table II, Sec. VI A): a_{min} and $a_{\overline{\text{MS}}}$ are the couplants in minimal subtraction and modified minimal subtraction, respectively, evaluated at $\mu = Q$. Varying μ in any RS (with the physical variable Q held fixed) corresponds to a variation of a, and traces the curve shown.

ent terms (see also Ref. 25).] Thus, *leading order gives only semiquantitative results*: One has to guess at the best value for the renormalization point, so the result is part calculation, part intuition. This situation is peculiar to leading order. Second- and higher-order approximations can, by the PMS criterion, be given a definite, quantitative meaning, because the calculation itself indicates which value of μ to choose.

(4). The optimization process implicitly determines the optimum value of μ. An explicit expression for $\bar{\mu}$ is not obtainable, but it is instructive to observe that for small $\bar{a}$ ($c\bar{a} \ll 1$), Eqs. (4.12) and (4.9) give

$$\bar{\tau} \simeq \rho_1 - \tfrac{1}{2}c\,, \tag{4.15}$$

and therefore,

$$b\ln(\bar{\mu}/\tilde{\Lambda}) \simeq b\ln(Q/\tilde{\Lambda}) - r_1(\text{calc}) - \tfrac{1}{2}c\,,$$

i.e.,

$$\bar{\mu} \simeq Q\exp\{-[\tfrac{1}{2}c + r_1(\text{calc})]/b\}. \tag{4.16}$$

Thus, as one expects, $\bar{\mu}$ is given by Q times a number which is of order unity—as long as $r_1(\text{calc})$ is not too big. A large result for $r_1(\text{calc})$ is an indication that the original choice of $\mu = Q$ was far too optimistic.

Another pertinent question is what sort of value the second-order coefficient $r_1(\bar{\tau})$ has in the optimum scheme. From Eq. (4.7) one has

$$\bar{r}_1 = -\tfrac{1}{2}c/(1 + c\bar{a})\,. \tag{4.17}$$

Note that $|\bar{r}_1|$ is always less than $\frac{1}{2}c$ so that, viewed in the optimum scheme, the breakdown of perturbation theory at low energies is due to the effective couplant $\bar{a}$ becoming large, and not due to the coefficients getting too large.[26]

(5). Another interesting question is what does the optimum result, $\mathcal{R}^{(2)}_{\text{opt}} \equiv \bar{a}(1 + \bar{r}_1\bar{a})$, look like when reexpanded in terms of the couplant a of the original scheme, i.e.,

$$\mathcal{R}^{(2)}_{\text{opt}} = \mathcal{R}^{(2)}(a) + a^3\Omega^{(2)} + a^4\Phi^{(2)} + \cdots\,. \tag{4.18}$$

The first two terms must, of course, agree with $\mathcal{R}^{(2)}(a) \equiv a(1 + r_1 a)$, but there will be definite, calculable correction terms beginning at order a^3. The coefficients $\Omega^{(2)}, \Phi^{(2)}, \ldots$ could be obtained by a brute-force expansion of the optimization formulas, but there is a simpler method, which I describe below. This will lead to an approximate optimization formula, or "improvement formula."

This exercise is, in a sense, unnecessary: I have already shown how to find $\mathcal{R}^{(2)}_{\text{opt}}$ exactly, so there is no need for an approximate method. However, there are two reasons for discussing this approach further: (i) because the improvement formula is very convenient for QED applications (where it gives results essentially identical to the optimization formula), and (ii) because the exact optimization formula in higher orders is rather complicated, and a step-by-step approach to the optimum may then be more practical; this approach is most easily explained at second order.

The coefficients $\Omega^{(2)}, \Phi^{(2)}, \ldots$ of Eq. (4.18) can most easily be obtained by using the fact that $\mathcal{R}^{(2)}_{\text{opt}}$ is independent of the initial choice of renormalization point, i.e., $\partial\mathcal{R}^{(2)}_{\text{opt}}/\partial\tau = 0$. It is important to be clear about the distinction between this requirement and the PMS criterion. The former is just the condition that $\mathcal{R}^{(2)}_{\text{opt}}$ is a RS invariant (and therefore independent of the arbitrary, initial choice of τ used in the calculation), while the latter decrees that the optimum is a *particular* invariant, namely the value of $\mathcal{R}^{(2)}$ at its stationary point [i.e., $\mathcal{R}^{(2)}_{\text{opt}} = \mathcal{R}^{(2)}(\bar{\tau})$ where $\bar{\tau}$ is characterized by $(\partial\mathcal{R}^{(2)}/\partial\tau)|_{\tau=\bar{\tau}} = 0$]. The invariance condition determines the form of $\Omega^{(2)}, \Phi^{(2)}, \ldots$; the PMS criterion fixes the boundary condition.

This can be seen as follows. Differentiating the RHS of Eq. (4.18) and equating the $a^3, a^4, \ldots$ terms to zero yields

$$\frac{\partial\Omega^{(2)}}{\partial\tau} = 2r_1 + c\,,$$

$$\frac{\partial\Phi^{(2)}}{\partial\tau} = 3\Omega^{(2)} + 2r_1 c\,, \tag{4.19}$$

etc.

The integration of these equations is easy since $\partial r_1/\partial\tau = 1$. The boundary condition is that the correction terms should vanish if $\tau = \bar{\tau}$, i.e.,

$$\Omega^{(2)}(\bar{r}_1) + \bar{a}\Phi^{(2)}(\bar{r}_1) + O(a^2) = 0\,. \tag{4.20}$$

Since the PMS condition, Eq. (4.7), gives

$$\bar{r}_1 = -\tfrac{1}{2}c + O(a)\,, \tag{4.21}$$

one obtains

$$\Omega^{(2)} = (r_1 + \tfrac{1}{2}c)^2\,, \tag{4.22}$$

$$\Phi^{(2)} = (r_1 + \tfrac{1}{2}c)(r_1^2 + 2r_1 c - \tfrac{1}{4}c^2)\,. \tag{4.23}$$

The positive semidefiniteness of $\Omega^{(2)}$ is no surprise, since $\mathcal{R}^{(2)}_{\text{opt}}$ is the largest of all the $\mathcal{R}^{(2)}$'s.

I stress that the improvement formula is much less satisfactory than the previous exact formulation. If one does elect to use this step-by-step approach, one must be sure that one has kept sufficient terms, so that $\mathcal{R}^{(2)}_{\text{opt}}$ is obtained to within the required *numerical* accuracy.

C. How to use the optimization formula

In this subsection, I summarize the procedure for optimizing second-order calculations of physi-

cal quantities, and present the necessary formulas for the general case. The aim is to show that the method is perfectly straightforward to apply in practice.

The results apply to any physical quantity that can be written as $A+B\mathcal{R}$, where $\mathcal{R}$ has the form[27]

$$\mathcal{R}=a^N(1+r_1a+\cdots), \tag{4.24}$$

and where A (the parton-model term, if any) and B (the leading-order coefficient) are independent of the RS. (This must be true for the self-consistency of perturbation theory.) A and B play no role in the optimization process. [N.B. In the moment analysis the quantities $-(Q/M_n)dM_n/dQ$ are of the above form.]

The first stage is to find the value of the appropriate RS invariant ρ_1. Given that the second-order coefficient r_1 in Eq. (4.24) has been calculated in some RS, one can construct ρ_1 from the formula

$$\rho_1\equiv\tau_0-r_1(\tau_0)/N, \tag{4.25}$$

with

$$\tau_0\equiv b\ln(\mu_0/\tilde{\Lambda}_0), \tag{4.26}$$

where μ_0 is the renormalization point used in the calculation, and $\tilde{\Lambda}_0$ is the QCD scale parameter appropriate to the RS employed. [$\tilde{\Lambda}$'s in different RS's are related exactly by a one-loop calculation,[1] so $\tilde{\Lambda}_0$ can always be expressed in terms of $\tilde{\Lambda}_{\rm ref}$ of some "reference" scheme. Here I pretend that the value of $\tilde{\Lambda}_{\rm ref}$ (and hence of $\tilde{\Lambda}_0$) is known. For a fuller discussion see Appendix A, paragraphs 3–5.] Although the calculated coefficient r_1 depends on the RS and renormalization point adopted in the calculation, the resulting ρ_1 is entirely independent of these arbitrary choices. In general ρ_1 will be a function of several physical variables which describe the kinematics, etc., of the process under consideration. Perturbation theory is reliable only if ρ_1 is large (i.e., $\rho_1\gg1$, or, more correctly, $\rho_1\gg|c|$).

Secondly, one must solve the equation

$$S_N(y)\equiv\ln y+\frac{1}{y}(1-y)\left[1+\frac{y}{(N+1)}\right]=\frac{\rho_1}{c} \tag{4.27}$$

for y. [To compare this with Sec. IV A, note that $y\equiv c\bar{a}/(1+c\bar{a})$, where $\bar{a}$ is the couplant in the optimum RS.] Finally one obtains the optimized second-order result from

$$\mathcal{R}^{(2)}_{\rm opt}=\frac{1}{c^N}\frac{y^N}{(1-y)^N}\left[1-\frac{N}{(N+1)}y\right]. \tag{4.28}$$

Equation (4.27) is readily solved by standard numerical methods, e.g., Newton's method converges nicely, for a small enough initial value of y. A simple technique for obtaining rough answers is to plot both $S_N(y)$ and $\mathcal{R}^{(2)}_{\rm opt}$ as functions of y on the same graph. For any input value of ρ_1/c one can then immediately read off the value of $\mathcal{R}^{(2)}_{\rm opt}$. This is illustrated in Fig. 6 for the $N=1$ case.

An alternative approach—not recommended for QCD applications, but convenient in QED—is to approach the optimum result stepwise from the naive result. That is, one first calculates the usual approximant

$$\mathcal{R}^{(2)}=a^N(1+r_1a) \tag{4.29}$$

in some RS. (This, too, requires solving a transcendental equation to express a, which is shorthand for $a^{(2)}$ here, as a function of $\tilde{\Lambda}$.) One can then evaluate the correction terms in the improvement formula

$$\mathcal{R}^{(2)}_{\rm opt}=\mathcal{R}^{(2)}+\Omega^{(2)}a^{N+2}+\Phi^{(2)}a^{N+3}+O(a^{N+4}). \tag{4.30}$$

The general formulas for the coefficients $\Omega^{(2)}$, $\Phi^{(2)}$ are

$$\Omega^{(2)}=\frac{(N+1)}{2N}(r_1+\tilde{c})^2, \tag{4.31}$$

$$\Phi^{(2)}=\frac{(N+1)}{6N^2}(r_1+\tilde{c})[(N+2)r_1^2+(5N+7)r_1\tilde{c}-(2N+1)\tilde{c}^2], \tag{4.32}$$

where $\tilde{c}\equiv cN/(N+1)$. Expressions for higher-order coefficients in the improvement formula can also be obtained reasonably easily, if required.

This stepwise approach works well if the couplant is small and if the initial RS is close to the optimum scheme. It is more convenient in QED and will usually give essentially the same result as the optimization formula. Moreover, if the cal-

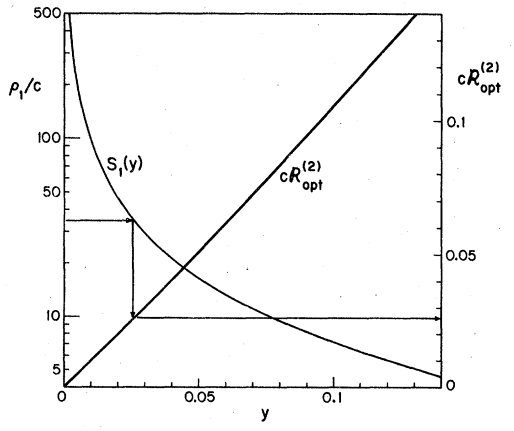

FIG. 6. A graphical solution to the second-order optimization formula for the $N=1$ case. The function $S_1(y)$ of Eq. (4.27) is plotted with respect to the logarithmic scale on the left-hand side. A graph of $c\mathcal{R}^{(2)}_{\rm opt}$, Eq. (4.28) referred to the right-hand scale, is superimposed. The arrows illustrate how, knowing ρ_1/c, one can obtain $c\mathcal{R}^{(2)}_{\rm opt}$.

culation of r_1 was performed in the canonical QED renormalization scheme, then the appropriate value of the couplant is just the fine-structure constant divided by π, so one avoids transcendental equations altogether. A QED example is discussed in Sec. VI B.

V. OPTIMIZATION IN HIGHER ORDERS

A. The optimization procedure in third order

Given the present situation with regard to QCD calculations, it might seem rather academic to consider the optimization procedure beyond second order. However, the future comes sooner than we imagine. Moreover, it is important theoretically to know how to deal with the RS-dependence problem in arbitrary order—if only to have a better understanding of second order, which is a rather special case. In this section, I therefore describe the more typical case of third order in detail, and sketch the generalization to higher orders.

I consider third-order results of the general form

$$\mathcal{R}^{(3)} = a^N(1 + r_1 a + r_2 a^2). \tag{5.1}$$

The notation here is very condensed: a is to be understood as $a^{(3)}(\tau; c_2)$, the third-order approximation to the full couplant, i.e., $a^{(3)}(\tau, c_2)$ is the solution to

$$\frac{\partial a^{(3)}}{\partial \tau} = \hat{\beta}^{(3)}(a^{(3)}) = -(a^{(3)})^2[1 + c a^{(3)} + c_2 (a^{(3)})^2]. \tag{5.2}$$

It also satisfies

$$\frac{\partial a^{(3)}}{\partial c_2} = \beta_2^{(3)}(a^{(3)}), \tag{5.3}$$

where $\beta_2^{(3)}(a)$ is the third-order approximation to the β_2 function defined in Eq. (3.21). It is therefore given by

$$\beta_2^{(3)}(a) = a^2(1 + ca + c_2 a^2) \int_0^a \frac{dx}{(1 + cx + c_2 x^2)^2}. \tag{5.4}$$

Since I am imagining here that the third-order calculations for c_2 and r_2 have been done (in some arbitrary RS—the same for both, of course), the functions $\hat{\beta}^{(3)}$ and $\beta_2^{(3)}$ are fully known.

Differentiating $\mathcal{R}^{(3)}$ with respect to τ and c_2 leads to

$$\frac{\partial \mathcal{R}^{(3)}}{\partial \tau} = \hat{\beta}^{(3)}(a)[N + (N+1) r_1 a + (N+2) r_2 a^2] a^{N-1} + a^{N+1}\{N + [Nc + (N+1) r_1] a\}, \tag{5.5}$$

$$\frac{\partial \mathcal{R}^{(3)}}{\partial c_2} = \beta_2^{(3)}(a)[N + (N+1) r_1 a + (N+2) r_2 a^2] a^{N-1} - N a^{N+2}, \tag{5.6}$$

where I have used the self-consistency equations

$$\frac{\partial r_1}{\partial \tau} = N, \quad \frac{\partial r_2}{\partial \tau} = Nc + (N+1) r_1, \quad \frac{\partial r_1}{\partial c_2} = 0, \quad \frac{\partial r_2}{\partial c_2} = -N, \tag{5.7}$$

which are required for the a^2 and a^3 terms to cancel. These equations are equivalent to the statement that the quantities

$$\rho_1 \equiv \tau - r_1/N, \qquad \rho_2 \equiv r_2 + N c_2 - \frac{(N+1)}{2N}\left(r_1 + \frac{N}{N+1} c\right)^2 \tag{5.8}$$

are RS invariants. Just as r_1 and r_2, ρ_1 and ρ_2 are known quantities which embody the information obtained from the diagram calculations but, unlike the coefficients, ρ_1 and ρ_2 are independent of the RS, and therefore should be regarded as more significant.

As in the second-order case, it proves to be convenient to change variables from τ to a using

$$\tau = \hat{K}^{(3)}(a; c_2) = \hat{K}^{(2)}(a) + c_2 \int_0^a \frac{dx}{(1 + cx)(1 + cx + c_2 x^2)}, \tag{5.9}$$

where

$$\hat{K}^{(2)}(a) = \frac{1}{a} + c \ln\left(\frac{ca}{1 + ca}\right), \tag{5.10}$$

as before. Equation (5.9) is obtained by integrating the β-function equation, Eq. (5.2), with the boundary condition (i.e., the definition of $\bar{\Lambda}$) described in Appendix A.

By substituting into Eq. (5.8) I can now express r_1, r_2 in terms of the RS variables a, c_2:

$$r_1 = N[\hat{K}^{(3)}(a; c_2) - \rho_1], \qquad r_2 = \rho_2 - N c_2 + \frac{N(N+1)}{2}\left[\hat{K}^{(3)}(a; c_2) - \rho_1 + \frac{c}{N+1}\right]^2. \tag{5.11}$$

Thus, given the values of the "physics parameters" N, ρ_1, ρ_2, one knows $\mathcal{R}^{(3)}$ completely as a function of the unphysical variables a, c_2:

$$\mathcal{R}^{(3)}(a, c_2) = a^N[1 + r_1(a, c_2) a + r_2(a, c_2) a^2]. \tag{5.12}$$

Obviously this is quite a complicated function. Its behavior in the vicinity of the optimum point $\bar{a}, \bar{c}_2$ is illustrated in Fig. 7: Basically it seems that, for $N = 1$, $\mathcal{R}^{(3)}$ is quasilinear as a function of c_2, and quasicubic as a function of a, as one might expect from the most naive considerations.

The optimum values of the unphysical variables correspond to the stationary point of $\mathcal{R}^{(3)}(a; c_2)$, according to the PMS criterion. The optimization

condition is therefore $\partial R^{(3)}/\partial(a, c_2)\big|_{a=\bar{a},\ c_2=\bar{c}_2} = 0$. These equations, which are just appropriate linear combinations of Eqs. (5.5) and (5.6) equated to zero, can be written as

$$[N\bar{c}_2 + (N+1)\bar{r}_1 c + (N+2)\bar{r}_2] + \bar{a}[(N+1)\bar{r}_1\bar{c}_2 + (N+2)\bar{r}_2 c] + \bar{a}^2(N+2)\bar{r}_2\bar{c}_2 = 0\,, \tag{5.13}$$

$$\int_0^{\bar{a}} \frac{dx}{(1+cx+\bar{c}_2x^2)^2} = \frac{N\bar{a}}{\{N+[Nc+(N+1)\bar{r}_1]\bar{a}\}}\,. \tag{5.14}$$

Since $\bar{r}_1, \bar{r}_2$ are known as functions of $\bar{a}, \bar{c}_2$ by Eqs. (5.11), one can, in principle, solve Eqs. (5.13) and (5.14) for $\bar{a}, \bar{c}_2$. One can then substitute back for $\bar{r}_1, \bar{r}_2$, and thereby determine the optimum result

$$R^{(3)}_{\rm opt} = \bar{a}^N(1+\bar{r}_1\bar{a}+\bar{r}_2\bar{a}^2)\,. \tag{5.15}$$

Obviously this procedure is quite horrendous analytically, but it may be quite straightforward to implement on a computer. The result is clearly RS independent, since it is completely determined by the RS invariants N, c, ρ_1, and ρ_2.

B. The improvement formula in third order

Since the exact optimization procedure is rather complicated, it may be useful to derive an improvement formula for the third-order case. In this approximate approach one writes the optimum result as the naive result in the original RS plus correction terms, i.e.,

$$R^{(3)}_{\rm opt} = R^{(3)} + \Omega^{(3)}a^{N+3} + O(a^{N+4})\,. \tag{5.16}$$

As in the second-order case, one can determine $\Omega^{(3)}$ from the requirement that $R^{(3)}_{\rm opt}$ is independent of the original choice of RS. One needs the power-series expansion of the $\beta_2^{(3)}(a)$ function, which can be obtained from Eq. (5.4), or from Eq. (A8) of Appendix A with $i=2$ and $c_3 = c_4 = \ldots = 0$. The leading terms are

$$\frac{\partial a}{\partial c_2} = \beta_2^{(3)}(a) = a^3 + O(a^5)\,. \tag{5.17}$$

By equating the order a^{N+3} terms in $\partial R^{(3)}_{\rm opt}/\partial(\tau; c_2) = 0$ and using Eqs. (5.5) and (5.6), one obtains

$$\frac{\partial\Omega^{(3)}}{\partial\tau} = [Nc_2 + (N+1)r_1 c + (N+2)r_2]\,, \qquad \frac{\partial\Omega^{(3)}}{\partial c_2} = -(N+1)r_1\,. \tag{5.18}$$

It is convenient here to change to r_1, r_2 as the RS variables by making use of Eqs. (5.7) and (5.8). This leads to

$$\frac{\partial\Omega^{(3)}}{\partial r_2} = \frac{(N+1)}{N}r_1\,, \qquad \frac{\partial\Omega^{(3)}}{\partial r_1} = \frac{(N+1)r_2}{N} - \frac{(N+1)(2N+1){r_1}^2}{2N^2} + \frac{cr_1}{N} + \frac{c^2}{2(N+1)} + \frac{\rho_2}{N}\,. \tag{5.19}$$

The boundary condition to be imposed is that $\Omega^{(3)} = 0$ at $r_1 = \bar{r}_1$, $r_2 = \bar{r}_2$, where the optimum point $\bar{r}_1$, $\bar{r}_2$ is characterized by the PMS condition. It is easy to see that, up to $O(a)$ corrections, this is equivalent to the vanishing of Eqs. (5.18), i.e.,

$$\bar{r}_1 + O(\bar{a}) = 0\,, \qquad N\bar{c}_2 + (N+2)\bar{r}_2 + O(\bar{a}) = 0\,. \tag{5.20}$$

[The vanishing of $\bar{r}_1$ in the small $\bar{a}$ limit is a consequence of the absence of an order a^4 term in $\beta_2(a)$. Notice that $\bar{r}_1$ here is not the same as $\bar{r}_1$ at second order. There is no inconsistency in this: As with the examples in Sec. II, the optimum point is different in different orders of approximation.]

Integration of Eqs. (5.19) yields

$$\Omega^{(3)} = r_1\left[\frac{(N+1)r_2}{N} - \frac{(N+1)(2N+1){r_1}^2}{6N^2} + \frac{cr_1}{2N} + \frac{c^2}{2(N+1)} + \frac{\rho_2}{N}\right]. \tag{5.21}$$

Finally, by resubstituting for ρ_2, one can write the result purely in terms of the directly calculated

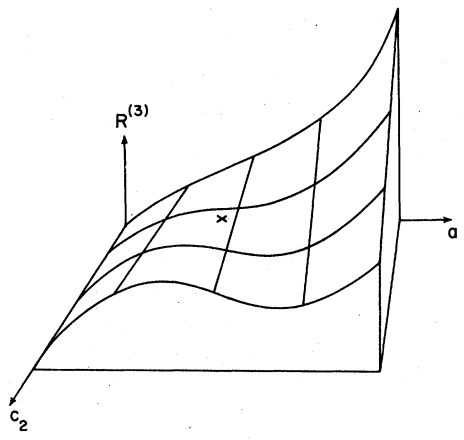

FIG. 7. An illustration of the behavior of the third-order approximant $R^{(3)}$, for $N=1$, as a function of the RS parameters a and c_2 in the vicinity of the optimum point (indicated by a cross) [See Eq. (5.12)]. The sketch is based on a numerical study using the same example as in Fig. 5 and Table II, assuming a moderate, positive value for ρ_2. Changing ρ_2 is qualitatively like a translation with respect to the c_2 axis.

quantities r_1, r_2, c_2:

$$\Omega^{(3)} = r_1\left[c_2 + \frac{(N+2)r_2}{N} - \frac{(N+1)(N+2)r_1^2}{3N^2} - \frac{cr_1}{2N}\right]. \tag{5.22}$$

Further coefficients in the improvement formula can also be obtained, if required, by straightforward, if tedious, calculation.

C. RS invariants and optimization in higher orders

In this subsection, I will briefly sketch how the optimization procedure works in higher orders. In $(k+1)$th order one will need the coefficients $r_1, \ldots, r_k$ in

$$\mathcal{R}^{(k+1)} = a^N(1 + r_1 a + \cdots + r_k a^k), \tag{5.23}$$

and the β-function coefficients $c, \ldots, c_k$ in

$$\frac{\partial a}{\partial \tau} = \hat{\beta}^{(k+1)}(a) = -a^2(1 + ca + c_2 a^2 + \cdots + c_k a^k), \tag{5.24}$$

where a is shorthand for $a^{(k+1)}$ here. All these coefficients must be calculated in the same RS, but it is immaterial which scheme is used.

The approximant $\mathcal{R}^{(k+1)}$ depends on RS through $(\tau; c_2, \ldots, c_k)$. From the self-consistency condition that $\partial\mathcal{R}^{(k+1)}/\partial(\tau; c_2, \ldots, c_k) = O(a^{N+k+1})$, one can determine the RS dependence of all the coefficients:

$$\frac{\partial r_l}{\partial \tau} = \sum_{m=0}^{l-1} (N+m) r_m c_{l-m-1}, \tag{5.25}$$

$$\frac{\partial r_l}{\partial c_j} = \frac{-1}{(j-1)} \sum_{m=0}^{l-j} (N+m) r_m W^j_{l-j-m}, \quad l \geq j$$

$$= 0, \quad l < j$$

where $c_0 \equiv r_0 \equiv W^j_0 \equiv 1$, and $c_1 \equiv c$. The W^j_n coefficients appear in the expansions of the $\beta_j(a)$ functions, and are given in Appendix A, Eq. (A8).

By integrating these differential equations, one can produce a set of RS invariants $\rho_1, \rho_2, \ldots, \rho_k$. In fact, only ρ_k will be new, in the sense that the others are known from lower orders. I have already given expressions for ρ_1 and ρ_2 in Eq. (5.8). (The first invariant, ρ_1, is rather special and needs to be large if perturbation theory is to work well.) The normalization, etc., of the higher-order invariants can be chosen such that

$$\rho_k \equiv r_k + \frac{N}{(k-1)} c_k - \Omega^{(k)}, \quad k \geq 2 \tag{5.26}$$

where $\Omega^{(k)}$ is the coefficient of the correction term in the previous (i.e., kth) order's improvement formula, and depends on r_i, c_i, $i = 1, \ldots, k-1$. It can readily be shown that the right-hand side of Eq. (5.26) is RS independent, i.e., satisfies $\partial\rho_k/\partial(\tau; c_2, \ldots, c_k) = 0$ [e.g., the reader can explicitly check this statement for ρ_3, using the expression for $\Omega^{(3)}$ given in Eq. (5.22)]. These RS invariants summarize the results of the diagram calculations in a complete, RS-independent form, and they naturally play an important role in "optimized perturbation theory."

Once the values of $\rho_1, \ldots, \rho_k$ have been calculated, one knows the exact form of $\mathcal{R}^{(k+1)}$ as a function of the RS variables $a; c_2, \ldots, c_k$. (Again it is convenient to swap τ for $a \equiv a^{(k+1)}$ by making use of the integrated β-function equation.) Applying the PMS criterion, one seeks the stationary point of this function. The condition for this, $\partial\mathcal{R}^{(k+1)}/\partial(a; c_2, \ldots, c_k)\big|_{\text{opt pt}} = 0$, yields the equations

$$\sum_{l=0}^{k} a^l \sum_{m=l}^{k} (N+m) r_m c_{k+l-m} = 0, \tag{5.27}$$

$$\int_0^a \frac{x^{j+2}\,dx}{[\hat{\beta}^{(k+1)}(x)]^2} = \frac{a^{j-1}}{(j-1)} \frac{\left[\sum_{l=0}^{k-j} a^l \sum_{m=0}^{l} (N+m) r_m W^j_{l-m}\right]}{\left[\sum_{l=0}^{k-1} a^l \sum_{m=0}^{l} (N+m) r_m c_{l-m}\right]},$$

$$(j = 2, \ldots, k).$$

These equations generalize Eqs. (5.13) and (5.14). It is to be understood that there should be an overbar above $r_1, \ldots, r_k$, $c_2, \ldots, c_k$, and a wherever they appear (explicitly or implicitly), since these are the equations for the optimum point $\bar{a}; \bar{c}_2, \ldots, \bar{c}_k$. Once these equations have been solved, one can immediately construct the optimum result $\mathcal{R}^{(k+1)}_{\text{opt}} \equiv \bar{a}^N(1 + \bar{r}_1\bar{a} + \cdots + \bar{r}_k\bar{a}^k)$.

D. Remarks

The following points are noted.

(1) An important question is whether the optimization procedure gives a unique result in every order. The following argument indicates that this is in fact the case. Consider the improvement formula for $\mathcal{R}^{(k+1)}_{\text{opt}}$, i.e., the expansion of $\mathcal{R}^{(k+1)}_{\text{opt}}$ in powers of the original couplant a. The coefficients $\Omega^{(k+1)}$, $\Phi^{(k+1)}$, ... of this series can, in principle, be calculated by a generalization of the method described in Sec. V B. By its very nature, this procedure clearly generates *unique* coefficients. This does not mean that Eqs. (5.27) have only one solution, but it does suggest that there is only one solution in the region of interest. [For instance, in second order, $\mathcal{R}^{(2)}(a)$ has a second stationary point at negative a, which is of no physical interest. It is close to, and an inevitable consequence of, the logarithmic singularity at $a = -1/c$.[28] In third order when c_2 is negative, $\mathcal{R}^{(3)}(a; c_2)$ acquires a logarithmic singularity at large a, and in the perturbatively inaccessible region beyond this wall of

singularities there may exist other, spurious, stationary points. The singularities of $\mathcal{R}^{(k+1)}$ correspond to zeros of the approximate β function $\beta^{(k+1)}(a)$, which obviously mark an outer boundary of the perturbative domain.]

I conclude that, for all practical purposes, the optimization procedure does lead to a unique result. I might add that nonuniqueness is not fatal to the philosophy of the PMS criterion: It occurs in the higher orders of the approximation schemes of Sec. II. The result is then a narrow range of values instead of a single value, the width of the range being a measure of the intrinsic error of the approximation. The special feature of the field-theory case is that each order introduces an extra unphysical parameter, which breaks the degeneracy that would otherwise occur, leading to a unique result.

(2) Secondly, I point out that the optimized result is independent of the specific choice of $v_i = c_i$ for the RS parameters, made in Sec. III B. Any consistent scheme for labeling RS's would lead to the same result. This is almost obvious: The value of $\mathcal{R}^{(k+1)}$ at its stationary point is invariant under changes of variables. [Formally one can prove this by showing that the PMS conditions $\partial\mathcal{R}^{(k+1)}/\partial v_i = 0$ $(i = 1, 2, 3, \ldots,$ with $v_1 \equiv \tau)$ are equivalent to the k equations $\partial\mathcal{R}^{(k+1)}/\partial(\tau; c_2, \ldots, c_k) = 0$ used above. The proof involves the observation that $\mathcal{R}^{(k+1)}$ is manifestly independent of c_{k+1}, $c_{k+2}, \ldots$ at fixed $\tau; c_2, \ldots, c_k$ (cf. Sec. III D), and the use of Eqs. (3.13)–(3.15) which are necessary for the labeling scheme to be consistent.] In this connection I remark that the RS with $c_2 = c_3 = \ldots = 0$ discussed, in another context, by 't Hooft[29] (if it exists[30]) does not have a special status in optimized perturbation theory.

(3) The effect of optimization on the high-order behavior of perturbation theory is an interesting theoretical question. It is usually believed that perturbation theory gives only an asymptotic series which ultimately diverges.[31] However, this paper casts some doubt on whether it is useful to regard perturbation theory as an expansion in powers of a fixed parameter: Certainly, written in terms of RS invariants, perturbative approximations are not polynomials in the couplant. Moreover, in optimized perturbation theory one not only calculates coefficients, but also readjusts the RS at each order in response to the results. Thus it is conceivable that perturbation theory might diverge in almost all RS's, yet converge in the optimum RS, towards which the optimization procedure guides one. This is sheer speculation, but it does emphasize that the ultimate behavior of perturbation theory is very much an open question.

VI. EXAMPLES AND SPECIAL CASES

A. Example: R in e^+e^- annihilation

For illustrative purposes, I give an example of the optimization process at work in second order. The example considered (not intended to be particularly realistic) is the QCD correction to the e^+e^- ratio R in a world with three massless quarks at an energy $Q = 100\,\tilde{\Lambda}_{\min}$, where $\tilde{\Lambda}_{\min}$ is the value of $\tilde{\Lambda}$ in the "minimal subtraction" (min) scheme.

Table II lists the approximate couplants $a^{(2)}$, the second-order coefficients r_1,[2,32] and the naive results $\mathcal{R}^{(2)}$, in two different RS's, min and $\overline{\text{MS}}$ (modified minimal subtraction). These quantities are all evaluated with the conventional choice $\mu = Q$.

The optimum result is computed at the bottom of the table, using Eqs. (4.25)–(4.28) of Sec. IV C. I used min scheme quantities to compute ρ_1, but the reader can easily check that ρ_1 is RS invariant, because $\tilde{\Lambda}_{\overline{\text{MS}}}$ and $\tilde{\Lambda}_{\min}$ are related by[1]

$$\ln(\tilde{\Lambda}_{\overline{\text{MS}}}/\tilde{\Lambda}_{\min}) = \tfrac{1}{2}(\ln 4\pi - \gamma_E) = 0.976\,904\,292\,, \quad (6.1)$$

which precisely compensates for the difference between $r_{1\min}$ and $r_{1\overline{\text{MS}}}$.

These results are illustrated in Fig. 5. It is important to remember that $a_{\min}$ and $a_{\overline{\text{MS}}}$ are the couplants *evaluated at* $\mu = Q$ in the min and $\overline{\text{MS}}$ schemes. If μ is allowed to vary in any scheme, then one traces out the whole of the curve shown in Fig. 5. In this example $\overline{\text{MS}}$ gives a better result than min, but this is not necessarily always the case: The relation $a_{\min} < a_{\overline{\text{MS}}}$ stays fixed, but the position of the maximum of $\mathcal{R}^{(2)}(a)$ with respect to them depends on the process.

This example shows perturbation theory in a fairly favorable light. At lower values of Q, and

TABLE II. A numerical example of the optimization process. The coefficients $r_{1\,\min}$ and $r_{1\,\overline{\text{MS}}}$ are taken from Refs. 2 and 32.

$$R = \left(3\sum_i q_i^2\right)(1 + \mathcal{R})$$

Example: $N_f = 3$, $\therefore b = \frac{9}{2}$, $c = \frac{16}{9}$; $Q/\tilde{\Lambda}_{\min} = 100$

min scheme: ($\mu = Q$)	$a^{(2)}_{\min} = 0.039\,088$ $r_{1\,\min} = 6.035\,890\,514$	$\mathcal{R}^{(2)}_{\min} = 0.048\,3$
$\overline{\text{MS}}$ scheme: ($\mu = Q$)	$a^{(2)}_{\overline{\text{MS}}} = 0.047\,963$ $r_{1\,\overline{\text{MS}}} = 1.639\,821\,202$	$\mathcal{R}^{(2)}_{\overline{\text{MS}}} = 0.051\,74$
Optimized result:		
$\rho_1 = b\ln(Q/\tilde{\Lambda}_{\min}) - r_{1\,\min} = 14.687\,375$		$\mathcal{R}^{(2)}_{\text{opt}} = 0.052\,565$
$y = 0.089\,092\,6$		
$(\therefore \bar{a} = 0.055\,016)$		

in other processes, the variations between results in different schemes can be even more dramatic. Notice that $\bar{a}$ and $\mathcal{R}^{(2)}_{\rm opt}$ are quite similar, indicating that optimized perturbation theory apparently converges quite well. One can perhaps estimate the intrinsic error in the second-order approximation as $\sim\bar{a}\,|\mathcal{R}^{(2)}_{\rm opt}-\bar{a}\,|$.

Finally, I can illustrate the workings of the improvement formula, Eqs. (4.30)–(4.32). Starting from the min result and successively adding the correction terms $\Omega^{(2)}a^3$ (= 0.0029) and $\Phi^{(2)}a^4$ (= 0.0009), with $a=a^{(2)}_{\rm min}$ here, gives

$$\mathcal{R}^{(2)}_{\rm min}=0.0483-0.0512-0.0521\,, \tag{6.2}$$

to be compared with the optimum result 0.052 565. The improvement formula works better starting from $\overline{\rm MS}$, since this scheme is closer to the optimum. The first and second correction terms are then 0.000 71 and 0.000 10, giving

$$\mathcal{R}^{(2)}_{\overline{\rm MS}}=0.051\,74-0.052\,44-0.052\,54\,. \tag{6.3}$$

B. Application to anomalous magnetic moments in QED

The anomalous magnetic moments of the electron and muon have been calculated to third order in QED and, for the muon, estimates of the fourth-order coefficient exist (see Ref. 33 and references therein). This provides a chance to test the optimization formulas by comparing the nth-order result, before and after optimization, with the known $(n+1)$th-order result. (This is a purely theoretical exercise: I am not concerned here with non-QED contributions, or with the comparison with experiment.) A slight cheat is involved in applying the previous results here: A proper treatment would require an extension of the present work to massive theories. However, I will ignore this deficiency here. This may not be too unreasonable since $(g-2)$ is finite as $m_e, m_\mu \to 0$, provided that the ratio m_e/m_μ is held fixed.

In QED there is a natural choice of RS in which the renormalization is performed on-shell and the couplant is α_{fs}/π, with α_{fs} the fine-structure constant. (For the purposes of this exercise I take α_{fs} as exactly 1/137.035 987.) The existing calculations are all based on this canonical RS. Table III presents a comparison of the normal perturbative results, taken from the review of Calmet *et al.*,[33] with the optimized results. Since α_{fs} is so small, it makes little difference whether one uses the exact or the approximate formula. I use the latter here for simplicity, quoting the correction term $\Omega^{(2)}a^3$ of Eqs. (4.30) and (4.31), and adding this to $\mathcal{R}^{(2)}$ to give $\mathcal{R}^{(2)}_{\rm opt}$. By comparing these second-order results with the third-order result $\mathcal{R}^{(3)}$, one can see that, in both electron and muon cases, the optimization does improve upon the canonical result, albeit marginally.

In the muon case the third-order result can itself be optimized and compared with the estimate of the fourth-order term. The correction term given by the improvement formula, Eqs. (5.16) and (5.22), is surprisingly large, = 214.7 a^4. However, this is very much in line with the best estimates of the actual fourth-order term. The reason for this success is not hard to find: A large part of the fourth-order coefficient is predictable from the third-order result by renormalization-group arguments,[34] and the optimization formula uses the same kind of information, although in a different way. In my approach the large value of r_2, caused by $\ln(m_e/m_\mu)$ terms, indicates that the optimum RS has a couplant $\bar{a}$, rather larger than α_{fs}/π, resulting in a larger value of $(g-2)$. There is an interesting parallel here with Ref. 35, which also argues that the effective coupling constant is larger than α_{fs} for the muon anomaly.

My conclusion from this exercise is that the optimization procedure does seem to work in the way it should. The improvement gained is quite small, but this is hardly surprising: There is every reason to expect that the canonical RS is a good scheme, well suited to the calculation being done, and hence close to the optimum RS in each case. This latter statement is rather less true in the muon case than in the electron case, as one might

TABLE III. The electron and muon anomalous magnetic moments in QED as a test of the optimization formulas. The numbers are taken from the review of Calmet *et al.*, Ref. 33.

$\mathcal{R}=(g-2)$

$c=\frac{3}{4}$, $c_2=-\frac{121}{96}$ (in canonical RS)

$a=\alpha_{fs}/\pi=2.322\,819\,67\times10^{-3}$

2nd order (electron)	
$r_1=-0.656\,956\,890$	$\mathcal{R}^{(2)}=2.319\,275\times10^{-3}$
correction term	
$a^3(0.0795)$	$\mathcal{R}^{(2)}_{\rm opt}=2.319\,276\times10^{-3}$
3rd order (electron)	
$r_2=2.376\pm0.05$	$\mathcal{R}^{(3)}=2.319\,304\,8\times10^{-3}$
2nd order (muon)	
$r_1=1.531\,564\,460$	$\mathcal{R}^{(2)}=2.331\,083\times10^{-3}$
correction term	
$a^3(3.6350)$	$\mathcal{R}^{(2)}_{\rm opt}=2.331\,129\times10^{-3}$
3rd order (muon)	
$r_2=48.90\pm0.1$	$\mathcal{R}^{(3)}=2.331\,696\,2\times10^{-3}$
correction term	
$a^4(214.7)$	$\mathcal{R}^{(3)}_{\rm opt}=2.331\,702\,5\times10^{-3}$
4th order (muon)	
$r_3\simeq257\pm126$	$\mathcal{R}^{(4)}=2.331\,703\,7(37)\times10^{-3}$

expect. A point to bear in mind is that even if one had used some noncanonical RS, such as minimal subtraction,[36]—where the naive results are poorer—one would still, via the optimization procedure, arrive at the same, excellent, optimized results.[37]

C. The case $N=-1$

The situation in which the physical quantity $\mathcal{R}$ behaves like a^{-1} in leading order has very special features. This case is of physical interest since it occurs for moments of the photon structure functions.[38,39] For $N=-1$ the coefficients $\Omega^{(2)}, \Phi^{(2)}$ of the improvement formula, Eq. (4.30), become divergent. The individual equations, (4.27) and (4.28), of the optimization formula also become singular, but together they yield

$$\mathcal{R}^{(2)}_{\rm opt}(N=-1)=\rho_1 . \tag{6.4}$$

This can be written as

$$\mathcal{R}^{(2)}_{\rm opt}(N=-1)=b\ln(Q/\tilde{\Lambda})+r_1(\mu=Q) , \tag{6.5}$$

which resembles the naive result $\mathcal{R}^{(2)\prime}_{-1}=a^{-1}(Q)+r_1(\mu=Q)$, except that $a(Q)$has its *leading*-order, not its second-order, form. This implies a purely logarithmic Q dependence, as in leading order.

However, this result must be treated with caution, because $\mathcal{R}^{(2)}_{-1}$ is a monotonic function,

$$\mathcal{R}^{(2)\prime}_{-1}(a)=\rho_1-c\ln[ca/(1+ca)] \tag{6.6}$$

(all derivatives are monotonic too), and the result quoted above corresponds to $\bar{a}\to\infty$, $\bar{r}_1=\rho_1$, which is clearly rather disturbing. The case $N=-1$ seems to lag one order behind the general case in that the approximant remains monotonic. It may be that the result in Eq. (6.4) is actually a lower bound on the exact result and it would be interesting to see if this could be proved.

There is no similar difficulty in third order. Remarkably, the optimization equations (5.13) and (5.14) for $N=-1$ can be solved explicitly. They become

$$\int_0^{\bar{a}}\frac{dx}{(1+cx+\bar{c}_2x^2)^2}=\frac{\bar{a}}{1+c\bar{a}} , \tag{6.7}$$

$$-\bar{c}_2+\bar{r}_2+\bar{a}\,\bar{r}_2c+\bar{a}^2\bar{r}_2\bar{c}_2=0 . \tag{6.8}$$

The solution to Eq.(6.7) is clearly $\bar{c}_2=0$, and therefore Eq. (6.8) gives $\bar{r}_2=0$. One now uses the definitions of the RS invariants, except that the definition of ρ_2 in Eq. (5.8) is obviously unsuitable for the case $N=-1$. Removing the singular, and irrelevant, c^2 term, I define

$$\rho_2'=r_2-c_2-cr_1 , \tag{6.9}$$

which is RS invariant and obtainable from calculations of r_1, r_2, c_2 in any RS. In the optimum scheme one has $\bar{r}_1=-\rho_2'/c$, so that $\bar{\tau}=\rho_1+\rho_2'/c$. Hence, $\bar{a}$ is given by

$$\hat{K}^{(2)}(\bar{a})=\rho_1+\rho_2'/c , \tag{6.10}$$

where I have used the fact that, since $\bar{c}_2=0$, $\hat{K}^{(3)}(\bar{a},\bar{c}_2)$ reduces to $\hat{K}^{(2)}(\bar{a})$. Combining all this, one finds

$$\begin{aligned}\mathcal{R}^{(3)}_{\rm opt}&\equiv\bar{a}^{-1}(1+\bar{r}_1\bar{a}+\bar{r}_2\bar{a}^2)\\&=\bar{a}^{-1}-\rho_2'/c ,\end{aligned} \tag{6.11}$$

where $\bar{a}$ is the solution to Eq. (6.10).

Notice that $\mathcal{R}^{(3)}_{\rm opt}$ has the same structure as a *second*-order approximant, suggesting again that the $N=-1$ case somehow lags one order behind the general case in its properties.

I note that the third-order optimization equations can also be solved explicitly for $N=-2$, although I do not know if this corresponds to a case of physical interest. The solution is $\bar{c}_2=0$, $\bar{r}_1=0$, $\bar{r}_2=\rho_2+c^2$, $\bar{\tau}=\rho_1$, so that

$$\mathcal{R}^{(3)\prime}_{\rm opt}(N=-2)=\bar{a}^{-2}+(\rho_2+c^2) \tag{6.12}$$

with $\bar{a}$ given by

$$\hat{K}^{(2)}(\bar{a})=\rho_1 . \tag{6.13}$$

D. Partially exponentiated perturbative approximations

It has been observed that in some cases second-order coefficients involve large π^2 terms, which can be proved to exponentiate[40,8] (or, at least, it can be shown that exponentiation is plausible). In such cases *one has an extra piece of information.* It is natural to make use of this by adopting partially exponentiated perturbative approximations,[40,8] i.e., in second order one would use

$$\mathcal{R}^{(2)}_E(a)\equiv a^Ne^{\kappa a}(1+s_1a) , \tag{6.14}$$

where κ is the (RS-invariant) part of the coefficient r_1, $=\kappa+s_1$, which is known, or supposed, to exponentiate.

One can apply the PMS criterion to this modified form of perturbation theory very easily. The generalization is very straightforward, so I need not go into details. I simply quote the formula for optimizing such partially exponentiated approximants in second order:

$$\mathcal{R}^{(2)}_E(\text{opt})=\frac{\bar{a}^Ne^{\kappa\bar{a}}}{(1+c\bar{a})}\frac{[(N+1)+c\bar{a}]}{[(N+1)+\kappa\bar{a}]} , \tag{6.15}$$

where $\bar{a}$ is given by

$$\hat{K}^{(2)}(\bar{a})+\frac{[Nc+\kappa(1+c\bar{a})]}{N(1+c\bar{a})[(N+1)+\kappa\bar{a}]}=\sigma_1 , \tag{6.16}$$

with σ_1 being the RS invariant

$$\sigma_1 \equiv \tau - \frac{1}{N} s_1$$

$$= \rho_1 + \frac{\kappa}{N} . \qquad (6.17)$$

Note that this is different from optimizing the conventional perturbative approximant and then taking that series and partially exponentiating it. However, there is no question about which is the proper procedure. One must first decide which approximation scheme one is using, i.e., the form of the approximants. One then selects from that set of approximants the one which is least sensitive to variations in the unphysical parameters.

VII. CONCLUSIONS

In the past few years it has become increasingly clear that perturbation theory in field theories is really rather more subtle than the old-fashioned idea of an expansion in powers of a fixed coupling constant. We have learned that the "coupling constant" can be more usefully considered as a function of a renormalization point chosen to suit the energy scale of the process in question. This is the so-called "renormalization-group-improved perturbation theory."[9] The present work carries this process one step further, providing a precise means of identifying the best choice of renormalization point. Moreover, it indicates how to tailor other aspects of the RS to suit the particular quantity being calculated. In this way the whole of the renormalization group[10,21]—not just the Gell-Mann–Low[14] subgroup associated with renormalization-*point* dependence—has a chance to be useful.

The connection with the renormalization group can be made clearer by restating the main argument in the following form. Physical quantities are RG invariants, a statement which translates into a set of RG equations:

$$\frac{\partial \mathcal{R}}{\partial \tau} = \left(\frac{\partial}{\partial \tau}\bigg|_a + \hat{\beta}(a) \frac{\partial}{\partial a} \right) \mathcal{R} = 0 \quad (\text{“}j = 1\text{”}) , \qquad (7.1)$$

$$\frac{\partial \mathcal{R}}{\partial c_j} = \left(\frac{\partial}{\partial c_j}\bigg|_a + \beta_j(a) \frac{\partial}{\partial a} \right) \mathcal{R} = 0 \quad (j = 2, 3, \ldots) . \qquad (7.2)$$

In ith-order perturbation theory only $(i-1)$ of these are nontrivial. These $(i-1)$ equations are satisfied *formally* in any RS, in the sense that the remainder term is of the same order as the terms already neglected. However, the remainder term is an indication of the direction in which the perturbation series is going, since one knows that the higher-order terms must conspire to cancel it exactly. The most reliable ith-order approximation corresponds to that $\mathcal{R}^{(i)}$ which satisfies Eqs. (7.1) and (7.2) exactly: This eliminates any *known* reason for the higher-order terms to be large. The PMS criterion simply requires the approximants to satisfy the RG equations. What could be a more natural way to make use of the renormalization group?

The optimized result (which is of the normal truncated-power-series form only in the optimum RS) is obtainable from perturbative calculations performed in *any* RS, and I have described the mechanics for doing this. I have named the proposed method optimized perturbation theory, to emphasize the conceptual differences from traditional ideas. The calculation of coefficients becomes only an intermediate step in computing the RS invariants $\rho_1, \ldots, \rho_i$, which then determine the optimized result through the optimization equations.

I stress that, particularly at second order, the method is quite straightforward and practical. There are obviously several immediate applications to existing QCD calculations which would help to put quantitative QCD phenomenology on a sounder theoretical basis. Further work is also needed on the extension of these ideas to massive field theories. Not only is this important for treating quark mass effects in QCD, but it would also open up applications to QED, electroweak gauge theories, and grand unified theories. The RS-ambiguity problem in these theories is, in principle, just as severe as in QCD, and is not always phenomenologically negligible either.[41]

Another problem to which the PMS criterion might also be applied is that of "factorization prescription dependence," discussed by Celmaster and Sivers.[42] This problem arises when dealing with hadrons in QCD perturbation theory by factorizing the collinear divergences of perturbation theory into parton distribution and/or fragmentation functions. The resulting cross sections depend on the "arbitrary" scale used in the factorization, as well as on the precise definition of the parton-hadron functions. This is strongly reminiscent of the RS-dependence problem,[42] and can probably be treated analogously. Presumably the factorization scale replaces the renormalization point, while the coefficients of the appropriate anomalous dimension play the role of the β-function coefficients. However, this conjecture requires further investigation.

Since this paper clearly stands or falls by the principle of minimal sensitivity, I close with some final remarks on this. The claim that the PMS criterion represents a general truth in approximation theory is not a statement susceptible of proof. The main reason for believing in it is the intrinsic logic of the argument, and because it uses, in a very direct way, the remarkable in-

variance property of the exact result—a piece of information that is otherwise wasted. As supporting evidence, I have also demonstrated that, in a number of examples, the PMS criterion clearly works.

I would also venture to argue on aesthetic grounds that the PMS criterion provides a natural way to harness the beauty and power of the renormalization group in perturbation theory. It forces one to consider the problem of the parametrization of the full renormalization group, the solution to which is unexpectedly simple and appealing. One is also led, slowly but inexorably, to discover the RS invariants $\rho_1, \rho_2, \ldots$. It seems hard to believe that such invariants could exist without being significant. Field theory has a habit of being strangely simple when one is doing the right thing (and ferociously difficult when one is not). Throughout this investigation I have been repeatedly surprised by the strangeness and simplicity of the mathematics: I take it as a confirmation of the soundness of the approach.

Note added in proof: Some aspects of the present work were remarkably foreshadowed in a paper by W. E. Caswell [Ann. Phys. (N.Y.) 123, 153 (1979)]. The paper treats the anharmonic oscillator by a method different from Halliday and Suranyi's, but with the same feature that finite-order results depend on an arbitrary parameter β. Caswell recognized that β should be chosen by requiring minimal sensitivity to small variations in β, and obtained very accurate results thereby. His concluding remarks on the implications for field theory make interesting reading.

ACKNOWLEDGMENTS

I thank Ian Halliday for arousing my interest in the work of Ref. 11, with such unexpected consequences. My thanks to Jacques Leveille and Eric Braaten for patiently listening to my ramblings, which helped in clarifying these ideas. In revising the original version of this work I was aided by the comments of several physicists. I am particularly grateful to Chris Llewellyn Smith for a very stimulating conversation, which influenced the paper in several respects. Part of this work was done at SLAC: My thanks to the theory group there for their hospitality—and to Jeff Mandula for provoking Sec. IIB. Finally, I thank Francis Halzen, Vernon Barger, and Taizo Muta for useful conversations and for much appreciated encouragement. This research was supported in part by the University of Wisconsin Research Committee with funds granted by the Wisconsin Alumni Research Foundation, and in part by the Department of Energy under Contract No. DE-AC02 76ER00881-185.

APPENDIX A: COUPLANTS, β FUNCTIONS, AND Λ-PARAMETERS

The following points are noted.

(1) It is more natural to write perturbation series in terms of (α_s/π) rather than $\alpha_s \equiv g^2/4\pi$, the so-called coupling constant. I use the symbol a for α_s/π, and refer to it as the couplant, both to distinguish it from α_s and to avoid the false implication that a is a constant. The couplant is a function of the renormalization point μ as well as depending on the RS through the parameters $c_2, c_3, \ldots$, i.e.,

$$a = a(\tau; c_2, c_3, \ldots), \quad \tau \equiv b \ln(\mu/\tilde{\Lambda}). \tag{A1}$$

The μ dependence of the couplant is governed by the β function, which I define here as

$$\mu \frac{\partial a}{\partial \mu} \equiv \beta(a), \tag{A2}$$

$$\beta(a) = -ba^2(1 + ca + c_2a^2 + c_3a^3 + \cdots). \tag{A3}$$

The coefficients b and c are fundamental constants of the theory. In massless QCD with N_f flavors[43,44]

$$b = \frac{(33 - 2N_f)}{6}, \tag{A4}$$

$$c = \frac{(153 - 19N_f)}{2(33 - 2N_f)}. \tag{A5}$$

(2) The higher-order β-function coefficients $c_2, c_3, \ldots$ are RS dependent, and can be used to label RS's. The dependence of a on these parameters is given by

$$\frac{\partial a}{\partial c_i} = \beta_i(a) \equiv -\hat{\beta}(a) \int_0^a \frac{x^{i+2}dx}{(\hat{\beta}(x))^2}, \tag{A6}$$

where $\hat{\beta}(a) \equiv \beta(a)/b$. These functions have power-series expansions

$$\beta_i(a) = \frac{1}{(i-1)} a^{i+1}(1 + W_1^i a + W_2^i a^2 + \cdots), \tag{A7}$$

with coefficients W_j^i given by the determinant

$$\begin{vmatrix} -\begin{vmatrix} -2 \\ 0 \end{vmatrix} c + \begin{vmatrix} -3 \\ 1 \end{vmatrix} c_2 - \begin{vmatrix} -4 \\ 2 \end{vmatrix} c_3 \cdots (-1)^j \begin{vmatrix} -j-1 \\ j-1 \end{vmatrix} c_j \\ 1 \quad -\begin{vmatrix} -1 \\ 1 \end{vmatrix} c + \begin{vmatrix} -2 \\ 2 \end{vmatrix} c_2 \cdots -(-1)^j \begin{vmatrix} -j+1 \\ j-1 \end{vmatrix} c_{j-1} \\ 1 \quad -\begin{vmatrix} 0 \\ 2 \end{vmatrix} c \cdots (-1)^j \begin{vmatrix} -j+3 \\ j-1 \end{vmatrix} c_{j-2} \\ 1 \quad -\begin{vmatrix} j-3 \\ j-1 \end{vmatrix} c \end{vmatrix}, \tag{A8}$$

where $\left|\begin{matrix} n \\ m \end{matrix}\right|$ is shorthand for $[(i+n)/(i+m)]$.

(3) Equation (A2) can be integrated to give

$$\tau \equiv b \ln(\mu/\tilde{\Lambda}) = \hat{K}(a) \equiv \int_0^a \frac{dx}{\hat{\beta}(x)} + (\text{infinite}) \text{ constant}. \tag{A9}$$

A particular choice of the infinite constant corresponds to a definition of $\tilde{\Lambda}$. It is natural to use a simple, known integral whose integrand has the same singularity structure as $1/\hat{\beta}(x)$ as $x \to 0$. The particular choice is purely a matter of convention. My definition of $\tilde{\Lambda}$ is based on the choice[45]

$$\text{infinite constant} = -\int_0^\infty \frac{dx}{\hat{\beta}^{(2)}(x)} = \int_0^\infty \frac{dx}{x^2(1+cx)}, \tag{A10}$$

so that

$$\tau = \hat{K}^{(2)}(a) + \int_0^a dx \left[\frac{1}{\hat{\beta}(x)} - \frac{1}{\hat{\beta}^{(2)}(x)}\right], \tag{A11}$$

where

$$\hat{K}^{(2)}(a) \equiv -\int_a^\infty \frac{dx}{\hat{\beta}^{(2)}(x)} = \frac{1}{a} + c \ln\left(\frac{ca}{1+ca}\right). \tag{A12}$$

The second term in Eq. (A11) is a convergent integral and gives a term $O(a)$. In second order, as implied by the notation, τ is given by $\hat{K}^{(2)}(a)$ as defined in Eq. (A12).

Since the integration of the β-function equation is performed at constant c_i, it is not obvious *a priori* that one should assume that the integration constant is independent of the c_i, as in Eq. (A10). The question comes down to whether this assumption is compatible with the limits of integration used in the definition of the β_i functions, which were chosen such that $\beta_i(a) = O(a^{i+1})$. This can be checked *a posteriori* by showing that the partial derivatives of the right-hand side of Eq. (A11) with respect to c_i $(i = 2, 3, \ldots)$ are zero, as they should be.

The conventional Λ parameter is defined in terms of an expansion of a in powers of $1/\ln(\mu^2/\Lambda^2)$ (see Ref. 4 for details). That definition corresponds to using a constant which differs from Eq. (A10) by a term $c \ln(b/2c)$, so that

$$b \ln(\Lambda/\tilde{\Lambda}) = -c \ln(b/2c)$$

i.e.,

$$\tilde{\Lambda} = \Lambda\left(\frac{2c}{b}\right)^{-c/b}. \tag{A13}$$

Thus the two parameters are related by a known, RS-independent factor, which however, depends on the number of quark flavors. (It varies between 1.10 and 1.18 for three and six flavors.) There is no physics reason for preferring one definition over the other. I elect to use $\tilde{\Lambda}$ rather than Λ for algebraic convenience, and to avoid the spurious reappearance of the constant b in the $\hat{K}$ functions, which I find unaesthetic.

(4) The definition of $\tilde{\Lambda}$ through Eq. (A11) is a RS-dependent definition, simply because a and $\hat{\beta}(a)$ are RS dependent. However, $\tilde{\Lambda}$'s in different RS's can be related *exactly* by a one-loop calculation. The argument is due to Celmaster and Gonsalves,[1] and I repeat it here for completeness.

Consider two different types of RS with couplants related by

$$a'(\mu) = a(\mu)[1 + v_1 a(\mu) + \cdots]. \tag{A14}$$

Note that this relation is written with the renormalization-point parameters (which may have quite different *meanings* in the two schemes) having the same *value*, μ. Writing Eq. (A11) in the two schemes, again with the same μ values, one has

$$\begin{aligned} \tau &\equiv b \ln(\mu/\tilde{\Lambda}) = \hat{K}^{(2)}(a) + O(a) \\ \tau' &\equiv b \ln(\mu/\tilde{\Lambda}') = \hat{K}^{(2)}(a') + O(a'). \end{aligned} \tag{A15}$$

Then, using Eqs. (A12) and (A14), one can quickly show that

$$\tau - \tau' = b \ln(\tilde{\Lambda}'/\tilde{\Lambda}) = \frac{1}{a} - \frac{1}{a'} + O(a) = v_1 + O(a). \tag{A16}$$

Since the argument is true for all μ, one can, in particular, take $\mu \to \infty$. The $O(a)$ term will then tend to zero because of asymptotic freedom. [For other theories one could work the same trick by taking $\mu \to 0$. The extra terms in Eq. (A16) actually cancel for any value of μ, but this fact is made obvious only by a shrewd choice of μ.] Therefore the relation between $\tilde{\Lambda}$ (or Λ) parameters in different RS's is given exactly by

$$\frac{\tilde{\Lambda}'}{\tilde{\Lambda}} = \frac{\Lambda'}{\Lambda} = \exp(v_1/b). \tag{A17}$$

This result is easy to understand. The definitions of the μ parameters in the two schemes are, in general, quite unrelated. The coefficient v_1 reflects this mismatch, and the relation between the $\tilde{\Lambda}$'s simply corrects for this so that the ratio $\mu/\tilde{\Lambda}$ has, loosely speaking, the same significance in each RS. (For example, the optimum value of $\mu/\tilde{\Lambda}$, at any order, is a RS invariant, but the optimum renormalization point $\bar{\mu}$ is scheme dependent. This is one reason for using τ rather than μ as a RS parameter.)

(5) Finally, I discuss the relation of the $\tilde{\Lambda}$ parameter to the one-parameter ambiguity of QCD. Viewed perturbatively, massless QCD is a theory with one, and only one, free parameter. In the original Lagrangian this free parameter is the bare coupling constant g_0. However, this turns out not to be finite, and the theory must be renormalized, i.e., *reparametrized.* In the renormalized theory the one-parameter ambiguity is manifested by the absence of a boundary condition for the β-function equation. The missing boundary-condition information is economically parametrized by $\tilde{\Lambda}$, as defined in paragraph 3 above, and so it is convenient to regard $\tilde{\Lambda}$ as the free parameter of QCD which must be fitted to experiment.[46]

However, since $\tilde{\Lambda}$ is a RS dependent concept, it is vital to be clear about *which* $\tilde{\Lambda}$ one is talking about. The simplest plan is to adopt $\tilde{\Lambda}$ in some convenient, "reference" scheme (such as $\overline{\text{MS}}$[4]) as *the* free parameter of QCD, always referring other $\tilde{\Lambda}$'s back to this one. The choice of a reference scheme is genuinely arbitrary: It is purely a bookkeeping convention and does not affect the physics at all.

The RS-independent predictions of optimized perturbation theory depend on the free parameter of QCD only through ρ_1. If the RS used in the calculation is not the same as the reference scheme, then one can rewrite ρ_1 as

$$\rho_1 = b\ln(\mu/\tilde{\Lambda}_{\text{ref}}) + b\ln(\tilde{\Lambda}_{\text{ref}}/\tilde{\Lambda}) - \frac{1}{N}r_1\,, \tag{A18}$$

where the second term is obtainable from the one-loop relation between the two schemes, as described in paragraph 4. In this way all purely perturbative QCD predictions can be expressed unambiguously in terms of the fit parameter $\tilde{\Lambda}_{\text{ref}}$.

The procedure for testing the theory is exactly the same in principle as for any theory with one free parameter, i.e., one calculates predictions for N physical quantities and compares them with experiment: This provides $(N-1)$ tests of the theory, and yields a best-fit value for $\tilde{\Lambda}_{\text{ref}}$. In practice, since QCD predictions are complicated, transcendental functions of $\tilde{\Lambda}_{\text{ref}}$, it would be hard to do a proper fitting procedure. However, one can always use the direct, unsophisticated method, i.e., simply evaluate the set of N predictions for various values of $\tilde{\Lambda}_{\text{ref}}$ in the likely range, and see which value works best. From this point of view, it is not too much of a distortion of the practical situation to pretend, as I do in the text, that $\tilde{\Lambda}_{\text{ref}}$ is a known quantity.

APPENDIX B: SPECIAL CONSIDERATIONS IN GAUGE THEORIES

The quantization of gauge theories requires a choice of gauge before a renormalization prescription can be defined. One might therefore suspect that extra parameters besides $\tau, c_2, c_3, \ldots$ are needed to fully characterize the RS, implying that, for gauge theories, the analysis described in the text is incorrect, or incomplete. This is not the case. As so often in gauge theories, the apparent extra degrees of freedom are essentially spurious. However, the problem requires some discussion.

Firstly, I can divide the types of RS's in gauge theories into two classes: "gauge invariant" and "gauge noninvariant." The former class includes minimal subtraction and an infinite number of other schemes related to it by

$$a = a_{\text{min}}(1 + v_1 a_{\text{min}} + v_2 a_{\text{min}}{}^2 + \cdots)\,, \tag{B1}$$

where $v_1, v_2, \ldots$ are gauge-invariant, but otherwise arbitrary quantities. For minimal subtraction it is known (see Ref. 47, Sec. 5.4) that

$$\left.\frac{\partial\beta}{\partial\xi}\right|_a = 0\,, \tag{B2}$$

where ξ is the renormalized gauge parameter. This means that the β-function coefficients in minimal subtraction are gauge invariant. Obviously all RS's in the class defined by Eq. (B1) will share this property. It follows, by the ubiquitous self-consistency argument of Sec. III D, that perturbative approximations to physical quantities must be gauge invariant in such schemes, i.e., gauge invariance is respected *order by order.* Thus, within the class of "gauge-invariant RS's," all gauge dependences have canceled before my analysis begins.

Renormalization schemes of the second class (e.g., those of the momentum subtraction type[1,6]) do not respect gauge invariance order by order. In such schemes there are two inequivalent ways of defining the β function, according to whether the *renormalized* or the *bare* gauge parameter is held fixed, i.e.,

$$\mu\left.\frac{\partial a}{\partial\mu}\right|_{\xi} \equiv \beta(a)\,, \tag{B3}$$

$$\mu\left.\frac{\partial a}{\partial\mu}\right|_{\xi_{\text{bare}}} \equiv \bar{\beta}(a)\,. \tag{B4}$$

(In gauge-invariant RS's $\beta = \bar{\beta}$.[47]) Since it is not in the spirit of this paper to deal with bare quantities, I use Eq. (B3) and not (B4) as the definition of the β function. The reader is warned that this differs from the standard literature[48]; in fact my $\beta, \bar{\beta}$ notation is opposite to that of Gross.[47] This is not mere perversity. The point is that the statement that the first *two* coefficients of the β function are RS independent is true *only* for β de-

fined by Eq. (B3).

[Otherwise one runs into a problem pointed out by Llewellyn Smith.[49] In the usual argument for the RS independence of the second β-function coefficient,[19] it is assumed that v_1 in

$$a' = a(1 + v_1 a + \cdots) \tag{B5}$$

is independent of μ. This is not true if the μ derivatives are taken with ξ_{bare} held constant, because the renormalized ξ is then a running (μ-dependent) gauge parameter. An extra term arises, namely

$$a^2\left(\mu\frac{\partial v_1}{\partial\mu}\bigg|_{\xi_{\text{bare}}}\right) = a^2\frac{\partial v_1}{\partial\xi}\left(\mu\frac{\partial\xi}{\partial\mu}\bigg|_{\xi_{\text{bare}}}\right), \tag{B6}$$

which is $O(a^3)$ since $\mu(\partial\xi/\partial\mu)$ begins at order a. Thus the second-order coefficients of $\bar{\beta}(a)$ and $\bar{\beta}'(a')$ differ by a term proportional to $\partial v_1/\partial\xi$. For the same reason the argument in Sec. III B that $\partial c/\partial(\text{RS}) = 0$ would break down if the β function were defined by Eq. (B4).]

With the adoption of the definition (B3) for the β function, the question of extra RS parameters becomes merely a semantic difficulty. To speak of scheme "A" (which might be, say, momentum subtraction based on the quark-gluon vertex[1,2,6]) as being one RS which is gauge dependent, is unhelpful. Rather, one should think of the Landau gauge and the Feynman gauge versions (for example) of scheme A as two *different* RS's. The fact that the two RS's are similarly defined and go under the same name is not relevant: What matters is that they have different values for $\tilde{\Lambda}$ and for the c_i parameters, and therefore yield different perturbative approximations.

Conversely, if two RS's have the same value for $\tilde{\Lambda}$ and for the c_i, then (for the same renormalization point μ) they always give the same perturbative approximations. The proof goes as before: Equal τ's and c_j's guarantee the equality of the $a^{(i)}$'s in the two RS's and the self-consistency argument shows that perturbative coefficients cannot depend on any parameter that $a^{(i)}$ does not depend on. Therefore, by varying τ and c_j, one ranges over the entire set of possible perturbative approximations: There are no extra RS parameters which need to be varied.

This conclusion can be reconfirmed in the following way. Suppose, to the contrary, that the gauge parameter ξ represented an extra, independent, RS parameter. The dependence of the couplant on ξ would have to be of the form

$$\frac{\partial a}{\partial\xi}\bigg|_{\mu} \equiv \beta_\xi(a) = \chi a^2 + \chi_1 a^3 + \cdots, \tag{B7}$$

because self-consistency requires that $\beta_\xi(a)$ be $O(a^2)$. The coefficients $\chi, \chi_1, \ldots$ can be ξ dependent, but cannot, by dimensional analysis, depend explicitly on μ. Since ξ is supposed to be an independent parameter, $\partial/\partial\xi\,|_\mu$ and $\partial/\partial\mu\,|_\xi$ must commute. Therefore

$$\mu\frac{\partial^2 a}{\partial\mu\,\partial\xi} = \beta(a)\beta'_\xi(a) \tag{B8}$$

$$= \mu\frac{\partial^2 a}{\partial\xi\,\partial\mu}$$

$$= \beta_\xi(a)\beta'(a) - b\left(a^4\frac{\partial c_2}{\partial\xi} + a^5\frac{\partial c_3}{\partial\xi} + \cdots\right), \tag{B9}$$

where the prime indicates differentiation with respect to a, regarding the series coefficients as fixed. From this one finds that

$$\beta_\xi(a) = -\frac{\chi}{b}\beta(a) + \sum_{i=2}^{\infty}\frac{\partial c_i}{\partial\xi}\beta_i(a), \tag{B10}$$

where the β_i functions are defined in Eq. (3.21). The first coefficient χ remains an undetermined constant of integration (which can be fixed by a one-loop calculation), and corresponds to the possibility of redefining the normalization of ξ.

The important point about Eq. (B10) is that it shows that $\beta_\xi(a)$ is a linear combination of the β and β_i functions. Thus, a satisfies a constraint of the form

$$\left[\frac{\partial}{\partial\xi} + L\left(\mu\frac{\partial}{\partial\mu}, \frac{\partial}{\partial v_i}\right)\right]a = 0, \tag{B11}$$

where L is a linear combination of derivatives with respect to the RS parameters $\mu; v_2, v_3, \ldots$. This simply means that ξ is not an *independent* parameter.

With my usual convention of identifying the RS parameters v_i with the c_i, the argument becomes even simpler. Equation (B9) simplifies because $\partial c_i/\partial\xi\,|_{c_2,c_3,\ldots} = 0$ by definition, so at constant $c_2, c_3, \ldots$, any variation in ξ is equivalent to a shift in μ.

APPENDIX C: COMPARISONS WITH SOME OTHER APPROACHES

In this appendix I comment on the relationship of the present work to a few of the alternative proposals for dealing with the RS-ambiguity problem.

(1) A very clear account of the nature and seriousness of the problem was given in the papers of Celmaster and Gonsalves,[1,2] which greatly influenced the present work. Their proposed solution—the use of a particular type of RS known as "momentum subtraction"—has been studied extensively; in particular, by Celmaster and Sivers.[6] The latter paper also contains a great deal of wisdom and good sense about the RS-dependence

problem and its importance. On the question of where the solution should be sought, there is a great difference in philosophy between their approach and mine. Nevertheless, in terms of their effects, the two approaches are far from being contradictory.

I would explain the virtue of the momentum-subtraction idea with a rather different emphasis from most authors. The principal ambiguity in perturbative QCD calculations lies in the choice of the renormalization point μ. This is the entirety of the problem at second order, and remains the most serious ambiguity in higher orders. The virtue of a momentum-subtraction RS is not so much that its couplant is intrinsically a "good expansion parameter," but rather that *it provides a framework within which one can make an educated guess as to the best value of* μ. In dimensional-renormalization schemes (such as $\overline{\text{MS}}$) one has no intuitive feeling for a good, *a priori* choice of μ because the "interpretation" of this parameter is obscure. Momentum-subtraction schemes, by contrast, are defined such that the radiative corrections to a particular vertex are absorbed into the couplant when the incoming momenta have particular values characterized by μ. Thus one can expect to obtain a well-behaved series with such a couplant if μ is chosen to be a "typical" momentum flowing through the relevant diagrams.

This argument is correct, but the problem with this approach is its inherent vagueness: The choice of μ is still left to one's intuition. The connection with the present work is that, in a momentum-subtraction scheme, one can expect the intuitive choice of μ and the PMS choice to be, in general, very similar. There are several indications that this is in fact the case: Ref. 6 tends to obtain small, and often negative, second-order coefficients, as one would also find with the PMS criterion (see also Ref. 8). I would argue that the PMS criterion provides the means to convert the "art" of choosing a RS, described in Ref. 6, into a "science."

(2) An interesting paper by Grunberg[5] is closer to the spirit of the present work in many ways. It too proposes to choose the RS *a posteriori*, such that the approximant has a particular property. Grunberg uses what I call the fastest apparent convergence criterion, requiring the calculated corrections to the leading-order result to vanish. He shows that this corresponds to a choice of μ, thereby connecting directly with the work of Wolfram.[25] However, as explained in Sec. II, I have reservations about the FAC criterion, which I believe is logically, and practically, less satisfactory than the PMS criterion. The former does not use the very special renormalization-group-invariance property of physical quantities which is so essential to the latter.

(3) Most other approaches make the implicit assumption that it is mandatory to use the same RS for every physical quantity. As I said in the Introduction, this assumption is false. Such approaches are therefore fundamentally misguided.

---

[1]W. Celmaster and R. J. Gonsalves, Phys. Rev. D 20, 1420 (1979); Phys. Rev. Lett. 42, 1435 (1979).

[2]W. Celmaster and R. J. Gonsalves, Phys. Rev. D 21, 3112 (1980); Phys. Rev. Lett. 44, 560 (1980).

[3]E. G. Floratos, D. A. Ross, and C. T. Sachrajda, Nucl. Phys. B129, 66 (1977); B139, 545 (E) (1978); B152, 493 (1979); M. Moshe, Phys. Lett. 79B, 88 (1978); 80B, 433 (E) (1980); A. N. Schellekens, Lett. Nuovo Cimento 24, 513 (1979); R. Barbieri, L. Caneschi, G. Curci, and E. d'Emilio, Phys. Lett. 81B, 207 (1979); A. Para and C. T. Sachrajda, *ibid.* 86B, 331 (1979); M. R. Pennington and G. G. Ross, *ibid.* 86B, 371 (1979); Oxford Report No. 19/80, 1980 (unpublished); L. F. Abbott and R. M. Barnett, Ann. Phys. (N.Y.) 125, 276 (1980); L. F. Abbott, Phys. Rev. Lett. 44, 1569 (1980); A. J. Buras, Rev. Mod. Phys. 52, 199 (1980); P. Pascual and R. Tarrach, Barcelona, Report No. UBFT-1-80, 1980 (unpublished); Nucl. Phys. B174, 123 (1980).

[4]A. J. Buras, E. G. Floratos, D. A. Ross, and C. T. Sachrajda, Nucl. Phys. B131, 308 (1977); W. A. Bardeen, A. J. Buras, D. W. Duke, and T. Muta, Phys. Rev. D 18, 3998 (1978).

[5]G. Grunberg, Phys. Lett. 95B, 70 (1980).

[6]W. Celmaster and D. Sivers, Phys. Rev. D 23, 227 (1981).

[7]P. M. Stevenson, Wisconsin Report No. DOE-ER/00881-153, 1980 (unpublished). A brief summary of this work is given in Phys. Lett. 100B, 61 (1981).

[8]C. H. Llewellyn Smith in *High Energy Physics—1980,* proceedings of the XXth International Conference, Madison, Wisconsin, edited by L. Durand and L. G. Pondrom (AIP, New York, 1981), p. 1346.

[9]See, for example, H. D. Politzer, Phys. Rep. 14C, 129 (1974). The idea has a long history, going back to Gell-Mann and Low (Ref. 14).

[10]E. C. G. Stueckelberg and A. Petermann, Helv. Phys. Acta 26, 449 (1953).

[11]I. G. Halliday and P. Suranyi, Phys. Rev. D 21, 1529 (1980); Phys. Lett. 85B, 421 (1979).

[12]The term "renormalization point" is used here to denote the arbitrary parameter μ, having the dimensions of mass, that must always be introduced when renormalizing a theory, such as massless QCD, with a scale-invariant Lagrangian. The "minimal subtraction" scheme does possess such a parameter, although it does not have an interpretation as a subtraction point as in older schemes (see Ref. 23).

[13]P. M. Stevenson, Ann. Phys. (N.Y.) (to be published).
[14]M. Gell-Mann and F. Low, Phys. Rev. 95, 1300 (1954).
[15]C. M. Bender and T. T. Wu, Phys. Rev. 184, 1231 (1969); Phys. Rev. D 7, 1620 (1973).
[16]I. G. Halliday, private communication.
[17]F. T. Hioe, D. MacMillen, and E. W. Montroll, Phys. Rep. 43C, 305 (1978).
[18]Several physicists have been kind enough to inform me of occurrences of noninvariant approximation schemes in other branches of physics. An insensitivity criterion, essentially equivalent to PMS, was employed by J. W. Dash and S. J. Harrington, Phys. Lett. 57B, 78 (1975) in the context of renormalization-point independence in the Reggeon calculus. A paper by L. P. Kadanoff and A. Houghton, Phys. Rev. B 11, 377 (1975) uses a FAC-type criterion, requiring two different approximations to agree, in evaluating critical properties of the two-dimensional Ising model. R-matrix theory in nuclear scattering provides another instance: Physical results should not depend on the radius of the sphere used to divide "internal" and "external" regions, in which expansions in different complete sets of eigenfunctions are used. For a review, see A. M. Lane and R. G. Thomas, Rev. Mod. Phys. 30, 257 (1958). Atomic physics provides another interesting example, where one uses perturbations about a one-electron atom with an "effective atomic number" $Z-s$. This is similar to the anharmonic-oscillator example in that it arises from adding and subtracting terms to the Hamiltonian. The optimum choice of the "screening parameter" s is different for different physical properties, and in this case the PMS criterion is equivalent to requiring the last, calculated term to vanish (FAC′) [see A. Dalgarno and A. L. Stewart, Proc. R. Soc. London A257, 534 (1960) and W. A. Sanders and J. O. Hirschfelder, J. Chem. Phys. 42, 2904 (1965)]. I thank, respectively, Professor J. W. Dash, Professor L. Bruch, Professor W. Friedman, and Professor S. T. Epstein for bringing these cases to my attention. No doubt there are many more such examples.
[19]See, for example, J. Ellis and C. T. Sachrajda, to appear in the proceedings of the 1979 Cargese Summer Institute [CERN Report No. TH 2782, 1979 (unpublished)]. See also Appendix B of Ref. 6.
[20]N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Wiley-Interscience, New York, 1959).
[21]A. Peterman, Phys. Rep. 53C, 157 (1979).
[22]Any two couplants are related by $a'=a[1+O(a)]$, so the notations $O(a^{i+1})$ and $O((a')^{i+1})$ are always interchangeable, a fact which is important to keep in mind in these formal arguments.
[23]G. 't Hooft, Nucl. Phys. B61, 455 (1973).
[24]M. Bače, Phys. Lett. 78B, 132 (1978).
[25]S. Wolfram, Caltech Report No. 68-690, 1978 (unpublished).
[26]I am assuming here that c is positive (i.e., $N_f \leq 8$ in QCD).
[27]It is beyond the scope of this paper to discuss the optimization of quantities explicitly involving phenomenological parton distribution and/or fragmentation functions. See the discussion in Sec. VII.
[28]If c is negative, which can happen in QCD if $9 \leq N_f \leq 16$, then the logarithmic singularity of $R^{(2)}(a)$ occurs at the positive value $a=-1/c$. The second stationary point of $R^{(2)}(a)$ then occurs in the uninteresting region beyond this singularity. Clearly one needs $a<1/|c|$ because of the threat of zero of the β function at or near $a=1/|c|$. [Note also that the integrals in Eq. (A10) and (A12) should be interpreted as principal values if c is negative.]
[29]G. 't Hooft, in *Deeper Pathways in High-Energy Physics*, proceedings of Orbis Scientiae 1977, Coral Gables, edited by A. Perlmutter and L. F. Scott (Plenum, New York, 1977), p. 699. See also W. Marciano and H. Pagels, Phys. Rep. 36C, 137 (1978), Sec. 7.2.1.
[30]See Appendix (2) of Ref. 36.
[31]However, the usual argument for the divergence of QED perturbation series [F. J. Dyson, Phys. Rev. 85, 631 (1952)] possesses a loophole. Dyson points out that the vacuum is unstable for α negative, no matter how small, and therefore QED results cannot be analytic at $\alpha=0$. This is quite correct, but a function can be nonanalytic at the origin and still have a convergent perturbation series. For example, $f(\alpha)\equiv 1/(1-\alpha)+\exp(-1/\alpha)$ has the series expansion $1+\alpha+\alpha^2+\cdots$, which converges—although not to the right answer: It misses the nonperturbative term. This function does have the property that it is unbounded for α small and negative. Bearing in mind what has been learned about instantons, etc., this may not be an unrealistic scenario for field theories.
[32]K. G. Chetyrkin, A. L. Kataev, and F. V. Tkachov, Phys. Lett. 85B; 277 (1979); M. Dine and J. Sapirstein, Phys. Rev. Lett. 43, 668 (1979).
[33]J. Calmet, S. Narison, M. Perrottet, and E. de Rafael, Rev. Mod. Phys. 49, 21 (1977).
[34]T. Kinoshita, Nuovo Cimento B51, 140 (1967); B. E. Lautrup and E. de Rafael, Nucl. Phys. B70, 317 (1974).
[35]R. Barbieri and E. Remiddi, Phys. Lett. 57B, 273 (1975).
[36]See, for example, R. Coquereaux, Ann. Phys. (N.Y.) 125, 401 (1980). This paper provides an instructive and thorough treatment of different RS's in QED.
[37]Since the optimization process is not linear, it should be applied to "elementary" quantities, not to arbitrary combinations of physical quantities (cf. the last paragraph of Sec. II A). The precise definition of elementary is problematic, but two common sense criteria seem to be adequate in practice. Clearly one does not want to mix quantities of different types, or quantities involving different energy scales. Also, one wants to use the longest possible perturbation series—taking ratios or differences in which the first-order terms cancel, prior to optimization, would discard some information which the optimization procedure could use. For both reasons I optimize the electron and muon anomalies separately, rather than, say, the $(g-2)$ of the electron and the difference between the muon and electron anomalies.
[38]E. Witten, Nucl. Phys. B120, 189 (1977); C. H. Llewellyn Smith, Phys. Lett. 79B, 83 (1978); W. R. Frazer, and J. F. Gunion, Phys. Rev. D 20, 147 (1979); R. J. Dewitt, L. M. Jones, J. D. Sullivan, D. E. Willen, and H. W. Wyld, Jr., *ibid.* 19, 2046 (1979); 20, 1751 (E) (1979).
[39]W. A. Bardeen and A. J. Buras, Phys. Rev. D 20, 166

(1979); 21, 2041 (E) (1980); D. W. Duke and J. F. Owens, *ibid.* 22, 2280 (1980).
[40]G. Parisi, Phys. Lett. 90B, 295 (1980); G. Curci and M. Greco, *ibid.* 92B, 175 (1980).
[41]Compare W. J. Marciano and A. Sirlin, Phys. Rev. D 22, 2695 (1980); M. Veltman, Phys. Lett. 91B, 95 (1980); F. Antonelli, M. Consoli, and G. Corbò, *ibid.* 91B, 90 (1980); V. Elias and T. N. Sherry, Toronto report, 1980 (unpublished).
[42]W. Celmaster and D. Sivers, Argonne Report No. ANL-HEP-PR-80-61, 1980 (unpublished).
[43]H. D. Politzer, Phys. Rev. Lett. 30, 1346 (1973); D. J. Gross and F. Wilczek, *ibid.* 30, 1343 (1973).
[44]D. R. T. Jones, Nucl. Phys. B75, 531 (1974); W. Caswell, Phys. Rev. Lett. 33, 244 (1974).
[45]This differs from my earlier definition in Ref. 7, which is unsatisfactory, and is inconsistent with the β_i functions (cf. the point to be made in the next paragraph). The earlier definition is, however, equivalent to the new definition in second order.
[46]This phenomenon, whereby an apparently scale-invariant theory ends up possessing a scale parameter, is known as "dimensional transmutation." See S. Coleman and E. Weinberg, Phys. Rev. D 7, 1888 (1973). For further discussion see Ref. 13, and Th. A. J. Maris, Rev. Bras. Física 6, 31 (1976).
[47]D. J. Gross, in *Methods in Field Theory*, edited by R. Balian and J. Zinn-Justin (North-Holland, Amsterdam, 1976).
[48]For example, W. Marciano and H. Pagels, Phys. Rep. 36C, 137 (1978).
[49]C. H. Llewellyn Smith, private communication.