Topics in perturbation theory

Barclay, David Thomas

Topics
in
Perturbation Theory

David Thomas Barclay
Department of Physics
University of Durham

A thesis submitted to the University of Durham
for the Degree of Doctor of Philosophy
July 1992

Abstract

In providing a means of progressively improving an initial estimate, perturbation series have become a ubiquitous tool in modern physics. However, and mainly because this stepwise process of improvement rapidly becomes increasingly involved, surprisingly little is known about the formal properties of the series obtained. This thesis therefore investigates some aspects of these properties and how they effect the application of these techniques, with an emphasis on quantum field theory and the phenomenology of e^+e^- colliders.

One of the better understood examples of a perturbative series is the WKB one which is widely used to approximate the energy levels of quantum mechanical systems. Recently much interest has centred on a modification of this, the SWKB series. Apart from (possibly) offering an improvement on the original, this is intrinsically interesting in being related to the supersymmetry of field theory. Furthermore, as Chapter 1 explains, there is a close connection between the cases where the initial estimate requires no correction and an important set of quantum mechanical problems (the "shape invariant" ones) which can be solved elegantly and completely.

The situation in field theory is more complicated, not least because the series for any particular problem is no longer unique. While this presents few theoretical difficulties, it has serious consequences when attempts are made to compare predictions with experiment. This obstacle is particularly severe in Quantum Chromodynamics and its fundamental constant (Λ_{QCD}) is therefore only roughly known at present. It will be argued that current responses to this are all imperfect, but that tests of the theory can be envisaged that circumvent the problem. This leads into questions concerning the origin of the divergences in the perturbation series – for although it may initially provide usefully improved estimates, the series probably breaks down eventually. Existing arguments about this topic are critically reviewed – and in one case substantially simplified – before an alternative one is proposed in some detail. By concentrating on a particularly restricted situation, the Common Effective Charge Approach simplifies matters to the extent that issues such as non-analyticity of functions and the potential accuracy of perturbative techniques in realistic applications can be conveniently investigated.

Acknowledgements

Foremost thanks clearly go to Chris Maxwell for his supervision, encouragement and patience over the last three years. Even when getting nowhere, arguing about some of each others more stupid ideas was great fun.

Thanks also to (in alphabetical order) Foster and Peter for putting up with me in a confined space over the same period.

Any futher listing would have to be so long, varied and inevitably incomplete that I'll have to thank the rest of you – friends both in the department and outside it who have made my stay in Durham so enjoyable and memorable – together. It's no less heartfelt for that.

I am grateful to the SERC for financing this research.

And apologies to Duncan (in particular) for the next page but one ...

DECLARATION

I declare that no material in this thesis has previously been submitted for a degree at this or any other university.

All the research in this thesis has been carried out in collaboration with Dr. C.J. Maxwell. That in Chapter 1 is partly summarised in the paper

D.T. Barclay and C.J. Maxwell, Phys.Lett.**A157**(1991)357.

Some of the material in Chapter 3 has previously appeared in

D.T. Barclay and C.J. Maxwell, Univ. of Durham preprint, DTP-92/26 (1992), submitted to Phys.Lett.**B**,

and some of that in Chapters 4 to 6 in

D.T. Barclay and C.J. Maxwell, Phys.Rev.**D45**(1992)1760.

A summary of part of Chapter 4 has previously appeared as

D.T. Barclay and C.J. Maxwell, Univ. of Durham preprint, DTP-91/72 (1991),

a condensed version of which is to appear in Phys.Rev.Lett. The copyright of this thesis rests with the author.

"The best amateur history, however entertaining, cannot enlarge the understanding or deepen the participation because it is written from outside, through a veil woven out of strangeness and wonderment. At its best it achieves sympathy and romantic love, but it cannot penetrate to fundamental explanation; at its common bad it is sentimental, ignorant and an insult to the intelligence."

G.R.Elton, *The Practice of History*

"**Quotation,** *n.*The act of repeating erroneously the words of another. The words erroneously repeated."

Ambrose Bierce

Contents

CHAPTER ONE

The SWKB Series

1.1 Supersymmetric Quantum Mechanics

Despite a complete lack of any evidence for experimental manifestations, supersymmetry (SUSY) is a dominant theme in present day particle physics on the basis of its spectacular technical sucesses in quantum field theory. Naturally this has tended to eclipse the parallel application of $N = 2$ SUSY to much simpler problems in quantum mechanics where the formalism can be shown to have verifiable consequences. In a superspace language this is an invariance under the transformations

$$\theta \to \theta + \epsilon, \qquad \overline{\theta} \to \overline{\theta} + \overline{\epsilon}, \qquad t \to t + i\overline{\epsilon}\theta + i\epsilon\overline{\theta} \tag{1.1}$$

on a scalar variable t and Grassmanian ones θ and $\overline{\theta}$. The SUSY generators

$$Q = \frac{\partial}{\partial\overline{\theta}} + i\theta\frac{\partial}{\partial t} \qquad \overline{Q} = \frac{\partial}{\partial\theta} + i\overline{\theta}\frac{\partial}{\partial t} \tag{1.2}$$

are used to define a Hamiltonian

$$H = \frac{1}{2}\{Q, \overline{Q}\} \tag{1.3}$$

and a SUSY algebra

$$\{Q, Q\} = \{\overline{Q}, \overline{Q}\} = 0 \qquad [Q, H] = [\overline{Q}, H] = 0. \tag{1.4}$$

In one dimension (now to be called x rather than t) this algebra has a particularly important representation introduced by Witten [1] in which

$$Q = \frac{1}{2}(\sigma_1 p + \sigma_2\phi(x)) \qquad \overline{Q} = \frac{1}{2}(\sigma_2 p - \sigma_1\phi(x)) \tag{1.5}$$

with σ_i the Pauli matrices and where the Hamiltonian

$$H = \frac{1}{2}(p^2 + \phi^2(x) + \hbar\sigma_3\phi'(x)) \tag{1.6}$$

acts on a two-component wavefunction. Momentum p is quantised canonically. The

function $\phi(x)$ is usually called the *superpotential* in these applications (it's not quite the superpotential of field theory) and is central to everything.

That importance is such that its superspace interpretation is worth returning to. A superfield

$$\Phi(t,\theta,\overline{\theta}) = q(t) + i\overline{\theta}\Psi(t) + i\theta\overline{\Psi}(t) + \theta\overline{\theta}B(t) \tag{1.7}$$

has a Lagrangian density

$$\mathcal{L} = \frac{1}{2}(D\Phi)(\overline{D}\Phi) - U(\Phi) \tag{1.8}$$

with derivatives

$$D = \frac{\partial}{\partial\overline{\theta}} - i\theta\frac{\partial}{\partial t} \qquad \overline{D} = \frac{\partial}{\partial\theta} - i\overline{\theta}\frac{\partial}{\partial t}. \tag{1.9}$$

$U(\Phi)$ is an arbitrary function with Taylor series expansion

$$U(\Phi) = U(0) + \Phi U'(0) + \ldots. \tag{1.10}$$

Integrating (1.8) over θ and $\overline{\theta}$ gives the actual Lagrangian and hence the Hamiltonian, from which the identification $U'(0) = \phi$ can be made.

Following Gendenshteĭn [2], the two-component Hamiltonian is fruitfully reinterpreted as two related Schrödinger problems

$$H_{\pm} = -\hbar^2\frac{d^2}{dx^2} + V_{\pm} \qquad (\sqrt{2m} = 1) \tag{1.11}$$

with potentials

$$V_{-} = \phi^2 - \hbar\phi' \tag{1.12a}$$

$$V_{+} = \phi^2 + \hbar\phi'. \tag{1.12b}$$

In handling the relations between these potentials it is useful to define generalised

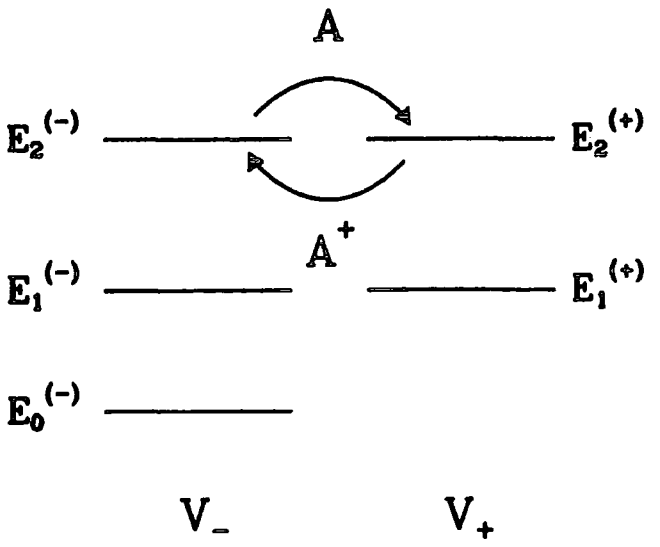

Figure 1.1 Partner Potentials

"raising and lowering" operators [3]

$$A = \hbar\frac{d}{dx} + \phi \qquad A^+ = -\hbar\frac{d}{dx} + \phi \tag{1.13}$$

in terms of which

$$H_- = A^+A \qquad H_+ = AA^+. \tag{1.14}$$

A simple and elegant pattern now emerges (Figure 1.1). If $\psi_n^{(-)}$ is the n-th eigenfunction of V_-, then $\psi_{n-1}^{(+)} = A\psi_n^{(-)}$ is the $(n-1)$-th eigenfunction of V_+ *with the same energy*, i.e.

$$E_n^{(-)} = E_{n-1}^{(+)}. \tag{1.15}$$

A^+ acts as the inverse operator to A. An exception to this pairing-up of the energy levels is the ground state of V_- which has no partner in V_+. However it is also special in that any potential which can be written as $\phi^2 - \hbar\phi'$ automatically has zero ground state energy; that it occurs here is merely a reflection of the characteristic absence of zero-point energies from supersymmetric theories. Many other features of this system have their origin in general SUSY results. The operators A and A^+ are closely related to the Q and $\overline{Q}$ which interrelate fermions and bosons, so one should perhaps think of there being a single potential, most of whose energy levels have a two-fold degeneracy corresponding to a fermionic degree of freedom. In this view A and A^+ change a boson into a fermion or vice versa without changing the energy.

This formalism evidently provides a quick method of generating nearly isospectral potentials: pick virtually (see below) any $\phi(x)$ and (1.12a) -(1.12b) give two such potentials. What is typically more useful, but much less easy, is to find the ϕ which gives a specified V_-, this involving the solution of the awkward Ricatti equation (1.12a). Actually this is equivalent to finding just the ground state wavefunction for the potential since

$$\phi(x) = -\hbar\left(\frac{\psi'_0}{\psi_0}\right) \tag{1.16}$$

$$\psi_0^{(-)}(x) = \exp\left(-\frac{1}{\hbar}\int \phi(x)dx\right). \tag{1.17}$$

The latter equation shows that the choice of ϕ cannot be quite arbitrary since $\psi_0^{(-)}$ must be normalisable and this entails that the superpotential be of predominantly odd parity in x as $|x| \to \infty$, a requirement very closely related to preventing the supersymmetry spontaneously breaking [1]. As a restriction this is not terribly severe and it is clearly satisfied by the ϕ of any previously allowed Schrödinger potential. Having found the ϕ for any particular potential there is always the possibility that the partner potential is easier to deal with – perhaps it has even been solved already. Although this will normally involve another Ricatti equation, V_+ will itself have another partner and so on; a complete solution to any of them will straightforwardly yield a solution to all the others. The only problem is to efficiently find the superpotential for any potential.

Before following these ideas in one special direction, it should be said that they have found uses in statistical, atomic and nuclear physics, often as a fresh interpretation of previously observed quantum mechanical regularities. In this sense supersymmetry is already a fact of nature. Details of these, overviews of the literature and fuller introductions to SUSY can be found in the reviews by Gendenshteĭn and Krive [4] and, at a slightly higher level, that by Lahiri, Roy and Bagchi [5]. Pedagogical introductions to the topics discussed in the rest of the chapter, including examples worked through in detail, can be found in the two articles by Dutt, Khare and Sukhatme [3][6].

1.2 Shape Invariance

One of the developments inspired by this way of treating Schrödinger's equation is an increased understanding of the simpler potentials whose spectrum can be solved for exactly. Contrary to the optimism conveyed in most elementary textbooks, potentials for which all the eigenvalues and eigenfunctions are known in closed form are actually rather scarce, with all of the simpler examples discovered within a few years of wave mechanics being proposed. The progressive and ongoing realisation that many of these have a common origin is therefore of some importance.

Again the key observation dates back to Gendenshteĭn's paper [2] which introduces the concept of a *shape invariant* potential. This is any $V_-(x, a_0)$ which depends on a finite set of parameters a_0 in such a way that the corresponding V_+ can be obtained by simply changing these parameters to a set a_1, i.e.

$$V_+(x, a_0) = V_-(x, a_1) + R(a_1) \tag{1.18}$$

where $R(a_1)$ is an arbitrary x-independent function. Or in terms of a compact notation for the superpotentials

$$\phi_0^2 + \hbar\phi'_0 = \phi_1^2 - \hbar\phi'_1 + R(a_1). \tag{1.19}$$

Clearly the ground state of $V_+(x, a_0)$ now has an energy $R(a_1)$, so this must be E_1 for $V_-(x, a_0)$. Provided the reparameterisation from a_0 to a_1 can be iterated, one can construct successive Hamiltonians each related by equations like (1.18) and straightforwardly prove by a continuation of this chain of argument that the spectrum of the original potential is

$$E_n(a_0) = \sum_{i=1}^{n} R(a_i). \tag{1.20}$$

By restricting oneself to potentials whose partners are essentially the same as the original, one has avoided having to partially solve a completely new problem each time a superpotential has to be found. The simplicity of the conclusion cannot be overemphasied: any solution $\phi(x, a_0)$ and $R(a_1)$ to (1.19) automatically has this

$\phi(x)$	Name	Reparameterisation
$\omega x - b$	Harmonic Oscillator	$\omega \to \omega$
$\omega x - \frac{(l+1)\hbar}{2x}$	3D Oscillator ($x > 0$)	$l \to l+1$
$\frac{e^2}{(l+1)\hbar} - \frac{(l+1)\hbar}{x}$	Coulomb ($x > 0$) (see [10])	$l \to l+1$
$A - Be^{-x}$	Morse	$A \to A - \hbar$
$A\tanh x + \frac{B}{A}$	Rosen-Morse	$A \to A - \hbar$
$A\tan x + \frac{B}{A}$	(see text)	$A \to A + \hbar$
$-A\coth x + \frac{B}{A}$	Eckart ($x > 0$)	$A \to A + \hbar$
$A\tanh x + B\,\mathrm{sech}\, x$		$A \to A - \hbar$
$A\coth x - B\,\mathrm{cosech}\, x$	($x > 0$)	$A \to A - \hbar$
$-A\cot x + B\csc x$	($0 < x < \pi$)	$A \to A + \hbar$
$A\tan x - B\cot x$	Pöschl-Teller I ($0 < x < \pi$)	$A \to A + \hbar, B \to B + \hbar$
$A\tanh x - B\coth x$	Pöschl-Teller II	$A \to A - \hbar, B \to B + \hbar$

Table 1.1. Shape Invariant Superpotentials

spectrum. Furthermore A^+ can be used to deduce the wavefunctions using

$$\psi_n(x, a_0) = N_0 A^+(x, a_0) A^+(x, a_1) \ldots A^+(x, a_{n-1}) \psi_0(x, a_n) \qquad (1.21)$$

where $\psi_0(x, a_n)$ is easily found using (1.17), and so a complete solution to the problem can be found [3].

How many solutions does (1.19) have ? Only the 12 listed in Table 1.1 have been found, but these contain all the very simple potentials known to have exact solutions. They are just those catalogued in more detail by Dutt et al [3], with the exception of $\phi(x) = A\tan x + B/A$, a trivial trigonometric generalisation of the Rosen-Morse potential independently pointed out in [7] and [8]. Like all the others, the reparameterisation involved is extremely simple. Searches have previously been made through rather large classes of potentials without finding any others, though any disappointment at this should be tempered, any exactly solvable potential which is not shape invariant being directly linked to a new infinite set of solvable partners using A and A^+ [9]. However solving one of these problems will usually involve a much more involved argument than that now applicable to those in Table 1.1.

In fact the concept of a reparameterisation symmetry encompassing this set of potentials was anticipated by Schrödinger's own *factorisation method* of 1940 [11]. Naturally this was not expressed in terms of a superpotential, but did require that any potential to which it was applied satisfied a condition equivalent to (1.19), except that the reparameterisations were restricted to a single one $a_1 = a_0 + 1$. A diligent search can normally reveal some sort of precedent for any innovation and Schrödinger's was itself only an independent rediscovery of Darboux's Theorem [12]. Development of the factorisation method largely died out after a review by Infeld and Hull [13] detailing the properties of the potentials in Table 1.1 showed that this exhausted the method's possibilities by considering polynomial dependences of ϕ on a_0. At first sight the restriction to a single parameter does not appear too significant since any simple reparameterisation $\{a_0\} \rightarrow \{a_1\}$ should be reexpressable as $a_0 = a_1 + 1$, e.g. the Pöschl-Teller potentials were thought of as one parameter potentials at that time. However, when shape invariance was introduced multiple parameters were explicitly allowed from the outset and this encourages the realisation that former ideas were possibly too restrictive. For instance one can imagine reparameterisations in which one parameter only changes on every second, or even on every prime, iteration and this cannot (at least self-evidently) be reduced to $a_0 = a_1 + 1$. Formulating all the theorems using this wider conception proved straightforward, although obviously no fully shape invariant examples have been found which actually require this extension. But this still necessitates finding an alternative to the Infeld-Hull exhaustion proof and the greater complexity now allowed to them suggests avoiding reparameterisation as its foundation and focussing on the x-dependence of ϕ instead.

In exploring the extent of shape invariance various ansätze have been proposed [2][7][9][13], all mainly inspired either empirically or by guesswork rather than any underlying theoretical principle. There is an unusual, previously unnoticed, potential which, as well as being interesting in its own right, indicates that new examples probably can't be generated simply by adding new parameters to existing ansätze. A heuristic way of thinking about (1.19) is that one must find ways of "hiding" the change of sign between the two sides and that this can really only be done by intermixing the ϕ^2 terms with the ϕ' ones. Thus all the ansätze are variations on letting $\phi' \simeq \phi^2$ in some loose sense, with sums of trigonometric functions, whose

derivatives can be equal to the cross-terms in the square, ideally suited to this. A 3-parameter version of this would be

$$\begin{aligned} \phi(x) &= Af(x) + Bg(x) + Ch(x) \\ g' &= fh \quad + \text{ cyclic perm.} \end{aligned} \tag{1.22}$$

Jacobi elliptic functions [14] are a generalisation of $\sin x$ and $\cos x$ to a triplet of functions $\mathrm{sn}x$, $\mathrm{cn}x$ and $\mathrm{dn}x$, hence the reason for considering

$$\phi(x) = A\frac{\mathrm{dn}x}{\mathrm{cn}x} + B(1-k^2)^{1/2}\frac{\mathrm{sn}x}{\mathrm{cn}x} + C(1-k^2)^{1/2}\frac{1}{\mathrm{cn}x} \tag{1.23}$$

where k is a parameter entering into the definition of these functions. As a potential this is a somewhat flat-bottomed well with infinitely high walls at finite x. Denoting two sets of parameters by $a_0 = \{A_0, B_0, C_0\}$ and $a_1 = \{A_1, B_1, C_1\}$, this satisfies (1.19) provided

$$\begin{aligned} A_1^2 + B_1^2 + C_1^2 &= A_0^2 + B_0^2 + C_0^2 \\ 2B_1C_1 - A_1 &= 2B_0C_0 + A_0 \\ &\quad +\text{cyclic perm.} \end{aligned} \tag{1.24}$$

One solution to this is

$$\begin{aligned} A_1 = A_0 - 1 \quad B_1 &= B_0 + 1 \quad C_1 = C_0 + 1 \\ -A_0 + B_0 + C_0 + \frac{3}{2} &= 0 \end{aligned} \tag{1.25}$$

and there are only two others, related to this one by cyclic permutations. The crucial point is that although ϕ_0 and ϕ_1 now satisfy (1.19), because

$$-A_1 + B_1 + C_1 + \frac{3}{2} \neq 0 \tag{1.26}$$

the reparameterisation cannot be repeated to find a ϕ_2 to satisfy (1.19) along with ϕ_1. The iteration necessary to prove (1.20) and to generate the complete solution has broken down.

This example exhibits a kind of *partial shape invariance.* Since only the first reparameterisation can be carried out, one can only find ($k' = (1-k^2)^{1/2}$)

$$\psi_0(a_0) = \left(\frac{1+\mathrm{sn}x}{\mathrm{cn}x}\right)^{-A_0}\left(\frac{k'+\mathrm{dn}x}{\mathrm{cn}x}\right)^{-B_0}\left(\frac{\mathrm{dn}x+k'\mathrm{sn}x}{\mathrm{cn}x}\right)^{-C_0} \tag{1.27}$$

$$\begin{aligned}\psi_1(a_0) = &\left(\frac{1+\mathrm{sn}x}{\mathrm{cn}x}\right)^{-A_0+1}\left(\frac{k'+\mathrm{dn}x}{\mathrm{cn}x}\right)^{-B_0+1}\left(\frac{\mathrm{dn}x+k'\mathrm{sn}x}{\mathrm{cn}x}\right)^{-C_0+1}\\ &\times\left((2A_0-1)\frac{\mathrm{dn}x}{\mathrm{cn}x}+(2B_0+1)k'\frac{\mathrm{sn}x}{\mathrm{cn}x}+(2C_0+1)k'\frac{1}{\mathrm{cn}x}\right)\end{aligned} \tag{1.28}$$

$$E_1(a_0) = 2k^2(A_0-B_0-1)+2B_0+1 \tag{1.29}$$

There are higher eigenfunctions, but these will only be accessible to other methods and their exact form is currently unknown. The above wavefunctions are legitimate provided the constraints

$$\begin{aligned}A_0+B_0+C_0 &> 0\\ 3A_0-B_0+C_0 &< 0\end{aligned} \tag{1.30}$$

imposed by normalisability are also satisfied. These formulae for ψ_1 and E_1 can be checked by substituting them into the actual Schrödinger equation; as expected this explicitly only works if A_0, B_0 and C_0 satisfy (1.25). Similar results hold for the other two solutions to (1.24).

Interest in this example ought not to centre on what we happen to learn about the solution – unlike complete ones, partial solutions to Schrödinger problems are very common [15] – and more on what it tells us about shape invariance. Clearly this is a counterexample to the Infeld-Hull proof, the loophole being that in some respects there are three parameters, in others only two. On the other hand, it appears that increasing the complexity of ϕ need not lead to any new fully shape invariant solutions (contrary to some earlier expectations [7][9]) and Table 1.1 could therefore be complete after all. A detailed examination of what is happening in this example shows that the number of terms an ansätz like (1.22) produces in (1.19) increases faster than the number of parameters, with each set of terms in (1.19) placing

constraints on these parameters. Extending (1.22) to four parameters is probably possible, although no set of functions generalising the Jacobi elliptic functions in the required way seem to have been defined so existence cannot be taken for granted. This solution would only just be possible and further generalisations seem very unlikely. It may be that a radically different ansätz could lead to new solutions, but there is now less cause for optimism.

1.3 The SWKB Series

One area newly invented as a result of SUSY QM is that of the supersymmetric WKB approximation, which is at least competative with its conventional forerunner, but also the basis for fresh insight into shape invariance. If the wavefunction is written as

$$\psi = e^{iS/\hbar} \tag{1.31}$$

and substituted into Schrödinger's equation, one finds that

$$S'^2 - i\hbar S'' + \phi^2 - \hbar\phi' = E \tag{1.32}$$

i.e.

$$S'^2 - i\hbar S'' + V_- = E. \tag{1.33}$$

The standard response is to expand S as a power series

$$S(x) = \sum_{n=0}^{\infty} (-i\hbar)^n S_n \tag{1.34}$$

and solve for the coefficients by collecting powers of $\hbar$. Depending on whether (1.32) or (1.33) is used, there are however two natural ways of treating the potential. In the original WKB approach V_- was regarded as a single $o(\hbar^0)$ object and an expansion derived accordingly. Once the superpotential is introduced as the fundamental function, however, (1.32) becomes a more sensible starting point with the potential now split into an $o(\hbar^0)$ term ϕ^2 and an $o(\hbar)$ one $-\hbar\phi'$ [16]. Although the series this leads to possesses certain new and important features, it is clearly related to the

WKB one and much of its development was modelled on that. Collecting powers gives the recurrence relation

$$S''_q = -\sum_{i=0}^{q+1} S'_i S'_{q-i+1}, \qquad q \geq 1$$
$$S'_0 = (E-\phi^2)^{1/2}, \qquad S'_1 = \frac{i\phi'}{2S'_0} + \frac{\phi\phi'}{2S'^2_0}. \tag{1.35}$$

As with the original [17], a quantisation condition can be derived

$$\int_a^b \sum_{n=0}^{\infty} (-i\hbar)^n S'_n dx = (n+1/2)\pi\hbar$$
$$\phi^2(a) = \phi^2(b) = E. \tag{1.36}$$

On calculating the $o(\hbar)$ term in this the first major difference emerges; to $o(\hbar)$ this condition is

$$\int_a^b (E-\phi^2)^{1/2} dx = n\pi\hbar \tag{1.37}$$

and hence it is trivially exact for the ground state ($a = b$, $n = 0$), in contrast to the WKB equivalent which almost invariably fails to reproduce even this.

One can go on to calculate corrections to this in powers of $\hbar$ [18]. Up to and including $\hbar^6$, this was first done by Adhikari et al [19] using a slightly different method than suggested by (1.35) – the known terms in the WKB quantisation condition were explicitly rearranged by expanding $V = \phi^2 - \hbar\phi'$. While this reinforces the connection between the series, it is as easy to start afresh using (1.35) and this method is more readily generalised. These terms have been independently recalculated in this manner – automated using a specially written FORTRAN program* –

* If this calculation were to be extended in future, the author would chose to do so using FORM or an equivalent. This alone indicates how quickly technology changes – the task proved beyond the capabilities of SMP.

to find [8]

$$\int_a^b (E-\phi^2)^{1/2}dx - \frac{\hbar^2 E}{6}\frac{d^2}{dE^2}\int_a^b \frac{\phi'^2}{(E-\phi^2)^{1/2}}dx$$

$$-\frac{\hbar^4 E}{128}\int_a^b (\frac{49E\phi'^4}{(E-\phi^2)^{11/2}} - \frac{140}{3}\frac{\phi'^4}{(E-\phi^2)^{9/2}} - 4\frac{\phi'\phi'''}{(E-\phi^2)^{7/2}})dx$$

$$+\frac{\hbar^6 E}{1024}\int_a^b (5005\frac{E^2\phi'^6}{(E-\phi^2)^{17/2}} - \frac{36036}{5}\frac{E\phi'^6}{(E-\phi^2)^{15/2}} + 2376\frac{\phi'^6}{(E-\phi^2)^{13/2}}$$

$$-924\frac{E\phi'^2\phi''^2}{(E-\phi^2)^{13/2}} + 720\frac{\phi'^2\phi''^2}{(E-\phi^2)^{11/2}} + 80\frac{\phi\phi''^3}{(E-\phi^2)^{11/2}} + 8\frac{\phi'''^2}{(E-\phi^2)^{9/2}})dx$$

$$+o(\hbar^8) = n\pi\hbar \tag{1.38}$$

in agreement with [19]. For convenience, this (rather than (1.34)) will be refered to as the *SWKB series.*

Note that the integration greatly simplified things, with repeated integration by parts causing numerous terms to combine or cancel. As will be explained later, the integration can be re-expressed as a contour one round a closed path which is why the endpoints can be safely discarded during this step and also why any term equal to an x-derivative will vanish. All the imaginary contributions naturally give zero, as can be demonstrated directly: if $S = A + iB$ then

$$B' = \frac{\hbar}{2}\frac{d}{dx}(\ln A') \tag{1.39}$$

which is zero after integration over x. Less obviously, all contributions with odd powers of $\hbar$ (apart from $o(\hbar)$, critically contributing the $\pi/2$ necessary to derive (1.37)) seem to vanish as well. For several simple ϕ this has been checked as far as $\hbar^{11}$ [20] and there is no reason to doubt its generality. No explanation for this is known, although heuristic ones will be discussed later.

How this new approximation compares to the WKB one has been the subject of extensive investigation, without any clear conclusion [3][5][19][20][21][22]. Aside

from its rather trivial success with the ground state and the important set of cases where (1.37) is exact which occupy the rest of the chapter, there is no particular reason to expect the SWKB series to be an improvement. In any application it is liable to suffer from the disadvantage that ϕ must be found or approximated as a first step. However it is clearly an advance to have two similar techniques so that results can be compared between them and our understanding improved.

1.4 Asymptotic Series

Is the expansion (1.38) convergent ? Almost certainly not, it being a long-established result that the WKB series is only asymptotic.* This is perhaps easiest to understand by modifying a famous argument from QED to be discussed in more detail in Chapter 4. Equation (1.32) is well-defined in the (semi-classical) limit $\hbar \to 0$, as is (1.31) since at large $|x|$ (1.32) ensures that S has the correct form so that ψ is normalisable. However if $\hbar < 0$, S remains the same in these regions, but the sign change in (1.31) means that ψ diverges and consequently the whole framework of the approximation breaks down. This indicates that it is non-analytic at $\hbar = 0$ and hence the series diverges. Since Chapter 4 will express reservations about this type of argument, it should be emphasised that the result can be rigorously proved by other means. In fact the approximation can be treated in a convergent fashion, but this approach is technically demanding and has yet to be extended to the SWKB case; the interested reader is refered to the Fromans' monograph [23].

With asymptotic series appearing throughout this thesis, it is worth briefly recalling their properties at the outset [24]. If a function $f(z)$ has a power series expansion $\sum r_n z^n$, then if $R_n(z)$ defined via

$$f(z) = r_0 + r_1 z + \ldots + r_{n-1} z^{n-1} + R_n(z) \tag{1.40}$$

has the property that, for each fixed n, $R_n(z)/z^n$ is bounded as $z \to 0$, the series is said to be an asymptotic expansion of $f(z)$. If so it can safely be added to or

* Unfortunately this thesis must deal with several distinct meanings of "asymptotic." A series may or may not be "asymptotic," but its coefficients will always have an "asymptotic" (i.e. large-order) behaviour. More confusingly, in field theory the "asymptotic" behaviour is likely to be that as $Q \to \infty$ and will refer to a function. These usages are so standard that it would be perverse to avoid them.

multiplied with another of the kind and it can also be integrated term-by-term, but possibly not so differentiated. An important case is when $r_n \sim n!$: then a maximum accuracy of $R(z) \sim (2\pi/z)^{1/2} e^{-1/z}$ is obtained by summing only the first $(1/z)e^{1/2z}$ terms. This can be extremely good provided z is small, but adding further corrections to this estimate only worsens it.

A feature of some significance in later chapters is that although $f(z)$ will have a unique expansion, $\sum r_n z^n$ does not have a unique function associated with it, if only because one can add perturbatively invisible terms like $e^{-1/z}$ to a function without changing the expansion. In certain circumstances, by placing additional constraints on what that function may be, a unique one can be identified and the best known of these methods will be defined in section 4.3.

1.5 Exact Quantisation Conditions

Even if there is little limitation in practical cases to the accuracy obtainable using corrections to it, those special problems where (1.37) gives the *exact* spectrum are of particular theoretical importance. This is especially so because of a result, originally proved by Dutt, Khare and Sukhatme [18], that this lowest-order SWKB approximation is exact when ϕ is shape invariant. Partly because that version obscured the central role of reparameterisation, partly because it was only proved to $o(\hbar)$, but mainly because of its importance to what follows we reproduce an alternative proof as presented in [8].

Truncating the expansion (1.38) at $o(\hbar^m)$ gives a quantisation condition of the form

$$F_m(a_0, E) = n\pi\hbar \tag{1.41}$$

obtained from a Schrödinger equation for V_-

$$S'^2 - i\hbar S'' + \phi_0^2 - \hbar\phi'_0 = E. \tag{1.42}$$

V_+ can be treated in the same way, but there are now two possible starting points

$$S'^2 - i\hbar S'' + \phi_0^2 + \hbar\phi'_0 = E \tag{1.43}$$

$$S'^2 - i\hbar S'' + \phi_1^2 - \hbar\phi'_1 = E - R(a_1). \tag{1.44}$$

(1.43) differs from (1.42) only with respect to a single change in sign. If the remaining odd powers of $\hbar$ vanish from the series as universally believed, this sign changes only effect on the derivation – (1.43) is (1.42) with $\hbar \to -\hbar$ and $i \to -i$, changes which leave the LHS of (1.38) unaltered – is that now the condition for V_+ is

$$F_m(a_0, E) = (n+1)\pi\hbar \tag{1.45}$$

so that to this order $E_n^{(-)}(a_0) = E_{n+1}^{(+)}(a_0)$. However, starting from (1.44) one derives a condition for V_+ of the form

$$F_m(a_1, E - R(a_1)) = n\pi\hbar. \tag{1.46}$$

For any particular value of E, these last two equations are asymptotic expansions in $\hbar^2$ of the same exact quantisation condition and they must therefore be equivalent. Thus, even after truncating at $o(\hbar^m)$, they must give the same (as yet not necessarily correct) spectrum for V_+. Inverting (1.41) gives a function

$$\begin{aligned} E_n(a_0) &= \epsilon_m(a_0, n) \\ \epsilon_m(a_0, 0) &= 0, \qquad \forall a_0 \end{aligned} \tag{1.47}$$

while doing the same to (1.46) gives

$$E_n(a_1) = \epsilon_m(a_1, n) + R(a_1). \tag{1.48}$$

Now

$$\epsilon_m(a_0, 1) = E_1(a_0) = E_0(a_1) = R(a_1). \tag{1.49}$$

Similarily, but also using reparameterisation

$$E_2(a_0) = E_1(a_2) = \epsilon_m(a_1, 1) + R(a_1) \tag{1.50}$$

$$= R(a_2) + R(a_1). \tag{1.51}$$

Inductively, the result of truncating at $o(\hbar^m)$ is a spectrum

$$E_n(a_0) = \sum_{i=1}^{n} R(a_i) \tag{1.52}$$

in agreement with (1.20). Returning to the actual SWKB series (1.38), this means that for a shape invariant ϕ all the $o(\hbar^2)$ and higher corrections will vanish order-by-order.

Note that although reparameterisation only enters at one point its role is crucial. Thus the partial shape invariance of (1.23) is too limited to enable the proof to work except for the first excited state which should be given exactly.

The next natural question is what is the converse ? For which ϕ is (1.37) exact ? To date, direct investigation has revealed no non-shape-invariant potentials for which this happens. That it may be a necessary condition for the SWKB corrections to be zero was first conjectured by Khare and Varshni [22]; evidence for the stronger result that these are only zero for the ϕ in Table 1.1 and hence that these are the only possible shape invariant potential was advanced in [8] and the rest of this chapter will be taken up with reinforcing this case.

Before discussing details it is worth reviewing what is known about the parallel case of the WKB series so as to discriminate between certain types of claim. One can obviously prove that the lowest-order result is exact in any special case if enough if enough is known to enable that result to be directly compared to the previously established correct spectrum, as the proof above does. A few such results are known for the WKB condition, most notably for the case of the harmonic oscillator where the additional result that all the corrections explicitly vanish can be proved [25]. However the most rigourous approach is to use the convergent version of the theory mentioned above and this has been done for all 8 potentials (some shape invariant, some not) believed to be exact [26]. While undeniably respectable, this need not be terribly enlightening: $V = Ae^{2ax} + Be^{-2ax}$ is known to be one, but no-one yet knows its exact spectrum. Where this approach is necessary is in ruling out potentials which have zero corrections, but for which the lowest-order condition gives the wrong spectrum. These can be constructed [26] by taking one of the exact solutions and then modifying the potential outside the classical turning points; since all WKB integrals are taken between these points this action neither alters the corrections nor the WKB spectrum, but it does change the correct answer. The loophole is that the energy levels shift by terms like $e^{-1/\hbar}$ which are perturbatively invisible. Note that this is only possible when there are a finite number of eigenvalues.

Here the criterion of "exactness" will be that the corrections are zero for arbitrary E, this last point ruling out such adjustments. Although the proof above merely proves that they vanish when $E = E_n$, it is shown below that all the poten-

tials in Table 1.1 satisfy this more stringent property. While there are cases where the above modifications can be made, it seems highly unlikely that any such ploy could preserve shape invariance. There is at least one case where this criterion is too stringent: for (1.23) the precise value of E_1 must enter in a special way. As any similar case must be equally special, these are ignored.

The central observation of this chapter is that the first few corrections are enough to restrict the set of potentials satisfying this criterion to those in Table 1.1 [8]. Taking the $o(\hbar^2)$ correction from (1.38), it can be rewritten as

$$\frac{d^2}{dE^2}\int_{-1}^{1}\frac{\phi'(\sqrt{E}u)}{(1-u^2)^{1/2}}du \tag{1.53}$$

where a change of variable from x to $u = \phi/\sqrt{E}$ has been made.* This can only be zero if $\phi'(\phi)$ has an even part which is no more than quadratic, i.e.

$$\phi' = a + b\phi^2 + \phi g(\phi) \tag{1.54}$$

where $g(\phi)$ is an even function. This is a necessary condition for any shape invariant ϕ. In conjunction with the definition of shape invariance slightly more is learnt about $g(\phi)$. Since normalisation requires that $\phi'(x) > 0$ as $|\phi| \to \infty$, the odd terms contained in g cannot be allowed to become too dominant at large ϕ, i.e.

$$\lim_{\phi\to\infty}\frac{g^2(\phi)}{\phi^2} \leq \text{constant}. \tag{1.55}$$

Unsurprisingly, requiring higher-orders in (1.38) to be zero places further restrictions on $g(\phi)$, but before pursuing that avenue it may help to understand the form taken in the known shape invariant cases. Remarkably, the 12 examples reduce to

$$g^2 = \begin{cases} ca \\ c(a + b\phi^2) \end{cases} \tag{1.56}$$

where a and b are exactly as in (1.54). Not only do they all satisfy this equation, integrating it up labouriously reproduces Table 1.1 while giving no additional solutions. Using it to evaluate the lowest-order quantisation condition also yields all

* This is only possible if $\phi'(\phi)$ has some odd component, but this is ensured by normalisability. The integrations with respect to E must be done first and are legitimate; although the limits are E-dependent their contributions cancel since $\phi^2(a) = \phi^2(b)$.

the correct spectra. A natural conjecture is therefore that (1.56) is a necessary and sufficient condition for shape invariance.

Whatever the truth of this, it certainly allows a simple proof that all the corrections in (1.38) are zero if ϕ appears in Table 1.1. A result of this sort was first proved by Raghunathan et al [27] for the Rosen-Morse potential, but all the examples can now be handled at once. Note that their conclusion that their argument generalises to all "solvable potentials," a wider class than that of all shape invariant ones, has been disproved [28]. The unique property of (1.56) is that differentiating $(n-1)$ times with respect to ϕ gives

$$\phi^{(n)}(\phi) = p(n+1) + p(n)g(\phi) \tag{1.57}$$

where $p(m)$ is an mth order, purely odd or even polynomial in ϕ, the details of which are irrelevant for the moment. Under multiplication

$$\phi^{(n)}\phi^{(m)} = p(n+m+2) + p(n+m+1)g(\phi). \tag{1.58}$$

Straightforward inspection of (1.35) and (1.38) shows that any term appearing in any correction has a restricted combination of factors and derivatives which the above formulae reduce to polynomials in ϕ of restricted power. Repeating the set of steps used to derive (1.53) then demonstrates that each of these terms is individually zero. By considering a few special terms in (1.38), the converse can be proved: terms in any correction are only seperately zero if ϕ satisfies (1.56).

This is important, but unnecessarily strong. For the lowest-order quantisation condition to be exact in the sense adopted here corrections are required to be zero order-by-order, not term-by-term. The problem thus reduces to that of considering conspiracies between terms in, say, $o(\hbar^m)$ which are each non-zero, but which manage to cancel.

1.6 Higher-Order Corrections

By examining the first two remaining corrections progress can be made in eliminating any possibility of conspiracy. Unfortunately, the corrections rapidly become complicated and the formulae involved in discussing when they are zero even more

so, hence the considerations in this section may appear unnecessarily labyrinthine, particularly since the conjectured answer is so simple. Although this attitude is probably justified – and the next section will describe some preliminary thoughts about an alternative motivated by just these frustrations – no economical approach to the problem is now known and so the only strategy is essentially to consider each possible conspiracy in turn.

Substituting (1.54) into the $o(\hbar^4)$ correction expressed as an integral over ϕ yields

$$\int d\phi \left[147E\frac{(a+b\phi^2)\phi^2 g^2}{(E-\phi^2)^{11/2}} - 140\frac{(a+b\phi^2)\phi^2 g^2}{(E-\phi^2)^{9/2}} - \frac{4}{(E-\phi^2)^{7/2}}\Big(g^2(a+7b\phi^2) \right.$$
$$\left. +(a+b\phi^2)\phi^2(2gg''+g'^2) + 2gg'\phi(3a+5b\phi^2)\Big)\right] \qquad (1.59)$$

as the non-trivial content ($g' \equiv dg/d\phi$ etc.), but this is too complicated to enable a condition on g to be spotted in the straightforward way it was with the $o(\hbar^2)$ equivalent. Thus the strategy chosen is to expand $g^2(\phi)$ as a power series in ϕ

$$g^2(\phi) = \beta_0 + \beta_1\phi^2 + \beta_2\phi^4 + \ldots \qquad (1.60)$$

and use the corrections to place constraints on the coefficients of this expansion. Making the substitution

$$\beta_n \to ab\left(\frac{b}{a}\right)^n \beta_n \qquad (1.61)$$

greatly simplifies the algebra in that it eliminates a and b from all the formulae. Since the reverse substitution easily reinstates them, the explicit dependence on a and b is dropped from here on. Deriving the constraints first involves integrating all terms in (1.59) with respect to E enough times (here five) so that the lowest power of S'_0 appearing is $(E-\phi^2)^{-1/2}$, then expanding all the g that appear. In doing so, one has to contend with the likes of $(dg/d\phi)^2$, which in this case is most conveniently dealt with by defining the expansion

$$\left(\frac{d(\phi g)}{d\phi}\right)^2 = \epsilon_0 + \epsilon_1\phi^2 + \epsilon_2\phi^4 + \ldots \qquad (1.62)$$

whose coefficients are related to the β_n by

$$\epsilon_n = (n+1)\beta_n + \frac{1}{\beta_0}\sum_{i=1}^{n-1}\{(i+1)(n-i+1)\beta_i - \epsilon_i\}\beta_{n-i}. \qquad (1.63)$$

Other combinations of the $g(\phi)$ and its derivatives are expressible in simple ways using just β_n and ϵ_n. One now has an expression mixing $(E-\phi^2)^{1/2}$ with series in ϕ^2. The change of variables used in the previous section, $\phi = \sqrt{E}u$, is repeated and then powers in E are collected. Each power has a coefficient which is now a set of simple integrals over u tied up with involved functions of the β_n. The integrals and then finally the derivatives with respect to E are done, eliminating the first few terms, but leaving a power series in E.

Formally at least, this power series allows the correction to be evaluated given ϕ' as a power series in ϕ, but for it to be useful in this way raises difficult questions concerning the convergence of both (1.60) and the series in E. While thinking of the series thus will be useful shortly, this is usually unnecessarily ambitious and we need only be interested in it as a formal expansion. To be precise, even if any of the series involved were to be badly divergent, the coefficients ought still to obey the derived formulae. Thus if some ϕ corresponding to (1.60) has a zero $o(\hbar^4)$ correction for all E, then all the coefficients of the power series in E must be zero. The point is that these zero coefficients are just functions of the β_n and hence can be used to derive necessary conditions on the expansion (1.60) of a shape invariant $g(\phi)$.

These condition are most conveniently expressed as a recurrence relation for the coefficients of a $g(\phi)$ giving a zero correction:

$$(2n-1)(2-5n)\beta_{n-1} + (-10n^2+33n-26)\beta_{n-2} + 6(\epsilon_{n-1}+\epsilon_{n-2}) = 0, \quad n > 3 \quad (1.63)$$

where the equation for each n derives from a separate power of E. As a simple cross-check, the form (1.56) (which has $\beta_n = 0$ and $\epsilon_n = -\epsilon_{n-1}$ for $n \geq 2$) clearly satisfies it. Note that β_0, β_1 and β_2 are not fixed by these equations, but all subsequent coefficients are uniquely determined in terms of them, so the set of all g having zero $o(\hbar^4)$ corrections can be parameterised by $g(\beta_0, \beta_1, \beta_2, \phi)$. By itself, this equation tells us very little – even the large n limit is not clearly understood and (in this

context) generating the series expansions doesn't really help us to get closer to these candidates for shape invariant potentials as functions. However, the main idea is evidently working and the set of functions which might have zero corrections throughout the SWKB series has already been drastically reduced. Thus encouraged the obvious next step is to repeat this exercise for the $o(\hbar^6)$ correction.

In terms of the manipulations involved the details are exactly the same as those for $o(\hbar^4)$, except that everything is a order of magnitude more complicated. This calculation was carried out by hand, but with substantial pieces of it checked using the Symbolic Manipulation Program (SMP). Details of the lengthy recurrence relation obtained are relegated to Appendix 1; the major point to register is that it gives the coefficients from β_5 onwards in terms of β_0, β_1, β_2, β_3 and β_4. The main remaining issue is whether any of the series so generated are consistent with any of those found from the previous correction.

However, at this point it may legitimately be asked how reliable, even with the benefit of SMP's involvement, are (1.63) and the formula in Appendix 1 ? For this reason a numerical check on them has been devised using the interpretation of their derivation as essentially calculating the corrections for an arbitrary $g(\phi)$. It relies on the fact that the coefficient of a power of E in the series for this answer only involves some of the β_n, e.g. in the $o(\hbar^4)$ case the E^{n+2} term contains contributions from β_0, β_1 ... β_{n+2} only. Previously this coefficient would be set to zero and thus a recurrence relation for β_{n+2} in terms of β_0, β_1 ... β_{n+1} obtained, but now we are interested in using it when the correction is non-zero. Suppose one picks some largish integer N and arbitrary values for β_0 and β_1, then uses the recurrence relation to calculate a series

$$g^{(N)}(\phi^2) = \beta_0 + \beta_1\phi^2 + \ldots + \beta_N\phi^{2N}. \tag{1.64}$$

If $N \to \infty$ this gives a series which when inserted in integral (1.59) should (passing over questions of convergence for the moment) produce zero overall, but which for finite values of N gives

$$I_1(g^{(N)}, E) = E^{N+1}\alpha_{N+1} + E^{N+2}\alpha_{N+2} + \ldots \tag{1.65}$$

for this integral – powers of E^N and lower having been set to zero by picking the

special values of β_n for $n \leq N$. In principle, $I_1(g^{(n)}, E)$ can be evaluated numerically for several (low) values of E and α_{N+1} estimated in this way. It can also be calculated algebraically and the relation between $\alpha_{N+1}(\beta_0, \beta_1 \ldots \beta_{N+1})$ and (1.63) is so close that agreement between the two versions constitutes a stringent check on the latter.

In practise, difficulties arise because of the non-integrable singularities in the denominator of (1.59). As is standard, these can be softened using integration with respect to E, leaving the correction in a form akin to (1.53), with an integral that must be evaluated for arbitrary E and then differentiated several times. From the present point of view this is unfortunate since numerically this new integral is dominated by the low powers of E eliminated by the differentiation and whose coefficients thus have little to do with (1.63), i.e. one now has

$$\begin{aligned} I_2(g^{(N)}, E) =& \alpha'_0 + E\alpha'_1 + E^2\alpha'_2 + E^3\alpha'_3 + E^4\alpha'_4 \\ &+ E^{N+1}\alpha'_{N+1} + E^{N+2}\alpha'_{N+2} + \ldots \end{aligned} \tag{1.66}$$

where these α'_n are simply related to α_n by factors involving n alone. A naïve solution would be to calculate the integral and then differentiate numerically, but this is too cumbersome and it is much easier to calculate

$$I_2(g^{(N)}, E) - I_2(g^{(M)}, E), \qquad M \gg N \tag{1.67}$$

in which these initial terms cancel. By using different values of N the recurrence relation can be thoroughly checked. At this point it is probably sufficient to say that the method succeeded in detecting several small errors in earlier versions of the algebra, mainly in connection with the $o(\hbar^6)$ correction. The final agreement was such as to dismiss any worries about convergence in this context.

A numerical search was instituted to search for consistent solutions of both recurrence relations, i.e. to find how many of the $g(\beta_0, \beta_1, \beta_2, \phi)$ given by (1.63) also satisfy the equation in Appendix 1 ? For any β_0, β_1 and β_2 the $o(\hbar^4)$ relation can give β_3, β_4, β_5, β_6, β_7 and β_8, then β_0, β_1, β_2, β_3 and β_4 can be inserted into the $o(\hbar^6)$ one to find β'_5, β'_6, β'_7 and β'_8. The question is are there non-trivial sets of

β_0, β_1 and β_2 such that

$$\beta_5 = \beta'_5, \quad \beta_6 = \beta'_6, \quad \beta_7 = \beta'_7, \quad \beta_8 = \beta'_8 \tag{1.68}$$

and this is equivalent to solving a (complicated) set of four non-linear equations in three variables. As such it is an awkward problem where the best hope of an efficient method is normally the straightforward 3D Newton-Raphson one [29], using "deflation" (placing artificial poles on top of known solutions) to prevent the iteration retreading explored regions. It was chosen to run the program as a search through a set of the three simplest equations with any solution of these being checked to see if $\beta_8 = \beta'_8$. During execution it relatively quickly located 12 solutions to the restricted set of equations, all of which then failed the last test, the program finally running on for 132 hours. Most of these candidates were clustered around the origin in $(\beta_0, \beta_1, \beta_2)$ space and an examination of the form of the functions involved shows them to have a lot of structure in this region, but to be featureless elsewhere. Obviously in this sort of situation there is always the danger that the search procedure is inefficient or is being misled and so no absolute assurance can be given that other solutions to the restricted set of equations do not exist. However these results do make the possibility that there are other ϕ with zero SWKB corrections very unlikely. Note that even if another solution were to be found it not only has to give $\beta_8 = \beta'_8$, but it also has to satisfy the infinite set of similar equations imposed by requiring that the recurrence relations agree to all orders in ϕ^2. And this is without worrying about the $o(\hbar^8)$ and higher corrections.

Therefore it is with some confidence that we can conclude that only superpotentials satisfying (1.56) can give zero SWKB corrections throughout the series. That equation is then both a sufficient and a necessary condition for shape invariance.

1.7 Towards An Analytic Proof

Two questions concerning the SWKB series are still in search of an elegant and insightful proof, assuming current conjectures to be true: why every second correction automatically vanishes and why (1.56) should be a necessary condition for the remaining ones to be zero ? Possible answers which connect both are the subject of this section. As will become clear this is no more than a sketch of a suggested solution and so on certain points rigour will not be pursued in detail.

What can be proved is that both statements are true in a large class of possible ϕ – those for which $\phi'(\phi)$ is only non-analytic along the real axis in the ϕ plane. This can be done by generalising the basic idea underlying the proof by Raghunathan et al [27], using the SWKB quantisation condition expressed as a contour integral, that the Rosen-Morse potential has no higher-order corrections. In the complex x plane the integral in (1.38) was originally the contour one running around the cut from a to b due to $(E-\phi^2)^{1/2}$ appearing in all the terms of the series [17]. By considering the ϕ dependence of the correction terms expressed as contour integrals over ϕ, they showed that all the integrands fell sufficiently quickly as the contour was expanded to infinity for all these terms to be zero. Crucially, the operation of expanding the contour can only succeed if there is no extra structure in outer reaches of the complex plane for it to get entangled in, a point that can be verified directly in the particular case they considered.

The argument here is that the same procedure can isolate a single term in the $o(\hbar^4)$ correction, thereby greatly simplifying the problem. Applying it to (1.59), one can prove that as $|\phi| \to \infty$, the integrand approaches

$$E\frac{\phi^4 g^2}{i\phi^{11}} + o(E^0), \tag{1.69}$$

this limit suppressing the E dependence arising from the denominators. Note that if this limit is possible (i.e. no obstructing poles or cuts in g^2), then no information is lost in taking it. If the correction is still to be zero for all E, then in this form we can see that this is only true provided

$$\oint_C \frac{g^2}{\phi^7} d\phi = 0 \tag{1.70}$$

where C is the contour at infinity. There are then several ways this can be exploited depending on how rigorous one wants to be, but perhaps the most honest is to reverse the expansion of the contour and recover

$$\int_{-\sqrt{E}}^{\sqrt{E}} \frac{(a+b\phi^2)g^2\phi^2}{(E-\phi^2)^{11/2}} d\phi, \tag{1.71}$$

a line integral which must be equal to (1.70), excluding irrelevant constants. By the

same line of argument as in section 1.5, this is only zero if

$$g^2 = \beta_0 + \beta_1\phi^2 + \beta_2\phi^4. \tag{1.72}$$

But this is inconsistent with (1.63) unless $\beta_1 = 1$ and $\beta_2 = 0$ which is just the known case (1.56).

A similar style of proof works for the terms with odd powers of $\hbar$, which as mentioned in section 1.3 appear to vanish for all ϕ. The WKB series displays a similar pattern as various proofs show [30][31], but almost invariably (see [25] for the interesting exception) these reduce to the observation that the quantisation condition cannot contain complex terms. Such a route is not available here, where $\hbar$ and i are no longer quite so closely related, but instead one can point out that in vanishing these terms preserve order-by-order the degeneracy between the spectra of V_- and V_+ that would otherwise be broken. This is extremely suggestive of the field theory result that if unbroken at tree-level, supersymmetry is likewise preserved in perturbation theory to all orders in $\hbar$ [32].* Unfortunately, the proof is an order-by-order result which proceeds at a Feynman diagram level and so does not help in finding an equivalent here. However the following approach is promising: let $S = A + \hbar B$ separate the odd and even terms in $\hbar$, then

$$\oint B dx = \oint \left(\frac{i}{2}\frac{d}{dx}(\ln A) + \frac{\phi'}{2A} \right) dx \tag{1.73}$$

has a first term which vanishes in the contour integral. To lowest order the other term gives the $o(\hbar)$ constant. This leaves

$$\oint \left(\frac{1}{A} - \frac{1}{A_0} \right) d\phi = \oint \frac{i\hbar^2 B' - \hbar^2 B^2}{AA_0(A + A_0)} d\phi \tag{1.74}$$

where $A_0 = (E - \phi^2)^{1/2}$, and in expanding the contour as before

$$A \to i\phi, \qquad B \to \frac{1}{\phi} \tag{1.75}$$

since

$$A^2 + \hbar^2 B - i\hbar^2 B' + \phi^2 = E$$

* In fact it was in response to this that Witten [1] considered models where SUSY is dynamically broken by (arbitrarily small) non-perturbative effects and was thus led to (1.6).

$$2AB - iA' - \phi' = 0 \tag{1.76}$$

so this integral becomes

$$\sim \oint \frac{d\phi}{\phi^3} = 0. \tag{1.77}$$

This handles all orders at once, but $1/A$ can be expanded as a series in $\hbar$ and thus one again sees that the deformation of the contour is only legitimate if $\phi'(\phi)$ puts no poles or cuts in the way. Note that the structures of $\phi''(\phi) = \phi'.d\phi'/d\phi$ and the higher derivatives are closely related to that of $\phi'(\phi)$, so we need only dicuss that.

At first sight the restriction to superpotentials where $\phi'(\phi)$ is analytic away from the real axis appears too severe and the above proofs therefore a dead end. However this is to overlook a host of constraints on the superpotential which could conceviably guarentee just such an outcome. For a start no cut on the real axis can extend to infinity. This is partly a consequence of normalisability – $\phi(x)$ may have turning points, between which $\phi'(\phi)$ will be multiply valued, but there is always a even number of them and at large $|x|$ the inverse is unambiguous, so $\phi'(\phi)$ is single-valued at large ϕ – and partly that of $\phi(x)$ being single-valued for real x. The latter point is usually only implicit, but it is necessary for ψ to be single-valued, as it must be on physical grounds. Any confined non-analyticity on the real axis is avoidable by taking a sufficiently large E such that the contour in, say, (1.73) encloses it to begin with.

Analytically continuing a real, single-valued $\phi(x)$ away from the real axis is a very well-known situation, the main feature of which is that $\phi(\overline{z}) = \overline{\phi(z)}$. Equally well-known is that this can give rise to poles, but these need not matter in $\phi'(\phi)$ where they appear at infinity. The circumstances concerning cuts is less clear; for instance, superficially one could have

$$\phi'(\phi) = \sqrt{1 + i\phi} + \sqrt{1 - i\phi} \tag{1.78}$$

except that this violates the implicit condition on $V(x)$ that it is defined classically and it is therefore somehow unnatural that that definition should inherently involve complex numbers. There is also (1.55) to be reproduced. Hence it is not quite so nonsensical to suggest that for any "sensible" superpotential, $\phi'(\phi)$ be analytic

except on part of the real axis. Furthermore, any of the additional criteria could explain why this type of argument needs only the $o(\hbar^4)$ correction to succeed, as opposed to that in the previous section which had to be augmented by the $o(\hbar^6)$ one – presumably all the series generated by (1.63) correspond to functions violating at least one of these criteria. If so, it would become certain that shape invariance is a property restricted to those superpotentials in Table 1.1.

Finally, there is the obvious question of which potentials does the original WKB approximation have zero corrections for. Apart from the case-by-case studies mentioned earlier [25][26], this has never been seriously addressed and it is perhaps surprising that greater progress has been made with the generally more complicated SUSY version. Unfortunately, setting $V = \varphi^2$ and finding $\varphi'(\varphi)$ for the known exact cases [26] produces only odd functions of φ, making the key change of variable from x to φ inadmissible and so halting the enterprise at the outset. That this did not happen before was a direct consequence of normalisability again and this in turn is intimately related to the preservation of SUSY in the model [1]. It may yet transpire that supersymmetry's most permanent legacy is a deeper understanding of the Schrödinger equation.

CHAPTER TWO

Massless Renormalisable Field Theory

2.1 Dimensional Transmutation

Theories in which we have some degree of confidence are normally investigated by trying to measure the unpredicted fundamental constants. For the moment suppose there is only one of these, a dimensionless coupling g. Actual predictions will either be of the type: if a dimensionless observable R (in some way dependent on g) has one value then the observable σ has another; or concern energy dependences. For instance it may be anticipated, simply from dimensional analysis [33], that R will satisfy

$$Q\frac{dR}{dQ} = -b\rho(R) \tag{2.1}$$

where Q is some external energy, ρ a function (in principle) given completely by the theory and $-b$ a product of malice aforethought, regardless of whatever value R might actually have at a particular energy, this latter needing a cooperative experimentalist before anything can be said about it. Integrating up this equation

$$\log Q + \text{constant} = -\int_{\infty}^{R} \frac{dx}{\rho(x)} = f(R(Q)) \tag{2.2}$$

where f, like ρ, is a function given to us (in principle) completely by theory. It is the constant here that is the minimal requirement from experiment. This can be done in two logically equivalent ways: either we can measure R at, say, $Q = 10$ GeV or we can find the energy Q at which, say, $R = 3.1415$. Conventionally, it is the latter that is done, albeit indirectly, by letting $Q = \Lambda$ be the energy at which

$$f(R(Q = \Lambda)) = 0 \tag{2.3}$$

so that

$$R(Q) = f^{-1}(\log(Q/\Lambda)) \tag{2.4}$$

is the full prediction of the theory which now allows us to anticipate what values R

will take at other energies. In ideal practice we would calculate f in closed form and then measure R at different energies to determine that at which (2.4) is satisfied and so find Λ in GeV. Measurable Λ is the only unknown parameter of the theory and as such replaces the coupling as the fundamental constant.

Now an innocent remark with large, awkward consequences. What if things had been done the other way round ? That would have introduced an energy $Q = \mu$, to be called the *renormalisation scale*, as that at which we chose to measure R and although

$$R(\mu) = f^{-1}(\log(\mu/\Lambda)) \tag{2.5}$$

depends on this choice,

$$R(Q) = f^{-1}(f(R(\mu)) + \log(Q/\mu)). \tag{2.6}$$

cannot, i.e. this formula is true for arbitrary μ. Similarily, neither can any other observable $\sigma(Q)$

Because all the equations above are dimensionless overall and the underlying theory could have been defined entirely in terms of massless quantities, it is perhaps surprising that the dimensionless parameter g has been replaced by a massive one Λ. In the process by which this happens, known as *dimensional transmutation*, the arbitrary scale μ plays a key role. Note that instead of materialising from nowhere, massive quantities have entered the the theory because it is being "probed" from outside by an external Q – for instance the 91 GeV centre of mass energy fed into collisions at LEP.

Of course to restrict ourselves to one observable is far too limiting and any other $\sigma(Q)$ is now given by some predicted function

$$\sigma(Q) = \sigma(Q/\mu, R(\mu)) \tag{2.7}$$

directly. For this, the condition that, contrary to appearances, σ can't depend on the choice of μ either,

$$\mu\frac{d\sigma}{d\mu} = 0 \tag{2.8}$$

can be expanded into

$$\left(\mu\frac{\partial}{\partial\mu} - \rho(R(\mu))\frac{\partial}{\partial x}\right)\sigma(Q/\mu, x)|_{x=R(\mu)} = 0. \tag{2.9}$$

Although this is all rather simple, we begin to see $\rho(x)$ emerge in a more important role.

Naturally physics is also independent of whichever observable R we choose to use as a reference and it is particularly easy to convert to another, say σ, for which instead of (2.1) we now have

$$Q\frac{d\sigma}{dQ} = -\rho_\sigma(\sigma(Q)). \tag{2.10}$$

Trivially

$$\rho_\sigma(\sigma) = \frac{d\sigma}{dR}\rho(R). \tag{2.11}$$

However, to satisfy the equivalent of (2.3) for σ will usually involve a different value of Λ, but numerically this can be compensated by a change of units altering the particular value of μ we have in mind. Conversely, changes in μ keeping Λ constant correspond to a change of reference quantity. Crucially, as is clear from dimensional analysis, changes in μ can also be compensated by changes in Q. Thus

$$\sigma(zQ, R(\mu), \mu) = \sigma(Q, R(\mu), \mu z^{-1}). \tag{2.12}$$

Using invariance under change from μ to $\mu' = \mu z^{-1}$, the right hand side becomes such that

$$\sigma(zQ, R(\mu), \mu) = \sigma(Q, R(\mu z), \mu). \tag{2.13}$$

Setting $Q = \mu$ (a new conceptual ingredient perhaps, but quite in keeping with the arbitrariness of μ) on the right and renaming $Qz = W$, we conclude

$$\sigma(W, R(\mu), \mu) = \sigma(\mu, R(W), \mu). \tag{2.14}$$

Hence the energy dependence of σ can be transferred from the function onto one of its arguments.

Thus far all functions have been considered as definite and given rather than as unknowns requiring calculation in some model. Between μ-invariance and (2.14), individual functions have turned out to be expressable in a variety of different, yet fundamentally equivalent, forms. However any attempt to calculate interesting functions invariably has to resort to some kind of approximation and it is unlikely that this will treat all these versions on an equal footing. This can be a powerful advantage. For instance, if $R(\mu)$ is known or believed to be small it is tempting to expand $\sigma(Q/\mu, R(\mu))$ as a perturbation series in powers of it, but it may be better to expand $\sigma(1, R(Q))$, particularly if one is interested in the high-energy properties of a theory in which $R(Q)$ decreases with energy. Or it may be that a special choice of μ gives a series with especially good convergence.

This is a freedom for which there is a price.

2.2 Inside A Quantum Field Theory

Simple dimensional analysis sufficed to derive (2.9) with unquestionable validity because σ itself was explicitly dimensionless and an observable. But should we choose to investigate an unmeasurable quantity Γ, perhaps as part of an intermediate step in the calculation of a measurable one, such considerations of what happens when we change units or renormalisation point are no longer a constraint. Actually, if Γ is unphysical we could a priori allow it to have virtually any properties we wish, violating any symmetry of nature. None of this matters so long as the calculational rules of the theory are respected and the observables are legitimate. However, faced with this potential anarchy we are forced to confront at least some of the details in an example of a massless, renormalisable theory. The simplest instance is that of a scalar field ϕ defined by the Lagrangian

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - g\phi^4. \tag{2.15}$$

That the quantum world is all that is the case (although weighted by the exponential of the classical action) was an observation which led Feynman to his "sum-over-histories" approach and path integral quantisation, the means of handling quantum fields to be adopted here. Standard presentations of this method are readily accessable [34] * and we do no more than outline the general framework.

* For an alternative, a discussion of canonical quantisation can be found in [35].

From the Lagrangian the classical action is

$$S[\phi] = \int d^4x \mathcal{L} \tag{2.16}$$

and this can be used to find the generating functional

$$Z[J] = \mathcal{M} \int [d\phi] \exp\{iS[\phi] + \int d^4x J(x)\phi(x)\} \tag{2.17}$$

where $Z[0] = 1$ and $J(x)$ is an arbitrary function.

Most difficulties of rigour arise in trying to define and calculate this path integral, the integrand of which is an entirely classical expression, but with the integration over all space-time configurations of the field. An immediate problem is that the integrand is oscillatory and hence not obviously convergent, so conventionally a Wick rotation is made from Minkowski to Euclidean space, the integral evaluated there and then continued back, with the hope that this is a sensible procedure. A more serious difficulty is that of evaluating the integral at all, in all but the most trivial of cases. Two approaches are popular: splitting $\mathcal{L}$ into a free field and an interaction term in order to expand the final amplitude for a process as a perturbation series; or breaking space-time up into a finite lattice and estimating $Z[J]$ as a sum over a set of randomly selected field configurations for a decreasing lattice spacing. For the moment however we assume that some method of obtaining $Z[J]$ is available and pass on to its significance and the remainder of the formalism.

And $Z[J]$ *is* central to calculating anything because it embodies all essential information concerning the theory. It is the vacuum-to-vacuum transition amplitude in the presence of a source J and simple intuition about this is correct: to understand the theory it is only necessary to master its behaviour for a closed box "prodded" in an arbitrary fashion. From it we can calculate any Green's function

$$G_N(x_1, ..., x_N) \equiv < 0|T\phi(x_1)\dots\phi(x_N)|0 > \tag{2.18}$$

where T denotes that the fields are time-ordered with $\phi(x_1)$ later than $\phi(x_N)$, using the formula

$$G_N(x_1, ..., x_N) = \frac{1}{i^N} \frac{\delta^N Z[J]}{\delta J(x_1)\dots\delta J(x_N)}|_{J=0}. \tag{2.19}$$

Note that $Z[J]$ is not a function of J: it's $J(x)$ that is a function while $Z[J]$ is a

functional, so the multiple differentiations are functional ones – Ryder [34] gives a concise introduction to the theory of functionals. Two important types within the set of all Green's functions are connected ones, generated from $iW[J] \equiv \log Z[J]$, and one-particle irreducible ones, $\Gamma^{(n)}$, found using

$$\Gamma[\phi_c] = \sum_{n=0}^{\infty} \frac{i^n}{n!} \int d^4x_1 \ldots \int d^4x_n \Gamma^{(n)}(x_1, \ldots, x_n)\phi_c(x_1)\ldots\phi_c(x_n) \tag{2.20}$$

$$\phi_c(x) \equiv \frac{\delta W[J]}{\delta J(x)} \tag{2.21}$$

$$\Gamma[\phi_c] \equiv W[J] - \int d^4x J(x)\phi_c(x). \tag{2.22}$$

Green's functions are an enormously important part of the mathematics in field theory and prime candidates for the unphysical quantities mentioned at the beginning of this section. However, the significant object for the description of experiment is the S-matrix; that is the set of all scattering amplitudes between initial states (as $t \to -\infty$) and final ones, represented as the unitary operator S relating incoming and outgoing free fields.

$$\phi_{out}(x) = S^\dagger \phi_{in}(x) S. \tag{2.23}$$

The closed form is the *reduction formula*

$$S =: \exp[\int \phi_{in}(z) K \frac{\delta}{\delta J(z)} dz] : Z[J]|_{J=0}. \tag{2.24}$$

Colons denote normal ordering in which all annihilation operators are written to the right of all creation ones and the operator K for each external particle is simply that which acting on a free field gives its equation of motion, e.g. $\partial^\mu \partial_\mu$ here in massless ϕ^4. Given S, the calculation of half-lives or cross-sections is purely a matter of kinematics independent of quantum field theory.*

As is well-known, none of this makes sense.

* An observation famously used by Heisenberg and successors in an attempt to subvert it [36].

Any attempt to calculate $Z[J]$ inevitably produces an infinite result which, and this is the critical aspect, passes through to the S-matrix. No attempt merely to regularise the earlier stages of a calculation can avoid this divergence in its answer. As is equally well-known, the source of this difficulty is that the Lagrangian (2.15) has been formulated in terms of an inherently unobservable, irretreivable remote parameter and field. The existence of divergences in the connection between the underlying theory and its physical manifestation naturally entails extra difficulties. Many of these are concerned with the initial regularisation of the theory to produce finite expressions which can then be freely manipulated during the actual process of *renormalisation.* Detailed examination reveals the expressions to be divergent because of the high-momenta, UV parts of the theory and so most regulators introduce a cut-off κ which will be allowed to become infinitely large at the end of the calculation. Since the regularised theory is intended to be a kind of approximation to the full theory in which μ plays so central a role, there must be a substitute scale and it is κ that fulfills this task.* Even so this approximation never replicates all the features of the original and it will in some way be unphysical, e.g. in breaking Lorentz or gauge invariance. This should not matter provided these aspects are not used in any calculation and regulators can normally be selected or developed to ensure this, the other major consideration being convenience. If such flagrant dangers are avoided, all regulators should give the same renormalised theory.

Having disposed of these preliminaries, the formal details of renormalisation and its consequences can be described. In terms of bare quantities our example Lagrangian was

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi_B)(\partial^\mu \phi_B) - g_B \phi_B^4, \tag{2.25}$$

bare quantities which it is assumed are related to their physical counterparts via

* Some regulators, notably dimensional regularisation, have limits which are dimensionless, but these also always need an arbitrary mass somewhere in the formulation. This is, slightly misleadingly, immediately identified with μ. As explained in section 2.1, μ is strictly a *renormalisation* scale and is unconcerned with any regularisation one. However, for exactly the same reasons as κ, they are formally similar and can be combined during renormalisation by a suitable choice of prescription (see section 2.3). Standard presentations do this implicitly.

renormalisation constants

$$\phi_B = \phi Z_\phi, \qquad g_B = g Z_1 Z_\phi^{-4} \tag{2.26}$$

Now

$$\mathcal{L} = \frac{1}{2}(\partial_\mu \phi)(\partial^\mu \phi) - g\phi^4 + \frac{1}{2}(Z_\phi^2 - 1)(\partial_\mu \phi)(\partial^\mu \phi) - (1 - Z_1) g\phi^4. \tag{2.27}$$

The terms to the right are refered to as *counterterms* and contain all the divergences. Their real usefulness only becomes apparent in order-by-order renormalisation using Feynman diagrams when diagrams arising from them will explicitly cancel the divergences in the diagrams of the bare theory. For present purposes we need only assume that both the bare parameters and the renormalisation constants can be made infinitesimal leaving finite physical parameters and an algorithm for calculating observables which gives finite answers. That only two (i.e. finitely many) renormalisation constants are required is what distinguishes the theory as renormalisable; for a discussion of non-renormalisable and super-renormalisable theories see Collins [37].

With a specific model and its innards lain out, we can return to the topic raised at the start of this section: what is the equivalent to (2.9) for unphysical quantities ? One particle-irreducible Green's functions provide a simple example of unmeasurable yet practical quantities and, from (2.18), these clearly transform under renormalisation as

$$\big(Z_\phi(\mu/\kappa, g_B)\big)^n \Gamma_B^{(n)}(q_i, g_B, \kappa) = \Gamma^{(n)}(q_i, g, \mu). \tag{2.28}$$

In our earlier discussion, we could assume that the renormalised quantity was independent of μ and while this is no longer possible, it can be replaced by the observation that the *bare* Green's function is also independent of it. Thus by differentiating (2.28) by μ

$$\left[\mu\frac{\partial}{\partial \mu} + \beta(g)\frac{\partial}{\partial g} - n\gamma(g)\right]\Gamma^{(n)}(q_i, g, \mu) = 0 \tag{2.29}$$

$$\beta(g) = \mu\frac{\partial g}{\partial \mu} \tag{2.30}$$

$$\gamma(g) = \mu \frac{\partial}{\partial \mu} \ln Z_\phi \tag{2.31}$$

analogously to (2.9), but with $\beta(g)$ replacing $-\rho(R)$ and the addition of a $n\gamma(g)$ term. Encouraged by this, can a generalisation of (2.14) be found now that the μ-dependence will be explicitly known in any particular case and the potential anarchy thus eliminated ? If $\Gamma^{(n)}$ has mass dimensions D, examining $q_i \to q_i e^t$ in combination with (2.29) yields the inhomogeneous Callan-Symanzik equation

$$\left[-\frac{\partial}{\partial t} + \beta(g) \frac{\partial}{\partial g} + D - n\gamma(g) \right] \Gamma^{(n)}(q_i e^t, g, \mu) = 0 \tag{2.32}$$

and its solution

$$\Gamma^{(n)}(q_i e^t, g(\mu), \mu) = \Gamma^{(n)}(q_i, g(\mu e^t), \mu) \exp\left[tD - n \int_0^t dt' \gamma(g(\mu e^{t'})) \right]. \tag{2.33}$$

The appearance of $\gamma(g)$ in (2.29) compared to (2.9) signals the existence of *anomalous dimensions*. In the case of an observable $\sigma(Q^2/\mu^2, R(\mu))$ and because the theory is massless, changes in μ and units could compensate transformations like $Q \to e^t Q$ in the external energies. This symmetry, like any other, has a Ward identity associated with it, but since $\Gamma^{(n)}$ is μ-dependent it need not hold for unobservable quantities and the classical symmetry is thus seen to develop quantum anomalies. As usual Coleman's [38] is the best account of the breaking of scale invariance and how it relates to dimensional analysis.

2.3 The Renormalisation Group

It is now desirable to unify and significantly extend the two versions of renormalisation presented so far. Section 2.1 presented the concept stripped down to its positivist essentials, couched as far as possible in terms of observable quantities (thereby avoiding any distinction between finite and infinite renormalisations) and in a form also intended to highlight certain features destined to recur in later chapters. The version just discussed is probably more familiar simply because it reflects the stance adopted during any calculation, but it is not one adapted for any realistic confrontation with experiment. Renormalised variables like g (and masses in more complicated examples) are loosely refered to as having their physical values, without

their measurement having been worried about. Mathematically this is avoided by introducing arbitrary *renormalisation prescriptions* like

$$g = i\Gamma^{(4)}(q_i = 0) \tag{2.34}$$

where $\Gamma^{(4)}(x_1, \ldots, x_4)$ has been transformed to momentum space. Such a decision amounts to ignoring the Lagrangian and defining g as being whichever quantity happens to satisfy (2.34). Formally this is perfectly acceptable since, as we have already noted, the observed physics cannot depend on our choice of variable to use as reference during renormalisation. However it completely confuses any attempt to measure the fundamental constant of the theory. In certain massive theories (including QCD) the situation is even worse: having specified the Lagrangian in terms of a set of masses, confinement of the constituent particles prevents any unambiguous determination of even their renormalised values. We must therefore find a bridge between the mathematically powerful formalism of section 2.2 and the operationally rigourous procedure of section 2.1.

Before showing how this is easily done, one apparent discrepency between the two versions must be explained. In section 2.2 two renormalisations – one of g and one of ϕ – were required. Closer examination reveals that that on ϕ, known as wavefunction renormalisation, has no effect on the S-matrix and thus the wavefunction renormalisation is carried out so that the renormalised field satisfies the standard commutation relations [39], thereby allowing the Feynman rules to be applied to (2.27) rather than (2.25). Only one fundamental constant has to be measured.

An analogy between $\rho(R)$ and $\beta(g)$ has already been noted and this can be tightened by partially differentiating (2.6) with respect to μ:

$$\mu \frac{\partial R(\mu)}{\partial \mu} = \left(\left. \frac{\partial F}{\partial x} \right|_{R(\mu)} \right)^{-1} = -\rho\big(R(\mu)\big) \tag{2.35}$$

as against (2.30). For convenience in dealing with perturbative expansions later, we introduce a *couplant*

$$a = \frac{g^2}{4\pi^2}. \tag{2.36}$$

In a marginal abuse of notation,* let

$$\mu \frac{\partial a}{\partial \mu} = \beta(a) \tag{2.37}$$

where $\beta(a)$ is now to be refered to as a β-function.

Not only will the physical significance of a ($\sim g^2$) depend on the choice of renormalisation prescription, the actual function β is so dependent as well. Indeed it is convenient to reverse this observation and use the β-function as a means of labelling its renormalisation prescription [40]. Also specify μ and the prescription is unique.

A prescription is also called a *renormalisation scheme* (RS).

This realisation that one can just define a scheme by specifying a β-function and a μ without having to worry about what condition corresponds to (2.3) in this prescription is of major conceptual importance, even if it does not directly help to calculate anything. It is a comprehensive way of expressing the extensive freedom we have in formulating a renormalisable theory without physics being in any way effected. This freedom is known as *renormalisation group* (RG) *invariance*, the group operation in question being that of changing from one RS to another. Indeed it is really only when expressed in terms of changing the (now arbitrary) β-function that we see that this is naturally an infinite group.

One important way of expressing the difference between schemes is given by the counterterms. As described above, counterterm renormalisation entailed splitting the Lagrangian into two halves, one of which contained all the divergences of the theory. Because finite terms can be absorbed into divergent ones without altering the nature of the latter, the last sentence cannot completely fix this procedure and a change in scheme just involves the transfer of a finite piece from one half of the Lagrangian to the other.

That this leaves the total Lagrangian unchanged is RG invariance. If the couplant and β-function are a and $\beta(a)$ in one scheme and a' and $\beta'(a')$ in another then

* Apart from a brief resurrection in section 4.5, $\beta(g)$ as defined in (2.30) will not appear again. Any confusion can only effect book-keeping, since $\beta(a)$ and $\beta(g)$ both play the same role in the theory.

(c.f. (2.11)) these must be related via

$$\beta(a) = \frac{da}{da'}\beta'(a') \tag{2.38}$$

using (2.37) above. With the full RG available, the problems raised at the start of the section simply melt away. Pick a scheme $\beta(a)$ and calculate $R(a)$ therein. There is evidently an RS in which

$$\beta'(x) = -\rho(x) \tag{2.39}$$

for $\rho(x)$ from (2.1), and where a comparision of (2.11) and (2.38) shows that $R = a'$. We are thus free to conduct calculation in whichever scheme we happen to find easiest and then coordinate these results with experiment via (2.38) and whichever observable we choose for R.

2.4 Perturbative Series

Beguiling though the formalism and its renormalisation may be, field theory is nothing unless a means of calculation. As already noted, the formulation of any practical technique at all is impossible without the acceptence of some degree of approximation and difficult even then. Precious few – perhaps only lattice theory, Schwinger-Dyson equations and dispersion relations aside from what follows – are of any significance and of these perturbation theory has been the overwhelmingly dominant tradition. Although this thesis will not resort to actually performing a conventional perturbative calculation, properties and problems generic to the results of these calculations are central to it and it is therefore appropriate to explain what a perturbation series is in this context.

Central to the method is the division of the Lagrangian (2.15) into free-field and perturbative terms

$$\mathcal{L} = \mathcal{L}_0 + g\mathcal{L}_1(\phi). \tag{2.40}$$

Naturally it has been ensured that the $g = 0$, free-field theory is solvable:

$$Z_0[J] = \exp(\frac{-i}{2}\int d^4x' \int d^4x J(x')\Delta_F(x'-x)J(x)) \tag{2.41}$$

$$\Delta_F(x'-x) = \int \frac{d^4p}{(2\pi)^4} e^{-ip.(x'-x)}(p^2+i\epsilon)^{-1}. \tag{2.42}$$

Δ_F is known as the Feynman propagator and the $i\epsilon$ is purely conventional, with $\epsilon \to 0^+$ at the end of the calculation. The central result to make the interacting theory tractable is

$$Z[J] = \mathcal{M} \exp\left(ig \int \mathcal{L}_1(-i\frac{\delta}{\delta J})dx \right) Z_0[J]. \tag{2.43}$$

Its proof is tedious and the reader is refered to Ryder [34] for details. Two remarks are in order: none of the remaining integrals are functional ones and note the way each ϕ in $\mathcal{L}_1$ has been replaced by a functional derivative. This expression is still rigourously exact, but remains uncalculable. The step from a non-perturbative theory to a perturbative one is finally taken and the exponential is expanded as a power series in g

$$Z[J] = \sum_{n=0}^{\infty} \frac{g^n i^n}{n!} \int d^4x_1 \ldots d^4x_n \mathcal{L}_1(-i\frac{\delta}{\delta J(x_1)}) \ldots \mathcal{L}_1(-i\frac{\delta}{\delta J(x_n)}) Z_0[J]. \tag{2.44}$$

With $Z_0[J]$ available as an explicit functional (2.41) and $\mathcal{L}_1(\phi) = \phi^4$, it is merely a matter of (extreme) patience to calculate any term in the series and by using the results of section 2.2 one can thus express any observable as an expansion in powers of g. And in a certain sense that is it: perturbative scalar quantum field theory as a logically complete structure.

However this present version bears no resemblance to the calculational tools of the practitioner. He or she uses (2.44) translated into topology and speaks of *Feynman rules* and *diagrams*. Strictly this translation introduces no new physics, yet it has been one of the most profound and fruitful reformulations in the science of our century. Since the visualisation of Green's functions and amplitudes as all possible networks with a specified number of ends and only certain kinds of vertex is one of the most widely disseminated ideas in physics and since its mathematical underpinings are incidental to the general properties of the series obtained, it is unnecessary to elaborate on the details. The basic idea has been explained better than we could hope to do here [41]. Of course each diagram is only a graphical mnemonic for a piece of algebra it is necessary to evaluate en route to a numerical coefficient.

With the theory being calculated order-by-order in perturbative expansions, one must renormalise in a similar fashion and it is now that the method of counterterms reveals its elegance. These can be treated as additional interaction terms and consequently give rise to extra diagrams distinct from the standard ones. In the original expansion diagrams containing closed loops typically contribute divergences and for each of these divergent graphs there is a new counterdiagram, also divergent but in such a way as to cancel the existing one up to finite amounts. One particular prescription, the $\overline{MS}$ one, has become dominant and is invariably used in phenomenology. Rather than being explicitly defined by its β-function (about which we currently know little), this scheme is specified by a calculational prescription. Handling diagrams with closed loops in the bare theory around which arbitrary momenta flow always reduces to integrals of which

$$\int \frac{d^4p(p^2)^r}{(p^2-a^2)^s} \tag{2.45}$$

is an example. The integration is over all the possible momenta in the loop and so the integral diverges. In dimensional regularisation the integral is continued to $n = 4 - \epsilon$ dimensions and evaluated there. Before taking the $\epsilon \to 0$ limit which removes the regulator, a $1/\epsilon$ pole must be removed by renormalisation. Apart from this pole, a group of constants always appears and so it is discarding

$$\frac{1}{\epsilon} + \ln 4\pi - \gamma_E \tag{2.46}$$

from all divergent integrals which defines the $\overline{MS}$ scheme. The *coupling* is written $\alpha_S^{\overline{MS}} = a\pi$ and the determination of this at a particular energy is often the immediate aim of any experiment. Because dimensional regularisation preserves gauge invariance, its development was a crucial step in proving the renormalisabilty of non-Abelian gauge theories and is the method of choice in the Standard Model, even though it remains to be rendered meaningful outside perturbation theory.

The extension of the complete programme above to the point where the Standard Model Lagrangian can be treated involves several new technical developments. Once the unfamiliar axioms of Grassman algebras are accepted, fermions can be included alongside bosons and in most theories, the inclusion of masses is trivial – the free-field Lagrangian can usually be solved for massive fields and for the purposes of

renormalisation the masses are merely additional (dimensionful) couplings.* But, crucially, an alternative technique for incorporating masses is apparently adopted in nature.

2.5 A Theory Of Almost Everything

The Lagrangian (2.15) of massless ϕ^4 theory was not intended as a serious candidate for a realistic theory and indeed it has proved too simple a structure to account for all the experimental evidence gathered to date concerning events on a scale smaller than, say, an apple. But not by much.

$$\mathcal{L} = \overline{L} i\gamma^\mu \partial_\mu L + \overline{R} i\gamma^\mu \partial_\mu R - g(\overline{q}\gamma^\mu T_a q) G^a_\mu - \frac{1}{4} G^a_{\mu\nu} G^{\mu\nu}_a \tag{2.47a}$$

$$- \frac{1}{4} W^a_{\mu\nu} W^{\mu\nu}_a - \frac{1}{4} B^a_{\mu\nu} B^{\mu\nu}_a \tag{2.47b}$$

$$- \overline{L}\gamma^\mu (g' \frac{1}{2} \tau_a W^a_\mu + g'' \frac{Y}{2} B_\mu) L - \overline{R}\gamma^\mu g'' \frac{Y}{2} B_\mu R \tag{2.47c}$$

$$+ |(i\partial_\mu - g' \frac{1}{2} \tau_a W^a_\mu - g'' \frac{Y}{2} B_\mu)\phi|^2 - V(\phi) \tag{2.47d}$$

$$- (G_1 \overline{L} \phi R + G_2 \overline{L} \phi_c R + \text{hermitean conjugate}) \tag{2.47e}$$

where

$$G^a_{\mu\nu} = \partial_\mu G^a_\nu - \partial_\nu G^a_\mu - g f_{abc} G^b_\mu G^c_\nu \tag{2.48a}$$

$$W^a_{\mu\nu} = \partial_\mu W^a_\nu - \partial_\nu W^a_\mu - g' f'_{abc} W^b_\mu W^c_\nu \tag{2.48b}$$

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{2.48c}$$

is just such a structure once the known fields and quantum numbers have been correctly assigned and two additional pieces (ϕ and part of q) are hypothesised. The elementary fields then correspond to three massive leptons (electron, muon and tau) and their neutrino companions, six quarks (up, down, strange, charm, bottom and top), the photon, the intermediate vector bosons ($W^\pm$ and Z^0), eight gluons and a Higgs boson. Yet more remarkably, this presentation of the *Standard Model* is needlessly explicit and somewhat hides the symmetries around which the Lagrangian has been formed, for it is no more than that of the $SU(3) \times SU(2) \times U(1)$

* Even in this case the arguments of section 2.1 roughly carry through [33], the infinitesimal bare masses being unable to provide an adequate mass scale.

non-Abelian gauge theory with its latter two symmetries spontaneously broken by a minimal Higgs mechanism.

A non-Abelian *gauge theory* is one invariant under the local gauge transformations of its fields

$$\psi(x) \rightarrow \exp(\frac{i}{2}\lambda_j \alpha_j(x))\psi(x) \tag{2.49a}$$

$$G^i_\mu(x) \rightarrow G^i_\mu(x) - \frac{1}{g}\partial_\mu \alpha_i(x) - f_{ijk}\alpha_j(x)G^k_\mu(x) \tag{2.49b}$$

where the $\alpha_i(x)$ are arbitrary functions and the λ_i are the generators of a particular Lie group. Its Lagrangian is now constrained to be

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu \partial_\mu - m)\psi - \frac{1}{2}g(\overline{\psi}\gamma^\mu \lambda_i \psi)G^i_\mu - \frac{1}{4}G^i_{\mu\nu}G^{\mu\nu}_i \tag{2.50}$$

with the field strength $G^i_{\mu\nu}$ given in terms of the gauge field G^i_μ by (2.48a). Quantising this theory is rather delicate precisely because of the extra symmetry which now relates infinite classes of field configurations, all sharing the same action, so that when the path integration is taken over all configurations a new divergence is introduced into the generating functional. Understanding now that the integration should only be over physically distinct configurations, gauge fixing terms must be added to the Lagrangian, but the reader is again directed to textbooks [34] for a full account of how this is done and why the solution is elegantly equivalent to the introduction of unphysical Faddeev-Popov ghost fields.

In most of the remainder our concern will be with circumstances where most of (2.47) is irrelevant and a good description of experiment is provided by *Quantum Chromodynamics* (QCD), defined by (2.50) with the gauge group taken to be colour $SU(3)$. Of the particles listed above only the quarks and gluons possess the colour quantum number and so their fellows can be temporarily forgotten while the gluons are assigned to the vector gauge particles G^i_μ $(i = 1, 8)$ and the quarks to the fermionic ψ. Perturbative calculations of QCD β-functions reveal one of the theories crucial properties, *asymptotic freedom*, by which it is meant that

$$b = \frac{33 - 2N_f}{6} > 0 \quad \text{for} \quad N_f < 17 \tag{2.51}$$

and so that (perturbatively) g decreases and hence we can anticipate an improvement in the applicability of perturbation theory as the energy increases. Historically

[42], the realisation that only non-Abelian gauge theories are asymptotically free was of prime importance in the proposal of QCD as a theory of the strong interaction. The property is possibly also responsible for ensuring the consistency of the renormalised theory, a complicated issue avoided in our discussion of renormalisability where it was assumed that the programme outlined was indeed sufficient to render observables finite. This has proved to be the case in ϕ^4 [37] and gauge theories [43] when they are defined in their perturbative versions, but the question of any non-perturbative generalisation is completely unresolved. With the short-distance, UV properties of the theory being the fundamental obstacle forcing infinite renormalisations upon us, it may be conjectured that only asymptotically free ones are well-behavied enough in this limit for consistency to be attained. That such a restrictive result may be appropriate is a serious possibility given detailed studies of ϕ^4 ([44] and references therein).

Conversely, at low energies QCD behaves in a more strongly interacting fashion and the perturbative approximation must ultimately break down. Indeed calculating the equivalent of (2.1) to the one-loop level in perturbation theory yields

$$\alpha_s(Q) \sim \frac{1}{\log Q^2/\Lambda^2} \tag{2.52}$$

indicating some sort of pathology near $Q \simeq \Lambda$. Since $\Lambda \simeq$o(100 MeV) roughly marks the mass range of the lightest hadrons constructed from the (effectively massless) quarks, the presence of this *Landau pole*, the breakdown of perturbation theory and the existence of these bound states are regarded as broadly related.

This pattern of asymptotic behaviours is reversed in Quantum Electrodynamics (QED), the unprecedentedly successful theory of low-energy interactions of charged particles mediated by photons, whose construction is analogous to that of QCD, but with the gauge group now the Abelian one $U(1)$ to give the familiar classical gauge invariance of Maxwell's equations. Now $b = -2N_f/3$ being negative means that the theory has a small coupling at low-energies. Naïvely this is taken to explain why QED achieves astonishing accuracy in its predictions, of which the best known is its anticipation of the electron's magnetic moment to one part in 10^7. Since this thesis has as one of its main concerns the ambiguities which prevent QCD being tested to this level of precision, there will be occasion (chapter 4) to examine

this commonplace sceptically. At the other end of the energy-scale, experimental evidence for the high-energy behaviour of pure QED is unlikely to be forthcoming for reasons suggested by the way in which the $U(1)$ of electrodynamics is indirectly incorporated into the Standard Model Lagrangian. As (2.47) and (2.48c) show, a $U(1)$ with gauge field B_μ and quantum number Y explicitly appears, but this Y is the weak hypercharge rather than the electric one and QED only emerges when (2.47) is rewritten to acknowledge that its symmetries are broken ones.

Although of enormous significance conceptually and in the development of ideas about renormalisability, *spontaneous symmetry breaking* is of no consequence in later chapters, where QCD is the main concern. It derives from the fact that every solution of a theory need not share the symmetries of the Lagrangian from whence it was generated, an eventuality which can immediately undermine the developments of section (1.4) where it was tacitly assumed that the free-field vacuum had been correctly identified prior to it being perturbed about. By forcing the free-field solution to have symmetries neither it nor the full solution possess, an accurate perturbative answer is unlikely to be forthcoming. Actually this is not an immediate problem in gauge theories, since the Lagrangian (2.50) does not exhibit this phenomenon. Unless an extra, Higgs, field ϕ is coupled into the theory. Only when it is realised that this can be done with the Lagrangian remaining gauge invariant, but in such a fashion that it solves the problem of how to introduce a massive gauge field – since mass terms like $M^2 G_\mu G^\mu$ are clearly not invariant under (2.49b) the gauge bosons had previously been required to be massless – does the idea become compelling. Details of how (2.47) is recast to display its equivalence to a theory containing an $SU(2)$ gauge theory with massive W and Z bosons can be found in any textbook [45]. To conjecture that the Lagrangian is unaltered by this breaking and hence still perturbatively renormalisable is natural [46], to prove altogether more difficult, but still feasible [43].

At the time of writing the large amounts of data generated by a quartet of collaborations (ALEPH, DELPHI, L3 and OPAL) operating at CERN's LEP e^+e^- collider are allowing a detailed comparision of the Standard Model against reality at an energy of 91 GeV where its three interactions form the heirarcy of strengths*

* "Strength" is normally a loose reference to the size of the a introduced in (2.36).

$$\alpha < \alpha_W \ll \alpha_S \tag{2.53}$$

the weak interaction being the broken $SU(2)$ portion of the model carried by the massive gauge bosons. Without aspiring to a similar accuracy, the testing of the electromagnetic and weak sectors of the theory has been approached with an attitude akin to that used in testing QED at low energies : perturbation theory can give an unambiguous framework in which the calculation of higher-orders merely leads to refined predictions of effectively unlimited precision. Much remains to be done by way of improving statistics and analyses and in the extension to higher energies, but it is clear that the $SU(2) \times U(1)$ part of the Standard Model stands unfalsified. However such an approach is much less convincing when adopted to test QCD. Since this thesis will be largely concerned with this problem, the force of which is to be explained in the next chapter, $SU(2) \times U(1)$ will play the incidental role of a precursor to the QCD processes of interest at LEP, although QED will occasionally appear as an example of a simple gauge theory.

The reader may be wondering what all this has to do with the fate of QED at high energies, but the very success of gauge symmetries as a guide in constructing the Standard Model has encouraged the widely held view that it in turn can be replaced by a simpler structure enclosing its three, at present still essentially distinct, interactions in a single one. Many examples of Grand Unified Theories (GUT's) have been proposed, all exploiting the remarkable observation that when extrapolated to high energies the ordering (2.53) of the couplings collapses, with the three becoming approximately equal at around 10^{15} GeV, and all of which explain this by postulating the existence of a simple gauge symmetry (originally $SU(5)$), manifest at very high energies, but spontaneously breaking at around this 10^{15} GeV into $SU(3) \times SU(2) \times U(1)$. Thus if QED were to be tested at high energies it would merge into an asymptotically free unified interaction well before the Landau pole of the pure theory might ever be encountered. Given current technology such scales are impossible to investigate directly, but what happens there should constrain certain aspects of phenomena already observed, in particular the relative values of the coupling (2.53) at 91 GeV. Crudely, the three theories incorporated into (2.47) act independently

---

The discussions of sections 2.1 and 2.3 may be thought of as motivating this as the random choice of a (hopefully typical) "cross-section".

below the GUT scale and so knowing the β-functions of these theories enables the running of their couplings to be predicted and their convergence at a unique energy checked, hence accurate measurements at LEP of these couplings (necessary as a boundary condition on this integration) is a test of the GUT hypothesis. When done in detail [47] the three are found to not quite meet. However this need only falsify the supposition that the Standard Model behaviour remains valid up until a single breaking scale and numerous extensions of this basic idea already existed – supersymmetric models, $SO(10)$, etc. – in which the breaking occurred at more than one scale. Any such generalisation almost invariably introduces an extra arbitrary parameter and the fit is such that the limited data can normally be accommodated by this extra degree of freedom, although the resulting models must always be physically reasonable and not predict too rapidly decaying a proton. While this limits the conclusions to be drawn, these fits are important as virtually our only means of anticipating what lies in store above a few TeV. In the foreseeable future the main experimental uncertainty involved will remain that on α_S at 91 GeV and an improvement in this measurement is thus paramount. It is to the vagueries of its determination that we now turn.

CHAPTER THREE

The Reliability Of Perturbation Theory

3.1 Introduction

Perturbative expansions, calculated to any finite order, violate one of the complete theory's essential properties, namely the freedom of any observable from dependence on the renormalisation point μ. As was remarked earlier, different choices for μ distribute the contents of a function $\sigma(q^2/\mu^2, a(\mu))$ amongst the terms of a perturbative series in $a(\mu)$ in different ways and – while this means that an infinitely wide range of superficially unrelated expansions are possible, a judicious choice from among which may greatly simplify the problem – it is the fact that such redistributions involve the whole series, whereas our knowledge of it is unavoidably partial, that creates severe difficulty in perturbation theory.

Because calculations of tree-level coefficients do not involve renormalisation, the problem only appears when the one-loop diagrams have been evaluated, but is then present in all orders thereafter. The seriousness of this dependence should not be underestimated; it is after all the hallmark of an unphysical quantity. Rather than estimating an unambiguous result, in most orders the perturbative truncation can provide any possible answer whatsoever – and all at the same time. In destroying a central feature of what we are trying to test, we appear to have forfeited the possibility of prediction. How then can the numerous experimental results claiming to verify QCD be explained ? Instead of succumbing to despair, it has been common practice to supplement the closely-defined procedures of perturbative calculation with a set of ad-hoc and informal rules of thumb motivated by, often vague, appeals to variously "reasonableness," "commonsense," "physical intuition" and "approximation theory" (see [48] for review]). Since it is usually unclear whether these alternatives are reinforcing, mere prejudice, or even mutually exclusive it is probable that a clear resolution of this confused situation will require not only a more detailed understanding of why they may be (un)justified, but also a fresh perspective. Though

the related debate as to why these issues were never a cause for concern in the application of QED is partly historical, it is also an ongoing physical one.

It is as well to state at the outset the author's guess that the standard strategies will ultimately prove to have been broadly correct (so that, for instance, $\Lambda_{\overline{MS}}$ is not likely to be 5 GeV), although possibly only because they were sufficiently conservative to prevent contradiction; as such they are a necessary and reasonable first step, a broad-brush solution. They do however raise the danger that in being too vague, too loose, they obscure otherwise discernable and important detail in data already gathered. Currently about half of the quoted error in global determinations of α_S is a consequence of the residual theoretical uncertainties left by these conventional approaches. If nothing else, the prospect of possibly reducing this justifies the challenge. That said, the view developed later in this chapter is only a partial answer and the enlargement on one part of it in Chapters 5-7 particularly tentative. But this will be true of any answer, a definite one being equivalent to an exact, non-perturbative treatment of the theory.

3.2 Definitions: Perturbation Series

Throughout the last chapter quantities like σ and $\beta(a)$ were taken to be exactly determinable functions instead of the truncated perturbative versions encountered in practice and it is now necessary to settle on a notation for these prior to embarking on an investigation of this truncation's consequences.

Suppose that a programme of evaluating Feynman diagrams is halted after the Nth order in the renormalised couplant $a \equiv \alpha_S/\pi$ has been found, so that $\sigma(a)$ is approximated by

$$\sigma^{(N)}(a^{(N)}) = K_1 a^{(N)} + K_2 {a^{(N)}}^2 + K_3 {a^{(N)}}^3 + \ldots + K_N {a^{(N)}}^N \qquad (3.1)$$

where the coefficients K_n are dependent on the choice of RS for $n > 1$, as is the approximation to the coupling $a^{(N)}(\mu)$, which is a solution of the truncated β-function equation in the same RS

$$\mu \frac{da^{(N)}}{d\mu} = -b{a^{(N)}}^2(1 + ca^{(N)} + c_2{a^{(N)}}^2 + \ldots + c_{N-1}{a^{(N)}}^{N-1}) \qquad (3.2)$$

$$\equiv -b{a^{(N)}}^2 B^{(N)}(a^{(N)}). \qquad (3.3)$$

Here b and c are the RS invariants the former defined in section 2.5 and the latter by

$$c = \frac{153 - 19N_f}{2(33 - 2N_f)} \text{ (QCD)}, \qquad c = \frac{3}{4} \text{ (QED)} \tag{3.4}$$

while the rest of the coefficients $c_2, \ldots, c_{N-1}$ are RS dependent. Note that any observable can have its perturbation series massaged into the form of (3.1) by subtracting constant terms from and taking powers of the original series. Given the existence of schemes (such as $\overline{MS}$) whose c_2 is known, but also of observables whose K_2 is the only coefficient known, it may be queried why both series have been truncated at the same order; might this not be discarding useful information ? This is however just the standard practice in even the simplest of perturbative calculations elsewhere and close reflection on the matter [49] justifies this by showing that the lowest order discarded in either series always dominates the truncation error.

It has already been seen (section 2.1) that an equation like (3.2) implies a mass scale $\tilde{\Lambda}$ such that $\mu/\tilde{\Lambda}$ is invariant under changes of unit, so it is convenient to define a new dimensionless variable

$$\tau \equiv b \ln\left(\frac{\mu}{\tilde{\Lambda}}\right) \tag{3.5}$$

in place of μ for use in labelling schemes. To Nth order, each RS now uniquely corresponds to a point in the abstract coordinate space $(\tau, c_2, \ldots, c_{N-1})$. Integrating equation (3.2) yields a transcendental equation [40]

$$\tau = \frac{1}{a^{(N)}} + c \ln\left(\frac{ca^{(N)}}{1 + ca^{(N)}}\right) + \int_0^{a^{(N)}} dx \left[-\frac{1}{x^2 B^{(N)}(x)} + \frac{1}{x^2(1 + cx)}\right] \tag{3.6}$$

with $a^{(N)}(\mu)$ in the relevant RS as its solution. In doing so a particular (infinite) boundary condition had to be chosen and this is the origin of the second term in the integrand cancelling the divergence in the reciprocal of the β-function. As should be clear on comparing equations (2.1)–(2.3) with these above, this decision is equivalent to defining the integration constant in section 2.1, a process which was responsible for the appearance of Λ. Here any finite modification of the term used to cancel the divergence can be absorbed into a redefinition of the parameter $\tilde{\Lambda}$ in τ. Unfortunately, the usual Λ_{QCD} was defined in a slightly different context [50] by

one of these slightly different conventions and, although the relation

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

$$= 1.15\Lambda_{QCD}, \qquad N_f = 5 \tag{3.6b}$$

between the two is rather simple, there is the possibility of confusion. In addition, the value changes with the number of effective flavours at each quark threshold [48] and some sources quote the four flavour $\Lambda^{(4)}_{QCD}$, others the $\Lambda^{(5)}_{QCD}$ relevant at LEP. When quoting numerical values we will always use the conventional $\Lambda^{(5)}_{QCD}$, but $\tilde{\Lambda}$ is too theoretically elegant not to be adopted elsewhere.

Using different schemes also gives different values of Λ: having moved to another μ, this can be compensated by changing units so that its numerical value returns to its original value, while altering Λ. However these values in different schemes are easily related using no more than a one-loop calculation [51]. To understand this consider a quantity

$$R = a(1 + r_1 a + \ldots) = a'(1 + r'_1 a' + \ldots) \tag{3.7}$$

calculated in two schemes RS and RS' whose couplings are related via

$$a' = a(1 + \nu_1 a + \nu_2 a^2 + \ldots). \tag{3.8}$$

Straightforwardly substituting gives $r_1 = \nu_1 + r'_1$. Integrating up the equivalent of (2.38) relating the β-functions of the two schemes involves an integration constant

$$C = \int^{a} \frac{dx}{\beta(x)} - \int^{a'} \frac{dx}{\beta'(x)} \tag{3.9}$$

which can be found by expanding in powers of a

$$C = \frac{1}{a} + c\ln\left(\frac{ca}{1+ca}\right) + o(a) - \frac{1}{a'} - c\ln\left(\frac{ca'}{1+ca'}\right) + o(a') \tag{3.10}$$

$$= \nu_1 + o(a) \tag{3.11}$$

and equating equal powers, so that one finds $C = \nu_1$. But this constant relates to the definition of Λ and in particular

$$\tau - \tau' = r_1 - r'_1. \tag{3.12}$$

This can be reformulated into the statement [40]: for each observable the combination $\rho_0 \equiv \tau - r_1$ calculated in one scheme has the same numerical value as the equivalent combination calculated in another, i.e. ρ_0 is an RS invariant.

Since this is only bound up with changes in τ, are there other invariants related to the full set of changes in the Renormalisation Group ? To see that there are, it is easiest just to construct an appropriate set [52]. They are based on the Effective Charge (EC) scheme [53] of the observable: the effective charge of σ is simply

$$R = \frac{\sigma(a')}{K_1} = a'(1 + r_1 a' + r_2 a'^2 + \ldots)$$
$$r_n \equiv \frac{K_{n+1}}{K_1} \tag{3.13}$$

in the scheme $(\tau, c_2, c_3, \ldots)$ with couplant a'; the EC scheme for σ is then the scheme $(\rho_0, \rho_2, \rho_3, \ldots)$ with couplant

$$a \equiv R \tag{3.14}$$

and β-function $\rho(a)$ where

$$\rho(x) = x^2(1 + \rho_1 x + \rho_2 x^2 + \ldots). \tag{3.15}$$

The standard formula

$$\rho(R) = \frac{dR}{da'}\beta(a') \tag{3.16}$$

relating these two schemes can be expanded in powers of a' to yield the coefficients $\rho_2, \rho_3, \ldots$ in terms of the c_n and the r_n in the original RS. The actual definition of the EC scheme is (3.14) alone, which is independent of $(\tau, c_2, c_3, \ldots)$ so

$$\rho_0 = \tau - r_1$$

$$\rho_1 = c$$
$$\rho_2 = c_2 + r_2 - r_1 c - r_1^2$$
$$\rho_3 = c_3 + 2r_3 - 4r_1 r_2 - 2r_1\rho_2 - r_1^2 c + 2r_1^3$$
$$\ldots \qquad (3.17)$$

is an infinite set of scheme invariants. Each observable has its own set of similar invariants. In the version truncated at the Nth order, knowing the perturbation series $(r_1, r_2, \ldots, r_{N-1})$ of $\sigma^{(N)}(a^{(N)})$ and the RS $(\tau, c_2, \ldots, c_{N-1})$ to that order is sufficient to calculate $(\rho_0, \rho_2, \ldots, \rho_{N-1})$ and no more. More generally, knowing any two of these three sets of coefficients allows the third to be deduced, making these invariants convenient for the translation of $(r_1, r_2, \ldots, r_{N-1})$ calculated in scheme $(\tau, c_2, \ldots, c_{N-1})$ into those which would have been found had another scheme $(\tau', c'_2, \ldots, c'_{N-1})$ been used.

As in all such cases, these invariants can be arbitrarily recombined into other sets of equally valid invariants, a process equivalent to defining the invariants as the β-function coeffients in whatever scheme gives a particular set of series coefficients, e.g. specifying that the series take the form $R = ae^a$, rather than the $R = a$ that it does in the EC scheme, and using whatever β-function coefficients this gives. Such redefinitions are possible, but unenlightening in that they lead to more complicated formulae throughout.

3.3 Definitions: Observables

Before turning to the problems involved in testing QCD, it will be as well to define several specific e^+e^- observables for use as the need arises. A mature approach to testing beyond tree-level is to choose observables for study mainly on the basis of their experimental and calculational convenience instead of trying, for instance, to isolate a "gluon jet." Indeed, once the simplistic belief that a 3-jet event is allowing one to "see" a gluon has been passed over, one can more or less define a "3-jet event" in any way one wishes, although of course, as many a textbook picture illustrates, some events do appear strikingly "three-jetty" and it would normally be perverse to choose a definition excluding these. Ultimately we might hope to be able to calculate every observable accurately, so the ones chosen to begin with need only be well understood and not of any great physical significance.

In the case of jet fractions, a range of definitions have been discussed, each of broadly the same experimental practicability, but with differing higher-order and hadronisation corrections. Some of these are simple in conception – e.g. hadrons must fall inside a cone of specified size in order to constitute a jet – others defy visualisation and are often cast in the form of an algorithm with the energy-momenta of particles in an event as input. By far the most significant subset is that of *JADE-type algorithms* [54] in which the experimentalists (actually their substantially swifter electronics) assign a number y_{ij} to each pair (i, j) of particles, locate the pair with the smallest value and then, provided that value satisfies

$$y_{ij} \leq y_c s \tag{3.18}$$

where y_c is an arbitrary parameter, these two particles are eliminated from further consideration, but have their energy and momentum combined to form a "pseudoparticle" which is included on an equal footing with the other particles when this procedure is repeated, as it is until all pairs fail (3.18). The total number of remaining particles and pseudoparticles is the number of jets in this event. Both y_{ij} and the rule for forming a pseudoparticle from its parents remain to be defined and different choices give rise to the variety of JADE-type algorithms used (Table 3.1). At first sight it may appear that the E algorithm is the most natural, but adding the 4-vectors of massless particles does not produce a massless pseudoparticle, while most theoretical work concerns massless partons, so it is usually felt to be advantageous to ensure that all particles are massless throughout, a requirement responsible for much of the variation in Table 3.1 [55]. All the algorithms are infra-red safe but subject to hadronisation corrections varying from the small to fairly large. Most of the current interest therefore centres on two algorithms, E0 and Durham, which both appear to have small ($< 5\%$) corrections [56].

Having divided the observed events into two, three etc. jet events, the ratios

$$f_n(y_c) \equiv \frac{\sigma(\text{n-jet})}{\sigma_{tot}} \tag{3.19}$$

known as jet fractions, are the actual observables. These have perturbative expan-

Name	y_{ij}	Recombination
E	$(p_i + p_j)^2$	$p_k = p_i + p_j$
E0	$(p_i + p_j)^2$	$E_k = E_i + E_j$
		$\mathbf{p}_k = \frac{E_k}{\lvert\mathbf{p}_i+\mathbf{p}_j\rvert}(\mathbf{p}_i + \mathbf{p}_j)$
JADE	$2E_iE_j(1 - \cos\theta_{ij})$	$p_k = p_i + p_j$
Durham	$2\min(E_i^2, E_j^2)(1 - \cos\theta_{ij})$	$p_k = p_i + p_j$

Table 3.1. Selected Jet Algorithms

sions in the couplant of the form

$$\begin{aligned} f_2 &= 1 - K_{21}a - K_{22}a^2 - K_{23}a^3 + \ldots \\ f_3 &= \quad K_{31}a + K_{32}a^2 + K_{33}a^3 + \ldots \\ f_4 &= \quad\quad K_{42}a^2 + K_{43}a^3 + \ldots \end{aligned} \tag{3.20}$$

Of the coefficients there only the one-loop K_{21}, the tree-level K_{31}, K_{42} and K_{53} [57], the two-loop K_{22} and the one-loop K_{32} are known fully for the common algorithms [58], although approximations of practical accuracy ($\sim$ 10%) are known for the higher tree-level ones [59]. Relatively simple examination of the formulae in Table 3.1 can reveal relationships between coefficients for different algorithms using kinematics only. Thus K_{31} is identical for both E0 and JADE, while it is larger for E.

Problems occur at small y_c because a 3-jet fraction will be structured thus

$$\begin{aligned} f_3 =& a(A_{10} + A_{11}\ln y_c + A_{12}\ln^2 y_c) \\ &+ a^2(A_{20} + A_{21}\ln y_c + A_{22}\ln^2 y_c + A_{23}\ln^3 y_c + A_{24}\ln^4 y_c) \\ &+ \ldots \end{aligned} \tag{3.21}$$

and for $y_c < 0.05$ the $\ln y_c$ terms become large. There are precidents for both this problem and its solution, most notably the recent improvements [60] in the treatment of both thrust (see below) and heavy jet masses, $M_H^{(T)}$ and $M_H^{(M)}$, by resumming the leading and next-to-leading logarithms. These developments are part of the application to QCD of ideas originally developed by Sudakov to deal with

bremsstrahlung in QED when the emission is so severely constrained kinematically that only soft, collinear photons can be radiated (see review by Catani [61]). Under these circumstances the total amplitude

$$1 + \sum_{n=1}^{\infty} \int dw(1,\ldots,n)\Theta(1,\ldots,n;y_c) \tag{3.22}$$

obtained by integrating the probability $dw(1,\ldots,n)$ for emitting n photons over the phase space $\Theta(1,\ldots,n;y_c)$ and then summing over the number of photons, becomes

$$1 + \sum_{n=1}^{\infty} \frac{1}{n!} \prod_{i=1}^{n} \int dw(i)\Theta(i,y_c) \tag{3.23}$$

$$= \exp\left(\int dw(i)\Theta(i,y_c) \right) \tag{3.24}$$

where the exponential only involves the description of single photon emission. This is dependent both on the factorisation of $dw(1,\ldots,n)$, the emission of n photons being that of n *independent* single ones, and also the factorisation of the phase space in the soft, collinear Sudakov limit. Although the first of these might have been expected to be inapplicable to coloured gluons, it can happen with certain QCD quantities. If similar arguments apply to jet fractions, when y_c is small the final term in each row of (3.21) will be large enough to destroy the approximation, but their coefficients will be related and all will be combinable into an exponential like (3.24). The details are complicated and work is still in progress, although well-advanced [62]. Difficulties arose when Brown and Stirling [63] pointed out that the second, phase space factorisation didn't occur for existing algorithms, basically because in them two soft gluons could be combined into a jet even though they would more be more naturally split between two other jets. Subsequent attention has focussed on two specially invented algorithms, the Durham [64] and Geneva [56] ones. It appears that the former exponentiates, whereas the latter probably does not [61]. Comparisions with data remain to be done for resummed calculations, but the $o(a^2)$ forms have been examined in detail with promising results [56].

An alternative to dividing an event up into jets is to classify its shape. For any event one can define the thrust

$$T = \max\left(\frac{\sum_a |\mathbf{p}_a.\hat{\mathbf{n}}|}{\sum_a |\mathbf{p}_a|} \right) \tag{3.25}$$

with the sums taken over all particle 3-momenta in the CM frame and the maximum is found by varying unit vector $\hat{\mathbf{n}}$. For a large event sample, a distribution in T can be established. The same final state particles can be used to define a tensor

$$\theta_{ij} = \frac{\sum_a p_a^i p_a^j / |\mathbf{p}_a|}{\sum_a |\mathbf{p}_a|} \tag{3.26}$$

which will have eigenvalues λ_n. Then the C-variable is

$$C = 3(\lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1), \tag{3.27}$$

varying from zero for back-to-back events upwards. Further details of both and the results for r_1 can be found in [65].

A less discriminating, but theoretically better understood, measure of purely hadronic events is the total hadronic cross-section, known as the R-ratio when normalised as

$$R \equiv \frac{\sigma(e^+e^- \to \text{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)}. \tag{3.28}$$

In the $\overline{MS}$ scheme this has the very well-known expansion [66] ($\mu^2 = q^2$, $N_f = 5$)

$$R = \frac{11}{3}(1 + a + 1.409a^2 - 12.8a^3 + o(a^4)) \tag{3.29}$$

$$= \frac{11}{3}(1 + \delta_{QCD}) \tag{3.30}$$

where δ_{QCD} is the form appropriate to (3.1) above. Close to the Z^0 peak there are substantial electroweak corrections to this QCD result, so the data to be used [67] is that taken at $Q = 34$ GeV where these can be neglected. On the peak, δ_{QCD} can be checked using the ratio of hadronic and leptonic decay widths

$$R_Z \equiv \frac{\Gamma_{had}}{\Gamma_{lep}} \tag{3.31}$$

$$= (19.97 \pm 0.03)(1 + \delta_{QCD}) \tag{3.32}$$

where the numerical factor is electroweak [68]. These massless QCD results can be modified to include heavy quark masses [69], but the changes are small ($K_1 = 1.05$,

$K_2^{\overline{MS}} = 1.41$) and insignificant in the present context. Unfortunately, because both R and R_Z are $o(1)$ in α_S, δ_{QCD} is small and its accurate measurement difficult, but they have the advantage that hadronisation can be ignored. For the previous observables, while being fairly small at the points that will be used later, the size of the hadronisation corrections varies across the distributions [65]. Except for that coming from lower energies and for R_Z (where the 1991 LEP average is used instead [70][71]), all data in the following comes from OPAL [70].

3.4 Claims And Misconceptions

In addressing the problem of scale dependence at NLO there are four attitudes discernable in the literature:

a) Treating μ as an unknown and fitting to the data for it.

b) Setting it to some physical scale in the problem, typically the centre of mass energy.

c) Picking a value such that

$$\mu \frac{dR}{d\mu} \simeq 0$$

i.e. minimising the dependence on the scale.

d) Using the ambiguity to make the series converge well, normally by adopting the EC scheme, also known as Fastest Apparent Convergence (FAC).

Although each of these in its pure form has its advocates (e.g. [72], [73],[40] and [53] respectively), it is more common for more than one to be appealed to, often as a justification for one of the others. In a recent global determination of α_S carried out by OPAL [70] all three appear in some form: an average is taken of the couplings obtained using a) and b), while checking that this estimate is consistent with c) and including for the first time the Durham jet algorithm with its soft μ-dependence alongside the usual JADE ones. As is clearly stated in doing so, this conservatism is a response to the lack of agreement on a correct procedure. Faced with the same problem the DELPHI Collaboration [71] basically consider variations in scale from $0.002E_{cm}$ to E_{cm} and then institute a complicated averaging procedure to handle the correlations between different observables.

Probably the most naïve approach involves fitting the scale to the data. However, simply fitting for μ and Λ to a single observable is impossible. Consider the one-loop truncation

$$\sigma^{(2)}(a^{(2)}) = K_1\left(a^{(2)} + a^{(2)^2}\frac{K_2}{K_1}\right) \tag{3.33}$$

$$= K_1(a^{(2)} + a^{(2)^2}(F(a^{(2)}) - \rho_0)) \tag{3.34}$$

where

$$F(x) = \frac{1}{x} + c\ln\left(\frac{cx}{1+cx}\right) \tag{3.35}$$

and

$$\mu\frac{da^{(2)}}{d\mu} = -ba^{(2)^2}(1 + ca^{(2)}) \tag{3.36}$$

$$r^{(2)} = F(a^{(2)}). \tag{3.37}$$

Figure 3.1 displays the generic parabolic shape of $\sigma^{(2)}(a^{(2)})$, with the dotted line representing the result of a typical measurement [48]. Normally both ρ_0 (i.e. Λ) and $a^{(2)}(\mu)$ are unknown so this single measurement does not suffice to determine them. Adding additional observables to the sample is of no help, there still being too many unknowns. Progress is only possible provided one is willing to make some sort of assumption relating μ values for *different* observables. Naturally the observables thus linked are ones already continuously connected, e.g. points in a thrust, y_c or C distribution are all fitted using a common μ. Innocuous as this may seem for a first approximation, it has a contradiction at its heart: to be interesting a distribution must have some strong variation – fitting a common μ and Λ to a set of points in a perfectly flat distribution is hardly likely to be an improvement on fitting them to one – yet this is exactly when one might imagine that "μ" varies widely across the distribution. It is also a scheme dependent approximation.

In all such fits the data is not being used to test QCD; it is being parameterised using a crude parameterisation which has some features in common with that theory. Parameterisations are certainly useful, but are usually a means of compressing information without any pretence at interpretation, often as an admission of ignorance

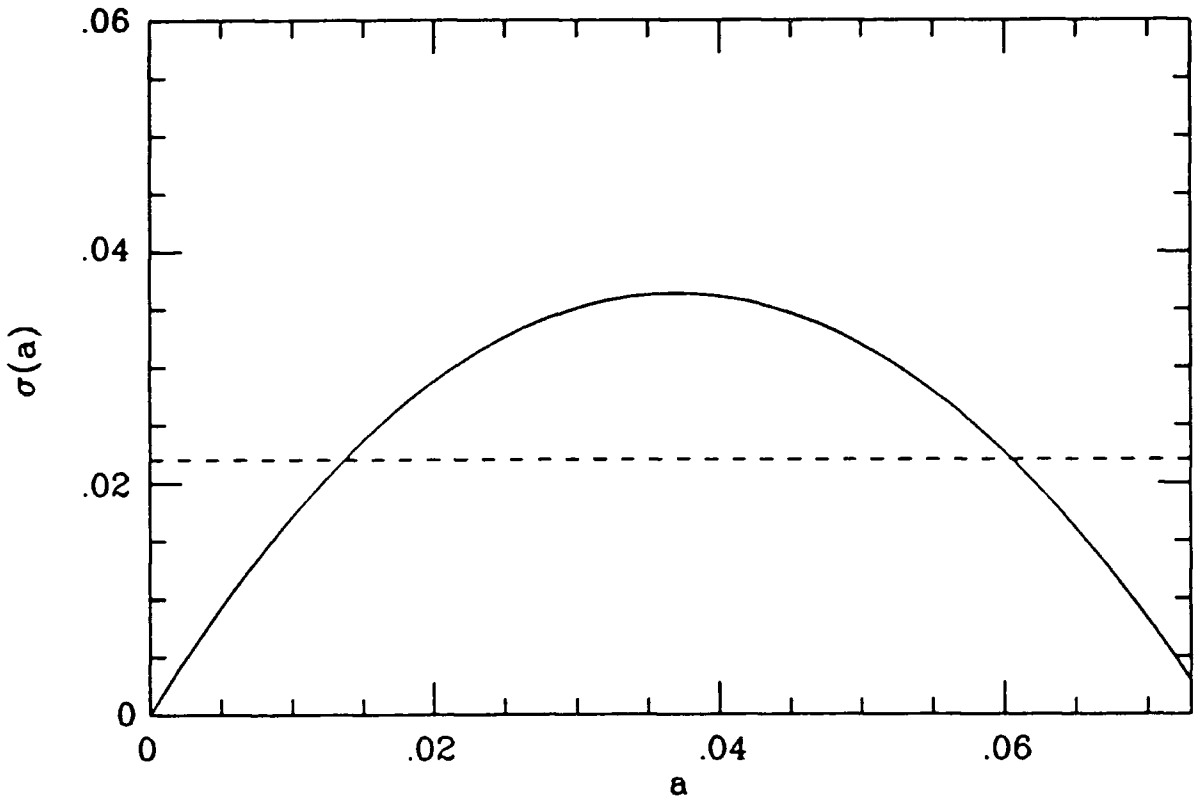

Figure 3.1 Scale Dependence of $\sigma^{(2)}$

about the underlying theory. As a mixture of theory and simple guesses, current fits are phenomenology in the original sense and as such an unreliable (although suggestive) basis for the measurement of a fundamental parameter in the complete theory, at least until the approximations can be cross-checked. For their sample of observables the OPAL Collaboration [70] find that fitting $x = \mu/E_{cm}$ gives values ranging from 0.008 for E0 jet rates and 0.017 for the C-variable through to 0.77 for oblateness. Quite apart from the $\Lambda_{\overline{MS}}$ values inferred, by themselves these numbers tell us nothing. By adopting some additional assumption about the "correct" scale it may be possible to claim something, but one has to be very careful. After all, one may be trying to distinguish between cases where the theory is wrong, the idea of a single scale is inadequate and where the assumption is incorrect.

Note that when the resummations mentioned above were done with thrust, where $\ln(1-T)$ terms are involved, expressions were obtained giving a good fit to the data over widely seperated T values with a fitted scale $\mu \sim Q$ and a greatly

reduced dependence on that scale at small $(1-T)$ [60][70]. Although minor problems matching results in the Sudakov limit to the conventional ones – with $M_H^{(T)}$ and $M_H^{(M)}$ the resummed results work well in the previously problematic small mass regions, but slightly spoil the existing agreement elsewhere [70] – remain, these are important advances which should greatly increase confidence in our ability to fit QCD predictions realistically. However, with the benefit of hindsight it should be emphasised however just how wrong earlier, and indeed most current, fits were. Using a single μ across an $o(a^2)$ distribution can be a terrible approximation.

Popular simply as an independent choice of scale, position b) is also often used as the extra assumption. The standard argument in its favour is that if there is only one energy scale Q in the problem then $\mu \sim Q$ is a plausible guess; a detailed understanding of the full physics involved may modify this, but probably by no more than numerical factors of $o(1)$. There are cases where $\mu = Q$ fits the data extremely well, notably in deep-inelastic scattering – Martin, Roberts and Stirling [74] report that even allowing some functional variation in μ fails to improve the agreement significantly – but these are the exception rather than the rule, particularly in e^+e^- reactions as we have seen [70]. Rather a lot of room for maneouvre is still left and so one finds arguments claiming to offer improvements by using more realistic physics. For instance, the e^+e^- jet fractions present a particular difficulty for the original estimate because fitting as described above gives values at least an order of magnitude less than the centre of mass energy Q. So it is sometimes argued that the relevant energy is actually $Q/3$, since this is the energy of each jet in a 3-jet event [56], or something smaller because the "essential" physics is presumed to take place at an even lower level in the event. Arguing about $Q/3$ as against Q in this context is to greatly overestimate the reliability of what is being done. Quite why this sort of dimensional argument, be it estimating the size of atoms, stars or mountains, is usually so successful when applied cautiously remains a mystery [75], but it can never distinguish a factor of 3 ($\sim \pi$) which could arise for any number of trivial reasons in an exact proof. Sometimes employed as a first draft, the method invariably works best when the underlying physics is already well understood and is no substitute for insight. Rephrasing this, its success depends on a degree of certainty that no other massive quantities are involved, when in fact one (Λ) is to hand. With the jet fractions it may be thought significant that the fitted

$\mu \sim \Lambda$, but this would suggest the involvement of low-energy non-perturbative effects when the actual hadronisation corrections appear small [65]. Finally, as its very dependence on μ indicates, the truncated series is possibly not a very physical object and one's intuition may break down – butchering the theory may introduce large factors spuriously. Similarily, any discussion of physical content in the dimensional continuation required to define the $\overline{MS}$ scheme must surely be tentative.

Note that it is not an argument for $\mu = Q$ to point out that otherwise $r_n(\mu/Q)$ will contain large logarithms of μ/Q. Until it has actually been calculated there is no way to know how $r_n(1)$ compares to the result for some other choice of μ. To suggest that it is likely to be "naturally" smaller is only the above reasoning in disguise.

The issue is further clouded by the precedent of QED where "natural" schemes (including a scale) are often appealed to as the reason why the choice was never a problem there. In fact, a large part of the explanation seems to be historical inertia. It can be rigourously proved in QED [76] that the classical Thomson cross-section for Compton scattering is recovered in the long-wavelength limit

$$\frac{d\sigma}{d\Omega} \to \frac{\alpha^2}{m_e^2}(\epsilon.\epsilon')^2, \qquad k \to 0. \tag{3.38}$$

The existence of this classical limit, a consequence of infrared freedom, enables perturbation theory to be applied at the very small energies appropriate to, say, solid-state physics where the electron charge can be precisely measured using the Josephson effect. Although our liberty to define the coupling in whatever way we choose still exists, it is only sensible, in preventing unnecessary confusion, to agree on a definition which can be compared to the usage elsewhere at this low energy and, once this is agreed, to present the results of all calculations in this RS. This is unobjectionable, provided it is remembered that this is no more than a convention. Unfortunately the description of this RS as the "natural" choice – which it is in the sense of least confusing – seems to have mutated into the folklore conviction that this is a "natural scheme" because it has some physical significance.* And so long

* An argument by Collins [77] claims that, unlike QED with the electron mass, QCD has no external mass scale and hence Λ_{QCD} is unmeasurable. This relies on

as perturbation theory gave such impressive agreement with experiment while using $\mu \sim m_e$ there was little reason to worry about the possibility of changing the RS. With the theory now being used at much higher energies with running couplings, these positions no longer accord with everyday practice. Thus the issue is not why there was not a scale dependence problem in QED, but why the scheme that happened to be picked was responsible for such reliable predictions ? Alternatively, why is QED so forgiving ? An answer to this, consistent with our attitude to QCD, will be given in the next section.

Turning to c), what does the size of the μ-dependence indicate about the reliability of an observable ? Contrary to folklore, for example

> "substantial renormalisation scale dependence ... indicates that perturbative corrections beyond the order calculated are not yet negligible," [56]

by itself it is rather uninformative [79]. Returning to the truncation (3.34) - (3.37), in realistic cases the second term in $F(a^{(2)})$ is small and in neglecting it none of the conclusions are changed. Now

$$\mu \frac{d\sigma}{d\mu} = 0 \quad \Rightarrow \quad a \sim \frac{1}{\rho_0} \tag{3.39}$$

and at this turning point

$$\mu^2 \frac{d^2\sigma}{d\mu^2} \simeq -\frac{2K_1 b^2}{{\rho_0}^3}\left(1 + \frac{c}{\rho_0}\right)^2, \tag{3.40}$$

so the flatness or otherwise of the μ-dependence at one-loop only depends on ρ_0. This is perhaps made more vivid by taking an example of current interest: the only reason that the Durham jet algorithm in fixed-order (not the resummed results) has smaller μ-dependence than the E0 one is because the one-loop correction (i.e. r_1) is smaller. Scale dependence is intimately bound up with the size of K_2 and is not an

---

a general observation [78] that if physics cannot depend on the choice of units, a "Theory of Everything" can only predict ratios of masses and is not, as claimed, a special consequence of dimensional transmutation. Since QCD is no more a TOE than QED, one is in fact still free to use m_e as a reference mass.

additional characteristic of the result as sometimes implied [56][61]. As an insight this is somewhat trivial, since it merely reiterates that a one-loop calculation by itself does not say anything about higher-orders. Uncalculated terms *are* unknown.

To be exact, the above derivation would need an exact value of Λ, so as to be able to find ρ_0. In practice this has to be replaced by one found from fitting with a common μ and only then can one study how stable this fit is under changes of that μ. Naturally one hopes that the fit shows only a small dependence on μ, but again care must be taken because there are two effects involved – that arising from assumming a common μ and the intrinsic μ-dependence of the truncation. Furthermore one must be wary of any claim to have estimated "the higher-order corrections" under any circumstances since this concept is highly scheme dependent. In fact studying the μ-dependence of a fit does tell one something about how large the corrections are, but only in a very restricted set of schemes. Although there are cases (the jet fractions at small y_c for example) where this information has been useful, it is not as significant as sometimes implied.

No such claim about higher-orders is made by Stevenson, the foremost proponent of c) in the form of his Principle of Minimal Sensitivity (PMS) [40][49]. This is the scheme-fixing prescription defined by

$$\mu\frac{dR^{(n)}}{d\mu} = 0, \qquad \frac{dR^{(n)}}{dc_j} = 0, \qquad j = 2, 3, \ldots (n-1). \tag{3.41}$$

Assessments of this extreme version of c) are readily available [48][80]; it is rather impervious to direct criticism because its justification lies outside field theory in the general notion that any approximation should be insensitive to the choice of any unphysical parameters. One could argue that as a desirable property of approximations it should be examined a posteriori, instead of being used as a basis for defining the approximation in the first place. There is also the reservation that it relies on the derivative of an asymptotic (?) power series which may not itself be asymptotic.

A remark general to all these positions is that none of them, even in combination, unambiguously add to our understanding. Were a reliable value of $\Lambda_{\overline{MS}}$ to be provided by other means, there would be little to be learnt in using it to investigate their conspicuous failures because the idea of a "correct" scale reproducing the

correct results is a ficticious one. As more orders are added to perturbation theory, one certainly expects that any of the extracted Λ's will approach the correct value, but one cannot anticipate that the "correct" scale will do the same – indeed as the dependence on μ decreases the fitted values will presumably become unstable. Paradoxically, if the idea of a "natural scale" were to be correct, we would learn less and less about it the better the theory is approximated.

Finally, simply as a choice of scheme (in this context) the EC one's only merit is the obvious opportunistic one. However, the formalism associated with it is of deeper significance than that.

3.5 Energy Dependence

If "higher-order corrections" have proved too elusive to help assess the reliability of perturbation theory, what can they be replaced by ? QCD makes two types of prediction about an observable R: how it relates to other observables and how it varies with energy. Investigations of a sample of observables at a single energy are normally coordinated via a Λ which should be universal, but as we have just seen this process is bedeviled by uncertainties concerning μ. In contrast, the energy dependence of R is a much cleaner prediction of the theory:

$$\begin{aligned} Q\frac{dR}{dQ} &= -b\rho(R) \\ &= -bR^2(1 + cR + \rho_2 R^2 + \ldots). \end{aligned} \tag{3.42}$$

The choice of notation here is deliberate and the function $\rho(R)$ is just the EC β-function.

This allows a simple test of the theory without any parallel of the μ-ambiguity. The situation is exactly that discussed in section 2.1 where it was seen that integrating up gives a prediction

$$R(Q) = f^{-1}(\log(Q/\Lambda)) \tag{3.43}$$

where Λ is the $\tilde{\Lambda}$ introduced above. By calculating diagrams as usual in whatever RS is most convenient and then using equations (3.17), as many terms as desired in the expansion of $\rho(R)$ (and hence f^{-1}) can be calculated. The only uncertainty is

the unavoidable one caused by neglecting terms in the expansion.* Measuring $R(Q)$ at a single Q then enables Λ to be infered.

It perhaps comes as little surprise to discover that this procedure is formally equivalent to using a particular RS and that it is the Effective Charge scheme at that. Interestingly, even though all schemes would then give the same answer, this scheme is probably the simplest to use were the theory to be applied non-perturbatively. Although Λ is what is wanted for comparisions between observables, for any particular quantity the same information is probably better encoded in the scheme invariant ρ_0. Now suppose that a complete calculation of some observable function $R(Q/\mu, \alpha_S(\mu))$ were to be available, how would ρ_0 be extracted ? The extrapolation of conventional attitudes would entail inverting this to find $\alpha_S(\mu)$ for some (now) arbitrary μ and then calculating ρ_0 via (3.6) and (3.17), assuming the β-function of whatever scheme has coupling α_s has also been calculated. However this is unnecessarily complicated and it is more natural to eliminate the first stage by taking R instead of α_s as the coupling, i.e. to use the EC scheme. Then R, once measured, is inserted straight into

$$\rho_0 = F(R) + \Delta\rho_0(R) \tag{3.44}$$

where $F(R)$ was defined in (3.35) and

$$\Delta\rho_0(R) \equiv \int_0^R dx \left(\frac{-1}{\rho(x)} + \frac{1}{x^2(1+cx)} \right). \tag{3.45}$$

Equation (3.44) also plainly displays where our ignorance resides when we compromise with a perturbative truncation. $F(R)$ is a trivial function of the measured

* Chýla [81] has erroneously argued that (3.42) avoids μ-dependence only at the expense of introducing a new ambiguity. As in Chapter 2 (3.43) can be rewritten as $R(Q) = f^{-1}(f(R(\mu))+\log(Q/\mu))$, but approximating f by perturbatively truncating ρ does not introduce a μ-ambiguity into it. The easiest way to see this is to note that the μ-independence of R to all orders does not depend on the actual form of ρ, so only approximating this function doesn't alter that. Truncations do introduce a μ-dependence into $\sigma(\mu, R(\mu))$, but this is an entirely separate issue.

value, but our knowledge of $\Delta\rho_0$ is restricted to its perturbative expansion

$$\Delta\rho_0(R) = \int_0^R \frac{dy(\rho_2 + \rho_3 y + \ldots)}{(1+cy)(1+cy+\rho_2 y^2 + \ldots)}. \tag{3.46}$$

At the one-loop level, the standard truncation implies that $\Delta\rho_0 = 0$ and this in conjunction with (3.44) gives values for ρ_0 and $\Lambda_{\overline{MS}}$.

To return to (3.42) and its consequences, inverting (3.43) also gives (3.44). However, the interpretation is now completely different.

In the conventional approach, once chosen the EC scheme (or any other choice of scheme) effectively ignores the μ-ambiguity thereby explaining why it is absent from the formalism, whereas now that ambiguity genuinely does not arise because *a separate problem is being considered.* Expressed differently, whatever formal resemblance

$$\mu\frac{\partial a}{\partial \mu} = \beta(a) \quad \text{and} \quad Q\frac{dR}{dQ} = \rho(R) \tag{3.47}$$

may have, they correspond to different physics. As a general idea, this sort of suggestion about using (3.43) to measure Λ has been made by Grunberg [53] (leading him to define EC schemes) and by Dhar and Gupta [82], but these proposals foundered on unnecessary confusion surrounding the interpretation of the EC formalism.

If the theory is truncated at the one-loop level (i.e. $\Delta\rho_0 = 0$), what sort of results are obtained ? An important comparison here is with the PMS result: this is also equivalent (as any scheme-fixing method must be) to specifying an approximation ρ'_0 to ρ_0,

$$\rho'_0 = F(R) + o(R). \tag{3.48}$$

Thus to a good approximation, at the one-loop (and in fact higher-loop [48][80]) level PMS and EC give indistinguishable values of ρ'_0 and adopting the latter formalism cannot reduce the substantial scatter (common to all known methods) found in the values of $\Lambda_{\overline{MS}}$ obtained using PMS [70]. There is however a major difference in the way this scatter is interpreted by the two approaches – unlike PMS, the EC approach can assign direct physical significance to it.

Comparing (3.42) and its integrated version (3.45) shows $\Delta\rho_0$ to be controlling the deviation from the approximation

$$Q\frac{dR}{dQ} \simeq -bR^2(1+cR) \tag{3.49}$$

to the running of R. Because of asymptotic freedom as $Q \to \infty$, $R \to 0$ and

$$F(R) \to \infty, \qquad \Delta\rho_0 \to 0. \tag{3.50}$$

At very large energies, one finds

$$|\Delta\rho_0| \ll |\rho_0| \tag{3.51}$$

and $\Delta\rho_0 = 0$ becomes an increasingly good prescription for extracting Λ; in particular the error involved is roughly

$$\left|\frac{\delta\Lambda}{\Lambda}\right| \simeq \frac{\rho_2}{b^2\ln(Q/\Lambda)} \quad \text{as} \quad Q \to \infty. \tag{3.52}$$

In any scheme, the scheme dependence problem reduces in severity as asymptotia is approached , eventually disappearing when the theory becomes trivial in the limit. Conversely, it is only because the theory must be applied sub-asymptotically that the problem exists at all and this is as much part of the issue as truncation. Unlike for the higher-order terms neglected by the truncation, there is a natural measure of the extent to which a particular observable is sub-asymptotic – none other than $\Delta\rho_0$ [79].

With the scale dependence problem mollified in such cases, one (at least initially) ought to concentrate on quantities satisfying (3.51). From this point of view the reversal in the sign of b and the direction in which the coupling runs in going from QCD to QED is largely irrelevant, so it is important that (3.51) is admirably obeyed in the standard tests of the latter theory. For instance, the anomalous magnetic moment of the electron has $\rho_0 \simeq 400$ and $\Delta\rho_0 \simeq 0.003$ when $Q \sim m_e$, to be compared to QCD at $Q \simeq M_Z$ when $\rho_0 \simeq 20$ and $\Delta\rho_0 = o(10)$ as we shall see below. The fundamental reason for the success of QED is thus seen to be that it has been tested sufficiently close to its asymptotic regime for the scheme dependence to be irrelevant, rather than non-existent.

Observable	$\Delta\rho_0$
f_2 E0 ($y_c = 0.10$)	0.99 ± 0.15
f_3 E0 ($y_c = 0.10$)	0.98 ± 0.21
f_2 D ($y_c = 0.10$)	2.04 ± 0.78
f_3 D ($y_c = 0.06$)	2.05 ± 1.35
Thrust ($T = 0.83$)	0.98 ± 0.69
C ($c = 0.38$)	1.47 ± 0.77
R_Z	7.16 ± 2.82
R	6.14 ± 2.79

Table 3.2. Selected $\Delta\rho_0$ Values

Thus instead of treating the EC formalism as merely another method of fixing a scheme in QCD and hence extracting $\Lambda_{\overline{MS}}$, it is more fruitful to reverse the logic and investigate the consequences of assuming some value of $\Lambda_{\overline{MS}}$. From (3.44) a measurement of R unambiguously gives $\rho_0 - \Delta\rho_0$ and the guess for $\Lambda_{\overline{MS}}$ along with a one-loop calculation in any RS determines ρ_0 and so one can find $\Delta\rho_0$ for different quantities. The particular value of $\Lambda_{\overline{MS}}$ is pretty incidental because other choices only correspond to uniform shifts in the values of $\Delta\rho_0$ for all the observables and so experiment and a one-loop calculation easily fix *relative* values of $\Delta\rho_0$, i.e. up to an unknown universal constant [79]. The scatter in the values of $\Lambda_{\overline{MS}}$ previously extracted now translates into a scatter in the values of $\Delta\rho_0$ which is to be interpreted as showing that different observables are approaching asymptotia at different rates.

For illustrative purposes, letting $\tilde{\Lambda}^{(5)}_{\overline{MS}} = 110$ MeV, the observables described in section 3.3 give the results in Table 3.2, which are also plotted in Figure 3.2. The points selected to represent distributions are typically ones with small errors. For T and C there is little variation across the distributions, but with the jet fractions there is substantial variation at small y_c where large corrections are anticipated, as is clear in Figure 3.3 where the full f_3 distribution for E0 is shown. Points were therefore selected from the large y_c region where $\Delta\rho_0$ is constant within the errors. Note that no scale uncertainty need be included in any of the errors, which are consequently purely experimental.

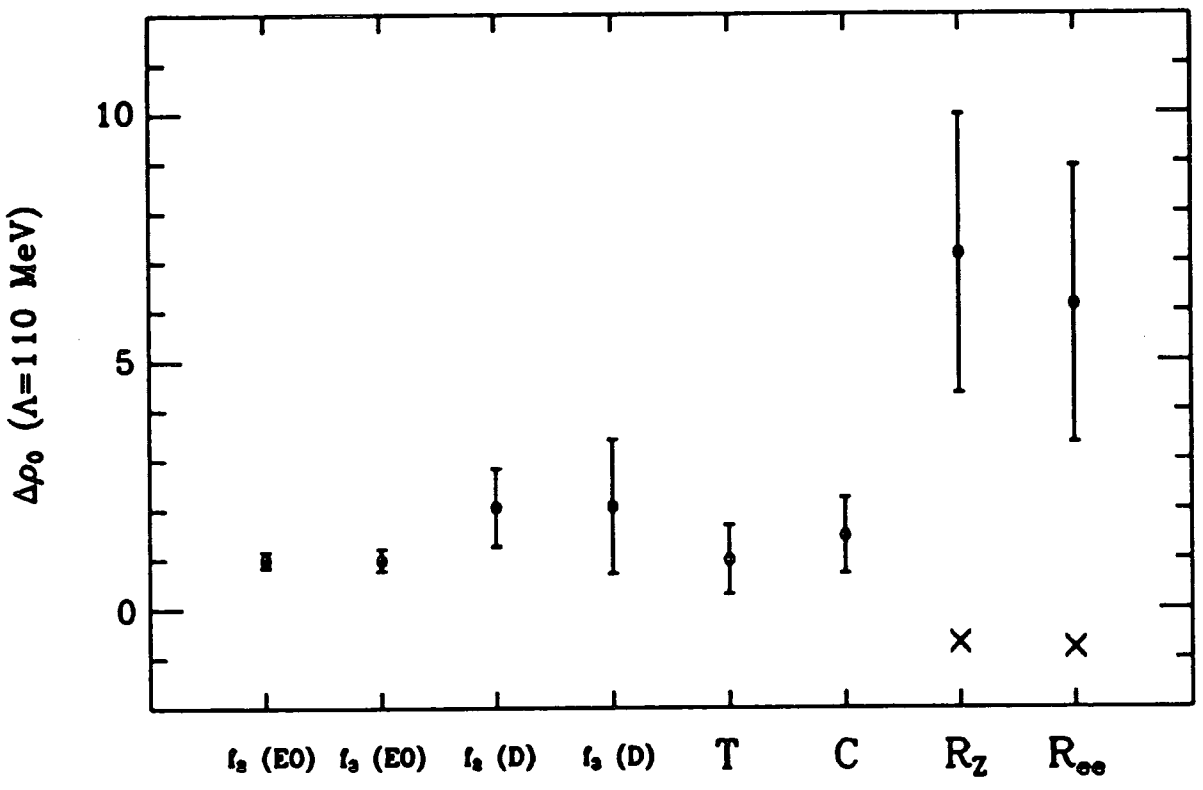

Figure 3.2 $\Delta\rho_0$ for selected observables assuming $\tilde{\Lambda}^{(5)}_{\overline{MS}} = 110$ MeV.

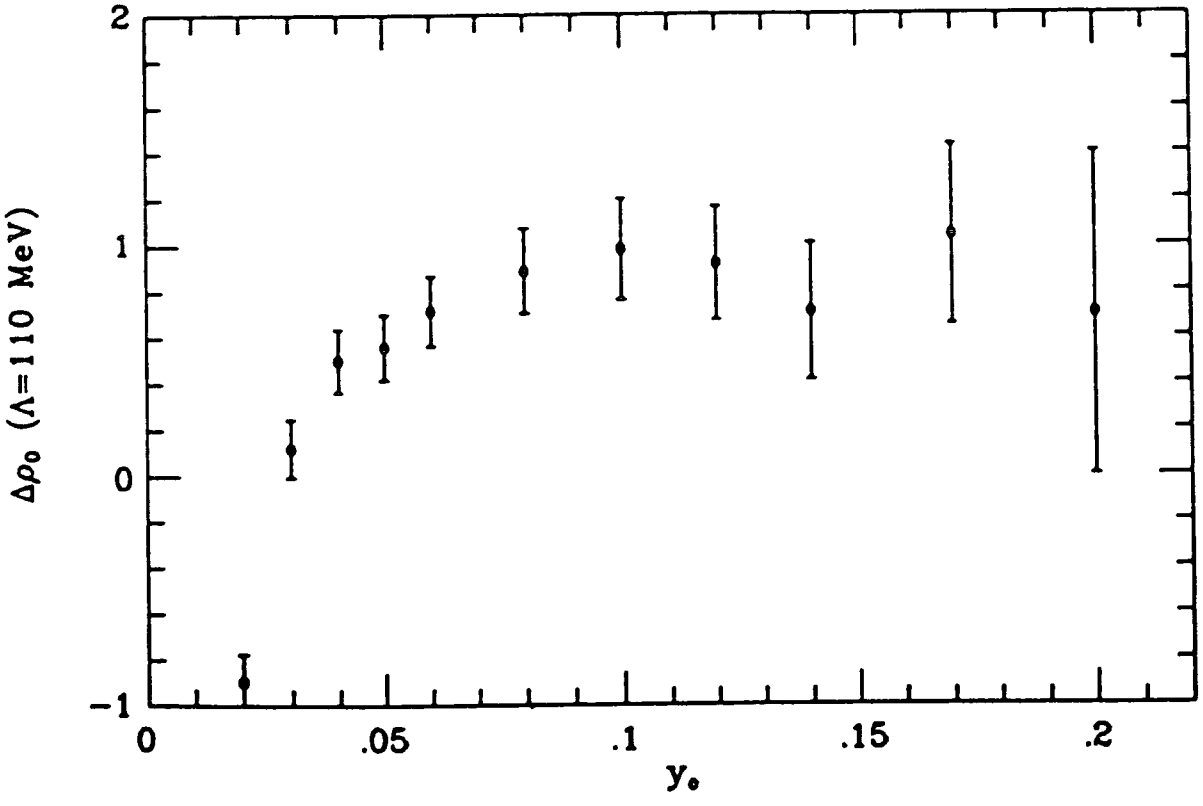

Figure 3.3 $\Delta\rho_0$ for E0 3-jet fraction

Returning to the question of significance, how does this differ from assuming a value for $\Lambda_{\overline{MS}}$ and then extracting some putatively "correct" scale μ for each observable ? Firstly, although the present data on energy dependence is slightly too crude for the purpose, in principle there is an independent check on a possible range for $\Delta\rho_0$ using (3.42). There are basic difficulties: in QED where the processes are liable to be clean the higher-order terms on the right-hand side of (3.42) will be small because $R \simeq \alpha$ is, while in QCD the corrections are large, but so are the errors. Aside from the inherent disadvantage of starting in $o(a^2)$ instead of $o(a)$ so that absolute sizes are smaller to begin with, practical pressures are in favour of accumulating large statistics at a single energy rather than aportioning beamtime amongst a range of energies. Bethke [55] displays the data available for the 3-jet fraction f_3 (at fixed y_c) with the intention of demonstrating that α_S runs convincingly in the manner predicted by QCD, although (3.42) suggests that a better interpretation is that f_3 itself runs. The points are somewhat scattered with relatively large errors, but are on the verge of usefulness. For instance, using JADE data [54] for the 3-jet fraction ($y_c = 0.08$) between Q =34 and 44 GeV, where the errors are slightly smaller then elsewhere, gives

$$Q\frac{dR}{dQ} = -0.022 \pm 0.008 \tag{3.53}$$

and

$$\Delta\rho_0 \simeq R\left(b(Q\frac{dR}{dQ})^{-1} + \frac{1}{R^2(1+cR)}\right) \tag{3.54}$$

then yields

$$-3.4 < \Delta\rho_0 < 8.7 \tag{3.55}$$

as a crude estimate, which translates into the bound [79]

$$\tilde{\Lambda}^{(5)}_{\overline{MS}} < 300\text{MeV}. \tag{3.56}$$

The advent of LEP 200 should help by providing an extended baseline and refining this sort of analysis for different variables may be able to produce a useable estimate of $\Lambda_{\overline{MS}}$ free of scale dependence uncertainties, as well as a cross-check between values of $\Delta\rho_0$ derived via (3.42) and those found from (3.45), thereby testing the consistency of QCD.

Secondly, unlike with μ, it is believed that $\Delta\rho_0$ can be calculated perturbatively with each additional order adding to our knowledge of it (at least until the divergence of a presumably asymptotic series sets in) rather than rendering it increasingly meaningless. However the omens for this are not good. By itself Figure 3.2 cannot decide which quantities have $|\Delta\rho_0| \simeq 0$, the freedom to shift everything up or down by changing Λ still existing. However if the data is correct this cannot be true for all the observables shown, with R and R_Z clearly requiring closer consideration. Fortunately, δ_{QCD} is the only series in the theory calculated to $o(a^3)$, and therefore $\Delta\rho_0$ can be approximated using

$$\Delta\rho_0^{NNLO} = \int_0^R dx \frac{\rho_2}{(1+cx)(1+cx+\rho_2 x^2)}, \tag{3.57}$$

these estimates appearing as crosses in Figure 3.2. Shifting the observed values down by increasing Λ could eliminate the difference, but only at the expense of increasing $|\Delta\rho_0|$ for f_3 (E0) beyond the crude bound (3.55) above. Although recognising that the discrepency is only about 2 standard deviations and hence subject to confirmation, a cautious conclusion would be to suggest that not all (if any) observables in QCD have reached asymptotia.

If conventional perturbation theory cannot yet reliably estimate $\Delta\rho_0$, other approaches must be turned to. This might be taken as the counsel of despair unless one such alternative – lattice gauge theory – were already close to feasibility. In a recent quenched-approximation calculation [83] of the 1P-1S charmonium splitting, comparision with the actual difference yielded the estimate $\tilde{\Lambda}^{(5)}_{\overline{MS}} = 110^{+36}_{-31}$ MeV. Although this must be treated with extreme caution – there are naturally difficulties in extrapolating from $\Lambda^{(0)}$, whatever this means, to $\Lambda^{(5)}$ and their inferred α_S is two standard deviations below the LEP average* – it is to be noted that this would suggest that most of the observables considered here are very close to asymptotia. This will be confirmed or falsified by improved measurements.

Finally, there is the possibility of some innovative basis being found for a prediction of $\Delta\rho_0$ in a particular case, probably for one where it is small. Chapters 5 through 7 will describe a speculative approximation to $\Delta\rho_0$ for the jet fractions which, amongst other features, has $\Delta\rho_0 \simeq 0$ as a consequence.

* But an average obtained using the methods criticised above.

CHAPTER FOUR

Divergence Proofs and Large Orders

4.1 Introduction

Whither perturbation theory if the only truly reliable way of determining Λ were to be through the use of lattice theory, a technique which ultimately aims to replace it ? A cynical response might be that expansions are only a set of numbers for which we have no intuition [84], but this is hardly adequate and supercomputer numbercrunching would be little advance from this point of view. In fact the demise of perturbation theory is not a serious prospect: different approaches will sometimes be competitive, sometimes complementary and it is not a complete capitulation to admit that one question cannot finally be decided by one of them. Furthermore the demarcation between perturbative and non-perturbative physics is rather fluid [85], so it is not entirely paradoxical to use properties of the series, normally their large-order behaviour, in an attempt to understand things outside perturbation theory. Working within perturbation theory, an understanding of how this approximation breaks down internally can teach us something external to it. As we shall see, most of the traffic normally flows in the opposite direction, but there are precedents (notably the diagrammatics that led to renormalons) for this sort of argument. Later chapters explore this sort of border territory further by proposing an hypothesis which can be formulated order-by-order, yet which has non-pertubative consequences; for the moment the focus is on relatively conventional approaches.

Regardless of one's attitude to the testing of QCD, questions of convergence and large-order behaviour encroach. Some of these are abstract and can probably be ignored in applying Feynman rules in new and ingenious ways to pressing problems, but it is probably as well to be reminded occasionally that our confidence in perturbative techniques at least partly consists of an act of faith – it remains to be proved that the series are even asymptotic, to be manipulated at will. If they are there is still the challenge of deciding how reliable they are, particularly given

the standard wisdom that their accuracy will begin to deteriorate after about $1/\alpha$ terms. The prospect of calculating, say, ten terms in a QCD series is unimaginably remote, but this is less so with five or six and at present we cannot be certain that such a staggeringly ambitious project would be worth embarking on if the intention is to improve the approximation. To ensure that the results would be shedding light on experiment requires an understanding of the series as a whole.

A wholly spurious approach to relating large and small order behaviours in an important quantity has recently been proposed by G.B.West [86], the details and faults of which it will be necessary to discuss in section 4.5, not least because its success would have resulted in the replacement of current techniques by large-order ones even in very low order calculations, a claim far more extreme than any we intend to make. Understanding the asymptotia of perturbation series can only ever be complementary to summing diagrams, not a substitute.

Finally, in at least one instance an acute phenomenological difficulty has led to a radical hypothesis about convergence. Influenced by his work on scheme dependence and his Principle of Minimal Sensitivity, Stevenson [87] argued that possible divergences in the perturbation series are irrelevant because we should actually only be concerned about the limit of a sequence of numerical approximations to a physical quantity rather than that of a set of partial sums. In any PMS-style approach we would be choosing a different scheme at each order in perturbation theory and using the truncated series – the full version of which is possibly divergent – in this temporary scheme to estimate an observable. The sequence of approximations is indeed a set of partial sums, but *partial sums of different series.* Our conventional intuition having broken down, he suggested that it might be possible for this sequence to converge to a sensible answer, which might or might not resemble the full non-perturbative result, and discussed several toy models mimicing this behaviour. Of more significance here – and indeed largely incidental to his new scenario – he went on to clearly highlight loopholes in the standard case that there are divergences in QCD series arising from vacuum effects.

4.2 Vacuum Instability Arguments

By a large margin the simplest and most physical argument claiming to prove

that perturbation series diverge is that devised by Dyson [88] for the case of QED. Any observable

$$R(a) = a(1 + r_1 a + r_2 a^2 + \ldots) \tag{4.1}$$

is an expansion in a couplant proportional to e^2, the square of the electron charge, and so for $a > 0$ we are dealing with a situation corresponding to conventional electrodynamics in which like-charges repel with a Coulomb force proportional to e^2. This theory we presume to be consistent, if only because of its relation to classical physics and our macroscopic experience. However, $a < 0$ implies an unfamiliar theory of attracting like-charges whose vacuum is catastrophically unstable: energy can be invested in pair-creating electrons and positrons in order to extract the unlimited return as separating clumps of exclusively particles or antiparticles implode. Something so drastic is happening around $a = 0$ that it is concluded that the function $R(a)$ is non-analytic there and hence has an expansion with zero radius of convergence about that point.

Apart from certain specific criticisms of its details – when $e^2 < 0$ the resultant complex charges force the Hamiltonian to be non-hermitean in contravention to the axioms of quantum theory – this argument displays failings general to most proofs of divergence and is thus worth analysing in some detail.

An immediate problem is that of RS dependence. At first sight the existence of schemes in which, say, $R(a) = a(1 + a)$ and is thus perfectly convergent appears to completely undercut the conclusion, but this is no more than an indication that the question has been badly formulated. Dealing with observable physics and largely relient on classical intuition, the argument cannot address such an RS dependent issue as how a particular series behaves. As we have seen (section 2.3) the perturbative expansion of an observable cannot be quoted in isolation, the corresponding β-function also being required as a specification of the scheme – although both of these can be summarised by a set of invariant ρ_n – and it is because there is more to QED than expansions of physical quantities that the proof should not directly associate discontinuites in the physics with expansion (4.1). Thus we could harmonise a convergent $R(a)$ with the non-analyticity at $a = 0$ by supposing it to partner some badly divergent β-function series, a possibility which undermines any attempt to conclude something about a scheme, e.g. $\overline{MS}$, where our ignorance of its β-function

is virtually complete.

An altogether more forceful criticism [87] also relies on realising that the full field theory has more content than its perturbative expression, although now in a rather different sense. The only formal mathematical step in the proof is the final one deducing that the series has zero radius of convergence because the function is non-analytic at $a = 0$. This is invalid, as the following simple example shows.

$$R(a) = \ln a . \exp(\frac{1}{a}(1 - e^{1/a})) + a(1 + r_1 a + r_2 a^2 + \ldots). \tag{4.2}$$

Despite having an essential singularity at the origin, this $R(a)$ could still be a function in QED and have a convergent expansion of the form (4.1) since the term containing the singularity is invisible in perturbation theory with it and all its derivatives vanishing as $a \to 0^+$. Although in this eventuality we could use the series to define an analytic function, this wouldn't necessarily be $R(a)$. Of course the existence of a non-perturbative term such as that in (4.2) is only to be expected from any realistic field theory, where an instanton could be responsible for an $e^{-1/a}$ contribution. On this point it is essential to re-emphasize the distinction between a function and its perturbative expansion. A function can be non-analytic at the origin, but have a convergent expansion which sums to give a different function; or have a divergent one from which the original function can be reconstructed; or have a divergent one telling us nothing about its source. All of these possibilities are consistent with the Dyson argument.

These objections retain their force when this informal proof is developed into detailed estimates of the large-order behaviour by showing that the non-analyticity takes the form in QCD of an $ie^{-1/|a|}$ cut along the negative a-axis (additional singularities close to $a = 0$ cannot be excluded, yet could invalidate any conclusion), a contour integral around which can be used to obtain the series coefficients, again assuming that this non-analyticity is visible in perturbation theory. Increasing mathematical sophistication and rigour brings with it the price that in QCD, to date at least, none of the firm conclusions relate to physical quantities, typically being applicable only to Green's functions, so their relevance is unclear. This is less of a restriction in $\phi^4(0+1)$ and Lipatov [89] pioneered these techniques by obtaining

$$c_n \sim (-1)^n n! n^{7/2} \tag{4.3}$$

as the behaviour of its momentum subtraction β-function terms for large n. The path integral and saddle-point methods needed to extract asymptotic behaviours are now highly refined [90] and beyond the scope of this discussion – although the proof to be criticised in section 4.5 mimics certain aspects in one-dimension – so we merely remark on these general reservations.

4.3 Borel Summation and Recoverability

Because saddle-point results invariably suggest that divergences in ϕ^4 and QED are alternating factorial (i.e. $\sim (-1)^n n!$) and those in QCD fixed factorial ($\sim n!$), alternative approaches are often formulated using Borel transforms. Given any function $R(a)$ and its series expansion

$$\sum_{n=0}^{\infty} r_n a^{n+1}, \tag{4.4}$$

one can define a Borel transform

$$F_B(z) \equiv \sum_{n=0}^{\infty} \frac{r_n z^n}{n!} \tag{4.5}$$

and a formal Borel integal

$$R_B(a) \equiv \int_0^{\infty} e^{-z/a} F_B(z) dz. \tag{4.6}$$

Even when the original series has zero radius of convergence, the transform (4.5) may have a finite one allowing (4.6) to exist by analytically continuing $F_B(z)$ to the whole real line. In such circumstances $R_B(a)$ is a plausible, and sometimes unique, sum of (4.4) in the sense of being a candidate for the function $R(a)$ it originally came from. For this reason the Borel transform has often been looked to as a means of reconstructing physical quantities from their divergent expansions and it may indeed work in ϕ^4 or QED since $r_n = (-1)^n n!$ has the well-defined Borel sum

$$R_B(a) = \int_0^{\infty} \frac{e^{-z/a}}{(1+z)} dz \tag{4.7}$$

$$= -e^{1/a} Ei(-1/a) \tag{4.8}$$

when expressed in terms of the exponential integral function [91]. Unfortunately, in the fixed factorial case the denominator in the above equation becomes $(1-z)$

and the sum is ill-defined, so the prospects in QCD are less promising. However, a converse holds and it is sufficient to show that $F_B(z)$ has a pole singularity on the positive z-axis to prove that the series (4.4) diverges fixed factorially; in fact each pole at x (> 0) contributes

$$r_n \sim \frac{n!}{x^n}, \tag{4.9}$$

so although the pole nearest the origin is sufficient, each subsequent one gives a sub-leading divergence.

In the functional integral approach each saddle point corresponds to a finite action classical solution of the Euclidean field equations, i.e. an instanton [92], and the standard results imply that in QCD a sequence of singularities is strung out along the positive z-axis in the Borel transform plane, the most significant one arising from the instanton solution of smallest action. However, these results strictly refer to the Borel-Laplace transform $F_{BL}(z)$ [87] defined via

$$R(a) = \int_0^\infty e^{-z/a} F_{BL}(z) dz \tag{4.10}$$

rather than $F_B(z)$ introduced in (4.5) and, as the example

$$F_{BL}(z) = e^z + \delta(1 - z), \tag{4.11}$$

giving

$$R(a) = \frac{a}{(1-a)} + a^{-1} e^{-1/a} \tag{4.12}$$

and thus

$$F_B(z) = e^z \tag{4.13}$$

demonstrates, a singularity in $F_{BL}(z)$ need not force $F_B(z)$ to have one or $R(a)$ to have a divergent expansion. This is merely a reiteration of the earlier observation that perturbatively invisible terms can undermine conventional arguments.

A different approach was attempted by 't Hooft [93] making use of the fact that very general constraints such as causality and the existence of hadronic bound states allow us to deduce information about the analyticity structure of Green's functions

in the complex momentum plane. Mapping this structure onto the a-plane, he claimed to show that the perturbation series for these functions must diverge, but in fact only proved [87] that the functions are not "Borel-recoverable" from their expansions, i.e. that

$$R(a) \neq R_B(a) \tag{4.14}$$

provided $R_B(a)$ is defined, or equivalently

$$F_B(z) \neq F_{BL}(z). \tag{4.15}$$

A possible reason for this is indeed that the series diverges so as to give $F_B(z)$ a singularity and render $R_B(a)$ ill-defined, but without further information the existence of non-perturbative terms – and much of the physics input to his argument involved bound states – is an equally possible explanation, c.f. (4.11) – (4.13).

4.4 Renormalons

However great their historical and technical significance, instantons are still essentially a classical artefact. They are also probably not responsible for the dominant singularity. This honour is reserved for the *renormalons* which can only appear in a quantised theory and about which much less is known – it is not even clear that they can exist in QCD. However, significant progress has very recently been made [94][95] and a great deal may be discovered in the near future.

In an infrared-free theory the situation is much more clear-cut and it was in QED that these divergences were first identified by Lautrup [96] and 't Hooft [93] while considering the n-bubble diagram contributing to the anomalous magnetic moment in nth order. That moment is associated by a dispersion relation to the vacuum polarisation $\Pi(k^2)$ and so the sum of these contributions is

$$A(a) = a \sum_{n=0}^{\infty} a^n \int_0^1 dx(1-x)\Big(-\Pi_2\Big(\frac{-x^2}{1-x}m^2\Big)\Big)^n, \tag{4.16}$$

the second-order vacuum polarisation $\Pi_2(k^2)$ having a definite functional form. Each term in the series can now be estimated and it turns out that at the nth order this one diagram will contribute $\sim n!$ to the coefficient. This is in striking contrast to the

instanton $n!$ which arises from the combinatorial increase in the number of diagrams instead of from a single graph. Summing the above series geometrically,

$$A(a) = a \int_0^1 dx \frac{1-x}{1 + a\Pi_2\left(\frac{-x^2}{1-x}m^2\right)}. \tag{4.17}$$

The structure of Π_2 is such that there is a pole for $a > 0$ and so $A(a)$ has a cut on the positive real axis. This same bubble summation (in the photon propagator) is directly responsible for QED having a Landau pole at high energies. Loosely this pole means that $A(a)$ is ambiguous at very high momenta; in Borel transform language the $n!$ produces a singularity with the consequence that in reconstructing $A(a)$ from the series the contour in (4.6) must be slightly distorted, introducing a term $\sim e^{-1/a}$ the details of which depend on precisely what contour is chosen. Resolving this ambiguity in Borel reconstruction is necessarily equivalent to a non-perturbative solution of the theory.

Apart from technical objections made by Litwin [97], one must have reservations about this argument, particularly because it concerns only a subset of diagrams and so could be subject to cancellations. However in a series of papers Parisi [98][99][100][101] argued that the divergence has a much deeper origin in the renormalisation. Working with a β-function $\beta(a) = a^2$ $(b < 0)$, Borel transforming the RG equation for a Green's function $\Gamma^{(n)}(p, a)$ gives

$$\left(-p\frac{\partial}{\partial p} - bz + n\right)B(p, z) = 0 \tag{4.18}$$

and hence a solution

$$B(p, z) = r(z)p^n\left(\frac{p}{\mu}\right)^{-bz} \tag{4.19}$$

for the Borel transform of that Green's function. The important piece here is the $(p/\mu)^{-bz}$ factor. If a Green's function is formed by multiplying two others, this factor propagates through the convolution theorem for products to be reproduced in the new transform. However, Green's functions are also interrelated by Schwinger-Dyson equations which are more complicated than products. For example, in ϕ^4 theory the 6-point function is related to the 4-point one by an equation of the

schematic form

$$\Gamma_6(0,a) = \int \frac{d^4p}{(p^2+m^2)^3} \left|\Gamma_4(p,-p,0,0;a)\right|^3 \tag{4.20}$$

or by its Borel transform equivalent

$$B_6(0,a) = \int \frac{d^4p}{(p^2+m^2)^3} I(p,z) \tag{4.21}$$

where in accordance with the above remarks $I(p,z) \sim (p/\mu)^{-bz}$. The integral giving B_6 is dependent for its convergence on the precise power here: as $z \to -2/b$ the integral diverges and thus the Borel transform of Γ_6 has a singularity at $z = -2/b$. Green's functions are so interdependent that a Borel singularity in one presumably appears in all the others and utilising information from all the Schwinger-Dyson equations, one finds that singularities appear at [98]

$$z = \frac{-2n}{b}, \qquad \forall n \in N, (b < 0) \tag{4.22}$$

i.e. evenly spaced along the positive real axis. These are the ultraviolet renormalons.

Here the choice of β-function isn't important and more detailed investigations show that the positions of the poles are fixed, although their strength has some dependence on at least the second β-function coefficient [85]. By using large-N expansions, Parisi and, later, others [90] were able to confirm all the above, except in the case of the $-2/b$ pole derived above where a conspiracy between diagrams occurs cancelling it [93]. This is obviously worrying, but there is no evidence for anything similar elsewhere. Moreover Parisi [99] derived a result connecting each pole with a local operator in the theory, thereby putting the subject on a firmer footing.

Although derived for an IR free theory, all these results survive in QCD with one major modification: the sign of b changes and so now all the poles lie on the negative real axis of the Borel plane where they do not effect large-order behaviours. A new set of poles, the infra-red renormalons, may appear in the old positions however. As their name suggests these are presumed to be produced by non-perturbative effects at low energy, but their status is much more controversial than their predecessors. In particular, Parisi's theorem [99] about operators crucially depended

on the Bogoliubov-Parasuik-Hepp (BPH) theorem which states that all UV divergences can be removed by counterterms of local operators, so in speculating about asymptotically-free theories he was forced to conjecture that a similar result holds for IR divergences [101]. This would necessarily involve non-local counterterms and its significance, even if true, remains a subject of debate [102].

A general reservation [103] about any such claim concerns the distinction between analytic and non-analytic RS's. Although both couplant a and transform $F_B(z)$ are RS-dependent, the Borel sum (4.6) is an invariant, at least formally. This however assumes that one can actually do the integral, which in turn requires that the couplant is defined, and this only happens if the β-function is analytic. Schemes like $\overline{MS}$ are probably non-analytic,* while the simplest example of an analytic scheme is that introduced by 't Hooft [93] where $c_k = 0$, $k \geq 2$, and the all-orders couplant is given by

$$\tau = \frac{1}{a} + c \ln \frac{ca}{1+ca}. \tag{4.23}$$

Unfortunately, one's intuition is that in applying a generalised BPH theorem the subtraction procedure is liable to produce a non-analytic β-function. The problem is that although the presence of a pole in the Borel transform derived in an analytic scheme guarantees the existence of the same pole in any other analytic scheme, this need not be the case for a non-analytic one. Thus the standard arguments need not be telling us anything about the actual Borel summability of QCD.

In this context, the realisation that there may be a simpler way of deriving the renormalon singularities is of particular significance. Developed in its original form by Brown and Yaffe [94], the proof considers a scalar function $f(-t)$ analytic across most of the $t = -q^2$ plane, such that in the deep Euclidean region, $t \to \infty$, it has an expansion in the running couplant $a(-t)$

$$f(-t) = a(-t) \sum_{n=0}^{\infty} \tilde{f}_n a(-t)^n \tag{4.24}$$

with real coefficients $\tilde{f}_n$. Analytically continue this from t real and negative to

* This is not to suggest that $\alpha_{\overline{MS}}$ is meaningless. Merely that within perturbation theory not enough is known about the $\overline{MS}$ scheme for it to form the basis for a summation.

$t = s + i0^+$, with s real and positive, the result of which will have an expansion

$$f(s) = a(s) \sum_{n=0}^{\infty} f_n a(s)^n. \tag{4.25}$$

In the simplest – though unphysical since it assumes $N_f = 153/19$ – RS the β-function is $\beta(a) = a^2$ and the running coupling is given by

$$a(-t) = \frac{a(\mu^2)}{1 + a(\mu^2)\frac{b}{2}\ln(-t/\mu^2)} \tag{4.26}$$

so the couplants before and after continuation are related via

$$\frac{1}{a(-t)} = \frac{b}{2}\ln\left(\frac{-t}{s}\right) + \frac{1}{a(s)} \tag{4.27}$$

$$-t = se^{-i\pi}.$$

This can be used to re-express (4.24) as an expansion in $a(s)$ which can be compared with (4.25) in order to deduce the f_n coefficients in terms of the $\tilde{f}_n$ ones. Athough this requires a messily complicated formula, Brown and Yaffe realised that the connection could be more elegantly expressed by introducing a dummy variable z:

$$\sum_{m=0}^{\infty} \frac{f_m z^m}{m!} = \exp\left(\frac{i\pi b}{2} z\right) \sum_{m=0}^{\infty} \frac{\tilde{f}_m z^m}{m!} \tag{4.28}$$

which is evidently a relation between the Borel transforms for the expansions of f in the physical and deep Euclidean regions. Extracting its imaginary part, recalling that the $\tilde{f}_n$ are real and introducing the notation $B[f](z)$ for the Borel transform of f, we find [94]

$$B[f(-t)](z) = \frac{B[Im(f(s))](z)}{\sin\frac{\pi b}{2}z}. \tag{4.29}$$

Thus unless the numerator on the right has a set of appropriate zeros, the Borel transform of $f(-t)$ has a set of poles in exactly the same places as predicted for the renormalons.

This conclusion must be immediately tempered with caution. One reaction is that the argument is trivial, akin to defining a function

$$G(z) \equiv (1-z)B[f(-t)](z) \tag{4.30}$$

and claiming that the Borel transform of $f(-t)$ has a pole at $z = 1$ unless $G(z)$ has a compensating zero there, so it should be emphasised that [94] derived the (essentially equivalent) result in the particular case where f is Π, the scalar part of the photon polarisation tensor, and $Imf = R/12\pi$, the normalised total hadronic e^+e^- cross-section, neither of which is artificial. Furthermore there is a precedent for this general type of result as part of existing approaches [98].

Another reaction is to note that (4.29) suggests another pole, at $z = 2/b$, where no renormalon is expected, there being no gauge-invariant local operator of dimension 2. The simplest resolution of this is just to assume that the Borel transform of R has a compensating zero there; this is itself new information, but there are no obvious phenomenological consequences. If however the pole is real, this could dramatically alter assumptions about the non-perturbative behaviour of QCD, the dominant non-perturbative corrections now falling as $1/Q^2$ rather than the $1/Q^4$ expected hitherto [85][95][104].

Superficially, the generalisation of (4.28) when using a less artificial β-function is complicated and in pursuing this Brown, Yaffe and Zhai [105] had to invent a version of the problem which treats the perturbation series as a vector in a Hilbert space to which ladder operators are applied. Arguing that a 't Hooft style scheme like (4.23) embodies all the essential physics, they adopt

$$\beta(a) = \frac{a^2}{1-ca} \tag{4.31}$$

for convenience. Even so, the details are still tortuous and their success depends on defining a modified Borel transform

$$\mathcal{F}(z) = \sum_{n=0}^{\infty} \frac{\Gamma(1+cz)}{\Gamma(n+1+cz)} f_n z^n \tag{4.32}$$

for which

$$f(a) = (1 - ca) \int_0^\infty \mathcal{F}(z) \frac{e^{-z/a}}{\Gamma(1 + cz)} \left(\frac{z}{a}\right)^{cz} dz. \tag{4.33}$$

An equivalent of (4.29) now holds for $\mathcal{F}(z)$ instead of $F_B(z)$, but it can be shown that any poles in the new transform are reproduced in the Borel one with the same positions. Furthermore, the strength of these poles is slightly modified in going from one transform to the other so that

$$F_B(z) \sim \frac{1}{(z - \lambda_i)^{1+c\lambda_i}}, \qquad z \simeq \lambda_i \tag{4.34}$$

in agreement with a QCD result previously derived heuristically by Mueller [85] for all renormalon poles in the theory. Note however that this particular agreement assumes that the transform on the right of (4.29) contains no poles – if it contains renormalon poles these strengths are altered. Thus the result is suggestive of Mueller's, but would also imply that the latter could only apply in certain circumstances.

Interesting though they are, the lengthy technical details of these current derivations only serve to seriously obscure the simplicity of what is happening. Everything is much more transparent if the Borel-Laplace transform or a modification thereof is used instead. Returning for clarity to $\beta(a) = a^2$, if (4.25) has transform $F_{BL}(z)$, then

$$f(s) = \int_0^\infty F_{BL}(z) e^{-z/a(s)} dz \tag{4.35}$$

$$= \int_0^\infty e^{\frac{i\pi b}{2}} F_{BL}(z) e^{-z/a(-t)} dz \tag{4.36}$$

the latter using (4.27), and this must also equal $f(-t)$ since it cannot matter non-perturbatively in which regime we choose to expand perturbatively. The crucial exponential factor is thus trivial. How this generalises is almost equally so: defining

$$G(a) = b \int^a \frac{da}{\beta(a)} \tag{4.37}$$

so that now (4.27) becomes

$$G(a(-t)) - G(a(s)) = -i\pi \tag{4.38}$$

and a new transform

$$f(a) = \int_0^\infty F(z)e^{-G(a)z}dz, \tag{4.39}$$

the same exponential factor appears. Equation (4.33) is just this new transform with (4.31) as the β-function; only the way a and z are tied up together in this is important and additional factors of either alone are irrelevant.

In this light, what is remarkable about [105] is not the derivation of the main result, but the properties of the modified transform, in particular that its poles remain fixed. This is a consequence of a formula [105]

$$F_B(z) = \oint \frac{dy}{2\pi i y}\left(1 + \frac{z/y}{(1-z/y)^{1+cy}}\right)\mathcal{F}(y) \tag{4.40}$$

relating the two transforms. A plausible conjecture is that a similar result holds when $\mathcal{F}(z)$ is replaced by the general transform (4.39), provided $\beta(a)$ is analytic. It is equally important that the new transform has a perturbative realisation like (4.32). In the Brown, Yaffe and Zhai proof, this is the foundation and (4.33) really only has the role of re-expressing the results in terms of the Borel transform via (4.40). Slight care is needed in using the Borel-Laplace transform route above since (4.35) and (4.36) strictly aren't defined when poles are present. However by restricting oneself to the cases where there are no non-perturbative terms and $F_{BL}(z) = F_B(z)$ one easily and rigourously derives a result in terms of F_{BL} which can be reinterpreted as a result like (4.28) concerning coefficients *which will remain true* even when the Borel-Laplace representation is ill-defined.

Simplifying the proof to this extent sharpens the question of what all this has to do with renormalons. These poles certainly appear indistinguishable from them, but do not obviously relate to the same physics. It is just conceivable that Π is a special choice and that, contrary to the standard arguments, its poles are somehow isolated and do not appear in the transforms of other functions in the theory. Issues like this have yet to be addressed, but there is certainly the prospect that when they are a better understanding of renormalons in general and a firmer basis for discussing their properties in QCD will result.

4.5 A Failed Approach

Brown and Yaffe originally derived (4.29) for Π and R while examining a controversial claim by West [86][106] that not only could the large-order behaviour of an expansion be determined using only momentum analyticity and renormalisation group invariance, but that the same answer could be highly competitive with existing low-order Feynman diagram calculations. The example involved the R-ratio of QCD discussed in section 3.3, where its expansion was seen to be of the form

$$R\left(\frac{q^2}{\mu^2}, a(\mu)\right) = (\sum Q_i^2) \sum_{n=0}^{\infty} r_n a^n, \tag{4.41}$$

West's asymptotic estimate for the coefficients being

$$r_n \sim -\frac{e^{1+b'}}{\pi}(\frac{eb}{2})^{n-1}\frac{\Gamma(n+b')}{(n+b')^2}, \qquad b' = \frac{2c}{b}. \tag{4.42}$$

When $N_f = 5$, this gives

$$r_3 \simeq -13.4 \tag{4.43}$$

with a claimed uncertainty of only 20%, to be compared with the

$$r_3 = -12.8 \tag{4.44}$$

of the full calculation [66]. By any standards this is startling: (4.44) involved several years of confusion * and hundred of hours of computer algebra spent in evaluating the graphs, while (4.42) required only a few pages of algebra. If justified, this agreement would render 40 years of calculational expertise obsolete and allow unparalled opportunities in applying perturbation theory.

Unfortunately this attempt to extrapolate an asymptotic estimate down to low orders dissolves with closer inspection. Although $N_f = 5$ is the number of flavours

* The original calculation of the 3-loop result [107] contained an error and gave 64.9 as the answer instead of the current -12.8, provoking the "large-coefficient" crisis. As an estimate (4.43) is all the more remarkable for having been published [106] prior to the correction of the exact result.

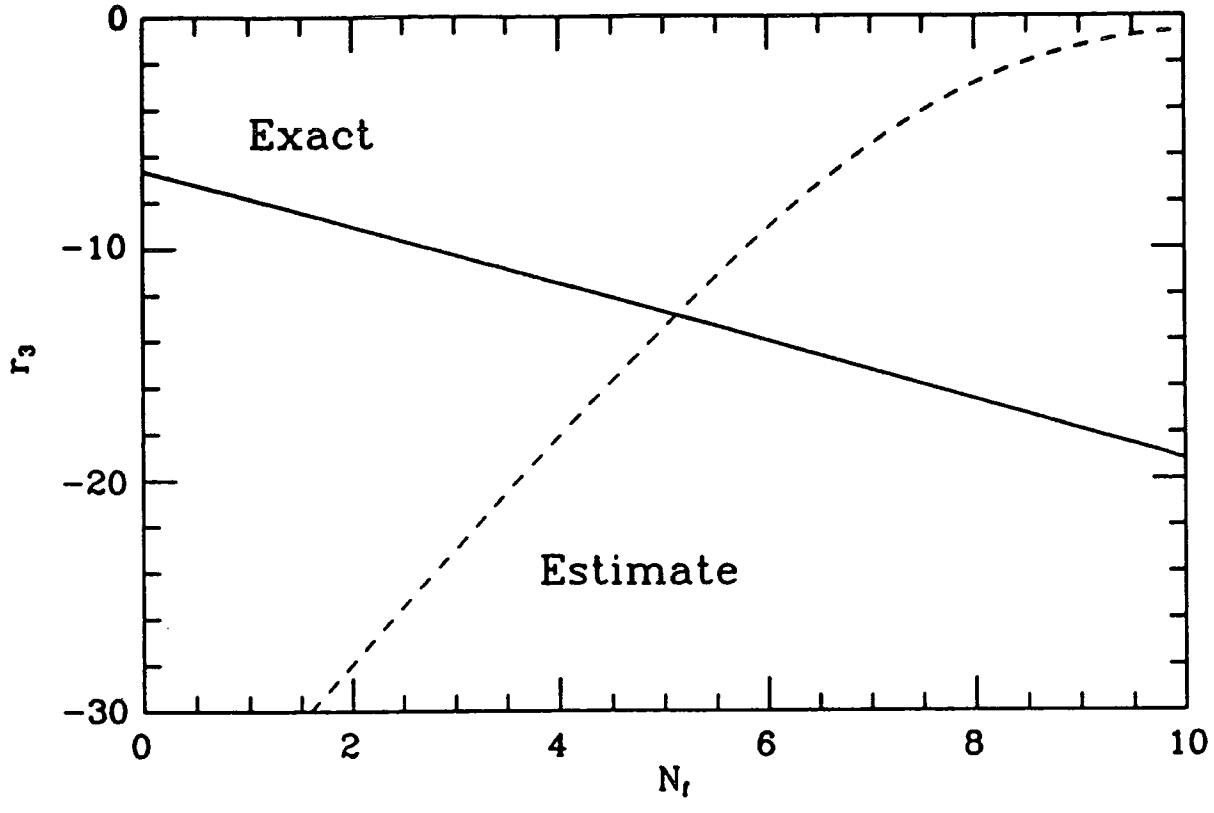

Figure 4.1 N_f dependences of exact result (4.45) and estimate (4.42) for r_3.

presently relevant at LEP and FNAL, we would expect the success of (4.43) as an approximation not to be dependent on an historical happenstance [108], [109], [110]: for arbitrary N_f the exact 3-loop result is

$$r_3 = (-6.637 - 1.200N_f - 0.005N_f^2), \tag{4.45}$$

omitting the negligible "light-by-light" term proportional to $(\sum Q)^2$, while the N_f dependence of (4.43) is easily found using (2.51) and (3.4), and these are compared in Figure 4.1. The excellent agreement for $N_f = 5$ is thus seen to be fortuitous – the two curves just happen to intersect near to $N_f = 5$. Although less physically motivated in that it is equivalent to completely changing the theory, one can also vary the number of colours, N_C, to similar effect [110]

Of course it might still be that for sufficiently large n (4.42) is the correct asymptotic result – as we will see, [86] grossly underestimates the first ($\sim 1/n$) correction term to (4.43) hence the disagreement in Figure 4.1 need not be reproduced for

$n \gg 3$. This is West's current position [111]. However, the derivation behind (4.42) is completely invalid and for a surprising reason. In order to explain why, it is first necessary to outline that derivation. West works with a fixed coupling $g^2 = 4\pi^2 a(\mu^2)$ and a β-function

$$\beta(g) = -g^3(b_1 + b_2 g^2 + \ldots) \tag{4.46}$$

whose coefficients correspond to

$$b_1 = \frac{b}{8\pi^2}, \qquad b_{n+1} = \frac{b_1 c_n}{(4\pi^2)^n} \tag{4.47}$$

in our usual notation. The claimed 20% uncertainty is obtained by retaining $O(b_2/b_1^2 n)$ terms while estimating the neglected scheme $O((b_3/b_1^3 n^2)$ ones. In $\overline{MS}$ for $N_f = 5$, where for $n = 3$ this neglects

$$\frac{b_3}{9b_1^3} = \frac{4c_2}{9b^2} \simeq 0.05 \tag{4.48}$$

compared to

$$\frac{b_2}{3b_1^2} = \frac{2c}{3b} \simeq 0.23 \tag{4.49}$$

and so appears justified.

From the photon polarisation tensor

$$\Pi_{\mu\nu}(q^2) = i \int d^4x e^{iq.x} < 0|T[j_\mu(x) j_\nu(0)]|0 > \tag{4.50}$$

$$= (q^2 g_{\mu\nu} - q_\mu q_\nu)\Pi(q^2) \tag{4.51}$$

we can find $R = 12\pi Im\Pi$ using the Optical Theorem and also introduce another function

$$D(q^2/\mu^2, g^2) \equiv q^2 \frac{\partial \Pi}{\partial q^2}. \tag{4.52}$$

Eliminating Π from these equations

$$ImD = \frac{g\beta(g)}{12\pi} \frac{\partial R}{\partial g^2} \tag{4.53}$$

so if D is expanded as

$$D(q^2/\mu^2, g^2) = \sum_{n=0}^{\infty} (-1)^n d_n (q^2/\mu^2) g^{2n} \tag{4.54}$$

the coefficients of the two series are related by

$$r_n = -(-4\pi^2)^{n+1} \frac{3}{n\pi b_1} [Im d_{n+1} + \frac{b_2}{b_1} Im d_n + \ldots]. \tag{4.55}$$

The actual physics of the argument contains only two assumptions, the first of which is the RG invariance of R under changes in μ, so that

$$R(q^2/\mu^2, a(\mu)) = f(z), \qquad z \equiv \frac{q^2}{\mu^2} e^{2K(g)} \tag{4.56}$$

where

$$2K(g) \equiv \int^g \frac{dg'}{\beta(g')}. \tag{4.57}$$

As in the previous section, the scheme in which $b_n = 0$, $n > 1$ (i.e. $c = 0$ and $c_n = 0$, $n > 1$) and $2K(g) = 1/b_1 g^2$ will be of particular importance.

Acceptance of the standard results [112] on the analyticity of $\Pi(q^2)$ is the other physics input. Causality is sufficiently constraining to be able to restrict any non-analyticity in the complex q^2-plane to the postive real axis, allowing the dispersion relation

$$\Pi(q^2) = \frac{1}{2\pi i} \int_C \frac{dq'^2}{(q'^2 - q^2)} \Pi(q'^2). \tag{4.58}$$

In terms of R and D this becomes

$$D(q^2/\mu^2, g^2) = \frac{q^2}{12\pi^2} \int_0^{\infty} \frac{dq'^2}{(q'^2 - q^2)^2} R(q'^2/\mu^2, g^2) \tag{4.59}$$

and – suppressing doubts about the legitimacy of this last equation – changing variables to z

$$D(q^2/\mu^2, g^2) = \frac{q^2}{\mu^2} e^{2K(g)} \int_0^{\infty} \frac{dz}{12\pi^2} \frac{f(z)}{[z - \frac{q^2}{\mu^2} e^{2K(g)}]^2}. \tag{4.60}$$

The strategy is to use this representation to get information on the asymptotic behaviour of the $Im d_n$, and hence of the r_n via (4.55). To this end one uses a Mellin

representation to continue $d_n(q^2/\mu^2)$ to $d(n, q^2/\mu^2)$ for complex values, by writing

$$d(s, q^2/\mu^2) = \int_C \frac{dg^2}{2\pi i}(-g^2)^{-1-s} D(q^2/\mu^2, g^2). \tag{4.61}$$

Using the representaion (4.60) for D, and interchanging the order of the integrations one has

$$d(s, q^2/\mu^2) = \int_0^\infty \frac{dz}{12\pi^2} f(z) \int_C \frac{dg^2}{2\pi i} \frac{(-g^2)^{-1-s} e^{t+2K(g)}}{(z - e^{t+2K(g)})^2}. \tag{4.62}$$

Here we have written $t = \ln(q^2/\mu^2)$. One then transforms this contour integral to the $k = 1/g^2$ plane and uses a saddle point technique to estimate

$$\int_{C'} \frac{dk}{2\pi i}(-k)^{s-1} \frac{e^{t+2K(k)}}{(z - e^{t+2K(k)})^2}. \tag{4.63}$$

Here C' is an anti-clockwise contour which straddles the cut along the positive real axis arising from $(-k)^{s-1}$. The claimed result is that a z-independent saddle point at $k = k_1 \sim b_1(s-1)$ dominates and

$$d(s, q^2/\mu^2) \simeq \int_0^\infty \frac{dz}{12\pi^2} f(z) \frac{(-k_1)^{s-1}}{[2\pi\phi(k_1)]^{1/2}} \frac{e^{t+2K(k)}}{(z - e^{t+2K(k)})^2} \tag{4.64}$$

where $\phi(k)$ is a specified function of $K(k)$ and its derivatives. This result is proportional to $D(q^2/\mu^2, 1/k_1)$ from (4.60) and hence one obtains

$$d(s, q^2/\mu^2) \simeq \left[\frac{2}{\pi\phi(k_1)}\right]^{1/2} k_1^{s-1} D(q^2/\mu^2, 1/k_1) \cos \pi s. \tag{4.65}$$

Taking the $s \to \infty$ limit of this,using (4.55) and converting back to our normal conventions leads to the estimate of equation (4.42).

In assessing the validity of this proof several approaches have been used. Brown and Yaffe [94] chose to ignore the details and addressed the general question of to what extent the assuptions of analyticity and RG invariance place restrictions on the relation between D and R and found nothing more powerful than (4.53), although in doing so they discovered (4.29) which can be thought of as resulting from expanding D in the deep Euclidean region. However interesting in itself, as a criticism of the argument above this approach is open to the objections that they may merely have failed to make full use of their premises or that West may have unconsciously added a crucial yet innocuous extra one.

A more decisive observation concerns the μ-dependence of the answer, although the whole question of scheme dependence in the proof is a murky one which has led to some confusion. At first sight it is unclear quite what RS has been used in the derivation since the answer only involves the scheme invariant constants b and c. However, discarding the $o(b_3/b_1^3 n^2)$ and higher terms is implicitly equivalent to defining the scheme as that in which the b_3 and higher β-function coefficients are zero, this being the scheme in which all these corrections are zero. As mentioned above, this scheme seems a good approximation to the $\overline{MS}$ one. But, contrary to earlier claims [86][106], this does not eliminate all the corrections [108]. For example, one of the final steps in the argument involves approximating $ImD(1,1/k_1)$ by the first term in its perturbative expansion

$$ImD(1,1/k_1) = Imd(2,1)\frac{1}{k_1^2} - Imd(3,1)\frac{1}{k_1^3} + o\left(\frac{1}{k_1^4}\right) \tag{4.66}$$

on the grounds that $k_1 \sim b_1(s-1)$ is large. The first correction term to this is proportional to

$$\frac{1}{n}\left(\frac{r_2}{2\pi^2} + \frac{b_2}{b_1}\right) \tag{4.67}$$

and does not involve the higher β-function coefficients. It is however scheme dependent via $r_2(\mu)$. Returning to the application of the result (4.42) at low n it is now possible to take up two distinct positions. Firstly, one can set $\mu = Q$ as prescribed in the $\overline{MS}$ scheme; then the original argument gives the good agreement for $N_f = 5$, but at the expense of a large correction when (4.67) is used with $r_2^{\overline{MS}} = 1.409$. Just this single term gives an 80% correction – significantly larger than the conservative 20% overall uncertainty claimed by West and completely destroying the validity at low n of the approximations. Alternatively, one could pick the $\mu \simeq Q/2$ such that (4.67) vanishes and (4.42) is exact. Unfortunately, there are presumably other corrections of this size involved and so quite what scheme one is picking will remain obscure if they are treated in the same way. Since there is bound to be a scheme in which (4.43) is correct, this is almost, if not quite, contentless.

The difficulties re-emerge in more serious guise asymptotically. Because of the $1/n$ factor in (4.67), this and all other corrections of this type are sub-asymptotic and so the estimate of r_n does indeed become independent of μ as $n \to \infty$. Brown,

Yaffe and Zhai [105] have pointed out that on very general grounds $d_n(q^2/\mu^2)$ cannot be independent of μ as indicated by (4.65) if the coefficients grow factorially, even once this limit has been taken. If the answer is inconsistent with RG invariance, the proof must be in error somewhere and, given the importance a corrected answer would have, the problem is now to locate and, if possible, amend that mistake. There is thus no substitute for a detailed examination of the proof such as was presented in [108], a fuller version of which we now turn to.

That something is seriously wrong inside the proof is most easily seen by evaluating (4.63) exactly for some special cases. When s is a positive integer the cut along the positive real axis necessary to render $(-k)^{s-1}$ single-valued is no longer needed and vanishes, allowing C' to be closed at infinity, thus making the integral vanish. Clearly this behaviour is not represented in its subsequent approximation

$$\frac{(-k_1)^{s-1}}{[2\pi\phi(k_1)]^{1/2}}\frac{e^{t+2K(k)}}{(z-e^{t+2K(k)})^2} \tag{4.68}$$

$$\phi(k_1)=\frac{s-1}{k_1}[\frac{K''(k_1)}{K'(k_1)}+2K'(k_1)-\frac{(s-2)}{k_1}] \tag{4.69}$$

which is non-zero for integer s. This can be confirmed in the specific case when $b_2=0$ and K is given by (4.57) as the integral can then be evaluated for a general s in terms of a generalised Riemann zeta function [113] to obtain

$$-\frac{1}{e^t}b_1^s\Phi(ze^{-t},s,1)\frac{\Gamma(s)\sin(\pi s)}{\pi}. \tag{4.70}$$

This gives zero for integer s both due to the $\sin\pi s$ factor and because

$$\lim_{s\to n}\frac{\Phi(x,s,1)}{\Gamma(1-s)}=0, \qquad n\neq 1. \tag{4.71}$$

Although $b_2=0$ is an unphysical assumption,* it is an unremarkable one in the original argument where there is no indication of a breakdown in the limit $b_2\to 0$.

* In a forthcoming reply [111], West has objected to its use in both [108] and [94]. As should be clear, the integral vanishes regardless of what value b_2 takes, but setting $b_2=0$ illustrates this behaviour rather nicely. Since the results in [94] have been generalised to $b_2\neq 0$, this objection also fails to address the issues raised there.

Clearly the method of steepest descent is giving the wrong answer for this example. This is presumably due to the saddle point lying on the positive axis with the path of steepest descent perpendicular to the cut along it preventing the contour from passing over the saddle point as required by the method. The alternative of a contour which doubles back is obviously at odds with the entire motivation for the saddle point technique. However the difficulty lies deeper than any possible problem in estimating the integral, since the exact form of (4.63) is evidently nonsensical in implying that $d(s)$ vanishes for integer s, something only possible if R were trivial.

Three aspects of the proof arouse suspicion initially, but two of these could have been avoided in a version of the proof less faithful to West's original. One is the change of variables from g^2 to $k = 1/g^2$ between (4.62) and (50) twisting the contour C from an innocuous one simply running around the cut along the positive axis into an awkward cardioid which is not obviously distortable into the C' assummed in [86]. But rather than having used (4.61) as a definition of $d(s, q^2/\mu^2)$, it would have been possible to continue the coefficients using

$$d(s, q^2/\mu^2) \equiv \int_{C'} \frac{dk}{2\pi i}(-k)^{s-1} D(q^2/\mu^2, 1/k) \tag{4.72}$$

so that k and C' carry through without modification. To forestall objections that this integrand diverges at infinity, we point out that C' can be defined as any contour with endpoints arbitarily close together at some finite point crossed by the cut. With this continuation we no longer have the special case (4.70), although the general point about $s = n$ still pertains. Secondly, instead of relying on (4.59) with its apparently non-integrable singularity we could have introduced

$$\Delta(q^2) \equiv \frac{-iq^2}{24\pi^2} \int_\Gamma \frac{dq'^2}{(q'^2 - q^2)^2} R(q'^2) \tag{4.73}$$

$$= q^2 \frac{\partial}{\partial q^2} Im\Pi \tag{4.74}$$

in its place, where Γ is the contour around the q^2-cut in R closed at infinity.

Thus we are left with a dubious interchange in the order of the integrations between (4.61) and (4.62) . Given the frequency with which such a step is used as a key ingredient of proofs in theoretical physics this is a somewhat disconcerting

suggestion, yet one which can be made more plausible by two simple examples. Modelling the first example on the above, let

$$D(g^2) = \int_0^\infty e^{-z/g^2} f(z) dz \tag{4.75}$$

replace (4.60), where $D(g^2)$ is clearly arbitrary since $f(z)$ is just its Borel-Laplace transform, so that

$$d(s) = \int_0^\infty f(z) dz \int_C \frac{dg^2}{2\pi i} e^{-z/g^2} (-g^2)^{1-s}. \tag{4.76}$$

For $s = n$, $d(n) = 0$ in contradiction to this arbitrariness because then the contour integral has zero residue from the essential singularity at $g^2 = 0$.* Examples can be very elementary:

$$\oint_C \int_0^1 \frac{1}{(w+z)^2} dw dz \neq \int_0^1 \oint_C \frac{1}{(w+z)^2} dz dw \tag{4.77}$$

if C is a circle of radius less than one.

In this light, how reliable are similar calculations ? For instance, in his introductory review on functional methods of obtaining large orders [114] Zinn-Justin blithely interchanges integrations in simple examples. But on closer inspection it is clear that all the integrations use saddle-point approximations even when analogous to the z-integration above, so they could have been done in either order, thereby allowing a useful check on consistency and both routes indeed agree in the few examples discussed. Because it depends on the k-integration returning something recognisable enabling D to be recovered from the z-integration, West's proof necessarily lacks this cross-check. Even so, the sparse mathematical results on this topic appear weak in non-trivial contexts and, along with the failure of the saddle-point method in a simple problem, the collapse of West's proof on these grounds suggests that results in this area should be treated with more caution.

* This example, introduced in [108], is discussed by West in [111] where he suggests replacing (4.61) with a line integral along the cut. However, this is only possible when the integrand is analytic at the origin, the line integral diverging otherwise, and it is implicit in both the main argument and the example that D is non-analytic there.

4.6 Conclusion

If this chapter is to have an overall conclusion, it must indeed be that all efforts to decide whether QCD perturbation series diverge must be treated sceptically. Such is the difficulty of stating anything for certain in field theory, there is perhaps a tendency to believe that the mathematics is too nasty for the answers to be anything so helpful as convergent. Without wishing to appear too optimistic instead, it can be pointed out that there are three criteria which one should hope that any approach would eventually satisfy and yet which all the above violate in some way

1. The series concerned should be that of a physical quantity and not just a Green's function.

2. The proof should be genuinely perturbative in that it can be expressed entirely in terms of perturbative coefficients. This is to avoid the possibility of perturbatively invisible terms.

3. The divergence must be provable for an analytic β-function.

In the next chapter a model incorporating certain features of QCD jet fractions will be described and the chapter after that will show without straying from these criteria that the series contained in it diverge.

CHAPTER FIVE

The Common Effective Charge Approach

5.1 The Assumptions

This and the remaining chapters are concerned with those situations where three observables f_2, f_3 and f_4 satisfy a constraint

$$f_2 + f_3 + f_4 = 1 \tag{5.1}$$

and have some dependence on a parameter y_c such that one of them, say $f_4(y_c)$, tends to zero in a limit $y_c \to y^*$ for kinematic reasons. By making two simple assumptions it will prove possible to reduce the content of this problem to a single equation [103] and later chapters will discuss solutions to this, showing that all the perturbation series diverge (Chapter 6) and that $\Delta\rho_0$ is small for the three observables (Chapter 7).

In the interests of generality and simplicity that key equation will be derived without particular reference to the details of what these observables might be and whether the assumptions may apply. However for readers familiar with their properties we signal our intention to focus the second half of the chapter on the e^+e^- jet fractions as an obvious example of the type of observable we have in mind. Thus the notation has been chosen to be consistent with these and f_2, f_3 and f_4 will sometimes be refered to as 2, 3 and 4-jet fractions respectively. It is also for this reason that their perturbative expansions are taken to be

$$\begin{aligned} f_2 &= 1 - K_{21}a - K_{22}a^2 - K_{23}a^3 + \ldots \\ f_3 &= \quad K_{31}a + K_{32}a^2 + K_{33}a^3 + \ldots \\ f_4 &= \qquad\quad K_{42}a^2 + K_{43}a^3 + \ldots \end{aligned} \tag{5.2}$$

Each of these observables has an associated effective charge

$$\tilde{f}_2 = \frac{1 - f_2}{K_{21}} \qquad \tilde{f}_3 = \frac{f_3}{K_{31}}$$

$$\tilde{f}_4 = \sqrt{\frac{f_4}{K_{42}}} \tag{5.3}$$

and also a set of scheme invariants, denoted by $(\rho_0^{(n)}(y_c), \rho_2^{(n)}(y_c), \rho_3^{(n)}(y_c), \ldots)$ for f_n, constructed in accordance with section 3.2 from the series

$$\tilde{f}_m = \sum_{n=0}^{\infty} r_n^{(m)} a^{n+1}. \tag{5.4}$$

Naturally, each also has an EC β-function

$$\rho^{(n)}(x) = x^2(1 + cx + \rho_2^{(n)}x^2 + \ldots + \rho_m^{(n)}x^m + \ldots) \tag{5.5}$$

satisfying

$$\rho^{(n)}(\tilde{f}_n) = \frac{d\tilde{f}_n}{d\overline{a}}\beta(\overline{a}) \tag{5.6}$$

in a particular scheme with coupling $\overline{a}$ (switching notation so that a can be reserved for the coupling in a scheme used extensively below). It is important to realise that the three functions $\rho^{(2)}$, $\rho^{(3)}$ and $\rho^{(4)}$ cannot be independent. For a start they must reflect the constraint (5.1). This must be true order-by-order in perturbation theory, an observation leading to the infinite set of equations relating the coefficients of (5.2) at each order

$$K_{31} - K_{21} = 0 \tag{5.6a}$$

$$K_{32} + K_{42} - K_{22} = 0 \tag{5.6b}$$

$$K_{33} + K_{43} - K_{23} = 0 \tag{5.6c}$$

$$\ldots$$

Deriving these relations involved no specific choice of scheme and so they can be reexpressed in terms of the $\rho_k^{(n)}$ and the tree-level coefficients alone:

$$K_{31} = K_{21} \tag{5.7a}$$

$$\rho_0^{(2)} - \rho_0^{(3)} = \frac{K_{42}}{K_{21}} \tag{5.7b}$$

$$\rho_0^{(4)} = \frac{1}{2}(\rho_0^{(2)} + \rho_0^{(3)}) - \frac{c}{2} - \frac{K_{21}}{2K_{42}}(\rho_2^{(2)} - \rho_2^{(3)}) \tag{5.7c}$$

$$\ldots$$

This is the first type of interdependence amongst the ρ-functions.

The next type concerns the $f_4 \to 0$ limit which leaves $\tilde{f}_2$ and $\tilde{f}_3$ equal for values of $y_c \geq y^*$. There one has the strict, but somewhat trivial, result that

$$\rho_0^{(2)} = \rho_0^{(3)} \qquad \rho^{(2)}(x) = \rho^{(3)}(x) \tag{5.8}$$

and so continuity implies that

$$(\rho_n^{(2)} - \rho_n^{(3)}) \to 0 \qquad \text{as} \qquad f_4 \to 0, \qquad n \geq 0 \tag{5.9}$$

further restricting the form of the functions. What happens to $\rho^{(4)}$, or equivalently $\tilde{f}_4$, in this limit ? Without doing the one-loop calculation the answer must strictly be that anything could happen since

$$\tilde{f}_4 = \overline{a}(1 + \frac{K_{43}}{2K_{42}}\overline{a} + \ldots) \tag{5.10}$$

possibly entails a $1/K_{42}$ pole as $y_c \to 0$. However $K_{42} \to 0$ is a consequence of the phase-space suppression of f_4 as a whole and it is thus probable that there is an overall factor reproducing this behaviour in each coefficient and giving the set of limits

$$\frac{K_{4n}}{K_{42}} \to \text{constant} \qquad \text{as} \qquad f_4 \to 0, \qquad n \geq 3 \tag{5.11}$$

If so, and it will be assumed hereafter, both $\tilde{f}_4$ and $\rho^{(4)}$ are well-defined in this limit.

The next step is most clearly explained if a particular RS is chosen in which to take it, although this is in no way necessary. That particular scheme is the 4-jet EC one* whose couplant is $a \equiv \tilde{f}_4$, so

$$f_4 = K_{42}a^2, \tag{5.12}$$

* A footnote for those worried that no such scheme exists for $y_c \geq y^*$. Each y_c value really gives a different definition of observable f_4 and hence of the 4-jet EC scheme, labelled by $\rho^{(4)}(y_c, x)$. The single-argumented function $\rho^{(4)}(y^*, x)$ has already been assumed to exist. Should one require it for $y_c \geq y^*$, one can take the "4-jet EC" scheme to be the one labelled by this function without worrying about identifying its effective charge with an observable.

and whose β-function is $\rho^{(4)}$. Now (5.1) takes the form

$$\tilde{f}_2 = \tilde{f}_3 + \frac{K_{42}}{K_{21}}a^2 \tag{5.13}$$

while (5.6) becomes

$$\rho^{(2)}(\tilde{f}_2) = \rho^{(4)}(a)\frac{d\tilde{f}_2}{da}. \tag{5.14}$$

There is an equivalent to this for $\tilde{f}_3$, but rather than display it individually we subtract it from this one to get

$$\rho^{(2)}\left(\tilde{f}_3 + \frac{K_{42}}{K_{21}}a^2\right) - \rho^{(3)}(\tilde{f}_3) = 2a\frac{K_{42}}{K_{21}}\rho^{(4)}(a). \tag{5.15}$$

If we let $\tilde{f}_2 \to \tilde{f}_3 \to R$ as $K_{42} \to 0$ and knowing from (5.8) that $\rho^{(2)} \to \rho^{(3)}$ in the same limit, then

$$\frac{d\rho^{(3)}(R)}{dR} + \frac{G(R)}{a^2} = \frac{2\rho^{(4)}(a)}{a} \tag{5.16}$$

where

$$G(R) = K_{21}\sum_{k=2}^{\infty} R^{k+2}\left(\lim_{K_{42}\to 0}\frac{\rho_k^{(2)} - \rho_k^{(3)}}{K_{42}}\right) \tag{5.17}$$

$$= \sum_{k=2}^{\infty} R^{k+2}\left(\lim_{K_{42}\to 0}\frac{\rho_k^{(2)} - \rho_k^{(3)}}{\rho_0^{(2)} - \rho_0^{(3)}}\right) \tag{5.18}$$

using (5.7b). At present nothing whatever is known for certain about this function in any of the relevant cases and it would require the calculation of perturbation series beyond their existing orders before anything definite could be said. On the grounds of simplicity, we propose that

$$G(x) \equiv 0 \tag{5.19}$$

by adopting the following assumption: both $\rho_k^{(2)}$ and $\rho_k^{(3)}$, regarded as functions of K_{42}, vary *smoothly*, i.e. with continous first derivative, through the 4-jet threshold. A mild extension of the exact continuity requirement (5.9) on these functions, this will be easily testable against future perturbative calculations and in the meantime appears eminently plausible. The consequences on the rest of the argument of weakening this assumption will be examined in section 7.6.

This *Smoothness Assumption* is actually the first unprovable statement required; once adopted it supercedes the earlier one about the existence of $\tilde{f}_4$ in the limit $f_4 \to 0$, since equations (5.7) now determine its finite coefficients unambiguously, e.g.

$$\rho_0^{(4)} = \rho_0^{(3)} - \frac{c}{2} - \frac{K_{21}}{2} \lim_{K_{42}\to 0}\left(\frac{\rho_2^{(2)} - \rho_2^{(3)}}{K_{42}}\right) \tag{5.20}$$

is the $K_{42} \to 0$ limit of (5.7c), which now yields the formula

$$r_1^{(4)} = r_1^{(3)} + \frac{c}{2}, \tag{5.21}$$

true in any scheme. This is merely a small taste of the perturbative expression of the set of non-perturbative equations that this assumption allows one to derive. (5.16) becomes

$$\frac{d\rho^{(3)}(R)}{dR} = \frac{2\rho^{(4)}(a)}{a} \tag{5.22}$$

whereas taking $K_{42} \to 0$ in the partner of (5.14) directly gives

$$\rho^{(3)}(R) = \frac{dR}{da}\rho^{(4)}(a). \tag{5.23}$$

Eliminating $\rho^{(4)}(a)$ between these equations, one gets

$$\frac{d\rho^{(3)}(R)}{da} = \frac{2\rho^{(2)}(R)}{a} \tag{5.24}$$

which can be integrated up to yield

$$\rho^{(3)}(R) = Ca^2 \tag{5.25}$$

where the integration constant can be fixed as $C = 1$ by noting that

$$\frac{\rho^{(3)}(x)}{x^2} \to 1 \qquad \frac{R(x)}{x} \to 1 \tag{5.26}$$

when $x \to 0$. The problem is thus reduced to the pair of equations

$$\rho^{(3)}(R) = a^2 \tag{5.27}$$

$$\rho^{(4)}(a) = a^2\frac{da}{dR}, \tag{5.28}$$

so specifying any one of the functions $\rho^{(4)}$, $\rho^{(3)} \equiv \rho^{(2)}$ and $R(a)$ is sufficient to enable the other two to be deduced.

An extra constraint is clearly necessary to actually determine these functions and this is naturally the most difficult step since it replaces a full calculation of part of the theory. Because the exact details of, say, $\rho^{(4)}$ are unknown, we must resort to a plausible guess and the simplest possibility is to set

$$\rho^{(3)}(x) = \rho^{(4)}(x) \equiv \rho(x) \tag{5.29}$$

so that there is a common EC β-function $\rho(x)$. Equations (5.27) and (5.28) then determine the perturbative coefficients ρ_k uniquely to all orders, and as we shall see in Chapter 7 the function itself is (probably) also solvable non-perturbatively. Taken literally, (5.29) is therefore a very restrictive and powerful assumption, but this is not our main intention and it should really be interpreted loosely – that is, $\rho^{(4)}$ should be a "similar" function to $\rho^{(3)}$. The next chapter will discuss in some detail how, when taken literally, this *Common Effective Charge Approximation* results in a divergent perturbation series for $\rho(x)$ with (probably) the fixed-factorial growth anticipated by earlier arguments (Chapter 4). In itself this is interesting and instructive (particularly since the divergence proof accords with the criteria in section 4.6), but the ultimate question must be how robust these conclusions are if much milder assumptions are adopted. This type of issue is addressed in Chapter 7 after the explanation of how non-perturbative information can be extracted.

As indicated by (5.21) relating $r_1^{(4)}$ and $r_1^{(3)}$, $\rho_0^{(4)} \simeq \rho_0^{(3)}$ and the assumption above is thus equivalent to assuming that the effective charge for 3-jet production is similar to that for 4-jet. That is, multijet production corresponds to tree-level diagrams with an approximately common effective charge at each vertex, which at least seems physically reasonable. Given our ignorance, that all the observables are treated on an equal footing is perhaps the most appealing aspect of (5.29).

However, the main justification for the assumption is that it enables one, however tentatively, to explore unknown territory in comparing the all-orders structure of the perturbation series with a *non-perturbative* solution and in beginning to understand how the different functions relate to each other. By providing a definite example that can be worked through in detail it also provides a foundation for more ambitious forays, both here and in the future. As a (hopefully not unrealistic) model of what is happening inside perturbation theory it may even have something to tell us about

applications thereof, but it cannot seriously be expected to describe experimental details.

Combining (5.29) with (5.27) and (5.28) from the previous section produces as a straightforward consequence the equation

$$u^2(x)\frac{du}{dx} = u^2(u(x)) \tag{5.30}$$

where $u(x)$ is the function such that

$$a = u(R) \tag{5.31}$$

$$\rho(x) = u^2(x). \tag{5.32}$$

It is therefore of central importance to the Common Effective Charge Approach (CECA) and it is to this that subsequent chapters will be devoted. Anticipating their conclusions, it is $u(x)$ that is proved to have a divergent expansion in x, with the divergences in other series following from this, and for all three observables

$$\Delta\rho_0 \simeq \Delta\rho_0^{NNLO}, \tag{5.33}$$

is the same for all three and small enough to be ignored in practice.

Finally, if another scheme had been chosen in which to do the derivation, how much would have survived explicitly, apart of course from the answer, $\rho(x)$, itself ? Repeating the above analysis more generally shows that

$$\tilde{f}_4 = u(\tilde{f}_3) \qquad y_c \to y^* \tag{5.34}$$

with $u(x)$ again the solution to (5.30), as one might have guessed. However, the relationship between $\tilde{f}_4$ and the couplant in the new scheme will necessarily be more complicated.

5.2 An Example

In deriving (5.30) all the physics has been distilled into an easily stated mathematical question: what are the solutions of this equation, if any ? This is the

immediate problem to be faced, but before giving an answer it is necessary to discuss whether the starting assumptions had any physical relevance in the first place. Otherwise the exercise merely becomes one in pure mathematics. Furthermore, a concrete example may help fill out particular details of the earlier, slightly abstract, presentation. However, what follows is hereby plainly advertised as a digression and any reader who wishes to omit it can retrieve the main thread of the CECA's development at the beginning of Chapter 6. The earlier health warning must also be repeated: the purpose is to show that the CECA *could* be realistic rather than to suggest that it *is* in this particular case.

Are there any sets of observables where

$$f_2 + f_3 + f_4 = 1 \tag{5.35}$$

and $f_4 \to 0$. Simple examples with both this constraint and threshold are, for suitably defined reaction rates, $e^+e^- \to n$ photons in QED, $X^+X^- \to n$ scalars in ϕ^4_{3+1} theory, where some extra charged fermions $X^\pm$ have been coupled in [115] , and $e^+e^- \to n$ jets + hard photon in QCD. These three (and particularly the last) have definitions modelled on the jet fractions for $e^+e^- \to n$ jets as predicted by QCD and as discussed in section 3.3. When these fractions are plotted against y_c broadly the same pattern is produced by each algorithm with f_2 and f_3 monotonically rising and falling respectively over most of the range as y_c increases [70]. Apart from f_2, all the others fall successively to zero for sufficiently large y_c. These thesholds are examples of the purely kinematic ones required by the CECA and, although the precise details will depend on the choice of algorithm, it is thus worth deriving the critical $y_c = y^*$ for the 4-jet one when

$$y_{ij} = 2E_iE_j - 2\mathbf{p}_i\cdot\mathbf{p}_j \tag{5.36}$$

as it is for the E0 algorithm. Working in the centre of mass frame, suppose one has four particles $(E_i, \mathbf{p}_j)$

$$\begin{aligned} \mathbf{p}_1 + \mathbf{p}_2 + \mathbf{p}_3 + \mathbf{p}_4 &= \mathbf{0} \\ E_1 + E_2 + E_3 + E_4 &= \sqrt{s}. \end{aligned} \tag{5.37}$$

For each of these there is an identity like

$$y_{12} + y_{13} + y_{14} = 2E_1\sqrt{s} \tag{5.38}$$

so that

$$2(y_{12} + y_{13} + y_{14} + y_{23} + y_{24} + y_{34}) = 2s. \tag{5.39}$$

Since this can only be a 4-jet event provided $y_{ij} \geq y_c s$ for all y_{ij}, one finds

$$y^* = \frac{1}{6}. \tag{5.40}$$

Generalising, an n-jet event is only possible if

$$y_c \leq \frac{2}{n(n-1)}. \tag{5.41}$$

There are thus no five jet events for $y_c \geq 0.1$, no four jet ones for $y_c \geq 1/6$ and only two jet ones when $y_c \geq 1/3$. These values will be slightly different for other algorithms, but note that the superficially different Durham one has in fact the same thresholds.

Althought the CECA assumes nothing about the position of the crucial 4-jet threshold, it is required that the 4-jet effective charge is well-defined in the $y_c \to y^*$ limit, i.e. that the phase space suppression does factor out of the power series. This seems to be bourne out in the data for the observed jet fractions. Figure 5.1 shows OPAL data [70] taken at 91 GeV using the E0 algorithm in which the convergence of $\tilde{f}_2$ and $\tilde{f}_3$ as y^* is approached is clearly seen, with $\tilde{f}_4$ displaying no indications of any imminent pathology there, although the very small absolute rates close to threshold prevent any definite settlement of the issue. At this point it should be re-emphasised that the Smoothness Assumption is a priori much more "natural" than (5.29) and so serious consideration should be given to any consequences this may have on its own.

Neither experiment nor theory can for the moment falsify $G(x) = 0$ as suggested above. For instance, considering $\rho_2^{(3)} - \rho_2^{(2)}$ in the 4-jet EC scheme reveals that

$$\lim_{K_{42}\to 0} K_{21}\frac{\rho_2^{(2)} - \rho_2^{(3)}}{K_{42}} = -c - \frac{2K_{22}}{K_{21}}. \tag{5.42}$$

Even though K_{22}/K_{21} has been calculated [58], only having the f_4 series to tree-level

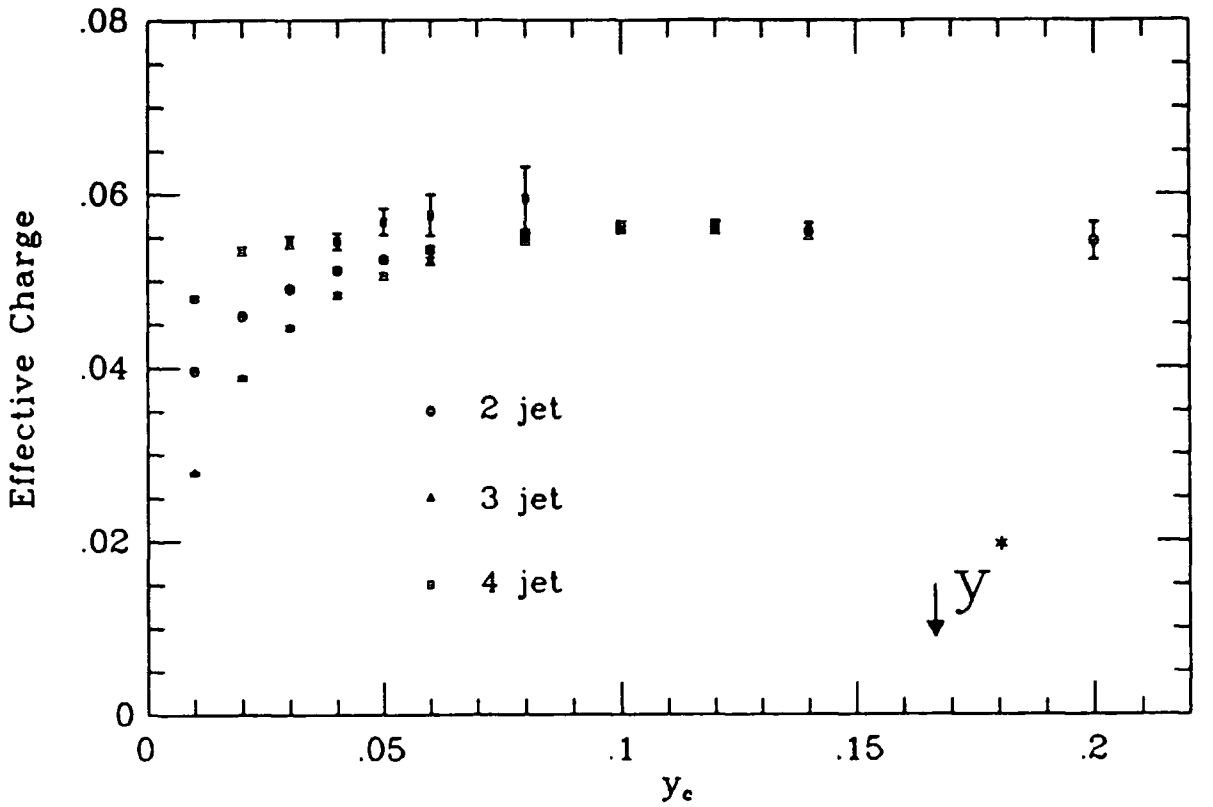

Figure 5.1 E0 effective charges

means that no-one knows which scheme this should be translated to for use in this non-invariant formula; a determination of K_{43} would however fix this first (invariant) coefficient of $G(R)$. Superficially this result will be algorithm dependent,but this is not inevitable since K_{22}/K_{21} in a specific scheme like $\overline{MS}$ *and* the definition of the 4-jet EC scheme would both change and could conceviably compensate for each other in doing so.

If $G(x) = 0$ is correct, $\rho^{(3)}(R) = a^2$ follows directly, where, as a reminder, R is just f_3/K_{31} measured at the 4-jet threshold and a is the limiting value of $\sqrt{f_4/K_{42}}$ there. As an exercise, one can plot out a against R using values obtained from jet fraction data taken at different energies. Due to the running of the quantities, this then maps out the function $\rho^{(3)}$. Doing this with data between 22 GeV [54] and 91 GeV [70] unfortunately only shows that the uncertainties are too large to distinguish the result from the simplest of all guesses $\rho^{(3)}(x) \simeq x^2$.

For completeness, we record that if the CECA were literally applied to the jet

fractions, its prediction (Chapter 7) that $\Delta\rho_0 \simeq 0$ would imply values of $\Lambda^{(5)}_{\overline{MS}}$ of the order of 150 GeV for both the E0 and the Durham algorithms. What is probably more significant is that it also predicts that

$$\Delta\rho_0 \simeq \Delta\rho_0^{NNLO} \tag{5.43}$$

encouraging the hope that this will be true for the jet fractions in general.

Although the 4-jet threshold occurs at relatively large y_c, the reader may be wondering whether the CECA has anything to do with the summation of large logarithms in y_c, particularly since they will both appear to be making statements about series to all orders in α_S. The short answer is no, but first a reservation must be expressed about the normal interpretation of these resummations as the summation of large terms in the expansion which otherwise spoil the convergence (see e.g. [61]). Since the series is probably only asymptotic, any such rearrangement is probably only formal. To take a simple example, the series

$$\begin{aligned}\sigma = 1 &+ a(A_0 + A_2L^2) \\ &+ a^2(B_0 + A_2A_0L^2 + \frac{1}{2}A_2^2L^4) \\ &+ \ldots\end{aligned} \tag{5.44}$$

where L denotes some large logarithm, could be rearranged as

$$\sigma = e^{aA_2L^2}(1 + aA_0 + a^2B_0 + \ldots). \tag{5.45}$$

If the series

$$1 + aA_0 + a^2B_0 + \ldots \tag{5.46}$$

is divergent, then so must be both of these versions. Should the exponential factor in (5.45) differ greatly from unity, summing the first few terms of one series will give a very different result from doing the same to the other. In QCD only one of these might be a good approximation – the other would start too far from the answer for enough terms to smoothly correct for this before the divergence sets in in both. For f_3 the exponential factor will be crucial at small y_c where $f_3 \ll 1$ experimentally

and (5.2) cannot have the correct starting point. In an asymptotic approximation the initial guess is much more important than the existence of large terms further down the series.

Although the resummations are invaluable in trying to reliably pick out the dominating contributions at small y_c and should thereby eliminate the major differences between large and small y_c, they fall short of describing the entire series – indeed from this point of view they are little better then a tree-level calculation – and it is for this reason that something like the CECA is necessary to tell us about the reliability of the series like (5.46) which still have to be faced.

CHAPTER SIX

Perturbative Consequences

6.1 Introduction

When presented with

$$u^2\frac{du}{dx} = u^2(u) \tag{6.1}$$

a natural reaction is to expand the function $u(x)$ as its perturbation series

$$u(x) = x\sum_{n=0}^{\infty} u_n x^n \tag{6.2}$$

and use the equation to derive a recurrence relation for the coefficients:

$$u_k = \frac{1}{k-1}\sum_{i=1}^{k-1}\{\rho_i C_{k-i}(\frac{u}{x})^i + u_i u_{k-i}\}$$

$$u_0 = 1 \quad , \quad u_1 = \frac{c}{2}$$

$$\rho_i = \sum_{j=0}^{i} u_j u_{i-j} \tag{6.3}$$

where $C_n(R)^i$ is to be interpreted as the x^n term in the series of $R(x)$ raised to the ith power obtained using Cauchy products. This chapter will have as its main concern the mathematical properties of this expansion, with the physical aspects of $u(x)$ accordingly taking a secondary role and with the non-perturbative ones ignored completely, as is the (unresolved) question of the function's existence, prior to being taken up in the next chapter. Thus (6.3) can be taken as the starting point for the discussion. As explained in Chapter 4, the received field theory belief is that such a series should be fixed-factorially divergent.

One special case should be mentioned: if $c = 0$ then the series truncates to give $u(x) = x$, which is also the only known exact analytic solution of (6.1), but which is trivial enough to be ignored and is anyway unphysical.

This aside, a few trivial observations can be made immediately. Direct calculation gives the first few terms as

$$u_0 = 1, \qquad u_1 = \frac{c}{2}, \qquad u_2 = \frac{3}{4}c^2, \qquad u_3 = \frac{13}{8}c^3, \quad \ldots \tag{6.4}$$

The c-dependence of the coefficients takes the particularly simple form

$$u_k = \omega_k c^k \tag{6.5}$$

with ω_k independent of c, as can easily be verified by induction, and so the problem reduces to that of investigating (6.3) for $c = \pm 1$. This series is an infinite one with its coefficients positive and monotonically increasing. A formal proof again proceeds by induction, noting that if the coefficients up to u_{n-1} are positive then so are all the terms in the equation for u_n and that one also has the equality

$$\omega_n > (\frac{n+1}{n-1})\omega_{n-1} \qquad n \geq 3 \tag{6.6}$$

obtained simply by dropping all terms which do not depend on ω_{n-1} from the recurrence relation.

With these lemmas in hand we can now prove one of this thesis' central results.

6.2 The Divergence Proof

To begin with it's easiest to prove the result for $c > 0$, for which it suffices to consider absolute convergence; the generalisation (given below) is then trivial.

Theorem The power series expansion of $u(x)$ has zero radius of convergence when $c > 0$.

Proof Because $u_{k+1} > cu_k$, the series cannot have an infinite radius of convergence, so suppose that it has a finite one $r > 0$. Within this radius of convergence

the series is assumed to be absolutely convergent to a sum which we denote by $U(x)$.

$$U(x) = x \sum_{n=0}^{\infty} u_n x^n. \qquad |x| < r \tag{6.7}$$

But since $u_k > 0$, $\forall k$ one has

$$U(x) > x + \frac{c}{2}x^2. \qquad 0 < x < r \tag{6.8}$$

An $x = X$ can then always be chosen such that $X < r$, but with

$$X + \frac{c}{2}X^2 > r. \tag{6.9}$$

Next recall that (6.3) was obtained from (6.1) by expanding both sides of the latter as power series and then rearranging terms. Well-known results state that differentiating a power series does not change the radius of convergence and that the Cauchy product of two series has a radius of convergence which is the smaller of the two for the original series. Applying these, the combination

$$U^2(x)\frac{dU}{dx}, \tag{6.10}$$

corresponding to the left-hand side of (6.1) expanded as a power series in x, is seen to have radius of convergence r. It is thus convergent at $x = X$. However, the right-hand side when expanded into a power series in x must be rearrangable (courtesy of the hypothesied absolute convergence) into

$$U^2(U(x)) = U^2(x)(1 + u_1 U(x) + u_2 U^2(x) + \ldots)^2 \tag{6.11}$$

which is only convergent at $x = X$, as it must be to enable the perturbation expansions on both sides of (6.1) to be equal, provided

$$U(X) < r \tag{6.12}$$

in contradiction with (6.8) and (6.9) above. The series thus cannot be absolutely convergent with either a finite or an infinite radius of convergence. •

Although a simple idea has already been laboured, two nuances of the proof require discussion. Firstly, the condition (6.12) is not universally true for all cases in which an expansion of $f(u(x))$ is under consideration, since cancellations can occur causing any one of the series involved to truncate, thereby rendering any considerations in terms of radii of convergence inappropriate. For instance, if

$$f(x) = x\sqrt{1-x} \qquad u(x) = 1 - x^2 \tag{6.13}$$

it is evidently not true that

$$|u(x)| < 1 \qquad \forall x > 0 \tag{6.14}$$

But this is not significant because $f(u(x))$ has collapsed to a polynomial in x and since the u_k are known to be positive and the expansion of $U^2(U(x))$ to be equal to the non-truncating one on the left-hand side, such a violation cannot occur here. Secondly and rather more crucially, we observe that the proof has not assumed any non-perturbative knowledge of $u(x)$, unlike the arguments discussed in Chapter 4. It may be objected that equating (6.10) and (6.11) looks suspiciously similar to the non-perturbative equality (6.1), but this is unfounded. Simply because the recurrence relation is derived from (6.1), even in the form (6.3) it must have buried inside a certain "structural" similarity to (6.1) which non-perturbative effects cannot destroy and it is this that the proof exploits. In principle everything could be directly reformulated in terms of (6.3), but such a version of the proof would only be much longer and obscure the basic idea.

As promised the series also diverges when $c < 0$. This is just a consequence of the fact that we're dealing with a power series and hence it either converges absolutely or it diverges.

6.3 $R(a)$ and $a(R)$

Knowing that $u(x)$ has a divergent expansion immediately tells one that $\rho(x) = u^2(x)$ must also have one and, as emphasised earlier (Chapter 3), ρ is a physical quantity directly measurable in experiments. However, even if a very restrictive definition of "physical quantity" is adopted which excludes ρ and only admits R, i.e. the jet fractions themselves, but not their energy variation, the criteria for

divergence proofs introduced in section 4.6 can still be adhered to, it now being straightforward to deduce that

$$R(a) = a\sum_{n=0}^{\infty} r_n a^n \tag{6.15}$$

in the 4-jet EC scheme is also a divergent expansion. Its coefficients are related to the u_n by the operation of "inversion" (sometimes "reversion"), the power series equivalent of inverting functions, and a standard result due to Cauchy states that if the power series for a function is divergent then so is the power series for the inverse of that function [116]. Although there are standard algorithms to actually calculate the coefficients of this inverted series these are complicated and the particular circumstances allow more elegant means. One obvious one is simply to set $c_i = \rho_i$, since this equality is at the heart of the CECA, in the infinite set of equations (3.17) and solve for the r_n, but in practice this is too awkward. It is much easier to expand

$$\frac{dR}{da} = \frac{a^2}{\rho(a)} \tag{6.16}$$

$$= \frac{1}{1 + \rho_1 a + \rho_2 a^2 + \ldots} \tag{6.17}$$

in powers of a to get

$$(n+1)r_n = -\sum_{k=1}^{n}(n-k+1)r_{n-k}\rho_k \tag{6.18}$$

or equivalently [117] and perhaps more elegantly

$$(n+1)r_n = (-1)^n \begin{vmatrix} \rho_1 & \rho_0 & 0 & \ldots & 0 \\ \rho_2 & \rho_1 & \rho_0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \cdots & \rho_0 \\ \rho_n & \rho_{n-1} & \rho_{n-2} & \cdots & \rho_1 \end{vmatrix}. \tag{6.19}$$

Incidentally, (6.17) could have been used to prove that $R(a)$ diverges since it is another standard theorem [116] that the radius of convergence of a reciprocal power series is equal to that of the original, unless that original has a zero in which case

it is less, so proving that $\rho(a)$ diverges is sufficient here. As an alternative way of calculating the coefficients, instead of expanding $\rho(R) = a^2$ in R as done previously, it can be expanded in powers of a to get the same answer as above. Of course, since its ρ_n are known it is only a matter of labourious calculation to find the series for R in *any* desired scheme.

Actually $u(x)$ itself corresponds to a "physical series" in the narrow sense assumed above. This is a consequence of equation (5.34) which shows that the function $u(x)$ relating the 3 and 4-jet effective charges is scheme invariant. Consider the 3-jet EC scheme instead of the 4-jet one and continue to let the 3 and 4-jet effective charges be R and a respectively; in this scheme the role of couplant and "observable" have been interchanged and the relevant expansion is now $a(R) \equiv u(R)$ instead of $R(a)$. Thus the expansion of f_4 in the 3-jet EC scheme as calculated in the CECA has zero radius of convergence.

Finally, in accordance with the last of the criteria in Chapter 4, it can be proved that the series for f_3 diverges in an analytic scheme. If the t'Hooft scheme (4.23) with couplant a' is used, then

$$a'^2(1 + ca')\frac{dR}{da'} = \rho^{(3)}(R) = u^2(R). \tag{6.20}$$

Suppose $a'(R)$ has a convergent series. Then da'/dR also has a finite radius of convergence, as does dR/da' considered as a series in R. Again the left-hand side is a convergent series in R while the other is not, giving the contradiction. If $a'(R)$ has zero radius of convergence, then so does $R(a')$.

6.4 The Large-Order Behaviour

Having somewhat indirectly proved in section 6.2 that it diverges, an obvious next step is to establish how the series for u behaves at large-orders and also just how quickly this asymptotic behaviour sets in. It is as well to state at the outset that no definite conclusion has been reached on the matter, but that the evidence is encouragingly consistent with an ultimately fixed-factorial growth. This section will be devoted to a discussion of the problems involved and to deriving the best available, if still loose, bound on the divergence and the following one to the results of numerical investigations into the coefficients.

What makes the recurrence relation (6.3) so intractable is that it relies on multiple Cauchy products and in particular the sums involved in evaluating them. If as suspected $u_n \sim n!$ at large n then at some level it has to be checked that if this is assumed for $n \leq k-1$ then the recurrence relation is consistent with it for $n = k$. However in doing this one quickly finds oneself having to deal with the summations like

$$\sum_{m=0}^{n} m!(n-m)! \tag{6.21}$$

which arise in trying to evaluate the coefficients in

$$(1 + x + 2!x^2 + \ldots + n!x^n + \ldots)^p. \tag{6.22}$$

Although it can legitimately be objected that taking the power of a divergent series like this one is an ill-defined operation, its formulation in terms of the Cauchy product is not and merely leads to a difficult problem in the manipulation of factorials, namely an adequate treatment of expressions like (6.21). Perhaps surprisingly no such treatment appears to exist.

Just restricting ourselves to (6.21), closed forms do exist. Products of factorials naturally suggest the beta-function $B(x,y)$ [14] and one is thus led to

$$\sum_{m=0}^{n} m!(n-m)! = (n+1)! \int_0^1 \frac{1-t^n}{(1+t)^{n+1}(1-t)} dt. \tag{6.23}$$

Alternatively, consideration of simple Dirichlet integrals gives

$$\sum_{m=0}^{n} m!(n-m)! = (n+2)! \int\int \frac{x^n - y^n}{x-y} x\,dx\,dy \tag{6.24}$$

the double integral being taken over a triangle in the (x, y) plane with vertices at (1,0), (0,1) and the origin. However, (6.21) relates to the $p = 2$ case of (6.22) and neither of these integrals provides any basis for the sort of iterative generalisation necessary to deal with arbitrary integer p.

Of course it is only the large-n limit of the recurrence relation (6.3) that is of interest at the moment and so it may be that, regardless of how interesting and

challenging it may be as an abstract mathematical problem, the precise result is unnecessary. Crudely one can argue that

$$\sum_{m=0}^{n} m!(n-m)! = 2n!(1+o(\frac{1}{n})) \tag{6.25}$$

with the subleading terms to be neglected in the remainder of any calculation. Unfortunately it has already been seen that naïvely extending this reasoning to (6.3) by discarding terms not including u_{n-1} (which must be "significantly" smaller) only leads to

$$u_n \sim cu_{n-1} \tag{6.26}$$

a result far too weak to be correct, if only because it indicates a convergent series. Marginal improvements on this theme are possible – picking out the $i = k-1$ term from (6.3) would suggest

$$2u_n \sim c\rho_{n-1} \tag{6.27}$$

yet $2u_n = c\rho_{n-1}$ only corresponds to the evidently convergent expansion for $u = c^{-1}(1-\sqrt{1-2xc})$, although whose radius of convergence at $1/2c$ is admittedly mildly better than in the original approximation – but to nowhere near the extent as to suggest a promising approach. One is therefore forced to the conclusion that there is rather more to the problem than the leading term during intermediate stages and having resolved to retain the first sub-leading term it becomes difficult to justify neglecting any of the others. Indeed, such a piecemeal approach again proves inadequate. In some as yet unclear fashion most of the recurrence relation (6.3) appears essential in generating the large-order behaviour.

Faced with this impasse one is perhaps tempted to waive the caution expressed in Chapter 4 and resort to applying a saddle point method to (6.1) on the grounds that *any* answer, even if possibly wrong, is better than none at all. But even with a foreknowledge of the next chapter not enough is certain about the properties of $u(x)$ to enable this to be done.

What then can be said with confidence about the large-order behaviour ? An upper bound on the series growth can be obtained and the rest of this section is

concerned with its derivation. Although based on the simple, even naïve, foundation of assuming a bound of the form $\omega_i \leq ba^i$ for part of the series, then using this to derive a weaker bound which also applies to the next coefficient up and iterating, the details of the proof are slightly delicate.

The first step is in the nature of proving a lemma. If $\omega_i \leq ba^i$, $i < k$ then

$$\omega_k \leq \frac{b^2a^k}{k-1}\sum_{i=1}^{k-1}\{b^i\frac{(k-1)!}{i!(k-i)!}+1\} \tag{6.28}$$

$$= \frac{b^2a^k}{k-1}\{b^2(1+b)^{k-2}(k+1+\frac{2}{b}) - b^k(k+1)+(k-1)\}. \tag{6.29}$$

The minus term can evidently be dropped as, furthermore, can the final one since it is smaller than it for $b > 1$, as is always the case in this proof, and so their sum is negative. Thus

$$\omega_k \leq \frac{b^4a^k}{k-1}(1+b)^{k-2}(k+1+\frac{2}{b}). \tag{6.30}$$

Some notation and the general strategy must now be introduced. Assume

$$\omega_i \leq \frac{1}{B_0}(A_0)^i, \qquad i \leq N \tag{6.31}$$

then use the lemma to show that

$$\omega_i \leq \frac{1}{B_1}(A_1)^i, \qquad i \leq N+1. \tag{6.32}$$

Iterating gives

$$\omega_i \leq \frac{1}{B_m}(A_m)^i, \qquad i \leq N+m \tag{6.33}$$

with the particular result

$$\omega_{N+m} \leq \frac{1}{B_m}(A_m)^{N+m}. \tag{6.34}$$

From (6.30), one has after n iterations

$$\omega_{N+n} \leq \frac{[A_{n-1}(1+\frac{1}{B_{n-1}})]^{N+n}}{B_{n-1}^2(1+B_{n-1})^2}\frac{N+n+1+2B_{n-1}}{N+n-1} \tag{6.35}$$

suggesting that one take

$$A_n = (1+\frac{1}{B_{n-1}})A_{n-1} \tag{6.36}$$

$$B_n = \frac{B_{n-1}^2(1+B_{n-1})^2}{\delta_n} \tag{6.37}$$

where δ_n has been required to satisfy

$$\delta_n > 1 + \frac{2(1+B_{n-1})}{N+n-1}. \tag{6.38}$$

The other important constraint on the iteration procedure is that the new bound cannot be allowed to violate earlier ones, a sufficient (and natural) condition for which is

$$\frac{1}{B_{n-1}}(A_{n-1})^i \leq \frac{1}{B_n}(A_n)^i, \qquad i \leq N+n-1. \tag{6.39}$$

From (6.36) above it is clear that $A_{n-1} < A_n$, so it is sufficient to enforce

$$B_n \leq B_{n-1} \tag{6.40}$$

to ensure this happens, particularly since $\omega_0 = 1$ means that $B_n \leq 1$. Everything now depends on a judicious choice of δ_n and B_n satisfying these formulae to produce as tight a bound as possible at very large n. From both the form of (6.36) and (6.34) the tightest sequence of bounds is clearly that given by the fixed point of (6.37)

$$B_n = B_{n-1} = B^* \tag{6.41}$$

corresponding to equality in constraint (6.40). For this to occur δ_n is set equal to a constant δ where

$$\delta > 1 + \frac{2(1+B^*)}{N} \tag{6.42}$$

the constraint (6.38) on δ_n now being increasingly well satisfied at large n. The fixed point of (6.37) itself requires

$$\delta = B^*(1+B^*)^2. \tag{6.43}$$

Now

$$A_n = \left(1 + \frac{1}{B^*}\right)^n A_0 \tag{6.44}$$

and

$$\omega_{N+m} \leq \frac{1}{B^*}\left(\left(1+\frac{1}{B^*}\right)^m\right)^{N+m} A_0^{N+m} \tag{6.45}$$

so the problem has been reduced to that of optimising the choice of B^*, or equiva-

lently δ. That choice is $B^* = 1$, $\delta = 4$, which duly satisfies (6.42) provided $N \geq 2$; take $N = 2$ and use (6.31), i.e.

$$\omega_i \leq \frac{(A_0)^i}{B_0}, \qquad i \leq 2 \tag{6.46}$$

to fix A_0 as small is as permissible given the first few terms in the series. This done, the final answer is

$$\omega_n \leq 2^{n^2-3n+2}(\sqrt{7})^n. \tag{6.47}$$

How good a bound is this ? Probably it is not too stringent because of the way the requirement that it apply thoughout the whole series necessarily restricts how effective it can be at large n where our real interest lies, even although the initial terms may ultimately be of little consequence in the recurrence relation. In particular, (6.40) arises only because the bound must encompass the first term.

This difficulty cripples analogous attempts to derive a lower bound, since it forces the equivalent of B_n to become larger and larger very quickly (there is now no fixed point B^*), eventually causing the bound to *fall* as n increases. The knowledge that the series must grow faster than any geometric one asymptotically aside, the best lower bound is roughly

$$\omega_n > \omega_{n-1}, \qquad n \geq 2 \tag{6.48}$$

which can only be described as extremely weak.

6.5 Numerical Results

Confronted with the intractability of the recurrence relation (6.3) when tackled analytically, an obvious recourse is to the calculation of as many terms of the series as possible on a computer in order to attempt an estimate of its large-order behaviour. In doing so the primary problem is to find an algorithm which makes the most efficient use of the limiting resources of time and memory. Clearly the largest decisions concern the handling of the multiple Cauchy products implied by the notation $C_n(\frac{u}{x})^m$ and indeed it is this which determines both the overall strategy adopted and the final performance of the program.

Given a truncated series' coefficients $(u_0, u_1, \ldots, u_N)$, one can straightforwardly calculate $(\rho_0, \rho_1, \ldots, \rho_N)$ and any subsequent powers one wishes, but one cannot go beyond the Nth term in any series – that requires knowledge of at least u_{N+1}. To calculate u_{N+1} using the recurrence relation in turn requires some knowledge of the Nth power of the truncated series, specifically $C_1(\frac{u}{x})^N$. This particular coefficient is actually sufficiently special not to need a complete calculation $(C_1(\frac{u}{x})^N = Nu_1)$, but the basic point remains: knowing $(u_0, u_1, \ldots, u_N)$ one can calculate all the powers up to the Nth one, find u_{N+1}, then iterate. Of course most of the calculation required in finding powers of $(u_0, u_1, \ldots, u_N, u_{N+1})$ is exactly that already done in finding powers of $(u_0, u_1, \ldots, u_N)$, so it's much quicker to store these series and then add to them later on than have to recalculate them entirely again and again. Thus the program is really calculating a square array corresponding to $C_n(\frac{u}{x})^m$, from part of which the actual u_n series can be read off:

$$\begin{array}{cccccc}
u_0 = 1 & u_1 = \frac{c}{2} & u_2 & u_3 & \ldots & u_N \\
\rho_0 = 1 & \rho_1 = c & \rho_2 & \rho_3 & \ldots & \rho_N \\
C_0^3 = 1 & C_1^3 = \frac{3}{2}c & C_2^3 & C_3^3 & \ldots & C_N^3 \\
C_0^4 = 1 & C_1^4 = 2c & C_2^4 & C_3^4 & \ldots & C_N^4 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
C_0^N = 1 & C_1^N = \frac{N}{2}c & C_2^N & C_3^N & \ldots & C_N^N
\end{array} . \tag{6.49}$$

Additional columns are added by using (6.3) and the table to find u_{N+1}, then filling in the rest from Cauchy's rule in the form

$$C_{N+1}\left(\frac{u}{x}\right)^m = \sum_{i=0}^{N+1} C_i \left(\frac{u}{x}\right)^{m-1} u_{N+1-i} \tag{6.50}$$

the application of which necessitates knowing all the rest of the array. Additional rows can be added merely by taking the Cauchy product of the top row with the bottom one. Both operations must be carried out prior to finding u_{N+2}.

The program itself is rather short, containing only 36 lines, and simply prints out the u_n as they are found. Its major storage requirement is the $N \times N$ quadruple array needed to be able to reach u_N.

All the operations involved are either arithmetic ones or calls to the main array and so in estimating the program's speed the main consideration would appear to be

the number of iterations around loops and how this grows as longer stretches of the series are explored. How many loops are needed to calculate u_N ? The operations of adding a row or a column alternate throughout execution, but for this purpose they can be considered separately. To find

$$\rho_n = \sum_{i=0}^{n} u_i u_{n-i} \tag{6.51}$$

requires n loops, as does $C_n(\frac{u}{x})^m$ provided the rows above are known. Hence if (u_0, ..., u_N) were given, the other rows could be calculated using

$$\frac{1}{2}N(N-1)(N+1) \tag{6.52}$$

steps. Only the steps involved in extending the top row using the recurrence relation still have to be included: to find u_n knowing the rest of the table requires $n-2$ steps, so

$$\frac{1}{2}(N-1)(N-2) \tag{6.53}$$

are needed in total for this. The running time of the program is thus expected to increase as

$$\frac{1}{2}(N-1)(N^2+2N-2) \tag{6.54}$$

$$\sim N^3. \tag{6.55}$$

In fact only the top left-hand half of the table is ever used in the recurrence relation and the rest could be omitted, but this only halves the memory requirement and reduces the running time by a factor 3, neither of which is a sufficiently dramatic improvement to be important.

Part of the reason for dwelling on this choice of strategy is that at first sight it runs counter to the conventional wisdom [118] concerning the efficient calculation of Cauchy products. Quick algorithms have been specifically designed for the problem of large powers of a single series and it might be thought that these would find an application here, but these were really intended for situations where the lower powers are of no interest. For instance, although Knuth [118] describes one which

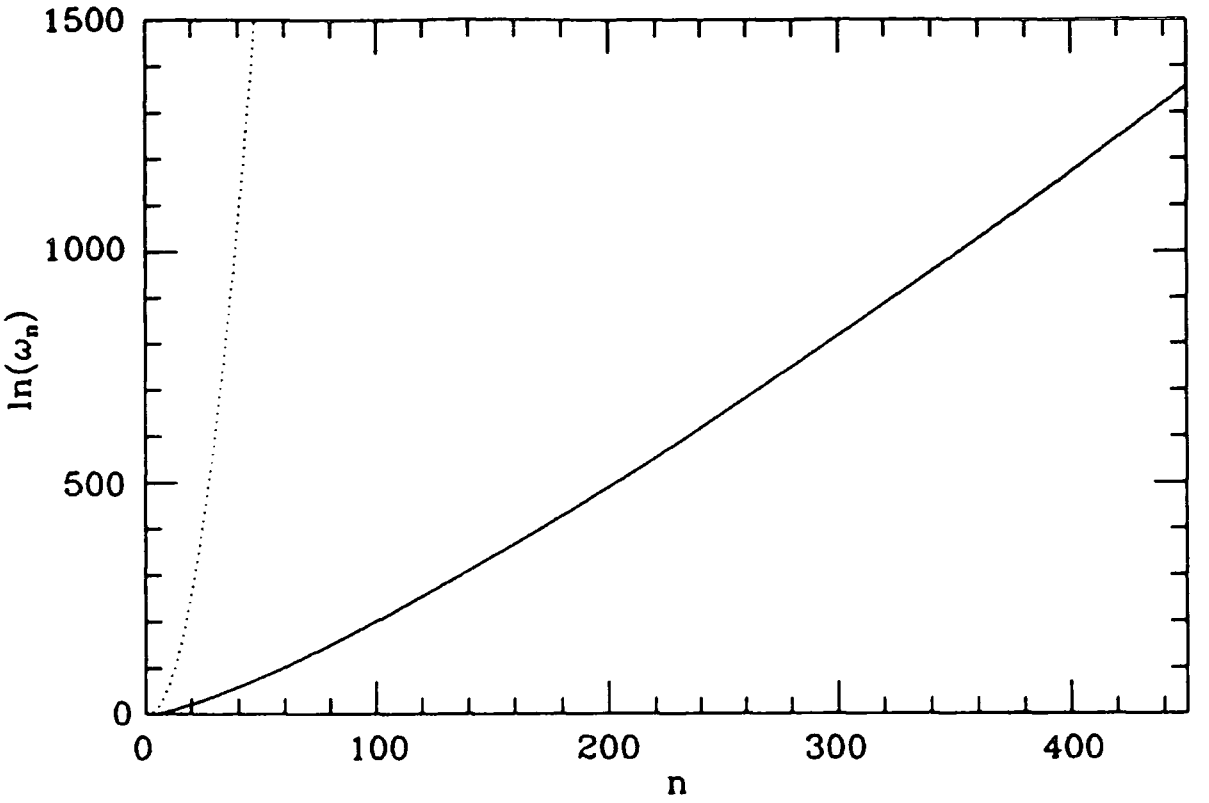

Figure 6.1 The series coefficients.

can determine C_N^N above using only $\sim N^2$ iterations, to be compared to the $\sim N^3$ above, it requires another $\sim N^2$ to find C_N^{N+1} instead of the N needed if the table has been *stored*. And since the rest of the table has already been needed in the recurrence relation it makes sense to do just that. If one *only* wanted C_N^N the alternative algorithm would be better, but we don't and it isn't, needing $\sim N^4$ iterations to find u_N. Although it would require much less ($\sim N$) memory, this would be at the expense of a prohibatively long running time.

Figure 6.1 displays the first 450 ω_n (merging to form the solid line) together with the bound (6.47) derived in the last section (the dotted line). That result is now seen to be rather weak, although the coefficients' growth is still precipitous and $\omega_{450} \simeq 10^{592}$. To obtain a better understanding of this growth, the function

$$D(C)^n \Gamma\left(\frac{n}{A} - B\right) \tag{6.56}$$

was fitted to each successive block of 50 coefficients and the results are plotted in

Figure 6.2. The most important conclusion is that the increase is close to $\Gamma(n+5)$ and probably converging to something like this. For field theory expectation to be satisfied it is only necessary for (6.56) to hold as $n \to \infty$ with $A \sim o(1)$, but these results are quite consistent with exactly $A = 1$ in that limit.

Hardly less interesting than this is the way that, say, the first 100 coefficients would be misleading about the ultimate asymptotic behaviour, this being significantly faster than one would suspect on that basis. This need not be discouraging since such a pattern could delay the breakdown of the series well beyond its first few orders and thereby give QCD much greater predictive accuracy than we imagine, at least in principle. This is in striking contrast to section 4.5 where we saw a conjectured asymptotic result used to estimate the first few terms in a series. Similar sorts of extrapolations are rather common, but Figure 6.2 suggests that more caution is appropriate. However, fitting a single curve to selected points in the entire range up to $n = 450$ gives $A \simeq 1.4$ and a curve indistinguishable by eye from Figure 6.1, so the difference should not be exaggerated. It should also be emphasised that the sensitivity of all these fits to the different parameters varies greatly – small changes in A produce vastly larger effects than equivalent changes in D – so that the precise values of C and D at any stage are more uncertain than those of A and B.

To summarise, the divergence of the series appears to be a non-Borel summable, fixed-factorial one.

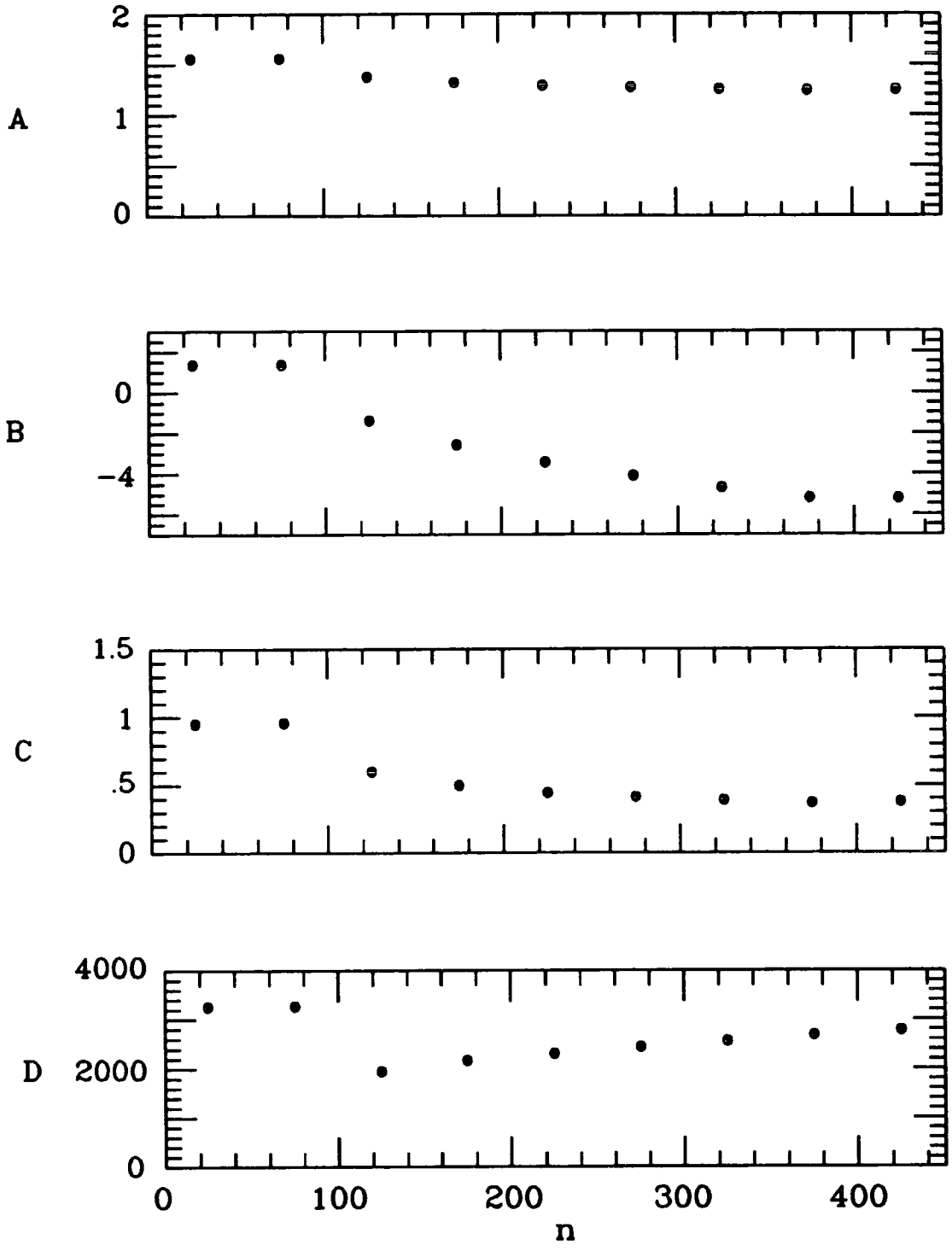

Figure 6.2 Results of fitting $DC^n\Gamma(n/A - B)$ to ω_n.

CHAPTER SEVEN

Non-Perturbative Consequences

7.1 Introduction

However challenging the problems it gives rise to, the perturbative solution to

$$u^2\frac{du}{dx} = u^2(u) \tag{7.1}$$

is only part of the story – its very promise is that of providing a non-perturbative solution. If such exists, and that remains to be proved, several properties of it can be deduced trivially, e.g.

$$\frac{du}{dx} > 0, \qquad \forall x \in \Re. \tag{7.2}$$

This assumes that any solution is real and, while it will turn out to be an inadequate supposition, this will probably be true in the region close to the origin of most interest to physics. It is also convenient to assume that $u(x)$ is single-valued for $x > 0$ (and $c > 0$), although this is probably equally simplistic. For a start the non-analyticity at the origin is probably due to a cut, possibly one extending along the negative real axis.

The trivial solution $u(x) = x$ can no longer be dismissed so lightly. It marks out a special line in the (u, x) plane which can only be crossed by a solution at the origin; at any other point $x = x_0$ where $u(x_0) = x_0$

$$\left.\frac{du}{dx}\right|_{x_0} = 1, \qquad \left.\frac{d^n u}{dx^n}\right|_{x_0} = 0, \qquad n > 1. \tag{7.3}$$

Which side of this a solution begins on at the origin depends on the sign of c with $c > 0$ giving ones above the line. It has already been seen that c was not fixed by the perturbative expansion and by substituting $u(x) \equiv x\omega(xc)$ into (7.1) one finds

$$\omega(xc) + xc\frac{d\omega(xc)}{d(xc)} = \omega^2\big(xc\omega(xc)\big) \tag{7.4}$$

as a reflection of this, so the complete set of non-perturbative solutions can be

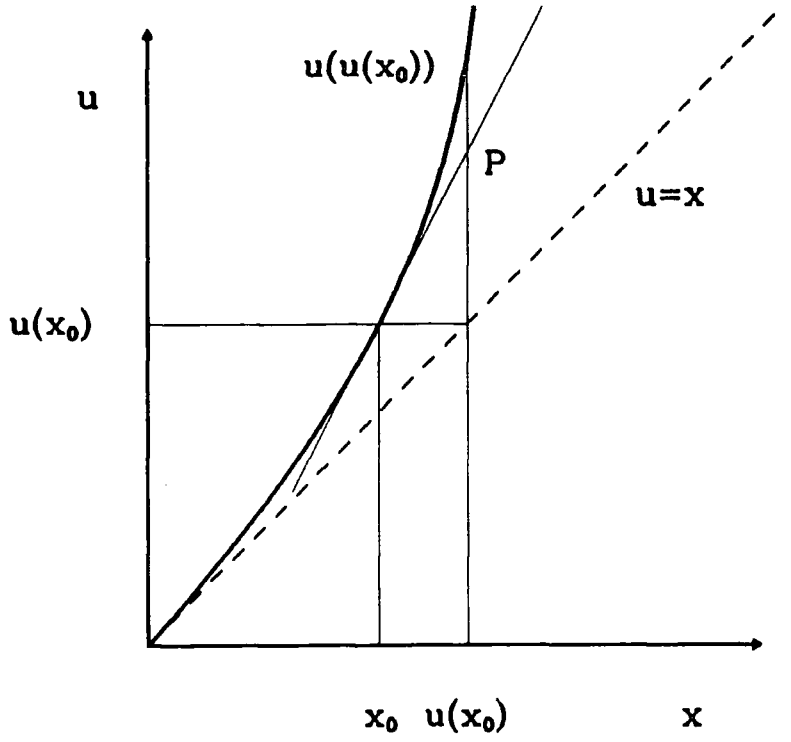

Figure 7.1 Construction I.

partitioned into infinite families, the members of which are parameterised by c. If c is allowed to become complex, this enables the analytic structure of solutions to be rotated as well as dilated. One special case is that if $u(x)$ satisfies (7.1), then so does $-u(-x)$ but with a different c.

The most awkward feature of (7.1) is the way $u(u(x))$ enters into it. If the solution is monotonic in accordance with (7.2) then its specification on any finite interval is insufficient for it to be checked that it does indeed satisfy (7.1) throughout that interval. This "overspill" problem will be a recurrent theme in this chapter. At first sight it may be thought that the most likely result is for a solution with $c > 0$ to start from the origin, then begin to get steeper and steeper as a sort of positive feedback sets in through the right-hand side of (7.1), since $u'(x) > 1$ and gets bigger the steeper the function is further out. That this cannot quite be the case is shown by the construction in Figure 7.1.

Letting $u \equiv u(x_0)$, $u' \equiv u'(x_0)$, the point P is given by

$$P = u + (u - x_0)u'. \tag{7.5}$$

If the growth of the solution does not ease off, then $P \leq u(u(x_0))$ and (7.1) at x_0 gives

$$u + (u - x)u' \leq u\sqrt{u'} \tag{7.6}$$

which can only have a solution for $\sqrt{u'}$ (noting that $u' > 0$) if

$$u^2 - 4(u - x)u \geq 0 \tag{7.7}$$

so that

$$u(x) \leq \frac{4x}{3}. \tag{7.8}$$

Interesting though this bound is (particularly that it is independent of c), it could be evaded by several things including the existence of some point where $u''(x) = 0$ or a pole in the solution at finite x. There is no effective equivalent for $c < 0$.

7.2 Numerical Solution

An important advance can be achieved by shifting attention away from (7.1) itself and onto the equation

$$\rho_0^{(4)} = \rho_0^{(3)} - \frac{c}{2} \tag{7.9}$$

derived in Chapter 5 instead. Introducing the function

$$f(x) = \frac{1}{x} + c\ln\Big(\frac{cx}{1+cx}\Big) + \int_0^x \Big(\frac{1}{y^2(1+cy)} - \frac{1}{\rho(y)}\Big)dy, \tag{7.10}$$

once the full CECA assumptions are applied, this equation becomes

$$f(a) = f(R) - \frac{c}{2} \tag{7.11}$$

which can be rewritten as

$$\int_R^a \frac{1}{\rho(x)}dx = \frac{c}{2}. \tag{7.12}$$

This is to be set alongside

$$\rho(R) = a^2. \tag{7.13}$$

These two equations suffice to fix the function $a = u(R)$, which must be a solution

to (7.1) both because of consistency and because

$$\frac{df}{dx} = \frac{-1}{\rho(x)} \tag{7.14}$$

allows that equation to be recovered directly.

The great advantage of this representation is that it suggests the possibility of finding a numerical solution: an initial guess for $\rho(x)$ can be substituted into (7.12) to obtain an approximation to $a(R)$ and hence hopefully a better approximation to $\rho(x)$ via (7.13). In practice the program developed to carry this out stores values of $u(x)$ at points evenly spaced in an interval $[0, X]$, then for every point R in this set the integral (7.12) is evaluated numerically with $a(R)$ initially set to the current value of $u(R)$, but then adjusted (using a NAGLIB routine for solving a transcendental equation) until (7.12) is satisfied. At the end of each iteration this new set $a(R)$ replaces the $u(R)$ used to find it. Unfortunately, one immediately runs into the "overspill" difficulty since one needs to be able to approximate the integrand at the upper limit where $x = a(X) > X$. This is dealt with by linearly extrapolating the integrand beyond the region where the function is stored numerically; this consistently underestimates $a(R)$ when solving (7.12), but overestimates the integrand in the same region on the next iteration, so provided $1/\rho(x)$ does not vary too rapidly this approximation is conceivably under control. Certainly when the program is run with X fairly small it converges to an apparently sensible result. In Figure 7.2 the solid line shows $g(x) \equiv 1/x^2(1+cx) - 1/\rho(x)$ for $c = 29/23$ (corresponding to QCD with $N_f = 5$) as obtained using 500 points.

This choice of function to display was motivated partly by its physical significance and partly because it's rather more interesting then $\rho(x)$ iself which merely stays close to $x^2(1+cx)$ throughout. Several important conclusions can be drawn from this graph.

Firstly, as expected the curve intersects the axis at $\rho_2 = 7/4c^2$ and is also sufficiently flat thereafter for

$$\Delta\rho_0^{NNLO}(R) \simeq \frac{7}{4}c^2 R \tag{7.15}$$

to be a very good approximation at small R, particularly when compared to $F(R) \simeq 1/R$ for $R \simeq \alpha_S$. Thus if one were to accept $\rho(x) = u^2(x)$ as the actual common

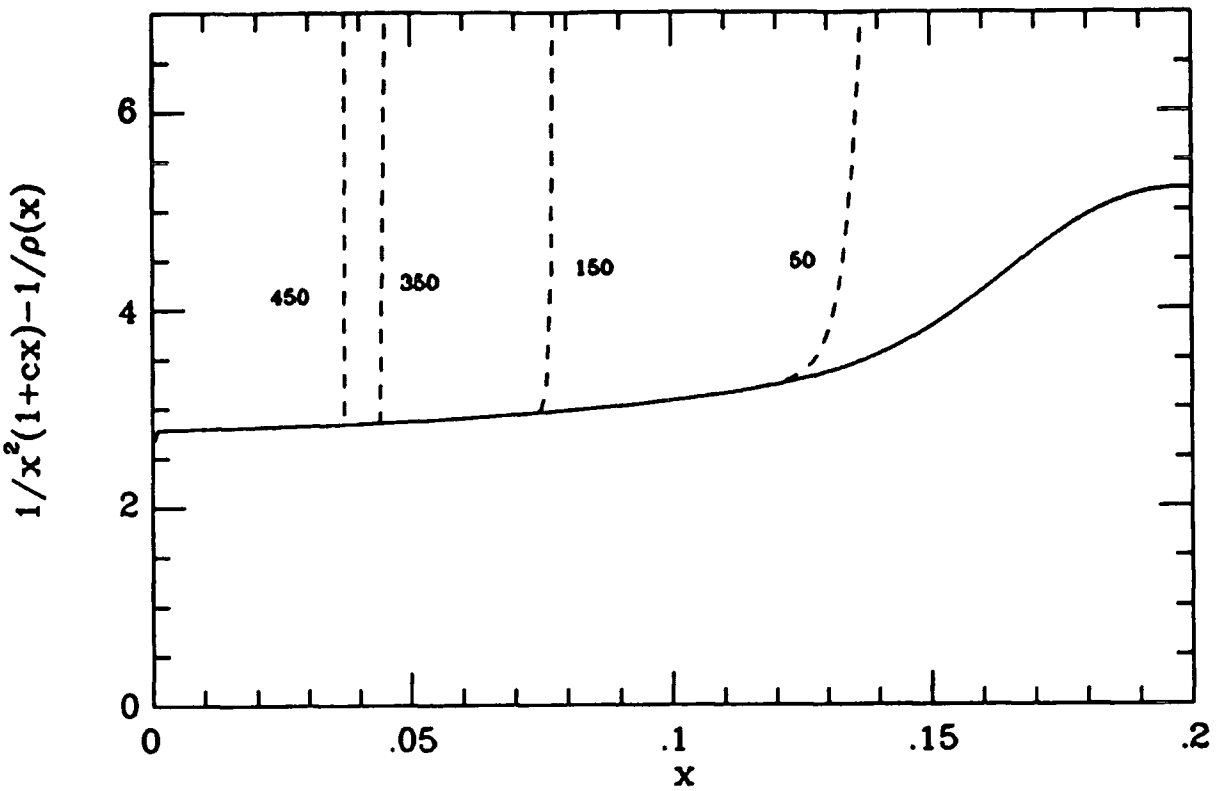

Figure 7.2 Numerical Results ($c > 0$)

β-function for one of the jet algorithms, one would have succeeded in finding a quantity for which $\Delta\rho_0$ is small since this would imply that its contribution to ρ_0 is less than 1% that of $F(R)$. Clearly, if that were to be accepted, one need not even be restricted to (7.15) and could use the numerical approximation to $\rho(x)$ and obtain a better estimate.

Secondly,it is possible to compare the new result with truncations of the perturbation series beyond this NNLO one. The dotted lines show how using the series for $u(x)$ truncated at different orders to estimate $g(x)$ compares to the actual function, the order of truncation being the number next to each line. At sufficiently large values of x the inadequacy of the approximation becomes apparent and the agreement with the correct result breaks down rather suddenly, the point at which this happens moving closer to the origin as more orders are added. Although the accuracy of the numerical solution is not good enough (due to the limitations of the NAGLIB routine) to confirm this, it is probable that truncating after 450 terms produces a better approximation close to $x = 0$ than the same series to 50 terms. This is exactly the behaviour expected for an asymptotic series – additional terms improve the accuracy in a smaller interval – and Figure 7.2 is therefore evidence

that the series discussed in the last chapter is behaving *like* one. For the purpose of applying QCD such a result is as encouraging as any formal proof that the series actually is asymptotic. It is also interesting that most of the truncations shown are well-behaved even at the comparitively large couplings relevant to contempory experiments, thereby contradicting the supposition that series must begin to diverge after only a handful of terms. This is presumably just a consequence of the strong sub-asymptotic (in the power series sense) effects noted in the last chapter.

7.3 Analytic Results

When the order of truncation is reduced to very small values, the perturbative approximation manages to reproduce $g(x)$ up until the maximum at $x \simeq 0.2$.* Beyond this point the function seems to be essentially non-perturbative and its form is therefore of particular interest. The numerical results indicate that $\rho(x)$ slowly falls back towards $x^2(1+cx)$, causing $g(x)$ to fall towards zero. However as the endpoint of the numerical calculation is increased past $x \simeq 0.5$ an instability develops in this new calculation and as the iterations of $g(x)$ continue a low wave is seen to sweep in from large x, growing as it does so into a very sharp peak which progressively destroys all the earlier results at small x, leaving $\rho(x) \simeq 0$ in its wake. Suspicion may fall on the extrapolation necessary to deal with "overspill" but, while this is related to it, the difficulty is much more fundamental than that.

To understand its origin it is useful to introduce the geometric reformulation of $u(x)$ illustrated in Figure 7.3a. $f(x)$ is any function such that

$$\frac{df}{dx} = \frac{-1}{u^2(x)}. \tag{7.16}$$

As such it is clearly related to the $f(x)$ in (7.10), which is just a solution of this equation with a particular choice of integration constant (c.f. the discussion of Chapter 3, of which this is just a special case), an integration constant which will be irrelevant in what follows. Because of this, (7.11) can be rewritten as

$$a = f^{-1}(f(R) - c/2) \tag{7.17}$$

the geometrical significance of which is just the arrowed path in the diagram. Triv-

* For c as in QCD. Solutions for other c scale in accordance with (7.4).

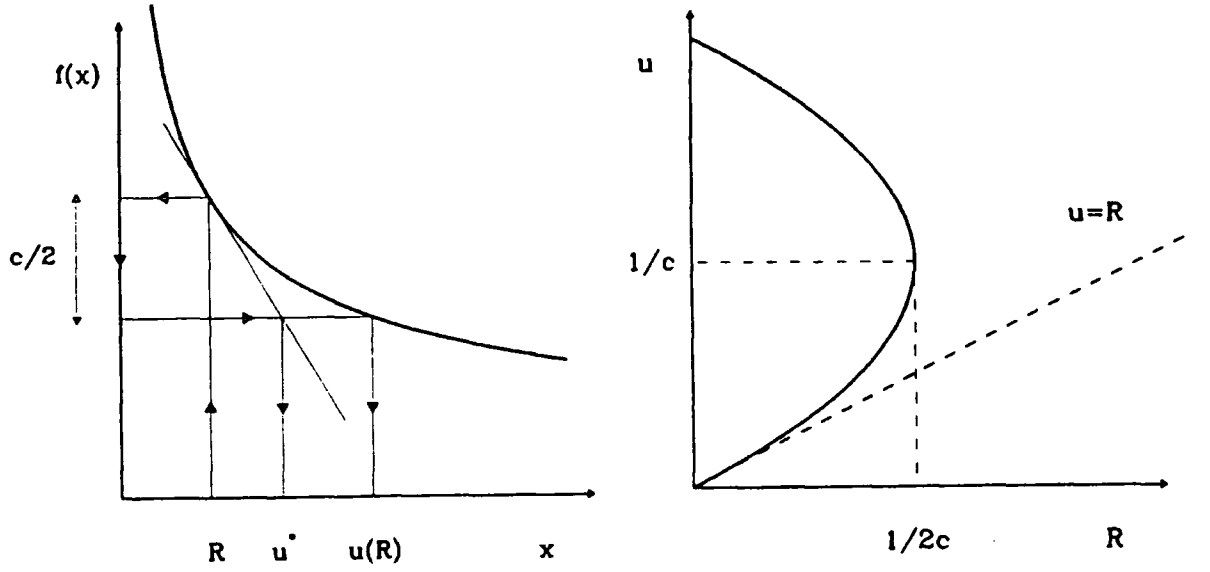

Figure 7.3 a) Construction II b) Bound on $u(x)$

ially

$$\frac{d^2f}{dx^2} = \frac{2}{u^3}\frac{du}{dx} \geq 0. \qquad x > 0 \tag{7.18}$$

so $f(x)$ is always convex, which means that u^* constructed using the tangent at $(R, F(R))$ as shown is always such that

$$u^* < u. \tag{7.19}$$

But this tangent's gradient is known from (7.16) and so one finds the crucial bound

$$x < u - \frac{u^2 c}{2} \tag{7.20}$$

displayed in Figure 7.3b. Alternatively (and rather quicker, although without an additional insight into the computer program), if $u(x)$ is monotonic then from (7.12)

$$\frac{c}{2} \leq \int_R^a \frac{1}{u^2(R)} dx = \frac{(a-R)}{a^2} \tag{7.21}$$

with the same result.

In extending it beyond $x = 1/2c$ the numerical method breaks down because it is trying to construct a solution which violates this bound. The extrapolation procedure is related to this since it essentially approximates $u(x)$ by x^* when this lies outside the range, except that a parabola is used in its construction (in adopting a linear approximation to $1/\rho(x)$, $f''(x)$ is being taken as constant) rather than the tangent; such modifications to the above argument are possible, but only marginally alter the final bound.

However this bound also decisively undermines the assumption that $u(x)$ is a real, single-valued function for all real $x > 0$. Nor is there a simple escape such as supposing that it follows the trend of the bound and doubles back before $x = 1/2c$; du/dx can only be less than zero if $u(x)$ is allowed to be complex. Tentative investigations of how the function may evade the bound have been attempted, but without any definite progress. The non-locality implied by the presence of $u(u)$ in (7.1) enters crucially into its structure and $u(x)$, if it exists at all, must be a complicated multi-branched complex function whose analytic properties are currently obscure.

Not that this need refute the results of the previous section nor should it be regarded as necessarily threatening the physical plausibility of the CECA. This would only be the case were it to be proved that $u(x)$ is complex on the positive real axis to within a short distance of the origin. Otherwise, particularly as no bound significantly tighter than (7.20) has been found,one can suppose that the behaviour will only become different when $R \simeq 1/2c$ which happens at energies low enough to be close to the leading-log Landau pole. That new aspects of the theory emerge at about this point is hardly a surprise and it is an interesting unanswered question whether the CECA remains physical at low energies. When applied to the interval shown in Figure 7.2, the computer calculation does not involve any assumptions about what happens at this larger $x \simeq 1/2c$ and so can be an accurate reflection of $u(x)$ in the region of interest at 91 GeV.

For future use, note that (7.12) can be recast on the real line into a more familiar form. Consider c as a function of u and x

$$xc = G\Big(\frac{u}{x} - 1\Big) \tag{7.22}$$

after which (7.12) can be changed into

$$G(x) = \int_0^x \frac{2dy}{(y+1)^2[1+G^{-1}((y+1)G(x))]^2}. \tag{7.23}$$

Apart from the appearance of G^{-1} in the denominator, this closely resembles a non-linear Volterra equation of the second kind. Unfortunately, there is a dearth of general results and literature on these equations [119][120], with each one giving rise to fresh difficulties and existence having to be proved case-by-case, but as a class they have an unusually wide variety of behaviours. For instance, even an example as simple as

$$G(x) = x + \int_0^x G^2(y)dy \tag{7.24}$$

only has a continuous solution if $x < \pi/2$ [120]. Note that this representation does not circumvent the problem of "overspill" since for any monotonic $G(x)$ – and if it's not, (7.23) is immediately ambiguous – one has

$$b < G^{-1}((1+b)G(b)) \tag{7.25}$$

so specifying $G(x)$ on $[0, b]$ does not determine the integrand on the same interval.

7.4 Negative c

If the construction in Figure 7.3a is examined more closely one realises that the difficulties arise at small $f(x)$, in particular with the tail as $x \to \infty$. For instance, it cannot be the case that $f(x) \to \kappa$, a constant, as $x \to \infty$, because this would imply the existence of an R_0 beyond which $u(R)$ would not be defined (the horizontal line needed to construct it passing below $f(x)$ for all x), yet $u(R)$ must be defined for $R > R_0$ since this region enters into the determination of $u(R)$ for $R < R_0$. It is the impossibility of scenarios like this which give rise to the bound. However, if $c < 0$ the arrowed path in Figure 7.3a reverses and these problems disappear, while no new ones emerge as $x \to 0$. This is simply a consequence of the fact that there is no longer any "overspill" since

$$u(x) < x, \qquad a(a) < a. \tag{7.26}$$

Repeating the argument above, the bound is now found to be

$$x < \frac{|c|}{2}u^2 + u. \tag{7.27}$$

This is not closed and so solutions can now be monotonic, single-valued and real.

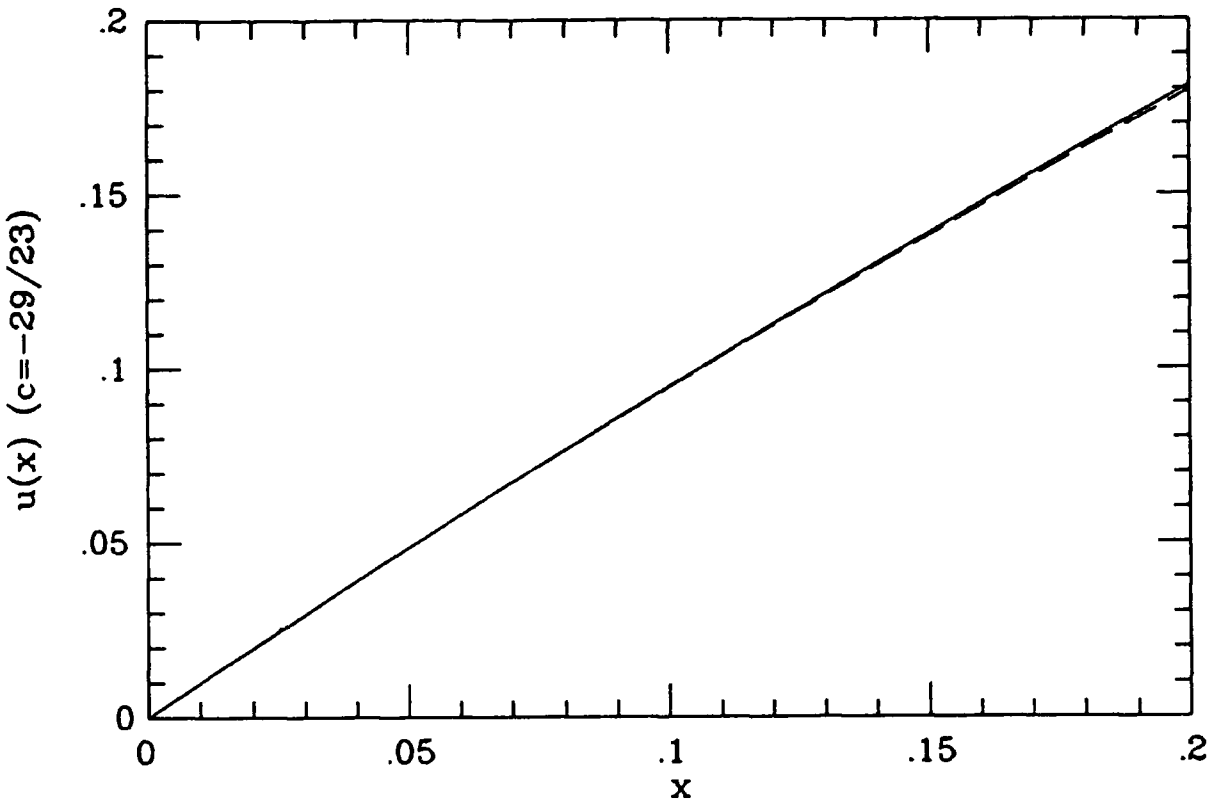

Figure 7.4 Numerical results for $c < 0$ together with bound.

Figure 7.4 displays the results of modifying the program so that it generates a numerical solution for $c < 0$. The barely discernable dashed line is this bound (7.27); any valid solution must lie above this line and a detailed comparision shows that the one found always does so. Unlike before, extending the program to larger x induces no breakdown in stability and the function continues to increase in the same featureless manner. An obvious check is to use it to numerically evaluate both sides of (7.1) to see if there is agreement; when done the agreement is one part in 10^6, comparable to the machine accuracy (using double precision). Similar agreement was achieved for the $c > 0$ solution in Figure 7.2, provided x was not approaching the bound.

Consistency of representation (7.23) requires that $G(x) \to -\infty$ as $x \to -1$, which implies that

$$\frac{u(R)}{R} \to 0 \quad \text{as} \quad R \to \infty, \quad c < 0. \tag{7.28}$$

Because everything maps into $x \in$[-1, 0] this version of the problem is especially useful in investigating the $R \to \infty$ limit. In this interval $G(x) < 0$ and monotonic;

by noting that $G^{-1}((1+y)G(x)) < 0$ when $y \in [x, 0]$, one obtains

$$G(x) \leq \frac{2x}{(1+x)}, \qquad \forall x < 0 \tag{7.29}$$

which in turn implies a bound on $u(R)$, although this happens to be a weaker one than (7.27) above. More interestingly, the integrand in (7.23) can be seen to be largest when $y = 0$, which leads to

$$G(x) \geq \frac{2x}{(1+x)^3}, \qquad \forall x < 0 \tag{7.30}$$

so that $G(x)$ is now tightly constrained. This does translate into new information about $u(R)$, since it is equivalent to

$$2\left(\frac{R}{u(R)}\right)^2 - 2\left(\frac{R}{u(R)}\right) - u(R)|c| \geq 0, \tag{7.31}$$

a bound on R of the form

$$R \geq \frac{U(R)}{2}(1 + \sqrt{1 + 2u(R)|c|}), \qquad c < 0. \tag{7.32}$$

Together with (7.27) this sandwiches $u(x)$ into a narrow band – this new bound is not shown in Figure 7.4 only because it is indistinguishable by eye from the solution just below it. As $R \to \infty$, the limiting behaviour is

$$\frac{2}{|c|}R^{1/2} \leq u(R) \leq \left(\frac{2}{|c|}\right)^{1/3} R^{2/3} \tag{7.33}$$

in agreement with (7.28) above.

Abstract situations where the CECA might be applied with $c < 0$ can be envisaged, notably in considering the process $X^+X^- \to n$ scalars when charged fermions have been coupled into ϕ^4_{3+1} theory. Apart from this, the symmetry mentioned in section 7.1 means that

$$u(-x, -c) = -u(x, c) \tag{7.34}$$

and so the solution shown for $c < 0$ is also really one for $c > 0$ on the negative real axis. In principle at least, this allows for the possibility of constructing a solution

with $x < 0$, $c > 0$, then continuing it round onto the positive real axis. How this actually works, and the bound (7.20) clearly indicates that some complications must enter, will depend on the detailed analytic structure, but it does at least open another avenue in trying to understand the physical solution. Since no existence proof for a non-trivial $u(x)$ is known, the importance of the $c < 0$ case is that it provides good evidence for (7.1) having an interesting solution at all.

However, in the absence of an existence proof, the outcome that $u(x)$ is either highly non-physical (most likely because it turns out to be inherently complex on the positive real axis) or even an impossibility must be considered. Although this would inevitably be very disappointing, the exercise would still teach us something: RG invariance alone is able to rule out some superficially "possible worlds." Since this is a much weaker condition than renormalisability itself, that would be surprising enough.

7.5 Weakening The Assumptions

The derivations of Chapter 5 depended on only two assumptions, that the ρ's are smooth at the 4-jet threshold and that they are equal there. Of these the first is the less restrictive and more natural and has as its main consequence the relation

$$\rho^{(3)}(R) = a^2 \tag{7.35}$$

whereas the second has turned out to be restrictive enough to leave no freedom in the theory apart from the value of Λ. This and the last chapter have described how they lead to the conclusion that the perturbation series are divergent yet $\Delta\rho_0 \simeq 0$. Whilst greater miracles have been known to occur in gauge field theories, it would be very surprising if both these assumptions, and particularly the second, were to turn out to be exact for any previously defined jet algorithm. A first test of this possibility will be the results for K_{43} obtained from the one-loop calculations presently underway [121]. Even without knowing the outcome of these, it is necessary to consider to what extent the conclusions may change if the assumptions have to be discarded or weakened.

Because it is clearly the stronger, and hence in greater need of justification, we begin by adjusting the assumption that $\rho^{(3)} = \rho^{(4)}$, while retaining (7.35). That

equation and

$$\rho^{(3)}(R)\frac{da}{dR} = \rho^{(4)}(a) \tag{7.36}$$

now define the problem, where $\rho^{(4)}$ is taken as arbitrary, apart from the non-trivial restriction that its perturbation series should start $a^2(1 + ca + \ldots)$. This situation is probably far too vague for anything useful in the way of general conclusions to be drawn. Probably the only special feature common to all in this subset of possible systems is that $\rho^{(3)}(x) \geq 0$, which will have consequences for any fixed points, e.g. if $\rho^{(4)}$ has a zero, R behaves like energy in the standard analysis of asymptotic behaviour and $a(R) \to$ constant as $R \to \infty$. However if $\rho^{(2)}$ has a zero, then $\rho^{(4)}$ must have more than one branch.

To solve for $\rho^{(3)}$ given $\rho^{(4)}$, one must take a reciprocal and integrate to find

$$R(a) = \int \frac{a^2}{\rho^{(4)}(a)} da \tag{7.37}$$

which must then be inverted and squared. Regarding $\rho^{(3)}$ and $\rho^{(4)}$ as power series, the properties of these operations can place restrictions on their radii of convergence via the theorems on reciprocation and inversion (integration and squaring have no effect) mentioned in section 6.3. In general all combinations of radii are possible, although a full description involves a messy set of special cases and it may be that certain possibilities can be excluded on general grounds, e.g. by specifying the behaviour of functions as $x \to 0$, or must entail certain features like the presence of fixed points. Thus for example both $\rho^{(3)}$ and $\rho^{(4)}$ can have infinite radii of convergence, but only if $R(a)$ is monotonic and there are no fixed points.

One interesting weakening of $\rho^{(3)} = \rho^{(4)}$ is to assume that $\rho_k^{(3)} = \rho_k^{(4)}$ but only for $k \geq N$, or in the limit $k \to \infty$. This evidently implies that their radii of convergence are equal and one can investigate what is necessary for these to be finite. One can apply the standard theorems to find that such a finite radius r is only possible provided

$$a(r) \leq r. \tag{7.38}$$

Note that there is no such $R = r$ for $\rho^{(3)} = \rho^{(4)}$ if $c \neq 0$; this is really a special case of noting that (7.38) will be violated if $\rho_k^{(3)} \geq 0$, $\forall k$. That equation probably doesn't

exhaust the consequences of imposing $\rho_k^{(3)} = \rho_k^{(4)}$ at large k and one can conjecture that it will always be violated in such cases. To progress further than this would require some formulation which, unlike any discussed in Chapter 6, untangles the behaviour at large-orders from the series' beginning.

Alternatively, if $\rho_k^{(4)} \geq \rho_k^{(3)}$, $\forall k$, then both $\rho^{(3)}$ and $\rho^{(4)}$ will diverge, since comparing the new recurrence relation to the old one shows that the equivalent of u_k for $\rho_k^{(4)} \geq \rho_k^{(3)}$ are greater than before.

Within broad limits, either $\rho^{(3)}$or $\rho^{(4)}$ can be selected at will and thus so can either $\Delta\rho_0(R)$ or $\Delta\rho_0(a)$. Note that because $\Delta\rho_0(\sigma)$ will be small when $\rho(x) \simeq x^2(1+cx)$ for $x < \sigma$, if $\rho^{(4)}$ is increasing very rapidly, then so is $\rho^{(3)}$ and both $\Delta\rho_0$'s will be large. However the CECA results described in this chapter encourage the hope that

$$\Delta\rho_0 \simeq \Delta\rho_0^{NNLO} \tag{7.39}$$

in all reasonable cases, even with a divergent series, but any general result *must* assume some sort of relation between $\rho^{(3)}$ and $\rho^{(4)}$.

7.6 Non-Smooth Versions

What are the consequences of weakening the smoothness assumption, i.e. letting

$$\lim_{K_{42}\to 0} \frac{\rho_k^{(2)} - \rho_k^{(3)}}{K_{42}} \neq 0, \tag{7.40}$$

but still requiring $\rho^{(3)} = \rho^{(4)}$? Since $R(a)$ will be a definite function, $G(R)$ can be replaced by $J(a)$ and equation (5.16) can be rewritten as

$$\frac{d\rho(R)}{dR} - J(a) = \frac{2\rho(a)}{a} \tag{7.41}$$

so that (5.27) is replaced by

$$\rho(R) = a^2 \exp\left(\int \frac{J(a)}{\rho(a)} da\right) \equiv a^2 e^{P(a)} \tag{7.42}$$

and (5.30) by

$$\frac{da}{dR} = \frac{a^2(a)}{a^2} \exp(P(a(a)) - P(a)). \tag{7.43}$$

Straightforwardly one still finds that $da/dR > 0$ and that $a = R$ is an uncrossable

line. The first of the limits (7.40)

$$L \equiv \lim_{K_{42}\to 0} \frac{\rho_2^{(2)} - \rho_2^{(3)}}{K_{42}} \tag{7.44}$$

is particularly important since (5.21) becomes

$$-L = c + 2\frac{K_{22}}{K_{21}} \tag{7.45}$$

$$= c + 2r_1 \tag{7.46}$$

where $R(a) = \sum r_n a^n$. For this reason, although c remains important, it is really

$$c' = c + L \tag{7.47}$$

that is the key parameter. In particular

$$\int_R^a \frac{dx}{\rho(x)} = \frac{c'}{2} \tag{7.48}$$

and, by analogy with (7.17), one has

$$a = f^{-1}(f(R) - c'/2) \tag{7.49}$$

where $f'(x) = 1/\rho(x)$ again. Regardless of the other limits (7.40), provided $L = -c$ this collapses to the trivial solution $R = a$. The physical reason for this is significant: the conventional CECA only gives a non-trivial $R(a)$ because, although $\rho(x)$ is common to all jet fractions, $\rho_0^{(3)} \neq \rho_0^{(4)}$ and so the effective charges cannot be exactly equal. Letting $L = -c$ removes this obstacle.

Most other consequences depend on the specific form of $J(x)$. An important class of cases is when

$$J(x) \geq x, \qquad \forall x < X \tag{7.50}$$

(for which $L < 0$), so that $P(x)$ is a monotonically increasing positive function. Then

$$\exp(P(a(a)) - P(a)) > 1 \tag{7.51}$$

when $X > a(a) > a$, so if $a(R)$ starts off similar to one of the old $c > 0$, $u(R)$ solutions (in fact that with $c = c'$) it will grow faster then it at larger R. There

are thus two competing factors governing the growth as compared to $u(R)$ with c itself unchanged: the new solution grows faster than an old one growing slower than $u(x,c)$. For sufficiently large L, c' becomes negative, this influence wins and $a(R)$ is permanently below the $a = R$ line, since for $a(a) < a$ the inequality (7.51) reverses. The dividing case is the $L = -c$ one. When (7.50) holds, a bound like (7.20) naturally exists.

This need not be the case when $J(x) < 0$, the derivation relying on $d\rho/dR > 0$ which can now be evaded. For any $P(x)$ that is monotonically decreasing, the solution above reverses and solutions tend to be forced towards the $a = R$ line when compared to $u(R,c')$. A major example here is when $P(a) \to -\infty$ as $a \to \infty$, so that $\rho(R) \to 0$ in that limit and the integrand in (7.48) increases without bound producing

$$a \sim R + \frac{c'}{2} R^2 e^{P(R)} \tag{7.52}$$

and hence $a(R)$ has $a = R$ as an asymptote.

It is difficult to be certain, but it seems unlikely that (the $L = -c$ cases aside) $\rho(x)$ can have any other than zero radius of convergence regardless of the form taken by $J(x)$. Although the grounds for believing this are intuitive, they are very general: as $a \to 0$, $J(a)$ becomes increasingly irrelevant in (7.41), its expansion starting one order higher than those of other terms, and this perturbative constraint presumably prevents it drastically changing the behaviour of other functions close to the origin. But this is exactly the region which determines whether the series converge or not, hence the optimism. In any particular case a proof modelled on that of Chapter 6 is probably possible, even though the expontentials complicate the general case excessively. Note that as part of this one can no longer rely on a simple c (or even c') dependence in the series coefficients. Similarly any attempt to develop an argument in the style of the last section has to contend with the more complicated relationship between $\rho(R)$ and $a(R)$, with the radius of convergence of $P(a)$ having to be brought into consideration.

For any $J(x)$ the problem can be solved numerically as before, although in doing so it is probably easier to classify cases using $P(x)$ instead. On the same grounds as above, one can expect that $\Delta\rho_0^{NNLO}$ will still be a good approximation. However,

if

$$P(a) = -La + p_2 a^2 + \ldots \tag{7.53}$$

then

$$\rho_2 = \frac{7}{4}c^2 + p_2 - \frac{1}{4}L^2 + Lc \tag{7.54}$$

so exactly how large this makes $\Delta\rho_0$ is no longer as simple as before.

Finally, we record a result whose derivation relies only on the $K_{42} \to 0$ limit of $\rho^{(4)}(a)$ being well-defined. Dropping the CECA assumption means that (7.42) becomes

$$\rho^{(3)}(R) = a^2 \exp\left(\int \frac{J(a)}{\rho^{(4)}(a)} da\right). \tag{7.55}$$

Again the most striking consequence here is that the 3-jet EC β-function must be greater than or (just possibly) equal to zero. This appears to be the most general consequence one can deduce from the normalisation of the jet fractions and RG invariance.

CONCLUSION

As regards the argument of Chapter 1, the conclusion is clear: there is now very strong numerical evidence that only the known shape invariant potentials give zero SWKB corrections and hence that

$$\phi' = a + b\phi^2 + c\phi \begin{cases} \sqrt{a} \\ \sqrt{a + b\phi^2} \end{cases}$$

is a necessary and sufficient condition for full shape invariance. This is certainly the most natural possibility and the ideas sketched in section 1.7 suggest that a detailed understanding of why it comes about is within reach.

No such note of finality is possible concerning series in field theory, nor probably will there ever be. Unlike mathematicians, physicists rarely have the liberty of asking questions they know to be capable of definite proof. However, to take an obvious example, it is more important (and more practical) to understand *why* perturbation series probably diverge than to settle the issue with full rigour. Even at this less ambitious level the topic continues to surprise. While probably not actually wrong , the standard (indeed textbook) explanations of Dyson and Lipatov increasingly look to be of secondary importance and the actual culprit to be the structure of RG invariance. Thus the divergence is caused by a general feature common to all renormalisable theories, rather than by particular solutions of the classical field equations. West was certainly premature to announce that analyticity and renormalisability alone could be responsible, but actually these appear not to fall too far short of sufficiency; Brown and Yaffe's result is a trivial consequence of these assumptions and by itself stongly suggests the existence of the renormalon singularities.

Similarly the Common Effective Charge Approach uses the properties of the RG invariance, expressed through the order-by-order scheme invariants, to establish the divergence in a wide set of cases. Here renormalisability is not augmented by analyticity and some additional assumption, but by a sequence of trial guesses at how $\rho^{(3)}$ and $\rho^{(4)}$ might be related. The simplest case, that of $\rho^{(3)} = \rho^{(4)}$ was discussed

in detail and it was shown that the series are not only divergent, but probably factorially so also. This is liable to be true much more generally. Futhermore, and this is one of the strengths of this approach, truncations of these series can be compared to the non-perturbatively generated functions that they are attempting to approximate. That simplest case is particularly encouraging in the way it suggests that not only can the first few terms in the series be a good approximation, but that the large-order behaviour sets in slowly enough for the breakdown due to the divergence to be delayed. If this pattern is found to hold in most cases, then perturbative QCD may be capable of much greater accuracy than anticipated hitherto. The only way to be sure of this would be to greatly increase the number of cases encompassed by the method; this will probably require a conceptual breakthrough if it is to be done efficiently, but the field is ripe for exploration.

Even if interpreted not quite so optimistically, these results presage well for sorting out how close to asymptotia (in energy) the various QCD observables are. The single NNLO calculation sits uncomfortably with current data, but several more theoretical results of this type should be able to clarify matters if in most circumstances $\Delta\rho_0^{NNLO}$ is a good approximation. Together with improved experimental results, extra NLO terms, analyses of energy dependence and refined lattice calculations, these hold out the prospect of greatly tightening our testing of the theory and hence the Λ extracted. But, whatever happens, QCD is likely to continue to surprise for some time to come.

APPENDIX 1

The recurrence relation for β_n derived from the $o(\hbar^6)$ correction is

$$
\begin{aligned}
&-8(2n+1)(2n+3)(35n^2-20)\beta_{n+1} \\
&\quad =(3360n^4+672n^3-1992n^2+696n+720)\beta_n \\
&\qquad +(3360n^4-11424n^3-10920n^2-16296n+720)\beta_{n-1} \\
&\qquad +(1120n^4-7840n^3-4120n^2-14840n+14160)\beta_{n-2} \\
&\qquad -72(2n+1)(7n+2)\Delta\epsilon_{n+1} \\
&\qquad +(-3024n^2-3528n-3600)\epsilon_n \\
&\qquad +(-3024n^2-4680n-8496)\epsilon_{n-1} \\
&\qquad +(-1008n^2-1944n-5040)\epsilon_{n-2} \\
&\qquad +216(\delta_{n-1}+3\delta_{n-2}+3\delta_{n-3}+\delta_{n-4}) \\
&\qquad +\sum_{i=0}^{n} f_1(n,i)\beta_i\beta_{n-i}+\sum_{i=0}^{n-1} f_2(n,i)\beta_i\beta_{n-i-1} \\
&\qquad -\sum_{i=0}^{n} 2(i+1)(480n+1296i+888)\beta_i\epsilon_{n-i} \\
&\qquad -\sum_{i=0}^{n-1}(2592i^2+960in+960n+2640i+912)\beta_i\epsilon_{n-i-1} \\
&\qquad +864\sum_{i=0}^{n}\epsilon_i\epsilon_{n-i}+864\sum_{i=0}^{n-1}\epsilon_i\epsilon_{n-i-1} \\
&\qquad +864\sum_{i=0}^{n-2}(i+3)(n-i-1)\beta_i\epsilon_{n-i-1} \\
&\qquad +864\sum_{i=0}^{n-1}(i+1)(n-i)\beta_i\epsilon_{n-i}
\end{aligned}
$$

and it holds for $n \geq 4$. Previously undefined pieces are

$$
\begin{aligned}
f_1(n,i) =&560n^4-896n^3-1700n^2+2592i^4-2808i^2-5184i^3n+5616i^2n^2 \\
&-1512i^2n-3024in^3+1512in^2+2808in-172n+192
\end{aligned}
$$

and

$$\begin{aligned} f_2(n,i) =& 560n^4 - 896n^3 - 3860n^2 + 2592i^4 + 5184i^3 + 5976i^2 - 5184i^3n \\ &+ 5616i^2n^2 - 3816i^2n - 3024in^3 - 2376in^2 - 6912in - 4996n \\ &+ 1944i - 600. \end{aligned}$$

The δ_n are the coefficients of

$$\left(\frac{d^2(\phi g)}{d\phi^2}\right)^2 = \phi^2(\delta_0 + \delta_1\phi^2 + \delta_2\phi^4 + \ldots)$$

and are analogous to the ϵ_n, with

$$\delta_n = \frac{1}{\epsilon_0}\Big(2\sum_{i=0}^{n}(i+1)(n-i+1)\epsilon_{i+1}\epsilon_{n-i+1} - \sum_{i=0}^{n-1}\delta_i\epsilon_{n-i}\Big).$$

$\Delta\epsilon_{n+1}$ is the part of ϵ_{n+1} without any dependence on β_{n+1}.

References

[1] E. Witten, Nucl.Phys.**B188**(1981)531.

[2] L.É. Gendenshteĭn, Pis'ma Zh.Eksp.Teor.Fiz.**38**(1983)299 [JETP Lett.**38**(1983) 356].

[3] R. Dutt, A. Khare and U.P. Sukhatme, Am.J.Phys.**56**(1988)163.

[4] L.É. Gendenshteĭn and I.V. Krive, Usp.Fiz.Nauk.**146**(1985)553 [Sov.Phys.Usp. **28**(1985)645].

[5] A. Lahiri, P.K. Roy and B. Bagchi, Int.J.Mod.Phys.**A5**(1990)1383.

[6] R. Dutt, A. Khare and U.P. Sukhatme, Am.J.Phys.**59**(1991)723.

[7] G. Lévai, J.Phys.**A22**(1989)689.

[8] D. Barclay and C.J. Maxwell, Phys.Lett.**A157**(1991)357.

[9] F. Cooper, J.N. Ginocchio and A. Khare, Phys.Rev.**D36**(1987)2458.

[10] A. Valance,T.J. Morgan and H. Bergeron, Am.J.Phys.**58**(1990)487.

[11] E. Schrödinger, Proc.R.Irish.Acad.**A46**(1940)9; *ibid.***A46**(1941)183.

[12] G. Darboux, C.R.Acad.Sci.(Paris)**94**(1882)1456.

[13] L. Infeld and T.E.Hull, Rev.Mod.Phys.**23**(1951)21.

[14] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series and Products*, Academic, 1980.

[15] P.G.L. Leach, J.Math.Phys.**25**(1984)2974.

[16] A. Comtet, A.D. Bandrauk and D.K. Campbell, Phys.Lett.**B150**(1985)159.

[17] J. Dunham, Phys.Rev.**41**(1932)721.

[18] R. Dutt, A. Khare and U.P. Sukhatme, Phys.Lett.**B181**(1986)295.

[19] R. Adhikari, R. Dutt, A. Khare and U.P. Sukhatme, Phys.Rev.**A38**(1988)1679.

[20] S.S Vasan, M. Seetharaman and K. Raghunathan, J.Phys.**A21**(1988)1897.

[21] A. Khare, Phys.Lett.**B161**131.

[22] A. Khare and Y.P. Varshni, Phys.Lett.**A142**(1989)1.

[23] N. Froman and P.O. Froman, *JWKB Approximation; Contributions to the Theory*, North-Holland, 1965.

[24] F.W.J. Olver, *Introduction to Asymptotics and Special Functions*, Academic, 1974.

[25] P.N. Argyres, Physics **2**(1965)131.

[26] C. Rosenzweig and J.B. Krieger, J.Math.Phys.**9**(1968)849.

[27] K. Raghunathan, M. Seetharaman and S.S Vasan, Phys.Lett.**B188**(1987)351.

[28] D. DeLaney and M.M. Nieto, Phys.Lett.**B247**(1990)301.

[29] F.S. Acton, *Numerical Methods That Work*, Harper & Row, 1970.

[30] C.M. Bender, K. Olaussen and P.S. Wang, Phys.Rev.**D16**(1977)1740.

[31] A.Z. Capri, *Nonrelativistic Quantum Mechanics*, Benjamin/Cummings, 1985, p.391.

[32] P. West, *Introduction to Supersymmetry and Supergravity*, World Scientific, 1986, p.207.

[33] P.M. Stevenson, Ann.Phys.**132**(1981)383.

[34] L.H. Ryder, *Quantum Field Theory*, Cambridge University Press, 1985;
R.J. Rivers, *Path Integral Methods in Quantum Field Theory*, Cambridge University Press, 1987.

[35] C. Itzykson and J.B. Zuber, *Quantum Field Theory*, McGraw-Hill, 1980.

[36] A. Pais, *Inward Bound*, Oxford University Press, 1986, p.497 ff.

[37] J.C. Collins, *Renormalisation*, Cambridge University Press, 1984.

[38] S. Coleman, *Aspects of Symmetry*, Cambridge University Press, 1985, chap. 3.

[39] A. Pais, *op. cit.*, p.462n.

[40] P.M. Stevenson, Phys.Rev.**D23**(1981)2916.

[41] R.P. Feynman, *QED:The Strange Theory of Light and Matter*, Princeton University Press, 1985.

[42] A. Pickering, *Constructing Quarks*, Edinburgh University Press, 1984.

[43] G 't Hooft, Nucl.Phys.**B33**(1971)173; *ibid* **B35**(1971)167.

[44] J.C. Collins, *op. cit.*, pp.118-9.

[45] F. Halzen and A.D. Martin, *Quarks and Leptons*, John Wiley, 1984.

[46] S. Weinberg, Phys.Rev.Lett.**19**(1967)1264;
A. Salam, in *Elementary Particle Theory*, ed. N. Svartholm, Almqvist and Wiksell, 1968, p.367.

[47] U. Amaldi, W. de Boer and H. Furstenau, Phys.Lett.**B260**(1991)447.

[48] D.W. Duke and R.G. Roberts, Phys.Rep.**120**(1985)275.

[49] P.M. Stevenson, Nucl.Phys.**B203**(1982)472.

[50] A.J. Buras, E.G. Floratos, D.A. Ross and C.T. Sachrajda, Nucl.Phys.**B131** (1977)308.

[51] W. Celmaster and R.J. Gonsalves, Phys.Rev.**D21**(1980)3112.

[52] A. Dhar, Phys.Lett.**B128**(1983)407;
C.J. Maxwell, Phys.Rev.**D29**(1984)2884;
P.M. Stevenson, Phys.Rev.**33**(1986)3130.

[53] G. Grunberg, Phys.Lett.**B95**(1980)70; Phys.Rev.**D29**(1984)2315.

[54] JADE Collaboration, S. Bethke et al., Phys.Lett.**B213**(1988)325.

[55] S. Bethke, J.Phys.**G17**(1991)1455.

[56] S. Bethke, Z. Kunszt, D.E. Soper and W.J. Stirling, Nucl.Phys.**B370**(1992)310.

[57] A. Ali et al., Nucl.Phys.**B167**(1980)454;
R.K. Ellis, D.A. Ross and A.E. Terrano, Nucl.Phys.**B178**(1981)421;
K. Fabricius, G. Kramers, G. Schierholz and I. Schmitt, Z.Phys.**C11**(1982)315;
K. Hagiwara and D. Zeppenfeld, Nucl.Phys.**B313**(1989)560.

[58] G. Kramer and B. Lampe, Z.Phys.**C39**(1988)101.

[59] C.J. Maxwell, Phys.Lett.**B225**(1989)425.

[60] S. Catani et al., Phys.Lett.**B263**(1991)491;
S. Catani, G. Turnock and B.R. Webber, Phys.Lett.**B272**(1991)368.

[61] S. Catani, CERN preprint, CERN-TH.6281/91(1991).

[62] S. Catani et al., Phys.Lett.**B269**(1991)432.

[63] N. Brown and W.J. Stirling, Phys.Lett.**B252**(1990)657.

[64] Yu.L. Dokshitzer, Workshop on Jets at LEP and HERA, Durham, 1990.

[65] Z. Kunszt and P. Nason, in Z. Kunszt et al.,*Z Physics at LEP1*, CERN Yellow Report 89-08, 1989, Vol.1.

[66] L.R. Surguladze and M.A. Samuel, Phys.Rev.Lett.**66**(1991)560;
S.G. Gorishny, A.L. Kataev and S.A. Larin, Phys.Lett.**B259**(1991)144.

[67] G. d'Agostini, W. de Boer and G. Grindhammer, Phys.Lett.**B229**(1989)160.

[68] D. Bardin et al., Nucl.Phys.**B351**(1991)1.

[69] K.G. Chetyrkin and J.H.Kühn, Phys.Lett.**B248**(1990)359.

[70] OPAL Collaboration, P.D. Acton et al., CERN-PPE/92-18.

[71] DELPHI Collaboration, P. Abreu et al., CERN-PPE/91-181/Rev.

[72] OPAL Collaboration, M.Z. Akrawy et al., Phys.Lett.**B235**(1990)389.

[73] S.J. Brodsky, G.P. Lepage and P.B. Mackenzie, Phys.Rev.**B28**(1983)228;
J.C. Collins, J.Phys.**G17**(1991)1547.

[74] A.D. Martin, W.J. Stirling and R.G. Roberts, Phys.Lett.**B266**(1991)173; Univ. of Durham preprint, DTP-92/16, RAL-92-021(1992).

[75] J.D. Barrow and F.J. Tipler, *The Anthropic Cosmological Principle*, Oxford University Press, 1986, *passim.*

[76] J.M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons*, Springer, 1976, p.246.

[77] J.C. Collins, *Renormalisation*, Cambridge University Press, 1984, pp.203-6.

[78] J.D. Barrow and F.J. Tipler, *op. cit.*, pp.291-5;
H. Reichenbach, *The Philosophy of Space and Time*, Dover, 1957.

[79] D.T. Barclay and C.J. Maxwell, Univ. of Durham preprint, DTP-92/26(1992), submitted to Phys.Lett.**B**.

[80] H.D. Politzer, Nucl.Phys.**B194**(1982)493;
M.R. Pennington, Phys.Rev.**D26**(1982)2048;

J.C. Wrigley, Phys.Rev.**D27**(1983)1965.

[81] J. Chýla, Phys.Rev**D31**(1985)2690.

[82] A. Dhar and V. Gupta, Phys.Rev.**D29**(1984)2822.

[83] A.X. El-Khadra, G. Hockney, A.S. Kronfelt and P.B. Mackenzie, FNAL preprint, FERMILAB-PUB-91/354-T(1992).

[84] R.P Feynman, Talk at Solvay Conference, 1959.

[85] A.H. Mueller, Nucl.Phys.**B250**(1985)327.

[86] G.B. West, Phys.Rev.Lett.**67**(1991)1388; Erratum **67**(1991)3732.

[87] P.M. Stevenson, Nucl.Phys.**B231**(1983)65.

[88] F.J. Dyson, Phys.Rev.**85**(1952)631.

[89] L.N. Lipatov, Zh.Eksp.Teor.Fiz.**72**(1977)411 [Sov.Phys.JETP.**45**(1977)216].

[90] J. Zinn-Justin, *Quantum Field Theory and Critical Phenomena*, Oxford University Press, 1989.

[91] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover, 1964.

[92] G 't Hooft, in *Orbis Scientiae: Deeper Pathways in High-Energy Physics*, Proceedings, Coral Gables, 1977, ed. by B. Kursunoglu, A. Perlmutter and W. Scott, Plenum, 1977.

[93] G. 't Hooft, in *The Whys of Subnuclear Physics*, Proceedings, Erice, 1977, ed. by A.Zichichi, Subnuclear Series Vol. 15, Plenum, 1977.

[94] L.S. Brown and L.G.Yaffe, Phys.Rev.**D45**(1992)R398.

[95] A.H. Mueller, Talk at Aachen Conference, 1992.

[96] B. Lautrup, Phys.Lett.**B69**(1977)109.

[97] C. Litwin, Phys.Lett.**B80**(1979)236.

[98] G. Parisi, in *Hadron Structure and Lepton-Hadron Interactions*, Proceedings, Cargèse, 1977, ed. by M. Lévy, Plenum, 1979.

[99] G. Parisi, Phys.Lett.**B76**(1978)65.

[100] G. Parisi, Phys.Rep.**49**(1979)215.

[101] G. Parisi, Nucl.Phys.**B150**(1979)163.

[102] M.C. Bergère and F. David, Phys.Lett.**B135**(1984)412.

[103] D.T. Barclay and C.J. Maxwell, Phys.Rev.**D45**(1992)1760.

[104] G. Altarelli, Talk at Aachen Conference, 1992.

[105] L.S. Brown, L.G. Yaffe and C. Zhai, Univ. of Washington preprint, UW-PT-92-07(1992).

[106] G.B. West, in Radiative Corrections, Proceedings of NATO workshop, Sussex, 1989, ed. by N. Dombey and F. Boudjema, Plenum, 1990.

[107] S.G. Gorishny, A.L. Kataev and S.A. Larin, Phys.Lett.**B212**(1988)238.

[108] D.T. Barclay and C.J. Maxwell, Univ. of Durham preprint, DTP-91/72(1991); to be published Phys.Rev.Lett.

[109] M.A. Samuel and E. Steinfelds, Oklahoma State Univ. preprint, OSU-RN-264(1992), submitted to Phys.Rev.Lett.

[110] A. Duncan and R. Willey, submitted to Phys.Rev.Lett.

[111] G.B. West, to be published Phys.Rev.Lett.

[112] J.D. Bjorken and S.D. Drell, *Relativistic Quantum Fields*, McGraw-Hill, 1965, chap. 18.

[113] I.S. Gradshteyn and I.M. Ryzhik, *op. cit.*, p.1075;
A. Erdélyi et al, *Higher Trancendental Functions*, McGraw-Hill, 1953, p.27;
C.A. Truesdell, Ann.of Math.**46**(1945)144.

[114] J. Zinn-Justin, Phys.Rep.**49**(1979)205.

[115] A.L. Kataev and M.D. Vardiashvili, Phys.Lett.**B221**(1989)377.

[116] T.J.I. Bromwich, *An Introduction to the Theory of Infinite Series*, Macmillan, 1926.

[117] J. Hagen, Amer.J.of Math.**5**(1882)236.

[118] D.E. Knuth, *The Art of Computer Programming*, Addison-Wesley, 1981, vol.2, p.506 ff.

[119] D.C. Curtis, private communication.

[120] L.B. Rall, in *Numerical Solution of Integral Equations*, ed. by L.M. Delves and J. Walsh, Oxford University Press, 1974.

[121] R.K. Ellis and W. Giele, in progress.