IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

On the finite difference between divergent sum and integral

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1981 J. Phys. A: Math. Gen. 14 1009

(http://iopscience.iop.org/0305-4470/14/5/021)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 129.252.86.83
The article was downloaded on 30/05/2010 at 15:46

Please note that terms and conditions apply.

J. Phys. A: Math. Gen. 14 (1981) 1009–1027. Printed in Great Britain

On the finite difference between divergent sum and integral

G Barton
University of Sussex, Brighton, BN1 9QH, UK

Received 7 November 1980

Abstract. We consider how, in a reasonably motivated manner, to define and then evaluate unique finite differences between individually divergent sums $S = \Sigma^{\infty} f(n)$ and integrals $I = \int^{\infty} \mathrm{d}n\, f(n)$. This is done by first replacing $f(n)$ by $f(n|\lambda) = f(n)\, g(n|\lambda)$, where g is a cutoff function ($g(n \to \infty|\lambda) = 0$) which obeys the permanence condition $g(n|\lambda \to \infty) = 1$, and then transforming $(S - I)$ into some convenient explicit functional $\mathscr{D}\{f(n|\lambda)\}$ which admits $\lim_{\lambda \to \infty} \mathscr{D}\{f(n|\lambda)\} = \mathscr{D}\{\lim_{\lambda \to \infty} f(n|\lambda)\} = \mathscr{D}\{f(n)\}$, where the final expression converges. There ought but there seems not to exist a convenient yet reasonably general theory for identifying admissible classes of summands and cutoffs and for deriving $\mathscr{D}\{f\}$, even though physicists have long dealt with simple cases *ad hoc*. No general theory is supplied here either, but three specialised prescriptions are presented for $\mathscr{D}\{f\}$; one based on the Abel–Plana formula, for suitably analytic f and g; another, less convenient, based on the Euler–Maclaurin formula for merely differentiable f and g; and the unconventional 'ε-averaging method' which this writer has used before but without detailed justification; it avoids the explicit introduction of a cutoff, and is especially convenient when it can be implemented at all, namely when S and I with finite limits are expressible in terms of familiar functions. The mutual compatibility of these methods is discussed. An example illustrates how they break down if the underlying cutoff violates some of the necessary conditions, even though it obeys the permanence condition and even though $\mathscr{D}\{f(n)\}$ converges. As illustrations and for the record, explicit differences are worked out for logarithmic, power-law, and exponential summands.

1. Introduction

In calculations on electromagnetic fields confined to limited regions, one often needs to sum certain functions of frequency over all normal modes, and then to subtract the corresponding sum (or rather integral) for an equal volume forming part of unbounded space. We shall symbolise such differences by D. For instance, when the summand is the quantum zero-point energy (i.e. half the frequency), D determines the Casimir effect whose archetype is the attraction between parallel perfectly conducting plates (Casimir 1948, see also Power 1964). When D represents the changes in the energy levels of a neutral molecule inserted between such plates (Barton 1970, 1979a), the summand is more complicated, containing contributions from both the electrostatic and the transverse-photon Green functions; and somewhat similar summands enter the recent calculation by Unwin and Critchley (1980) of the difference between the ground-state Lamb shifts of an atom situated in ordinary space and in certain multiply-connected spaces.

Unfortunately, in practically every interesting case the sum and the integral in question, taken separately, diverge at high frequency; hence some further physically

0305-4470/81/051009+19\$01.50 © 1981 The Institute of Physics

motivated prescription is needed to make D mathematically well-defined and calculable. For any given system it is usually not too difficult to devise a satisfactory prescription *ad hoc*, by analysing just how the normal-mode decomposition of the field propagators enters the formulae, and then subtracting the corresponding expression constructed with the unconfined (free-space) propagators (see for instance Brown and Maclay 1969, Unwin and Critchley 1980; and in the spherical-shell problem, Milton *et al* 1978). But there is another point of view from which, paradoxically, such a detailed approach overshoots the mark: for often one expects on physical grounds that modes with very high frequencies (i.e. short wavelengths) should be irrelevant, and that, subject only to some weak restrictions, *any* method should serve which eliminates their contributions. For example, at short wavelengths any conducting plates become transparent, and one feels that the precise manner in which transparency sets in should be irrelevant, say to the Casimir effect, at plate separations large compared with the wavelengths at which it does set in. Often in such cases it seems natural to multiply the summand by some cutoff function (convergence factor), to calculate the requisite difference between sum and integral, and then to let the cutoff frequency recede to infinity (take the no-cutoff limit). For instance, this is the attitude to the Casimir effect taken by Fierz (1960) and Lukosz (1971), who adopt a specially convenient cutoff (namely an exponential), clearly implying that no generality is lost by this specific choice; it is also the attitude taken by Boyer (1968) to the case of a spherical shell.

Appeal to a cutoff immediately raises two interrelated problems. The first is a question of principle: given the summand, just how weak mathematically can the restrictions on classes of physically reasonable cutoff functions be, while still guaranteeing that all cutoffs in the class yield the same result for D in the no-cutoff limit? The second problem is more practical: to find formulae from which one can actually evaluate the no-cutoff limit of D without any explicit mention of the cutoff. This writer is not equipped to tackle the first problem with the generality one could wish for, and the present paper is devoted mainly to the second problem. However, it turns out that by solving the second problem subject to relatively stringent conditions (differentiability or analyticity) on both cutoff and summand, some limited light can be cast even on the first problem.

In its simplest form, i.e. for one-dimensional rather than multiple sums, the mathematical problem so set up is, luckily, self-contained, and here we shall not pursue its implications for other and more difficult questions about surface effects in fields confined to multidimensional finite regions of various shapes, and obeying various boundary conditions; for these wider problems see for instance Balian and Bloch (1970, 1971, 1972); Balian and Duplantier (1977, 1978); Barton (1979b); and especially the clear, comprehensive and fully referenced review by Baltes and Hilf (1976).

Section 2 formulates the mathematical question, after a brief sketch of its provenance, and introduces the cutoff functions which play a crucial role in the argument although they are absent from the final expressions. Since we lack a general theory, we can provide only partial answers, by establishing various explicit formulae for D; their region of validity, and the method for obtaining them, naturally depend on the rate of growth of the summand, and on the assumptions made about its smoothness and about the cutoff function. Section 3 derives a generalised form of the Abel–Plana formula, which is perhaps the most useful single result, and which solves the problem subject to fairly strong assumptions about analyticity. Section 4 describes a new and unconventional cutoff procedure called the 'ε-averaging method'. This is usually the easiest method to implement provided that, without cutoff factors, the partial sum and integral can be found in terms of familiar functions. The writer has used this method before, though without detailed justification (Barton 1970, 1979b; Babiker and Barton 1972a,b); its compatibility with the Abel–Plana method is discussed briefly at the start of § 4, and its links with a third method in appendix 1. Appendix 1 describes this third method, based on the famous Euler–Maclaurin formula, and applicable to differentiable functions with suitably bounded higher derivatives. Though less flexible than the two preceeding methods this has perhaps been the most widely used, probably because of its distinguished ancestry, and because often it does readily yield the asymptotic expansion of D in powers of some suitable parameters like the size of the system, which may be all that is required. Further, the Euler–Maclaurin prescription is interesting as a matter of principle, because it can yield convergent expressions for D under weaker conditions than the Abel–Plana method, even if the result is less convenient for actually evaluating D. The appendix also shows that the ε-averaging and the Euler–Maclaurin methods are compatible; the compatibility of the latter with the Abel–Plana method is discussed by Hardy (1949). Section 5 illustrates the methods of §§ 3 and 4, and derives explicit expressions for D in the simplest and most commonly required cases of logarithmic, power-law, and exponential summands. Some illustrations of the Euler–Maclaurin method are given in appendix 1. Appendix 2 gives a cautionary example of an illegitimate cutoff function.

Of course yet other methods can be and have been developed, e.g. from the Poisson summation formula (see also Unwin and Critchley 1980); but, though useful in some problems, they are generally less flexible than those considered here.

Finally we risk stressing again at the start what will certainly become obvious by the end of the paper, that for the differences in the title we have failed to supply as general and elegant a theory as they deserve, and as deeper mathematics could no doubt provide. From such a theory the partial results given here should eventually follow as easily recognisable special cases, and the compatibility of the various explicit prescriptions for D should be demonstrable from the outset, unhedged by the various special provisos on which the arguments in this paper have to rely. However, the summands that we can accommodate, and the restricted classes of cutoffs specified as we go along, do more than cover the problems that have occurred so far in applications.

2. Formulation of the problem

To see the background of the problem, consider for simplicity the scalar wave equation in one dimension, $\partial^2\psi/\partial x^2 - c^{-2}\,\partial^2\psi/\partial t^2 = 0$, subject to periodic boundary conditions† over a length L: $\psi(0, t) = \psi(L, t)$. The normal modes are $\psi_n = \exp \mathrm{i}(k_n x - \Omega_n t)$, where $k_n = 2\pi n/L$ and $\Omega_n = c|k_n|$. Defining, for any function $F(\Omega)$ of frequency, $F(\Omega_n) \equiv f(n)$, a sum of F over all modes reads $\Sigma_{n=-\infty}^{\infty} f(n)$. If, as we now assume, F is a smooth function of Ω, then, as L increases and the spacing between successive allowed values of Ω shrinks indefinitely, $\Sigma_{-\infty}^{\infty} f(n)$ may be replaced by $\int_{-\infty}^{\infty} \mathrm{d}n\, f(n) = (L/2\pi)\int_{-\infty}^{\infty} \mathrm{d}k\, F(c|k|)$; the last form shows that in the limit the sum becomes proportional to L, which is the physical reason why we interpret the integral $\int \mathrm{d}n\, f(n)$ as the contribution to the 'sum' over all modes ascribable to each unit length of the unbounded x axis. For convenience we shall work with *half* this sum; we call it S, the corresponding

† Non-scalar fields, or different boundary conditions, or both, make a great difference in a wider context, (cf the fourth paragraph of § 1), but no essential difference to the specific mathematical question to be considered here.

integral I, and their difference D, all these being functionals of $f(n)$. (When the functional dependence of D on f needs stressing, we shall write $D=\mathscr{D}\{f\}$; otherwise we ease the notation by retaining the symbol D.) Thus:

$$D=S-I=\sum_{n=0}^{\infty}{}' f(n)-\int_0^{\infty} \mathrm{d}n\, f(n). \tag{2.1}$$

The primed sum, and a doubly-primed sum needed later, are defined by

$$\sum_{n=0}^{\infty}{}' f(n)=\tfrac{1}{2}f(0)+\sum_{n=1}^{\infty} f(n), \tag{2.2}$$

$$\sum_{n=N}^{M}{}'' =\tfrac{1}{2}f(N)+\sum_{h=n+1}^{M-1} f(n)+\tfrac{1}{2}f(M). \tag{2.3}$$

Naturally, the definition (2.1) makes sense only if S and I converge†. But in many physically relevant cases they diverge; simple examples of $f(n)$ which we shall consider are

$$(n+\eta)^{-1};\ \ln(n+\eta);\ n^{p};\ \exp(\alpha n);\ \exp(n^{p}), \tag{2.4}$$

where η, p, α are constants‡. Then the problem is modified by replacing $f(n)$ in (2.1) by

$$f(n|\lambda)=f(n)g(n|\lambda) \tag{2.5}$$

and (2.1) itself by

$$D(\lambda)=S(\lambda)-I(\lambda)=\sum_{n=0}^{\infty}{}' f(n|\lambda)-\int_0^{\infty} \mathrm{d}n\, f(n|\lambda). \tag{2.6}$$

The cutoff function $g(n|\lambda)$ may depend on one, or on several, or even on an infinity of parameters collectively denoted by the single index λ; it must be a sufficiently differentiable function of n, and together with its derivatives $g^{(S)}\equiv\partial^{S}g/\partial n^{S}$ it must satisfy several conditions. Some of these conditions we state now; others, peculiar to particular methods for obtaining D, will emerge retrospectively later on.

First, it must be possible to define a no-cutoff limit, represented symbolically as $\lambda\to\infty$, with the property§

$$\left.\begin{array}{ll} g(n|\lambda)\to 1 & \qquad (2.7a)\\ g^{(S)}(n|\lambda)\to 0, \quad s=1,2,3,\ldots & \qquad (2.7b)\end{array}\right\}\ \text{as } \lambda\to\infty \text{ for fixed } n.$$

Second, as $n\to\infty$ for fixed λ, g and the $g^{(s)}$ must vanish fast enough to ensure the convergence of sums and integrals where they occur. Third, in the end-result of each method (as symbolised in equation (2.8) below) such convergence must be uniform in λ, this being in effect a retrospective condition on g. In § 3, certain analyticity conditions

† Some elementary methods for evaluating D in such cases are reviewed by Boas and Stutz (1971); for others see Hardy (1949).

‡ In practice, a parameter like η is usually proportional to the system size L, since the expression $(n+\eta)$ would enter the original summand F as $(k_n+K)=(2\pi/L)(n+LK/2\pi)$, with K some characteristic dimensional parameter independent of L. Similarly α would be proportional to L^{-1}. Consequently, extreme values of L call for asymptotic approximations for D.

§ The conditions (2.7) are evident analogs of the familiar 'permanence' conditions of summability methods, ensuring that in the limit the cutoff prescription does not alter the value of D in cases where S and I are well defined even without a cutoff. Unfortunately the writer has not succeeded in extracting from the analogies with summability theories any help with the present problem.

will be imposed as well. Perhaps the most popular example of a cutoff is $\exp(-n/\lambda)$, with $\lambda \to \infty$ understood literally.

With the cutoff, $S(\lambda)$, $I(\lambda)$ and $D(\lambda)$ in (2.6) are all finite but functions of λ. The general mathematical problem would be to explore the most liberal conditions on $f(n)$ and $g(n|\lambda)$ under which $\lim_{\lambda\to\infty} D(\lambda) \equiv D$ exists and has a finite value independent of any further details of g. But, as explained in § 1, here we pursue only a more modest goal: namely, under suitable conditions on f and g, to *transform* (2.6), while λ is still finite, into some more convenient explicit functional of f:

$$D(\lambda) = \mathscr{D}\{f(n|\lambda)\} \tag{2.8}$$

such that

$$\lim_{\lambda\to\infty} \mathscr{D}\{f(n|\lambda)\} = \mathscr{D}\{\lim_{\lambda\to\infty} f(n|\lambda)\} = \mathscr{D}\{f(n)\} \equiv D \tag{2.9}$$

makes sense. In particular, the final form $\mathscr{D}\{f(n)\}$ must converge, which is a condition on $f(n)$ alone; but beyond this, conditions on $g(n|\lambda)$ are implied by the existence of the leftmost limit, and by the postulated uniformity of convergence which legitimises the first equality in (2.9). Appendix 2 illustrates what can happen if these less obvious constraints are ignored. For classes of summands f and cutoffs g for which our programme can be implemented, it provides a partial answer to the more general question too, since the end result in (2.9) is manifestly independent of the underlying $g(n|\lambda)$. Nevertheless the physical relevance of the result must remain doubtful unless a reasonably wide class of cutoffs can be identified at least in principle.

As a preliminary it is often convenient to subdivide $D(\lambda)$ in (2.6) as follows, at some point N independent of λ (L and R stand for left and right):

$$D(\lambda) = D_{\mathrm{L}}(N|\lambda) + D_{\mathrm{R}}(N|\lambda), \tag{2.10}$$

$$D_{\mathrm{L}}(N|\lambda) = \sum_{n=0}^{N}{}'' f(n|\lambda) - \int_0^N \mathrm{d}n f(n|\lambda), \tag{2.11}$$

$$D_{\mathrm{R}}(N|\lambda) = \sum_{n=N}^{\infty}{}' f(n|\lambda) - \int_N^\infty \mathrm{d}n f(n|\lambda). \tag{2.12}$$

Observe that $D(\lambda) = D_{\mathrm{L}}(\infty|\lambda) = D_{\mathrm{R}}(0|\lambda)$. In D_{L} with its finite upper limit N the no-cutoff limit can be taken trivially:

$$D_{\mathrm{L}}(N|\infty) = \sum_{n=0}^{N}{}'' f(n) - \int_0^N \mathrm{d}n f(n). \tag{2.13}$$

The point is that any simplifying assumptions about f need apply only for $n \geqslant N$. For instance, in this paper we need not pursue the implications of any singularities in $f(n)$ at finite n, such for instance as ensue if the parameter η in (2.4) is negative; though these can be important in applications (Barton 1970), from our present point of view all such singularities can be confined to the nonproblematic component D_L of D. Similarly, we are not now interested in any divergences as $n \to 0$, and in case of difficulty at the origin (e.g. if $f(n)$ is an inverse power) we simply confine attention to D_{R} with $N = 1$, say†.

Two general points need stressing. First, in defining the problem the summand $f(n)$ is to be specified for (at least) all real positive n; in other words, both the summand of S

† We shall write $D_{\mathrm{R}}(N|\infty)$ simply as $D_{\mathrm{R}}(N)$ whenever the context clearly identifies the argument as N rather than λ.

and the integrand of I are given *a priori*, and we need not struggle to render unique the continuation of $f(n)$ from integer to continuous n.†

Second, we must note an important logical reservation in cases where the rate of growth of the summand at infinity is controlled by a numerical parameter, like p or α in the examples (2.4). If D can be evaluated in closed form as a function of this parameter, say p, in a region where S and I, separately, converge, then it becomes tempting to jump to the conclusion that analytic continuation in p must yield the correct value of D for all values of p. Although borne out in the simple examples we study in § 5, at present this expectation lacks the force of argument, since on physical grounds D is defined through the no-cutoff limit, as explained earlier in this section, and since analytic continuation in such a parameter has not so far been justified from the basic definition, except *a posteriori* in special cases. To provide a reasonably general justification under stated conditions is an interesting problem but beyond our scope here. Meanwhile, divergent cases must be investigated in their own right, even when analytic continuation happens to provide a heuristic preliminary guess.

3. The generalised Abel–Plana formula

The basic idea is to establish a convenient formula for $D_R(N|\lambda)$, equation (2.12), by requiring $f(n|\lambda)$, and therefore f and g, to have rather strong analyticity properties, as specified below. For mathematical details we refer to Hardy (1949), who derives the Abel–Plana (AP) formula (3.4) in full; here we shall eventually need the generalised form (3.3), of which (3.4) is a special case.

Let $f(z|\lambda)$ be an analytic function of $z=x+\mathrm{i}y$, free of singularities (except at infinity, when the no-cutoff limit is taken later), either in the positive half-plane $x\geqslant x_0$ for some $x_0<N$ (N is some non-negative integer), or at least to the right of the wedge formed by the lines $y=\pm(x-x_0)\tan\phi$, $0<\phi\leqslant\pi/2$, where the requisite values of ϕ will appear from the context; and let $f(z|\lambda)$ vanish fast enough for subsequent contour integrals to draw no contributions from their arcs at infinity.

The integral on the right of (2.12) can now be rewritten as half the sum of two other integrals, running rightwards from $z=N$ to infinity along straight lines $L_\pm$ of slopes $\pm\tan\phi$ respectively:

$$\int_N^\infty \mathrm{d}n f(n|\lambda)=\tfrac{1}{2}\int_0^\infty \mathrm{d}\rho\{\mathrm{e}^{\mathrm{i}\phi}f(N+\rho\,\mathrm{e}^{\mathrm{i}\phi}|\lambda)+\mathrm{e}^{-\mathrm{i}\phi}f(N+\rho\,\mathrm{e}^{-\mathrm{i}\phi}|\lambda)\}. \tag{3.1}$$

Next, the as yet unprimed sum in (2.12) is rewritten as $(2\pi i)^{-1}\oint \mathrm{d}z\,\pi\cot(\pi z)f(z|\lambda)$ along an anticlockwise contour enclosing the real axis from $z=N$ to $+\infty$; and this contour is then distorted to run from infinity to $N+\varepsilon\exp(\mathrm{i}\phi)$ along the line L_+, anticlockwise along a circular arc of radius ε from $N+\varepsilon\exp(\mathrm{i}\phi)$ to $N+\varepsilon\exp\mathrm{i}(2\pi-\phi)$, and from there to infinity along the line L_-. In the limit $\varepsilon\to0$ this yields

$$\sum_{n=N}^{\infty} f(n|\lambda)=f(N|\lambda)(1-\phi/\pi)+(2\mathrm{i})^{-1}\int_0^\infty \mathrm{d}\rho\{-\mathrm{e}^{\mathrm{i}\phi}\cot(\pi\rho\,\mathrm{e}^{\mathrm{i}\phi})f(N+\rho\,\mathrm{e}^{\mathrm{i}\phi}|\lambda)$$
$$+\mathrm{e}^{-\mathrm{i}\phi}\cot(\pi\rho\,\mathrm{e}^{-\mathrm{i}\phi})f(N+\rho\,\mathrm{e}^{-\mathrm{i}\phi}|\lambda)\}, \tag{3.2}$$

where we have used $\cot\pi(N+z)=\cot\pi z$. Substituting (3.1) and (3.2) into (2.12),

† In § 4 we do have to continue the partial sums $S(\nu)=\Sigma'^{\nu}_{n=1}f(n)$ to continuously variable values of ν.

including the correction term $-f(N|\lambda)/2$, and rearranging, we obtain

$$D_{\mathrm{R}}(N|\lambda)=f(N|\lambda)(\tfrac{1}{2}-\phi/\pi)+\int_0^\infty \mathrm{d}\rho\left\{\frac{\mathrm{e}^{\mathrm{i}\phi}f(N+\rho\,\mathrm{e}^{\mathrm{i}\phi}|\lambda)}{[\exp(-2\pi\mathrm{i}\rho\,\mathrm{e}^{\mathrm{i}\phi})-1]}+\frac{\mathrm{e}^{-\mathrm{i}\phi}f(N+\rho\,\mathrm{e}^{-\mathrm{i}\phi}|\lambda)}{\exp(2\pi\mathrm{i}\rho\,\mathrm{e}^{-\mathrm{i}\phi})-1}\right\}. \tag{3.3}$$

This is the generalised AP formula, i.e. the AP prescription for the functional $\mathscr{D}\{fn|\lambda)\}$ of equation (2.8). Accordingly, the no-cutoff limit $D_{\mathrm{R}}(N|\infty)$ is obtained simply by replacing $f(z|\lambda)\to f(z)$ on the right of (3.3), provided the result converges; the underlying class of cutoffs contains those which have the required analyticity properties and which can ensure the convergence of $S(\lambda)$ and $I(\lambda)$ as well as the uniform convergence of (3.3). Appendix 2 shows what havoc can be created by a cutoff with the wrong analyticity properties.

Most often one uses $N=1$ or $N=0$; the original AP formula as given by Hardy (1949) is the special case with $N=1$ and $\phi=\pi/2$, namely

$$D_{\mathrm{R}}(1|\infty)=\mathrm{i}\int_0^\infty \mathrm{d}\rho\frac{1}{\mathrm{e}^{2\pi\rho}-1}\{f(1+\mathrm{i}\rho)-f(1-\mathrm{i}\rho)\}. \tag{3.4}$$

Since $D=D_{\mathrm{R}}(0)$ we can, if the result makes sense, write D itself as

$$D=\mathrm{i}\int_0^\infty \mathrm{d}\rho\frac{1}{\mathrm{e}^{2\pi\rho}-1}\{f(\mathrm{i}\rho)-f(-\mathrm{i}\rho)\}. \tag{3.5}$$

Section 5 gives explicit applications, but some preliminary examples may help in assessing the scope of these formulae. Logarithms and powers are covered by (3.4) or (3.5); to verify that satisfactory underlying cutoffs *exist*, one need think only of $g=\exp(-n/\lambda)$. These prescriptions cover also exponentials $f=\exp[(a+\mathrm{i}b)n]$ provided $|b|<2\pi$; but with $|b|\geqslant 2\pi$ they diverge, and so, irremediably, do the generalised formulae (3.3) (with $f(z|\lambda)\to f(z)$) for any allowed choice of $\phi(0<\phi\leqslant\pi/2)$. In this case we are forced to fall back on the ε-averaging method described in the next section. On the other hand, it is easily verified that (3.3) converges even for $f=\exp(an^p)$ with real $a>0$ and any $p>1$, provided we chose ϕ so that $\cos(p\phi)<0$, i.e. $\pi/2p<\phi<3\pi/2p$. With exponential summands it takes a little more care to see that suitable underlying cutoffs could be found. Perhaps it is simplest to think of $g=\exp[-(n/\lambda)^{p+\delta}]$, with small positive δ; such a δ restricts the range allowed for ϕ, but there is no difficulty in finally taking the limit $\delta\to 0$, thus arriving at the choices given above†.

4. The ε-averaging method

We proceed to describe a somewhat unorthodox method which has some advantages over the AP prescription:

(i) it may have more intuitive appeal;

(ii) it covers some summands for which the AP prescription diverges, e.g. $\exp[(a+\mathrm{i}b)n]$ with $|b|\geqslant 2\pi$; and

† Taking the limit $\delta\to 0+$ is just a calculational device making it easier to evaluate the final integral, whose value is independent of δ; this limit has nothing to do with the no-cutoff limit which is taken earlier, namely when $f(z|\lambda)$ in (3.3) is replaced by $f(z)$.

(iii) it is often easier to implement, as we shall see in § 5, especially if one is armed with Stirling's formula but lacks a comprehensive table of integrals for dealing with those thrown up by the AP prescription. Compatibility with the AP method (when both methods converge) is proved here only for the restricted class of summands expressible as Laplace transforms, $f(n)=\int \mathrm{d}sF(s)\exp(sn)$; then the two prescriptions agree because, as we shall see from §5 and from equation (4.10) below, they yield the same result for $\exp(sn)$. Appendix 1 contains a somewhat more general proof of compatibility with the Euler–Maclaurin method; and we note for completeness that the compatibility of the latter with the AP prescription (when both converge) can be readily established by adapting the method given by Hardy (1949) (though he spells it out only for the case where S and I themselves converge).

The basic idea is to start with the particularly simple cutoff

$$g(n|\lambda)=\theta(\lambda-n) \tag{4.1}$$

where θ is the step function ($\theta(x\geqslant 1)=1$, $\theta(x<0)=0$), and then to seek a way to compensate for the obviously unphysical consequences of the sharp step, while still benefiting from its simplicity.

In this section, λ denotes a single cutoff parameter; the limit $\lambda\to\infty$ is understood literally, with λ increasing *continuously*; $[\lambda]$ is the integer part of λ, and we define

$$\lambda\equiv[\lambda]+\varepsilon\equiv\nu+\varepsilon, \qquad 0\leqslant\varepsilon<1. \tag{4.2}$$

The subdivision into D_{L} and D_{R} is not required here, and we define

$$D(\lambda)=\left(\sum_{n=0}^{\infty}{}'-\int_0^{\infty}\mathrm{d}n\right)\theta(\lambda-n)f(n)$$

$$=\left(\sum_{n=0}^{\nu}{}'-\int_0^{\lambda}\mathrm{d}n\right)f(n)\equiv S(\nu)-I(\lambda), \tag{4.3}$$

where

$$S(\nu)=\tfrac{1}{2}f(0)+\sum_{n=1}^{\nu}f(n). \tag{4.4}$$

Note the arguments of $S(\nu)$ and $I(\lambda)$.

We must now improve on the obvious but naive attempt to obtain D by taking the limit $\lambda\to\infty$ of (4.3) as it stands, for in the interesting cases this limit does not exist. (The reader may appreciate the following argument more readily by explicitly working through the example $f(n)=n$.) Indeed, if we allow λ to increase continuously, we notice that, on account of its component $S(\nu)$, $D(\lambda)$ oscillates violently as λ advances from one integer to the next. Hence we try to secure convergence by suitably averaging $D(\lambda)$ over a finite range of λ before allowing $\lambda\to\infty$. The smoothing process should assign no special role to integer values of λ; we can see this by recalling from the start of § 2 that $n=L\Omega/2\pi c$ and that the physical cutoff should be a smooth function of Ω independently of the precise value of L. The method to be given arose from an attempt to implement these ideas as simply as possible; with hindsight one realises that other similar prescriptions, though perhaps equally plausible *prima facie*, would either fail to secure convergence for any wide class of summands, or would fail to produce a prescription for D compatible with the fundamental definition of § 2 in cases where the latter leads to the AP or the Euler–Maclaurin prescriptions.

Regarding the sum $S(\nu)$ in (4.4), we observe that by virtue of (4.2) ν and ε are of course functions of λ. Nevertheless we now formulate the smoothing processes as follows. Write $S(\nu) = S(\lambda - \varepsilon)$; treat ε as if it were a continuous variable independent of λ, and formally average $S(\lambda - \varepsilon)$ over the permitted range $0 \leqslant \varepsilon < 1$, keeping λ fixed; call the result $\bar{S}(\lambda)$, and substitute it for $S(\nu)$ in (4.3); and finally take the limit $\lambda \to \infty$. In other words we define

$$\bar{S}(\lambda) \equiv \int_0^1 \mathrm{d}\varepsilon\, S(\lambda - \varepsilon), \tag{4.5a}$$

$$\bar{D} \equiv \lim_{\lambda \to \infty} \{\bar{S}(\lambda) - I(\lambda)\}. \tag{4.5b}$$

The compatibility argument showing that $\bar{D} = D$ was mentioned at the beginning of this section.

When this ε-averaging method (4.5) can be implemented at all, it is generally very easy to do so; the reason lies in the fact that right from the start it is permissible to approximate all expressions appropriately to large λ and ν, and to drop all terms which can be recognised as due to vanish in the limit $\lambda \to \infty$ and $\nu \to \infty$. It is also useful to keep in mind that if $\lim_{\nu \to \infty} S(\nu) = S(\infty)$ is finite, then $\lim_{\lambda \to \infty} \bar{S}(\lambda) = S(\infty)$.

The crucial step of treating $S(\nu)$ as if ν were continuously variable clearly depends on an analytic continuation of f and S; in principle as in practice this is achieved through Taylor series, which need converge only within a circle or just over unit radius around λ, i.e. around any arbitrarily large value of n. But if for simplicity we assume or pretend for the moment that the Taylor series for $f(n)$ converges for all n, then we can gain some useful insight into the mechanism whereby the ε-averaging process secures convergence. To this end we define the derivative operator $\Delta : \Delta^r f = f^{(r)}$, and symbolise the Taylor expansion by $f(n) = \exp(n\Delta) f(0)$. Then we can write

$$S(\nu) = \left\{ -\tfrac{1}{2} + \sum_{n=0}^{\nu} \mathrm{e}^{n\Delta} \right\} f(0) = \left\{ -\tfrac{1}{2} + \frac{\mathrm{e}^{(\nu+1)\Delta} - 1}{\mathrm{e}^{\Delta} - 1} \right\} f(0). \tag{4.6}$$

To calculate $\bar{S}(\lambda)$ from (4.6) and (4.5a) we need

$$\int_0^1 \mathrm{d}\varepsilon\, \mathrm{e}^{(\lambda - \varepsilon + 1)\Delta} = \mathrm{e}^{\lambda\Delta}(\mathrm{e}^{\Delta} - 1)/\Delta;$$

substitution into (4.6) yields

$$\bar{S}(\lambda) = \left\{ -\tfrac{1}{2} + \mathrm{e}^{\lambda\Delta}\frac{1}{\Delta} - \frac{1}{\mathrm{e}^{\Delta} - 1} \right\} f(0). \tag{4.7}$$

Expressing $I(\lambda)$ in the same symbolism we have

$$I(\lambda) = \int_0^{\lambda} \mathrm{d}n\, f(n) = \left\{ \int_0^{\lambda} \mathrm{d}n\, \mathrm{e}^{n\Delta} \right\} f(0) = (\mathrm{e}^{\lambda\Delta} - 1)(1/\Delta) f(0). \tag{4.8}$$

The λ-dependent terms cancel between $\bar{S}(\lambda)$ and $I(\lambda)$, so that the subsequent limit $\lambda \to \infty$ becomes redundant; substituting (4.7) and (4.8) into (4.5) we obtain

$$\bar{D} = \left\{ -\tfrac{1}{2} + 1/\Delta - \frac{1}{\mathrm{e}^{\Delta} - 1} \right\} f(0). \tag{4.9}$$

It is easy to see that $\bar{D}_{\mathrm{R}}(N)$ is obtained from (4.9) simply by replacing $f(0)$ on the right by $f(N)$.

An immediate and important application of (4.9) is to the summand $f(n) = \exp(\alpha n)$; since this is an eigenfunction of Δ belonging to eigenvalue α, we have, rigorously in this case,

$$\bar{\mathscr{D}}\{e^{\alpha n}\} = \{-\tfrac{1}{2} + 1/\alpha - 1/(e^{\alpha} - 1)\}. \tag{4.10}$$

Of course the same formula would emerge directly from (4.5) without any appeal to the symbolic operations with Δ. It is discussed further in § 5.3.

5. Examples

We aim to illustrate the methods described in §§ 3 and 4, to compare them as to convenience, and also to assemble those results most likely to be needed in practice. The integrals we quote are standard in the sense that they can be found in large enough collections of tables (e.g. Dwight 1961 and Groebner and Hofreiter 1958 jointly); of course this does not imply that they are easy to evaluate from first principles. The examples on powers and exponentials should be considered in the light of the remarks at the end of § 2 on the lack of *a priori* justification for evaluating D by analytic continuation.

5.1. *Logarithms and related summands*

Consider the summand $f(n) = \ln(n + \eta)$, taking $\eta > 0$ to avoid singularities (but note the remarks in the paragraph below equation (2.13)).

The AP formula (3.5) gives

$$\begin{aligned} D &= i\int_0^\infty d\rho \frac{1}{e^{2\pi\rho} - 1}\{\ln(\eta + i\rho) - \ln(\eta - i\rho)\} \\ &= -2\int_0^\infty d\rho \frac{1}{e^{2\pi\rho} - 1}\tan^{-1}(\rho/\eta), \end{aligned} \tag{5.1}$$

whence

$$\mathscr{D}\{\ln(n + \eta)\} = -\{\ln\Gamma(\eta) - (\eta - \tfrac{1}{2})\ln\eta + \eta - \tfrac{1}{2}\ln 2\pi\}. \tag{5.2}$$

From this we can obtain the result for $f(n) = (n + \eta)^{-1}$ by differentiating with respect to η:

$$\mathscr{D}\{(n + \eta)^{-1}\} = -\psi(\eta) + \ln\eta - \tfrac{1}{2}\eta, \tag{5.3}$$

where $\psi(z) \equiv d\ln\Gamma(z)/dz$. For later use we record also the result for $D_R(N = 1)$, obtainable from (5.3) by replacing $\eta \to (\eta + 1)$:

$$f(n) = 1/(n + \eta): \qquad D_R(N = 1) = -\psi(\eta + 1) + \ln(\eta + 1) - \tfrac{1}{2}(\eta + 1), \tag{5.4}$$

and its limit as $\eta \to 0$,

$$f(n) = 1/n: \qquad D_R(N = 1) = \gamma - \tfrac{1}{2}, \tag{5.5}$$

where $\gamma = -\psi(1)$ is Euler's constant.

If the prospect of evaluating the integral (5.1) from first principles does not appeal, we can fall back on the ε-averaging method. The programme is to find $S(\nu) = S(\lambda - \varepsilon)$;

approximate it for large λ; average over ε to find $\bar{S}$; calculate $I(\lambda)$ for large λ; and to take the difference. Thus:

$$S(\nu) = -\tfrac{1}{2}\ln\eta + \sum_{n=0}^{\nu}\ln(n+\eta)$$

$$= -\tfrac{1}{2}\ln n + \ln\Gamma(\nu+\eta+1) - \ln\Gamma(\eta). \tag{5.6}$$

Only the term $\ln\Gamma(\nu+\eta+1)$ is affected by ε-averaging, and as $\lambda\to\infty$ we have, by Stirling's formula

$$\ln\Gamma(\lambda+\eta+1-\varepsilon) = \tfrac{1}{2}\ln 2\pi - \lambda + (\lambda-\varepsilon+\eta+\tfrac{1}{2})\ln\lambda + \mathrm{O}(\lambda^{-1}).$$

Under $\int_0^1 \mathrm{d}\varepsilon \ldots$, we have $\varepsilon\to\frac{1}{2}$ while the other terms remain unaffected. Substituting the result into (5.6) we obtain

$$\bar{S}(\lambda) = \{-\tfrac{1}{2}\ln\eta - \ln\Gamma(\eta) + \tfrac{1}{2}\ln 2\pi - \lambda + (\lambda+\eta)\ln\lambda\} + \mathrm{O}(\lambda^{-1}). \tag{5.7}$$

Next,

$$I(\lambda) = \int_0^{\lambda} \mathrm{d}n\,\ln(n+\eta) = \{(\lambda+\eta)\ln(\lambda+\eta) - \lambda - \eta\ln\eta\}$$

$$= \{\eta - \eta\ln\eta - \lambda + (\lambda+\eta)\ln\lambda\} + \mathrm{O}(\lambda^{-1}). \tag{5.8}$$

Combining (5.7) and (5.8) we see that the λ-dependent terms cancel (apart perhaps from those of $\mathrm{O}(\lambda^{-1})$), and in the limit $\lambda\to\infty$ we reproduce the result (5.2). Notice that nothing more abstruse has been needed than Stirling's formula; in particular, recondite integrals like (5.1) have been sidestepped.

5.2. *Powers: $f(n) = n^p$*

In order to allow convergent and divergent cases to be compared (i.e. to avoid the singularity at $n = 0$), in this subsection we focus attention on $D_{\mathrm{R}}(N = 1)$, (D_{R} for short), i.e. symbolically

$$D_{\mathrm{R}} = \left(\sum_1^{\infty}{}' - \int_1^{\infty}\mathrm{d}n\right)n^p = -\tfrac{1}{2} + \sum_1^{\infty} n^p - \int_1^{\infty}\mathrm{d}n\,n^p. \tag{5.9}$$

We shall need

$$\zeta(z) = \sum_{n=1}^{\infty} n^{-z}, \qquad \operatorname{Re} z > 1, \tag{5.10}$$

and the functional relation for the ζ-function:

$$\Gamma(z)\zeta(z) = \zeta(1-z)2^{z-1}\pi^z/\cos(\pi z/2); \tag{5.11}$$

and recall that $\zeta(z)$ has a pole at $z = 1$, where it behaves like

$$\zeta(z) = 1/(z-1) + \gamma + \mathrm{O}(|z-1|). \tag{5.12}$$

5.2.1. When $p < -1$, sum and integral converge separately, and

$$D_{\mathrm{R}} = -\tfrac{1}{2} + \zeta(-p) + 1/(p+1). \tag{5.13}$$

5.2.2. When $p = -1$, the result is already known from (5.5); by virtue of (5.12) it agrees with what one might have guessed (but not established) by analytic continuation from the convergent case (5.13).

5.2.3. When $p>-1$ we need one of our prescriptions. The ε-averaging method applied directly is not convenient because there is no simple expression for $\Sigma_1^{\nu} n^p$, unless p happens to be an integer. On the other hand the AP method is convenient only when $p \geqslant 0$, so that (3.5) can be used. When $p<0$, divergence at $\rho=0$ forces one to use (3.4), which yields an unmanageable integral. Hence we reserve the intermediate range $-1<p<0$ for separate treatment.

5.2.4. When $p \geqslant 0$, we use formula (3.5) which yields D rather than D_{R}:

$$
\begin{aligned}
D &= \mathrm{i} \int_0^\infty \mathrm{d}\rho \frac{1}{\mathrm{e}^{2\pi\rho}-1}\{(\mathrm{i}\rho)^p - (-\mathrm{i}\rho)^p\} \\
&= -2\mathrm{i} \sin(\pi p/2) \int_0^\infty \mathrm{d}\rho \frac{\rho^p}{\mathrm{e}^{2\pi\rho}-1}. \qquad (5.14)
\end{aligned}
$$

Notice that D vanishes when p is an even integer. Now there is no avoiding the integral: one obtains

$$D = -2 \sin(\pi p/2)\Gamma(p+1)\zeta(p+1)/(2\pi)^{p+1},$$

which by the aid of (5.11) reduces to

$$D = \zeta(-p). \qquad (5.15)$$

Subtracting from this

$$D_{\mathrm{L}} = \tfrac{1}{2} - \int_0^1 \mathrm{d}n\, n^p = \tfrac{1}{2} - \frac{1}{p+1} \qquad (5.16)$$

we recover for $D_{\mathrm{R}} = D - D_{\mathrm{L}}$ the expected expression (5.13).

5.2.5 When $-1<p<0$, we are forced to use a special trick, exploiting within the framework of the ε-averaging method the result given by Hardy (1949, equation 13.10.7):

$$\lim_{\nu\to\infty} \left\{ \sum_1^{\nu}{}' n^p + \tfrac{1}{2} - \nu^{p+1}/(p+1) \right\} = \zeta(-p), \qquad p<0. \qquad (5.17)$$

Noting that

$$\int_1^\lambda \mathrm{d}n n^p = \frac{1}{p+1}(\nu^{p+1}-1) + \int_\nu^\lambda \mathrm{d}n n^p,$$

and with an eye on (5.17), we write (cf equation (4.3))

$$
\begin{aligned}
D_{\mathrm{R}}(N=1|\lambda) &\equiv \left(\sum_1^{\nu}{}' - \int_1^\lambda \mathrm{d}n \right) n^p \\
&= \left\{ \sum_1^{\nu}{}' n^p + \tfrac{1}{2} - \frac{\nu^{p+1}}{p+1} \right\} - \tfrac{1}{2} + \frac{1}{p+1} - \int_\nu^\lambda \mathrm{d}n n^p. \qquad (5.18)
\end{aligned}
$$

As $\lambda \to \infty$ and $\nu = [\lambda] \to \infty$, the final term in (5.18) vanishes (since $p<0$); and since the contents of the curly brackets approach the finite limit $\zeta(-p)$, they are replaced by this value after ε-averaging (recall here the paragraph following equation (4.5)). Accordingly, we have again recovered the expected result (5.13).

5.3. *Exponentials:* $f(n) = \exp(\alpha n)$

We revert to calculating D itself (rather than D_{R}), i.e. symbolically

$$D = \left(\sum_0^\infty{}' - \int_0^\infty \mathrm{d}n \right) \mathrm{e}^{\alpha n} = -\tfrac{1}{2} + \sum_0^\infty \mathrm{e}^{\alpha n} - \int_0^\infty \mathrm{d}n\, \mathrm{e}^{\alpha n}, \qquad \alpha = a + \mathrm{i}b. \tag{5.19}$$

In the convergent case, $a < 0$, equation (5.19) yields directly

$$D = -\tfrac{1}{2} - 1/(\mathrm{e}^{\alpha} - 1) + 1/\alpha. \tag{5.20}$$

We have already seen, at the end of § 4, that the ε-averaging method gives this same result for *all* values of α. By contrast, the AP prescription (3.5) yields

$$D = \mathrm{i} \int_0^\infty \mathrm{d}\rho \frac{1}{\mathrm{e}^{2\pi\rho} - 1} \{ \mathrm{e}^{-b\rho + \mathrm{i}a\rho} - \mathrm{e}^{b\rho - \mathrm{i}a\rho} \} \tag{5.21}$$

which converges only when $|b| < 2\pi$; and one can easily check that formulae of the type (3.3), with whatever choice of ϕ, also diverge beyond this range of b. However, when convergent, (5.21) does reduce to (5.20); this agreement is important since it underlies the argument (at the start of § 3) about the wider compatibility of the AP and the ε-averaging methods.

The poles of D at $\alpha = \mathrm{i}(2\pi r)$, $(r = \pm 1, \pm 2, \ldots)$ reflect a peculiar tuning of the original summand $\mathrm{F}(\Omega)$ to the length L of the system. From the discussion at the start of § 2 we see that $f(n) = \exp(\alpha n)$ implies $F(\Omega) = \exp(A\Omega)$, with $cA = (\alpha L/2\pi)$; and $\alpha = \mathrm{i}(2\pi r)$ implies $cA = \mathrm{i}rL$.

5.4. *Exponentials:* $f(n) = \exp(an^p)$

We consider only real positive a, and $p > 1$; the prescriptions of § 3 are then appropriate. With $1 < p \leqslant 3$, the condition $\pi/2p < \phi < 3\pi/2p$ still admits the convenient choice $\phi = \pi/2$, i.e. use of the simple AP prescription (3.5); but with $p > 3$ one is forced to chose $\phi < \pi/2$, and to fall back on the generalised version (3.3). Then the obvious choice is $\phi = \pi/p$, and (3.3) yields, for $D = D_{\mathrm{R}}(N = 0)$:

$$D = (\tfrac{1}{2} - 1/p) + \int_0^\infty \mathrm{d}\rho\, \mathrm{e}^{-a\rho^p} \left(\frac{\mathrm{e}^{\mathrm{i}\pi/p}}{\exp(-2\pi\mathrm{i}\rho\mathrm{e}^{\mathrm{i}\pi/p}) - 1} + \frac{\mathrm{e}^{-\mathrm{i}\pi/p}}{\exp(2\pi\mathrm{i}\rho\, \mathrm{e}^{-\mathrm{i}\pi/p}) - 1} \right) \tag{5.22}$$

No further simplification seems possible, except for an asymptotic expansion when a is large.

Appendix 1. The Euler–Maclaurin formula

A1.1 *Introduction*

The Euler–Maclaurin formula (EM in the following) has been relegated to the appendix because, for actually evaluating D, it is generally less convenient than the other methods†. Nevertheless it merits attention, partly because by tradition it is the first tool that comes to hand for dealing with the difference between sum and integral, partly because in several cases it does readily provide a convergent expression for D at least in

† Its relative popularity is due to the fact that it can usually yield the asymptotic expansion in parameters like η when $f(n) = \ln(n + \eta)$ or $(n + \eta)^{-1}$.

principle, and partly because it has links both with the ε-averaging and the AP prescriptions; its compatibility with the latter is discussed by Hardy (1949), and its compatibility with the former will be proved in § A1.4. This proof will naturally not involve formal manipulations like those in which we indulged at the end of § 4.

We shall need the Bernouilli numbers defined by

$$t/(e^t-1)\equiv 1-\tfrac{1}{2}t+\sum_{s=1}^{\infty}(-1)^{s+1}\tilde{B}_s t^{2s}/(2s)!$$

$$\equiv\sum_{s=0}^{\infty}B_s t^s/s!, \tag{A1.1}$$

the sums converging for $|t|<2\pi$. Comparison shows that

$$B_0=1, B_1=-\tfrac{1}{2},\quad B_{2s+1}=0, B_{2s}=(-1)^{s+1}\tilde{B}_s \quad \text{for } s\geqslant 1. \tag{A1.2}$$

Other numerical values are $B_2=\frac{1}{6}$, $B_4=-\frac{1}{30}$, $B_6=\frac{1}{42}$. As $s\to\infty$ one has

$$(2\pi)^{2s}B_{2s}/2(2s)!\sim(-1)^{s+1}(1+O(2^{-2s})). \tag{A1.3}$$

The B_s satisfy the remarkable identity

$$(B+1)^q=B_q+\delta_{q1}, \tag{A1.4}$$

where on the left one identifies $B^q=B_q$. Finally one has

$$B_{s+1}/(s+1)=-\zeta(-s), \qquad s=1,2,3,\ldots. \tag{A1.5}$$

A.2. The Euler–Maclaurin prescription

Our basic tool is the Euler–Maclaurin identity valid for any sufficiently differentiable function $h(n)$:

$$\sum_{n=n_1}^{n_2}{}'' h(n)-\int_{n_1}^{n_2}\mathrm{d}n\,h(n)$$

$$=\sum_{s=2}^{2k}\frac{B_s}{s!}[h^{(s-1)}(n_2)-h^{(s-1)}(n_1)]+\int_{n_1}^{n_2}\mathrm{d}n\,h^{(2k)}(n)\chi_{2k}(n), \tag{A1.6}$$

$$\chi_{2k}(n)=2(-1)^k\sum_{m=1}^{\infty}\cos(2\pi mn)/(2\pi m)^{2k}. \tag{A1.7}$$

Identifying $h(n)=f(n|\lambda)$ we now impose the following conditions on it, i.e. on its factors $f(n)$ and $g(n|\lambda)$. It must be possible to choose $2k$ so that: $f(n|\lambda)$ is differentiable $2k$ times; all the derivatives in (A1.6) vanish as $n_2\to\infty$; the derivative $f^{(2k)}(n|\lambda)$ vanishes fast enough for the integral on the right to converge as $n_2\to\infty$; and this convergence is uniform in λ, so that lim $\lambda\to\infty$ exists and may be taken under the integral. (Appendix 2 shows what can happen if this condition is ignored.) The specific consequences for $f(n)$ itself will be spelled out below. Under these conditions we can allow $n_2\to\infty$: the terms $h^{(s-1)}(n_2)$ all vanish, and the difference between Σ'' and Σ' disappears; writing $n_1=N$, we obtain, in the notation of equation (2.10–13),

$$D_R(N|\lambda)=-\sum_{s=2}^{2k}\frac{B_s}{s!}f^{(s-1)}(N|\lambda)+R_{2k}(N|\lambda), \tag{A1.8a}$$

$$R_{2k}(N|\lambda) = \int_N^\infty \mathrm{d}n f^{(2k)}(n|\lambda)\chi_{2k}(n). \tag{A1.8b}$$

Unfortunately, as a rule the remainder† R_{2k} in (A.8) fails to vanish as $k \to \infty$, and the series can be used at best as an asymptotic series in some suitable parameter, for instance in η when the summand is $f = \ln(n + \eta)$.

The no-cutoff limit (2.7) implies $f^{(r)}(N|\lambda) \to f^{(r)}(N)$ as well as $f(N|\lambda) \to f(N)$. We now take this limit in (A1.8), having first chosen $2k$ large enough to allow (A1.8*b*) to continue to converge in the limit (see below); that such a choice of $2k$ be possible is the chief restriction on the summands $f(n)$ amenable to the EM method. This yields

$$D_\mathrm{R}(N) = -\sum_{s=2}^{2k} \frac{B_s}{s!} f^{(s-1)}(N) + R_{2k}(N), \tag{A1.9}$$

$$R_{2k}(N) = \int_N^\infty \mathrm{d}n\, f^{(2k)}(n)\chi_{2k}(n). \tag{A1.10}$$

This is our end result, namely, after addition of D_L, the EM prescription for the functional $\mathscr{D}\{f(n)\}$ of equation (2.8).

As regards the convergence of the integral in (A1.10) note the following. By virtue of (A1.8) the factor $\chi_{2k}(n)$ is a periodic function of period unity, and its integral over a period vanishes‡. Hence, by breaking up the integration at conveniently chosen points, $R_{2k}(N)$ can be expressed as an oscillating series which converges provided the factor $f^{(2k)}(n)$ decreases monotonically for large enough n; this follows from the standard argument from Fresnel diffraction theory (e.g. Born and Wolf 1975, who credit the idea to Schuster 1891; for recent references to the summation of alternating series, see e.g. Johnsonbaugh 1979). Accordingly, the $f(n)$ covered by the EM prescription (A1.9–10) include all algebraic and logarithmic functions bounded by some power n^p as $n \to \infty$, since for all these $f^{(2k)}$ is bounded by n^{p-2k}, and we need merely choose k so that $p - 2k < 0$. That f belong to this class is sufficient but not necessary, as is shown by the example of exponentials $\exp(\alpha n)$ with $|\alpha| < 2\pi$ in § A1.3.

A1.3. Applications

For the actual evaluation of D, the EM prescription is convenient only if the series in (A1.9) either terminates, or if it converges and can be summed in closed form, which happens only in the following two cases.

When $f(n) = n^p$, with p a positive integer, the series terminates at $s = p + 1$. Choosing $2k \geq p + 1$, R_{2k} vanishes and (A1.9) becomes

$$D_\mathrm{R}(N = 1) = -\frac{1}{P+1} \sum_{s=2}^{p+1} \binom{p+1}{s} B_s.$$

† The word 'remainder' reflects the role of R with respect to the index $2k$, and not with respect to the argument N; no question arises of letting $N \to \infty$, since the remark following equation (2.12) shows that this would merely remove D_R from the argument, and transfer any difficulties, unchanged, to D_L.

‡ In fact $\chi_{2k}(n) = -B_{2k}(n)/(2k)!$, where the Bernouilli polynomials $B_s(n)$ are defined, for $0 \leq n \leq 1$, by

$$t\,\mathrm{e}^{nt}/(\mathrm{e}^t - 1) = \sum_{s=0}^{\infty} B_s(n) t^s/s!,$$

and for other n by periodic continuation.

Adding and subtracting the terms with $s=0,1$ (cf (A1.2)), and using in turn (A1.4) and (A1.5) we obtain

$$\begin{aligned} D_{\mathrm{R}}(N=1) &= -\frac{1}{p+1}\left\{\sum_{s=0}^{p+1}\binom{p+1}{s}B_s - 1 + \frac{p+1}{2}\right\} \\ &= -\frac{1}{p+1}\left\{B_{p+1} - 1 + \frac{p+1}{2}\right\} \\ &= -\tfrac{1}{2} + \zeta(-p) + \frac{1}{p+1}, \end{aligned} \tag{A1.11}$$

which is our known result (5.13).

When $f(n)=\exp(\alpha n)$, the series in (A1.9) becomes

$$-\sum_{s=2}^{2k} B_s\alpha^{s-1}/s! = -(1/\alpha)\left\{\sum_{s=0}^{2k} B_s\alpha^s/s! - 1 + \alpha/2\right\}. \tag{A1.12}$$

Comparison with (A1.1) shows that, provided $|\alpha|<2\pi$, we can allow $k\to\infty$ (when the remainder term must vanish), obtaining

$$D_{\mathrm{R}}(N=1) = -\tfrac{1}{2} + 1/\alpha - 1/(\mathrm{e}^{\alpha}-1), \tag{A1.13}$$

which is the known result (4.10)

A.4. Compatibility with the ε-averaging method

Compatibility will be shown by proving that the ε-averaging prescription (4.3–5) can be recast in the form (A1.9–10) with $N=0$, provided both prescriptions make sense, and subject for technical reasons to certain restrictions spelled out following equation (A1.21) below. We start by rewriting $D(\lambda)$, equation (4.3), by aid of the EM identity (A1.6); of course in the latter we now retain the correction terms for the finite upper limit, including the difference $f(\nu)/2$ between Σ' and Σ'', and a term $-\int_\nu^\lambda \mathrm{d}n f(n)$ to allow for the difference between the upper limits of the integrals. Grouping the terms by hindsight:

$$\begin{aligned} D(\lambda) = &\left\{-\sum_{s=2}^{2k}\frac{B_s}{s!}f^{(s-1)}(0) + \int_0^{\nu}\mathrm{d}n\, f^{(2k)}(n)\chi_{2k}(n)\right\} \\ &+\left\{\sum_{s=2}^{2k}\frac{B_s}{s!}f^{(s-1)}(\nu) + \tfrac{1}{2}f(\nu) - \int_{\nu}^{\lambda}\mathrm{d}n\, f(n)\right\}. \end{aligned} \tag{A1.14}$$

As before, the index $2k$ is chosen so that the first integral converges as $\nu\to\infty$; i.e., $f^{(2k)}(n)$ must eventually decrease monotonically as $n\to\infty$.

As instructed in (4.5), we replace ν by $\lambda-\varepsilon$ and carry out the integration over ε at fixed λ; meanwhile, we drop at once any terms that will vanish as $\nu\to\infty$ and $\lambda\to\infty$. On the right of (A1.14), the first sum is unaffected since it does not involve ν. As regards the remainder term (involving χ_{2k}), since by assumption the integral converges as $\nu\to\infty$, the ε-averaging makes no difference to it ultimately, and we immediately replace ν by ∞ in this term. Hence, under ε-averaging the contents of the first pair of curly brackets in (A1.14) become identical to the desired end result, namely to (A1.9, 10) with $N=0$.

It remains only to show that the contents of the second pair of curly brackets in (A1.14) vanish under ε-averaging, namely that

$$\lim_{\lambda\to\infty} W(\lambda) \equiv \lim_{\lambda\to\infty} \int_0^1 \mathrm{d}\varepsilon \left\{ \sum_{s=2}^{2k} \frac{B_s}{s!} f^{(s-1)}(\lambda-\varepsilon) + \tfrac{1}{2} f(\lambda-\varepsilon) - \int_{\lambda-\varepsilon}^{\lambda} \mathrm{d}n f(n) \right\} = 0. \tag{A1.15}$$

The crucial step is to expand $f^{(s-1)}(\lambda-\varepsilon)$, $f(\lambda-\varepsilon)$ and $f(n)$ in Taylor series around λ. We need only assume, as already in § 4, that these series converge in a circle of just over unit radius around λ. The integrations in (A1.15) are then carried out explicitly, and we collect together all the terms involving derivatives of a given order. After these straightforward manipulations, $W(\lambda)$ becomes

$$W(\lambda) = \sum_{q=0}^{\infty} \frac{(-1)^q f^{(q)}(\lambda)}{(q+2)!} V(q), \tag{A1.16}$$

$$V(q) = \tfrac{1}{2} q - \sum_{s=2}^{\min(2k,q+1)} \binom{q+2}{s} B_s = - \sum_{s=0}^{\min(2k,q+1)} \binom{q+2}{s} B_s, \tag{A1.17}$$

where in the last step we have added and subtracted the terms with $s = 0, 1$, and have used (A1.2). If $q+1 \leqslant 2k$, then, adding and subtracting also the term with $s = q+2$, and using the identity (A1.4), we find

$$V(q) = -\sum_{s=0}^{q+2} \binom{q+2}{s} B_s + B_{q+2} = 0, \qquad (q+1 \leqslant 2k). \tag{A1.18}$$

Hence the effective lower limit of the sum in (A1.16) is $q = 2k$, and we must finally prove the vanishing of

$$\lim_{\lambda\to\infty} \sum_{q=2k}^{\infty} \frac{(-1)^q f^{(q)}(\lambda)}{(q+2)!} V(q). \tag{A1.19}$$

Now equation (A1.17) entails (writing C for any constant independent of q)

$$|V(q)| < C \sum_{s=0}^{2k} \binom{q+2}{s} < C(q+2)^{2k}; \tag{A1.20}$$

using this in (A1.19) we obtain

$$|W(\lambda)| < C \sum_{q=2k}^{\infty} |f^{(q)}(\lambda)| (q+2)^{2k}/(q+2)!. \tag{A1.21}$$

To proceed we must assume† that $f^{(2k)}(z)$ decreases monotonically when $\mathrm{Re}\, z \to \infty$ not only for real z, but for all z in a strip $|\mathrm{Im}\, z| < \sigma < \sigma_0$, $\sigma_0 > 1$; (recall that it is already assumed that f is analytic in such a strip, since its Taylor series must converge within a circle of radius $\sigma > 1$ around any sufficiently large λ). Then, writing $q = 2k + r$, and with Γ a circle of radius σ $(1 < \sigma < \sigma_0)$ around $z = \lambda$, we have

$$\begin{aligned} f^{(2k+r)}(\lambda)/r! &= (1/2\pi \mathrm{i}) \int_\Gamma \mathrm{d}z f^{(2k)}(z)/(z-\lambda)^{r+1}, \\ |f^{(2k+r)}(\lambda)| &< (\max_{\sigma,\lambda} |f^{(2k)}|) r!/\sigma^r, \end{aligned} \tag{A1.22}$$

† The appeal to the following argument arose from a discussion with Dr J S Plaskett.

where $(\max_{\sigma,\lambda}|f^{(2k)}|)$ is the maximum value of $|f^{(2k)}|$ on Γ. Finally, by (A1.21,22)

$$|W(\lambda)| < C(\max_{\sigma,\lambda}|f^{(2k)}|)\sum_{r=0}^{\infty}\frac{r!(2k+2+r)^{2k}}{(2k+2+r)!}\frac{1}{\sigma^{r}}. \tag{A1.23}$$

The series converges (since by assumption $\sigma > 1$) and hence $W(\lambda)$ does indeed vanish as $\lambda \to \infty$, by virtue of our assumption about $f^{(2k)}$.

Appendix 2. An inadmissible cutoff

A simple but instructive example shows what can happen if a cutoff violates the analyticity assumption of the Abel–Plana prescription or the uniform-convergence assumption of the Euler–Maclaurin prescription. The expression

$$g(n|\lambda) = \lambda^2/[(\lambda - n)^2 + \gamma^2] \tag{A2.1}$$

contravenes both. (By contrast, the cutoff $\lambda^2/[(\lambda + n)^2 + \gamma^2]$ leads to no difficulties.) Of course, if one is in a position to *choose* cutoffs, the pathology we shall uncover merely alerts one to avoid such a g. But if one is concerned with a problem where the cutoff is actually known, and is given by (A2.1), then the conclusion is more important: for then it tells us that the true result for D is not obtainable by the AP or EM prescriptions, even though g satisfies the permanence conditions (2.7), and even if the final formulae $\mathscr{D}\{f(n)\}$ of both prescriptions converge.

The danger is perhaps clearest in the AP method, for g has manifest poles in the positive half-plane at $\lambda = n \pm \mathrm{i}\gamma$. If nevertheless we deform the contour as in § 3, then in addition to the right-hand side of say (3.5) we obtain an extra contribution Δ from these two poles:

$$\Delta(\lambda) = \frac{\pi\lambda^2}{\gamma}\left\{\frac{f(\lambda+\mathrm{i}\gamma)}{\exp(2\pi\gamma - 2\pi\mathrm{i}\lambda) - 1} + \frac{f(\lambda - \mathrm{i}\gamma)}{\exp(2\pi\gamma + 2\pi\mathrm{i}\lambda) - 1}\right\}. \tag{A2.2}$$

From here on we confine ourselves for simplicity to the trivial summand $f(n) = 1$, where the result with any admissible cutoff is $D = 0$, by virtue of (5.14) with $p = 0$. By contrast, with the cutoff (A2.1) the true result is

$$\begin{aligned} D(\lambda) = &-\int_0^\infty \frac{\mathrm{d}\rho}{(\mathrm{e}^{2\pi\rho} - 1)}\cdot\frac{4\lambda^3\rho}{[(\lambda^2 - \rho^2 + \gamma^2)^2 + 4\lambda^2\rho^2]} \\ &+\frac{\pi\lambda^2}{\gamma}\left\{\frac{1}{\exp(2\pi\gamma - 2\pi\mathrm{i}\lambda) - 1} + \frac{1}{\exp(2\pi\gamma + 2\pi\mathrm{i}\lambda) - 1}\right\}. \end{aligned} \tag{A2.3}$$

The first term is what one would have obtained by applying the AP prescription without recognising the illegitimate nature of the cutoff; and as $\lambda \to \infty$ this term does of course vanish. By contrast, the second term has no limit, continuing to oscillate indefinitely with an amplitude proportional to λ^2.

Since (A2.3) is exact for $D(\lambda)$, it must equal the Euler–Maclaurin expression (A1.8) for the same quantity, demonstrating *a posteriori* that the final EM prescription (namely (A1.9,10) with $N = 0$), which is identically zero, cannot be the limit of (A1.8), which we can write as

$$D(\lambda) = -\frac{B_2}{2!}f^{(1)}(0|\lambda) + \int_0^\infty \mathrm{d}n f^{(2)}(n|\lambda)\chi_2(n). \tag{A2.4}$$

Of course $f^{(1)}(0|\lambda) = 2\lambda^3/(\lambda^2+\gamma^2)^2$ does vanish as $\lambda \to \infty$; but the integral R_2 in (A2.4) has no limit, whence it cannot approach

$$\int_0^\infty \mathrm{d}n \lim_{\lambda\to\infty} f^{(2)}(n|\lambda)\chi_2(n) = 0.$$

Indeed, one can see directly that R/λ^2 will continue to oscillate indefinitely with increasing λ, if one recalls that $\chi_2(n)$ is a periodic function, while $f^{(2)}(n|\lambda)/\lambda^2$ is essentially a fixed shape of width γ localised near $n = \lambda$.

References

Babiker M and Barton G 1972a,b *Proc. R. Soc.* A **326** 255, 277
Balian R and Bloch C 1970 *Ann. Phys., NY* **60** 401
—— 1971 *Ann. Phys., NY* **64** 271
—— 1972 *Ann. Phys., NY* **69** 76
Balian R and Duplantier B 1977 *Ann. Phys., NY* **104** 300
—— 1978 *Ann. Phys., NY* **112** 165
Baltes H P and Hilf E R 1976 *Spectra of finite systems, Bibliographisches Institut, Mannheim*
Barton G 1970 *Proc. R. Soc.* A **320** 251
—— 1979a *Proc. R. Soc.* A **367** 117
—— 1979b *Rep. Prog. Phys.* **42** 963
Boas R P and Stutz C 1971 *Am. J. Phys.* **39** 745
Born M and Wolf E 1975 *Principles of Optics*, 5th edn, § 8.2 (Oxford: Pergamon)
Boyer T H 1968 *Phys. Rev.* **174** 1764
Brown L S and Maclay G J 1969 *Phys. Rev.* **184** 1272
Casimir H S G 1948 *Kon. Ned. Akad. Wetenshap. Proc.* **51** 793
Dwight H B 1961 *Tables of integrals and other mathematical data*, 4th edn (New York: Macmillan)
Fierz M 1960 *Helv. Phys. Acta* **33** 855
Groebner W and Hofreiter N 1958 *Integraltafel II*, 2nd edn (Vienna: Springer)
Hardy G H 1949 *Divergent series*, ch 13 (London: OUP)
Johnsonbaugh R 1979 *Am. Math. Monthly* **86** 637
Lukosz W 1971 *Physica* **56** 101
Milton K A, DeRaad L L and Schwinger J 1978 *Ann. Phys., NY* **115** 388
Power E A 1964 *Introductory quantum electrodynamics*, § 3.4 (London: Longmans)
Schuster A 1891 *Phil. Mag.* **31** 77
Unwin S D and Critchley R 1980 *Proc. R. Soc.* A **372** 297