Zeta-function regularization is uniquely defined and well

1994 J. Phys. A: Math. Gen. 27 L299

(http://iopscience.iop.org/0305-4470/27/9/010)

J. Phys. A: Math. Gen. 27 (1994) L299–L304. Printed in the UK

LETTER TO THE EDITOR

Zeta-function regularization is uniquely defined and well

E Elizalde†

Division of Applied Mechanics, Norwegian Institute of Technology and Department of Physics, University of Trondheim, N-7034 Trondheim, Norway

Received 18 January 1994

Abstract. Hawking's zeta-function regularization procedure is shown to be rigorously and uniquely defined, thus putting an end to the spreading lore about different difficulties associated with it. Basic misconceptions, misunderstandings and errors which keep appearing in important scientific journals when dealing with this beautiful regularization method—and other analytical procedures—are clarified and corrected.

This paper is a defence of Hawking's zeta-function regularization method [1], against the different criticisms which have been published in important scientific journals and which seem to conclude (sometimes with exaggerated emphasis) that the procedure is ambiguous, ill-defined, and that it possesses even more problems than the well known ones which afflict, for example, dimensional regularization. Our main purpose is to clarify, once and for all, some basic concepts, misunderstandings, and also errors which keep appearing in the physical literature, about this method of zeta-function regularization and other analytical regularization procedures. The situation is such that, what is in fact a most elegant, well defined, and unique—in many respects—regularization method, may look now to the non-specialist as just one more among many possible regularization procedures, plagued with difficulties and ill-definiteness.

We shall not review here, in detail, the essentials of the method, nor give an exhaustive list of the papers that have pointed out 'difficulties with the zeta-function procedure'. For this purpose we refer the specialized reader to a book to appear soon [2] (see also [3]). Instead, we shall restrict ourselves to some specific points which are at the very heart of the matter and which may be interesting to a much broader audience.

The method of zeta-function regularization is uniquely defined in the following way. Take the Hamiltonian H, corresponding to our quantum system, plus boundary conditions, plus possible background field and including a possibly non-trivial metric (because we may live in a curved spacetime). In mathematical terms, all this boils down to a (second-order, elliptic) differential operator A plus corresponding boundary conditions. The spectrum of this operator A may or may not be calculable explicitly, and in the first case may or may not exhibit a beautiful regularity in terms of powers of natural numbers. Whatever the situation, it is a well established mathematical theorem that to *any* such operator a *zeta function* ζ_A can be defined in a rigorous way. The formal expression of this definition is

$$\zeta_A(s) = \operatorname{Tr} e^{-s \ln A} . \qquad (1)$$

† On leave of absence from and permanent address: Department ECM, Faculty of Physics, Barcelona University, Diagonal 647, 08028 Barcelona, Spain. E-mail address: eli@ebubecm1.bitnet

Let us stress that this is completely *general*. Moreover, the present function is by no means the only kind of zeta function known to mathematicians, for whom this concept has a wider character (see, for instance [4]). The zeta function $\zeta_A(s)$ is generally a meromorphic function (develops only poles) in the complex plane, $s \in \mathbb{C}$. Its calculation usually requires complex integration around some circuit in the complex plane, the use of the Mellin transform, and the calculation of invariants of the spacetime metric (the Hadamard–Minakshisundaram–Pleyel–Seeley–DeWitt–Gilkey–... coefficients) [5].

In the particular case when the eigenvalues of the differential operator A—or equivalently, the eigenvalues of the Hamiltonian (with the boundary conditions, etc taken into account)—can be calculated explicitly (let us call them λ_n and assume they form a discrete set), the expression of the zeta function is given by

$$\zeta_A(s) = \sum_n \lambda_n^{-s} . \tag{2}$$

Now, as a particular case of this (already particular) case, when the eigenvalues are of one of the forms: (i) an, (ii) $a(n+b)$ or (iii) $a(n^2+b^2)$, we obtain, respectively, the so-called (i) (ordinary) Riemann zeta function ζ_{R} (or simply ζ), (ii) Hurwitz (or generalized Riemann) zeta function ζ_{H}, and (iii) (a specific case of the) Epstein–Hurwitz zeta function ζ_{EH}. Finally, depending on the physical magnitude to be calculated, the zeta function must be evaluated at a certain particular value of s. For instance, if we are interested in the vacuum or Casimir energy, which is simply given as the sum over the spectrum

$$E_{\mathrm{C}} = \frac{\hbar}{2} \sum_n \lambda_n \tag{3}$$

this will be given by the corresponding zeta function evaluated at $s = -1$:

$$E_{\mathrm{C}} = \frac{\hbar}{2} \zeta_A(-1) . \tag{4}$$

Normally, the series (3) will be divergent, and this will involve an analytic continuation through the zeta function. That is why such regularization can be termed as a particular case of analytic continuation procedure. In summary, the zeta function of the quantum system is a very general, uniquely defined, rigorous mathematical concept, which does not admit either interpretations nor ambiguities.

Let us now come down to the concrete situations which have motivated this article. In [6], when calculating the Casimir energy of a piecewise uniform closed string, Brevik and Nielsen where confronted with the following expression [6, equation (52)]

$$\sum_{n=0}^{\infty} (n+\beta) \tag{5}$$

which is clearly infinite. Here, the zeta-function regularization procedure consists of the following. This expression comes about as the sum over the eigenvalues $n+\beta$ of the Hamiltonian of a certain quantum system (here the transverse oscillations of the mentioned string), i.e. $\lambda_n = n+\beta$. There is little doubt about what to do: as clearly stated above, the corresponding zeta function is

$$\zeta_A(s) = \sum_{n=0}^{\infty} (n+\beta)^{-s} . \tag{6}$$

Now, for Re $s > 1$ this is the expression of the Hurwitz zeta function $\zeta_{\mathrm{H}}(s;\beta)$, which can be analytically continued as a meromorphic function to the whole complex plane. Thus, the zeta function regularization method unambiguously prescribes that the sum under consideration should be assigned the following value:

$$\sum_{n=0}^{\infty}(n+\beta)=\zeta_{\rm H}(-1;\beta)\,. \tag{7}$$

The wrong alternative (for obvious reasons, after all that has been said before), would be to argue that 'we might as well have written

$$\sum_{n=0}^{\infty}(n+\beta)=\zeta_{\rm R}(-1)+\beta\zeta_{\rm R}(0) \tag{8}$$

which gives a different result'. In fact, that 'the Hurwitz zeta function (and not the ordinary Riemann)' was the one 'to be used' was recognized by Li *et al* [7], who reproduced in this way the correct result obtained by Brevik and Nielsen by means of a (more conventional) exponential cut-off regularization. However, the authors of [7] were again misunderstanding the main issue when they considered their method as being a *generalization* of the zeta-regularization procedure (maybe just because the generalized Riemann zeta function appears!). Quite the contrary, this is just a *specific and particularly simple* application of the zeta function regularization procedure.

Of course, the method can be viewed as just one of the many possibilities of analytic continuation in order to give sense to (i.e. to regularize) infinite expressions. From this point of view, it is very much related with the standard dimensional regularization method. In a very recent paper Svaiter and Svaiter [8] have argued that, being such close relatives, these two procedures even share the same type of diseases. But precisely to cure the problem of the dependence of the regularized result on the kind of the extra dimensions (artificially introduced in dimensional regularization) was, let us recall, one of the main motivations of Hawking for the introduction of a new procedure, i.e. zeta-function regularization, in physics [1]. So we seem to have been caught in a devil's staircase.

The solution to this paradox is the following. Actually, there is no error in the examples of [8] and the authors know perfectly what they are doing, but their interpretation of the results may be the origin of a great deal of confusion among non-specialists. To begin with, it might look at first sight as if the concept itself of analytical continuation would *not* be *uniquely* defined. Given a function in some domain of the complex plane (here, normally, a part of the real line or the half plane Re $s > a$, a being some abscisa of convergence), its analytic continuation to the rest of the complex plane (in our case, usually as a meromorphic function, but this need not be so in general) *is* uniquely defined. Putting it plainly, a function *cannot* have two different analytic continuations. What Svaiter and Svaiter do in their examples is simply to start in each case from two different functions of s and then continue each of them analytically. Of course, the result is different. In particular, these functions are

$$f_1(s)=\sum_{n=0}^{\infty}n^{-s} \tag{9}$$

versus

$$f_2(s)=\sum_{n=0}^{\infty}n\left(\frac{n}{a}+1\right)^{-(s+1)} \tag{10}$$

continued to $s=-1$, in the first example, which corresponds to a Hermitian massless conformal scalar field in 2D Minkowski spacetime with a compactified dimension, and

$$g_1(s)=\sum_{n=0}^{\infty}n^{-3s} \tag{11}$$

continued to $s=-1$, versus

$$g_2(s)=\sum_{n=0}^{\infty} n^3\left(\frac{n}{a}+1\right)^{-s} \tag{12}$$

continued to $s=0$, in the second example, in which the vacuum energy corresponding to a conformally coupled scalar field in an Einstein universe is studied. Needless to say, the number of possibilites to define 'different analytic continuations' in this way is literally infinite. What use one can make of them remains to be seen.

However, *what is absolutely misleading is to conclude from those examples that* analytical regularizations 'suffer from the same problem as dimensional regularization', precisely the one that Hawking wanted to cure! This has no meaning. In the end, dimensional regularization is also an analytical procedure! One must realize that zeta-function regularization is perfectly well defined, and has little to do with these arbitrary analytic continuations '*à la* ζ' in which one changes at will any exponent at any place with the only restriction that one recovers the starting expression for a particular value of the exponent s.

The facts are as follows.

(i) There exist infinitely many different analytic regularization procedures, dimensional regularization and zeta-function regularization being just two of them.
(ii) Zeta-function regularization is, as we have seen, a specifically defined procedure, provides a unique analytical continuation and (sometimes) a finite result.
(iii) Therefore, zeta-function regularization does not suffer, in any way, from the same kind of problem (or a related one) as dimensional regularization.
(iv) This does not mean, however, that zeta-function regularization has *no* problems, but they are of a different kind; the first already appears when it turns out that the point (let us say $s=-1$ or $s=0$) at which the zeta function must be evaluated turns out to be precisely a *pole* of the analytic continuation. This and similar difficulties can be solved, as discussed in detail in [9]. Eventually, as a final step, one has to resort to renormalization group techniques [10].
(v) Zeta-function regularization has been extended to higher loop order by McKeon and Sherry under the name of operator regularization and there also some difficulties (concerning the breaking of gauge invariance) appear [11].
(vi) But, in the end, the fundamental question is: which of the regularizations that are being used is *the one* chosen by nature? In practice, one always tries to avoid answering this question, by checking the finite results obtained with different regularizations and by comparing them with classical limits which provide well known, physically meaningful values. However, one would be led to believe that in view of its uniqueness, naturalness and mathematical elegance, zeta-function regularization could well be *the one*. Those properties are certainly to be counted among their main virtues, but (oddly enough) in some sense also as its drawbacks: we do not manage to see clearly how and what infinites are thrown away, something that is evident in other more pedestrian regularizations (which are actually equivalent in some cases to the zeta one, as pointed out, for example, in [8]).

The final issue of this paper will concern the practical application of the procedure. Actually, aside from some very simple cases (among those, the ones reviewed here), the use of the procedure of analytic continuation through the zeta function requires a good deal of mathematical work [2]. It is no surprise that it has been associated with mistakes and errors so often [12]. One which often repeats itself can be traced back to (1.70) of the celebrated book by Mostepanenko and Trunov [13] on the Casimir effect:

$$\frac{a^2}{\pi^2}\sum_{n=1}^{\infty}\left(n^2+\frac{a^2m^2}{\pi^2}\right)^{-1}=\frac{1}{2m^2}\left(-1+\frac{am}{\pi}\coth\frac{am}{\pi}\right) \tag{13}$$

in other words (for $a=\pi$ and $m=c$)

$$\sum_{n=1}^{\infty}\left(n^2+c^2\right)^{-1}=\frac{1}{2c^2}\left(-1+c\coth c\right). \tag{14}$$

That (14) is not right can be observed by simple inspection. The corrected formula reads

$$\sum_{n=1}^{\infty}\left(\pi^2n^2+c^2\right)^{-1}=\frac{1}{2c^2}\left(-1+c\coth c\right). \tag{15}$$

The integrated version of this equality, namely,

$$\sum_{n=-\infty}^{\infty}\ln\left(n^2+c^2/\pi^2\right)=2c+2\ln\left(1-\mathrm{e}^{-2c}\right) \tag{16}$$

under the specific form

$$T\sum_{n=-\infty}^{\infty}\ln\left[(\omega_n)^2+(q_l)^2\right]=q_l+2T\ln\left(1-\mathrm{e}^{-q_l/T}\right) \tag{17}$$

with $\omega_n=2\pi nT$ and $q_l=\pi l/R$, has been used by Antillón and Germán in a very recent paper [14, equation (2.20)], when studying the Nambu–Goto string model at finite length and non-zero temperature. Now this equality is again formal. It involves an analytic continuation since there is no sense in integrating the left-hand side term by term: we get a divergent series.

A rigorous way to proceed is as follows. The expression on the left-hand side happens to be the simplest form of the inhomogeneous Epstein zeta function (usually called the Epstein–Hurwitz zeta function [15]). This function is quite involved and different expressions for it (including asymptotical expansions very useful for accurate numerical calculations) have been given in [15] (see also [16]). In particular

$$\begin{aligned}\zeta_{\mathrm{EH}}(s;c^2)&=\sum_{n=1}^{\infty}\left(n^2+c^2\right)^{-s}\\&=-\frac{c^{-2s}}{2}+\frac{\sqrt{\pi}\,\Gamma(s-1/2)}{2\Gamma(s)}c^{-2s+1}+\frac{2\pi^s c^{-s+1/2}}{\Gamma(s)}\sum_{n=1}^{\infty}n^{s-1/2}K_{s-1/2}(2\pi nc)\end{aligned} \tag{18}$$

which is reminiscent of the famous Chowla–Selberg formula (see [4]). Derivatives can be taken here and the analytical continuation in s again presents no problem.

The usefulness of zeta-function regularization is without question [17, 2, 5]. It can give immediate sense to expressions such as $1+1+1+\cdots=-\frac{1}{2}$, which turn out to be invaluable for the construction of new physical theories, as different as Pauli–Villars regularization with infinite constants (advocated by Slavnov [18]) and mass generation in cosmology through Landau poles (used by Yndurain [19]). The Riemann zeta function was termed by Hilbert in his famous 1900 lecture as the most important function of the whole of mathematics [20]. It will probably remain so in the Paris congress of AD 2000, but now maybe with quantum-field physics adhered to.

I am very grateful to Professor I Brevik and to Professor K Olaussen for many illuminating discussions and also to them and to Professor L Brink for the hospitality extended to me at the Universities of Trondheim and Göteborg, respectively. This work has been supported by DGICYT (Spain) and by CIRIT (Generalitat de Catalunya).

References

[1] Brink L and Nielsen H B 1973 *Phys. Lett.* **45B** 332
Hawking S W 1977 *Commun. Math Phys.* **55** 133
Dowker J S and Critchley R 1976 *Phys. Rev.* D **13** 3224
Brown L S and MacLay G J 1969 *Phys. Rev.* **184** 1272

[2] Elizalde E, Odintsov S D, Romeo A, Bytsenko A A and Zerbini S *Zeta Regularization Techniques with Applications* (Singapore: World Scientific) to be published

[3] Schwarz A S 1993 *Quantum Field Theory and Topology* (New York: Springer)

[4] Chowla S and Selberg A 1949 *Proc. Natl Acad. Sci. USA* **35** 317
Iyanaga S and Kawada Y (eds) 1977 *Encyclopedic Dictionary of Mathematics* vol II (Cambridge, MA: MIT Press) pp 1372ff

[5] Elizalde E, Leseduarte S and Zerbini S 1993 Mellin transform techniques for zeta-function resummations *University of Barcelona Preprint* UB-ECM-PF 93/7

[6] Brevik I and Nielsen H B 1990 *Phys. Rev.* D **41** 1185

[7] Li X, Shi X and Zhang J 1991 *Phys. Rev.* D **44** 560

[8] Svaiter B F and Svaiter N F 1993 *Phys. Rev.* D **47** 4581

[9] Blau S K, Visser M and Wipf A 1988 *Nucl. Phys.* B **310** 163

[10] Elizalde E and Kirsten K *J. Math. Phys.* to appear

[11] McKeon D G C and Sherry T N 1987 *Phys. Rev. Lett.* **59** 532; 1987 *Phys. Rev.* D **35** 3854
Rebhan A 1989 *Phys. Rev.* D **39** 3101

[12] Elizalde E and Romeo A 1989 *Phys. Rev.* D **40** 436

[13] Mostepanenko V M and Trunov N N 1990 *Casimir Effect and its Applications* (Moscow: Energoatomizdat) (in Russian)

[14] Antillón A and Germán G 1993 *Phys. Rev.* D **47** 4567

[15] Elizalde E 1990 *J. Math. Phys.* **31** 170

[16] Brevik I and Clausen I 1989 *Phys. Rev.* D **39** 603

[17] Actor A 1991 *J. Phys. A: Math. Gen.* **24** 3741
McAvity D M and Osborn H 1993 *Nucl. Phys.* B **394** 728

[18] Slavnov A, personal communication

[19] Yndurain F, personal communication

[20] Reid C 1970 *Hilbert* (Berlin: Springer) pp 82