String Perturbation Theory Diverges

David J. Gross and Vipul Periwal
Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544
(Received 11 April 1988)

We prove that perturbation theory for the bosonic string diverges for arbitrary values of the coupling constant and is not Borel summable. This divergence is independent of the existence of the infinities that occur in the theory due to the presence of tachyons and dilaton tadpoles. We discuss the physical implications of such a divergence.

PACS numbers: 11.17.+y

If perturbative string theory were to make sense, string theory would have nothing to do with physical reality, since there are many features of all perturbative treatments of string theory that are not shared by the real world. It is therefore important to determine the range of validity of string perturbation theory, as well as to develop nonperturbative tools that transcend it.

Much has been made of the fact that string perturbation theory is a topological expansion, and that (for closed-string theories) there is only a single "term" to calculate at each order in this expansion. For this reason it may seem unlikely that perturbation theory diverges. After all, the ubiquitous divergences of perturbative expansions in quantum field theory arise because of the $n!$ number of n-loop Feynman diagrams.[1] However, the single h-loop string Feynman graph is integrated over all of the moduli space of Riemann surfaces with h handles, a complicated space over which we really do not have much control, and therefore such naive statements have dubious validity. Furthermore, string theory contains within it, at low energies, ordinary field theory, which always yields divergent perturbative expansions. Thus we might expect similar divergent behavior in string theory.

In this Letter, we prove explicitly that bosonic string perturbation theory diverges at $\sum g^h h!$, where h is the number of handles. This reassures us that perturbation theory, for the bosonic string at least, has zero radius of convergence. Even more, such an expansion is not even (Borel) summable. We will argue below that this type of behavior could be an indication of the nonperturbative instability of the vacuum.

We shall consider the simplest amplitude, the partition function, and we shall show that it grows factorially in h, for large h. This is accomplished in two steps: (a) controlling the behavior of the string integrand uniformly over moduli space, so that we can give a *lower bound* on the integrand, showing that the integrand does not decrease like $1/h!$; and (b) estimating the volume moduli space for a given h, to show that it increases like $h!$, so that we can derive a *lower bound* on the integral which increases factorially in h.

We use the Selberg zeta-function description of the string integrand,[2] namely the partition function of the bosonic closed string is given by $cg^h Z(2)Z'(1)^{-13}d\mu_{\mathrm{WP}}$, where $Z(s)$ is the Selberg Z function, $d\mu_{\mathrm{WP}}$ is the Weil-Petersson measure on moduli space, c is independent of the genus, and all other constants that grow geometrically with h have been absorbed into a redefinition of the coupling g. As is well known, $Z'(1) = \det'(\Delta)$ (Δ is the scalar Laplacian, and the prime denotes the removal of the zero eigenvalue), and $Z(2)$ is the ghost determinant. One reason for our considering the partition function is that it is a manifestly positive quantity, and hence one can make unambiguous statements about the divergence of the perturbation series. Our arguments can be extended trivially to the two-point amplitude for the trace of the graviton at zero momentum, which is another example of a real amplitude. We certainly expect that the divergence of perturbation theory is, as in ordinary field theory, universal and will occur in any string amplitude and in any string theory.

As is well known, there are divergences in the bosonic string because of the tachyon, and because of the dilaton tadpole. These show up as bad behavior of the string integrand at points in moduli space where geodesics pinch off. These points constitute the so-called degeneration locus. They give rise to infinities that are an indication of the perturbative instabilities of the flat-space vacuum. We remove these divergences by introducing a genus-independent cutoff on the minimum length of geodesics. This restricts moduli space and renders the loop amplitudes finite. We shall then show that the cutoff h-loop amplitude is bounded from below by $h!$, uniformly in the cutoff. The reason it makes sense to do this is that the infrared divergences of the individual terms have nothing to do with the $h!$ growth of these terms and thus with the divergence of the series. This is similar to the situation in quantum field theory, where the divergences of perturbation theory arise from different sources than the ultraviolet divergences of individual terms in the series. We are unfortunately unable to cope with the problems introduced by fermionic degrees of freedom; after all, even in ordinary field theory it is very hard to estimate the divergences of perturbation theory in the presence of fermions. We would expect, nonetheless, that a similar divergence would arise in the case of the superstring or the heterotic string, where there are no infrared divergences at all.

We first discuss bounding the string integrand. Note that everything in this paper that relates to Riemann surfaces is described in terms of constant ($=-1$) curvature metrics. A will always stand for the area of the Riemann surface; since we shall deal solely with compact Riemann surfaces without boundary, $A=2\pi(2h-2)$. We shall be concerned only with the scalar Laplacian. Recall that the Selberg Z-function form of the string partition function arises from the usual ζ function

$$\zeta(s)\equiv \mathrm{tr}(\Delta^{-s})=\frac{1}{\Gamma(s)}\int dt\, t^{s-1}\mathrm{tr}[\exp(-t\Delta)].$$

(We ignore the presence of zero modes for the moment.) The function $\theta(t)\equiv\mathrm{tr}[\exp(-t\Delta)]$ is just the trace of the heat kernel; application of the Selberg trace formula to the heat kernel leads to McKean's formula[3]

$$\frac{Z'(s)}{Z(s)}=(2s-1)\int_0^\infty dt\exp[-s(s-1)t][\theta(t)-Ak(t)]. \tag{1}$$

Here, $k(t)[\equiv k(z,z;t)]$ is the heat kernel on the hyperbolic plane (evaluated at coinciding points). Explicitly,

$$k(t)\equiv\frac{\sqrt{2}\exp(-t/4)}{(4\pi t)^{3/2}}\int_0^\infty\frac{\exp(-b^2/4t)}{(\cosh b-1)^{1/2}}b\,db\sim\frac{1}{4\pi t}-\frac{1}{12\pi}+O(t)\quad \text{as } t\to 0^+.$$

We are interested in (1) as $s\to 1$ (or 2), and therefore in the behavior of the integrand near $t=0$ and $t=\infty$ (for $s\to 1$). Let us first consider the short-time behavior of the heat kernel. Define $\hat\theta(t)\equiv\theta(t)-1$, where we are subtracting the contribution of the trivial constant eigenfunction. Recall that

$$\theta(t)\sim\frac{A}{4\pi t}-\frac{A}{12\pi}+O(t)\quad\text{as } t\to 0.$$

The coefficients of positive powers of t in this asymptotic expansion are integrals of polynomials in the curvature tensor. For constant-curvature metrics, we can therefore write the entire asymptotic expansion as $\theta(t)\sim A\times p(t)$, where p is entirely independent of the manifold, and only reflects the *local* properties of constant-curvature metrics in two dimensions. A simple consequence of this fact is

$$f(t)\equiv\hat\theta(t)-Ak(t)\sim -1+o(t^n) \tag{2}$$

for small t, where n can be made arbitrarily large.

We are staying away from the degeneration locus; by a theorem of Schoen, Yau, and Wolpert,[4] this implies that the lowest nontrivial eigenvalue of Δ is bounded away from zero, by a constant depending only on the genus and on the length of the shortest disconnecting curve on the surface. We denote this lower bound by $\rho\equiv\rho(h)>0$, so that $\hat\theta(t)=O(\exp(-\rho t))$ for large t.

We may now integrate (1), arriving at

$$\frac{d}{ds}\ln\frac{Z(s)}{s(s-1)}=-\frac{d}{ds}\int_0^\infty\exp[-ts(s-1)][f(t)-f(0)]\frac{dt}{t}$$
$$-f(0)\frac{d}{ds}\left[\int_0^\infty\{\exp[-ts(s-1)]-\exp[-ta(a-1)]\}\frac{dt}{t}+C\right], \tag{3}$$

where a and C are constants that we have to determine. The form in which we have displayed (3) is one in which both integrals on the right-hand side make sense. To determine the constants, we note that

$$\int_0^\infty\frac{dt}{t}[\exp(-ta)-\exp(-tb)]=-\ln\left(\frac{a}{b}\right)$$

and that $Z(\infty)=1$, arriving at

$$\ln\frac{Z(s)}{s(s-1)}=-\int_0^\infty\frac{dt}{t}\{\exp[-ts(s-1)]f(t)+\exp(-t)\}. \tag{4}$$

We may now take the limit $s\to 1^+$ since the right-hand side makes sense. In this limit, the left-hand side is just $\ln Z'(1)$.

We are interested in an upper bound on $Z'(1)$ and a lower bound on $Z(2)$. As mentioned above, we need to show that the string integrand does not decrease too rapidly for large h. Equation (4) implies

$$Z'(1)=\exp\left[-\int_0^\infty\frac{dt}{t}[\hat\theta(t)-Ak(t)+\exp(-t)]\right].$$

Using (2), we break the integral into two parts, so that we have

$$Z'(1) < \exp(\epsilon)\exp\left[-\int_\epsilon^\infty \frac{dt}{t}[\hat{\theta}(t) - Ak(t) + \exp(-t)]\right] = C_\epsilon \tilde{c}_\epsilon^A \exp\left[-\int_\epsilon^\infty \frac{dt}{t}\hat{\theta}(t)\right].$$

This is enough for our purposes because $\hat{\theta} > 0$. Hence, apart from a factor that depends on $h = 2 + A/4\pi$ in an acceptable exponential manner, we conclude that $Z'(1)$ cannot become "too large" at large h. A similar analysis of (4) at $s = 2$ leads to

$$Z(2) > Z(1+\delta) > Z'(1)K(\delta,\epsilon)^A M(\delta,\epsilon),$$

where $K(\delta,\epsilon)$ and $M(\delta,\epsilon)$ are constants independent of the genus. This completes the desired lower bound on the integrand.

We come now to the second part of our argument. What is the volume of moduli space for surfaces? Clearly, since we have shown that the string integrand is bounded by c^h, to show that perturbation theory diverges we must exhibit a factorial growth in h of the volume of the moduli space of surfaces of genus h.

One way to construct Riemann surfaces is to use trivalent graphs (i.e., graphs with three edges emanating from every vertex) as glueing diagrams for pairs of pants. These lead to distinct Riemann surfaces if the pairs of pants all have short geodesics. It is known from the work of Bollobás[5] that the number of isomorphism classes of trivalent graphs without loops or multiple edges increases factorially, i.e., is $O(h!)$. Recall that the Weil-Petersson Kähler form, in terms of length-twist coordinates, is given by $\omega = \sum l\, dl \wedge d\alpha$, the sum running over the $3h-3$ geodesics associated with the decomposition of the surface into pairs of pants. This implies that the Weil-Petersson volume of a small (hyper)annulus (because we are excluding the degeneration locus) in moduli space goes as c^{3h-3}, where c is a constant depending only on the length of the shortest geodesic we include in the annulus and on the thickness of the annulus (both of which may be chosen to be constants independent of the genus). We conclude that the volume of moduli space is bounded below by $c^{3h-3}c'h!$. This concludes our proof that the h-loop bosonic partition function, cut off in the infrared, is bounded below by $k'k^h h!$ and thus that the perturbative expansion diverges as $\sum g^{2h} h!$.

The fact that every term in the series is positive implies that the series is not Borel summable. In other words the Borel transform of a series $F(g) = \sum g^h f_h$ where $f_h \approx h!$, namely $F_{\text{Borel}}(g) = \sum g^h f_h / h!$, is singular on the real axis, which prevents us from recovering the original function as $F(g) = \int_0^\infty F_{\text{Borel}}(gt) e^{-t}$, without additional (nonperturbative) information of how to integrate about the singularity. This is just as well, for if string perturbation theory were Borel summable then many properties that hold order by order in perturbation theory would be true of the complete theory.[6] This would certainly be true of linear relations between S-matrix amplitudes, so that we could prove, for example, that supersymmetry could not be broken.

What is the origin and meaning of this divergence? It is similar in form to the ever-present divergence of quantum field theory. These can be seen to arise because one is expanding the functional integral $\int \mathcal{D}\Phi \exp[\mathcal{L}_0(\Phi) + \mathcal{L}_{\text{int}}(\Phi)]$ in powers of the interaction Lagrangean, $\mathcal{L}_{\text{int}}$, which dominates the free Lagrangean $\mathcal{L}_0$ for large fields no matter how small the coupling. Indeed, the nature of the divergence can be deduced for a given Lagrangean, say $\mathcal{L} = \phi K \phi + g\phi^p$, by considering the functional integral for a *zero*-dimensional path integral $\int d\phi \exp[-(\phi K \phi + g\phi^p)]$, whereby we conclude that a $g\phi^p$ interaction leads to a perturbation series that diverges as $\sum (-g)^n [(p-2)n/2]!$. The divergence that we have found, $\sum g^{2h} h!$, is indicative of a $g\phi^3$ interaction, consistent with the form of certain string field theories.[7] A more physical interpretation of the divergence of the perturbative expansion of quantum field theory is that of Dyson,[8] i.e., that it is due to the instability of the theory for negative (or imaginary) values of the couplings. Indeed, the high-order behavior of the perturbative series can be deduced from a semiclassical analysis of the vacuum instability for negative couplings.[9] It would be of great interest to pursue this line of thought for string theory.

What is the meaning of the fact that the perturbation series is not Borel summable? In ordinary quantum mechanics and in quantum field theory such a situation is an indication of the instability of the perturbative vacuum. Thus, in most cases, a non-Borel-summable perturbation series can be traced to the existence of instantons (which indicate vacuum mixing through tunneling), or Euclidean bounces (which are an indication of the decay of a false vacuum) or renormalons[10] (which are an indication of ultraviolet troubles). This is reasonable, since if perturbation theory is summable, one can sum it, without nonperturbative input, and the vacuum should be qualitatively correct. We therefore suggest that the non-Borel-summable perturbative expansion of string theory is similarly an indication of nonperturbative effects that destabilize the perturbative ground state. Since we might expect a similar divergence for the heterotic string, we conclude that the enormous multitude of classical heterotic solutions might all be unstable, and that the truly stable (and perhaps unique) ground state is picked out by the nonperturbative dynamics that destabilizes them. It would be of great interest to explore this idea with the hope of identifying this nonperturbative mechanism.

We thank P. Doyle, C. McMullen, and W. Thurston for their generosity in helping us understand the vagaries of moduli spaces and the spectra of Laplacians. One of us (V.P.) wishes to thank E. D'Hoker, R. de la Llave, E. Lieb, A. Uribe, and K. Aoki for several conversations. We acknowledge conversations with R. Penner during the early stages of this work.

[1]A. Jaffe, Commun. Math. Phys. **1**, 127 (1965).

[2]E. D'Hoker and D. H. Phong, Nucl. Phys. **B269**, 205 (1986).

[3]H. P. McKean, Comm. Pure Appl. Math. **25**, 225 (1972).

[4]R. Schoen, S.-T. Yau, and S. A. Wolpert, in *Proceedings of the American Mathematical Society Symposium on the Geometry of the Laplace Operator,* edited by R. Osserman and A. Weinstein (American Mathematical Society, Providence, RI, 1980), p. 279.

[5]B. Bollobás, J. London Math. Soc. (2) **26**, 201 (1982), and European J. Combin. **1**, 311 (1980).

[6]B. Simon, Phys. Rev. Lett. **28**, 1145 (1972).

[7]M. Kaku and K. Kikkawa, Phys. Rev. D **10**, 1110, 1823 (1974); M. Kaku, Phys. Rev. D **10**, 3943 (1974); E. Cremmer and J.-L. Gervais, Nucl. Phys. **B76**, 209 (1974), and **B90**, 410 (1975); M. B. Green and J. H. Schwarz, Nucl. Phys. **B64**, 205 (1973), and **B69**, 77 (1974), and **B83**, 413 (1974); E. Witten, Nucl. Phys. **B268**, 253 (1986).

[8]F. J. Dyson, Phys. Rev. **85**, 631 (1952).

[9]L. N. Lipatov, Pis'ma Zh. Eksp. Teor. Fiz. **24**, 179 (1976) [JETP Lett. **24**, 157 (1976)]; see also E. Brézin, J. C. Le Guillou, and J. Zinn-Justin, Phys. Rev. D **15**, 1544 (1977).

[10]G. 't Hooft, in *The Whys of Subnuclear Physics,* edited by A. Zichichi (Plenum, New York, 1979).