Theta Dependence in the Large N Limit of Four-Dimensional Gauge Theories

Edward Witten
School of Natural Sciences, Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540
(Received 20 July 1998)

The θ dependence of pure gauge theories in four dimensions can be studied using a duality of large N gauge theories with string theory on a certain spacetime. Via this duality, one can argue that for every θ, there are infinitely many vacua that are stable in the large N limit. The true vacuum, found by minimizing the energy in this family, is a smooth function of θ except at $\theta = \pi$ where it jumps. This jump is associated with spontaneous breaking of CP symmetry. Domain walls separating adjacent vacua are described in terms of wrapped six-branes. [S0031-9007(98)07077-X]

PACS numbers: 11.15.Pg, 11.25.Hf, 12.38.Aw

In weak coupling, the dependence of four-dimensional gauge theories on the theta angle is computed via instantons. An instanton or anti-instanton contribution is proportional to $\exp(-8\pi^2/g^2)\exp(\pm i\theta)$. For example, in spontaneously broken gauge theories, in which instantons have a characteristic maximum size, and a characteristic effective coupling, the θ dependence is determined by an instanton expansion.

For unbroken asymptotically free gauge theories, the situation is rather different. In such theories, at the classical level, instantons come in all sizes. The infrared behavior of the instanton gas is difficult to understand, and it is not clear what effective description should be used at long wavelengths to describe the theta dependence, or other aspects of the physics.

One sharp way to pose the question is to consider the large N limit of an SU(N) gauge theory [1]. The large N limit is an important avenue for understanding the dynamics of pure gauge theory, or gauge theory with a small number of light quark flavors, in four dimensions. The large N limit is attained by taking $N \to \infty$ with $\lambda = g^2 N$ fixed. Thus the amplitude for an instanton or anti-instanton of definite size is weighted by a factor of $\exp(-8\pi^2 N/\lambda)$, and it appears that instanton effects would vanish exponentially for $N \to \infty$. However, there are a variety of reasons to believe that, because of infrared divergences, this is not so. (In contrast, $\mathcal{N} = 4$ super-Yang-Mills theory, which is scale invariant rather than asymptotically free, does not have these infrared divergences, and does have exponentially small θ-dependent effects, which can be computed via instantons [2]. These instantons are related to the anti–de Sitter/conformal field theory (AdS/CFT) correspondence to -1-branes [3].) For example, if light quarks are included, then the θ dependence can be seen in current algebra [4,5], by reinterpreting some old pre-QCD computations [6]. One can show that if chiral symmetry breaking survives in the large N limit, then so does the theta dependence. Moreover, the most plausible interpretation of how the solution of the U(1) problem of QCD fits into the $1/N$ expansion implies that in the pure gauge theory, the θ dependence of the ground state energy is present to leading order in $1/N$ [7]. (An explanation of exactly what is meant by "leading order" will emerge below.) These arguments suggest that, as in some two-dimensional models where somewhat similar questions can be asked [8,9], the θ dependence of pure gauge theory in four dimensions (with or without a small number of matter fields in the fundamental representation of the gauge group) is present in the leading order of the $1/N$ expansion.

If so, one can draw an interesting deduction about the form of the theta dependence [5]. In any theory with $N \times N$ matrix fields Φ_i, the large N limit is obtained by taking a Lagrangian of the form

$$L(\Phi_i) = NS(\Phi_i; w_\alpha), \tag{1}$$

where S is independent of N and the w_α are parameters such as bare masses and coupling constants. With the normalization in (1), the large N limit is obtained by keeping w_α fixed as $N \to \infty$.

Now instead, the most general renormalizable Lagrangian for gauge fields in four dimensions takes the form

$$L = \frac{N}{4\lambda} \operatorname{Tr} F_{\mu\nu}F^{\mu\nu} + \frac{\theta}{16\pi^2} \epsilon^{\mu\nu\alpha\beta} \operatorname{Tr} F_{\mu\nu}F_{\alpha\beta}\,. \tag{2}$$

The normalization is chosen so that θ is an angular variable. The general recipe of (1) would tell us to set $\theta = N\psi$ and keep ψ fixed for $N \to \infty$. In the large N limit, the vacuum energy E is proportional to N^2 (as the number of degrees of freedom is of that order). So we expect $E(\psi) = N^2 h(\psi)$ for some function h which should have a limit as $N \to \infty$. In terms of θ, that means

$$E(\theta) = N^2 h(\theta/N)\,. \tag{3}$$

In addition, E must obey

$$E(\theta) = E(\theta + 2\pi)\,. \tag{4}$$

These conditions are, however, practically incompatible: a smooth function of θ/N cannot be invariant to $\theta \to \theta + 2\pi$ unless it is constant.

The most plausible way out seems to be [5] that $E(\theta)$ is a multibranched function because of many candidate vacuum states that all become stable (but not degenerate) for $N = \infty$. Such behavior occurs in many two-dimensional models [10] (including [9] some with $1/N$ expansions that raise issues like those we are discussing here). In the kth vacuum, the energy would be

$$E_k(\theta) = N^2 h\big((\theta + 2\pi k)/N\big). \tag{5}$$

The truly stable vacuum would be found, for each θ, by minimizing E_k with respect to k. The actual vacuum energy would be therefore

$$E(\theta) = N^2 \min_k h\big((\theta + 2\pi k)/N\big). \tag{6}$$

This function is periodic in θ, but (if h is not constant) it is not smooth—at some value of θ there is a jump between two different "branches."

Under a CP transformation, one has $\theta \to -\theta$. So in particular CP is a symmetry if and only if θ equals 0 or π. Hence $h(\theta) = h(-\theta)$. Note that CP acts by $k \to -k$ at $\theta = 0$, and by $k \to -1 - k$ at $\theta = \pi$.

Moreover, $E(\theta)$ has its absolute minimum at $\theta = 0$, because precisely at $\theta = 0$ the integrand of the Euclidean space path integral is real and positive. (At $\theta = 0$ all contributions to the path integral receive positive weights; at $\theta \neq 0$, this is not so because the instanton factor is $e^{i\theta}$. The Euclidean space path integral in volume V computes $\exp[-VE(\theta)]$, so $E(\theta)$ is minimized by maximizing the Euclidean space path maximal, which happens when the weights are all positive.) If the vacuum is unique at $\theta = 0$, then the minimum in (6) occurs for $k = 0$ (otherwise k and $-k$ would both contribute). Moreover, one expects that $d^2h/d\theta^2 \neq 0$ at $\theta = 0$ because of arguments involving the U(1) problem in the theory with quarks [7], or simply because of the absence of a symmetry that would make this quantity vanish. If so, we can set $h(\theta) = C\theta^2 + \ldots$, where C is positive and the higher order terms do not contribute to (6) in leading order in $1/N$. Thus, one conjectures that the large N structure of the vacuum energy is

$$E(\theta) = C \min_k (\theta + 2\pi k)^2 + O(1/N), \tag{7}$$

with some constant C. This function exhibits a nonanalyticity at $\theta = \pi$, which we associate with a jump between two vacua (with $k = 0$ and $k = -1$) and the spontaneous breaking of CP invariance.

The computation.—In what dynamical approximation to gauge theory can one hope to check the ideas that were just reviewed? All known approaches to the dynamics of four-dimensional gauge theory in which one can exhibit any of the difficult properties like confinement and the mass gap involve replacing the theory by a simpler theory which is hoped to be in the same universality class. There are several candidates for what the simpler theory can be. One candidate is lattice gauge theory. This framework makes it possible to compute a great deal, especially upon using computer simulation to get away from the strong coupling limit. But it is not very convenient for discussing the θ dependence, particularly if one wishes to probe issues in which the invariance under $\theta \to \theta + 2\pi$ is important. Another possibility is to consider the realization of four-dimensional gauge theory via M-theory five-branes [11], where again, at the cost of replacing the theory of interest by a simplified version that is hopefully in the same universality class, one can demonstrate the mass gap and confinement. We will instead study the problem in yet a third framework in which one can exhibit the mass gap and confinement in the context of a simplified version of four-dimensional gauge theory. This involves a circle of ideas connected with the correspondence between conformal field theory and quantum gravity on anti–de Sitter space [12–14].

To get to the specific issues of interest here as quickly as possible, we will begin as in Sect. 4 of [15] with type IIA superstring theory on $M = \mathbf{R}^4 \times \mathbf{S}^1 \times \mathbf{R}^5$, with N parallel four-branes whose world volume is $V = \mathbf{R}^4 \times \mathbf{S}^1 \times x$; here x is a point in $\mathbf{R}^5$. We assume that the "spin structure" is such that fermions change sign in going around the $\mathbf{S}^1$. Then the theory on the branes is at low energies a pure U(N) gauge theory in four dimensions. It follows from the general AdS/CFT correspondence that the large N behavior of the SU(N) part of this gauge theory can be studied by studying weakly coupled string theory on the supergravity solution X which these branes generate. Topologically $X = \mathbf{R}^4 \times D \times \mathbf{S}^4$, where D is a two-dimensional disk. The change in topology from M to X is crucial (along with the fact that X is a smooth manifold without branes), both in the explanations of confinement and the mass gap in [15] and in the discussion below of the θ dependence. The metric of X is

$$ds^2 = \frac{8\pi}{3}\eta\lambda^3 \sum_{i=1}^{4}(dx^i)^2 + \frac{8}{27}\eta\lambda\pi\left(\lambda^2 - \frac{1}{\lambda^4}\right)d\psi^2 + \frac{8\pi}{3}\eta\lambda\frac{d\lambda^2}{\lambda^2 - 1/\lambda^4} + \frac{2\pi}{3}\eta\lambda d\Omega_4^2. \tag{8}$$

Here x^i are coordinates on $\mathbf{R}^4$, λ and ψ (with $1 \le \lambda \le \infty$, $0 \le \psi \le 2\pi$) are polar coordinates on D (note that $\lambda = 1$ is the origin of the polar coordinates, the "center" of D), and $d\Omega_4^2$ is the metric of a unit four-sphere. η is a parameter which determines how far one is from conventional four-dimensional gauge theory; for $\eta \gg 1$, the string theory on X can be studied via long wavelength supergravity, while asymptotically free gauge theory is expected to emerge in the opposite limit $\eta \to 0$. In the present paper, we work in an approximation of long wavelength supergravity, so to compare with gauge theory, we must assume that the system has no phase transition as a function of η.

How can we include θ in the formulation of gauge theory via four-branes? This can be done quite simply by including the U(1) gauge field that arises in type IIA superstring theory from the Ramond-Ramond sector. We will in this paper denote that field as a, and its field strength as $f_{ij} = \partial_i a_j - \partial_j a_i$. Let us reconsider type IIA superstring theory on $M = \mathbf{R}^4 \times \mathbf{S}^1 \times \mathbf{R}^5$ with the wrapped four-branes of world volume V. The low energy world volume effective Lagrangian of the four-branes has a term

$$\Delta L = \int_V a \wedge \frac{\mathrm{Tr}\, F \wedge F}{8\pi^2}. \tag{9}$$

(The most familiar manifestation of this term is that instantons on the four-brane are charged with respect to a—they carry zero-brane charge.) Here F is the U(N) field strength. We now modify the type IIA vacuum so that $f = 0$, but

$$\int_{\mathbf{S}^1} a = \theta_a \tag{10}$$

is possibly nonzero. (The left hand side is gauge-invariant modulo $2\pi\mathbf{Z}$, so we interpret θ_a as an angle.) At low energies in four dimensions, (9) reduces to a theta term in the gauge theory action, and the four-dimensional Yang-Mills theta angle is

$$\theta = \theta_a. \tag{11}$$

What we have done so far is just learn how to include θ in the four-brane description of four-dimensional gauge theory. Now we go over to the dual description in terms of supergravity (or string theory) on X. In doing so, we must bear in mind that the parameters of the theory are determined by specifying the type IIA vacuum far away from the branes, that is at large λ, and then the behavior at small λ is determined by the supergravity equations on X (or, if η is small, the full string theory on X) and encodes the behavior of the gauge theory. For example, in the original description with branes on M we assumed that $f = 0$. In the dual description on X, the analogous statement is that $f = 0$ for $\lambda \to \infty$. Likewise, (10) should be interpreted to mean that $\int_{\mathbf{S}^1} a = \theta_a = \theta$ if the integral is taken at large λ. If we combine these conditions (plus Stokes' theorem $\int_D f = \lim_{\lambda\to\infty} \int_{\mathbf{S}^1} a$, where $\int_D f$ is, of course, defined as $\int_D d\lambda\, d\psi\, f_{\lambda\psi}$), we learn that $\int_D f = \theta_a = \theta \bmod 2\pi\mathbf{Z}$. The $2\pi\mathbf{Z}$ indeterminacy arises because the left hand side is a well-defined real number, but θ is an angle. Hence

$$\int_D f = \theta + 2\pi k \tag{12}$$

for some integer k.

Maxwell's equations for the f field have a normalizable zero mode in which the only nonzero component is

$$f_{\lambda\psi} = \frac{6}{\lambda^7}. \tag{13}$$

The normalization has been chosen so that $\int_D f = 2\pi$. Hence, we can find a solution of (12) obeying Maxwell's equations in the simple form

$$f_{\lambda\psi} = (\theta + 2\pi k)\frac{3}{\pi\lambda^7}. \tag{14}$$

The back reaction on the geometry produced by this f field is negligible in the limit that N goes to infinity with fixed $\theta + 2\pi k$. The reason for this is that the classical action of the spacetime is of order N^2 (like the vacuum energy of the large N gauge theory to which it is dual). The kinetic energy of the f field is of relative order $1/N^2$ as it is simply

$$\int d^{10}x \sqrt{g}\, f_{ij} f^{ij}, \tag{15}$$

with no factors of N or of the string coupling constant. So to lowest order in $1/N$, we obey (12) and the classical equations of motion simply by solving for f in the fixed spacetime X.

The θ and k dependent part of the vacuum energy is found by evaluating (15) with f as given in (14). It thus takes the form foreseen in the introduction:

$$E_k(\theta) = C(\theta + 2\pi k)^2, \tag{16}$$

with some positive constant C that is independent of N. The vacuum energy for given θ is obtained by minimizing this with respect to k,

$$E(\theta) = C \min_k (\theta + 2\pi k)^2. \tag{17}$$

We have obtained precisely the structure anticipated in the introduction. For given θ, there are infinitely many vacua, labeled by the choice of an integer k. The vacuum energy $E(\theta)$ is a smooth function of θ except at $\theta = \pi$, where there is a jumping between two solutions ($k = 0$ and $k = -1$). This jumping represents spontaneous breaking of CP at $\theta = \pi$.

Finally, we should check that the vacua labeled by k are all stable for $N \to \infty$ and find a framework for estimating their lifetime for finite N. The essential point is to describe the domain wall separating vacua with adjacent values of k. The decay of a k vacuum involves nucleation of a "bubble" with a smaller value of $|\theta + 2\pi k|$ and hence a lower energy density; this bubble is bounded by a domain wall. If the energy per unit area of such a domain wall is large for $N \to \infty$, then the k-vacua have lifetimes that go to infinity for $N \to \infty$.

In fact, the domain wall is constructed simply by compactifying a type IIA six-brane on the $\mathbf{S}^4$ factor in $X = \mathbf{R}^4 \times D \times \mathbf{S}^4$. Let w be a point in D and let C be a codimension one surface in $\mathbf{R}^4$ given by, say, $x^3 = 0$ (with x^3 one of the space coordinates). Consider a six-brane whose world volume is $Q = C \times w \times \mathbf{S}^4$. The value of k jumps by ± 1 in crossing such a six-brane (the sign depends on orientations and on the direction of crossing the six-brane). The defining property of a type IIA six-brane is that if E is a two-surface that has linking number 1 with the six-brane world volume, then $\int_E f = 2\pi$. In the present case, we can take

$E = D_1 - D_2$, where D_1 and D_2 are copies of D that are respectively to the "left" or "right" of C. (In other words, $D_i = y_i \times D \times z$, where z is a point in $\mathbf{S}^4$, and the y_i are points in $\mathbf{R}^4$ to the left or right of C.) So

$$\int_{D_1} f - \int_{D_2} f = \pm 2\pi\,, \tag{18}$$

and this means that k jumps by ± 1 in crossing the six-brane. The domain wall is thus not an ordinary soliton, as one might naively have thought, but a D-brane. In particular, color flux tubes associated with quark confinement can terminate on the boundaries between different vacua, just as they can terminate [11] on chiral domain walls in $\mathcal{N} = 1$ supersymmetric gauge theory in four dimensions. (Both kinds of domain wall involve a 2π jump in θ.)

Now the energy density of a type IIA six-brane is for weak coupling of order $1/\lambda_{st}$ (λ_{st} is the type IIA string coupling constant), and, as a result, in the large N limit it is of order N. An instanton describing the decay of a k-vacuum can be constructed as a sort of six-brane bubble and has an action that grows as a power of N. So the lifetime of a k-vacuum, for any given k, is exponentially long for $N \to \infty$.

This work was supported in part by NSF Grant No. PHY-9513835. The author thanks S.-J. Rey for comments.

[1] G. 't Hooft, Nucl. Phys. **B72**, 461 (1974).
[2] M. Bianchi, M. B. Green, S. Kovacs, and G. Rossi, hep-th/9807033.
[3] T. Banks and M. B. Green, hep-th/9804170.
[4] V. Baluni, Phys. Rev. D **19**, 2227 (1979).
[5] E. Witten, Ann. Phys. (N.Y.) **128**, 363 (1980).
[6] R. Dashen, Phys. Rev. D **3**, 1879 (1971).
[7] E. Witten, Nucl. Phys. **B156**, 269 (1979).
[8] A. D'Adda, M. Luscher, and P. DiVecchia, Nucl. Phys. **B146**, 63 (1979).
[9] E. Witten, Nucl. Phys. **B149**, 285 (1979).
[10] S. Coleman, Ann. Phys. (N.Y.) **101**, 239 (1976).
[11] E. Witten, Nucl. Phys. **B500**, 3 (1997).
[12] J. Maldacena, hep-th/9711200.
[13] S. Gubser, I. R. Klebanov, and A. M. Polyakov, Phys. Lett. B **428**, 105 (1998).
[14] E. Witten, hep-th/9802150.
[15] E. Witten, hep-th/9803131.