Commun. Math. Phys. 147, 191–197 (1992)

Communications in Mathematical Physics

The Semiclassical Limit for Gauge Theory on S^2 ★

Ambar Sengupta

Department of Physics, Princeton University, Princeton, NJ 08544-0708, USA

Received July 1, 1991

Abstract. It is shown that the Yang-Mills measure $Z_h^{-1}e^{-S(\omega)/h}[D\omega]$, where $h>0$, describing gauge fields on the two-sphere converges to a probability measure on the moduli space of Yang-Mills connections on S^2, as $h\to 0$.

1. Introduction

In this paper we prove that the quantum Yang-Mills measure $d\mu_{\mathrm{YM}}^T(\omega) = \frac{1}{Z_T}e^{-S(\omega)/T}[D\omega]$ (notation to be explained in Sect. 2) for gauge fields over the two-sphere S^2 converges, as $T\to 0$, to a probability measure μ_{YM}^T on the set of minima of the Yang-Mills action functional S. The measure μ_{YM}^T has been constructed and studied in [Se 1, 2] (and, from a different point of view, by Fine in [F]) for a wide class of gauge groups. On the other hand, the minima of the Yang-Mills action S for gauge fields over S^2 are also well-understood [AB, G, FH, Se 1, NU]. In Sect. 2 we summarize the relevant results that are known and in Sect. 3 we describe the limiting process.

2. Classical and Quantum Yang-Mills on S^2

Let G be a compact connected Lie group with a fixed bi-invariant metric $\langle \cdot , \cdot \rangle_{\mathfrak{g}}$ on its Lie algebra $\mathfrak{g}$.

Equip S^2 with a Riemannian metric. If E is a Borel subset of S^2 we denote by $|E|$ its area as given by the area-measure $d\sigma$ induced by the metric. For the geometric discussions we will visualize S^2 as the usual sphere sitting in R^3 and we will equip it with a north pole n, a south pole s, and the hemispheres N and S which intersect in the equator $\mathscr{E}$. We will often refer to the meridians – these are the usual meridians

★ This work was partially supported by NSF Grants DMS-8922941, and PHY-8912067

on $S^2 \subset R^3$ running from n to s. We fix a basepoint $e_0 \in \mathscr{E}$ and denote by M_0 the meridian through it. We will work with a principal G-bundle $\pi: P \to S^2$. We fix a point u on the fiber $\pi^{-1}(n)$. The space of smooth connections on P will be denoted by $\mathscr{A}$, the group of automorphisms of P covering the identity map on S^2 by $\mathscr{G}$, and the subgroup of all those automorphisms in $\mathscr{G}$ which fix the fiber over n by $\mathscr{G}_n$. The quotients $\mathscr{C} = \mathscr{A}/\mathscr{G}$ and $\mathscr{C}_n = \mathscr{A}/\mathscr{G}_n$ will be of basic importance. If ω is a connection on P then we denote its curvature by Ω^ω. Consider any $m \in S^2$. If $e_1, e_2 \in T_pP$, where p is any point on the fiber $\pi^{-1}(m)$ and e_1, e_2 project to an orthonormal basis of T_mS^2 then the number $\|\Omega^\omega\|^2(m) = \|\Omega^\omega(e_1, e_2)\|_g^2$ is independent of the choice of p and (e_1, e_2). The Yang-Mills action $S(\omega)$ is defined to be $\int_{S^2} \|\Omega^\omega\|^2 d\sigma$, where $d\sigma$ is the area measure on S^2. The value $S(\omega)$ depends only on the class $[\omega] \in \mathscr{C}$. So S is naturally defined on the quotients $\mathscr{C}_n$ and $\mathscr{C}$.

Choose any trivializations over the hemispheres N and S which agree at the basepoint $e_0 \in \mathscr{E}$ and let $\phi: \mathscr{E} \to G$ be the transition function. Then the homotopy class of ϕ, as a loop based at $e \in G$, specifies the topology of the bundle P (see [St]). We denote this homotopy class by $[P] \in \pi_1(G, e)$.

Recall that u is a fixed point on the fiber over n. If C is a piecewise smooth closed loop in S^2 based at n then we denote by $g_u(C; \omega)$ the holonomy around C for the connection ω, with initial point u. We will often drop the subscript u in $g_u(C; \omega)$. Given C (and u) the value $g_u(C; \omega)$ depends only on the class $[\omega] \in \mathscr{C}_n$ and, conversely, the values $g_u(C; \omega)$ for all C as described above specify the class $[\omega] \in \mathscr{C}_n$ uniquely.

Recall that if $\gamma: [a, b] \to G$ is a piecewise smooth path then its *energy* is $\int_a^b \|d\gamma/dt\|^2 dt$.

The following relates the minima of S to minimum energy geodesics on G:

Theorem 2.1. *Let $[\omega] \in \mathscr{C}_n$ be a minimum of $S(\cdot)$. Then there is a unique minimum energy geodesic $\gamma^\omega: [0, |S^2|] \to G$ in the homotopy class $[P]$ such that if C is any piecewise smooth closed loop in S^2 based at n, which bounds, in the positive sense, a region $E_C \subset S^2$, then:*

$$g_u(C; \omega) = \gamma^\omega(|E_C|).$$

Conversely, if γ is a minimum energy geodesic $[0, |S^2|] \to G$ in $[P]$ then there is a unique $[\omega] \in \mathscr{C}_n$ such that $\gamma = \gamma^\omega$.

Thus there is a one-to-one correspondence between the set $\mathscr{C}_n^0$ of minima of S on $\mathscr{C}_n$ and the set $\Gamma_0^{[P]}$ of minimum energy geodesics in the homotopy class $[P]$. By taking the quotient of both sides by suitable actions of G one obtains a one-to-one correspondence between the set $\mathscr{C}^0$ of minima of S on $\mathscr{C}$ and the conjugacy classes of minimum energy loops in $[P]$.

Proof. See any of the references cited in Sect. 1 in this context. We give a brief sketch of the argument in [AB]. The Yang-Mills variational equations, in this situation, say that the curvature is a covariant constant and this can be used to show that the equation of parallel-transport corresponds to that of a geodesic on G. One then computes that $S(\omega)$ is proportional to the energy of the corresponding geodesic. □

Note that if the bundle P is trivial then the minimum of S is 0 and is given by the flat connections.

Now we turn to the quantum description. We will use the results of [Se 2] (which extends ideas used in [GKS] and [Dr] for gauge fields on the plane to those over S^2). In that work the Euclidean quantum field measure μ_{YM} representing gauge fields over S^2 was constructed for gauge groups G with compact universal cover (G compact semi-simple, for example) and for G abelian. If G is a general compact connected group covered by the product of a compact simply connected group H with N copies of the real line, and the metric on g is the product of the usual metric on R^N and an invariant metric on the Lie algebra of H then the theory extends in a straightforward way to the gauge group G as well. The discussion below applies to such situations. In the quantum setting, the space $\mathscr{C}_n$ of gauge equivalence classes of smooth connections is replaced by a larger space $\overline{\mathscr{C}_n}$. On $\overline{\mathscr{C}_n}$ is defined the Yang-Mills probability measure which has the heuristic form $d\mu_{YM} = Z^{-1}e^{-S(\omega)}[D\omega]$, where $[D\omega]$ denotes the pushforward of "Lebesgue measure" on $\mathscr{A}$ to $\mathscr{C}_n$, and Z is a "normalizing constant" insuring that $\mu_{YM}(\mathscr{C}_n)=1$. We now pause to give a summary description of $\overline{\mathscr{C}_n}$ (details may be found in [Se 2]) – this material is not essential to the understanding of the discussions that follow it.

The sphere S^2 is divided into the two hemispheres N and S, as before, intersecting in the equator $\mathscr{E}$; a base meridian M_0 is fixed and this meridian intersects $\mathscr{E}$ at the point e_0. Let us first consider the part P_N of P which is over N. Fix a point u on the fiber over n and corresponding to any connection ω on P_N define a section ("radial gauge") s_ω^N of P_N by parallel-translating u along meridial lines. Define $F^\omega: N \to g$ by requiring that $(s_\omega^N)^*\Omega^\omega = F^\omega d\sigma$. Let $\mathscr{C}_n^N$ denote the quotient of the space of connections on P_N by the group of gauge-transformations which fix the fiber over n. Then the assignment $[\omega] \mapsto F^\omega$ sets up a well-defined bijective correspondence between $\mathscr{C}_n^N$ and the space X_N of smooth g-valued functions on N. By use of this map it is standard practice to identify the Yang-Mills measure for gauge fields over N with Gaussian measure on the space X_N described heuristically by a density proportional to $e^{-\|F\|^2_{L^2(D;g)}}$ (the space X_N has a natural inner-product structure and hence, informally, a "Lebesgue measure" defined on it; the density just referred to is with respect to this Lebesgue measure). To be quite precise the Gaussian measure is defined on some Banach space $\bar{X}_N$ containing X_N but we will write X_N instead of $\bar{X}_N$. The F^ω is now replaced by the following stochastic analog: for any Borel set $E \subset N$, there is a Gaussian random variable $F(E)$ on X_N, taking values in g, which is the analog of $\int_E F^\omega d\sigma$. We now outline how parallel-translation is defined in this context. Consider a well-behaved curve $C:[a,b] \to N$ and, for each $t \in [a,b]$, denote by C_t the loop based at n obtained by following the meridial segment from n to $C(a)$, followed by C up to time t and then followed by the meridial segment back to n. If $g(C_t;\omega)$ denotes the holonomy, with initial point u, around C_t with respect to a smooth connection ω then it is an immediate consequence of the definition of parallel-translation that $g_a=e$ and $dg_t = -dM_t g_t$, where M_t is the integral of F^ω over the region E_t whose positive boundary is formed by C_t. To obtain the quantum analog we replace the differential equation by its stochastic form (interpreting it as a Stratonovich stochastic differential equation) and take M_t to be $F(E_t)$. Put another way, g_t describes Brownian motion on G with time clocked by $|E_t|$ instead of t. We say that the random variable g_b describes stochastic parallel translation along the entire curve C. We can carry out an exactly analogous procedure over S, using a section s_ω^S and obtaining a space X_S corresponding to X_N. The transition function between the sections s_ω^N and s_ω^S can be taken as the (random) function $\phi: \mathscr{E} \to G$ given by $\phi(m) = g_N(e_0m)g_S(e_0m)^{-1}$, where $g_N(e_0m)$ gives the stochastic parallel-transport along the part of $\mathscr{E}$ from e_0 to m with respect to the connection as viewed from N and $g_S(e_0m)$ is the corresponding quantity for S. For the Yang-Mills space $\overline{\mathscr{C}_n}$ for S^2 we take the product probability space $X_N \times X_S$ and *condition* the measure so that ϕ describes a loop in G in the homotopy class $[P]$. Thus is obtained $\overline{\mathscr{C}_n}$ and μ_{YM} on $\overline{\mathscr{C}_n}$. If C is a well-behaved curve in either N or S then $g(C)$ has been defined as a random variable on X_N or X_S and, viewing it as a random variable on the product $X_N \times X_S$ in the natural way, $g(C)$ is defined as a random variable on $\overline{\mathscr{C}_n}$ and it is well-defined under the measure μ_{YM}. If C is a closed loop based at n but passing through both hemispheres then $g(C)$ is defined by breaking up C into pieces in N and S and with appropriate factors involving the transition function ϕ introduced at the points where C crosses from one hemisphere to the other.

The quantum analog of the holonomy is a random variable $g(C):\overline{\mathscr{C}_n} \to G:\omega \mapsto g(C;\omega)$ associated to a closed loop in S^2 based at n. Due to technical (but conceptually irrelevant) reasons one has to restrict to a certain class of curves C. For our purposes a *curve* or *curve segment* in S^2 will always mean a piecewise smooth map of a compact interval in the real line into S^2. Let us say that a curve segment in S^2 is a *basic segment* if it is smooth one-to-one and either lies entirely on a meridian or intersects each meridian in at most one point (if the latter condition is satisfied we say that the curve is *horizontal*); a collection of basic segments is a *basic collection* if it contains finitely many segments and any two segments in the collection either do not intersect or intersect at one or both endpoints only. We say that a set of $\mathscr{S}$ of curves in S^2 is *admissible* if (i) there is a basic collection such that every curve in $\mathscr{S}$ is made up of a finite number of segments each drawn from the basic collection, (ii) $\mathscr{S}$ is non-empty but finite, and (iii) no curve in $\mathscr{S}$ is a point curve. The random variable $\omega \mapsto g(C;\omega)$ is defined whenever $\{C\}$ is admissible. For our purposes, the σ-algebra on $\mathscr{C}_n$ will be taken to be the one generated by the $g(C)$'s.

We will always work with an admissible collection $\mathscr{S} = \{C_1, \ldots, C_m\}$ of loops in S^2 all based at n. The rest of this section describes a way to compute the joint distribution of the random variables $g(C_i;\omega)$. The strategy is to construct a collection of special loops (called lassos) $L_1, \ldots, L_K$ such that each C_i is essentially a composite of a number of the L_i's (and reversed L_i's) so that $g(C_i)$ is the product of the corresponding $g(L_i)$'s [and $g(L_i)^{-1}$'s]. Thus if we know the joint distribution of the $g(L_i)$'s (under the probability masure μ_{YM}) then we would know that of the $g(C_i)$'s, too.

We draw enough meridians $M_0, \ldots, M_k$ so that the curves C_i are broken up into segments which together with the segments from the M_j form a basic collection. We label the M_i's in increasing order of the angles they make with the fixed initial meridian M_0. A *lasso* is a closed loop formed in the following way from five legs: (i) follow a meridial segment from n along some meridian M_i to the initial point of some horizontal segment σ running from M_i to M_{i+1} (here, as always, $M_{n+1} = M_0$) or until the south pole s is reached; (ii) then follow σ till it reaches M_{i+1}; (iii) move "back" along M_{i+1} towards n until the final point of some horizontal segment σ' (running from M_i to M_{i+1}) is reached or until n is reached in case there are no segments like σ'; (iv) follow σ' in reverse until M_i; (v) finally, return to n back along M_i. Note that in degenerate examples some of these legs would be absent; for example, if s is reached in step (i) then step (ii) is not necessary. Having defined a lasso we observe that the lassos can be arranged in a natural sequence $L_1, \ldots, L_K$ such that the composite curve $L_K \ldots L_1$ (read from right to left) reduces to the constant curve at n after all segments that are traversed consecutively in opposite directions are dropped. For example, one can start with L_1 as the lasso with its first ("long") leg reaching all the way along M_0 to s, L_2 as the lasso with its first leg along M_0 but "closest" to s after L_1, etc. If L is a lasso and we drop from L part of its first leg and all of its last leg then we obtain a simple closed loop (the little "square" at the head of the lasso) – we denote by $|L|$ the area of the region enclosed (in the positive sense) by this closed loop at the "tip" of L.

The following result involves the Brownian loop $[0, |S^2|] \to G$, based at e, in the homotopy class $[P]$. This is obtained by projecting onto G the corresponding Brownian bridge process on the universal cover of G. To be precise, the Brownian loop we deal with here is described by a probability measure on the space $\Lambda_{|S^2|}$ of continuous loops $[0, |S^2|] \to G$ based at e and in the homotopy class $[P]$; the basic random variables on $\Lambda_{|S^2|}$ are the maps $\gamma \mapsto \gamma(t)$, where $t \in [0, |S^2|]$. The set $\Lambda_{|S^2|}$ is a metric space under uniform convergence.

Theorem 2.2. *The G^K-valued random variable $\omega \mapsto (g(L_1; \omega), \ldots, g(L_K; \omega))$ on $\overline{\mathscr{C}_n}$ has the same distribution as $\gamma \mapsto (\gamma_{t_1}, \gamma_{t_2}\gamma_{t_1}^{-1}, \ldots, \gamma_{t_K}\gamma_{t_{K-1}}^{-1})$, where $t_i = |L_1| + \ldots + |L_i|$, and $[0, |S^2|] \to G : t \mapsto g_t$ is a Brownian loop in G, based at $e \in G$, in the homotopy class $[P]$.* □

Proof. See [Se 2]. □

3. The Limiting Process

We wish to consider the probability measure constructed in the same way as $d\mu_{\text{YM}}$ except with $S(\cdot)$ scaled to $S(\cdot)/T$, where $T > 0$. That is, we consider the measure $d\mu_{\text{YM}}^T = Z_T^{-1} e^{-S(\omega)/T} [D\omega]$.

There is an easy way to see how the measure μ_{YM}^T is related to μ_{YM}. Instead of the metric ds^2 on S^2 that we have been working with, introduce a new metric $ds'^2 = T ds^2$. Then the corresponding area-measures $d\sigma$ and $d\sigma'$ are related by $d\sigma' = T d\sigma$. Now recall that $S(\omega) = \int_{S^2} \|\Omega^\omega\|^2 d\sigma$, where $\|\Omega^\omega\|^2$ is the function on S^2 whose value at a point m is given by $\|\Omega^\omega(e_1, e_2)\|_g^2$, where (e_1, e_2) are tangent vectors to P at some point on $\pi^{-1}(m)$ and which project to a basis of $T_m S^2$ which is orthonormal *with respect to the metric ds^2*. Thus $S'(\omega)$, the corresponding object for the metric ds'^2, is related to $S(\omega)$ by: $S'(\omega) = S(\omega)/T$. This suggests that the measure μ_{YM}^T should be constructed just as μ_{YM} except all areas should be scaled by T. Both the probability space $\overline{\mathscr{C}_n}$ and the σ-algebra are the same as before but now we have a new probability measure μ_{YM}^T on $\overline{\mathscr{C}_n}$. Thus, if $\mathscr{S} = \{C_1, \ldots, C_m\}$ is an admissible collection of curves in S^2 and $L_1, \ldots, L_K$ is the sequence of lassos constructed as in Sect. 2, then the random variables $g(C_i)$ are products of the $g(L_j)$'s and $g(L_k)^{-1}$'s as before, but the joint distribution of the $g(L_i)$'s is as described in Proposition 3.1 below.

We denote by Λ_a the space of continuous loops $[0, a] \to G$, based at e, lying in the homotopy class $[P]$. The standard Brownian loop in G in the homotopy class $[P]$ is described by a probability measure $\mu_{[0,a]}$ on Λ_a. If $t \in [0, a]$ then $\gamma \mapsto \gamma(t)$ is a random variable on Λ_a (and these variables generate the σ-algebra on Λ_a). On the other hand, for $T > 0$, one also has a probability measure μ_T on Λ_a such that, for any $t_1, \ldots, t_l \in [0, a]$, the random variable $\gamma \mapsto (\gamma_{t_1}, \ldots, \gamma_{t_l})$ has the same distribution under μ_T as does $\gamma \mapsto (\gamma_{Tt_1}, \ldots, \gamma_{Tt_l})$ as a random variable on the space Λ_{Ta} with the measure $\mu_{[0, Ta]}$. Put another way, the measure $\mu_{[0, Ta]}$ describes the standard

Brownian loop $[0, Ta] \to G$ (in the homotopy class $[P]$) whereas μ_T is a measure on loops $[0, a] \to G$ (in [P]) which is related to $\mu_{[0,Ta]}$ by time scaling. In our usage, $a = |S^2|$.

Using Theorem 2.2 and the discussion above we can then formulate the relationship between μ_{YM}^T and μ_T as follows:

Proposition 3.1. *The G^K-valued random variable $\omega \mapsto (g(L_1; \omega), \ldots, g(L_K; \omega))$ on $\mathscr{C}_n$ has the same distribution with respect to the measure μ_{YM}^T as $\gamma \mapsto (\gamma_{t_1}, \gamma_{t_2}\gamma_{t_1}^{-1}, \ldots, \gamma_{t_K}\gamma_{t_{K-1}}^{-1})$ on $\Lambda_{|S^2|}$ has under the measure μ_T.* □

We now invoke the following result proved by Molchanov [Mo] and Hsu [H]:

Theorem 3.2. *The sequence of probability measures μ_T on $\Lambda_{|S^2|}$ converges weakly to a probability measure μ_0 which is concentrated on the set $\Gamma_0^{[P]}$ of minimum energy geodesic loops $[0, |S^2|] \to G$, based at e, in the homotopy class $[P]$.* □

Proof. See Sect. 5 of [Mo] or Theorem 4.2 of [Hsu]. □

Note that a minimum energy loop in G is described by a smooth map of S^1 into G.

Combining Theorem 2.2 with Proposition 3.1 we see that for any bounded continuous function f on G^K the expectation value $\int_{\mathscr{C}_n} f(g(L_1; \omega), \ldots, g(L_K; \omega)) d\mu_{YM}^T(\omega)$ converges, as $T \to 0$, to $\int_{\Gamma_0^{[P]}} f(\gamma(|L_1|), \gamma(|L_2|)\gamma(|L_1|)^{-1}, \ldots, \gamma(|L_K|)\gamma(|L_{K-1}|)^{-1}) d\mu_0(\gamma)$. Recalling the correspondence (Theorem 2.1) between $\mathscr{C}_n^0$ and $\Gamma_0^{[P]}$ we see that the measure μ_0 can be transferred to a probability measure μ_{YM}^0 on $\mathscr{C}_n^0$ and then we have as $T \to 0$:

$$\int_{\mathscr{C}_n} f(g(L_1; \omega), \ldots, g(L_K; \omega)) d\mu_{YM}^T(\omega) \to \int_{\mathscr{C}_n^0} f(g(L_1; \omega), \ldots, g(L_K; \omega)) d\mu_{YM}^0(\omega).$$

Now recall that the L_i's were constructed as tools for computing $g(C_i; \omega)$, where the C_i's constitute an admissible collection $\{C_1, \ldots, C_m\}$ of closed curves in S^2 all based at n. Now if f is a bounded continuous function on G^m then $f(g(C_1; \omega), \ldots, g(C_m; \omega))$ is of the form $F(g(L_1; \omega), \ldots, g(L_K; \omega))$ for some bounded continuous function F on G^K, since each $g(C_i)$ is a product of some $g(L_j)$'s and some $g(L_k)^{-1}$'s. Thus we have:

Theorem 3.3. *There is a probability measure μ_{YM}^0 on $\mathscr{C}_n^0$ such that for any admissible collection $\{C_1, \ldots, C_m\}$ of closed loops in S^2 based at n, as $T \to 0$*

$$\int_{\mathscr{C}_n} f(g(C_1; \omega), \ldots, g(C_m; \omega)) d\mu_{YM}^T(\omega) \to \int_{\mathscr{C}_n^0} f(g(C_1; \omega), \ldots, g(C_m; \omega)) d\mu_{YM}^0(\omega). \quad \square$$

By taking only those f which are invariant under the replacement $f \mapsto f^g$, for every $g \in G$ [where $f^g(x_1, \ldots, x_m) = f(gx_1g^{-1}, \ldots, gx_mg^{-1})$], we obtain the analogous result for the full quotient spaces $\mathscr{C}$ and $\mathscr{C}^0$.

References

[AB] Atiyah, M., Bott, R.: Proc. Indian Acad. Sci. (Math. Sci.) **90**, 11–20 (1981)

[Dr] Driver, B.K.: Commun. Math. Phys. **123**, 575–616 (1989)

[F] Fine, D.: Commun. Math. Phys. **134**, 273 (1990)

[FH] Friedrich, Th., Habermann, L.: Commun. Math. Phys. **100**, 231–243 (1985)

[G] Graveson, J.: Commun. Math. Phys. **127**, 597–605 (1990)
[GKS] Gross, L., King, C., Sengupta, A.: Ann. Phys. **194**, 65–112 (1989)
[NU] Nahm, W., Uhlenbeck, K.: The equivalence of quantized gauge fields on S^2 and the quantum mechanics of a particle moving on the group maniold. Preprint
[Se 1, 2] Sengupta, A.: The Yang-Mills measure for the two-sphere (Ph.D. Thesis, Cornell University), June 1990; J. Funct. Anal. (to appear)
[St] Steenrod, N.: The topology of fiber bundles. Princeton, NJ: Princeton University Press 1951

Communicated by K. Gawedzki