Global gauge in a non-Abelian theory

M. A. Solov'ev
P. N. Lebedev Physics Institute, Academy of Sciences of the USSR, Moscow

(Submitted 25 August 1983)
Pis'ma Zh. Eksp. Teor. Fiz. **38**, No. 8, 415–417 (25 October 1983)

A global gauge is topologically forbidden in a non-Abelian theory only for sufficiently smooth fields. The corresponding boundary is established by the Sobolev criterion. Below this boundary, the customary Coulomb conditions are not gauge conditions, even locally.

PACS numbers: 11.15 – q, 02.40.Vh

In this paper we attempt to show how the particular space of fields which is considered affects the gauge situation in a non-Abelian theory.[1,2] Here we have the peculiar circumstance that Singer's topological prohibition of a global gauge applies only to sufficiently smooth fields, but as soon as we cross the boundary below which this prohibition does not apply we find that the conditions customarily used (the Coulomb, Lorentz, and background conditions) cease to be gauge conditions even locally. Clarifying this question would be useful for going beyond perturbation theory. For example, a refinement of Gribov's suggestion regarding the path-integration region by means of a variational principle[3,4] uses a metric determined by the form of the kinetic part of the Lagrangian. The space L^2 corresponding to this metric lies below the prohibition boundary, and the gauge-transformation group has several singularities here, as does the variation principle itself. We will see below that the topology of this group changes. Furthermore, this group ceases to be a Lie group; i.e., the terms discarded from the infinitesimal-transformation expansion $e^w = 1 + w + ...$ are in fact of the same order as w.

The Sobolev spaces[6] L^p_k constitute the scale customarily adopted for tracing the changes in topology as the number of functions increases. These spaces consist of fields all of whose derivatives of order up to k have an integrable pth power. As the indices p and k become smaller, the space becomes broader. We denote by G_{k+1}^{p} the corresponding group of gauge transformations. Their derivatives are pth integrable up to order $k+1$. The boundary of interest here is defined by the inequality

$$p(k+1)>n\,. \tag{1}$$

Here n is the dimensionality of the manifold on which the field is defined. Other, stronger restrictions are also used, but it is condition (1) which is of importance for a non-Abelian theory. For spaces (1) the Coulomb condition serves as a local gauge condition.[6] To show that the theorem of Ref. 2 also holds for these spaces, it is sufficient to consider the simplest case of the SU(2) theory on the S^3 sphere. Since SU(2) is homomorphic to S^3, the gauge-transformation group G in this case consists of mappings $S^3 \to S^3$. Under condition (1), these mappings are continuous by virtue of the Sobolev theorem[5]; they permit a continuous deformation into each other only if they are of identical power. Accordingly, G breaks up into a countable number of components. The center Z of this group consists of two elements: constant mappings $S^3 \to I$ and $S^3 \to -I$. For essentially any field the stability subgroup coincides with Z. In other words, the factor group G/Z acts freely on the fields of general position, which are still called irreducible. The set of these fields is connected; i.e., any two of these fields can be connected by a continuous path which also consists of irreducible fields. This is true for all p and k. The simple and essential meaning of the theorem of Ref. 2 is as follows: If a global gauge did exist, then the set of irreducible fields could be continuously mapped onto G/Z by associating with each field that transformation which sends it to the gauge surface (this transformation would be unique by virtue of the freedom of action). This could not be the case, however, since if it were then any two elements of G/Z could be connected by a continuous path (by taking the image of the path in the field space), but this group is unconnected, as is G.

If $P(k+1)<n$, the situation is different. It is simple to see from the same example that in the topology of a space of this sort any continuous gauge transformation could be deformed into an identity; i.e., all the components of G would merge to form a single component. Without any loss of generality, we may assume that $g{:}S^3 \to \mathrm{SU}(2)$ maps the north pole N into a unit matrix. We assume $0<t\leqslant 1/2$. We imagine that the neighborhood $0<\theta<\pi t$ of the point N is stretched out into a sphere $(1-t)/t$ times. Correspondingly, the additional part of the sphere is compressed by an equal factor. What happens to the mapping g during this deformation in the limit $t \to 0$? In the small neighborhood $\pi(1-t)\leqslant\theta\leqslant\pi$ of the south pole, we have a function compressed by a factor of $(1-t)/t \sim 1/t$, while elsewhere on the sphere we have the function found from the small neighborhood of N, which is therefore approximately the same as I. We thus have $\int \mathrm{tr}(g_t - I)^+ \to 0$. We can evaluate $\int \mathrm{tr}\partial_\mu g_t \partial^\mu g_t^+$ in an analogous way: The integral over the region $0\leqslant\theta<\pi(1-t)$ falls off as t^2 (the function is stretched out, but the measure of the region is bounded), while the integral over the remainder behaves as t^3/t^2 (the function is compressed, and the measure of the region varies $\sim t^3$). We thus have $g_t \to I$ in the G_1^2 topology. These arguments also hold for the sphere S^n. The norm $\|\cdot\|_{p,k+1}$ is determined primarily by the integral of the $(k+1)$th derivative over the neighborhood of south pole. Since its measure varies $\sim t^n$, while the derivative appears raised to the power p, we have $\|g_t - I\|_{p,k+1} \to 0$ for $p(k+1)<n$. When we cross

boundary (1), the topological prohibition of a global gauge is thus lifted. On the other hand, a fundamentally new entity arises here: The structure of the smooth stratification disappears, both that of the original stratification on which the fields were connected sets and that of the infinite-dimensional stratification generated by the action of the group G on the irreducible fields.

Let us consider the space L^2 of quadratically summable fields with the scalar product $(A,B) = \int \mathrm{tr} A_\mu B^{\mu^+}$. Semenov-Tyan-Shanskiĭ and Franke[3] and Zwanziger[4] singled out from the orbits generated by G those fields which are the closest in the L^2 metric to some fixed field B. We note that if an extremum on an orbit turned out to be unique this would mean the construction of a global gauge. A calculation of the first variation, which must be understood as a variation with respect to smooth subgroups (1), quickly shows that the extremal fields satisfy the background condition $D(B)(A-B)=0$. First, however, we must show that such fields exist. (Their existence is assumed in Ref. 4 without proof, and the arguments of Ref. 3 are insufficient.) To show that an extremum is reached in L^2 we consider the orbit $A_0^g, g \in G_1^2$. We introduce $a = \inf_g \|A_0^g - B\|$, and we distinguish on the orbit those points $A_i = A_0^{g_i}$ such that $\|A_{i+1} - B\| \leqslant \|A_i - B\|$, $\lim \|A_i - B\| = a$. The sequence A_i is bounded in accordance with the norm of L^2, and from it we can single out a weakly converging subsequence. We denote the weak limit by A. The distance from A to B does not exceed a, since a strongly closed sphere is also weakly closed. Accordingly, all we have to show is that at least one of the weak limits lies on the orbit A_0. We see from the expression

$$\partial_\mu g_i = g_i A_\mu^i - A_\mu^0 g_i \tag{2}$$

that the sequence g_i is of bounded norm, $\|\cdot\|_{2,1}$. From this sequence we also extract a weakly converging subsequence. We denote the limit by g. For the subsequences we use the same notation, A_i, g_i. By virtue of Roellich's lemma,[5] we have $g_i \to g$ in accordance with the norm $\|\cdot\|_2$. We can then also prove $A = A_0^g$. Let us take the scalar product of (2) with the field ϕ, assuming it to be smooth and with a small carrier. We then take the limit. At the left we find $(\partial_\mu g, \phi_\mu)$ by virtue of the weak convergence of g_i. The second term on the right gives us $(A_0 g, \phi)$, for the same reason. We write the former as $(g_i - g)A_i + gA_i$. We then have $(gA_i, \phi) \to (gA, \phi)$ by virtue of the weak convergence of A_i, while we have $(g_i - g, \phi_\mu A_{i)}^{\mu +}$ by virtue of Roellich's lemma, since $\|\phi_\mu A_i^{\mu+}\|_2 \leqslant C$. Since the ϕ's of this type are dense everywhere, we conclude that A and A_0^g coincide as elements of L^2. Each field with a finite L^2 norm can therefore be mapped by a transformation from G_1^2 onto the plane $D(B)(A-B)=0$.

We now show that the background condition is not a gauge condition, even locally, for L^2. No matter what L^2 neighborhood of B we take, it will contain different gauge-equivalent fields which satisfy this condition. In the case of spaces (1), in contrast, this cannot happen if the neighborhood is sufficiently small.[6] As before, it is sufficient to consider the SU(2) theory and the Coulomb gauge ($B=0$). Using Gribov's notation,[1] we take the "spherically symmetric" field

$$A_j(x) = f(r)\,\hat{n}\,\frac{\partial \hat{n}}{\partial x_j}\,. \tag{3}$$

Here $\hat{n} = \Sigma_{j=1}^{i}(\chi_j/r\,\sigma_j$, where the σ_j are the Pauli matrices. Field (3) satisfies the transversality condition $\partial_j A_j = 0$. We apply the transformation $g = \exp\{\alpha(r)\hat{n}\}$ to it. A direct calculation yields $\|A\|^2 = 16\pi \int f^2(r)dr$ and

$$\|A^g\|^2 = \|A\|^2 + 16\pi \int [(2f+1)\sin^2\alpha + \alpha'^2 r^2]\,dr. \tag{4}$$

It is not difficult to specify fields (3) which are arbitrarily close to $B = 0$ and for which the integral (4) can be made negative. In this case the extremal point of the orbit, which also satisfies the transversality condition, lies even closer to $B = 0$. We set $f(r) = -\theta(r_0 - r)r^{-\beta}$, where $0 < \beta < 1/2$, and θ is the unit step function. We obviously have $\|f\|_2 \to 0$ in the limit $r_0 \to 0$. We take $\alpha(r) = \alpha_0(r/\varepsilon - 1)$, where α_0 is a standard bell-shaped function with a carrier in the interval $(-1,1)$, and we assume $\alpha_0 < \pi/2$. We then have $\alpha \geqslant \sin\alpha \geqslant 2\alpha/\pi$. In the limit $\varepsilon \to 0$, the positive part of integral (4) falls off linearly with ε, while the negative part falls off in proportion to $\varepsilon^{1-\beta}$ and is dominant. These arguments remain valid when we replace f by λf for all $\lambda > 0$. The set of extremal points of the orbits, which is obviously convex and closed, is thus not an absorbing set; i.e., a linear subspace constructed on this set will not cover the $\partial_i A_j = 0$ plane.

The topology thus changes below boundary (1), and this boundary must be taken into account in determining the actual configuration space of a non-Abelian theory.

I am deeply indebted to B. L. Voronov and V. Ya. Faĭnberg for a useful discussion.

[1]V. N. Gribov, in: Fizika élementarnykh chastits. XII shkola LIYaF (Physics of Elementary Particles. Proceedings of the Twelfth School of the Leningrad Institute of Nuclear Physics), LIYaF, Leningrad, 1977; V. N. Gribov, Nucl. Phys. **B139**, 1 (1978).
[2]I. M. Singer, Commun. Math. Phys. **60**, 7 (1978).
[3]M. A. Semenov-Tyan-Shanskiĭ and V. A. Franke, in: Voprosy kvantovoĭ teorii polya i statisticheskoĭ fiziki. 3 (Questions of Quantum Field Theory and Statistical Physics. 3), Nauka, Leningrad, 1982, p. 159.
[4]D. Zwanziger, Phys. Lett. **114B**, 337 (1982); Nucl. Phys. **B209**, 336 (1982).
[5]R. Narasimhan, Analysis on Real and Complex Manifolds (Russ. transl. Mir, Moscow, 1971).
[6]K. K. Uhlenbeck, Commun. Math. Phys. **83**, 31 (1982).

Translated by Dave Parsons
Edited by S. J. Amoretty