

Mass in Quantum Yang-Mills Theory

(Comment on a Clay Millenium Problem)

L. D. Faddeev

St. Petersburg Department of Steklov
Mathematical Institute.

Among seven problems, proposed for XXI century by Clay Mathematical Institute [1], there are two stemming from physics. One of them is called "Yang-Mills Existence and Mass Gap." The detailed statement of the problem, written by A. Jaffe and E. Witten [2], gives both motivation and exposition of related mathematical results, known until now. Having some experience in the matter, I decided to complement their text by my own personal comments¹ aimed mostly to mathematical audience.

1 What is Yang-Mills field

Yang-Mills field bears the name of the authors of the famous paper [4], in which it was introduced into physics. From mathematical point of view it is a connection in a fiber bundle with compact group G as a structure group. We shall treat the case when the corresponding principal bundle E is trivial

$$E = M_4 \times G$$

and the base M_4 is a four dimensional Minkowski space.

In our setting it is convenient to describe the Yang-Mills field as one-form A on M_4 with the values in the Lie algebra \mathcal{G} of G :

$$A(x) = A_\mu^a(x) t^a dx^\mu.$$

Here x^μ , $\mu = 0, 1, 2, 3$ are coordinates on M_4 ; t^a , $a = 1, \dots, \dim G$ — basis of generators of \mathcal{G} and we use the traditional convention of taking sum over indices entering twice.

Local rotation of the frame

$$t^a \rightarrow h(x) t^a h^{-1}(x),$$

¹The first variant was published in [3]. In this new version more details are given in the description of renormalization.

where $h(x)$ is a function on M_4 with the values in G induces the transformation of the A (gauge transformation)

$$A(x) \rightarrow h^{-1}(x)A(x)h(x) + h^{-1}dh(x) = A^h(x).$$

Important equivalence principle states, that a physical configuration is not a given field A , but rather a class of gauge equivalent fields. This principle essentially uniquely defines the dynamics of the Yang-Mills field.

Indeed, the action functional, leading to the equation of motion via variational principle, must be gauge invariant. Only one local functional of second order in derivatives of A can be constructed.

For that we introduce the curvature — two form with values in \mathcal{G}

$$F = dA + A^2,$$

where the second term in RHS is exterior product of one-form and commutator in \mathcal{G} . In more detail

$$F = F_{\mu\nu}^a t^a dx^\mu \wedge dx^\nu,$$

where

$$F_{\mu\nu}^a = \partial_\mu A_\nu - \partial_\nu A_\mu + f^{abc} A_\mu^b A_\nu^c$$

and f^{abc} are structure constants of \mathcal{G} entering the basic commutation relation

$$[t^a, t^b] = f^{abc} t^c.$$

The gauge transformation of F is homogenous

$$F \rightarrow h^{-1} F h,$$

so that the 4-form

$$\mathcal{A} = \text{tr } F \wedge F^* = F_{\mu\nu}^a F_{\mu\nu}^a d^4x$$

is gauge invariant. Here F^* is a Hodge dual to F with respect to Minkowskian metric, and d^4x is corresponding volume element. It is clear, that S contains the derivatives of A at most in second order. The integral

$$\mathcal{S} = \frac{1}{4g^2} \int_{M_4} \mathcal{A} \tag{1}$$

can be taken as an action functional. The positive constant g^2 in front of the integral is a dimensionless parameter which is called a coupling constant.

Let us stress, that it is dimensionless only in the case of four dimensional space-time.

Remind that in general the dimension of physical quantity is a product of powers of 3 fundamental dimensions [L] — length, [T] — time and [M] — mass with usual units of cm, sec and gr. However in relativistic quantum physics we have two fundamental constants — velocity of light c and Planck constant \hbar and use the convention, that $c = 1$ and $\hbar = 1$, reducing the possible dimensions to the powers of length [L]. The Yang-Mills field has dimension $[A] = [L]^{-1}$, the curvature $[F] = [L]^{-2}$, the volume element $[d^4x] = [L]^4$, so that integral in \mathcal{S} is dimensionless. Now full \mathcal{S} should be dimensionless, as it has the same dimension as \hbar , thus g^2 has dimension zero.

We see, that \mathcal{S} contains terms in powers of A of degrees 2, 3, 4

$$\mathcal{S} = \mathcal{S}_2 + \mathcal{S}_3 + \mathcal{S}_4,$$

which means that Yang-Mills field is selfinteracting.

Among many approaches to quantizing the Yang-Mills theory the most natural is that of the functional integral. Indeed, the equivalence principle is taken into account in this approach by integrating over classes of equivalent fields. So we shall use this approach in what follows. There is no place here to describe in detail this purely heuristic method of quantization, moreover it hardly will lead to a solution of Clay Problem. However it will be very usefull for an intuitive explanation of this problem, which we shall do here.

2 What is mass

It was the advent of the special relativity which has given a natural definition of mass. A free massive particle has the following expression of the energy ω in terms of its momentum p

$$\omega(p) = \sqrt{p^2 + m^2},$$

where m is called mass. In quantum version mass appears as a parameter (one out of two) of the irreducible representation of the Poincare group (group of motion of the Minkowski space).

In quantum field theory this representation (insofar as m) defines a one-particle space of states \mathcal{H}_m for a particular particle entering the full spectrum of particles. The state vectors in such a space can be described as functions $\psi(p)$ of momentum p and $\omega(p)$ defines the energy operator.

The full space of states has the structure

$$\mathcal{H} = \mathbb{C} \oplus \left(\sum_i \oplus \mathcal{H}_{m_i} \right) \oplus \cdots,$$

where one dimensional space \mathbb{C} corresponds to the vacuum state and \cdots mean spaces of many-particles states, being tensor products of one-particle spaces. In particular if all particles in the system are massive the energy has zero eigenvalue corresponding to vacuum and then positive continuous spectrum from $\min m_k$ till infinity. In other words the least mass defines the gap in the spectrum. The Clay problem requires the proof of such a gap for the Yang-Mills theory.

We see an immediate difficulty. In our units m has dimension $[m] = [L]^{-1}$. But in the formulation of the classical Yang-Mills theory no dimensional parameter entered. On the other hand, the Clay Problem requires, that in quantum version such parameter must appear. How come?

I decided to write these comments exactly for the explanation how quantization can lead to appearance of dimensional parameter when classical theory does not have it. This possibility is connected with the fact, that quantization of the interacting relativistic field theories leads to infinities — appearance of the divergent integrals which are dealt with by the process of renormalization. Traditionally these infinities were considered as a plague of the Quantum Field Theory. One can find very strong words denouncing them, belonging to the great figures of several generations, such like Dirac, Feynmann and others. However I shall try to show, that the infinities in the Yang-Mills theory are beneficial — they lead to appearance of the dimensional parameter after the quantization of this theory.

This point of view was already emphasized by R. Jackiw [5] but to my knowledge it is not shared yet by other specialists.

Sidney Coleman [6] coined a nice name "dimensional transmutation" for the phenomenon, which I am going to describe. Let us see what all this means.

3 Dimensional transmutation

The most direct way to introduce the functional integral is to consider the generating functional for the scattering operator. This functional depends

on the initial and final configuration of fields, defined by the appropriate asymptotic condition. In naive formulation these asymptotic configurations are given as solutions A_{in} and A_{out} of the linearized classical equations of motion. Through these solutions the particle interpretation is introduced via well defined quantization of the free fields. However the more thorough approach leads to the corrections, which take into account the selfinteraction of particles. We shall see below, how it is realized in some consistent way.

Very formally the generating functional $W(A_{\text{in}}, A_{\text{out}})$ is introduced as follows

$$e^{iW(A_{\text{in}}, A_{\text{out}})} = \int_{A \rightarrow \substack{A_{\text{in}}, t \rightarrow -\infty \\ A_{\text{out}}, t \rightarrow +\infty}} e^{iS(A)} dA, \quad (2)$$

where $S(A)$ is the classical action (1). Symbol dA denotes the integration and we shall make it more explicit momentarily.

The only functional integral one can deal with is a gaussian. To reduce (2) to this form and, in particular, to identify corresponding quadratic form we make shift of integration variable

$$A = B + ga,$$

where the external variable B should take into account the asymptotic boundary conditions and new integration variable a has zero incoming and outgoing components.

We can consider both A and B as connections, then a will have only homogeneous gauge transformation

$$a(x) \rightarrow h^{-1}(x) a h(x).$$

However, for fixed B the transformation law for a is nonhomogeneous

$$a \rightarrow a^h = \frac{1}{g}(A^h - B). \quad (3)$$

Thus the functional $S(B + a) - S(B)$ is constant along such “gauge orbits”. Integration over a is to take this into account. We shall denote $W(A_{\text{in}}, A_{\text{out}})$ as $W(B)$, having in mind that B is defined by $A_{\text{in}}, A_{\text{out}}$ via some differential equation. Here is the answer detailing the formula (2)

$$e^{iW(B)} = e^{iS(B)} \int \exp i \left\{ S(B + a) - S(B) + \int \frac{1}{2} \text{tr}(\nabla_\mu a_\mu)^2 dx \right\} \\ \times \det((\nabla_\mu + ga_\mu) \nabla_\mu) \prod_x da(x). \quad (4)$$

Here we integrate over all variables $a(x)$, considered as independent coordinates. Furthermore, ∇_μ is a covariant derivative with respect to connection B

$$\nabla_\mu = \partial_\mu + B_\mu.$$

The quadratic form $\frac{1}{2} \int (\nabla_\mu a_\mu)^2 dx$ regularizes the integration along the gauge orbits (3) and the determinant provides the appropriate normalization. This normalization was first realized by V. Popov and me [7] with additional clarification by 't Hooft [8]. I refer to physical literature [9], [10] for all explanations. One more trick consists in writing the determinant in terms of the functional integral

$$\det(\nabla_\mu + ga_\mu)\nabla_\mu = \int \exp i \left\{ \int \text{tr}((\nabla_\mu + ga_\mu)\bar{c}\nabla_\mu c) dx \right\} \prod_x d\bar{c}(x)dc(x)$$

over grassman algebra with generators $\bar{c}(x)$, $c(x)$ in the sense of Berezin [11]. These anticommuting field variables play only accessory role, there are no physical degrees of freedom, corresponding to them.

The resulting functional which we should integrate over $a(x)$, $\bar{c}(x)$, $c(x)$ assumes the form

$$\begin{aligned} \exp i \left\{ \frac{1}{2} (M_1 a, a) + (M_0 \bar{c}, c) + \frac{1}{g} \Gamma_1(a) \right. \\ \left. + g \Gamma_3(a, a, a) + g^2 \Gamma_4(a, a, a, a) + g \Omega_3(\bar{c}, c, a) \right\}, \end{aligned} \quad (5)$$

where we use short notations for the corresponding linear, quadratic, cubic and quartic forms in variables a and \bar{c} , c . The linear form $\Gamma_1(a)$ is defined via the classical equation of motion for the field $B_\mu(x)$

$$\Gamma_1(a) = \int \text{tr}(\nabla_\mu F_{\mu\nu}(x) a_\nu(x)) dx, \quad (6)$$

forms Γ_3 , Γ_4 and Ω_3 are given by

$$\Gamma_3 = \int \text{tr} \nabla_\mu a_\nu [a_\mu, a_\nu] dx, \quad (7)$$

$$\Gamma_4 = \frac{1}{4} \int \text{tr}([a_\mu, a_\nu])^2 dx, \quad (8)$$

$$\Omega_3 = \int \text{tr} \nabla_\mu \bar{c} [a_\mu, c] dx \quad (9)$$


and operators M_1 and M_0 of the quadratic forms look like

$$M_1 = -\nabla_\rho^2 \delta_{\mu\nu} - 2[F_{\mu\nu}, \cdot], \quad (10)$$

$$M_0 = -\nabla_\rho^2. \quad (11)$$


The equation on the external field B in the naive approach would be classical equation of motion, assuring vanishing of $\Gamma_1(a)$. This would correspond to the stationary phase method. However we shall make a different choice taking into account the appropriate quantum corrections.

It is instructive to use the simple pictures (Feynman diagrams) to visualize the objects (6)–(9). For the forms Γ_1 , Γ_3 , Γ_4 and Ω_3 they look as vertices with external lines, number of which equals number of fields $a(x)$, $\bar{c}(x)$, $c(x)$



$$\begin{array}{cccc} \Gamma_1 & \Gamma_3 & \Gamma_4 & \Omega_3 \end{array} \quad (12)$$

The Green functions G_1 and G_0 for operators M_1 and M_0 are depicted as simple lines



$$\begin{array}{cc} G_1 & G_0 \end{array} \quad (13)$$

Each end of lines in (12) and (13) bears indices x, μ, a or x, a characterizing fields $a_\mu^a(x)$ and $\bar{c}^a(x)$, $c^b(x)$. The arrow on line $- \rightarrow -$ distinguishes fields \bar{c} and c . Note, that Green functions are well defined due to homogeneous boundary conditions for $a(x)$, $\bar{c}(x)$, $c(x)$.

Now simple combinatorics for the gaussian integral which we get from (4) expanding the exponent containing vertices in a formal series, gives the following answer

$$\begin{aligned} \exp iW(B) &= \exp iS(B) (\det M_1)^{-1/2} \det M_0 \\ &\times \exp \left\{ \sum \text{connected closed graphs} \right\}, \end{aligned} \quad (14)$$

where we get graph by saturating the ends of vertices by lines, corresponding to the Green functions. The term “closed” means, that graph have no external lines.

We shall distinguish weakly and strongly connected graphs. The weakly connected graph can be made disconnected by crossing one line. (In physical literature the term “one particle reducible” is used for such graph.)

The quantum equation of motion, which we impose on B , can be depicted as

$$\times \text{---} + \text{---} \bigcirc \text{---} = 0, \quad (15)$$

where the second term in the LHS is a sum of strongly connected graphs with one external line. In the lowest approximation it looks as follows

$$\times \text{---} + g \text{---} \bigcirc \text{---} + g \text{---} \bigcirc \text{---} = 0.$$

With this understanding the expression for $W(B)$ is given by the series in the powers of the coupling constant g^2

$$\begin{aligned} W(B) = & \frac{1}{g^2} \int \text{tr}(F \wedge F^*) + \ln \det M_0 - \frac{1}{2} \ln \det M_1 \\ & + g^2 \left(\text{---} \bigcirc \text{---} + \text{---} \bigcirc \text{---} + \text{---} \bigcirc \text{---} \right) \\ & + \sum_{k=3}^{\infty} g^{2(k-1)} (\text{strongly connected graphs with } k \text{ loops}). \end{aligned} \quad (16)$$

From now on we use an “euclidian trick” here, changing x_0 to ix_0 , so that M_0 and M_1 become elliptic operators. This answer can be considered as an alternative definition of the functional integral (2). Two natural questions can be asked: 1. are the individual terms in (16) well defined?; 2. does the series converge? Whereas we know almost nothing about the second question, the answer to the first one is quite instructive. Here we are confronted with the problem of divergences and renormalization.

Let us turn to the zero order in g^2 term in (16). It is given by determinants of operators M_1 and M_0 , which clearly diverge and must be regularized. The trivial regularization is the subtraction of an infinite constant, corresponding for the dets for $B = 0$. Then we can use the formula

$$\ln \det M_i(B) - \ln \det M_i(0) = - \int_0^\infty \frac{dt}{t} \text{Tr} (e^{-M_i(B)t} - e^{-M_i(0)t}), \quad i = 0, 1. \quad (17)$$

The Green functions $D_i(x, y; t)$ of the parabolic equations

$$\frac{d D_i}{d t} + M_i D_i = 0, \quad D_i|_{t=0} = I \delta(x - y) \quad (18)$$

has the well known expansion for small t

$$D(x, y; t) = \frac{1}{4\pi^2 t^2} e^{-\frac{|x-y|^2}{4t}} (a_0(x, y) + t a_1(x, y) + t^2 a_2(x, y) + \dots),$$

where the coefficients a_0, a_1, a_2, \dots are functionals of B . (Let me remind, that we deal with 4-dimensional space-time.) Trace in (17) means

$$\int \text{tr} D(x, x; t) dx. \quad (19)$$

The coefficient a_0 is the holonomy for connection B along the straight line, connecting points x and y . Clearly $a_0(x, x)$ equals unity and so its contribution disappears from (17) due to the subtraction of $\exp -M(0)t$. Now $a_1(x, x)$ for the operator M_0 vanishes and same is true for $\text{tr} a_1(x, x)$ for M_1 . So what remains is the contribution of a_2 to (17) which diverges logarithmically in the vicinity of $t = 0$. The expansion is valid for small t , so we divide the integration in (17) as

$$\int_0^\infty = \int_0^\mu + \int_\mu^\infty \quad (20)$$

and regularize the first integral as

$$\begin{aligned} & \int_\epsilon^\mu \frac{dt}{t} \int \text{tr} a_2(x, x) dx + \\ & + \int_0^\mu \frac{dt}{t} \int (\text{tr} D(x, x; t) - \text{tr} D(x, x; t)|_{B=0} - \text{tr} a_2(x, x) + \mathcal{O}(t^2)) dx. \end{aligned}$$

In this way we explicitly separated the infinite part proportional to $\ln \epsilon/\mu$. (In physical literature one uses large momentum cutoff Λ instead of short auxilliary time ϵ ; the $\ln \epsilon/\mu$ looks like $-2 \ln \Lambda/m$, where m has dimension of mass.)

Now observe, that $\int \text{tr} a_2(x, x) dx$ is proportional to the classical action $\int \text{tr}(F \wedge F^*)$. It follows from general considerations of gauge invariance and dimensionlessness, but can be found also explicitly together with the corresponding numerical coefficient. We get

$$\begin{aligned} W(B) = & \frac{1}{4} \left(\frac{1}{g^2} + \frac{11}{48\pi^2} C(G) \ln \frac{\epsilon}{\mu} \right) \int \text{tr}(F \wedge F^*) \\ & + \text{finite zero order terms} + \text{higher order loops.} \end{aligned} \quad (21)$$

Here $C(G)$ is a value of a Casimir operator for group G in the adjoint representation.

Now we invoke the idea of renormalization a-la Landau and Wilson: the coupling constant g^2 is considered to be a function of regularizing parameter ϵ in such a way that coefficient in front of the classical action stay finite when $\epsilon \rightarrow 0$

$$\frac{1}{g^2(\epsilon)} + \beta \ln \frac{\epsilon}{\mu} = \frac{1}{g_{\text{ren}}^2}, \quad \beta = \frac{11}{3} \frac{C}{16\pi^2}. \quad (22)$$

This can be realized only if the coefficient β is positive, which is true in the case of the Yang-Mills theory. Of course $g^2(\epsilon) \rightarrow 0$ in this limit.

Similar investigation can be done for the quantum equation of motion (15), the one loop diagrams are divergent, but the infinite term is proportional to the classical equation of motion, so that (15) acquires the form

$$\nabla_\mu F_{\mu\nu} + g_{\text{ren}}^2 (\text{finite terms}) = 0.$$

Higher loops contribute corrections to the renormalization (22), however their influence is not too drastic. I can not explain this here and mention only, that it is due to important general statement, according to which the logarithmic derivative of $g^2(\epsilon)$ over ϵ does not depend on ϵ explicitly

$$\frac{d g^2(\epsilon)}{d \ln \epsilon} = \beta(g^2(\epsilon)),$$

where

$$\beta(g) = \beta g^3 + \mathcal{O}(g^5).$$

This relation is called the renormalization group equation; it follows from it and requirement, that the renormalized charge does not depend on ϵ , that the correction to (22) have form $\ln \ln \epsilon/\mu$ and lower.

We stop here the exposition of the elements of quantum field theory and return to our main question of mass. We have seen, that important feature of the definition of $W(B)$ and equations of motion was the appearance of the dimensional parameter. Thus the asymptotic states, which characterize the particle spectrum, depend on this parameter and can be associated with massive particles. Let us stress, that the divergences are indispensable for this, they lead to breaking of the scale invariance of classical theory.

In our reasoning it was very important, that divergences have logarithmic character, which is true only for the 4-dimensional space-time. All this and positivity of the coefficients β in (22) distinguishes the Yang-Mills theory

as a unique quantum field theory, which has chance to be mathematically correct.

It is worth to mention, that the disenchantment in quantum field theory in the late 50-ties and 60-ties of last century was connected with the problem of the charge renormalization. In the expressions, similar to (22), for all examples, fashionable at that time, the coefficient β was negative. It was especially stressed by Landau after investigation of the most successful example of quantum field theory — quantum electrodynamics. The realization in the beginning of 70-ties of the fact, that in Yang-Mills theory the coefficient β is positive, which is due to 't Hooft, Gross, Wilchek and Politzer, changed the attitude of physicists towards the quantum theory and led to the formulation of Quantum Chromodynamics. (This dramatic history can be found in [12].)

Conclusion

We have seen, that the quantization of the Yang-Mills field theory leads to a new feature, which is absent in the classical case. This feature — “dimensional transmutation” — is the trading of the dimensionless parameter g^2 for the dimensional one μ with dimension $[L]^2$. Also we have seen, that on a certain level of rigour, the quantization procedure is consistent. This gives us hope, that the Clay problem is soluble. Of course, the real work begins only now. I believe, that the promising direction is the investigation of the quantum equation of motion, which should enable to find solutions with nontrivial mass. One possibility will be search for solitonic solutions. Some preliminary formulas in this direction can be found in [13].

I hope, that this text could be stimulating for a mathematician seriously interested in an actual problem of the modern theoretical physics.

References

- [1] Clay Mathematics Institute Millenium Prize Problems,
<http://www.claymath.org/prizeproblems/index.htm>
- [2] A. Jaffe and E. Witten Quantum Yang-Mills Theory,
<http://www.claymath.org/prizeproblems/yang-mills.pdf>

- [3] L. Faddeev, Mass in Quantum Yang-Mills Theory (comment on a Clay Millenium Problem), *Bull. of Brazil Math. Soc.; New Series*, **33(2)**, 1-12.
- [4] C. N. Yang and R. Mills, Conservation of Isotopic Spin and Isotopic Gauge Invariance, *Phys. Rev.* **96**, 191–195, (1954).
- [5] R. Jackiw, What Good are Quantum Field Theory Infinities, *Mathematical Physics 2000* (ed. A. Fokas et al., Imperial College Press, London, 2000)
- [6] S. Coleman, Secret Symmetries: An Introduction to Spontaneous Symmetry Breakdown and Gauge Fields: Lecture given at 1973 Intern. Summer School in Phys. Ettore Majorana. Erice (Sicily), 1973, *Erice Subnucl. Phys.*, 1973.
- [7] L. D. Faddeev and V. Popov, Feynman Diagrams for Yang-Mills field, *Phys. Lett.* **B 25**, 29-30 (1967).
- [8] G. 't Hooft, Renormalization of massless Yang-Mills fields, *Nucl. Phys.* **B 33**, 173-199, 1971.
- [9] L. D. Faddeev and A. A. Slavnov, Gauge Fields: An Introduction to Quantum Theory, *Frontiers in Physics* **83**, Addison-Wesley (1991).
- [10] M. E. Peskin and D. V. Schroeder, An Introduction to Quantum Field Theory, Addison-Wesley (1995).
- [11] F. A. Berezin, The Method of Secondary Quantization, (in russian), Moscow, Nauka, 1965.
- [12] D. Gross, The discovery of asymptotic freedom and the emergence of QCD, in *At the frontier of particle physics - Handbook of QCD*, ed. M. Shifmann, World Scientific 2001.
- [13] L. Faddeev and A. J. Niemi, Aspects of Electric and Magnetic Variables in SU(2) Yang-Mills Theory, *Phys. Lett.* **B 525**, (2002) 195-200.