Mass in Quantum Yang–Mills Theory (Comment on a Clay Millennium Problem)

L. D. Faddeev

St. Petersburg Department of Steklov Mathematical Institute, St. Petersburg, Russia faddeev@pdmi.ras.ru

Summary. Among seven problems, proposed for the XXI century by the 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 comments[1] aimed mostly to mathematical audience.

1 What is the Yang–Mills Field?

The Yang–Mills field bears the name of the authors of the famous paper [4], in which it was introduced into physics. From a 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 a one-form A on M_4 with the values in the Lie algebra $\mathfrak{G}$ of G:

$$A(x) = A^a_\mu(x) t^a \mathrm{d}x^\mu \ .$$

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

Local rotation of the frame

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

[1] The 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)

$$A(x) \to h^{-1}(x)A(x)h(x) + h^{-1}\mathrm{d}h(x) = A^h(x) \ .$$

The 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 the 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 – a two-form with values in $\mathfrak{G}$

$$F = \mathrm{d}A + A^2 \ ,$$

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

$$F = F^a_{\mu\nu} t^a \mathrm{d}x^\mu \wedge \mathrm{d}x^\nu \ ,$$

where

$$F^a_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + f^{abc} A^b_\mu A^c_\nu$$

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

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

The gauge transformation of F is homogenous

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

so that the 4–form

$$\mathcal{A} = \mathrm{tr}\, F \wedge F^* = F^a_{\mu\nu} F^a_{\mu\nu} \mathrm{d}^4 x$$

is gauge invariant. Here F^* is a Hodge–dual to F with respect to the 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

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

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.

Recall that the dimension of a physical quantity is in general a product of powers of three fundamental dimensions – length [L], time [T] and mass [M], with usual units of cm, sec and gr, respectively. However, in relativistic quantum physics we have two fundamental constants – the velocity of light c

and the Planck constant $\hbar$ and we 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 $[A] = [\mathrm{L}]^{-1}$, the curvature $[F] = [\mathrm{L}]^{-2}$, the volume element $[\mathrm{d}^4x] = [\mathrm{L}]^4$, so that an integral in $\mathcal{S}$ is dimensionless. Now, all of $\mathcal{S}$ 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 two, three and four

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

which means that the Yang–Mills field is self-interacting.

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 will hardly lead to a solution of the Clay Problem. However, it will be very useful for an intuitive explanation of this problem, which we shall do here.

2 What is Mass?

It was the advent of special relativity which lead to 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 the quantum version mass appears as a parameter (one out of two) of the irreducible representation of the Poincaré group (the 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 the one-dimensional space $\mathbb{C}$ corresponds to the vacuum state and $\cdots$ means 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 $\min m_k$ to 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 $[m] = [\mathrm{L}]^{-1}$, but in the formulation of the classical Yang–Mills theory no dimensional parameter entered. On the other hand, the Clay Problem requires that in the quantum version such parameter must appear. How come?

I decided to write these comments exactly for the explanation how quantization can lead to the appearance of the 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 as Dirac, Feynmann and others. However I shall try to show that the infinities in the Yang–Mills theory are beneficial – they lead to the 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 the name "dimensional transmutation" for this phenomenon, which I am now 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 a 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 self-interaction of particles. We shall see below how it is realized in some consistent way.

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

$$\mathrm{e}^{\mathrm{i}W(A_{\mathrm{in}},A_{\mathrm{out}})} = \int_{A \to \substack{A_{\mathrm{in}}, t\to -\infty \\ A_{\mathrm{out}}, t \to +\infty}} \mathrm{e}^{\mathrm{i}S(A)} \mathrm{d}A \ , \tag{2}$$

where $S(A)$ is the classical action (1). The symbol $\mathrm{d}A$ 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 the corresponding quadratic form we make a shift of the integration variable

$$A = B + ga \ ,$$

where the external variable B should take into account the asymptotic boundary conditions and the 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

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

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

$$a \to a^h = \frac{1}{g}(A^h - B) \, . \tag{3}$$

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 $W(B)$, keeping in mind that B is defined by A_{in}, A_{out} via some differential equation. Here is the answer detailing the formula (2)

$$\begin{aligned} \mathrm{e}^{\mathrm{i}W(B)} = \mathrm{e}^{\mathrm{i}S(B)} \int \exp \mathrm{i} \left\{ S(B+a) - S(B) + \int \frac{1}{2} \operatorname{tr}(\nabla_\mu a_\mu)^2 \mathrm{d}x \right\} \\ \times \det\big((\nabla_\mu + g a_\mu)\nabla_\mu\big) \prod_x \mathrm{d}a(x) \, . \end{aligned} \tag{4}$$

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 $\frac{1}{2} \int (\nabla_\mu a_\mu)^2 \mathrm{d}x$ 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 the physical literature [9], [10] for all explanations. One more trick consists in writing the determinant in terms of the functional integral

$$\det(\nabla_\mu + g a_\mu)\nabla_\mu = \int \exp \mathrm{i} \{ \int \operatorname{tr}\big((\nabla_\mu + g a_\mu)\bar{c}\nabla_\mu c\big)\mathrm{d}x \} \prod_x \mathrm{d}\bar{c}(x)\mathrm{d}c(x)$$

over the 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

$$\begin{aligned} \exp \mathrm{i} \Big\{ \frac{1}{2}(M_1 a, a) + (M_0 \bar{c}, c) + \frac{1}{g} \varGamma_1(a) \\ + g \varGamma_3(a,a,a) + g^2 \varGamma_4(a,a,a,a) + g \varOmega_3(\bar{c}, c, a) \Big\} \, , \end{aligned} \tag{5}$$

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 $B_\mu(x)$

$$\Gamma_1(a) = \int \mathrm{tr}(\nabla_\mu F_{\mu\nu}(x) a_\nu(x)) \mathrm{d}x \ , \tag{6}$$

the forms Γ_3, Γ_4 and Ω_3 are given by

$$\Gamma_3 = \int \mathrm{tr}\, \nabla_\mu a_\nu [a_\mu, a_\nu] \mathrm{d}x \ , \tag{7}$$

$$\Gamma_4 = \frac{1}{4} \int \mathrm{tr}\big([a_\mu, a_\nu]\big)^2 \mathrm{d}x \ , \tag{8}$$

$$\Omega_3 = \int \mathrm{tr}\, \nabla_\mu \bar{c} [a_\mu, c] \mathrm{d}x \tag{9}$$

and the operators M_1 and M_0 of the quadratic forms look like

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

$$M_0 = -\nabla_\rho^2 \ . \tag{11}$$

The equation on the external field B in the naive approach would be the classical equation of motion, assuring that $\Gamma_1(a)$ vanishes. 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, the number of which equals the number of fields $a(x)$, $\bar{c}(x)$, $c(x)$

Γ_1 Γ_3 Γ_4 Ω_3 (12)

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

G_1 G_0 (13)

Each end of the lines in (12) and (13) bears indices x, μ, a or x, a characterizing the fields $a_\mu^a(x)$ and $\bar{c}^a(x)$, $c^b(x)$. The arrow on line - ->- - distinguishes the 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

$$\exp \mathrm{i}W(B) = \exp \mathrm{i}S(B)(\det M_1)^{-1/2} \det M_0 \times \exp\{\textstyle\sum \text{connected closed graphs}\} , \tag{14}$$

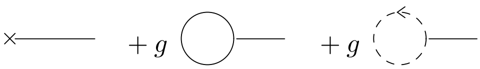
where we get a graph by saturating the ends of vertices by lines, corresponding to the Green functions. The term “closed” means that a graph has no external lines.

We shall distinguish weakly and strongly connected graphs. A 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

$$= 0 , \tag{15}$$

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

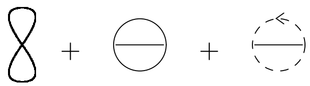


$$= 0 .$$

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

$$W(B) = \frac{1}{g^2}\int \mathrm{tr}(F \wedge F^*) + \ln\det M_0 - \frac{1}{2}\ln\det M_1$$

$$+ g^2 (\quad)$$



$$+ \sum_{k=3}^{\infty} g^{2(k-1)} (\text{strongly connected graphs with } k \text{ loops}) . \tag{16}$$

From now on we use a “Euclidean trick” here, changing x_0 to $\mathrm{i}x_0$, 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 to the dets for $B = 0$. Then we can use the formula

$$\ln\det M_i(B) - \ln\det M_i(0) = -\int_0^\infty \frac{\mathrm{d}t}{t}\,\mathrm{Tr}\big(\mathrm{e}^{-M_i(B)t} - \mathrm{e}^{-M_i(0)t}\big) , \tag{17}$$

where $i = 0, 1$. The Green functions $D_i(x, y; t)$ of the parabolic equations

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

has the well known expansion for small t

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

where the coefficients $a_0, a_1, a_2, \ldots$ are functionals of B. (Recall that we deal with 4–dimensional space–time.) The trace in (17) means

$$\int \operatorname{tr} D(x, x; t) \mathrm{d}x \, . \tag{19}$$

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 $\operatorname{tr} a_1(x, x)$ 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

$$\int_0^\infty = \int_0^\mu + \int_\mu^\infty \tag{20}$$

and regularize the first integral as

$$\begin{aligned} &\int_\epsilon^\mu \frac{\mathrm{d}t}{t} \int \operatorname{tr} a_2(x, x) \mathrm{d}x \\ &\quad + \int_0^\mu \frac{\mathrm{d}t}{t} \int \left(\operatorname{tr} D(x, x; t) - \operatorname{tr} D(x, x; t)|_{B=0} - \operatorname{tr} a_2(x, x) + \mathcal{O}(t^2)\right) \mathrm{d}x \, . \end{aligned}$$

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

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

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

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

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

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

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

A 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

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

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 the important general statement, according to which the logarithmic derivative of $g^2(\epsilon)$ over ϵ does not depend on ϵ explicitly

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

where

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

This relation is called the renormalization group equation; it follows from it and the requirement that the renormalized charge does not depend on ϵ and 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 the important feature of the definition of $W(B)$ and the 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 the breaking of the scale invariance of the 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 a chance to be mathematically correct.

It is worth to mention that the disenchantment in quantum field theory in the late fifties and sixties of the 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. This was especially stressed by Landau after investigation of the most successful example of quantum field theory – quantum electrodynamics. The realization in the beginning of seventies 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 $[\mathrm{L}]^2$. We have also 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 a promising direction is the investigation of the quantum equation of motion, which should enable us to find solutions with nontrivial mass. One possibility will be the 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 Millennium 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: Bull. of Brazil Math. Soc.; New Series **33(2)**, 1
4. C. N. Yang, R. Mills: Phys. Rev. **96**, 191 (1954)
5. R. Jackiw: What Good are Quantum Field Theory Infinities. In: *Mathematical Physics 2000*, ed by A. Fokas et al. (Imperial College Press, London 2000)
6. S. Coleman: Secret Symmetries: An Introduction to Spontaneous Symmetry Breakdown and Gauge Fields. In: *Aspects of Symmetry, Selected Erice Lectures* (Cambridge University Press, Cambridge 1985)
7. L. D. Faddeev, V. Popov: Phys. Lett. B **25**, 29 (1967)
8. G. 't Hooft: Nucl. Phys. B **33**, 173 (1971)
9. L. D. Faddeev, A. A. Slavnov: Gauge Fields: An Introduction to Quantum Theory. In *Frontiers in Physics*, vol 83 (Addison–Wesley 1991)
10. M. E. Peskin, D. V. Schroeder: *An Introduction to Quantum Field Theory* (Addison–Wesley 1995)
11. F. A. Berezin: *The Method of Secondary Quantization*, (in russian), (Nauka, Moscow 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 by M. Shifmann (World Scientific 2001)
13. L. Faddeev, A. J. Niemi: Phys. Lett. B **525**, 195 (2002)