Yang–Mills existence and mass gap

Sourav Chatterjee

Lagrangian mechanics

Lagrangian mechanics looks at the trajectory of a physical system and assigns an action to the trajectory.

For example, if $\gamma : \mathbb{R} \to \mathbb{R}^3$ is the trajectory of a particle of mass m moving in a potential V, then the action is formally

$$S(\gamma) = \int_{-\infty}^{\infty} \left(\frac{1}{2} m \|\dot{\gamma}(t)\|^2 - V(\gamma(t)) \right) dt.$$

The classical trajectory of the system must be a critical point of the action. That is, we need

$$\frac{\delta S}{\delta \gamma} = 0.$$

For the above system, the equation is $m\ddot{\gamma}(t) = -\nabla V(\gamma(t))$, which is the familiar equation of motion from Newtonian mechanics.

Path integral formulation of quantum mechanics

When we move from classical to quantum mechanics, we shift from deterministic to random trajectories.

Instead of trying to calculate the exact value of some real-valued function $F(\gamma)$ of the trajectory γ, we should now aim to compute the expected value of $F(\gamma)$.

In the path integral formulation of quantum mechanics, this is formally given by

$$\langle F(\gamma) \rangle = \frac{\int F(\gamma) e^{\frac{\mathrm{i}}{\hbar} S(\gamma)} \mathcal{D}\gamma}{\int e^{\frac{\mathrm{i}}{\hbar} S(\gamma)} \mathcal{D}\gamma},$$

where $\mathrm{i} = \sqrt{-1}$ and $\hbar$ is Planck's constant, and $\int \ldots \mathcal{D}\gamma$ denotes integration over the space of all trajectories, which is often not mathematically well-defined.

From quantum to classical

When we apply the path integral formulation in the classical world, $\hbar$ is tiny.

Heuristically, this implies that $e^{\frac{\mathrm{i}}{\hbar}S(\gamma)}$ will have huge fluctuations if we move γ slightly away from critical points of S.

Thus, again heuristically, the primary contribution to the path integrals will come from small neighborhoods of the critical points.

So, if there is a unique critical point $\widehat{\gamma}$, then $\langle F(\gamma)\rangle$ should be $\approx F(\widehat{\gamma})$ if $\hbar$ is tiny.

Maxwell's equations in vacuum

An electromagnetic field consists of an electric field $\mathbf{E}$ and a magnetic field $\mathbf{B}$. This means, at each time $t \in \mathbb{R}$ and each point $x \in \mathbb{R}^3$, we have a vector $\mathbf{E}(t, x) \in \mathbb{R}^3$ and a vector $\mathbf{B}(t, x) \in \mathbb{R}^3$.

Maxwell's equations for an electromagnetic field in vacuum govern the behavior of $(\mathbf{E}, \mathbf{B})$:

$$\nabla \cdot \mathbf{E} = 0 \qquad \text{(Gauss's law for electricity)},$$
$$\nabla \cdot \mathbf{B} = 0 \qquad \text{(Gauss's law for magnetism)},$$
$$\nabla \times \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(Faraday's law of induction)},$$
$$\nabla \times \mathbf{B} = \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \qquad \text{(Ampère's law in vacuum)},$$

where $c =$ speed of light.

Gauge field

Henceforth, we will work in units where $\hbar = c = 1$.

Question: Can we obtain Maxwell's equations as the equations for the critical point of an action?

Yes. But the procedure is a bit complicated.

First, we introduce an abstract, unphysical object called a gauge potential or gauge field $A : \mathbb{R}^4 \to (\mathrm{i}\mathbb{R})^4$.

Let us write $A = (-\mathrm{i}\phi, \mathrm{i}\mathbf{A})$, where $\phi : \mathbb{R}^4 \to \mathbb{R}$ and $\mathbf{A} : \mathbb{R}^4 \to \mathbb{R}^3$. Then A gives rise to $(\mathbf{E}, \mathbf{B})$ via

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t}, \quad \mathbf{B} = \nabla \times \mathbf{A}.$$

(But the correspondence between A and $(\mathbf{E}, \mathbf{B})$ is not one-to-one, which is why A is unphysical.)

Action of the gauge field

Henceforth, we will index elements of $\mathbb{R}^4$ as $x = (x_0, x_1, x_2, x_3)$, with x_0 denoting the time coordinate.

Define the action of $A = (A_0, A_1, A_2, A_3)$ as

$$S(A) = -\frac{1}{2g_0^2} \int_{\mathbb{R}^4} \sum_{0 \le i < j \le 3} \eta_i \eta_j (\partial_i A_j(x) - \partial_j A_i(x))^2 dx,$$

where $\eta_0 = 1$, $\eta_1 = \eta_2 = \eta_3 = -1$, and g_0 is a parameter known as the coupling constant.

It turns out that A is a critical point of S if and only if the $(\mathbf{E}, \mathbf{B})$ obtained from A solves Maxwell's equations.

(This was discovered by Weyl in the 1920s, following contributions by Minkowski and Hilbert.)

Generalization by Yang and Mills

The action defined by Weyl allowed quantization of the electromagnetic field, eventually leading to the development of quantum electrodynamics by Tomonaga, Schwinger, Feynman, and Dyson in the 1940s.

In the 1950s, Chen-Ning Yang and Robert Mills realized that the theory generalizes elegantly to allow quantization of two of the remaining three fundamental forces, namely, the strong force and the weak force.

The first key realization was that the gauge field $A : \mathbb{R}^4 \to (\mathrm{i}\mathbb{R})^4$ in Weyl's theory can be seen as a $\mathfrak{u}(1)$-valued 1-form

$$A = \sum_{j=0}^{3} A_j \, dx_j,$$

since $\mathfrak{u}(1) = \mathrm{i}\mathbb{R}$.

Generalization by Yang and Mills, continued

The next realization was that the term $\partial_i A_j - \partial_j A_i$ appearing in Weyl's action is simply F_{ij}, where the $\mathfrak{u}(1)$-valued 2-form

$$F = \sum_{0 \le i < j \le 3} F_{ij}\, dx_i \wedge dx_j$$

is the curvature form of A.

The final realization of Yang and Mills was that the theories for the weak force and the strong force can be obtained simply by replacing $\mathfrak{u}(1)$ by $\mathfrak{su}(2)$ and $\mathfrak{su}(3)$, respectively, keeping all else the same!

The Yang–Mills action

Let $\mathfrak{g}$ be any matrix Lie algebra.

Consider a theory where the gauge field A is a $\mathfrak{g}$-valued 1-form on $\mathbb{R}^4$. The curvature form of A is the $\mathfrak{g}$-valued 2-form

$$F = \sum_{0 \le i < j \le 3} F_{ij}\, dx_i \wedge dx_j,$$

where

$$F_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j].$$

The Yang–Mills action of A is

$$S(A) = -\frac{1}{2g_0^2} \int_{\mathbb{R}^4} \sum_{0 \le i < j \le 3} \eta_i \eta_j \operatorname{Tr}(F_{ij}(x)^2) dx,$$

where $\eta_0 = 1$ and $\eta_1 = \eta_2 = \eta_3 = -1$.

The Yang–Mills existence problem

Roughly speaking, the Yang–Mills existence problem is to give a rigorous mathematical meaning to expectation values of the form

$$\langle F(A) \rangle = \frac{\int F(A) e^{iS(A)} \mathcal{D}A}{\int e^{iS(A)} \mathcal{D}A},$$

where $F(A)$ is some function of A and $\int \ldots \mathcal{D}A$ denotes integration over the space of all fields.

We will now discuss a more precise version of the question.

State space

Let $\gamma : \mathbb{R} \to \Omega$ be the trajectory of a physical system, where Ω is the space in which the system takes value in a given instant of time.

Let $S(\gamma)$ be the action of the trajectory γ.

It often turns out, heuristically, that if we are interested in evaluating $\langle F(\gamma(t)) \rangle$ — that is, the expected value of some function of the value of the trajectory at a given instant of time (using path integrals), then the result does not depend on t, and is given by $\int_\Omega F d\mu$ for some probability measure μ on Ω.

In this case, we take $\mathcal{H} = L^2(\mu)$ to be the Hilbert space for our quantum system. That is, elements of $\mathcal{H}$ represent the possible states of the system at an instant of time.

Time evolution groups

Suppose we have such a state space $\mathcal{H} = L^2(\mu)$. Then, again heuristically, it often turns out that for any $F, G \in \mathcal{H}$ and $t_1, t_2 \in \mathbb{R}$, the expectation value

$$\langle \overline{F(\gamma(t_1))} G(\gamma(t_2)) \rangle$$

depends only on the difference $t = t_2 - t_1$, and moreover, there is a unitary operator $U(t) : \mathcal{H} \to \mathcal{H}$ such that

$$\langle \overline{F(\gamma(t_1))} G(\gamma(t_2)) \rangle = \langle U(t)G, F \rangle_{\mathcal{H}}.$$

It is also usually true that $(U(t))_{t \in \mathbb{R}}$ is a group of unitary operators, meaning that $U(s+t) = U(s)U(t)$ for all $s, t \in \mathbb{R}$.

Constructive field theory

It turns out that conversely, if one can construct the group $(U(t))_{t\in\mathbb{R}}$, then one can give a meaning to the expectation values defined earlier using path integrals.

Constructive field theory is a branch of mathematical physics developed in the 1960s and 1970s by many contributors (Wightman, Symanzik, Nelson, Glimm, Jaffe, Osterwalder, Schrader,...) that attempts to rigorously construct $\mathcal{H}$ and $(U(t))_{t\in\mathbb{R}}$ for various quantum field theories.

A widely accepted set of rules known as the Wightman axioms determine the properties that such a system should have, to be compatible with physical reality.

Stone's theorem

Suppose we have a group of unitary operators $(U(t))_{t\in\mathbb{R}}$ on a Hilbert space $\mathcal{H}$.

Suppose that this group is strongly continuous, meaning that for each $x \in \mathcal{H}$, the map $t \mapsto U(t)x$ is continuous.

Then, there is a theorem from functional analysis known as Stone's theorem, which says that there is a self-adjoint operator $H : \mathcal{D} \to \mathcal{H}$, where $\mathcal{D}$ is a dense subspace of $\mathcal{H}$, such that $U(t) = e^{-\mathrm{i}tH}$ for all t. Conversely, given such an H, $U(t) := e^{-\mathrm{i}tH}$ is a strongly continuous group of unitary operators.

Thus, the problem of constructing $(U(t))_{t\in\mathbb{R}}$ is equivalent to constructing the Hamiltonian operator H.

Appearance of probability theory

The key idea in constructive field theory is as follows. Recall that we aim to define H, which then generates $U(t) = e^{-itH}$ by Stone's theorem. Also, recall that $\mathcal{H} = L^2(\mu)$ for some probability measure μ on the state space Ω.

Consider a stationary Markov process $\{X_t\}_{t\in\mathbb{R}}$ with state space Ω and stationary distribution μ.

The process defines a Markov semigroup of symmetric linear operators $(P(t))_{t\geq 0}$ on $\mathcal{H}$ as

$$P(t)F(x) = \mathbb{E}(F(X_t)|X_0 = x).$$

(This is a semigroup because $P(s+t) = P(s)P(t)$ for $s, t \geq 0$.)

By the Hille–Yosida theorem, if the semigroup is strongly continuous, then there is a self-adjoint operator H such that $P(t) = e^{-tH}$. Key idea: Construct $\{X_t\}_{t\in\mathbb{R}}$ such that this H is the same as the H we want for our quantum system.

The Osterwalder–Schrader axioms

For quantum field theories, the state space Ω is usually a space of distributions on some vector space.

For example, for Yang–Mills theories, Ω should be some space of $\mathfrak{g}$-valued distributional 1-forms.

That is, instead of A being defined pointwise, we can only talk about average values of A over regions of nonzero volume.

Thus, we need to construct a suitable Markov process on such a space of distributions.

The Osterwalder–Schrader axioms give a set of conditions under which one can construct the corresponding quantum system, starting from this Markov process, following the route outlined earlier.

Mass gap

A quantum system consisting of a Hilbert space $\mathcal{H}$ and a self-adjoint Hamiltonian H is said to have a mass gap if there a gap between the lowest and second-lowest elements in the spectrum of H.

The second part of the Clay problem is to show that certain Yang–Mills theories have a mass gap.

For example, $\mathrm{SU}(3)$ theory in 4D (i.e., the theory of the strong force) is supposed to have a mass gap.

The reason for this nomenclature is that a mass gap corresponds to particles in the theory having nonzero mass. (I cannot explain the details here, due to limited time.)

For example, $\mathrm{U}(1)$ theory in 4D, which is the theory of photons, does not have a mass gap because photons are massless.

On the probability side, one needs to show that the corresponding Markov process has a spectral gap.

State of affairs in constructive field theory

Constructive quantum field theory was successfully used to construct a number of quantum field theories, such as:

- Free scalar fields in all dimensions.
- ϕ^4 theory in 2D and 3D.
- Yang–Mills theories in 2D.
- Abelian Yang–Mills theories in 2D, 3D, and 4D.

However, we do not yet have constructions of

- Any non-Abelian Yang–Mills theory in $d \geq 3$.
- Any interacting quantum field theory in 4D.

Why do we need a rigorous theory?

In the absence of a rigorous foundation for Yang–Mills theories, what do theoretical physicists do?

The only recourse for theoretical work is to apply perturbative arguments, taking the coupling parameter g_0 and doing a kind of Taylor expansion around $g_0 = 0$.

This has been hugely successful, earning several Nobel prizes. There is also a degree of rigorous mathematical justification for the perturbative calculations, e.g., in the work of Kevin Costello.

However, there are fundamental problems — most notably mass gap and quark confinement — which are believed to be beyond the reach of perturbation theory. A non-perturbative construction would be valuable for this reason.

Lattice gauge theories

Euclidean Yang–Mills theories are the Markov processes that we need to construct for applying the machinery of constructive field theory to construct quantum Yang–Mills theories and solve the Clay problem.

Unfortunately, we do not even know how to construct these stochastic objects for 4D non-Abelian theories (or even 3D non-Abelian theories).

Lattice gauge theories are discrete approximations of Euclidean Yang–Mills theories. These are mathematically well-defined.

Numerical success of lattice gauge theories

Lattice gauge theories can be used to calculate various physical quantities to a high degree of accuracy via numerical computation.

This is a very large enterprise, active since the mid-2000s, consisting of numerous collaborations that coordinate under the umbrella of USQCD (the US lattice quantum chromodynamics consortium) in the US, and sister organizations in the UK and Europe.

They have produced numbers for hadron masses and many other quantities that match experimental results at the few-percent (and sometimes sub-percent) level.

Bałaban's work

In the 1980s, Tadeusz Bałaban, in a long series of papers, attempted to construct 3D and 4D Euclidean Yang–Mills theories as continuum limits of lattice gauge theories.

As far as I know, the effort has remained unfinished; Bałaban proved that a crucial property known as ultraviolet stability holds, but the construction of the Euclidean Yang–Mills theories was not completed.

To solve the Clay problem, one needs to therefore, either

(a) understand Bałaban's work and complete it (a formidable task), or

(b) come up with a different approach (perhaps an even more formidable task).

Recent progress

In the last ten years, the probability community has spent quite a bit of effort on the Yang–Mills problem and related topics.

Two main strands have emerged:

- Gauge-string duality in lattice gauge theories.
- Stochastic quantization of Yang–Mills theories.

We will now briefly discuss these and conclude the talk.

Gauge-string duality

Gauge-string duality is the idea that certain calculations in Yang–Mills theories — which are theories of quantum fields and particles — can be transferred to calculations in string theories — which are theories of gravity.

Gauge-string duality has been a very active area in theoretical physics for the last 30 years, but has remained largely out of the reach of rigorous mathematics.

I have a paper published in 2019 which shows that calculations in certain lattice gauge theories can be transferred to certain 'lattice string theories'.

Significant generalizations and new results on this have been achieved in a series of papers in the last three years by Scott Sheffield and collaborators (Cao, Nissim, Park, Pfeffer, Yu).

Future directions in rigorous gauge-string duality

The recent results are all for lattice models. Proving gauge-string duality in the continuum remains an open problem.

A key roadblock is that we can prove the duality results only in the so-called strong coupling regimes of lattice gauge theories. One has to move to the weak coupling regime before hoping to pass to the continuum limit.

A key reason why gauge-string duality has been so successful in theoretical physics is that string theories allow calculations that Yang–Mills theories do not. The lattice string theories, so far, have not been greatly helpful with calculations.

Stochastic quantization

Recall that Euclidean Yang–Mills theories are random fields needed for constructing quantum Yang–Mills theories using the tools from constructive field theory.

In the early 1980s, Parisi and Wu came up with a novel idea for constructing Euclidean Yang–Mills theories (and other Euclidean field theories).

Their idea was to construct a Markov process whose stationary distribution would be the law of the Euclidean Yang–Mills theory.

Note the subtlety here: The Euclidean Yang–Mills theory is itself supposed to be a Markov process; Parisi and Wu proposes to construct a Markov process, which at each instant of time, is an entire Markov process representing the Euclidean YM theory.

In other words, it's a Markov process of Markov processes.

Recent progress in stochastic quantization

Strangely, although the new problem looks more difficult than the original, it seems to be more amenable to certain modern methods.

The Parisi–Wu prescription for the Markov process required for Euclidean YM theory is a singular stochastic partial differential equation (SPDE), known as the stochastic Yang–Mills heat flow.

Due to the great progress in singular SPDEs in the last decade, originating in the works of Martin Hairer, we now have the tools to construct some of these SPDEs. This includes:

- Construction of short-time solutions to 2D and 3D stochastic Yang–Mills heat flows. (Chandra, Chevyrev, Hairer, Shen, Bringmann, Cao,...)
- New proofs of difficult old results such as the construction of ϕ^4 theory in 3D. (Gubinelli, Hofmanová)

Remaining challenges

The SPDEs need to have global-in-time solutions before one can try to establish the existence of stationary distributions. This is not yet available for 3D stochastic Yang–Mills heat flow. (I have heard that it is work in progress for the 2D flow.)

The methods, as of now, do not seem to apply to 4D theories.

Conclusion

I hope I have been able to give the audience some idea of what the problem of Yang–Mills existence and mass gap is about.

The problem, though extremely hard, does not seem completely out of reach. Recent developments have been encouraging.

There is a chance that it may be solved in the next 10 to 15 years, but who knows?

Everything that I have talked about so far, will be explained in detail in a book that I am writing; it stands at 400 pages at this moment, and may grow to around 500. I will try to finish it soon.

Thank you!