An introduction to gauge theories for probabilists

Sourav Chatterjee

Part I: Quantum Yang–Mills theories

Maxwell's equations of electromagnetism

- Equations governing the evolution of an electric field interacting with a magnetic field.
- Electric field: $\mathbf{E} = (E_1, E_2, E_3)$ is a function of space and time.
- Magnetic field: $\mathbf{B} = (B_1, B_2, B_3)$.
- Jointly evolve over time, satisfying the equations:

$$\nabla \cdot \mathbf{B} = 0, \quad \frac{\partial \mathbf{B}}{\partial t} = -\nabla \times \mathbf{E},$$
$$\nabla \cdot \mathbf{E} = 0, \quad \frac{\partial \mathbf{E}}{\partial t} = \nabla \times \mathbf{B}.$$

Gauge theoretic framework for Maxwell's equations

- Instead of the fields **E** and **B**, consider a single field $A = (A_0, A_1, A_2, A_3)$ of on $\mathbb{R}^4$, where each A_i is a map from $\mathbb{R}^4$ into the imaginary axis. The field A is known as a gauge field.
- For $0 \le j, k \le 3$, define a matrix-valued function $F = (F_{jk})_{0 \le j,k \le 3}$ as
$$F_{jk} := \frac{\partial A_k}{\partial x_j} - \frac{\partial A_j}{\partial x_k}.$$
This is known as the curvature form of the gauge field A.
- Define **E** and **B** as
$$F = \begin{pmatrix} 0 & -\mathrm{i}E_1 & -\mathrm{i}E_2 & -\mathrm{i}E_3 \\ \mathrm{i}E_1 & 0 & \mathrm{i}B_3 & -\mathrm{i}B_2 \\ \mathrm{i}E_2 & -\mathrm{i}B_3 & 0 & \mathrm{i}B_1 \\ \mathrm{i}E_3 & \mathrm{i}B_2 & -\mathrm{i}B_1 & 0 \end{pmatrix}.$$
- Note that **E** and **B** are real fields.

Gauge theoretic derivation of Maxwell's equations

- It can be easily checked that the above definitions of **E** and **B** automatically ensure that the first two of Maxwell's equations are satisfied by **E** and **B** for arbitrary gauge field A (Bianchi identity).
- What about the other two equations?
- Define

$$S(A) := -\int_{\mathbb{R}^4} \sum_{j,k=0}^{3} F_{jk}(x)^2 dx,$$

assuming that the integral is finite. Note that $S(A) \geq 0$.

- Suppose that we want to find the local minimizers of $S(A)$.
- This is a variational problem. The Euler–Lagrange equations for this problem are precisely the third and fourth Maxwell equations.

Generalization

- Instead of a field $A : \mathbb{R}^4 \to (\mathrm{i}\mathbb{R})^4$, we can consider a field whose components are $\mathfrak{g}$-valued, where $\mathfrak{g}$ the Lie algebra of some matrix Lie group $G \subseteq U(N)$ for some N.
- That is, $A = (A_0, A_1, A_2, A_3)$ is a map from $\mathbb{R}^4$ into $\mathfrak{g}^4$.
- The previous case corresponded to the case $G = U(1)$, for which $\mathfrak{g} = \mathrm{i}\mathbb{R}$.
- Define F as

$$F_{jk}(x) = \frac{\partial A_k}{\partial x_j} - \frac{\partial A_j}{\partial x_k} + [A_j(x), A_k(x)],$$

so that F is a 4×4 array of elements of $\mathfrak{g}$.
- Define the Yang–Mills action

$$S_{\mathrm{YM}}(A) = -\int_{\mathbb{R}^4} \sum_{j,k=0}^{3} \mathrm{Tr}(F_{jk}(x)^2)dx.$$

- If G is a subgroup of $U(N)$ for some N, then the elements of $\mathfrak{g}$ are skew-Hermitian, and hence $S_{\mathrm{YM}}(A)$ is nonnegative.

Why generalize?

- Four fundamental interactions of nature:
  - Electromagnetism: Interactions between electric fields, magnetic fields and light.
  - Weak interactions: Interactions between sub-atomic particles that cause radioactive decay.
  - Strong interactions: Interactions between quarks that hold them together to form protons and neutrons, etc.
  - Gravity.
- The Standard Model of quantum mechanics unifies the first three interactions using the framework of gauge theories.
- Replacing $U(1)$ by $SU(3)$ gives the equations for the strong force. (Quantum chromodynamics.)
- Replacing $U(1)$ by $SU(2) \times U(1)$ gives the equations for a combined theory of electromagnetic and weak interactions. (Electroweak theory.)
- The Standard Model is more complicated than this, because it also involves matter fields. We will not talk about that.

Quantization

- Maxwell's equations are classical, in the sense that they predict a deterministic evolution.
- The quantum mechanical world is necessarily random.
- In the Lagrangian formulation of classical physics, we assign an action $S(\gamma)$ to each possible trajectory γ of a classical system.
- The system follows the trajectory that minimizes the action. This gives Newton's equations of motion.
- Path integral quantization: Roughly speaking, put a complex measure with density proportional to $e^{-\frac{i}{\hbar}S(\gamma)}$ on the space of trajectories, where $\hbar$ is Planck's constant.
- Leads to Schrödinger's equation.
- When $\hbar \to 0$, the measure concentrates near the action-minimizing trajectory due to cancellations caused by rapid oscillations away from the minimizer.
- Hard to make rigorous in many cases of interest.

Euclidean field theories

- Remove the i in the exponent. Gives a probability measure, which looks more manageable mathematically. These are known as Euclidean field theories.
- Often, physicists consider it enough to study Euclidean field theories, because the results for quantum theories can then be calculated by 'analytic continuation'.
- Euclidean field theories are natural objects for probabilists. We will therefore be content to study these, and leave aside the quantum versions.
- Example: Set of trajectories are functions from $[0,1]$ into $\mathbb{R}$, with $f(0) = f(1) = 0$. Action:

$$S(f) = \frac{1}{2}\int_0^1 f'(x)^2 dx.$$

- The probability measure on this space that has density proportional to $e^{-S(f)}$ is the law of the Brownian bridge. (What does this mean?? BB is not even differentiable!)

Brownian bridge as an Euclidean field theory

- Consider a discretization of the problem. Instead of $f : [0,1] \to \mathbb{R}$, consider $f : \{0, 1/n, 2/n, \ldots, n/n\} \to \mathbb{R}$, with $f(0) = f(1) = 0$. Then the space of trajectories is just $\mathbb{R}^{n-1}$.
- Replace $f'(x)$ by

$$\frac{f(x+1/n) - f(x)}{1/n}.$$

- Discretize the action S as

$$S(f) = \frac{1}{2n} \sum_{j=0}^{n-1} \left(\frac{f((j+1)/n) - f(j/n)}{1/n} \right)^2$$
$$= \frac{n}{2} \sum_{j=0}^{n-1} (f((j+1)/n) - f(j/n))^2.$$

- Pick a random f from the probability density on $\mathbb{R}^{n-1}$ with density proportional to $e^{-S(f)}$. Easy: As $n \to \infty$, f converges to a Brownian bridge.

Euclidean Yang–Mills theories

- Let us return to our previous setting:
  - G is a matrix Lie group contained in $U(N)$ for some N, $\mathfrak{g}$ its Lie algebra.
  - Let $\mathcal{A}$ be the space of G-connections on $\mathbb{R}^4$, namely, the set of all $A : \mathbb{R}^4 \to \mathfrak{g}^4$.
  - Define the curvature form F and the Yang–Mills action $S_{\mathrm{YM}}(A)$ as before.
- Heuristically, an Euclidean Yang–Mills theory on $\mathbb{R}^4$ is a probability measure on $\mathcal{A}$ with density proportional to $e^{-\beta S_{\mathrm{YM}}(A)}$, where β is the coupling parameter.
- Given the Brownian bridge example, one may hope to make sense of this object too, by passing from a discrete to a continuum limit.
- Unfortunately, that is an open problem (the Yang–Mills existence problem).

Gauge transforms

- Take any $A \in \mathcal{A}$ and any $U : \mathbb{R}^4 \to G$.
- For $j = 0, 1, 2, 3$, let

$$A'_j(x) = U(x)A_j(x)U(x)^{-1} + \frac{\partial U}{\partial x_j}U(x)^{-1}.$$

- The field $A' = (A'_0, A'_1, A'_2, A'_3)$ is a gauge transform of A. Each choice of U induces a gauge transform.
- It is a fact that A is also a gauge transform of A'. The fields A and A' are gauge equivalent. This is an equivalence relation.
- The space $\mathcal{A}$ of G-connections is not physically relevant. Rather, the set of gauge equivalence classes is the physically relevant space.
- For example, in Maxwell theory, two gauge equivalent connections give the same electric and magnetic fields.

Gauge invariance

- A function f on $\mathcal{A}$ is called gauge invariant if $f(A) = f(A')$ whenever A and A' are gauge equivalent.
- A function is gauge invariant if and only if it can be lifted to a function on the space of gauge equivalence classes.
- For example, it is not difficult to check that the Yang–Mills action S and the curvature form F are gauge invariant.
- Any physical observable must be gauge invariant. For example, in Maxwell theory, if a quantity is not gauge invariant, it cannot be expressed as a function of the electric and magnetic fields.
- The curvature form is a physical observable that can actually be measured in experiments.

Gauge fixing

- Gauge fixing is the process of choosing one member from each gauge equivalence class.
- This is a thorny issue in the continuum setting, because it is usually impossible to make such a choice that has desirable properties, such as smoothness (see Gribov ambiguity).
- Physicists get around this problem by focusing only on perturbative analysis, which is not hindered by problems with global gauge fixing.
- Gauge fixing is much easier in the discrete setting. More later.

More on gauge fixing

- After fixing the gauge, it suffices to look at the induced probability measure on the equivalence classes.
- Since the Yang–Mills action is gauge invariant, one can write $S_{\mathrm{YM}}(C)$ for a gauge equivalence class C.
- The probability density at C is then $e^{-\beta S_{\mathrm{YM}}(C)} \cdot \mathrm{size}(C)$.
- Physicists have an ingenious way of calculating $\mathrm{size}(C)$ using infinitesimal transforms (akin to measuring the length of a circle by measuring how much it curves locally).
- The resulting formula is a determinant, which is then expressed as a Gaussian integral involving Grassmann variables. These variables are called Faddeev–Popov ghosts.
- Such calculations have so far proved to be useful only for perturbative analysis.

Wilson loop variables

- Let γ be a piecewise smooth closed path γ in $\mathbb{R}^4$.
- Take any $A \in \mathcal{A}$. The Wilson loop observable W_γ is defined as

$$W_\gamma = \mathrm{Tr}\biggl(\mathcal{P} \exp\biggl(\int_\gamma \sum_{j=1}^4 A_j dx_j\biggr)\biggr),$$

  where $\mathcal{P}$ is the path-ordering operator. (Will explain soon.)
- In differential geometric terminology, the term inside the trace in the above display is the holonomy of A along the closed path γ.
- Alternatively, it is the parallel transport of the identity matrix along γ by the connection A.

Wilson loops, from discrete to continuous

- We will now see a relatively straightforward definition of W_γ, without the involvement of differential geometric language.
- Take a smooth G-connection form A.
- Take some small $\epsilon > 0$. Replace $\mathbb{R}^4$ by $\epsilon\mathbb{Z}^4$.
- Let e_1, e_2, e_3, e_4 denote the standard basis vectors of $\mathbb{R}^4$.
- To an edge $(x, x + \epsilon e_j)$ on $\epsilon\mathbb{Z}^4$, attach the unitary matrix $e^{\epsilon A_j(x)}$. To the reverse edge $(x + \epsilon e_j, x)$, attach the inverse matrix $e^{-\epsilon A_j(x)}$.
- Approximate γ by a closed path of directed edges in $\epsilon\mathbb{Z}^4$.
- Approximate W_γ by the trace of the product of the matrices attached to the edges of this path.
- Then, as $\epsilon \to 0$, these approximations tend to a limit, which is exactly our W_γ.

Quark confinement

- Recall that quarks are the constituents of certain massive particles, such as neutrons and protons.
- Free quarks are never observed in nature. This is known as the phenomenon of quark confinement. No convincing theoretical argument exists.
- Understanding quark confinement was Wilson's original motivation behind the definition and study of Wilson loops.

Quark confinement and Wilson loops

- Using physical arguments, Wilson showed that the potential energy $V(R)$ of a static quark-antiquark pair at distance R from each other is given by the formula

$$V(R) = -\lim_{T\to\infty} \frac{1}{T} \log\langle W_{\gamma_{R,T}}\rangle,$$

  where $\gamma_{R,T}$ is a rectangle of length T and breadth R, and $\langle\cdot\rangle$ denotes expectation in $SU(3)$ Euclidean Yang–Mills theory.
- If $V(R) \to \infty$ as $R \to \infty$, then conservation of energy implies that the quark-antiquark pair can never separate beyond a finite distance.
- Wilson conjectured that for $SU(3)$ theory,

$$|\langle W_{\gamma_{R,T}}\rangle| \le C_1 e^{-C_2 RT},$$

  where C_1 and C_2 do not depend on R and T.
- Clearly, this implies that $V(R) \ge C_2 R$, and therefore quarks confine.

Wilson's area law

- More generally, it is conjectured that for a large class of 4D non-Abelian gauge theories,

$$|\langle W_\gamma \rangle| \le C_1 e^{-C_2 \mathrm{area}(\gamma)}$$

  for any (nice) closed loop γ, where area(γ) is the minimal surface area enclosed by γ.
- This is still not proved, even in the physics sense.
- Of course, to prove this rigorously, one has to first define quantum Yang–Mills theories, which is itself an open problem.

Strong coupling and weak coupling

- The parameter β is the square-root of the inverse of a parameter called the 'coupling constant' in physics.
- Consequently, small β corresponds to 'strong coupling' and large β corresponds to 'weak coupling'.
- The terminology may be slightly confusing for probabilists, because correlations are more rapidly decaying in the strong coupling regime.
- The reason for the terminology is that in the strong coupling regime, the non-Abelian effects become more pronounced. As a consequence, the non-Abelian part becomes more 'strongly coupled' with the Abelian (Gaussian) part of the theory.
- What does this mean? Next slide.

The coupling constant

- Recall that probability density of the gauge field A is proportional to $e^{-\beta S_{\mathrm{YM}}(A)}$, where

$$S_{\mathrm{YM}}(A) = -\int_{\mathbb{R}^4} \sum_{j,k=0}^{3} \mathrm{Tr}\left[\left(\frac{\partial A_k}{\partial x_j} - \frac{\partial A_j}{\partial x_k} + [A_j(x), A_k(x)]\right)^2\right] dx.$$

- Rescale the field A as $B = \sqrt{\beta} A$, and let $g = \beta^{-1/2}$.
- Then it is easy to check that the probability density of the rescaled field B is proportional to $e^{-S'(B)}$, where

$$S'(B) = -\int_{\mathbb{R}^4} \sum_{j,k=0}^{3} \mathrm{Tr}\left[\left(\frac{\partial B_k}{\partial x_j} - \frac{\partial B_j}{\partial x_k} + g[B_j(x), B_k(x)]\right)^2\right] dx.$$

- The parameter g is called the coupling constant. It controls the non-Abelian effect through the term $g[B_j(x), B_k(x)]$. If this term is absent, then the theory is purely Gaussian (non-interacting, or trivial, as physicists would say).

Perturbative expansions

- The standard way for physicists to handle non-Abelian quantum Yang–Mills theories is by expanding perturbatively in powers of g.
- Since the term not involving g is purely Gaussian, one can (in principle) do the calculations for the perturbative expansions using Gaussian moment computations (Feynman diagrams).
- Quite complicated, but has been done. Involves renormalization — first regularizing the theory in some way and then adjusting the parameters so that the terms in the perturbative expansion have finite limits when the regularization is taken away.
- The perturbative expansions have even been made rigorous to a large extent (see monograph by Costello).
- However, nonperturbative analysis is required for many important questions (such as quark confinement).
- Hope: Nonperturbative analysis using lattice gauge theories.

Part II: Lattice gauge theories

Lattice gauge theories

- Lattice gauge theories are discrete versions of quantum Yang–Mills theories, proposed by Wilson (1974).
- We will now define lattice gauge theories in arbitrary dimension $d \geq 2$.
- As before, let G be a matrix Lie group contained in $U(N)$ for some N.
- Let $\Lambda \subseteq \mathbb{Z}^d$ be a finite set.

Definition of lattice gauge theories

- Suppose that for any two adjacent vertices $x, y \in \Lambda$, we have a group element $U(x, y) \in G$, with $U(y, x) = U(x, y)^{-1}$.
- Let $G(\Lambda)$ denote the set of all such configurations.
- A square bounded by four edges is called a plaquette. Let $P(\Lambda)$ denote the set of all plaquettes in Λ.
- For a plaquette $p \in P(\Lambda)$ with vertices x_1, x_2, x_3, x_4 in anti-clockwise order, and a configuration $U \in G(\Lambda)$, define

$$U_p := U(x_1, x_2)U(x_2, x_3)U(x_3, x_4)U(x_4, x_1).$$

- The Wilson action of U is defined as

$$S_{\mathrm{W}}(U) := \sum_{p \in P(\Lambda)} \mathrm{Re}(\mathrm{Tr}(I - U_p)).$$

- More generally, physicists substitute $\rho(U_p)$ in place of U_p, where ρ is a representation of G. We will not worry about that level of generality.

Definition of lattice gauge theory

- Let σ_Λ be the product Haar measure on $G(\Lambda)$.
- Given $\beta > 0$, let $\mu_{\Lambda,\beta}$ be the probability measure on $G(\Lambda)$ defined as

$$d\mu_{\Lambda,\beta}(U) := \frac{1}{Z} e^{-\beta S_{\mathrm{W}}(U)} d\sigma_\Lambda(U),$$

  where Z is the normalizing constant.
- This probability measure is called the lattice gauge theory on Λ for the gauge group G, with inverse coupling strength β.
- An infinite volume limit of the theory is a weak limit of the above probability measures as $\Lambda \uparrow \mathbb{Z}^d$.
- The infinite volume limit may or may not be unique.
- Open problem: The uniqueness (or non-uniqueness) is in general unknown for lattice gauge theories in dimension four when β is large.

An approximation for products of matrix exponentials

- Recall the Baker–Campbell–Hausdorff formula for products of matrix exponentials:
$$e^B e^C = e^{B+C+\frac{1}{2}[B,C]+\cdots}.$$
- Iterating this gives, for any m and any $B_1, \ldots, B_m$,
$$e^{B_1} \cdots e^{B_m} = e^{\sum_{a=1}^m B_a + \frac{1}{2}\sum_{1\le a<b\le m}[B_a,B_b]+\cdots}.$$
- Suppose that the B_i are skew-Hermitian. Then all terms within the exponential are skew-Hermitian and therefore have purely imaginary traces.
- Thus, if the entries of $B_1, \ldots, B_m$ are of order ϵ and if the entries of $B_1 + \cdots + B_m$ are of order ϵ^2, then
$$\begin{aligned}&\mathrm{Re}(\mathrm{Tr}(I - e^{B_1}\cdots e^{B_m}))\\ &= -\frac{1}{2}\mathrm{Tr}\biggl[\biggl(\sum_{a=1}^m B_a + \frac{1}{2}\sum_{1\le a<b\le m}[B_a, B_b]\biggr)^2\biggr] + O(\epsilon^5).\end{aligned}$$

From LGT to Euclidean YM theory: Wilson's heuristic

- Recall the discretization of the Brownian bridge that we discussed before.
- Just as in that problem, discretize $\mathbb{R}^d$ as $\epsilon\mathbb{Z}^d$ for some small ϵ.
- Take a smooth G-connection form A on $\mathbb{R}^d$. Recall that this is just a smooth map from $\mathbb{R}^d$ into $\mathfrak{g}^d$, where $\mathfrak{g}$ is the Lie algebra of G.
- Let $e_1, \ldots, e_d$ denote the standard basis vectors of $\mathbb{R}^d$.
- For a directed edge $(x, x + \epsilon e_j)$ of $\epsilon\mathbb{Z}^d$, define

$$U(x, x + \epsilon e_j) := e^{\epsilon A_j(x)},$$

  and let $U(x + \epsilon e_j, x) := U(x, x + \epsilon e_j)^{-1}$.
- This defines a configuration of unitary matrices assigned to directed edges of $\epsilon\mathbb{Z}^d$.

Wilson's heuristic, contd.

- Take any $x \in \epsilon\mathbb{Z}^d$ and any $1 \le j < k \le d$, and let

$$x_1 = x,\ x_2 = x + \epsilon e_j,\ x_3 = x + \epsilon e_j + \epsilon e_k,\ x_4 = x + \epsilon e_k.$$

- Let p be the plaquette formed by the vertices x_1, x_2, x_3, x_4.
- Then

$$\begin{aligned} U_p &= U(x_1, x_2)U(x_2, x_3)U(x_3, x_4)U(x_4, x_1) \\ &= e^{\epsilon A_j(x_1)} e^{\epsilon A_k(x_2)} e^{-\epsilon A_j(x_4)} e^{-\epsilon A_k(x_1)}. \end{aligned}$$

- Writing

$$A_k(x_2) = A_k(x + \epsilon e_j) = A_k(x) + \epsilon \frac{\partial A_k}{\partial x_j} + O(\epsilon^2)$$

and using a similar Taylor expansion for $A_j(x_4)$, we get

$$A_j(x_1) + A_k(x_2) - A_j(x_4) - A_k(x_1) = \epsilon\left(\frac{\partial A_k}{\partial x_j} - \frac{\partial A_j}{\partial x_k}\right) + O(\epsilon^2).$$

Wilson's heuristic, contd.

- Thus,

$$\begin{aligned}
&\mathrm{Re}(\mathrm{Tr}(I - U_p)) \\
&= -\frac{1}{2}\epsilon^4 \,\mathrm{Tr}\biggl[\biggl(\frac{\partial A_k}{\partial x_j} - \frac{\partial A_j}{\partial x_k} + [A_j(x), A_k(x)]\biggr)^2\biggr] + O(\epsilon^5) \\
&= -\frac{1}{2}\epsilon^4 \,\mathrm{Tr}(F_{jk}(x)^2) + O(\epsilon^5).
\end{aligned}$$

- This gives a heuristic justification for the formal approximation

$$S_{\mathrm{W}}(U) \approx -\frac{\epsilon^{4-d}}{4} S_{\mathrm{YM}}(A).$$

- The above heuristic was used by Wilson to justify the approximation of Euclidean Yang–Mills theory by lattice gauge theory, scaling the inverse coupling strength β like ϵ^{4-d} as the lattice spacing $\epsilon \to 0$.

Scaling in dimension four

- The most important dimension is $d = 4$, because spacetime is four-dimensional.
- In the above formulation, β does not scale with ϵ at all when $d = 4$.
- Currently, however, the general belief in the physics community is that β should scale like some multiple of $\log(1/\epsilon)$ in dimension four.

Other actions

- The Wilson action is not the only one that occurs in the literature.
- Any action that is gauge invariant (to be discussed later) and heuristically scales to the Yang–Mills action as the lattice spacing goes to zero, is acceptable.
- Notice that

$$S_W(U) = \prod_p f_\beta(U_p),$$

where $f_\beta(U_p) = e^{-\beta \mathrm{Re}(\mathrm{Tr}(I - U_p))}$ is a measure of how far U_p is from I, with rate of decay away from I controlled by β.
- An alternative is the heat kernel action, where f_β is replaced by ρ_t, where ρ_t is the probability density of Brownian motion on G at time t, started from I. Here $t = 1/\beta$.
- In the special case of $G = U(1)$, the heat kernel action has an explicit form and is known as the Villain action.

Wilson loops in lattice gauge theories

- Suppose that we have a lattice gauge theory on $\Lambda \subseteq \mathbb{Z}^d$ with gauge group G.
- Given a loop $\gamma \subseteq \Lambda$ with directed edges $e_1, \ldots, e_m$, the Wilson loop variable W_γ is defined as

$$W_\gamma := \mathrm{Tr}(U(e_1)U(e_2)\cdots U(e_m)).$$

- We have discussed earlier why this can be considered to be a discrete approximation of Wilson loops in continuum theories.

Area law in lattice gauge theories

- It is believed that for a large class of non-Abelian gauge groups (including $SU(3)$, in particular), 4D lattice gauge theories satisfy Wilson's area law.
- Namely, for any (nice) loop $\gamma \subseteq \Lambda$,

$$|\langle W_\gamma \rangle| \le C_1 e^{-C_2 \mathrm{area}(\gamma)},$$

  where $\mathrm{area}(\gamma)$ is the minimal surface area enclosed by γ and C_1, C_2 depend only on β and G.
- Unlike the continuum version, this is a well-posed mathematical problem since lattice gauge theories are well-defined objects.
- Would be a big breakthrough to prove this even for rectangles.

Area law at strong coupling (Osterwalder–Seiler)

- When β is small (strong coupling), the area law was rigorously proved for rectangular loops by Osterwalder and Seiler (1978). Any dimension, essentially any gauge group.
- However, this is not sufficient for quark confinement. To establish quark confinement, one has to at least show that the area law holds for all large enough β.
- (More accurately, one has to show that the area law should hold in the continuum limit, which involves taking $\beta \to \infty$.)
- Moreover, confinement is not supposed to hold in 4D $U(1)$ gauge theory, since an electromagnetic field does not confine matter.
- However, the Osterwalder–Seiler theorem implies that the area law holds in strongly coupled 4D U(1) lattice gauge theory.
- Thus, strongly coupled lattice gauge theories do not represent physical reality. Understanding the large β regime is essential for making physical conclusions.

Deconfinement transition (Guth–Fröhlich–Spencer)

- The area law at small β for 4D $U(1)$ lattice gauge theory was a vexing issue for some time, until Guth (1980) gave an argument showing that the area law does not hold at large β for 4D $U(1)$ lattice gauge theory.
- A fully rigorous proof was produced by Fröhlich and Spencer (1982) for 4D $U(1)$ lattice gauge theory with Villain action.
- They showed that in fact, when β is large, 4D $U(1)$ lattice gauge theory satisfies

$$\langle W_\gamma \rangle \ge C_1 e^{-C_2 \text{perimeter}(\gamma)}.$$

- This implies that the quark-antiquark potential $V(R)$ remains bounded as $R \to \infty$. Hence, quarks need not be confined.
- The Guth–Fröhlich–Spencer theorem is known as the deconfinement transition of 4D $U(1)$ lattice gauge theory.

The Göpfert–Mack theorem

- Is 4D really important for understanding confinement?
- Indeed, it is. Göpfert and Mack (1982) showed that 3D $U(1)$ lattice gauge theory with Villain action satisfies the area law at all β.
- In other words, $U(1)$ theory has no deconfinement transition in 3D.

Why Villain action?

- The proofs of Fröhlich–Spencer, as well as Göpfert–Mack and many others, use the Villain action instead of the Wilson action for mathematical convenience.
- This is because $U(1)$ theory with Villain action is dual to certain other theories (akin to Kramers–Wannier duality for the Ising model) that convert large β problems to small β problems using the character expansion for Abelian groups.
- This trick, unfortunately, is only available for Abelian theories. Non-Abelian duality has not proved to be mathematically tractable till now.
- Since the most important open problems are all about non-Abelian theories, we will not go into this any further.

Gauge transforms for lattice gauge theories

- Gauge transformations are particularly simple to describe for lattice gauge theories.
- Suppose that we have lattice gauge theory on $\Lambda \subseteq \mathbb{Z}^d$ and a gauge configuration $U \in G(\Lambda)$.
- A gauge transform is induced by a map h from the vertex set of Λ into G.
- The transform of U by h is defined as

$$U'(x, y) := h(y)U(x, y)h(x)^{-1}.$$

- The configurations U and U' are called gauge-equivalent. Clearly, this is an equivalence relation. The correspondence with the continuum version can be checked in the usual way.
- A function f is called gauge-invariant if $f(U) = f(U')$ whenever U and U' are gauge-equivalent.
- It is easy to see that the Wilson action and the heat kernel action are gauge-invariant. Wilson loop variables are also gauge-invariant.

Gauge fixing

- Due to the finiteness of the problem, it is much easier to fix the gauge for lattice gauge theories.
- In fact, there is a general prescription:
  - Take any rooted spanning tree T of Λ and a gauge configuration U.
  - If r is the root, define $h(r) := I$.
  - If y is node of T and x is its parent, inductively define $h(y)$ such that $h(y)U(x,y)h(x)^{-1} = I$.
  - Let U' be the gauge transform of U by h, so that $U'(x,y) = I$ for all edges (x,y) of T.
- It is not difficult to show that for any given T, the process of obtaining U' from U is a procedure for choosing one member from each gauge equivalence class.
- In other words, if U_1 and U_2 are gauge equivalent, they will lead to the same U'.
- It is also easy to show that all gauge equivalence classes have exactly the same size. No Faddeev–Popov ghosts.

2D theories

- Suppose that Λ is a rectangle in $\mathbb{Z}^2$.
- Let T be the spanning tree consisting of all horizontal edges and the vertical edges on the left boundary. Let the bottom-left vertex be the root.
- Gauge fixing makes the matrices attached to all edges in T equal to I.
- Thus, a gauge equivalence class can be represented as a configuration U where $U(x, y) = I$ whenever (x, y) is an edge of T.
- Since all equivalence classes have the same size, the probability density on this reduced space of configurations is the same as the original one.
- Thus, the expected value of a gauge-invariant observable, such as a Wilson loop variable, is the same in the gauge fixed model as in the original one.

2D theories, contd.

- In the gauge-fixed configurations, only vertical edges (except the left-most ones) are not equal to I.
- These edges can be classified according to height. Let E_n be those at height n. Enumerate E_n from left to right as $\{e_{n,1}, e_{n,2}, \ldots\}$.
- Then the probability density is proportional to

$$\prod_n \prod_i \exp\biggl(-\beta \mathrm{Re}(\mathrm{Tr}(I - U(e_{n,i+1})U(e_{n,i})^{-1}))\biggr).$$

- This shows that for any n, $U(e_{n,1}), U(e_{n,2}), \ldots$ is a random walk with i.i.d. increments on G with a kernel depending on β, and these random walks are independent of each other.
- This observation makes 2D theories amenable to probabilistic analysis, such as for proving the area law, passing to the continuum limit, and computing Wilson loop expectations (Gross, Driver, Sengupta, Lévy, Hall, Kemp, Norris, ...).

2D theories, contd.

- The random walk connection makes it possible to use stochastic calculus to study the continuum limits in 2D.
- Things can get quite interesting if one considers 2D theory on closed surfaces instead of $\mathbb{R}^2$.
- Unfortunately, the 2D methods do not extend to three and higher dimensions.
- In particular, the decomposition into random walks does not work in higher dimensions.

So, what to do in higher dimensions?

- Largely open, other than the results mentioned earlier (Osterwalder–Seiler, Guth–Fröhlich–Spencer, Göpfert–Mack).
- Main problems:
  - Prove area law at large β.
  - Prove decay of correlations at large β. Exponential decay would be ideal, but no decay has been proved.
  - Construct the continuum limit. Monumental work of Bałaban and others on this topic. Based on the framework of constructive quantum field theory. However, the question is not yet settled. In particular, it is not clear what the limit object is, nor how one can calculate Wilson loop expectations in the limit. For more details, see my survey 'Yang–Mills for probabilists' on arXiv.
- There is, however, a certain set of recursive equations for Wilson loops that hold in any dimension. This is our next topic.

Part III: The master loop equations and gauge-string duality

The master loop equations

- The master loop equations for lattice gauge theories, also known as the Makeenko–Migdal equations, were discovered by Makeenko & Migdal (1979).
- The equations represent the expectation of a Wilson loop variable as a polynomial in several other Wilson loop variables (in any dimension).
- The derivation was not rigorous. In particular, it involved an unproved 'factorization ansatz' that assumed that the expectation of a product of Wilson loop variables approximately factorize into a product of expectations.
- The equations were reduced to a set of equations involving finitely many edges by Eguchi & Kawai (1982).
- Possibly due to problems with the original Makeenko–Migdal equations, problems arose in the Eguchi–Kawai reduction. Currently, physicists believe in a modified system, known as the twisted Eguchi–Kawai equations.

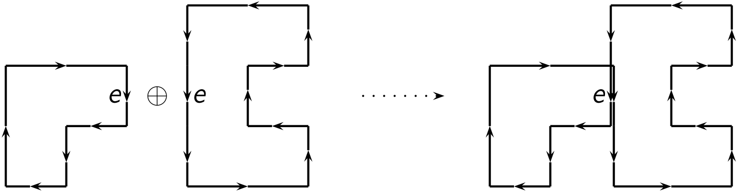
Rigorous derivation of the master loop equations

- In 2D, master loop equations can be formulated and rigorously proved in the continuum setting. This was done by Lévy (2011) and more recently by Driver, Hall and Kemp (2016).
- In three and higher dimensions, the master loop equations for $SO(N)$ lattice gauge theory were rigorously derived for the first time in C. (2015). These equations will be described in the next few slides.
- In the rigorous version derived in C. (2015), the basic objects are not Wilson loop expectations, but expectations of products of Wilson loop variables.
- The factorization ansatz is not assumed. The equations hold for any finite N.

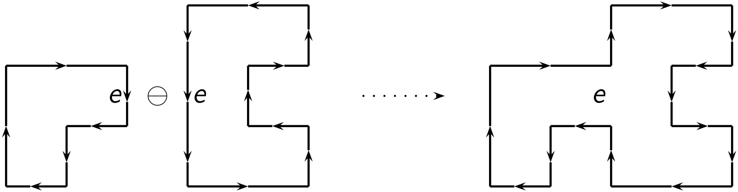
A string theory on the lattice

- The statement of the master loop equations from C. (2015) requires some preparation.
- The first step is the definition of a string theory on $\mathbb{Z}^d$.
- In physics, a string theory prescribes an action for the trajectory of a collection of closed loops (instead of a collection of particles) evolving in time. We will do something analogous on the lattice.
- For us, a string is a finite collection of closed loops in $\mathbb{Z}^d$.
- A string evolves over time.
- At each time step, a component loop may become slightly deformed, or two loops may merge, or a loop may split into two, or a loop may twist at a bottleneck, or nothing may happen.
- Each operation has two subtypes. Total of eight possible operations (nine, if we count doing nothing as an operation).
- These operations are described in the following slides.

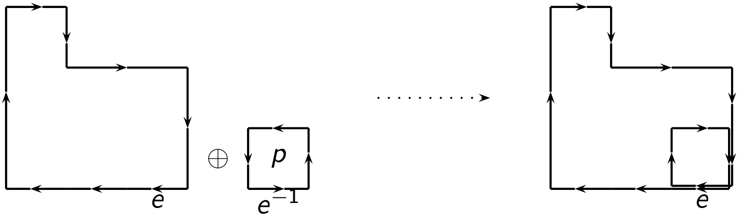
Positive merger



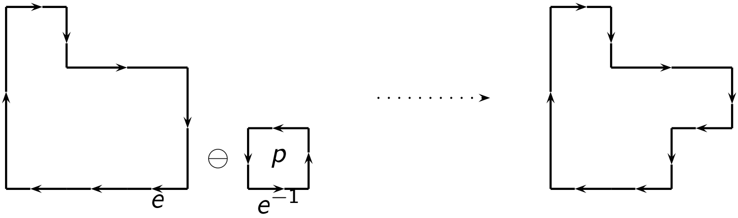
Negative merger



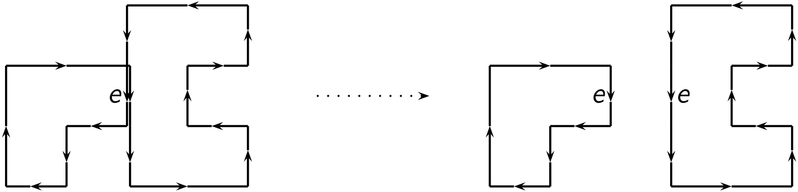
Positive deformation



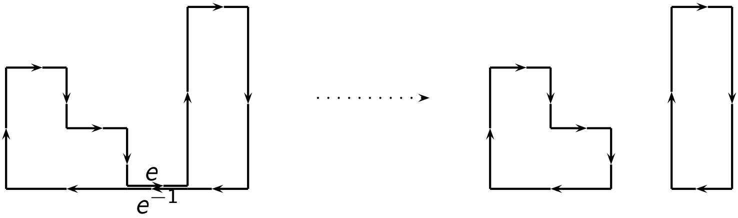
Negative deformation



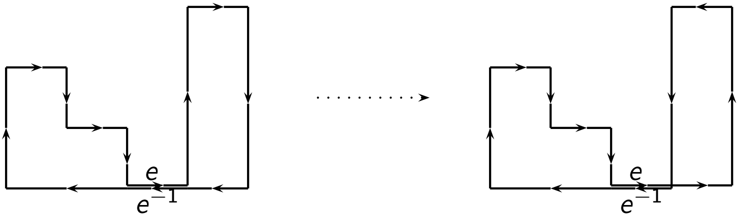
Positive splitting



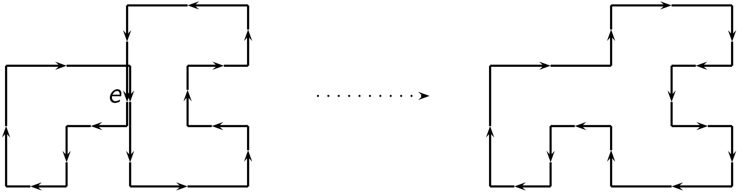
Negative splitting



Positive twisting



Negative twisting



Operations on lattice strings

- After performing one of the eight operations, we get a new string s', which may contain a different number of loops.
- Let

$$
\begin{aligned}
\mathbb{D}^+(s) &:= \{s' : \; s' \text{ is a positive deformation of } s\},\\
\mathbb{D}^-(s) &:= \{s' : \; s' \text{ is a negative deformation of } s\},\\
\mathbb{S}^+(s) &:= \{s' : \; s' \text{ is a positive splitting of } s\},\\
\mathbb{S}^-(s) &:= \{s' : \; s' \text{ is a negative splitting of } s\},\\
\mathbb{M}^+(s) &:= \{s' : \; s' \text{ is a positive merger of } s\},\\
\mathbb{M}^-(s) &:= \{s' : \; s' \text{ is a negative merger of } s\},\\
\mathbb{T}^+(s) &:= \{s' : \; s' \text{ is a positive twisting of } s\},\\
\mathbb{T}^-(s) &:= \{s' : \; s' \text{ is a negative twisting of } s\}.
\end{aligned}
$$

- What is the action in our string theory? We will talk about it later.

The master loop equation

Theorem (C., 2015)

Consider $SO(N)$ lattice gauge theory on $\mathbb{Z}^d$, for any $N \geq 2$ and $d \geq 2$, and any β. For a string $s = (\ell_1, \ell_2, \ldots, \ell_n)$, define

$$\phi(s) := \frac{\langle W_{\ell_1} W_{\ell_2} \cdots W_{\ell_n} \rangle}{N^n}.$$

Let $|s|$ be the total number of edges in s. Then

$$\begin{aligned}(N-1)|s|\phi(s) &= \sum_{s' \in \mathbb{T}^-(s)} \phi(s') - \sum_{s' \in \mathbb{T}^+(s)} \phi(s') + N \sum_{s' \in \mathbb{S}^-(s)} \phi(s') \\ &\quad - N \sum_{s' \in \mathbb{S}^+(s)} \phi(s') + \frac{1}{N} \sum_{s' \in \mathbb{M}^-(s)} \phi(s') - \frac{1}{N} \sum_{s' \in \mathbb{M}^+(s)} \phi(s') \\ &\quad + \beta \sum_{s' \in \mathbb{D}^-(s)} \phi(s') - \beta \sum_{s' \in \mathbb{D}^+(s)} \phi(s').\end{aligned}$$

First step in the proof: Stein equation for $SO(N)$

- In the probability literature, the Stein identity for a standard Gaussian variable Z says that $\mathbb{E}(Zf(Z)) = \mathbb{E}(f'(Z))$ for all f.
- This is just integration by parts. Extensions of it are called 'Ward identities' or 'Schwinger–Dyson equations' in physics.
- The following theorem generalizes this to $SO(N)$. Let $\mathbb{R}^{N\times N}$ denote the space of all $N \times N$ real matrices.

Theorem (C., 2015)

Let f and g be C^2 functions in a neighborhood of $SO(N)$ in $\mathbb{R}^{N\times N}$, and let $\mathbb{E}(\cdot)$ denote expectation with respect to the Haar measure. Then

$$\mathbb{E}\biggl(\sum_{i,k} x_{ik}\frac{\partial f}{\partial x_{ik}} g\biggr) = \frac{1}{N-1}\mathbb{E}\biggl(\sum_{i,k}\frac{\partial^2 f}{\partial x_{ik}^2} g - \sum_{i,j,k,k'} x_{jk}x_{ik'}\frac{\partial^2 f}{\partial x_{ik}\partial x_{jk'}} g + \sum_{i,k}\frac{\partial f}{\partial x_{ik}}\frac{\partial g}{\partial x_{ik}} - \sum_{i,j,k,k'} x_{jk}x_{ik'}\frac{\partial f}{\partial x_{ik}}\frac{\partial g}{\partial x_{jk'}}\biggr).$$

More about the Stein equation

- The Stein equation displayed in the previous slide can be written much more compactly as the integration-by-parts identity

$$\int_{SO(N)} (g\Delta f + \langle \nabla g, \nabla f\rangle) d\eta = 0,$$

  where Δ and ∇ are the appropriately defined Laplacian and gradient operators on the manifold $SO(N)$, and η is the normalized Haar measure.
- The previously displayed version is the above equation written in extrinsic coordinates, considering $SO(N)$ as a submanifold of $\mathbb{R}^{N\times N}$.
- This was pointed out to me by Thierry Lévy. The proof in the paper uses a different idea.
- The version in extrinsic coordinates is, however, quite important for subsequent calculations.

Second step in the proof: A property of Wilson loop variables

- Let ℓ be a loop and e be an edge of ℓ.
- Denote the $(i,j)^{\mathrm{th}}$ element of $U(e)$ by q_{ij}.
- Let m is the number of occurrences of e and e^{-1} in ℓ.
- Then, it is not hard to show that

$$mW_\ell = \sum_{i,j} q_{ij} \frac{\partial W_\ell}{\partial q_{ij}}.$$

- This holds because W_ℓ is a homogeneous polynomial of degree m in the q_{ij}'s, if all other matrices are held constant. Whenever $p(x_1, \ldots, x_k)$ is a homogeneous polynomial of degree m in variables $x_1, \ldots, x_k$, we have

$$\sum_{i=1}^{k} x_i \frac{\partial p}{\partial x_i} = mp(x_1, \ldots, x_k).$$

Completing the proof

- Let $s = (\ell_1, \ldots, \ell_n)$ be a string.
- Let e be an edge of ℓ_1. Let m be the number of occurrences of e and e^{-1} in ℓ_1. Denote the $(i,j)^{\mathrm{th}}$ element of $U(e)$ by q_{ij}.
- Let g be the product of $W_{\ell_2} W_{\ell_3} \cdots W_{\ell_n}$ with the probability density of $SO(N)$ lattice gauge theory.
- If $\langle \cdot \rangle$ is expectation in the lattice gauge theory and $\mathbb{E}(\cdot)$ is expectation with respect to Haar measure, then by the previous slide,

$$m\langle W_{\ell_1} W_{\ell_2} \cdots W_{\ell_n} \rangle = m\mathbb{E}(W_{\ell_1} g) = \mathbb{E}\biggl(\sum_{i,j} q_{ij} \frac{\partial W_{\ell_1}}{\partial q_{ij}} g\biggr).$$

- One can now apply the Stein equation to the right side.
- Somewhat miraculously, the result is the expected value of a sum of products of Wilson loops (in LGT).
- There is nothing special about ℓ_1 and e. The master loop equation is obtained by averaging over all loops and edges.

Remarks

- The eight string operations arise naturally from the above calculation.
- There seems to be no particular reason why the integration by parts should yield an expression involving expected values of products of Wilson loops. It is important to have an intuitive explanation for this computational result.
- My former student Jafar Jafarov worked out the master loop equations for $SU(N)$ theory along the same lines. Interestingly, the equations and the string operations are somewhat different for $SU(N)$.
- Open problem: Is there a general result that works for other Lie groups?

The large N limit

- An important direction of research, pioneered by 't Hooft (1974), is to look at very high dimensional gauge groups, such as $SO(N)$ or $SU(N)$ with $N \to \infty$.
- This is known as large N gauge theory.
- The reason for taking $N \to \infty$ is that it simplifies many calculations.
- 't Hooft's work also pioneered the theory of planar diagrams and their connections with random matrices.

Planar diagrams

- Recall the perturbative calculations in quantum Yang–Mills theories, described in the first lecture.
- It was mentioned that the calculations reduce to Gaussian moment computations, which can be handled using Feynman diagrams.
- 't Hooft realized that a slight modification of the Feynman diagram method converts the diagrams into surfaces with edges, vertices and faces.
- Then, 't Hooft suggested that the parameter β should be replaced by $N\beta$. This is known as the 't Hooft scaling.
- This reparametrization has the interesting effect that when N is taken to ∞, the perturbative series expansion becomes a series in powers of $1/N$. This is known as the $1/N$ expansion.
- The interesting outcome of the above exercise is that the k^{th} term in the expansion is a sum over surfaces of genus k. This is the beginning of planar diagram theory.

Planar diagrams in mathematics

- 't Hooft's planar diagram theory has inspired a large amount of rigorous mathematics. Examples include the works of Eynard and Orantin, as well a large body of random matrix literature produced by Guionnet and collaborators.

Wilson loop expectations in the large N limit

- 't Hooft's theory is perturbative. Let us look at what happens in the non-perturbative setting.
- Recall that a string s is a collection of loops $(\ell_1, \ldots, \ell_n)$. Let

$$\phi_N(s) = \frac{\langle W_{\ell_1} W_{\ell_2} \cdots W_{\ell_n} \rangle}{N^n}.$$

- Here the expectation is taken with respect to $SO(N)$ lattice gauge theory on $\mathbb{Z}^d$, by taking a weak limit of theories on finite cubes.
- Suppose that like 't Hooft, we replace β by $N\beta$, and send $N \to \infty$.
- Does $\phi_N(s)$ converge to a limit? If so, can we calculate it?
- Observe that since $|\phi_N(s)| \le 1$, and there are countably many strings, one can always produce a subsequential limit simultaneously for all strings by a diagonal argument.

Master loop equation in the large N limit

- Let $\phi(s)$ be a subsequential limit of $\phi_N(s)$ (simultaneously for all s, under the 't Hooft scaling).
- Then ϕ satisfies the limiting master loop equation

$$|s|\phi(s) = \sum_{s' \in \mathbb{S}^-(s)} \phi(s') - \sum_{s' \in \mathbb{S}^+(s)} \phi(s') + \beta \sum_{s' \in \mathbb{D}^-(s)} \phi(s') - \beta \sum_{s' \in \mathbb{D}^+(s)} \phi(s').$$

- Note that only the splitting and deformation terms have survived, and the twists and mergers have disappeared in the large N limit.
- As before, note that this holds for any β.
- If this system of equations has a unique solution, then we may conclude that $\phi_N(s) \to \phi(s)$ as $N \to \infty$.

Existence of the large N limit at strong coupling

- It was shown in C. (2015) that the system of limiting master loop equations indeed has a unique solution when β is sufficiently small.
- Therefore, $\phi_N(s)$ indeed converges to a limit for every s when β is small.
- Open problem: Show that the limit exists for any β.

Sketch of proof

- Let ϕ and ψ be two solutions of the limiting master loop equations.
- Let Δ be the set of all finite sequences of integers.
- If $\delta, \delta' \in \Delta$, we will say that $\delta \le \delta'$ if the two sequences have the same length and δ' dominates δ in each component.
- For a string $s = (\ell_1, \ell_2, \ldots, \ell_n)$, let $\delta(s) = (|\ell_1|, |\ell_2|, \ldots, |\ell_n|)$.
- For $\delta \in \Delta$, let

$$D(\delta) := \sup_{s : \delta(s) \le \delta} |\phi(s) - \psi(s)|.$$

- For $\delta = (\delta_1, \ldots, \delta_n) \in \Delta$, let

$$\iota(\delta) := \delta_1 + \cdots + \delta_n - n.$$

- For $\lambda \in (0, 1)$, let

$$F(\lambda) := \sum_{\delta \in \Delta} \lambda^{\iota(\delta)} D(\delta).$$

Proof sketch contd.

- First, it is shown that $F(\lambda) < \infty$ if λ is sufficiently small.
- Next, repeated applications of the master loop equations produce the inequality

$$F(\lambda) \le \left(4\lambda^3 + 4\lambda + \frac{4\beta d}{\lambda^4} + \frac{4\beta d}{1-\lambda}\right) F(\lambda).$$

- This shows that if λ is small and β is small enough (depending on λ and d), then the coefficient of $F(\lambda)$ on the right is less than 1.
- Due to the finiteness of $F(\lambda)$, this implies that $F(\lambda) = 0$.
- Consequently, $D(\delta) = 0$ for all δ, and hence $\phi = \psi$.

Quantum gravity

- There is no quantum theory of gravity. Finding such a theory is the holy grail of modern physics.
- Quantum effects are important at very small distances, where gravity is unimportant.
- Gravity is felt at large distances, between very massive bodies, where quantum effects are negligible.
- So, the unification seems like only an academic goal, except that...
- Both gravity and quantum effects manifest themselves in black holes. Small black holes can potentially form when particles collide with each other at very high speeds, as in particle accelerators.

More about quantum gravity

- In classical physics, particles move along deterministic trajectories.
- In quantum physics, the trajectories are probabilistic (although the notion of probability is replaced by complex probability amplitudes).
- General relativity is a classical theory, in the sense that the structure of curved spacetime is deterministic.
- A quantum theory of gravity would replace the fixed spacetime by a randomly fluctuating spacetime (random Riemannian manifold).
- The most promising approach: String theory.
- Roughly speaking, strings moving randomly trace out random surfaces. Higher dimensional strings, known as branes, trace out higher dimensional random manifolds.

Gauge-string duality

- There is a theoretical result of Weinberg and Witten (1980) that it is impossible to generate gravity using quantum field theories in the traditional sense (hard to explain without going into details).
- However, Weinberg and Witten have the unstated assumption that one is looking for both theories in the same dimension.
- In 1997, Maldacena made the remarkable discovery that certain quantum field theories are 'dual' to certain string theories in *one dimension higher!*
- Duality means that any calculation in one theory corresponds to some calculation in the other theory.
- Maldacena's discovery is known as AdS-CFT duality or gauge-string duality. The principle of going to one dimension higher is known as the holographic principle.
- Elements of gauge-string duality were already present in 't Hooft's planar diagram theory.

Gauge-string duality in $SO(N)$ lattice gauge theory

- A duality between $SO(N)$ lattice gauge theory with large N, and the lattice string theory described earlier, was discovered in C. (2015). This may be the first rigorously derived gauge-string duality result. This result will be described in the next few slides.
- In our lattice string theory, call the trajectory of a string vanishing if it vanishes in the finite number of steps.
- Let $X = (s_0, s_1, \ldots, s_k)$ be a vanishing trajectory.
- The action of X is defined as

$$w(X) := w(s_0, s_1)w(s_1, s_2) \cdots w(s_{k-1}, s_k),$$

where $w(s_i, s_{i+1})$ is defined as follows.

Weight of a transition

- Let s be a string and $|s|$ be the total number of edges in s.
- If s evolves into s' after an operation, define the weight of the transition from s to s' as

$$w(s,s') := \begin{cases} 1 & \text{if } s' = s, \\ -1/|s| & \text{if } s' \in \mathbb{T}^+(s) \cup \mathbb{S}^+(s) \cup \mathbb{M}^+(s), \\ 1/|s| & \text{if } s' \in \mathbb{T}^-(s) \cup \mathbb{S}^-(s) \cup \mathbb{M}^-(s), \\ -\beta/|s| & \text{if } s' \in \mathbb{D}^+(s), \\ \beta/|s| & \text{if } s' \in \mathbb{D}^-(s). \end{cases}$$

- As mentioned in the previous slide, the action of a vanishing trajectory is computed by multiplying the weights of the successive transitions.

The duality result

Theorem (C., 2015)

Consider $SO(N)$ lattice gauge theory on $\mathbb{Z}^d$ with coupling parameter $N\beta$. Let s be a string and let

$$\phi_N(s) = \frac{\langle W_{\ell_1} W_{\ell_2} \cdots W_{\ell_n} \rangle}{N^n},$$

as before. Then, if β is small enough, we have

$$\lim_{N\to\infty} \phi_N(s) = \sum_{X \in \mathcal{X}(s)} w(X),$$

where $\mathcal{X}(s)$ is the set of all vanishing trajectories starting at s whose steps consist of only splitting and deformations, and $w(X)$ denotes the action of a trajectory X, as defined earlier. Moreover, the infinite sum on the right is absolutely convergent.

Remarks

- The condition that trajectories can only have deformations and splittings, and no mergers, ensure that the surfaces traced out by the trajectories have genus zero.
- In later work with Jafar Jafarov, we computed the full $1/N$ expansion. The higher order terms involves sums over trajectories that trace out surfaces of high genus.
- Although the formulas look similar to 't Hooft's sums over planar diagrams, they are actually quite different. 't Hooft's analysis and the subsequent body of work look at perturbative expansions in the continuum theory, whereas this result is a rigorous non-perturbative theorem for lattice gauge theories.

Proof sketch

- For a string $s = (\ell_1, \ldots, \ell_n)$, define its 'index'

$$\iota(s) := |\ell_1| + \cdots + |\ell_n| - n\,.$$

- We prove facts about strings by induction on the index.
- The key results that help us carry out the inductions is that $\iota(s) > 0$ for any nonempty s, and if s' is a string that is produced by splitting s, then $\iota(s') < \iota(s)$.

Proof sketch, contd.

- The next step is to define a collection of coefficients $a_k(s)$, one for each nonnegative integer k and string s, using a certain inductive definition (by induction on k and $\iota(s)$, as described above) that guarantees the following two properties:
  - For β sufficiently small, the power series

$$\phi(s) := \sum_{k=0}^{\infty} a_k(s)\beta^k$$

    converges absolutely for any s.
  - The function ϕ satisfies the limiting master loop equations.
- These two properties and the uniqueness of the solution of the master loop equation imply that $\phi_N \to \phi$.

Proof sketch, contd.

- For each $k \geq 0$ and string s, let $\mathcal{X}_k(s)$ be the set of all vanishing trajectories with only deformations and splittings that start at s and have exactly k deformations.
- We show by induction that for any k and s,

$$a_k(s)\beta^k = \sum_{X \in \mathcal{X}_k(s)} w(X).$$

- Once we have this, it follows that

$$\phi(s) = \sum_{k=0}^{\infty} \sum_{X \in \mathcal{X}_k(s)} w(X).$$

- The absolute convergence of the right side requires a different argument that I will omit here.

Open problems

- Develop a continuum version of these results.
- At least, develop a version for large β.
- There are results of Basu & Ganguly (2016) that allow exact computations of the sum over trajectories in 2D using combinatorial connections and free probability theory. Extension to higher dimensions?
- As mentioned before, even establishing the existence of the limit for large β would be a significant achievement.

And finally, let me conclude with the hope that someone in this audience will solve one of the many important open problems posed in these lectures. This is important not only for personal fame and glory, but also for actually understanding the universe we live in.