Integrability From Four Dimensions

Edward Witten

Baxter Celebration - September 9, 2025

I am going to explain today a surprising perspective – due originally to Kevin Costello, and further extended by Costello, Masahito Yamazaki, and me – on the classical Yang-Baxter equation. By now, this work is not new, but the perspective may be unfamiliar to many.

Some references:

K. Costello, "Supersymmetric Gauge Theory and the Yangian," arXiv:1303.2632

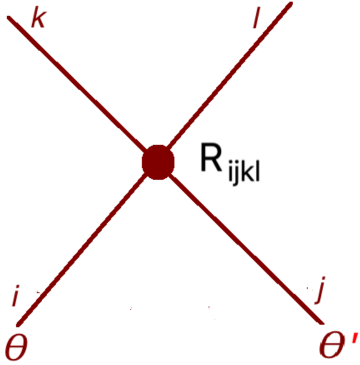
E. Witten, "Integrable Lattice Models from Gauge Theory," arXiv:1611.00592

K. Costello, E. Witten and M. Yamazaki, "Gauge Theory and Integrability, I," arXiv:1709.09993

K. Costello and D. Gaiotto, "Q Operators are 't Hooft Lines," arXiv:2103.01835.

Because time is limited, I will be very schematic in explaining what the Yang-Baxter equation is and why it is important, assuming that most of you are already familiar with the Yang-Baxter equation.

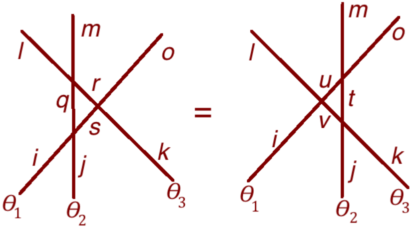
The basic picture is this one:



Here in Baxter's interpretation of the equation, i, j, k, l label the possible states of a classical spin (in a model of classical statistical mechanics in two dimensions). The spins live on lines, as shown. Each line has a "spectral parameter" θ or θ', and where two lines cross, the statistical sum gets a factor R_{ijkl}, which depends on the difference $\theta - \theta'$.

(In the quantum many-body interpretation of the Yang-Baxter equation, developed by Yang, Faddeev, A. and Al. Zamolodchikov, and many others, the labels i, j, k, l represent particle types, θ is the momentum or "rapidity" of a particle, and $R_{ijkl}(\theta - \theta')$ is an S-matrix element.)

The Yang-Baxter equation is a cubic equation for $R_{ijkl}(\theta)$, very schematically

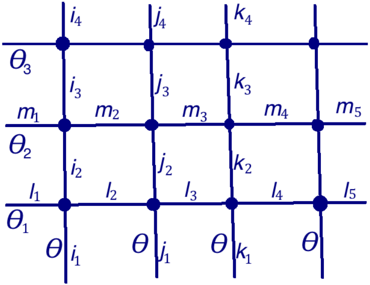


or in formulas

$$R_{23}R_{13}R_{12} = R_{12}R_{13}R_{23}.$$

In the picture, it is understood that one sums over the intermediate spins or particle types (that is, over the labels q, r, s on the left or u, t, v on the right) and in a crossing of two lines $a, b = 1, 2, 3$ labeled by rapidities θ_a, θ_b, one inserts an appropriate factor $R_{ab}(i, j, k, l; \theta_a - \theta_b)$.

For every solution of Yang-Baxter, one can construct an integrable spin system, described by this picture:

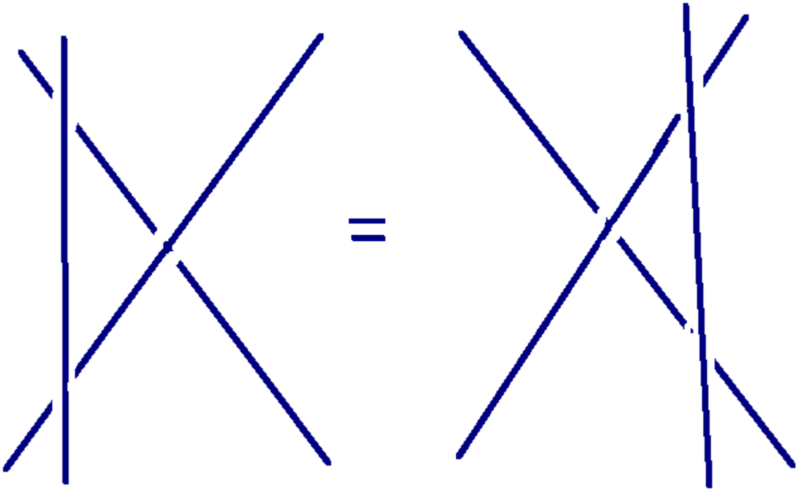


Here the vertical and horizontal lines are labeled by rapidities θ or θ_i, a line segment is labeled by a spin state $i, j, k, \ldots$, and a crossing is labeled by the appropriate R-matrix element.

To get the partition function, one sums over all the labels $i, j, k, \ldots$, weighting each choice by the product of the corresponding R-matrix elements, one factor at each vertex.

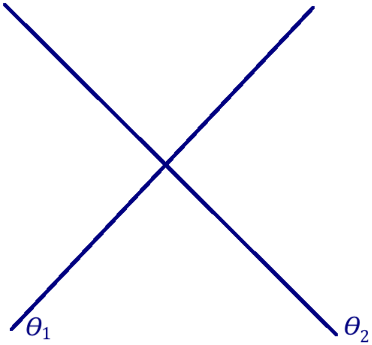
Perhaps the most obvious question about the Yang-Baxter equation is "why" solutions of this highly overdetermined equation exist. My goal today is to explain a perspective on this question. As a clue, the usual solutions of Yang-Baxter are determined by the choice of a simple Lie group G and an irreducible representation ρ. Why does that data lead to a solution of Yang-Baxter?

There is another area in which one finds something a lot like the Yang-Baxter equations. This is the theory of knots in three dimensions. Here is one of the Reidemeister moves:

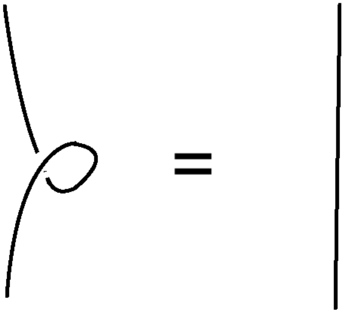


The resemblance to the Yang-Baxter equation is obvious, but there are also conspicuous differences:

(1) In knot theory, one strand passes “over” or “under” the other, while Yang-Baxter theory is a purely two-dimensional theory in which lines simply cross, with no “over” or “under”:



(2) In knot theory, there is another Reidemeister move that has no analog for Yang-Baxter:



These two points refer to structure that is present in knot theory and not in Yang-Baxter theory. But there is also an important difference in the opposite direction:

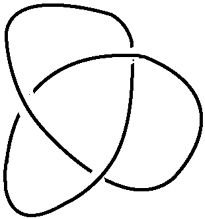
(3) In Yang-Baxter theory, the spectral parameter is crucial, but it has no analog in knot theory.

Despite these differences, there is an obvious analogy between the Yang-Baxter equation and the first Reidemeister move of knot theory, so let us pursue this a little bit. The usual solutions of Yang-Baxter depend, as I've said, on the choice of a simple Lie group G and an irreducible representation ρ. There are knot invariants that depend on the same data. To define them at least formally, consider a three-dimensional gauge theory with gauge group G. In mathematical language, let M be a three-manifold, $E \to M$ a G-bundle, and A a connection on G. Then one has the *Chern-Simons function*

$$\mathrm{CS}(A) = \frac{k}{4\pi} \int_M \mathrm{Tr} \left(A \mathrm{d} A + \frac{2}{3} A \wedge A \wedge A \right),$$

where k is a parameter that must be an integer to make the Feynman path integral well-defined.

A quantum field theory with this action is a "topological quantum field theory," since there is no metric tensor in sight. Let us just take the three-manifold M to be $\mathbb{R}^3$, and let $K \subset \mathbb{R}^3$ be an embedded knot.



We pick an irreducible representation ρ of G, and define

$$W_\rho(K) = \mathrm{Tr}_\rho P \exp\left(\oint_K A\right)$$

i.e. the trace, in the representation ρ, of the holonomy of A around K. This is the "Wilson loop operator."

Physically, it is part of the amplitude for a charged particle in the representation ρ, coupling to the gauge field A, to propagate around the path K in spacetime.

The usual "quantum knot invariants" can be defined via the expectation value of the Wilson operator, $\langle W_\rho(K)\rangle = \langle \mathrm{Tr}_\rho P \exp(\oint_K A)\rangle$. For $G = SU(2)$ and ρ the two-dimensional representation of $SU(2)$, the invariant we get this way is the Jones polynomial, the prototype of the quantum invariants of knots.

From these quantum invariants, one cannot really extract the usual solutions of the Yang-Baxter equation since one is missing the spectral parameter.

How can we modify or generalize Chern-Simons gauge theory to include the spectral parameter? A naive idea is to replace the finite-dimensional gauge group G with its loop group $\mathcal{L}G$. The loop group is the group of G-valued functions $g(\theta)$ of an angular variable θ. They are multiplied in the obvious way by pointwise multiplication. It is important that here we take the loop group itself, and not its central extension, which one often encounters in two-dimensional quantum field theory, string theory, and statistical mechanics. The central extension would force us to construct infinite-dimensional representations, but the loop group itself has some very simple representations: the "evaluation" representations that "live" at a particular value $\theta = \theta_0$ along the loop. In such a representation, a loop $g(\theta)$ just acts according to its value $g(\theta_0)$ at $\theta = \theta_0$. We hope that θ_0 will be the spectral parameter label carried by a particle in the solution of the Yang-Baxter equation.

Taking the gauge group to be a loop group means that the gauge field $A = \sum_i A_i(x)\mathrm{d}x^i$ now depends also on θ and so is $A = \sum_i A_i(x, \theta)\mathrm{d}x^i$. Note that there is no $\mathrm{d}\theta$ term so this is not a full four-dimensional gauge field. The Chern-Simons action has a generalization to this situation:

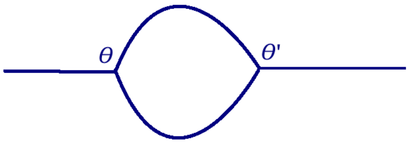
$$I = \frac{k}{4\pi} \int_{M \times S^1} \mathrm{d}\theta \, \mathrm{Tr} \left(A\mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right).$$

This is perfectly gauge-invariant.

What goes wrong is that because there is no $\partial/\partial\theta$ in the action, the "kinetic energy" of A is not elliptic and the perturbative expansion is not well-behaved. The propagator is

$$\langle A_i(\vec{x},\theta)A_j(\vec{x}',\theta') = \frac{\epsilon_{ijk}(x-x')^k}{|\vec{x}-\vec{x}'|^2}\delta(\theta-\theta')$$

with a delta function because the kinetic energy was not elliptic, and because of the delta function, loop amplitudes will be proportional to $\delta(0)$:



This loop will come with a factor $\delta(\theta-\theta')^2 = \delta(\theta-\theta')\delta(0)$.

What Kevin Costello did was to cure this problem via a very simple deformation. Take our three-manifold to be $\mathbb{R}^3$, and write x, y, t for the three coordinates of $\mathbb{R}^3$, so overall we have x, y, t, and θ. Costello combined t and θ into a complex variable

$$z = \varepsilon t + \mathrm{i}\theta.$$

Here ε is a real parameter. The theory will reduce to the bad case that I just described if $\varepsilon = 0$. As soon as $\varepsilon \neq 0$, its value does not matter and one can set $\varepsilon = 1$. I just included ε to explain in what sense we are making an infinitesimal deformation away from the ill-defined Chern-Simons theory of the loop group.

One replaces $\mathrm{d}\theta$ (or $(k/4\pi)\mathrm{d}\theta$) in the naive theory with $\mathrm{d}z$ (or $\mathrm{d}z/\hbar$) and one now regards A as a partial connection on $\mathbb{R}^3 \times S^1$ that is missing a $\mathrm{d}z$ term (rather than missing $\mathrm{d}\theta$, as before). The action is now

$$I = \frac{1}{\hbar} \int_{\mathbb{R}^3 \times S^1} \mathrm{d}z \, \mathrm{Tr} \left(A \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right).$$

We've lost the three-dimensional symmetry of standard Chern-Simons theory, because of splitting away one of the three coordinates of $\mathbb{R}^3$ and combining it with θ. We still have two-dimensional diffeomorphism symmetry. However, as we discussed when we were comparing Yang-Baxter theory to knot theory, Yang-Baxter theory does not have three-dimensional symmetry, only two-dimensional symmetry. Modifying standard Chern-Simons theory in this fashion turns out to have exactly the right properties to give Yang-Baxter theory rather than knot theory: the three-dimensional diffeomorphism invariance is reduced to two-dimensional diffeomorphism invariance, but on the other hand, now there is a complex variable z that will turn out to be the spectral parameter.

I've described the action so far on $\mathbb{R}^2 \times \mathbb{C}^*$ where $\mathbb{C}^* = \mathbb{R} \times S^1$ (parametrized by $z = \varepsilon t + \mathrm{i}\theta$), with the complex 1-form $\mathrm{d}z$. The theory works just as well if $\mathbb{C}^*$ is replaced by $\mathbb{C}$ or by an elliptic curve (the quotient of $\mathbb{C}$ by a lattice of rank 2). The three cases turn out to correspond to rational, trigonometric, and elliptic solutions of Yang-Baxter.

The first point is that this theory has a sensible propagator and a sensible perturbation expansion. The basic reason for a sensible propagator is that on $\mathbb{R} \times \mathbb{R}$ or $\mathbb{R} \times S^1$ parametrized by t and θ, the operator $\partial/\partial t$ that appeared in the naive action of $\mathcal{L}G$ is not elliptic, but the operator $\partial/\partial\overline{z}$ that appears in the deformed version is elliptic. After a suitable gauge-fixing, the propagator (on $\mathbb{R}^2 \times C$ for the rational model, i.e. $C = \mathbb{C} \cong \mathbb{R}^2$) is

$$\langle A_i(x,y,z)A_j(x',y',z')\rangle = \varepsilon_{ijkz} g^{kl} \frac{\partial}{\partial x^l}\left(\frac{1}{(x-x')^2+(z-z')^2+|z-z'|^2}\right)$$

where i, j, k take the values $x, y, \overline{z}$ and the metric on $\mathbb{R}^4 = \mathbb{R}^2 \times C$ is $\mathrm{d}x^2 + \mathrm{d}y^2 + |\mathrm{d}z|^2$. As Costello proves, there is no difficulty in doing perturbation theory.

Now we consider Wilson operators, that is holonomy operators

$$\mathrm{Tr}_\rho P \exp \oint_\ell A$$

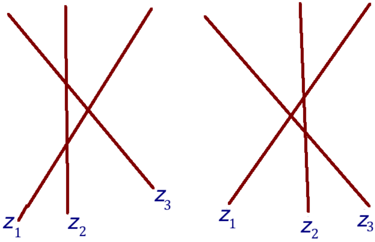
where ℓ is a loop in $\Sigma \times C$. Here Σ is the topological two-manifold, and C is a complex Riemann surface (with the differential $\omega = \mathrm{d}z$). But we only have a partial gauge field or connection

$$A = A_x \mathrm{d}x + A_y \mathrm{d}y + A_{\overline{z}} d\overline{z}$$

so we would not know how to do any parallel transport in the z direction. (We cannot interpret A as a gauge field with $A_z = 0$ because this condition would not be gauge-invariant, and quantizing the theory requires gauge-invariance. We have to interpret it as a theory with A_z undefined, so we cannot do parallel transport in the z direction.) This means that we must take ℓ to be a loop that lies in Σ, at a particular value of z.

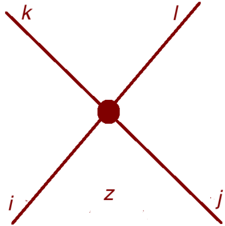
That is actually what we want for Yang-Baxter theory: z is the spectral parameter at which the given knot "lives." In Yang-Baxter theory, the spectral parameter is indeed in general a complex parameter.

Now let us consider some lines that meet in Σ in the familiar configuration associated to the Yang-Baxter equation:



Two-dimensional diffeomorphism invariance means that we are free to move the lines around as long as we don't change the topology of the configuration. But assuming that z_1, z_2, and z_3 are all distinct, it is manifest that there is no discontinuity when we move the middle line from left to right even when we do cross between the two pictures. Thus two configurations of Wilson operators that differ by what we might call a Yang-Baxter move are equivalent.

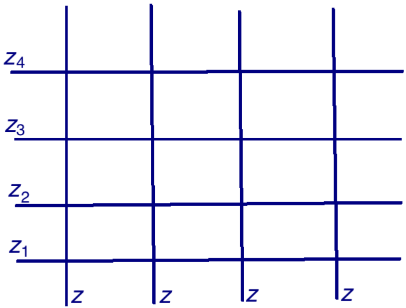
But why is there as elementary a picture as in the lattice spin systems, where one can evaluate the path integral by labeling each line by a basis element of the representation ρ and each crossing by a local factor $R_{ij,kl}(z)$?



It turns out that this has a simple proof using two facts: (1) the theory is "topological" in the Σ direction, meaning that it does not depend on a length scale; (2) by power-counting, the theory is infrared-free.

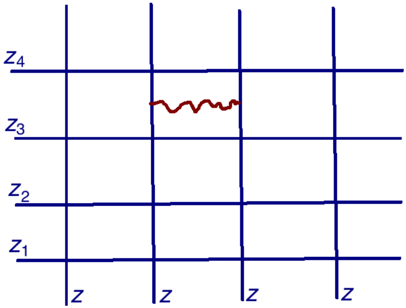
A metric on $\Sigma \times \mathbb{C}$ entered only when we fixed the gauge to pick a propagator. Recall that we used the metric $\mathrm{d}x^2 + \mathrm{d}y^2 + |\mathrm{d}z|^2$. We could equally well scale up the metric along Σ by any factor and use instead $e^B(\mathrm{d}x^2 + \mathrm{d}y^2) + |\mathrm{d}z|^2$ for very large B.

That means that when you look at this picture



which is drawn in $\Sigma = \mathbb{R}^2$ (with the lines being labeled by points $z_i \in \mathbb{C}$) you can consider the vertical lines and likewise the horizontal lines to be very far apart (compared to $z - z_i$ or $z_i - z_j$).

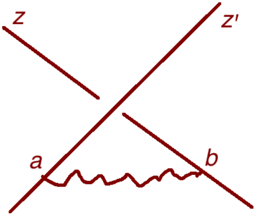
In that limit, effects that involve a gauge boson exchange between two nonintersecting lines are negligible:



One should worry about gauge boson exchange from one line to itself

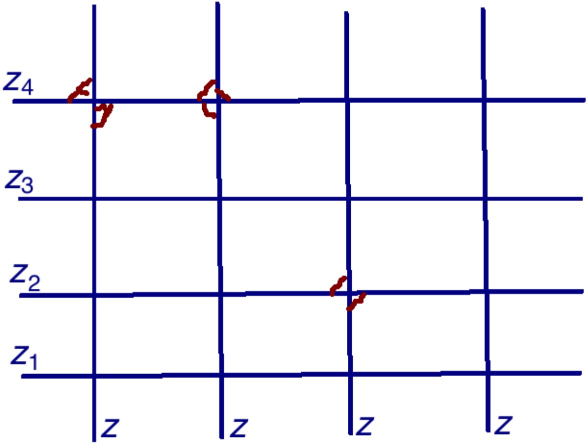
because then the distance $|a - b|$ need not be large. Such effects correspond roughly to "mass renormalization" in standard quantum field theory. In the present problem, in the case of a straight Wilson line, the symmetries do not allow any interesting effect analogous to mass renormalization.

When two lines cross we get an integral



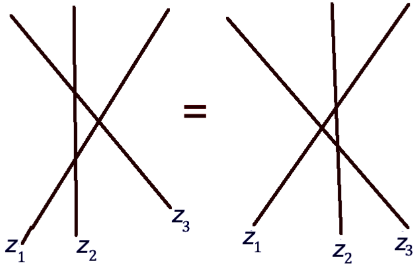
over a and b that converges, and receives significant contributions only from the region $|a|, |b| \lesssim |z - z'|$. I will say what it converges to in a few minutes.

Now when we study a general configuration such as the one related to the integrable lattice models



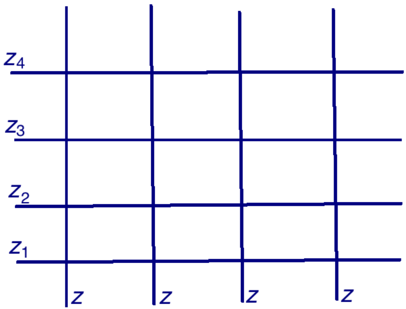
we can draw very complicated diagrams, but the complications are all localized near one crossing point or another.

The diagrams localized near one crossing point simply build up a universal R-matrix associated to that crossing, and the discussion makes it obvious that the Yang-Baxter equation



is obeyed.

Moreover, this makes it clear that the path integral in the presence of the configuration of Wilson operators associated to the integrable lattice models



can be evaluated by the standard rules – label each vertical or horizontal line segment by a basis vector $i, j, k, \ldots$ of the representation ρ, and include the appropriate R-matrix element at each crossing; then sum over all such labelings.

With a little more work, one can show that this construction accounts for the standard rational, trigonometric, and elliptic solutions of the Yang-Baxter equation.

The story goes farther that I have been able to explain today. For example, a slight extension of the picture accounts for the "modified Yang-Baxter equation" of Felder. Costello and D. Gaiotto showed that Baxter's Q operator (which is a key tool in actually solving for the partition function of one of these spin systems) arises in the four-dimensional gauge theory as the "'t Hooft operator," a basic ingredient in gauge theory. And Costello and Yamazaki used the same setup to account for many properties of integrable models of many-body physics in $1+1$ dimensions. I suspect that much more can be done.

Personally, what I find most satisfying about this perspective is that it gives a unified picture of many different phenomena, and in some sense, it explains “why” the highly overdetermined Yang-Baxter equation has non-trivial solutions or equivalently why the corresponding integrable systems exist.