A New Look At Integrable Spin Systems

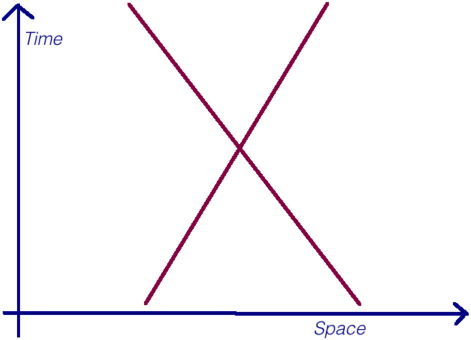
Edward Witten

Lecture at Strings 2016, Beijing

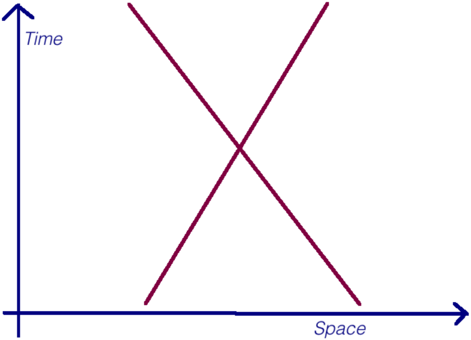
You have just heard about Kevin Costello's approach to integrable spin systems via gauge theory.

You have just heard about Kevin Costello's approach to integrable spin systems via gauge theory. I will be saying a little more about it.

This is meant to be a spacetime picture of elastic scattering of two particles



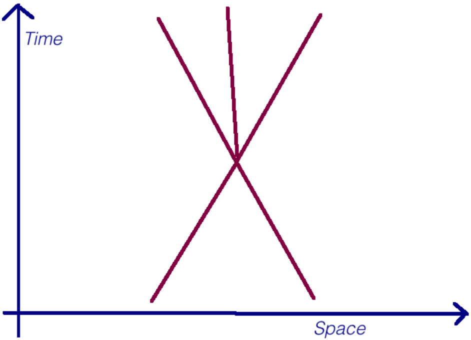
This is meant to be a spacetime picture of elastic scattering of two particles



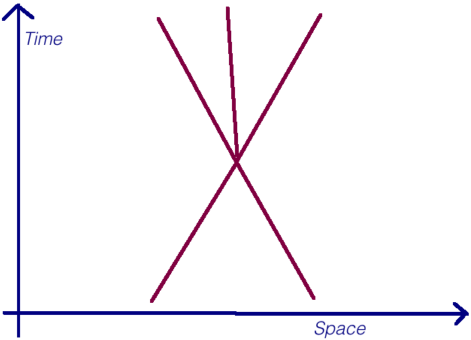
Because of conservation of energy and momentum, the outgoing particles go off at the same slope (same velocity) as the incoming particles. There are time delays that I have not tried to draw.

In a typical relativistic quantum field theory, there are particle production processes, which are a large part of what makes quantum field theory interesting.

In a typical relativistic quantum field theory, there are particle production processes, which are a large part of what makes quantum field theory interesting. Here is a picture of two particles going to three:



In a typical relativistic quantum field theory, there are particle production processes, which are a large part of what makes quantum field theory interesting. Here is a picture of two particles going to three:

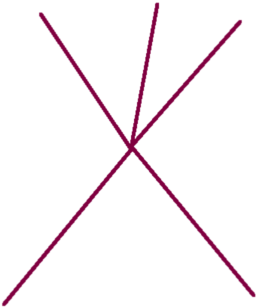


The symmetries of typical relativistic field theories allow such processes and they happen all the time in the real world.

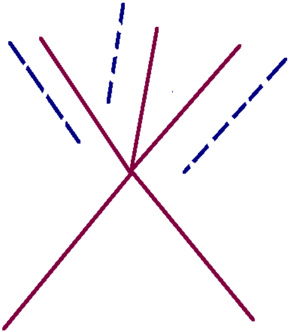
However, in two spacetime dimensions, there are “integrable” field theories that have extra symmetries that move a particle in space by an amount that depends on its velocity.

However, in two spacetime dimensions, there are "integrable" field theories that have extra symmetries that move a particle in space by an amount that depends on its velocity. Then particle production is not possible:

However, in two spacetime dimensions, there are "integrable" field theories that have extra symmetries that move a particle in space by an amount that depends on its velocity. Then particle production is not possible:

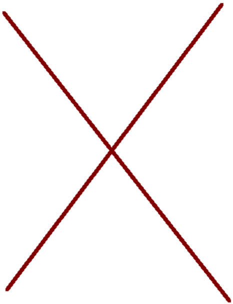


However, in two spacetime dimensions, there are “integrable” field theories that have extra symmetries that move a particle in space by an amount that depends on its velocity. Then particle production is not possible:

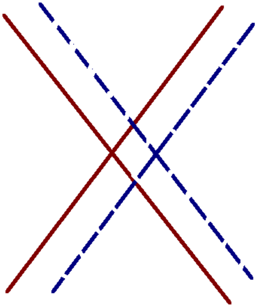


since several generic lines in the plane do not intersect.

Two particle scattering does happen even in integrable systems, since two lines in the plane do generically intersect:



Two particle scattering does happen even in integrable systems, since two lines in the plane do generically intersect:



How do we characterize a particle?

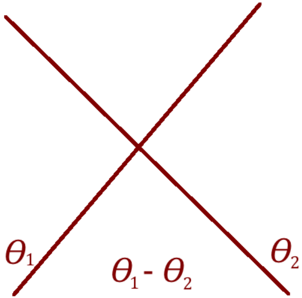
How do we characterize a particle? A particle has a velocity, or better, in relativistic terms, a "rapidity" θ

$$\begin{pmatrix} E \\ p \end{pmatrix} = m \begin{pmatrix} \cosh\theta \\ \sinh\theta \end{pmatrix}$$

How do we characterize a particle? A particle has a velocity, or better, in relativistic terms, a “rapidity” θ

$$\begin{pmatrix} E \\ p \end{pmatrix} = m \begin{pmatrix} \cosh\theta \\ \sinh\theta \end{pmatrix}$$

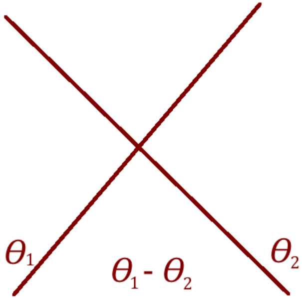
Scattering of two particles with rapidities θ_1 and θ_2 depends only on the rapidity difference $\theta = \theta_1 - \theta_2$:



How do we characterize a particle? A particle has a velocity, or better, in relativistic terms, a "rapidity" θ

$$\begin{pmatrix} E \\ p \end{pmatrix} = m \begin{pmatrix} \cosh\theta \\ \sinh\theta \end{pmatrix}$$

Scattering of two particles with rapidities θ_1 and θ_2 depends only on the rapidity difference $\theta = \theta_1 - \theta_2$:

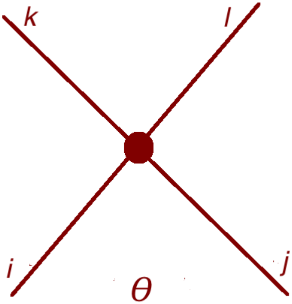


(Note that the slope with which I draw a line depends on the rapidity of the particle in question.)

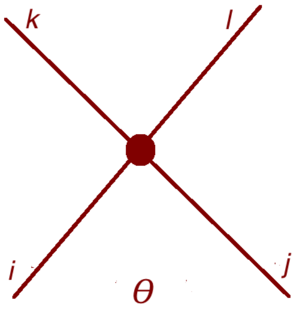
But the amplitude for scattering of two particles of rapidities θ_1 and θ_2 is in general not only a function of the rapidity difference $\theta = \theta_1 - \theta_2$ because there may be several different "types" of particles of the same mass.

But the amplitude for scattering of two particles of rapidities θ_1 and θ_2 is in general not only a function of the rapidity difference $\theta = \theta_1 - \theta_2$ because there may be several different "types" of particles of the same mass. An obvious reason for this is that the theory might have a symmetry group G and the particles may be in an irreducible representation ρ of G.

The picture is then more like this:

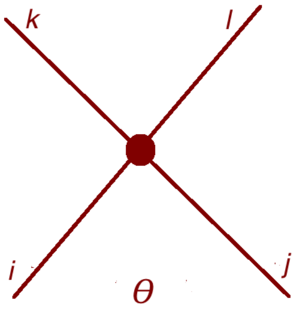


The picture is then more like this:



Here $i, j, k.l$ can be understood to represent basis vectors in the representation ρ. We write $R_{ij,kl}(\theta)$ for the quantum mechanical "amplitude" that describes this process.

The picture is then more like this:

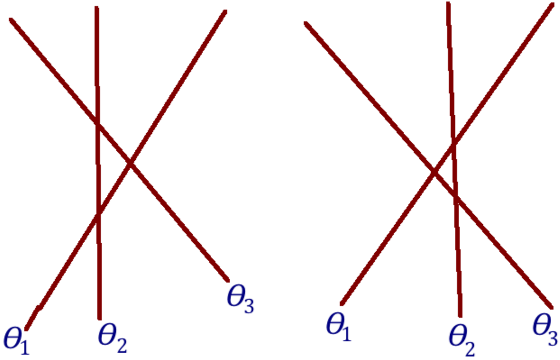


Here $i, j, k.l$ can be understood to represent basis vectors in the representation ρ. We write $R_{ij,kl}(\theta)$ for the quantum mechanical "amplitude" that describes this process. It is usually called the R-matrix.

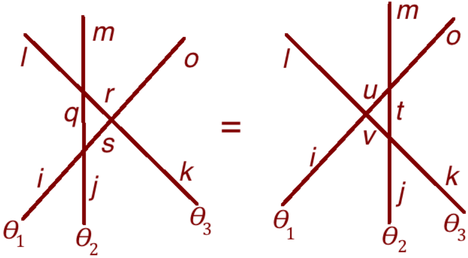
The real fun comes when we consider three particles in the initial and final state.

The real fun comes when we consider three particles in the initial and final state. Since we can move them relative to each other, leaving their slopes (or rapidities) fixed, we can assume that there are only pairwise collisions.

The real fun comes when we consider three particles in the initial and final state. Since we can move them relative to each other, leaving their slopes (or rapidities) fixed, we can assume that there are only pairwise collisions. But there are two ways to do this and they must give equivalent results:



In more detail, equivalence of these pictures leads to the celebrated "Yang-Baxter equation"



which schematically reads

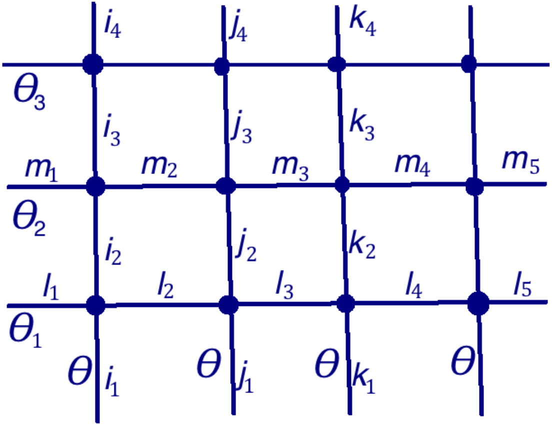
$$R_{23}R_{13}R_{12} = R_{12}R_{13}R_{23}.$$

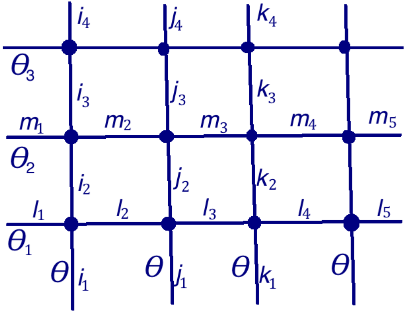
The traditional solutions of the Yang-Baxter equation – as discovered by Bethe, Lieb, Yang, Baxter, Fadde'ev, Drin'feld and many others – are classified by the choice of a Lie group G and a representation ρ, subject to (1) some restrictions, and (2) the curious fact that in many important cases (like the 6-vertex model of Lieb and the 8-vertex model of Baxter) a model associated to a given group G does not actually have G symmetry.

The traditional solutions of the Yang-Baxter equation – as discovered by Bethe, Lieb, Yang, Baxter, Fadde'ev, Drin'feld and many others – are classified by the choice of a Lie group G and a representation ρ, subject to (1) some restrictions, and (2) the curious fact that in many important cases (like the 6-vertex model of Lieb and the 8-vertex model of Baxter) a model associated to a given group G does not actually have G symmetry. In fact, there are three broad classes of solutions of Yang-Baxter – "rational, trigonometric, and elliptic" – and only the rational ones have G symmetry.

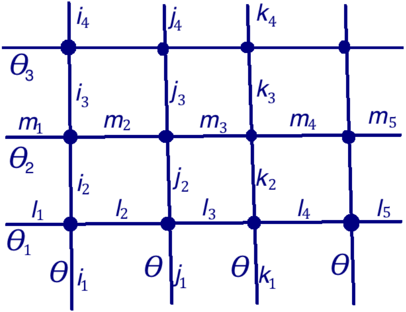
I've motivated this introduction by talking about relativistic scattering, but the same solutions of the Yang-Baxter equation are used for physical models of a completely different sort.

I've motivated this introduction by talking about relativistic scattering, but the same solutions of the Yang-Baxter equation are used for physical models of a completely different sort. The ones most relevant today are the integrable lattice systems of statistical mechanics, which are constructed directly from a solution of the Yang-Baxter equation:

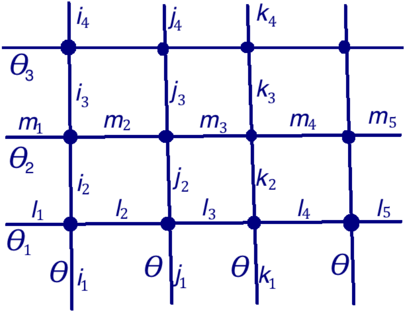




i_4 j_4 k_4
θ_3
i_3 j_3 k_3
m_1 m_2 m_3 m_4 m_5
θ_2
i_2 j_2 k_2
l_1 l_2 l_3 l_4 l_5
θ_1
θ i_1 θ j_1 θ k_1 θ



To explain this rather busy picture, the vertical and horizontal lines are labeled by rapidities θ or θ_i, a line segment is labeled by a basis vector $i, j, k, \ldots$ of the representation ρ, and a crossing is labeled by the appropriate R-matrix. The problem of statistical mechanics is to compute the "partition function" by summing over labels, with each set of labels being weighted by the product of the appropriate R-matrix elements.



To explain this rather busy picture, the vertical and horizontal lines are labeled by rapidities θ or θ_i, a line segment is labeled by a basis vector $i, j, k, \ldots$ of the representation ρ, and a crossing is labeled by the appropriate R-matrix. The problem of statistical mechanics is to compute the "partition function" by summing over labels, with each set of labels being weighted by the product of the appropriate R-matrix elements. The problem turns out to be solvable because "the transfer matrices commute," which means that (using the Yang-Baxter equation) the horizontal lines can be moved up and down past each other.

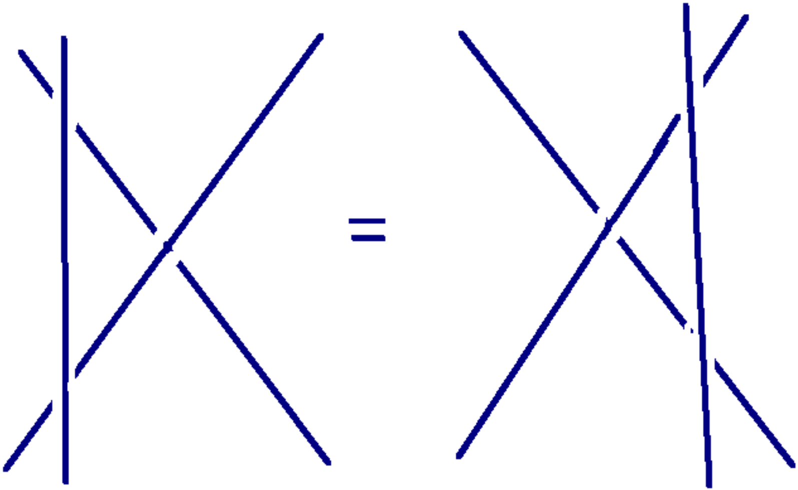
Perhaps the most obvious question about the Yang-Baxter equation is “why” solutions of this highly overdetermined equation exist.

Perhaps the most obvious question about the Yang-Baxter equation is "why" solutions of this highly overdetermined equation exist. There is another area in which one finds something a lot like the Yang-Baxter equations.

Perhaps the most obvious question about the Yang-Baxter equation is "why" solutions of this highly overdetermined equation exist. 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.

Perhaps the most obvious question about the Yang-Baxter equation is "why" solutions of this highly overdetermined equation exist. 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:

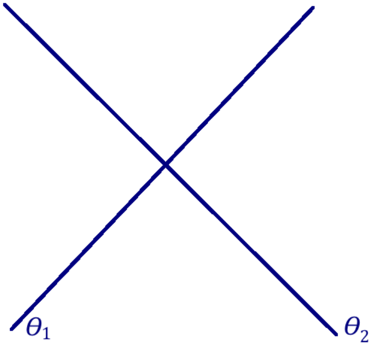
Perhaps the most obvious question about the Yang-Baxter equation is "why" solutions of this highly overdetermined equation exist. There is another area in which one finds something a lot like the Yang-Baxter equations. This is the theory of knots in three dimensions. Here is one of the Reidemeister moves:



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

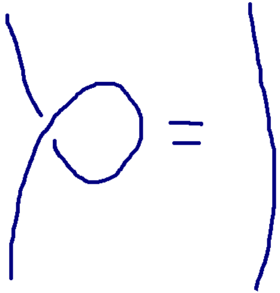
The resemblance to the Yang-Baxter equation is obvious, but there are also conspicuous differences:

(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:

(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.

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:

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.

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.

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.

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, let M be a three-manifold, $E \to M$ a G-bundle, and A a connection on G.

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, 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*

$$\mathrm{CS}(A) = \frac{1}{4\pi} \int_M \mathrm{Tr} \left(A \mathrm{d} A + \frac{2}{3} A \wedge A \wedge A \right).$$

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, 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*

$$\mathrm{CS}(A) = \frac{1}{4\pi} \int_M \mathrm{Tr} \left(A \mathrm{d} A + \frac{2}{3} A \wedge A \wedge A \right).$$

I have normalized it so that it is gauge-invariant mod $2\pi\mathbb{Z}$.

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, 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*

$$\mathrm{CS}(A) = \frac{1}{4\pi} \int_M \mathrm{Tr}\left(A \mathrm{d} A + \frac{2}{3} A \wedge A \wedge A \right).$$

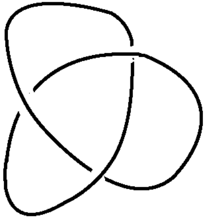
I have normalized it so that it is gauge-invariant mod $2\pi\mathbb{Z}$. In quantum mechanics, the "action" must be well-defined mod $2\pi\mathbb{Z}$, so we can take

$$I = k\mathrm{CS}(A), \quad k \in \mathbb{Z}.$$

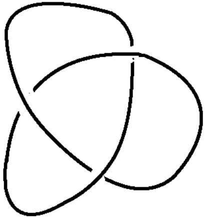
A quantum field theory with this action is a "topological quantum field theory," since there is no metric tensor in sight.

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.

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.



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 let

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

i.e. the Wilson loop operator in the representation ρ.

The usual "quantum knot invariants," of which the prototype is the Jones polynomial of a knot, can be defined via the expectation value of the Wilson operator, $\langle W_\rho(K)\rangle = \langle \mathrm{Tr}_\rho P \exp(\oint_K A)\rangle$.

The usual “quantum knot invariants,” of which the prototype is the Jones polynomial of a knot, can be defined via the expectation value of the Wilson operator, $\langle W_\rho(K)\rangle = \langle \mathrm{Tr}_\rho P \exp(\oint_K A)\rangle$. From the knot invariants that one makes this way, one cannot really extract the usual solutions of the Yang-Baxter equation since one is missing the spectral parameter.

The usual "quantum knot invariants," of which the prototype is the Jones polynomial of a knot, can be defined via the expectation value of the Wilson operator, $\langle W_\rho(K)\rangle = \langle \mathrm{Tr}_\rho P \exp(\oint_K A)\rangle$. From the knot invariants that one makes this way, one cannot really extract the usual solutions of the Yang-Baxter equation since one is missing the spectral parameter. However, in a sense from these knot invariants one can extract a special case of the Yang-Baxter solutions: in which the spectral parameter is taken to $i\infty$.

How can we modify or generalize Chern-Simons gauge theory to include 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 $\mathcal{L}G$.

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 $\mathcal{L}G$. We parametrize the loop by an angle θ.

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 $\mathcal{L}G$. We parametrize the loop by an angle θ. The loop group has "evaluation" representations that "live" at a particular value $\theta = \theta_0$ along the loop.

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 $\mathcal{L}G$. We parametrize the loop by an angle θ. The loop group has "evaluation" representations that "live" at a particular value $\theta = \theta_0$ along the loop. We hope that this will be the spectral parameter label θ_0 carried by a particle in the solution of the Yang-Baxter equation.

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 $\mathcal{L}G$. We parametrize the loop by an angle θ. The loop group has "evaluation" representations that "live" at a particular value $\theta = \theta_0$ along the loop. We hope that this will be the spectral parameter label θ_0 carried by a particle in the solution of the Yang-Baxter equation. (It is important that we take the loop group, not its central extension.)

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 $A = \sum_i A_i(x,\theta)\mathrm{d}x^i$.

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.

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.

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,

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

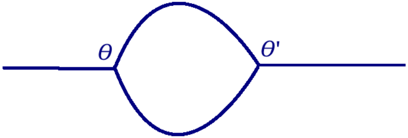
$$\langle A_i(\vec{x},\theta)A_j(\vec{x}',\theta') = \frac{\epsilon_{ijk}(x-x')^k}{|\vec{x}-\vec{x}'|^2}\delta(\theta-\theta')$$

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

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

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

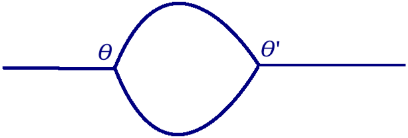
with a delta function because the kinetic energy was not elliptic, and because of the delta function, loops will be proportional to $\delta(0)$:



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, loops 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.

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.

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

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

Here ε is a real parameter. The theory will reduce to the bad case that I just described if $\varepsilon = 0$.

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

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

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 $\varepsilon = 1$.

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

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

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).

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'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.

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, but only two-dimensional symmetry.

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, but 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 $z = t + \mathrm{i}\theta$), with the complex 1-form $\mathrm{d}z$.

I've described the action so far on $\mathbb{R}^2 \times \mathbb{C}^*$ where $\mathbb{C}^* = \mathbb{R} \times S^1$ (parametrized by $z = t + \mathrm{i}\theta$), with the complex 1-form $\mathrm{d}z$. The classical action makes sense more generally on $\Sigma \times C$, where Σ is any smooth (oriented) two-manifold and C is a complex Riemann surface endowed with a holomorphic 1-form ω:

$$I = \frac{1}{\hbar} \int_{\Sigma \times C} \omega \wedge \mathrm{Tr}\left(A \wedge \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right).$$

I've described the action so far on $\mathbb{R}^2 \times \mathbb{C}^*$ where $\mathbb{C}^* = \mathbb{R} \times S^1$ (parametrized by $z = t + \mathrm{i}\theta$), with the complex 1-form $\mathrm{d}z$. The classical action makes sense more generally on $\Sigma \times C$, where Σ is any smooth (oriented) two-manifold and C is a complex Riemann surface endowed with a holomorphic 1-form ω:

$$I = \frac{1}{\hbar} \int_{\Sigma \times C} \omega \wedge \mathrm{Tr} \left(A \wedge \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right).$$

It turns out, however, that to get a quantum theory, one wants ω to have no zeroes.

I've described the action so far on $\mathbb{R}^2 \times \mathbb{C}^*$ where $\mathbb{C}^* = \mathbb{R} \times S^1$ (parametrized by $z = t + \mathrm{i}\theta$), with the complex 1-form $\mathrm{d}z$. The classical action makes sense more generally on $\Sigma \times C$, where Σ is any smooth (oriented) two-manifold and C is a complex Riemann surface endowed with a holomorphic 1-form ω:

$$I = \frac{1}{\hbar} \int_{\Sigma \times C} \omega \wedge \mathrm{Tr}\left(A \wedge \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right).$$

It turns out, however, that to get a quantum theory, one wants ω to have no zeroes. Intuitively this is because a zero of ω is equivalent to a point at which $\hbar \to \infty$.

I've described the action so far on $\mathbb{R}^2 \times \mathbb{C}^*$ where $\mathbb{C}^* = \mathbb{R} \times S^1$ (parametrized by $z = t + \mathrm{i}\theta$), with the complex 1-form $\mathrm{d}z$. The classical action makes sense more generally on $\Sigma \times C$, where Σ is any smooth (oriented) two-manifold and C is a complex Riemann surface endowed with a holomorphic 1-form ω:

$$I = \frac{1}{\hbar} \int_{\Sigma \times C} \omega \wedge \mathrm{Tr} \left(A \wedge \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right).$$

It turns out, however, that to get a quantum theory, one wants ω to have no zeroes. Intuitively this is because a zero of ω is equivalent to a point at which $\hbar \to \infty$. By constrast, there is no problem with poles of ω.

I've described the action so far on $\mathbb{R}^2 \times \mathbb{C}^*$ where $\mathbb{C}^* = \mathbb{R} \times S^1$ (parametrized by $z = t + \mathrm{i}\theta$), with the complex 1-form $\mathrm{d}z$. The classical action makes sense more generally on $\Sigma \times C$, where Σ is any smooth (oriented) two-manifold and C is a complex Riemann surface endowed with a holomorphic 1-form ω:

$$I = \frac{1}{\hbar} \int_{\Sigma \times C} \omega \wedge \mathrm{Tr} \left(A \wedge \mathrm{d}A + \frac{2}{3} A \wedge A \wedge A \right).$$

It turns out, however, that to get a quantum theory, one wants ω to have no zeroes. Intuitively this is because a zero of ω is equivalent to a point at which $\hbar \to \infty$. By constrast, there is no problem with poles of ω. At a pole of ω, effectively $\hbar \to 0$.

So C has to be a complex Riemann surface that has a differential ω with possible poles, but with no zeroes.

So C has to be a complex Riemann surface that has a differential ω with possible poles, but with no zeroes. The only three options are $\mathbb{C}$, $\mathbb{C}/\mathbb{Z} \cong \mathbb{C}^*$, and $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$, which is a Riemann surface of genus 1.

So C has to be a complex Riemann surface that has a differential ω with possible poles, but with no zeroes. The only three options are $\mathbb{C}$, $\mathbb{C}/\mathbb{Z} \cong \mathbb{C}^*$, and $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$, which is a Riemann surface of genus 1. It turns out that these three cases correspond to the three traditional classes of solutions of the Yang-Baxter equation – rational, trigonometric, and elliptic.

The first point is that this theory has a sensible propagator and a sensible perturbation expansion.

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.

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 (for the rational model, i.e. on $\mathbb{R}^2 \times C$ with $C = \mathbb{C} \cong \mathbb{R}^2$) 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 $\mathrm{d}x^2 + \mathrm{d}y^2 + |\mathrm{d}z|^2$.

With this propagator, the perturbative expansion is well-defined, as Costello proves.

With this propagator, the perturbative expansion is well-defined, as Costello proves. This is a tricky point: the theory is actually unrenormalizable by power cointing, so on that basis, one would not expect a well-behaved quantum theory.

With this propagator, the perturbative expansion is well-defined, as Costello proves. This is a tricky point: the theory is actually unrenormalizable by power cointing, so on that basis, one would not expect a well-behaved quantum theory. However, it has no possible counterterms, because all local gauge-invariant operators vanish by the classical equations of motion.

With this propagator, the perturbative expansion is well-defined, as Costello proves. This is a tricky point: the theory is actually unrenormalizable by power cointing, so on that basis, one would not expect a well-behaved quantum theory. However, it has no possible counterterms, because all local gauge-invariant operators vanish by the classical equations of motion. Anyway Costello proves, using a fairly elaborate algebraic machinery of BV quantization, that the theory has a well-defined perturbation expansion.

Now we consider Wilson operators, that is holonomy operators

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

where ℓ is a loop in $\Sigma \times C$.

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 $\omega = \mathrm{d}z$).

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 connection

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

so we would not know how to do any parallel transport in the z direction.

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 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.)

Now we consider Wilson operators, that is holonomy operators

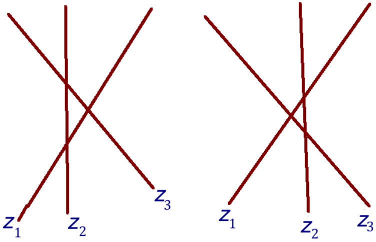
$$\mathrm{Tr}_\rho P \exp \oint_\ell A$$

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 connection

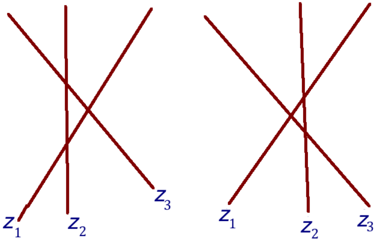
$$A = A_x \mathrm{d}x + A_y \mathrm{d}y + A_{\overline{z}} d\overline{z}$$

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.

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

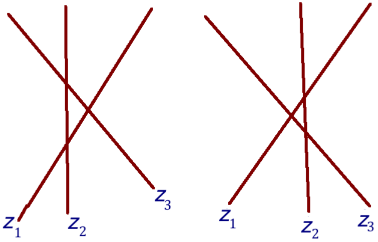


Now let us consider some lines that meet in Σ in the familiar configuration associated to the Yang-Baxter equation:



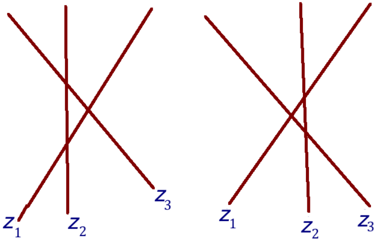
Two-dimensional diffeomorphism invariance means that we are free to move the lines around as long as we don't change the topology of the configuration.

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



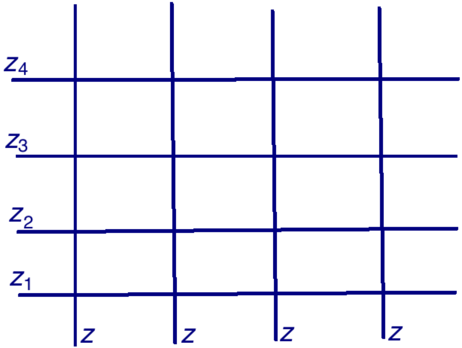
Two-dimensional diffeomorphism invariance means that we are free to move the lines around as long as we don't change the topology of the configuration. But assuming that z_1, z_2, and z_3 are all distinct, it is manifest that there is no discontinuity when we move the middle line from left to right even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

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



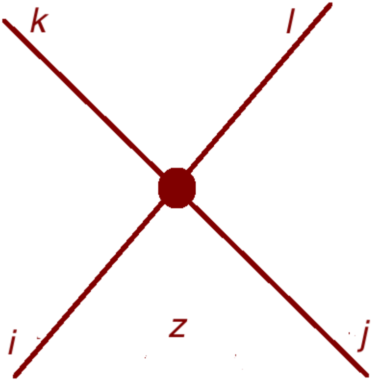
Two-dimensional diffeomorphism invariance means that we are free to move the lines around as long as we don't change the topology of the configuration. But assuming that z_1, z_2, and z_3 are all distinct, it is manifest that there is no discontinuity when we move the middle line from left to right even 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.

Likewise, in the configuration associated to integrable lattice spin systems



we can move the horizontal lines up and down at will.

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 $R_{ij,kl}(z)$?



This is a little tricky and depends on picking the right boundary condition, but there is a way to make it work for each of the three choices of C, corresponding to rational, trigonometric, and elliptic solutions of the Yang-Baxter equation.

Easiest to explain is the rational case, in which the Riemann surface is $C = \mathbb{C}$, the complex plane.

Easiest to explain is the rational case, in which the Riemann surface is $C = \mathbb{C}$, the complex plane. We require that the gauge field A on $\Sigma \times \mathbb{C}$ goes to 0 at infinity in the $\mathbb{C}$ direction, and likewise in quantizing we divide only by gauge transformations that approach 1 at infinity along $\mathbb{C}$.

Easiest to explain is the rational case, in which the Riemann surface is $C = \mathbb{C}$, the complex plane. We require that the gauge field A on $\Sigma \times \mathbb{C}$ goes to 0 at infinity in the $\mathbb{C}$ direction, and likewise in quantizing we divide only by gauge transformations that approach 1 at infinity along $\mathbb{C}$. Then one finds that the classical solution $A = 0$ has no deformations (up to gauge transformation) and also no automorphisms (unbroken gauge symmetries).

So on $\Sigma \times \mathbb{C}$, when we expand around the trivial solution $A = 0$, there are no deformations or automorphisms of this trivial solution and hence the perturbative expansion is straightforward.

So on $\Sigma \times \mathbb{C}$, when we expand around the trivial solution $A = 0$, there are no deformations or automorphisms of this trivial solution and hence the perturbative expansion is straightforward. It gives a simple answer because the theory is infrared-trivial, which is the flip side of the fact that it is unrenormalizable by power-counting.

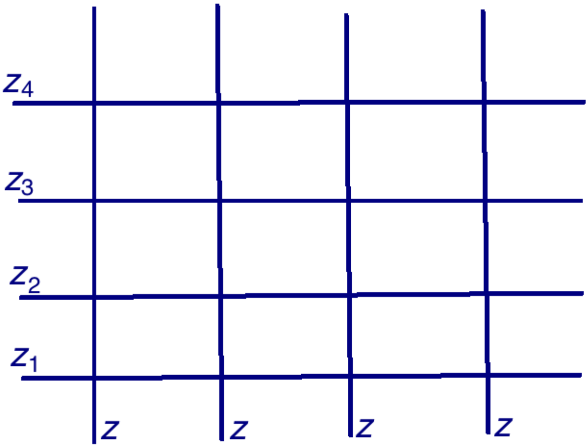
So on $\Sigma \times \mathbb{C}$, when we expand around the trivial solution $A = 0$, there are no deformations or automorphisms of this trivial solution and hence the perturbative expansion is straightforward. It gives a simple answer because the theory is infrared-trivial, which is the flip side of the fact that it is unrenormalizable by power-counting. That means that effects at "long distances" in the topological space are negligible.

I put the phrase "long distances" in quotes because two-dimensional diffeomorphism invariance means that there is no notion of distance on the topological space Σ (the first factor of $\Sigma \times \mathbb{C}$).

I put the phrase "long distances" in quotes because two-dimensional diffeomorphism invariance means that there is no notion of distance on the topological space Σ (the first factor of $\Sigma \times \mathbb{C}$). A metric on $\Sigma \times \mathbb{C}$ entered only when we fixed the gauge to pick a propagator. Recall that we used the metric $\mathrm{d}x^2 + \mathrm{d}y^2 + |\mathrm{d}z|^2$.

I put the phrase “long distances” in quotes because two-dimensional diffeomorphism invariance means that there is no notion of distance on the topological space Σ (the first factor of $\Sigma \times \mathbb{C}$). 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.

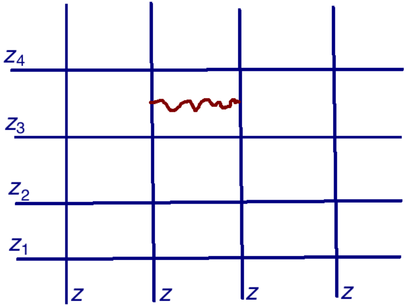
That means that when you look at this picture



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 such a situation, in an infrared-free theory, effects that involve a gauge boson exchange between two nonintersecting lines are negligible:

In such a situation, in an infrared-free theory, 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

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

because then the distance $|a - b|$ need not be large.

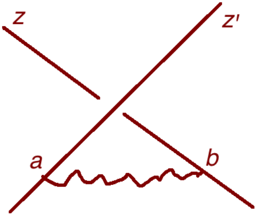
One should worry about gauge boson exchange from one line to itself

because then the distance $|a - b|$ need not be large. Such effects correspond roughly to "mass renormalization" in standard quantum field theory.

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

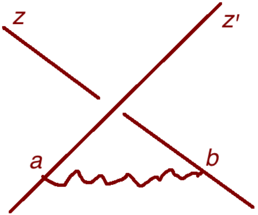
because then the distance $|a - b|$ need not be large. Such effects correspond roughly to "mass renormalization" in standard quantum field theory. In the present problem, in the case of a straight Wilson line, the symmetries do not allow any interesting effect analogous to mass renormalization.

When two lines cross we get an integral



over a and b that converges, and receives significant contributions only from the region $|a|, |b| \lesssim |z - z'|$.

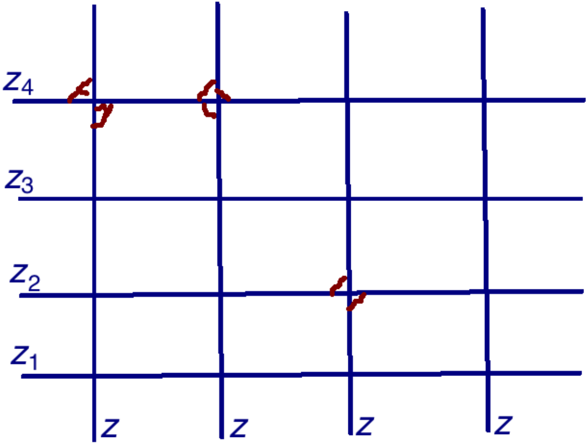
When two lines cross we get an integral



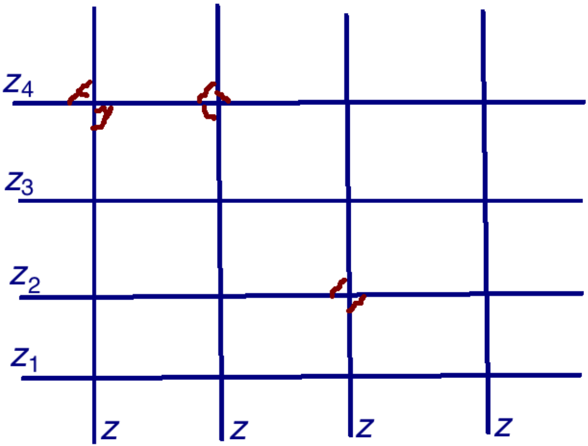
over a and b that converges, and receives significant contributions only from the region $|a|, |b| \lesssim |z - z'|$. I will say what it converges to in a few minutes.

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

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

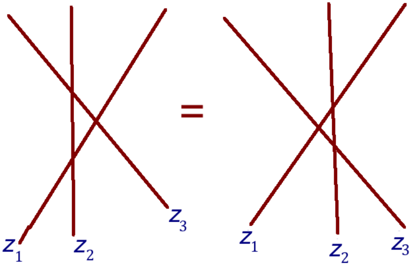


Now when we study a general configuration such as the one related to the integrable lattice models



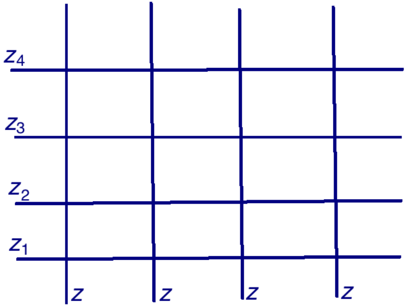
we can draw very complicated diagrams, but the complications are all localized near one crossing point or another.

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.

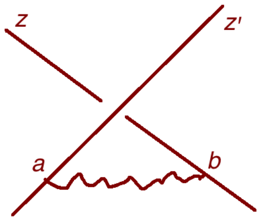
Moreover, this makes it clear that the path integral in the presence of the configuration of Wilson operators associated to the integrable lattice models



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

But why is the R-matrix obtained this way the standard rational solution of the Yang-Baxter equation? (or the standard trigonometric or elliptic one, if we had done one of those cases).

In his paper, Costello explicitly evaluates the lowest order correction in $R = 1 + \hbar r + \mathcal{O}(\hbar^2)$ from this diagram

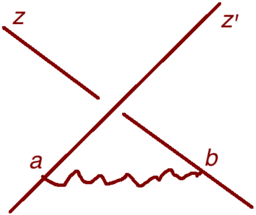


and gets the standard answer

$$r = \frac{\sum_a t_a t'_a}{z - z'}$$

(where t_a, t'_a, $a = 1, \ldots, \dim G$ are the generators of the Lie algebra of G acting in the two representations).

In his paper, Costello explicitly evaluates the lowest order correction in $R = 1 + \hbar r + \mathcal{O}(\hbar^2)$ from this diagram



and gets the standard answer

$$r = \frac{\sum_a t_a t'_a}{z - z'}$$

(where t_a, t'_a, $a = 1, \ldots, \dim G$ are the generators of the Lie algebra of G acting in the two representations). Once the first order deformation is known, the whole story follows from general arguments of Drinfeld and others.

One last comment.

One last comment. Costello's theorem is purely about perturbation theory, but his theorem shows that, in this particular theory (and rather exceptionally), perturbation theory converges.

One last comment. Costello's theorem is purely about perturbation theory, but his theorem shows that, in this particular theory (and rather exceptionally), perturbation theory converges. As a physicist, one would want to give an *a priori* "nonperturbative definition" of the theory, which would have the claimed perturbative expansion.

One last comment. Costello's theorem is purely about perturbation theory, but his theorem shows that, in this particular theory (and rather exceptionally), perturbation theory converges. As a physicist, one would want to give an *a priori* "nonperturbative definition" of the theory, which would have the claimed perturbative expansion. To do this, it appears that one has to go to string theory and use a certain "brane" system, the D4-NS5 system.

One last comment. Costello's theorem is purely about perturbation theory, but his theorem shows that, in this particular theory (and rather exceptionally), perturbation theory converges. As a physicist, one would want to give an *a priori* "nonperturbative definition" of the theory, which would have the claimed perturbative expansion. To do this, it appears that one has to go to string theory and use a certain "brane" system, the D4-NS5 system. This step would involve ideas somewhat similar to the ones I used in relating the Jones polynomial to the D3-NS5 system.