Introduction to integrable quantum field theory

Hilary term 2006, week 3 to 8, Friday 11am, Fisher room, Denis Wilkinson Building

Benjamin Doyon

Rudolf Peierls Centre for Theoretical Physics

Oxford University

1 Introduction

This is an introductory course on integrable quantum field theory. Being given the time that we have, I will not be able to go into all details, but I plan to introduce all the main concepts. I also understand that many of you may not have studied QFT in depth and may not use it commonly, so I will try to make all concepts as clear as it is possible from basic principles.

I will just talk about QFT in one infinite space dimension at zero temperature, because this is where the techniques of integrability I will talk about apply.

1.1 Integrability

The subject of integrability is, first, both modern and old, and second, sometimes very precise and sometimes quite intuitive. I will talk essentially about the modern intuitive part of it, IQFT, because it is, to my opinion, the most interesting physically, and also because mathematically it offers the opportunity of a more precise understanding of QFT.

But first, what is integrability?

The very concept of integrability was introduced in 1855 by Liouville, in classical mechanics with finitely many degrees of freedom. As he showed, a characteristics of an integrable system is that one can solve it exactly. He introduced the idea that an integrable system possesses as many independent conserved charges in involution as there are degrees of freedom. A conserved charge Q is a quantity that Poisson-commute with the Hamiltonian:

$$\{H,Q\}=0$$

and conserved charges in involution Poisson-commute with one another:

$$\{Q_m,Q_n\}=0\ .$$

Then, he showed that these charges constrain the dynamics a lot (the system must lie on “hypertori” in phase space), and that one can use them to solve by quadrature.

Now, this has limited interest in modern physics, and we would like to generalize to other systems. This is where integrability becomes more subtle.

- First, what happens in classical field theory? There are infinitely many degrees of freedom, so we would need infinitely many conserved charges, but how big an infinity? I don't think there is yet a clear answer to this question. Also, solving by quadrature just does not work anymore, the problem is infinitely huge! Fortunately, a deeper understanding has been gained in the sixties: the concept of Lax pairs takes the principal role, and the inverse scattering method provides a solution method, but things become, at first sight, less intuitive.

- The problem may seem deeper with generalization to quantum mechanics: there, the Hamiltonian is a matrix, say 2^N by 2^N in a chain of interacting spin-1/2 degrees of freedom. But we can always in principle diagonalize it so there is always 2^N other matrices that commute with it and with each other; nevertheless, we cannot always find exact solutions... What do we mean by "enough conserved charges"? The ideas of integrability in classical mechanics and field theory give a good intuition, and there are here many powerful techniques developed (Bethe ansatz, quantum inverse scattering method), but, to my opinion, there is no clear fundamental principle of integrability yet.

Now, quantize a classical field theory, or take the scaling limit of a quantum chain, and you get a QFT. Both problems mentioned above about the conserved charges seem to be present, and although the techniques mentionned can be generalized, they become quite involved. Happily, in relativistic QFT, there are new structures, so that new techniqes can be developed and the theory of integrability can be based only on the effect of conserved charges from fundamental principles of QFT.

Things become nice for two reasons:

- Locality. In quantum field theory, we have the principle of locality in order to solve the second problem mentioned above. That is, conserved charges should be local. What it means is that they should be distributed evenly on the whole space

  $$Q = \int dx\, q(x)$$

  and that the charge density at one point $q(x)$ should be quantum mechanically independent from the energy density $h(x')$ at a different point:

  $$[q(x), h(x')] = 0 \ .$$

  Only in integrable models of QFT are there infinitely many local conserved charges. It is natural that having enough of independent conserved densities, the dynamics should be simpler.

- Relativistic invariance and other space-time symmetries. This constrains the spectrum and the dynamics very much, and allows to "count" the conserved charges by looking at their spin (under the boost transformation). The set of spins for which there are conserved charges is a characteristics of an integrable system.

Thanks to these two reasons, there is quite a direct link between the presence of local conserved charges in integrable QFT and the properties that allow us to solve integrable models. In fact, it turns out that just two conserved charges of higher spin are necessary for integrability (although we always find infinitely many in integrable models)!

Note that locality is also a principle that can be used to identify integrable quantum chains, and that relativistic invariance can also be present classically. But only in IQFT are these two principles so strong.

1.2 QFT

Besides these advantages, relativistic quantum field theory is very useful by itself. In the case of integrable models, it is most useful as a description for the universal part of the behavior of quantum systems (chains, edges, impurities, etc.) near to a second order quantum phase transition, or of classical statistical 2D systems (Ising model, vertex models) near to a second order thermal phase transition.

But what is a (1+1-dimensional) local relativistic quantum field theory? It is quantum mechanics:

- A Hilbert space $\mathcal{H}$;
- An operator H acting on $\mathcal{H}$, called the Hamiltonian, which is diagonalizable and whose eigenvalues are bounded from below; the vector associated with the lowest eigenvalue is the vacuum, $|\text{vac}\rangle \in \mathcal{H}$.

with additional properties:

- Relativistic invariance: operators P (the momentum operator) and B (the boost operator) which satisfy
$$[H,P]=0\ ,\quad [B,P]=iH\ ,\quad [B,H]=iP\ .$$
- Locality: one can write
$$H=\int dx\,h(x)\ ,\quad P=\int dx\,p(x)\ ,\quad B=\int dx\,b(x)$$
such that two sets of things are true:
$$\begin{array}{ll} 1. & [P,h(x)]=i\frac{\partial}{\partial x}h(x)\ ,\quad [P,p(x)]=i\frac{\partial}{\partial x}p(x) \\ 2. & [h(x),h(x')]=[p(x),p(x')]=[b(x),b(x')]=[h(x),p(x')]=[h(x),b(x')]=[p(x),b(x')]=0 \end{array}$$
for $x\neq x'$. The first equation says that x is a space variable and that the energy and momentum densities are homogeneous in space.

As usual in quantum mechanics (Heisenberg picture), time is just a parameter associated to the Hamiltonian that is generated when we “evolve” an operator $\mathcal{O}$ through
$$\mathcal{O}(t)=e^{iHt}\mathcal{O}e^{-iHt}\ .$$

Usually, one is given a Hamiltonian in terms of fields that satisfy a canonical algebra. In order to evaluate measurable quantities, one has to go further:

- Construct the Hilbert space $\mathcal{H}$, essentially defined such that the resulting Hamiltonian is bounded from below, and calculate the eigenvalues of the Hamiltonian;
- Find all local fields (observables) $\mathcal{O}(x)$, defined such that

  $$[P,\mathcal{O}(x)] = i\frac{\partial}{\partial x}\mathcal{O}(x)\ , \quad [h(x),\mathcal{O}(x')] = 0 \quad (x\neq x')$$

  and evaluate their matrix elements, in particular the correlation functions: time-ordered matrix element of products of them between vacuum states $\langle \mathrm{vac}|\mathcal{T}(\mathcal{O}(x,t)\mathcal{O}(x',t'))|\mathrm{vac}\rangle$. In fact, a quantum field theory can also be seen as a set of correlation functions of local fields, with appropriate properties.

This looks like a huge task, but in integrable QFT we can go a long way towards its completion.

A note on our terminology. We will distinguish two types of operators acting on the Hilbert space: the "local fields", which are as defined above, and the "local charges", which are just integration, on space, of local fields (the associated charge densities). Hamiltonian and momentum operator are examples of local charges. In principle, in the integrand there can be also some space-time-dependent density multiplying the local field; for instance, this is the case of the boost operator. However, we will not need this kind of charges, so when we will talk about a "local charge" we will always assume that it is uniform in space and time.

1.3 Short example of application of a QFT

In order to see how a QFT can be related to an actual condensed matter model, here is an example of how a correlation function in a quantum field theory is related to a correlation function in a quantum chain. Take the so-called "XXZ" chain in its massive regime:

$$H = J\sum_j \left(\sigma_j^{\mathrm{x}}\sigma_{j+1}^{\mathrm{x}} + \sigma_j^{\mathrm{y}}\sigma_{j+1}^{\mathrm{y}} + \Delta\sigma_j^{\mathrm{z}}\sigma_{j+1}^{\mathrm{z}}\right)$$

for $J>0$, where $\sigma_j^{\mathrm{x,y,z}}$ are Pauli matrices associated to independent sites $j\in\mathbb{Z}$. The massive regime is for $\Delta>1$, and it means that (zero-temperature) correlation functions decrease exponentially at large distances, for instance

$$\langle\sigma_j^{\mathrm{y}}\sigma_k^{\mathrm{y}}\rangle_{XXZ}\propto e^{-\frac{|j-k|}{\xi}} \quad \text{as} \quad |j-k|\to\infty$$

for some *correlation length* ξ. The correlation length is a function of the parameter Δ of the model. It turns out that it tends towards infinity as $\Delta\to 1^+$. That means that there is a second order (quantum) phase transition in the model at that point, and we can calculate universal quantities using quantum field theory. This goes as follows:

$$\lim_{\Delta\to1^+}\xi^{-2d}\langle\sigma_{[mx\xi]}^{\mathrm{y}}\sigma_{[mx'\xi]}^{\mathrm{y}}\rangle_{XXZ} = \langle\mathrm{vac}|\mathcal{O}(x)\mathcal{O}(x')|\mathrm{vac}\rangle$$

where the limit is taken keeping the real parameters x and x' fixed (and $[mx\xi]$ mean that we must take the integer part of $mx\xi$). On the right-hand side, we have a correlation function in a QFT with a mass m, and the operator $\mathcal{O}(x)$ is an appropriate local field ("representing" σ^{y} in the QFT). The unique (positive) number d making the limit finite is called the dimension of the field $\mathcal{O}$. The QFT correlation function is a "scaling function", and is universal, in that it depends on the details of the initial quantum chain just through a finite number of parameters (models with very different interactions and basic constituents could give the same QFT). In this case, the QFT is integrable (it is related to the quantum sine-Gordon model).

1.4 Factorized scattering theory

Factorized scattering theory applies most naturally to massive models, which are those for which the correlation functions of local field decrease exponentially fast at large space-like distance.

The steps in the theory of factorizable scattering are as follows:

- Find the exact scattering from analytic properties and factorization (Yang-Baxter equations);
- Find the exact form factors of local fields (that is, their matrix elements in asymptotic states) from the knowledge of the scattering matrix and from assumed analytic properties;
- Calculate correlation functions from the form-factor expansion (Lehman expansion).

Hence, even if we are interested in response functions in a condensed matter system, for instance, we still need to know everything about the scattering matrix and the particles of the QFT describing the scaling limit.

2 Factorizable S-matrix

2.1 Asymptotic states in massive models

The structure of the Hilbert space can in fact be described to a great extent without knowing the details of the Hamiltonian, thanks to relativistic invariance. In general, in local quantum field theory, the Hilbert space is a module for the algebra of space symmetries. A one-particle sub-module is a unitary irreducible module. Here, we have relativistic invariance, and a unitary irreducible module characterized by the real number m (and possibly by other quantum numbers if internal symmetries are present) has basis

$$|\theta, m\rangle \quad (\theta \in \mathbb{R})$$

such that

$$H|\theta,m\rangle = m\cosh(\theta)|\theta,m\rangle \ , \quad P|\theta,m\rangle = m\sinh(\theta)|\theta,m\rangle \ , \quad B|\theta,m\rangle = -i\frac{\partial}{\partial\theta}|\theta,m\rangle \ .$$

This generates an irreducible representation with Casimir

$$H^2 - P^2 = m^2 \ .$$

The variable θ, the rapidity, is just a convenient way of parametrizing solutions to this equation which is the usual relativistic dispersion relation. It is natural that we have such a sub-module in $\mathcal{H}$, because if there is just one stable particle, it cannot interact so it will just propagate freely.

Now, we want quantum field theory to describe the scattering and interaction of particles around a given time and around a given point, and so we assume that if there are many particles, at a finite time before or after the "experimentation", they are separated by a finite distance. Hence, as we go back in time, they whether get more separated, or some may stay together forever, forming bound states. If we see the bound states as particles themselves, at the infinite past, we have many particles infinitely separated, so that they all propagate freely. The associated states, the *in* states, correspond to tensor products of the one-particle states:

$$|\theta_1, m_1; \theta_2, m_2; \ldots\rangle^{in} \ .$$

Certainly the same is true in the infinite future, giving the *out* states:

$$|\theta_1, m_1; \theta_2, m_2; \ldots\rangle^{out} \ .$$

Those form two bases for the Hilbert space, with energy

$$H|\theta_1, m_1; \theta_2, m_2; \ldots\rangle^{in,out} = \sum_k m_k \cosh(\theta_k) |\theta_1, m_1; \theta_2, m_2; \ldots\rangle^{in,out}$$

Note that if in the infinite past we have a pure *in* state, then most likely in the infinite future we will have a linear combinations of *out* states. The overlaps between *in*-states and *out*-states form the scattering matrix, or S-matrix.

Of course, quantum mechanically, a particle cannot have both a definite momentum and a definite position. To be more precise, the construction of such states goes as follows. First, identify a local field $\Psi(x)$ that "creates" the particle in which you are interested. This means that it has the quantum numbers of the particle of interest, and that the Fourier transform of the two-point function

$$\langle \text{vac}|\mathcal{T}(\Psi(x,t)\Psi(0,0))|\text{vac}\rangle$$

($\mathcal{T}$ is the time-ordering symbol, bringing the earliest operator to the right – here we assume Ψ to be a real field without any charge) has a pole, as function of the square two-momentum $E^2 - p^2$, at the square of the mass m^2 of the particle, and no other singularity at lower values. In superrenormalizable theories, such a field can be uniquely determined by requiring that it be of lowest dimension. This means that correlation functions of such fields will satisfy the Klein-Gordon equation (or other equations depending on the representation of the Lorentz group associated to the particle) asymptotically at large times

$$(m^2 - \partial_x^2 + \partial_t^2)\langle \Psi(x,t) \cdots \rangle = O\left(e^{-m'\sqrt{t^2-x^2}}\right) \quad \text{as} \quad t^2 - x^2 \to \infty$$

where m' is the lowest mass, greater than m, of a state created by this field, and $\cdots$ represents fields at other fixed positions. Consider the operators (wave packets operators)

$$A(\theta)^{(in,out)} = \lim_{L\to\infty} \lim_{t\to\mp\infty} \int dx \, (f_\theta(x,t)\partial_t\Psi(x,t) - \partial_t f_\theta(x,t)\Psi(x,t)) \tag{2.1}$$

with

$$f_\theta(x,t) = \exp\left[im\cosh(\theta)\, t - im\sinh(\theta)\, x - \frac{(x-\coth(\theta)t)^2}{L^2} \right] \ , \tag{2.2}$$

as well as their hermitian conjugate $A^\dagger(\theta)^{(in,out)}$. Here we take Ψ to be bosonic with spin 0; simple modifications occur for fermions and for other spins. The operators $A(\theta)^{(in)}$, $A^\dagger(\theta)^{(in)}$ satisfy canonical commutation relations, $[A(\theta)^{(in)}, A^\dagger(\theta')^{(in)}] = 2\pi\delta(\theta-\theta')$, and similarly for the operators $A(\theta)^{(out)}$, $A^\dagger(\theta)^{(out)}$. They are also eigenoperators of the Hamiltonian, $[H, A^\dagger(\theta)^{(in,out)}] = m\cosh(\theta) A^\dagger(\theta)^{(in,out)}$. Similar operators can be defined for all particles of the theory, and operators corresponding to different particle types commute with each other. The Hilbert space is the Fock space over the algebra of all such *in*-operators, which is isomorphic to the Fock space over the algebra of all *out*-operators.

In fact, in the wave-packet description above, we chose quite arbitrarily the centers of the wave packets; we have specified them by making all particles "collide at one point" under an extrapolation of their free trajectories. Defining the operators $A(\theta)^{(in,out)}$ by shifting slightly the central positions of the wave packets changes when the particles "would collide", the impact parameter – some particles could collide first. That is, we could have used the definition

$$f_\theta(x,t) = \exp\left[im\cosh(\theta)\, t - im\sinh(\theta)\, x - \frac{(x-x_0-\coth(\theta)(t-t_0))^2}{L^2} \right] \ . \tag{2.3}$$

Any choice of $x_0 - \coth(\theta)t_0$ leads to a different but legitimate basis of the Hilbert space, and the overlap between the associated *in*-states and *out*-states form a scattering matrix depending on the impact parameters.

In equations, writing from now on an index a for describing the type of particle (mass, quantum numbers) instead of explicitly writing the mass, the scattering matrix is defined by

$$\begin{aligned} &|\theta_1,\theta_2,\ldots;\text{imp. param. in}\rangle^{in}_{a_1,a_2,\ldots} = \\ &\sum_{a_1',a_2',\ldots} \int d\theta_1' d\theta_2' \cdots S_{a_1,a_2,\ldots}^{a_1',a_2',\ldots}(\theta_1,\theta_1',\theta_2,\theta_2',\cdots;\text{imp. param. in,out}) \cdot \\ &\qquad |\theta_1',\theta_2',\ldots;\text{imp. param. out}\rangle^{out}_{a_1',a_2',\ldots} \end{aligned}$$

2.2 Local conserved charges: case of the free boson

An integrable model has an infinity of local conserved charges. It is very instructive to see what they look like in a simple example. The most simple example possible of conserved charges in an integrable model are those of the free boson. The Hamiltonian of the free boson is

$$H = \int dx\, \frac{1}{2}\left[\pi^2(x) + \left(\frac{\partial}{\partial x}\phi(x)\right)^2 + m^2\phi^2(x) \right]$$

with the two local fields $\phi(x)$ and $\pi(x)$ satisfying

$$[\phi(x), \pi(x')] = i\delta(x-x') \ .$$

The momentum operator is simply expressed as

$$P = -\int dx\, \pi(x) \frac{\partial}{\partial x} \phi(x) \ .$$

Now, there are many very simple conserved charges that can be constructed:

$$P_{(n)} = (-1)^{\frac{n+1}{2}} \int dx\, \pi(x) \left(\frac{\partial}{\partial x} \right)^n \phi(x)$$

for any n positive and odd, which have action

$$[P_{(n)}, \phi(x)] = \left(i \frac{\partial}{\partial x} \right)^n \phi(x) \ , \quad [P_{(n)}, \pi(x)] = \left(i \frac{\partial}{\partial x} \right)^n \pi(x) \ .$$

Other charges are

$$H_{(n)} = (-1)^{\frac{n-1}{2}} \int dx\, \frac{1}{2} \left[\pi(x) \left(\frac{\partial}{\partial t} \right)^n \phi(x) - \phi(x) \left(\frac{\partial}{\partial t} \right)^n \pi(x) \right]$$

for odd positive n, where the time derivatives $\partial/\partial t$ should really be replaced by what is obtained from the equation of motion: $\partial/\partial t \cdot = i[H, \cdot]$. These charges have actions

$$[H_{(n)}, \phi(x)] = \left(-i \frac{\partial}{\partial t} \right)^n \phi(x) \ , \quad [H_{(n)}, \pi(x)] = \left(-i \frac{\partial}{\partial t} \right)^n \pi(x)$$

with the same meaning for the time derivatives. In similar ways, one can imagine how to construct local conserved charges whose action on the fields $\phi(x)$ and $\pi(x)$ are of the type

$$[Q_{(n,k)}, \phi(x)] = \left(i \frac{\partial}{\partial x} \right)^n \left(-i \frac{\partial}{\partial t} \right)^k \phi(x) \ , \quad [Q_{(n,k)}, \pi(x)] = \left(i \frac{\partial}{\partial x} \right)^n \left(-i \frac{\partial}{\partial t} \right)^k \pi(x)$$

with $n + k$ positive and odd. They are not all independent, because of the equations of motion. Indeed, we always have, for instance, $(-\partial^2/\partial t^2 + \partial^2/\partial x^2)\phi(x) = m^2 \phi(x)$. In order to get the set of independent charges, it is convenient to use the coordinates $z = x - t$ and $\bar{z} = x + t$. Then, one can combine the charges $Q_{(n,k)}$ to form charges that act through

$$\left(\frac{\partial}{\partial z} \right)^{n'} \left(\frac{\partial}{\partial \bar{z}} \right)^{k'} \ .$$

But clearly, $\partial/\partial z\, \partial/\partial \bar{z} = -m^2$ as a consequence of the equations of motion, so that we are left with only $\partial/\partial z$ derivatives or $\partial/\partial \bar{z}$ derivatives. Hence, the independent conserved charges are Q_s for odd (positive or negative) integer s whose action on the fields $\phi(x)$ and $\pi(x)$ are examplified by

$$[Q_s, \phi(x)] = \left(\frac{\partial}{\partial z} \right)^s \phi(x) \ (s > 0) \ , \quad [Q_s, \phi(x)] = \left(\frac{\partial}{\partial \bar{z}} \right)^s \phi(x) \ (s < 0) \ .$$

Under the action of the boost operator B, these charges have spin s, that is, $[B, Q_s] = sQ_s$. This is most intuitively seen using mode operators that we saw earlier in the most general constext, with commutation relations $[A(\theta),A^\dagger(\theta')] = 4\pi\delta(\theta-\theta')$. Here, the Hamiltonian and momentum operators are

$$H = \int \frac{d\theta}{4\pi}\, m\cosh(\theta)A^\dagger(\theta)A(\theta)\ ,\quad P = \int \frac{d\theta}{4\pi}\, m\sinh(\theta)A^\dagger(\theta)A(\theta)\ ,$$

and the conserved charges of well-defined spins that we alluded to are simply written, in terms of modes, as

$$Q_s = \int \frac{d\theta}{4\pi}\, e^{s\theta}A^\dagger(\theta)A(\theta)\ .$$

In particular, $Q_{\pm1} = m^{-1}(H\pm P)$.

Here it seems not to be possible to construct local conserved charges of even spins. For free fermions, similar arguments can be made, but in this case we will have conserved charges of all integer spins. That is, in the free theories, we have infinitely many independent local Hermitian conserved charges indexed by integer spins (we just constructed even ones for real bosons). The same type of charge occur in any free theory, with or without internal degrees of freedom.

2.3 Elastic scattering

Let us now derive one consequence of the presence of an infinity of local conserved charges in involution in interacting models. Recall that those are local charges that commute with the Hamiltonian and with each other. Besides the Hamiltonian, there is one that is trivial: the momentum. From these two charges, it is convenient, as above, to build two different, with better transformation properties under boost, and without scale:

$$Q_1 = m^{-1}(H+P)\ ,\quad Q_{-1} = m^{-1}(H-P)$$

where m is the mass of one of the particles of the theory (arbitrarily chosen, this is just a normalisation of the operator; for instance, take the lowest mass). These charges have eigenvalues

$$Q_{\pm1}|\theta_1,\theta_2,\ldots\rangle^{in,out}_{a_1,a_2,\ldots} = \sum_k q_{a_k}e^{\pm\theta_k}|\theta_1,\theta_2,\ldots\rangle^{in,out}_{a_1,a_2,\ldots}$$

for $q_{a_k} = m_{a_k}/m$ where m_{a_k} is the mass of the particle of type a_k, and transform with spin ±1 under boost.

Remember what locality of a charge essentially means: the charge density $q(x)$ commutes with the Hamiltonian density $h(x')$ at different points $x\neq x'$. In order to derive consequences of conserved charges on the scattering matrix, we need something just slightly stronger: the charge density $q(x)$ must commute with the "fundamental" fields creating particles $\Psi(x')$ at different points:

$$Q = \int dx\, q(x)\ ,\quad [q(x),\Psi(x')] = 0\quad (x\neq x')\ .$$

We say that not only the charge Q is local (that is, its density $q(x)$ is a local field), but also that it is local *with respect to the fundamental fields* (that is, its density is a field local with respect to the fundamental fields). I will come back to the various extension of the concept of locality a bit later in the course.

Then, a conserved charge applied on an asymptotic state with many particles will act independently on each of them. In fact, the conserved charges that we are interested in are Hermitian local conserved charges that act on asymptotic states (essentially very separated particles propagating freely) in the same way as conserved charges of free models (see previous sub-section) act on free particles. That is, they are conserved charges corresponding to some space-time symmetries, and can always be made to commute with possible internal symmetry charges of the model. As we saw, we can always arrange the charges to have definite spin, so that by relativistic covariance, we have, for some infinite set of spins s,

$$Q_s|\theta_1,\theta_2,\ldots\rangle^{in,out}_{a_1,a_2,\ldots} = \sum_k q^{(s)}_{a_k} e^{s\theta_k} |\theta_1,\theta_2,\ldots\rangle^{in,out}_{a_1,a_2,\ldots} \ .$$

Note that these charges are automatically in involution.

A comment about impact parameters is in order here. We are not writing them explicitly in the states above, but the action of a charge may involve a modification of the impact parameters (something which we do not see in free theories because the states are independent of impact parameters). We will analyse this in more details below, but for now we just keep in mind that we can choose them as we wich, in such a way that the conditions we derive below are valid for scattering matrices of any impact parameters.

The values of the spins s for which there is a conserved charge and the numbers $q^{(s)}_{a_k}$ are usually good fingerprint of a given integrable model. Now, in the equation above, take the *in* states, and sandwich the whole equation with the conjugate *out* state

$${}_{a'_1,a'_2,\ldots}^{\quad\quad out}\langle\theta'_1,\theta'_2,\ldots| \ .$$

We see that the overlap between an *in* state and an *out* state is non-zero only if the following equation is satisfied:

$$\sum_k q^{(s)}_{a_k} e^{s\theta_k} = \sum_k q^{(s)}_{a'_k} e^{s\theta'_k} \ .$$

This is an infinite set of equations (for all spins s occuring in the model) for an indefinite number of the rapidities θ'_k and particle types a'_k, being given the rapidities θ_k and the particle types a_k. One solution is obviously

$$\{\theta'_k\} = \{\theta_k\} \ , \quad q^{(s)}_{a'_k} = q^{(s)}_{a_k} \text{ for } \theta'_k = \theta_k \ .$$

In fact, if we require the solution for the out-going rapidities to be $\{\theta'_k\} = \{\theta_k\}$ for all sets of in-going rapidities $\{\theta_k\}$, then the only possibility for the particle types is $q^{(s)}_{a'_k} = q^{(s)}_{a_k}$ for $\theta'_k = \theta_k$. This is elastic scattering: the number of particles going out is the same as that coming in, and they have the same rapidities, with possible exchanges and possible modifications of the particle types.

We would indeed like to conclude elastic scattering, but for this we need to do a little bit more analysis from this argument. In principle there may be other solutions than $\{\theta'_k\} = \{\theta_k\}$, depending on how the numbers $q_{a_k}^{(s)}$ behave as function of s. The simplest way to see this is to construct the following analytic function of α, defined by its expansion at large α:

$$f_{a_k}(\alpha) = \sum_{s>0} e^{-s\alpha} q_{a_k}^{(s)}$$

where the sum is over all spins that occur in the model; we assume this series to be convergent for α large enough. Our condition, for positive spins, then becomes

$$\sum_k f_{a_k}(\alpha - \theta_k) = \sum_k f_{a'_k}(\alpha - \theta'_k) \;\forall\, \alpha$$

(we can do something similar for negative spins; if we have parity symmetry, it is sufficient to look at positive spins). If the function $f_{a_k}(\alpha)$ has only one singular point on the real line at the same position for all particle types a_k, then the only possibility, for any real $\{\theta_k\}$, is indeed $\{\theta'_k\} = \{\theta_k\}$.

Take, for instance, the situation where there is only one particle type, or when there are many but thay all fall into one multiplet of some internal symmetry. Then, we must have $q_{a_k}^{(s)} = q_{a'_k}^{(s)}$ so that they can all be set to 1. If, for instance, only odd spins are involved (this is the situation in the $SU(2)$-Thirring model, for instance), the analytical function $f(\alpha)$ is

$$f(\alpha) = \frac{1}{2\sinh(\alpha)} \;.$$

This indeed only has one pole on the real line, and the conclusion follows.

But take, for instance, a fictitious case where we only have odd spins, where for a particle type 1, we have $q_1^{(s)} = 1$, and where for a particle type 2, we have $q_2^{(s)} = 1 + e^{-s\beta}$ for $s > 0$. Then clearly, if there are just two *in* rapidities and if we have with $\theta_1 = \theta_2 + \beta$ and $a_1 = a_2 = 2$, then we can choose to have only one *out* rapidity with $\theta'_1 = \theta_1$ and $a'_1 = 2$, and the equation is satisfied for all α.

This fictitious situation has never been seen in any model, to my knowledge. In order to establish elastic scattering from this argument, we must check in explicit models what the $q_a^{(s)}$ look like. Admittedly, then, this argument is not as strong as we would like it to be, but generically, we see a singularity only at $\alpha = 0$ in $f(\alpha)$ for $\alpha \in \mathbb{R}$. From this we may conclude that the obvious solution is the only one for real rapidities:

$$\{\theta'_k\} = \{\theta_k\} \;, \quad q_{a'_k}^{(s)} = q_{a_k}^{(s)} \text{ for } \theta'_k = \theta_k \;.$$

2.4 Factorized scattering: Yang-Baxter equations

It is convenient here to think about the full evolution operator $\hat{S}$. It is formally $\lim_{t\to\infty} e^{-iHt}$; more precisely, when acting on an *in*-state, one must take the time in the definition of the operators $A(\theta)^{(in)}$ to be of much smaller absolute value than that of the time t above, involved in the definition of S. Then, one doesn't get only a phase (as expected naively; recall that *in*-states are eigenstates of the Hamiltonian), but rather, the state transforms into a linear combinations of states representing what is left after the scattering. More precisely,

$$\begin{aligned}&\hat{S}|\theta_1,\theta_2,\ldots;\text{imp. param. in}\rangle^{in}_{a_1,a_2,\ldots} = \\ &\sum_{a'_1,a'_2,\ldots}\int d\theta'_1 d\theta'_2\cdots S^{a'_1,a'_2,\ldots}_{a_1,a_2,\ldots}(\theta_1,\theta'_1,\theta_2,\theta'_2,\cdots;\text{imp. param. in,out})\,\cdot \\ &\qquad|\theta'_1,\theta'_2,\ldots;\text{imp. param. out}\rangle^{in}_{a'_1,a'_2,\ldots}\ .\end{aligned}$$

Note that on the right-hand side, we have a linear combination of *in*-states (we chose this basis, but we could have chosen as well the *out*-basis to define the operator $\hat{S}$). It will be convenient and sufficient for our purposes to choose the "same" impact parameters for *in* and *out* particles; that is, for given impact parameters of *in* states, we choose the impact parameters of *out* particles to be those obtained by an extension of the trajectory of the *in* particles all the way to the *out* particles as if they were not interacting. Then, the S-matrix just depends on one set of impact parameters. It is given by the matrix element

$$\begin{aligned}&S^{a'_1,a'_2,\ldots}_{a_1,a_2,\ldots}(\theta_1,\theta'_1,\theta_2,\theta'_2,\cdots;\text{imp. param.}) = \\ &{}_{a'_1,a'_2,\ldots}^{\quad\ \ in}\langle\theta'_1,\theta'_2,\ldots;\text{imp. param.}|\hat{S}|\theta_1,\theta_2,\ldots;\text{imp. param.}\rangle^{in}_{a_1,a_2,\ldots}\ .\end{aligned}$$

The main point is then to realize that the operator $\hat{S}$ commutes with all conserved charges of the theory:

$$\begin{aligned}&{}_{a'_1,a'_2,\ldots}^{\quad\ \ in}\langle\theta'_1,\theta'_2,\ldots;\text{imp. param.}|e^{i\alpha Q_s}\hat{S}e^{-i\alpha Q_s}|\theta_1,\theta_2,\ldots;\text{imp. param.}\rangle^{in}_{a_1,a_2,\ldots} = \\ &{}_{a'_1,a'_2,\ldots}^{\quad\ \ in}\langle\theta'_1,\theta'_2,\ldots;\text{imp. param.}|\hat{S}|\theta_1,\theta_2,\ldots;\text{imp. param.}\rangle^{in}_{a_1,a_2,\ldots}\ .\end{aligned} \tag{2.4}$$

We can exploit this formula by analysing in more details than above the action of a conserved charge on asymptotic states. More precisely, we will see that the scattering matrix is in fact independent of impact parameters, and this will lead to Yang-Baxter equation.

From the formula (2.1) for the operators destroying/creating asymptotic *in*-states, let us compute

$$[Q_s,A(\theta)^{(in)}] = \lim_{L\to\infty}\lim_{t\to-\infty}\int dx\,(f_\theta(x,t)\partial_t[Q_s,\Psi(x,t)] - \partial_t f_\theta(x,t)[Q_s,\Psi(x,t)])$$

where the function $f_\theta(x,t)$ is given in (2.2). We can evaluate the commutators involved, because the conserved charges act like higher-spin conserved charges on free particles. For instance, for positive spins we have

$$[Q_s,A(\theta)^{(in)}] = -\frac{(-i)^s q^{(s)}}{m^s}\lim_{L\to\infty}\lim_{t\to-\infty}\int dx\,(f_\theta(x,t)\partial_t(\partial_x-\partial_t)^s\Psi(x,t)-\partial_t f_\theta(x,t)(\partial_x-\partial_t)^s\Psi(x,t))\ .$$

We can move the derivatives $(\partial_x-\partial_t)$ on the fields to the same derivatives on the wave packet $f_\theta(x,t)$ with an overall sign $(-1)^s$. This is obvious for the space derivatives: we do this by integration by part. For the time derivatives, first make the replacement $\partial_t^n \mapsto [iH,[iH,[iH,\ldots]]]$ where there are n commutators. Taking these commutators out of the integrals and using

$[iH, A(\theta)^{(in)}] = -im\cosh(\theta)A(\theta)^{(in)}$ gives then a factor $(-im\cosh(\theta))^n$. But this is just $(-1)^n$ times n time-derivatives applied on the wave packet $f_\theta(x,t)$, which shows that we can indeed move the derivatives $(\partial_x - \partial_t)$ as said. We then get

$$-\frac{i^s q^{(s)}}{m^s} \lim_{L\to\infty} \lim_{t\to-\infty} \int dx\, ((\partial_x - \partial_t)^s f_\theta(x,t)\partial_t\Psi(x,t) - \partial_t(\partial_x - \partial_t)^s f_\theta(x,t)\Psi(x,t)) \ .$$

To leading and first sub-leading order in $1/(mL)$, we have

$$-\frac{i^s q^{(s)}}{m^s}(\partial_x - \partial_t)^s f_\theta(x,t) = -q^{(s)} e^{s\theta} \left[1 + \frac{2s}{imL\sinh(\theta)}\frac{x - \coth(\theta)t}{L} + O((mL)^{-2})\right] f_\theta(x,t) \ .$$

We recognize the leading term as giving the correct eigenvalue of Q_s on asymptotic states, but we want to analyse the effect of the additional term. In order to calculate

$$e^{i\alpha Q_s} A(\theta)^{(in)} e^{-i\alpha Q_s} \ ,$$

we compute

$$\begin{aligned} &\sum_{n=0}^{\infty} \frac{1}{n!}\left(-i\alpha\frac{i^s q^{(s)}}{m^s}\right)^n (\partial_x - \partial_t)^{ns} f_\theta(x,t) \\ &= \exp\left(-i\alpha q^{(s)} e^{s\theta}\right)\left[1 - \frac{2i\alpha q^{(s)} s e^{s\theta}}{imL\sinh(\theta)}\frac{x - \coth(\theta)t}{L} + O((mL)^{-2})\right] f_\theta(x,t) \\ &= \exp\left[-i\alpha q^{(s)} e^{s\theta} - \frac{2i\alpha q^{(s)} s e^{s\theta}}{imL\sinh(\theta)}\frac{x - \coth(\theta)t}{L} + O((mL)^{-2})\right] f_\theta(x,t) \ . \end{aligned}$$

The right-hand side can be written

$$\exp\left(-i\alpha q^{(s)} e^{s\theta}\right) \exp\left[iM\cosh(\theta)\, t - iM\sinh(\theta)\, x - \frac{(x - x_0 - \coth(\theta)(t - t_0))^2}{L^2} + O((mL)^{-2})\right]$$

with

$$m\cosh(\theta)t_0 - m\sinh(\theta)x_0 = \alpha q^{(s)} s e^{s\theta} \ .$$

Since this was obtained by exponentiation of $(\partial_x - \partial_t)^s$, we must have $x_0 = -t_0 \equiv l$.

We recognize this as, up to the factor $\exp\left(-i\alpha q^{(s)} e^{s\theta}\right)$ containing the eigenvalue of Q_s, simply the wavepacket (2.3) with non-zero impact parameters. The impact parameter is given by

$$ml = -\alpha q^{(s)} s e^{(s-1)\theta} \quad (s > 0) \ .$$

If $s = 1$, it is clearly independent of θ, but otherwise, the impact parameters depend on the rapidity θ.

A similar derivation holds for negative spins. The action of Q_s on a field $\Psi(x,t)$ in this case is

$$[Q_s, \Psi(x,t)] = -\frac{(-i)^s q^{(s)}}{m^{|s|}}(\partial_x + \partial_t)^{|s|}\Psi)(x,t) \quad (s < 0)$$

and we will have $x_0 = t_0 \equiv l$. The impact parameter then takes the value

$$ml = -\alpha q^{(s)} s e^{(s+1)\theta} \quad (s < 0) \ .$$

Hence, we found that

$$e^{i\alpha Q_s} A(\theta)^{(in)} e^{-i\alpha Q_s} = e^{-i\alpha q^{(s)} e^{s\theta}} \ A(\theta)^{(in)}\Big|_{\text{impact parameter } l}$$

and the same thing holds, for conjugate eigenvalue, for the operator $A^\dagger(\theta)^{(in)}$. Looking back at (2.4), we realize that the scattering matrix is invariant under such changes of impact parameters.

It is not *a priori* clear that using enough conserved charges, we can bring the impact parameters to any value we want, hence showing that the scattering matrix is independent of impact parameters. However, there is a simple and strong argument that can do this. First, we must realise that the 2-particle S-matrix does not depend on the impact parameters. Indeed, by convention, we chose the *out* impact parameters to agree with the *in* impact parameters (which can always be done in elastic scattering). Then, space and time translation invariance means that the 2-particle S matrix is independent of impact parameters. Consider then the 3-particle to 3-particle scattering, with scattering matrix

$$S_{a_1,a_2,a_3}^{a_1',a_2',a_3'}(\theta_1,\theta_2,\theta_3;l_1,l_2,l_3) \ .$$

Here we use elastic scattering and write only the three rapidities involved. The particle types a_1, a_2, a_3 are for the *in*-particles of rapidity $\theta_1, \theta_2, \theta_3$ respectively, and similarly for a_1', a_2', a_3' for the *out*-particles. The parameters l_1, l_2, l_3 are the associated impact parameters. Consider the set of rapidities $(\theta_1, \theta_2, \theta_3)$ for which the three impact parameters resulting from the action of Q_s,

$$-\alpha m_{a_i}^{-1} q_{a_i}^{(s)} s e^{(s-1)\theta} \ , \quad i = 1,2,3$$

are not all equal to each other. There is only possibly one set $(\theta_1, \theta_2, \theta_3)$ for which this is not so for a given spin s; for this one, we must use another conserved charge of a different spin and verify that the inequality holds. If we have parity invariance in the theory, we can use the conserved charge of spin $-s$ (we have $q_a^{(-s)} = q_a^{(s)}$ in parity invariant models), and we are guaranteed that it will work. Then, we can use invariance under shift of impact parameters for each particle, and with α large enough, the impact parameters become so different that the particle will only meet pair by pair at points very distant from each other: say particles 1 and 2 first, then particle particles 1 and 3, finally particles 2 and 3. Since these meeting points are very far from each other, by locality the 3-particle S-matrix decomposes itself into products of 2-particle S-matrices:

$$S_{a_1,a_2,a_3}^{a_1',a_2',a_3'}(\theta_1,\theta_2,\theta_3;l_1,l_2,l_3) = S_{a_1,a_2}^{c,b}(\theta_1,\theta_2) S_{c,a_3}^{a_1',d}(\theta_1,\theta_3) S_{b,d}^{a_3',a_2'}(\theta_2,\theta_3)$$

(with implicit sum over repeated indices). Since this does not depend anymore on the initial impact parameters, this shows that the scattering matrix does not depend on them. Hence, we can also choose them so that we get the opposite situation: particles 2 and 3 meet first, then particles 1 and 3, finally particles 1 and 2 (equivalently, we could have taken α large with opposite sign). The two ways of decomposing it give the same value:

$$S^{c,b}_{a_1,a_2}(\theta_1,\theta_2)S^{a_1',d}_{c,a_3}(\theta_1,\theta_3)S^{a_2',a_3'}_{b,d}(\theta_2,\theta_3) = S^{b,c}_{a_2,a_3}(\theta_2,\theta_3)S^{d,a_3'}_{a_1,c}(\theta_1,\theta_3)S^{a_1',a_2'}_{d,b}(\theta_1,\theta_2) \tag{2.5}$$

This equation (or set of equations) is called *Yang-Baxter equation.*

Now consider the multi-particle scattering. We can apply a similar argument and write it as a product of a $n-1$-particle S-matrix times a 2-particle S-matrix. Recursively, we then get a product of 2-particle S-matrices. Again this is true for any initial impact parameters, so that it does not depend on them. Repeated use of Yang-Baxter equation insures that we can write it as any decomposition into 2-particle S-matrices.

Note that we only needed *two* conserved charges of spin higher than 1 to do these manipulations! Using similar arguments, Parke (1980) was able to actually prove *elastic scattering* as well as factorizability, using two local conserved charges of higher spin.

It is worth noting that similar arguments could be attempted in higher dimension. The presence of higher-spin conserved charges indeed give there the independence form impact parameters, and in higher than on space dimension, this means that we can choose the trajectory to all avoid each other. Hence the theory has to be trivial: free fermions or free bosons. This is the essence of Coleman-Mandula theorm (1967).

2.5 Analytic and other properties of the two-particle S-matrix

Besides satisfying the Yang-Baxter equation (2.5), the two-particle S-matrix must satisfy other, simpler equations that are consequences of general properties of QFT, as well as "bootstrap" equations, which are like Yang-Baxter equations but which concern bound states.

The first equation comes from Lorentz invariance, and says that only the difference of rapidities matters:

$$S^{b_1,b_2}_{a_1,a_2}(\theta_1,\theta_2) = S^{b_1,b_2}_{a_1,a_2}(\theta_1-\theta_2)\ .$$

The second is about the fact that the transformation from *in*-states to *out*-states is unitary. From $|\theta_1,\theta_2\rangle^{(in)}_{a_1,a_2} = \sum_{b_1,b_2} S^{b_1,b_2}_{a_1,a_2}(\theta_1-\theta_2)|\theta_1,\theta_2\rangle^{(out)}_{b_1,b_2}$ and orthonormality of both *in*-states and *out*-states, we have

$$\sum_{b_1,b_2} S^{b_1,b_2}_{a_1,a_2}(\theta)\left(S^{b_1,b_2}_{c_1,c_2}(\theta)\right)^* = \delta^{c_1}_{a_1}\delta^{c_2}_{a_2} \tag{2.6}$$

(for real θ). Note that the unitarity relation for non-integrable models is not this, because there is in general more intermediate states, with more than 2 particles.

For the other properties, we need to assume time-reversal symmetry. This says that if we reverse the direction of time and if all particles are replaced by their corresponding anti-particle, then the scattering amplitudes are the same. The consequence for the two-particle scattering matrix is

$$\text{(time-reversal symmetry)}\qquad S^{b_1,b_2}_{a_1,a_2}(\theta) = S^{\bar a_2,\bar a_1}_{\bar b_2,\bar b_1}(\theta)$$

(for real rapidities). The exchange of horizontal position of the indices occurs because if a particle is travelling to the right, then under time reversal, its anti-particle travels to the left. The notation $\bar{a}$ means that we must take the particle index corresponding to the anti-particle of that of particle type a. Generically, there will be a conjugation matrix $C_{a,b}$ such that

$$A^{\bar{a}} = C_{a,b}A^b \ , \quad A_{\bar{a}} = C^{a,b}A_b \ ,$$

with property

$$C_{a,b}C^{b,c} = \delta_a^c \ .$$

The most important properties of the scattering matrix is its analytical structure. General principles of QFT (or of the theory of analytic S-matrices) say the following:

In a general model of (1+1-dimensional) QFT, as an analytic function of the Mandelstam variable

$$s = m_1^2 + m_2^2 + 2m_1m_2\cosh(\theta_1 - \theta_2) \ ,$$

the two-particle to two-particle scattering matrix (preserving the masses – such a scattering is always elastic from the kinematics) is a multi-valued function, and possesses a Riemann sheet with only three branch points where it is otherwise meromorphic. On this sheet, called the physical sheet, the branch points are at $(m_1+m_2)^2$, $(m_1-m_2)^2)$ and ∞, the cuts are on the real line avoiding the interval $[(m_1-m_2)^2, (m_1+m_2)^2]$, and the only possible poles are on this interval. This sheet is characterised by the fact that the physical values of the scattering matrix are just above the cut on the interval $[(m_1+m_2)^2, \infty]$ of the real line.

The prescription "just above the cut" for the physical values of the scattering can be understood as coming from Feynmann's prescription for the propagator. The poles on the real line, between the cuts, correspond to possible bound states between the two particles involved in the scattering. The restriction to 1+1-dimensional QFT implies that the 2-particle scattering matrix only depends on the Mandelstam variable s.

Back in the $\theta = \theta_1 - \theta_2$ plane, the physical sheet corresponds to the strip $\text{Im}(\theta) \in [0, \pi]$, the branch points are $\theta = 0, i\pi, \infty$, and the cuts run along the lines $\text{Im}(\theta) = 0, \pi$ connecting to infinity. They may be chosen to run towards the left or towards the right. The poles corresponding to bound states lie on the line $\text{Re}(\theta) = 0$ in the physical strip.

An important relation valid on the physical sheet is called "Hermitian analyticity", or when there is parity invariance, "real analyticity". It simply says that the complex conjugate of the scattering matrix on the physical sheet is the scattering matrix on the complex conjugate argument still on the physical sheet:

$$\left(S_{a_1,a_2}^{b_1,b_2}(\theta)\right)^* = S_{b_2,b_1}^{a_2,a_1}(-\theta^*) \ . \tag{2.7}$$

Notice how the real part of the rapidity changes its sign. This is because $s \mapsto s^*$ on the physical sheet corresponds to $\theta \mapsto -\theta^*$. This relation can be understood heuristically as follows. Consider the path-integral formulation of the scattering matrix, with real rapidities and $\theta_1 > \theta_2$:

$$S_{a_1,a_2}^{b_1,b_2}(\theta) = \int_{\substack{\Psi=\sum_k A_{a_k} e^{iE_k t - ip_k x} \ (t\to-\infty) \\ \Psi=\sum_k A_{b_k} e^{iE_k t - ip_k x} \ (t\to+\infty)}} [d\Psi] e^{iS[\Psi]} \ .$$

Here, the sum specifying the asymptotic conditions on the fundamental field has to be understood as giving the spacetime-dependent phase factors for wave packets ordered, on the space slice at $t=-\infty$, from left to right with increasing index k. Taking the complex conjugate of this gives

$$\left(S^{b_1,b_2}_{a_1,a_2}(\theta)\right)^* = \int_{\substack{\Psi=\sum_k A_{a_k} e^{-iE_kt+ip_kx}\ (t\to-\infty)\\ \Psi=\sum_k A_{b_k} e^{-iE_kt+ip_kx}\ (t\to+\infty)}} [d\Psi] e^{-iS[\Psi]}\ .$$

On the other hand, time-reversal invariance can be written

$$S^{b_1,b_2}_{a_1,a_2}(\theta) = S^{\bar a_2,\bar a_1}_{\bar b_2,\bar b_1}(\theta) = \int_{\substack{\Psi=\sum_k A_{a_k} e^{-iE_kt-ip_kx}\ (t\to+\infty)\\ \Psi=\sum_k A_{b_k} e^{-iE_kt-ip_kx}\ (t\to-\infty)}} [d\Psi] e^{-iS[\Psi]} \quad (\theta_1>\theta_2)\ .$$

Comparing gives (2.7) for θ real and positive, and analytical continuation proves it for θ on the physical strip.

If there is parity invariance:

$$\text{(parity invariance)}\qquad S^{b_1,b_2}_{a_1,a_2}(\theta) = S^{b_2,b_1}_{a_2,a_1}(\theta)$$

then Hermitian analyticity implies real analyticity, because we always have CPT invariance (charge-parity-time-reversal invariance) from general principles of QFT:

$$\text{(CPT invariance)}\qquad S^{b_1,b_2}_{a_1,a_2}(\theta) = S^{a_1,a_2}_{b_1,b_2}(\theta)\ .$$

Now we may combine Hermitian analyticity with the unitarity relation of integrable models derived above, in order to obtain

$$S^{b_1,b_2}_{a_1,a_2}(\theta)S^{c_2,c_1}_{b_2,b_1}(-\theta^*) = \delta^{c_1}_{a_1}\delta^{c_2}_{a_2}\ . \tag{2.8}$$

This relation is usually called "unitarity" for the 2-particle S-matrix in integrable models. It is important to recall that $-\theta^*$ means the analytical continuation from θ two $-\theta^*$ counterclockwise around the point 0 (that is, for physical initial $\theta>0$, we always stay on the physical strip to reach $-\theta$).

Another relation, again consequence of general principles of QFT, is crossing symmmetry. It essentially says that quantizing the theory in a scheme where the time and space arrows are rotated by $\pi/2$ gives the same scattering amplitudes. For the 2-particle scattering $S^{b_1,b_2}_{a_1,a_2}(\theta_1-\theta_2)$, the slope of the world line of particle k is $\coth(\theta_k)$. Making a $\pi/2$ counter-clockwise rotation amounts to $\theta_k\mapsto i\pi/2-\theta_k$, to be understood as analytical continuation on the physical strip, and after this rotation particle 2 seems like propagating in reverse time. We can then use time-reversal symmetry in order to transform it to an anti-particle propagating correctly, and we must remember that when complex values of rapidities are involved, time-reversal symmetry involves complex conjugation of the rapidities (understood, again, on the physical strip of $\theta_1-\theta_2$). Hence, we find

$$S^{b_1,b_2}_{a_1,a_2}(i\pi-\theta) = S^{\bar a_2,b_1}_{\bar b_2,a_1}(\theta) \tag{2.9}$$

Now it is possible to understand the nature of the branch points at $\theta=0,i\pi$ on the physical strip. Start at $\theta>0$ real, and use the unitarity relation (2.8) in order to travel to $\theta<0$ real, always staying on the physical strip. Using again (2.8) with $\theta < 0$ real, we get the analytical continuation all the way around the point 0. But this is equal to the initial value:

$$S_{a_1,a_2}^{b_1,b_2}(\theta)S_{b_2,b_1}^{c_2,c_1}(-\theta^*)[S_{c_1,c_2}^{d_1,d_2}(\theta)]_{\text{anal. cont. around } 0} = S_{a_1,a_2}^{d_1,d_2}(\theta) = [S_{a_1,a_2}^{d_1,d_2}(\theta)]_{\text{anal. cont. around } 0}$$

(we used unitarity in two different ways for contracting the two different pairs of S-matrices). Hence, the point $\theta = 0$ is in fact not a branch point, it is a regular point without any singularity; this means that the corresponding branch point in the s-plane is a square-root branch point. By crossing symmetry, the same is true for the point $\theta = i\pi$.

This means that the S-matrix is a meromorphic function of θ, its poles on the physical strip can only be on the imaginary axis, and once it is determined on the physical sheet, it is determined everywhere on the θ-plane. Also, if the poles and their residues are known on the physical strip and on the strip $\text{Im}(\theta) \in [-\pi, 0]$, then it is determined everywhere.

Finally, there are additional conditions coming from possible bound states. A bound state manifests itself by a virtual particle being created in a two-particle scattering process, and gives rise to a pole in the S-matrix at the corresponding imaginary value of the rapidity. This value is sole consequence of the kinematics. For two particles of masses m_1 and m_2 forming a particle of mass m, it is

$$\cosh(\theta) = \frac{m^2 - m_1^2 - m_2^2}{2m_1m_2} \ . \tag{2.10}$$

The residue at the pole is purely imaginary, iR. If $R > 0$, this is a bound state in the "direct" channel, if $R < 0$, it is in the "crossed" channel (as if the virtual particle were travelling faster than the speed of light!). This would not be of much use if it were not for the resulting "bound-state Yang-Baxter" relations. They are not as easy to derive from local conserved charges, and I will not go into any detail of there derivation; but they are easy to state. They can just be seen as consistency in the different ways of decomposing a 3-particle scattering, where two of the particles (say particle 1 and 2) form a virtual bound-state particle, into products of 2-particle scattering. The assumption is that any bound state is a particle that is part of the asymptotic state spectrum (this assumption is called "nuclear democracy"), hence particle 3 can scatter with the bound-state of particles 1 and 2 as if it were an asymptotic particle, and this must give the same result as when particle 3 scatters with particles 1 then 2 separately.

2.6 Zamolodchikov-Faddeev algebra

A convenient way of representing the properties of factorized scattering is using the Zamolodchikov algebra. It is the exchange algebra generated by elements $Z_a(\theta)$, which are parametrised by particle type a and rapidity θ. This algebra is used to represent states in integrable QFT by associating to a state an element of the enveloping algebra of Zamolodchikov's algebra. The horizontal position of each factor corresponds to the position of the particles in the scattering process at a given time. For instance, the *in*-state, where wave packets are ordered from the most positive rapidity at the left to the most negative one at the right, is represented by the product (an element of the enveloping algebra)

$$Z_{a_1}(\theta_1)\cdots Z_{a_n}(\theta_n)\ , \quad \theta_1 > \cdots > \theta_n\ .$$

Since, as we saw, we can choose the impact parameters so that scattering occurs 2 particles by 2 particles at points very separated from each other, we can define "intermediate states", which are neither *in* nor *out*, where wave packets travel freely and are ordered in such a way that some may never meet in the future (going in opposite directions after having scattered already). Such states would not make sense in ordinary quantum field theory, or at least would be of no interest, because in this case such a situation never happens by construction (the asymptotic states are initially defined such that particles meet in a finite region of space-time); it only happens, or is useful, thanks to independence from impact parameters. Then, in general, such intermediate states will be represented by

$$Z_{a_1}(\theta_1)\cdots Z_{a_n}(\theta_n)\ , \quad \text{any ordering of } \theta_1,\ldots,\theta_n\ .$$

Physical scattering tells us that the Zamolodchikov algebra elements must satisfy the exchange relation

$$Z_{a_1}(\theta_1)Z_{a_2}(\theta_2) = S_{a_1,a_2}^{b_1,b_2}(\theta_1-\theta_2)Z_{b_2}(\theta_2)Z_{b_1}(\theta_1)\ , \quad \theta_1 > \theta_2\ . \tag{2.11}$$

In order to verify associativity of this algebra, it is sufficient to check that the two ways of obtaining $Z_{a_3}(\theta_3)Z_{a_2}(\theta_2)Z_{a_1}(\theta_1)$ from $Z_{a_1}(\theta_1)Z_{a_2}(\theta_2)Z_{a_3}(\theta_3)$ are consistent. They are indeed consistent thanks to the Yang-Baxter equation (2.5).

The relation (2.11) was written for $\theta_1 > \theta_2$. The exchange relation for $\theta_1 < \theta_2$ follows from it by exchanging θ_1 and θ_2, but if we want to write it in the same form as (2.11), we need to define $S_{a_1,a_2}^{b_1,b_2}(\theta_1-\theta_2)$ for $\theta_1 < \theta_2$. The algebra tells us that it should be defined such that the following equation is sastisfied:

$$S_{a_1,a_2}^{b_1,b_2}(\theta_1-\theta_2)S_{b_2,b_1}^{c_2,c_1}(\theta_2-\theta_1) = \delta_{a_1}^{c_1}\delta_{a_2}^{c_2}\ .$$

This is just a consequence of the algebra, but we may wonder is the analytical continuation of the physical 2-particle scattering matrix $S_{a_1,a_2}^{b_1,b_2}(\theta_1-\theta_2)$ from the region $\theta_1 > \theta_2$ to the region $\theta_1 < \theta_2$ would give a function that satisfies this relation. It is not *a priori* clear that this must be so (and certainly it is not true of a 2-particle scattering matrix in non-integrable models), but since the states we constructed are actual asymptotic-like state of integrable models, we could expect that the relation above is true for the 2-particle scattering matrix as an analytical function of the rapidities. Comparison with the previous sub-section shows that it is so.

It is just a small step from representing states using the enveloping algebra of Zamolodchikov's algebra to constructing the Hilbert space as a module for a slightly extended algebra, Zamolodchikov-Faddeev algebra. It is defined by the relations

$$\begin{aligned}
Z_{a_1}(\theta_1)Z_{a_2}(\theta_2) - S_{a_1,a_2}^{b_1,b_2}(\theta_1-\theta_2)Z_{b_2}(\theta_2)Z_{b_1}(\theta_1) &= 0 \\
\bar{Z}^{a_1}(\theta_1)\bar{Z}^{a_2}(\theta_2) - S_{b_1,b_2}^{a_1,a_2}(\theta_1-\theta_2)\bar{Z}^{b_2}(\theta_2)\bar{Z}^{b_1}(\theta_1) &= 0 \\
Z_{a_1}(\theta_1)\bar{Z}^{a_2}(\theta_2) - S_{b_2,a_1}^{a_2,b_1}(\theta_2-\theta_1)\bar{Z}^{b_2}(\theta_2)Z_{b_1}(\theta_1) &= \delta_{a_1}^{a_2}\delta(\theta_1-\theta_2)\ .
\end{aligned} \tag{2.12}$$

The Hilbert space can then be constructed as a Fock space over this algebra, with vacuum defined by $\bar{Z}^a(\theta)|\text{vac}\rangle = 0$, and with conjugation defined by $Z_a(\theta)^\dagger = \bar{Z}^a(\theta)$. Consistency of these relations is a consequence of Yang-Baxter relation (2.5) and unitarity (2.8), and consistency with the Hermitian structure is consequence of Hermitian analyticity (2.7).

2.7 Simple examples, CDD ambiguity, and defining QFT from the scattering matrix

2.7.1 Recapitulation of the requirements

The problem of finding the scattering matrix is now reduced to the problem of solving a set of equations along with analytical conditions (a Riemann-Hilbert problem). We want to find a function $S_{a_1,a_2}^{b_1,b_2}(\theta)$ that is meromorphic and that satisfies:

1. Yang-Baxter equation (2.5);
2. Hermitian analyticity (2.7);
3. unitarity (2.8);
4. crossing symmetry (2.9);
5. bound state consistency relations when poles on the physical strip (2.10) are present.

The solutions to this problem can be characterized by the possible structure of solutions to the Yang-Baxter equations.

2.7.2 A diagonal example

The simplest structure is diagonal scattering: $S_{a_1,a_2}^{b_1,b_2}(\theta) = \delta_{a_1}^{b_1}\delta_{a_2}^{b_2} S_{a_1,a_2}(\theta)$ (no sum over repeated indices!). This trivially solves Yang-Baxter relations, and we are left only with the other requirements. The most general solution to requirements of points 2, 3 and 4 above is the following:

$$S_{a,b}(\theta) = \prod_{x\in X_{a,b}} \frac{\sinh\left(\frac{1}{2}(\theta + i\pi x)\right)}{\sinh\left(\frac{1}{2}(\theta - i\pi x)\right)} \tag{2.13}$$

where $X_{a,b}$ is a set of a finite number of complex numbers lying, for instance, in the strip $\text{Re}(x) \in [-1,1]$. If these numbers are all in the strip $\text{Re}(x) \in [-1,0]$, then there are no poles on the physical strip, and this immediately is a valid scattering matrix! An example of such a model is the so-called sinh-Gordon model. It is a model with only one particle in the spectrum of asymptotic states and, besides the mass of this particle, one additional free real parameter b. The scattering amplitude is

$$S(\theta) = \frac{\tanh\left(\frac{\theta}{2} - \frac{i\pi b^2}{2(b^2+1)}\right)}{\tanh\left(\frac{\theta}{2} + \frac{i\pi b^2}{2(b^2+1)}\right)}$$

The classical action of the sinh-Gordon model is written in terms of a single real b1osonic field ϕ:

$$\mathcal{A} = \int dx\,dt\,\left\{\frac{1}{16\pi}\left[(\partial_t\phi)^2-(\partial_x\phi)^2\right]-2\mu\cosh(b\phi)\right\}\ .$$

The mass of the particle is proportional to a power of μ. The fact that the scattering amplitude above corresponds to this action is really a conjecture, but I will discuss below possible ways of going from classical action to a solution to the equations for the scattering matrix.

If poles are present on the physical strip, we must additionaly solve the consistency relations for bound states. I will not go at all into this quite extensive subject...

2.7.3 A non-diagonal example

Another class of solutions to the Yang-Baxter equations are those where two types particles (that we will characterize by two charges $+/-$), scatter into one another, with scattering matrix of the form

$$S(\theta) = \begin{array}{cc} \begin{array}{cccc} ++ & +- & -+ & -- \end{array} & \\ \left(\begin{array}{cccc} a(\theta) & 0 & 0 & 0 \\ 0 & b(\theta) & c(\theta) & 0 \\ 0 & c(\theta) & b(\theta) & 0 \\ 0 & 0 & 0 & a(\theta) \end{array}\right) & \begin{array}{c} ++ \\ +- \\ -+ \\ -- \end{array} \end{array}$$

The pairs of signs on the right are the lower indices, and those on the top are the upper indices of the scattering matrix $S_{a_1,a_2}^{b_1,b_2}(\theta)$. The Yang-Baxter equation then amounts to only two equations:

$$\begin{aligned} a(\theta_{12})b(\theta_{13})c(\theta_{23}) &= b(\theta_{23})c(\theta_{12})c(\theta_{13})+a(\theta_{13})b(\theta_{12})c(\theta_{23}) \\ a(\theta_{12})a(\theta_{23})c(\theta_{13}) &= a(\theta_{13})c(\theta_{12})c(\theta_{23})+b(\theta_{12})b(\theta_{23})c(\theta_{13}) \end{aligned}$$

where $\theta_{ij}\equiv\theta_i-\theta_j$. From this, one then needs to solve the four other requirements written above. A solution that has no pole on the physical strip is that of the so-called $SU(2)$-Thirring model. It is a model with two particles of equal mass and of opposite $SU(2)$ spin, transforming under the fundamental representation of $SU(2)$, without any other free parameter than the mass and having a scattering matrix as above with

$$a(\theta)=\frac{\Gamma\left(\frac{1}{2}-\frac{i\theta}{2\pi}\right)\Gamma\left(\frac{i\theta}{2\pi}\right)}{\Gamma\left(\frac{1}{2}+\frac{i\theta}{2\pi}\right)\Gamma\left(-\frac{i\theta}{2\pi}\right)}\ ,\quad b(\theta)=a(\theta)\frac{\theta}{i\pi-\theta}\ ,\quad c(\theta)=a(\theta)\frac{i\pi}{i\pi-\theta}\ .$$

The classical action of this model is written in terms of a Dirac spinor Ψ with extra index in the fundamental representation of $SU(2)$:

$$\mathcal{A}=\int dx\,dt\,\left(\bar\Psi\gamma^\mu\partial_\mu\Psi-\frac{g}{2}\sum_m\bar\Psi\gamma_\mu\sigma^m\Psi\bar\Psi\gamma^\mu\sigma^m\Psi\right)$$

where σ^m are Pauli matrices and γ^μ are Dirac gamma-matrices. The parameter g, which is classically dimensionless, is in fact a running coupling constant under the action of the renormalisation group, and the scale associated to this running gives rise to the physical mass of the particle; this process is called “dimensional transmutation”.

2.7.4 CDD ambiguity

It is important to note that in general, any solution to the 5 requirements above can always be multiplied by a factor (2.13) for $\text{Re}(x) \in [-1,0]$, giving another solution. This is called a CDD factor (from Castillejo, Dalitz and Dyson, 1956), and the resulting ambiguity for the scattering matrix is the CDD ambiguity. The sinh-Gordon example is a pure CDD factor[1]. Resolving the CDD ambiguity cannot be done solely using the techniques of factorized scattering theory; one needs to do perturbative calculations of the scattering matrix or other, non-perturbative checks.

2.7.5 From classical action to scattering matrix

We have seen some examples of solutions to the set of requirements for the scattering matrix. We also related these examples to classical actions. How to figure out such a relation? This is in general a very complicated problem for which there is no general solution. However, we can often go a long way towards its solution in certain situations.

1. Sometimes it is possible to determine the spectrum of asymptotic particles (masses and types) from semi-classical calculations on the classical action. This occurs when we can write down explicitly *classical solutions* to the equation of motion. Every asymptotic state corresponds to a given solution to the classical equation of motion. Indeed, suppose we have an action depending on a field ϕ (hence an equation of motion for this field). Then we can formulate the matrix elements of operators between two given asymptotic states via a path integral with asymptotic conditions on ϕ at $t = \pm\infty$ corresponding to the large-time limit of appropriate classical solutions[2]. The most pictorial situation is when the classical solutions are soliton-like. Then, we understand that the asymptotic particles are actually just quantization of the solitons (bumps or kinks that seem to propagate locally in classical solutions without being affected by one another at large distances)[3].

If both *in* and *out* asymptotic states are the same, then, in order to evaluate approximately these matrix elements, we can just make a saddle-point expansion of the integrand around the required classical solution; this is the WKB approximation scheme. Of course this is an approximate method in general, but the magic occurs when we want to calculate the masses of the particles and of the bound states that they can form using WKB approximation: it turns out that the results to leading order are exact in integrable models! There is no proof of that, it is more an “empirical” result, but there is an understanding as follows. Recall that in integrable models of classical mechanics (where everything is clear), the orbit on phase space

[1] Hence, its scattering amplitude is just a CDD factor times the scattering amplitude of a free massive Majorana fermion!

[2] To make a connection with the previous description of asymptotic states via a “fundamental fields” creating particles, a fundamental field for some particles is essentially a field such that its large-time asymptotic in some classical solutions looks, in a sense that can be made precise, like separated free plane waves.

[3] Again in connection to the concept of “fundamental field”, the field, say ϕ, whose classical configurations are solitons is not the “fundamental field” associated to the particles corresponding to its solitons. Indeed, solitons do not look like plane waves at large time! Generically, the fundamental field for such particles is a very complicated functional of ϕ.

always falls onto tori. But we know that for systems of this type, the "old quantization method" from Bohr, by which we simply quantize periodic orbits by requiring that the total phase space area be a multiple of $\hbar$, is actually correct: take the example of the hydrogen atom (which is an integrable model). This old quantization method is essentially the leading order of the WKB approximation. Transferring all this to integrable QFT (neglecting the details about the infinity of degrees of freedom!) we may understand the result above. Once we have the masses of the particles and the bound states they can form, there is little ambiguity in the solution to the requirements above that we must take, except mainly for the CDD ambiguity. To fix all ambiguities, one generecally needs to perform some perturbative calculations or checks of other types.

2. Another way of making quite explicit calculations from the classical action is by applying the methods of "coordinate Bethe ansatz." This aims at solving the system explicitly by constructing the Hilbert space starting from a simple reference state ("pseudo-vacuum") and by making a simple ansatz for the excited states. After lengthy and tedious calculation this sometimes leads to the full spectrum and even to the scattering matrix itself.

3. Finally, it is also sometimes possible to have an explicit, integrable "lattice regularisation" of the QFT, or quantum-chain regularisation. Integrable quantum chains and lattice models can be solved by the method of algebraic Bethe ansatz, and, although from this solution it is still very hard to go back to the QFT model and often not known how to do that, we may still have enough intuition about the spectrum in order to determine the scattering matrix.

2.7.6 The "scattering matrix" formulation of QFT

Recall how I defined a QFT: a Hilbert space with a local Hamiltonian bounded from below and local space-time symmetry generators. Now that we have an idea about how to go from a given Hamiltonian to the scattering matrix, it is worth noting that a local massive QFT can equivalently be defined by giving a set particles and masses (hence having the Hilbert space and the action of the Hamiltonian on it) and a scattering matrix with specific properties. The specific properties are complicated in general (this is the theory of the analytic S-matrix), but in integrable models, they are just as described above. Note that just giving a set of particles and masses, although it directly gives the Hilbert space and the action of the Hamiltonian on it, does not say anything about the Hamiltonian density, which is essential for describing *locality* of the QFT. Essentially, all the locality properties of the QFT are hidden into the scattering matrix. This is quite non-trivial; for instance, how do we define local fields if we don't have an explicit Hamiltonian density? How do we calculate their correlation functions? These questions are what is assessed in the next chapter.

3 Form factors

We now start with the assumption that we know the exact scattering matrix, and that it is elastic and factorizable. Of course, the scattering matrix is not directly useful for most calculations related to experimental situations. The objects that are of direct importance are the correlation functions of local fields. In particular, the two-point function,

$$\langle \text{vac}|\mathcal{O}_1(x,t)\mathcal{O}_2(0,0)|\text{vac}\rangle \ ,$$

is a quantity often required, as it is related to the response function of the system at one point once it is disturbed at another point. The calculation of two-point functions generated a lot of research in integrable QFT. Probably the most fruitful idea is to start from a representation of the two-point function coming from inserting a complete set of energy eigenstates between the operators:

$$\begin{aligned}
&\langle \text{vac}|\mathcal{O}_1(x,t)\mathcal{O}_2(0,0)|\text{vac}\rangle \\
&\quad = \sum_{n=0}^{\infty}\sum_{a_1,\ldots,a_n}\int \frac{d\theta_1\cdots d\theta_n}{(4\pi)^n n!}\langle \text{vac}|\mathcal{O}_1(x,t)|\theta_1,\ldots,\theta_n\rangle^{in}_{a_1,\ldots,a_n}\ {}^{in}_{a_1,\ldots,a_n}\langle\theta_1,\ldots,\theta_n|\mathcal{O}_2(0,0)|\text{vac}\rangle \\
&\quad = \sum_{n=0}^{\infty}\sum_{a_1,\ldots,a_n}\int \frac{d\theta_1\cdots d\theta_n}{(4\pi)^n n!}e^{-iE_nt+ip_nx}\langle \text{vac}|\mathcal{O}_1(0,0)|\theta_1,\ldots,\theta_n\rangle^{in}_{a_1,\ldots,a_n}\ {}^{in}_{a_1,\ldots,a_n}\langle\theta_1,\ldots,\theta_n|\mathcal{O}_2(0,0)|\text{vac}\rangle
\end{aligned}$$

where

$$E_n = \sum_{k=1}^{n} m_{a_k}\cosh(\theta_k)\ , \quad p_n = \sum_{k=1}^{n} m_{a_k}\sinh(\theta_k)\ .$$

This is the basis for the usual Källén-Lehmann spectral decomposition. The *in*-states are, as usual, wave packets at minus infinite time with particles ordered from left to right by decreasing rapidity (so that above, the leftmost particle is not necessarily θ_1). Of course, we could have used as well the *out* basis.

In integrable systems, we know that states $|\theta_1,\ldots,\theta_n\rangle$ can be defined for any ordering of the rapidities by asymptotic-like states that occur between two-body scattering events far apart from each other. When $\theta_1 > \cdots > \theta_n$ they agree with the *in*-state, when $\theta_1 < \cdots < \theta_n$ they agree with the *out*-states, but for other orderings, they just form a different basis for the Hilbert space. It is not hard to see, from the unitarity of the scattering matrix written as (2.6), that we can just use all these bases and symmetrise over the rapidities:

$$\begin{aligned}
&\langle \text{vac}|\mathcal{O}_1(x,t)\mathcal{O}_2(0,0)|\text{vac}\rangle \\
&\quad = \sum_{n=0}^{\infty}\sum_{a_1,\ldots,a_n}\int \frac{d\theta_1\cdots d\theta_n}{(4\pi)^n n!}e^{-iE_nt+ip_nx}\langle \text{vac}|\mathcal{O}_1(0,0)|\theta_1,\ldots,\theta_n\rangle_{a_1,\ldots,a_n}\ {}_{a_1,\ldots,a_n}\langle\theta_1,\ldots,\theta_n|\mathcal{O}_2(0,0)|\text{vac}\rangle\ .
\end{aligned}$$

The matrix elements

$$F^{\mathcal{O}}_{a_1,\ldots,a_n}(\theta_1,\ldots,\theta_n) = \langle \text{vac}|\mathcal{O}(0,0)|\theta_1,\ldots,\theta_n\rangle_{a_1,\ldots,a_n}$$

involved in this expression are called *form factors*. They are really a generalisation of the form factors used in usual QFT, because of the intermediate bases that we have. The main point is that these matrix elements are actually meromorphic in the rapidities. They have the following properties, which form what is called a *Riemann-Hilbert problem*:

1. Meromorphicity: as functions of the variable $\theta_i - \theta_j$, for any $i,j \in \{1,\ldots,n\}$, they are analytic inside $0 < \mathrm{Im}(\theta_i - \theta_j) < 2\pi$ except for simple poles;

2. Relativistic invariance:

$$F^{\mathcal{O}}_{a_1,\ldots,a_n}(\theta_1+\beta,\ldots,\theta_n+\beta) = e^{s(\mathcal{O})\beta} F^{\mathcal{O}}_{a_1,\ldots,a_n}(\theta_1,\ldots,\theta_n)$$

   where $s(\mathcal{O})$ is the spin of $\mathcal{O}$;

3. Generalized Watson's theorem:

$$F^{\mathcal{O}}_{a_1,\ldots,a_j,a_{j+1},\ldots,a_n}(\theta_1,\ldots,\theta_j,\theta_{j+1},\ldots,\theta_n) = S^{b_j,b_{j+1}}_{a_j,a_{j+1}}(\theta_j-\theta_{j+1}) F^{\mathcal{O}}_{a_1,\ldots,b_{j+1},b_j,\ldots,a_n}(\theta_1,\ldots,\theta_{j+1},\theta_j,\ldots,\theta_n)$$

4. Locality:

$$F^{\mathcal{O}}_{a_1,\ldots,a_{n-1},a_n}(\theta_1,\ldots,\theta_{n-1},\theta_n+2\pi i) = (-1)^{f_{\mathcal{O}} f_\Psi} e^{2\pi i \omega(\mathcal{O},\Psi)} F^{\mathcal{O}}_{a_n,a_1,\ldots,a_{n-1}}(\theta_n,\theta_1,\ldots,\theta_{n-1})$$

   where $f_{\mathcal{O}}$ is 1 if $\mathcal{O}$ is fermionic, 0 if it is bosonic, Ψ is the fundamental field associated to the particle a_n, and $\omega(\mathcal{O},\Psi)$ is the *semi-locality index* (or mutual locality index) of $\mathcal{O}$ with respect to Ψ (to be defined below);

5. Kinematic pole: as function of the variable θ_n, there are poles at $\theta_j + i\pi$ for $j \in \{1,\ldots,n-1\}$, with residue

$$\begin{aligned} &iF^{\mathcal{O}}_{a_1,\ldots,a_n}(\theta_1,\ldots,\theta_n) \sim C_{a_n,b_j} \frac{F_{a_1,\ldots,\hat{a}_j,\ldots,a_{n-1}}(\theta_1,\ldots,\hat{\theta}_j,\ldots,\theta_{n-1})}{\theta_n-\theta_j-i\pi} \times \\ &\Big(\delta^{b_1}_{a_1}\cdots\delta^{b_{j-1}}_{a_{j-1}} S^{b_{j+1},c_j}_{a_{j+1},a_j}(\theta_{j+1}-\theta_j) S^{b_{j+2},c_{j+1}}_{a_{j+2},c_j}(\theta_{j+2}-\theta_j)\cdots S^{b_{n-1},b_j}_{a_{n-1},c_{n-3}}(\theta_{n-1}-\theta_j) - \\ &(-1)^{f_{\mathcal{O}} f_\Psi} e^{2\pi i\omega(\mathcal{O},\Psi)} \delta^{b_{n-1}}_{a_{n-1}}\cdots\delta^{b_{j+1}}_{a_{j+1}} S^{c_j,b_{j-1}}_{a_j,a_{j-1}}(\theta_j-\theta_{j-1}) S^{c_{j-1},b_{j-2}}_{c_j,a_{j-2}}(\theta_j-\theta_{j-2})\cdots S^{b_j,b_1}_{c_3,a_1}(\theta_j-\theta_1)\Big) \end{aligned}$$

   where a hat means omission of the argument.

6. Bound-state poles: there are additional poles in the strip $0 < \mathrm{Im}(\theta_i - \theta_j) < \pi$ if bound states are present, and these are the only poles in that strip (I will not go into any detail about this).

I will not say anything about point 6, except that something similar occurs here as for the scattering matrix when bound states are present. It is believe that the set of all solutions to points 1 to 6 form the set of all local fields of an integrable QFT. We see, then, that the scattering matrix is enough to define local fields (if we add the property of crossing symmetry, discussed below, in order to have matrix elements with excited states on both sides), and eventually to evaluate their correlation funcitons (through the form factor expansion).

Before going into an explanation of points 1 to 5, though, I have to introduce the concept of semi-locality.

3.1 Local fields and semi-locality

As I said above, a local field $\mathcal{O}(x)$ is by definition a field that commutes with the Hamiltonian density $h(x)$ at space-like distances:

$$[h(x),\mathcal{O}(x')] = 0 \quad (x \neq x') \ .$$

I also introduced the concept of respective locality: a field $\mathcal{O}_1(x)$ is local with respect to another field $\mathcal{O}_2(x)$ if they commute (for bosonic fields) or anti-commute (for fermionic fields) at space-like distances:

$$[\mathcal{O}_1(x),\mathcal{O}_2(x')] = 0 \quad \text{or} \quad \{\mathcal{O}_1(x),\mathcal{O}_2(x')\} = 0 \quad (x \neq x') \ .$$

In integrable QFT, a concept attached to local fields, as important as their anomalous dimension or spin, is that of semi-locality index. Two fields $\mathcal{O}_1, \mathcal{O}_2$ are *semi-local* with respect to each other with index $\omega_{1,2}$ if they satisfy

$$\mathcal{O}_1(x)\mathcal{O}_2(x') = (-1)^{f_1 f_2} e^{-2\pi i \omega_{1,2}\Theta(x-x')}\mathcal{O}_2(x')\mathcal{O}_1(x) \quad (x \neq x')$$

or if they satisfy

$$\mathcal{O}_1(x)\mathcal{O}_2(x') = (-1)^{f_1 f_2} e^{2\pi i \omega_{1,2}\Theta(x'-x)}\mathcal{O}_2(x')\mathcal{O}_1(x) \quad (x \neq x')$$

where $\Theta(x-x')$ is Heaviside's step function (1 for $x > x'$, zero for $x < x'$) and $(-1)^{f_1 f_2}$ is (-1) if both operators are fermionic, 1 otherwise.

Recall that in Feynmann's path integral formulation of QFT, a product of operators inside a vacuum expectation value $\langle \text{vac}|\mathcal{O}_1(x_1,t_1)\cdots\mathcal{O}_n(x_n,t_n)|\text{vac}\rangle$ has to be time-ordered $t_1 > \cdots > t_n$ in order to be represented as the functional integral

$$\int [d\Psi] e^{iS[\Psi]}\mathcal{O}_1(x_1,t_1)\cdots\mathcal{O}_n(x_n,t_n) \ .$$

From this, it is clear that semi-locality just says that the following holds inside any functional integral with other fields at times different from 0:

$$\begin{aligned}
&\lim_{\varepsilon\to0^+}\left[\int[d\Psi]e^{iS[\Psi]}(\cdots\mathcal{O}_1(x,\varepsilon)\mathcal{O}_2(x',0)\cdots)\right] \\
&\quad = e^{\pm 2\pi i\omega_{1,2}\Theta(\pm(x'-x))}\lim_{\varepsilon\to0^+}\left[\int[d\Psi]e^{iS[\Psi]}(\cdots\mathcal{O}_1(x,-\varepsilon)\mathcal{O}_2(x',0)\cdots)\right] \qquad (x\neq x')
\end{aligned}$$

where the signs ± are synchronized (and the equation holds for one choice of sign only), this being valid for bosonic as well as fermionic fields.

Recall also that in Feynmann's path integral formalism, it is necessary to put a slight imaginary part to the time variables. This means that the equation above will be valid also when the small real parameter ε is replaced by $-i\varepsilon$ in the arguments of the fields, the correlation functions being defined by analytical continuation to imaginary time.

$$\begin{aligned}
&\lim_{\varepsilon\to0^+}\left[\int[d\Psi]e^{iS[\Psi]}(\cdots\mathcal{O}_1(x,-i\varepsilon)\mathcal{O}_2(x',0)\cdots)\right] \qquad (3.1)\\
&\quad = e^{\pm 2\pi i\omega_{1,2}\Theta(\pm(x'-x))}\lim_{\varepsilon\to0^+}\left[\int[d\Psi]e^{iS[\Psi]}(\cdots\mathcal{O}_1(x,-i\varepsilon)\mathcal{O}_2(x',0)\cdots)\right] \qquad (x\neq x')
\end{aligned}$$

Then, since with an imaginary time the distance between two operators is always space-like, we can re-formulate semi-locality in a much clearer way. Consider the functional integral

$$F(x,t) = \int [d\Psi] e^{iS[\Psi]} (\cdots \mathcal{O}_1(x,t)\mathcal{O}_2(0,0)\cdots)$$

as a function of x and t, and analytically continue it in the variable t from a space-like region $|x| > |t|$ to purely imaginary values of time $t = -i\tau$, real τ. The action $S[\Psi]$ (or its analytical continuation to imaginary times) gives rise to a natural notion of local real displacement of fields in x and τ through the equations of motion. Then, semi-locality says that the function F "locally displaced" from a point (x,τ) around the point $x = 0, \tau = 0$ counterclockwise all the way back to x, τ gives rise to a phase:

$$F(e^{i\alpha} z, e^{-i\alpha}\bar{z})\Big|_{\alpha:0\mapsto 2\pi} = e^{2\pi i\omega_{1,2}} F(z,\bar{z})$$

where we use the variables $z = x + i\tau$, $\bar{z} = x - i\tau$. To be more precise, the function $F(z,\bar{z})$ in fact should have a cut on $z \in \mathbb{R}^+$ or on $z \in \mathbb{R}^-$ according to (3.1). But "local displacements" means that we consider its continuation through that cut obtained from the (finite) limit as we approach the cut of its normal derivative (given by the equation of motion); in other words, we consider its value on a covering of the Euclidean plane with a branch point at $(0,0)$.

How can such fields exist in a model? Not all models possess fields that exhibit the property of semi-locality. Important cases where such fields exist are when the model has a global $U(1)$ (or a subgroup thereof) symmetry. In order to understand, it is simpler to go to the Euclidean theory: consider the analytical continuation to imaginary times as described above for all fields in a correlation function. Then, correlation functions are described by the functional integral (we denote $\mathcal{O}^E(x,\tau)$ the fields in Euclidean theory corresponding to $\mathcal{O}(x,-i\tau)$ understood by analytical continuation as explained above, and we denote by S_E the Euclidean action obtained from analytical continuation)

$$\int [d\Psi] e^{-S_E[\Psi]} \mathcal{O}_1^E(x_1,\tau_1)\cdots\mathcal{O}_n^E(x_n,\tau_n)\ .$$

Suppose the action $S_E[\Psi]$ has a $U(1)$ symmetry; take for instance Ψ to be a complex boson, the $U(1)$ symmetry acting by $\Psi \mapsto e^{i\alpha}\Psi$. Consider a field $\mathcal{O}^E$, called *twist field* (associated to the $U(1)$ symmetry); it simply inserts of a source of magnetic charge, dual to the electric charge associated to the $U(1)$ symmetry. It is defined as follows (here we must imagine that there is another magnetic charge at infinity; it doesn't matter because this other field at infinity really just affects the normalisation of correlation functions):

$$\int [d\Psi] e^{-S_E[\Psi]} \mathcal{O}^E(0,0)\mathcal{O}_1^E(x_1,\tau_1)\cdots\mathcal{O}_n^E(x_n,\tau_n) = \int_{\mathcal{C}(0,0)} [d\Psi] e^{-S_E[\Psi]} \mathcal{O}_1^E(x_1,\tau_1)\cdots\mathcal{O}_n^E(x_n,\tau_n) \tag{3.2}$$

where $\mathcal{C}(0,0)$ is a quasi-periodicity condition on the fields Ψ in the functional integral, stating that

$$\mathcal{C}(0,0)\ :\ \left\{ \begin{array}{l} \Psi(x,0^-) = e^{2\pi i\omega}\Psi(x,0^+) \quad (x>0) \\ \Psi(x,0^-) = \Psi(x,0^+) \quad (x<0)\ . \end{array} \right. \tag{3.3}$$

More precisely, the functional integral is defined by taking away the ray $x > 0$, evaluating the functional integral with fixed boundary conditions on each side of the cut with (3.3) satisfied, then integrating over all the configurations on each sides of the cut keeping (3.3) satisfied. It is obvious that the field $\mathcal{O}^E$ thus defined has semi-locality index ω with respect to the fundamental field Ψ, and semi-locality index 0 with respect to any field that is $U(1)$-invariant. Also, the field $\mathcal{O}^E$ is a local field since the Hamiltonian density $h(x)$ is certainly $U(1)$-invariant, hence $\mathcal{O}^E$ is local with respect to it.

A consequence of it being a local field is that the functional integral is in fact independent of the shape of the cut from 0 to ∞ on each side of which the quasi-periodicity condition is imposed, up to a $U(1)$ transformation of the fields that the cut may cross while moving. Indeed, consider cuting not only on $A:\ x > 0$, but also on a path B starting and ending on the line $x > 0$, isolating a region bounded by B and by the segment $C \subset A$. Fix configurations along all these cuts with the quasi-periodicity condition (3.3) across A and with a condition of continuity across B. Integrating over all configurations on these cuts certainly gives back the functional integral (3.2) (puting the cut at B with continuity condition across it and suming over the configurations on it is like puting a complete set of states, hence does not change the path integral). But, since one region is isolated, we can make a $U(1)$ transformation in that region, which preserves the action in that region and change the fixed boundary conditions in a simple way: the new conditions across the cut B are quasiperiodicity conditions similar to (3.3), and the condition across C is continuity. Integrating over all configurations on the cuts now gives a functional integral with a modified cut, $B \cup (A \setminus C)$, across which the quasi-periodicity condition holds. The only difference is in the transformation of the fields present inside the isolated region, which shows the assertion.

Further insight can be gained into such fields making a change of variable in order to make the cut disappear. Indeed, consider the path integral on the right-hand side of (3.2) and apply a $U(1)$ transformation on the fields inside a (infinite) triangular region bounded by $A:\ x > 0$, by $B:\ z = e^{i\alpha}l$, $l > 0$ and by infinity (the triangular region is, for instance, above the cut A). Since this is a symmetry transformation, the action is invariant under this transformation up to three changes: a contribution on the boundary B of the triangle coming from the kinetic (derivative) term in the action, the modification of the boundary condition just above A, and the $U(1)$ transformations of fields that may be inside this triangle. We can choose the $U(1)$ transformation in such a way that the boundary condition just above A is equal to that just below it. Then, the path integral becomes a usual path integral without cut, but with insertion of the contribution of the action along B:

$$\mathcal{O}^E(0,0) = e^{2\pi i\omega \int_B dx^\mu j^\nu \epsilon_{\mu,\nu}} \tag{3.4}$$

where we denoted the contribution as an integration along B of the component of a current perpendicular to B (obviously, here we would need an appropriate regularisation of the exponential in order to define this field properly). We showed that by moving B through fields they get $U(1)$-transformed. Moving B can be done here simply by puting the exponential of the integral of the current, $e^{i\omega\int_L dx^\mu j^\nu \epsilon_{\mu,\nu}}$, along a loop L that shares a border with a segment of B and such that, on this segment, the integral goes in opposite direction as the integral in $e^{i\omega\int_B dx^\mu j^\nu \epsilon_{\mu,\nu}}$ above. Hence, we find that the insertion of

$$e^{2\pi i\omega\int_L dx^\mu j^\nu \epsilon_{\mu,\nu}}$$

has the effect of a $U(1)$ transformation of the fields surrounded by L. This means that the current j^ν is nothing else than the Noether current associated to the $U(1)$ invariance (and the derivation above may be seen as a way of defining this current). This makes it even more clear that the field is like the insertion of a magnetic charge.

Note that it is simple to extend this construction to twist fields associated to any global symmetry of the model. The form factor equations, however, have not been derived for such fields yet.

An example of such a twist field is the field in the free massive Majorana theory corresponding to the spin variable in the Ising model. The Majorana theory can be defined by its Euclidean action

$$S_E[\psi,\bar\psi] = -i\int d^2x\ \left[\psi\bar\partial\psi - \bar\psi\partial\bar\psi + 2m\bar\psi\psi\right]$$

where $\partial \equiv (1/2)(\partial/\partial x - i\partial/\partial\tau)$ and $\bar\partial \equiv (1/2)(\partial/\partial x + i\partial/\partial\tau)$. The fermion fields $\psi,\bar\psi$ are both real and are governed by the equations of motion

$$\bar\partial\psi = \frac{m}{2}\bar\psi\ ,\quad \partial\bar\psi = \frac{m}{2}\psi\ .$$

This means that any correlation function $\langle\mathrm{vac}|\mathcal{O}_1(x_1,\tau_1)\cdots\mathcal{O}_n(x_n,\tau_n)\psi(x,\tau)|\mathrm{vac}\rangle$ will satisfy, as function of x and τ, the equation of motion $\partial_x^2+\partial_\tau^2-m^2=0$.

The action has the Z_2 symmetry $\psi\mapsto-\psi$, $\bar\psi\mapsto-\bar\psi$. We can then define a twist field σ associated to this symmetry as above. In the free-field context, the definition means that the correlation function

$$\langle\mathrm{vac}|\sigma(0,0)\psi(x,\tau)\psi(x',\tau')|\mathrm{vac}\rangle$$

(we put two fermion fields so that it is non-zero) is, as function of both x,τ and x',τ', a solution to the equation of motion on a double covering of the Euclidean plane with branch point at $0,0$, and with quasi-periodicity condition according to which it gains a sign when going once around $0,0$. Along with the condition that as x,τ go to $0,0$, we get the least singular behavior, the solution is unique up to a normalisation. The condition on the least singular behavior comes from the fact that many fields, beside the twist fields introduced above, have a the same twist-field effect: their descendants under the free fermion operator algebra. But they have higher scaling dimension, so that the twist field as defined above is uniquely characterised by further imposing the condition of least singular behavior on the correlation functions. The field σ corresponds to the scaling limit of the spin variable in the lattice Ising model near criticality.

3.2 Explanation of the form factor properties

3.2.1 Analyticity and Watson's theorem

Properties 2 is obvious from relativistic invariance. Property 3 is also obvious from our definition of the states $|\theta_1,\ldots,\theta_n\rangle_{a_1,\ldots,a_n}$ as, generically, intermediate "asymptotic-like" states. However, the real statement of Watson's theorem, which is not obvious, is a part of Property 1: the fact that the regions $\theta_j-\theta_{j+1}>0$ and $\theta_j-\theta_{j+1}<0$ are analytical continuation of one another through at least a small strip containing the real axis. In the two-particle case, this is consequence of general (assumed) analytical properties of QFT. An heuristic argument can be given on the lines of the argument given above for understanding (2.7). Consider the path-integral formulation of the following matrix element, with $\theta_1>\theta_2$:

$$ {}_{a_1,a_2}^{\ \ out}\langle\theta_1,\theta_2|\mathcal{O}|\text{vac}\rangle = \int_{\Psi=\sum_k A_{a_k} e^{iE_k t - ip_k x}\ (t\to+\infty)} [d\Psi] e^{iS[\Psi]}\ . $$

Here, $\mathcal{O}$ is a local field at the point $0,0$. Again, the sum specifying the asymptotic conditions on the fundamental field has to be understood as giving the spacetime-dependent phase factors for wave packets ordered, on the space slice $t=-\infty$, from left to right with increasing index k. Taking the complex conjugate gives

$$ \langle\text{vac}|\mathcal{O}|\theta_1,\theta_2\rangle_{a_1,a_2}^{out} = \int_{\Psi=\sum_k A_{a_k} e^{-iE_k t + ip_k x}\ (t\to+\infty)} [d\Psi] e^{-iS[\Psi]} $$

and time-reversal invariance leads to

$$ \langle\text{vac}|\mathcal{O}|\theta_1,\theta_2\rangle_{a_1,a_2}^{out} = \int_{\Psi=\sum_k A_{a_k} e^{iE_k t + ip_k x}\ (t\to-\infty)} [d\Psi] e^{iS[\Psi]} $$

This is just the matrix element $\langle|\mathcal{O}|\theta_1',\theta_2'\rangle_{a_1,a_2}^{in}$ analytically continued to from $\theta_1'>\theta_2'$ to $\theta_1'=-\theta_1$, $\theta_2'=-\theta_2$ (that is, the main assumption here is that the path integral with wave-packet asymptotic prescription in the "wrong order" is just the analytical continuation of that in the "right order"). In other words, if there are no particle production (as in integrable systems, or else if the energies of the particles are small enough), then we have, for $\theta_1>\theta_2$,

$$ \langle\text{vac}|\mathcal{O}|\theta_1,\theta_2\rangle_{a_1,a_2}^{in} = S_{a_1,a_2}^{b_1,b_2}(\theta_1-\theta_2)\langle\text{vac}|\mathcal{O}|\theta_2',\theta_1'\rangle_{a_2,a_1}^{in}\Big|_{\theta_2'\mapsto\theta_2,\theta_1'\mapsto\theta_1} $$

where the analytical continuation is now from $\theta_2'>\theta_1'$. This is Watson's theorem

$$ F_{a_1,a_2}^{\mathcal{O}}(\theta_1,\theta_2) = S_{a_1,a_2}^{b_1,b_2}(\theta_1-\theta_2)F_{b_2,b_1}^{\mathcal{O}}(\theta_2,\theta_1) $$

in the 2-particle case; that is, the analytical continuation agrees with our choice of basis of asymptotic states. This was derived quite generally, but in integrable systems we have the intermediate basis, where processes occur 2-particle by 2-particle. Although the derivation above would not work as is, the important point is that if we change the rapidities in such a way that two wave packets, first in the correct order for an *in*-state, become colinear and then are in the wrong order, doing that by analytical continuation, then we get a factor of the scattering matrix associated to this 2-particle process. We can do that independently for every two-particle processes in our intermediate states of integrable systems, and this gives generalized Watson's theorem.

Another fact which is not obvious is the other part of Property 1: the fact that only simple poles may occur all the way up to the line of imaginary part 2π. This is called the "maximal analyticity assumption". That no other singularities occur all the way up to the line of imaginary part π is from standard principles of QFT, but the maximal analyticity assumption is really just an assumption, and for integrable models, the current way of thinking is to take this for granted, and to verify that the form factors obtained under this assumption (and the other form factor properties) give rise to local fields of the theory.

3.2.2 Crossing symmetry and locality property

For explaining Points 4 and 5, it will be useful to have in mind particles that possess a $U(1)$ charge, so that we consider Ψ, the fundamental field associated to a particle, to be a complex field. We will take $\mathcal{O}$ to be the associated twist field with semi-locality index ω defined as above.

The locality property, point 4, is essentially a consequence of two applications of crossing symmetry.

Crossing symmetry, a general property of QFT, stipulates that analytically continuing the rapidity variable θ_n to $\theta_n + i\pi - i0^+$ in the matrix element

$$\langle \text{vac}|\mathcal{O}|\theta_1, \ldots, \theta_{n-1}, \theta_n\rangle^{(in)}_{a_1,\ldots,a_{n-1},a_n} ,$$

with initially $\theta_1 > \cdots > \theta_n$, gives the matrix element

$${}^{\;out}_{\bar{a}_n}\langle \theta_n - i0^+|\mathcal{O}|\theta_1, \ldots, \theta_{n-1}\rangle^{(in)}_{a_1,\ldots,a_{n-1}}$$

understood as analytical continuation again. The latter matrix element, without the $-i0^+$, is really a distribution for real rapidities. This distribution, in terms of θ_n, has a part supported at other rapidities $\theta_1, \ldots, \theta_{n-1}$, and a part supported on the complement on $\mathbb{R}$. Essentially, the part supported on the points $\theta_1, \ldots, \theta_{n-1}$ is interpreted as coming from particles going through from *in*- to *out*-state without interacting with the operator. The analytical continuation above, with the $-i0^+$, is the analytical continuation of the part supported away from these points.

Crossing symmetry can be understood by slowly bringing the wave packet associated to θ_n around the operator $\mathcal{O}$ counterclockwise from the *in*-state to the *out*-state. This motion is implemented by taking first θ_n to $-\infty$, then adding $i\pi/2$, then bringing its real part to $+\infty$, then adding again $i\pi/2$ and finally bringing its real part back to its initial value. In this final motion, we need to avoid the position of the rapidities $\theta_1, \ldots, \theta_n$, by keeping a slight negative imaginary part. Straightening the path (under the assumption that there are no branch points in the rapidity plane) gives the statement above. With the operator $\mathcal{O}$ being a twist field, we need, for this to be valid, to take the cut to its left (although the shape of the cut does not affect the correlation functions, it affects some of its matrix elements on the Hilbert space), so that the wave packet does not cross it.

Moving the wave packet from an *out* position to an *in* position, a different expression of crossing symmetry allows us to start from

$$ {}_{\bar{a}_n}^{out}\langle\theta_1 + i0^+|\mathcal{O}|\theta_2,\ldots,\theta_n\rangle^{(in)}_{a_2,\ldots,a_1} $$

with $\theta_1 > \cdots > \theta_n$ and analytically continue to $\theta_1 \mapsto \theta_1 + i\pi$, giving back

$$ \langle\mathrm{vac}|\mathcal{O}|\theta_1,\theta_2,\ldots,\theta_{n-1},\theta_n\rangle^{(in)}_{a_1,a_2,\ldots,a_n} \ . $$

In doing so, however, the wave packet will need to cross the cut of the operator $\mathcal{O}$ if it is semi-local with respect to the associated fundamental field (we assume throughout that this cut is on the left). This gives a factor with the semi-locality index. In fact, in a more precise treatment of this, the factor of semi-locality would come from commuting mutually semi-local operators, and it is always accompanied by a factor $(-1)^{f_{\mathcal{O}} f_\Psi}$ [4].

With an appropriate analyticity assumption, included in the assumption of maximal analyticity, the two values

$$ {}_{\bar{a}_n}^{out}\langle\theta_1 \pm i0^+|\mathcal{O}|\theta_2,\ldots,\theta_n\rangle^{(in)}_{a_2,\ldots,a_1} $$

are analytical continuation of each other everywhere on $\theta_1 \in [\theta_2,\theta_n]$ (no cut occurs, or at least we can correctly go around it). Then, we can apply twice crossing symmetry and we find Point 4 with the understanding that the analytical continuation of θ_n is from a value with $\theta_1 > \cdots > \theta_n$ towards $\theta_n + 2\pi i$ with $\theta_n > \theta_1 > \cdots > \theta_{n-1}$. Since in integrable models the form factors are meromorphic, this can be extended to the real part of θ_n staying the same.

3.2.3 Kinematic pole

The kinematic pole, point 5, is, in its simplest form, a consequence of general principles of QFT (although one needs the analyticity assumption of the paragraph above). Consider the two-particle case:

$$ iF^{\mathcal{O}}_{a_1,a_2}(\theta_1,\theta_2) \sim C_{a_2,a_1}\frac{1-e^{2\pi i\omega}}{\theta_2-\theta_1-i\pi} $$

where we used the fact that $(-1)^{f_{\mathcal{O}} f_\Psi} = 1$ (because, even if Ψ is fermionic, in order to have a non-zero two-particle form factor, $\mathcal{O}$ must bosonic: $f_{\mathcal{O}} = 0$). In order to derive this equation, we will look at the behavior around and at $\theta_1 = \theta_2$ of the different but related matrix element

$$ g_{\bar{a}_2,a_1}(\theta_2,\theta_1) = \langle\mathrm{vac}|A_{\bar{a}_2}(\theta_2)^{(out)}\mathcal{O}A^\dagger_{a_1}(\theta_1)^{(in)}|\mathrm{vac}\rangle \ . $$

As we said above, it is generically not an analytical function, but rather a distribution. The part of the distribution supported away from $\theta_1 = \theta_2$ has an analytical continuation that may

[4] That this factor appears in doing the latter crossing $\theta_1 + i0^+ \mapsto \theta_1 + i\pi$ or in doing the former one $\theta_n \mapsto \theta_n + i\pi - i0^+$ is really a matter of definition of the phase of the asymptotic states (which is not fixed by their normalisation).

have a singularity at this point. The part supported at $\theta_1 = \theta_2$ comes from particles going through from *in*- to *out*-state without interacting with the operator, and the singularity of the part supported away comes from particles being affected only by the cut giving the semi-locality of the operator $\mathcal{O}$.

We could write this matrix element using the definition (2.1) of the operators $A(\theta)^{(in,out)}$. Instead, we will use what is sometimes called the "extrapolating fields":

$$\begin{aligned} g_{\bar{a}_2,a_1}(\theta_2,\theta_1) &= \langle \mathrm{vac}| \int dx_2\, e^{-ip_2x_2} \left(\partial_t \tilde{\Psi}_{\bar{a}_2}(x_2,0)^{(out)} - iE_2 \tilde{\Psi}_{\bar{a}_2}(x_2,0)^{(out)}\right)\, \mathcal{O}\ \times \\ &\quad \times \int dx_1\, e^{ip_1x_1} \left(\partial_t \tilde{\Psi}^\dagger_{a_1}(x_1,0)^{(in)} + iE_1 \tilde{\Psi}^\dagger_{a_1}(x_1,0)^{(in)}\right) |\mathrm{vac}\rangle \end{aligned} \tag{3.5}$$

where, as usual, $p_1 = m\sinh(\theta_1)$, $E_1 = m\cosh(\theta_1)$, etc. The extrapolating fields are defined by

$$\tilde{\Psi}_a(x,t)^{(in/out)} = \frac{1}{4i\pi} \int d\theta\, \left[e^{ip(\theta)x - iE(\theta)t} A_a(\theta)^{(in/out)} + e^{-ip(\theta)x+iE(\theta)t} A^\dagger_{\bar{a}}(\theta)^{(in/out)} \right]\ .$$

These fields are not generically local, and very different from the local fundamental field associated to the particle a, which would have the form, in terms of *in* operators,

$$\begin{aligned} \Psi_a(x,t) &= \int d\theta \left[e^{ip(\theta)x} \left(e^{-iE(\theta)t} A_a(\theta)^{(in)} + \right.\right. \\ &\quad + \int d\theta_1 d\theta_2 \delta(p(\theta) - p_1 - p_2) e^{-i(E_1+E_2)t} u_a^{a_1,a_2}(\theta_1,\theta_2) A_{a_1}(\theta_1)^{(in)} A_{a_2}(\theta_2)^{(in)} + \\ &\quad + \int d\theta_1 d\theta_2 \delta(p(\theta) + p_1 - p_2) e^{-i(-E_1+E_2)t} v_a^{a_1,a_2}(\theta_1,\theta_2) A^\dagger_{\bar{a}_1}(\theta_1)^{(in)} A_{a_2}(\theta_2)^{(in)} + \\ &\quad \left. + \int d\theta_1 d\theta_2 \delta(p(\theta) + p_1 + p_2) e^{-i(-E_1-E_2)t} w_a^{a_1,a_2}(\theta_1,\theta_2) A^\dagger_{\bar{a}_1}(\theta_1)^{(in)} A^\dagger_{\bar{a}_2}(\theta_2)^{(in)} + \cdots \right) + \\ &\quad \left. + e^{-ip(\theta)x} \left(e^{iE(\theta)t} A^\dagger_{\bar{a}}(\theta)^{(in)} + \cdots \right) \right] \end{aligned}$$

and a similar expression exists in terms of *out* operators (only in free theories are extrapolating fields equal to fundamental fields). Note that the particle types involved on the right-hand side must be such that the $U(1)$ charge is the same on both sides. Inverting, one would obtain

$$\tilde{\Psi}_a(x,t) = \Psi_a(x,t) + \int dx_1 dx_2\ \left[K_a^{a_1,a_2}(x,x_1,x_2) \Psi_{a_1}(x_1,t) \Psi_{a_2}(x_2,t) + \cdots \right]$$

for some $K_a^{a_1,a_2}(x,x_1,x_2)$, where the $\cdots$ mean terms of similar type involving one time-derivative of the fields, complex conjugate fields, and more and more factors. Again, on the right-hand side, the $U(1)$ charge is the same as that of the left-hand side. The important point is that the quantities $K_a^{a_1,a_2}(x,x_1,x_2)$ vanish exponentially fast as x_1 or x_2 are brought far from x:

$$K_a^{a_1,a_2}(x,x_1,x_2) \propto e^{-m|x-x_1|} \quad (|x-x_1| \to \infty)\ , \quad K_a^{a_1,a_2}(x,x_1,x_2) \propto e^{-m|x-x_2|} \quad (|x-x_2| \to \infty)\ .$$

That is, although the extrapolating fields are not strictly local, they are "of finite extent", in the sense that they have tails that decrease exponentially.

Now, the behavior near and at $\theta_1 = \theta_2$, for finite $p_1 + p_2$, is obtained from (3.5) by looking at the region $|x_1 + x_2| \to \infty$ with $|x_1 - x_2|$ finite. Since the extrapolating fields are then very far from the field $\mathcal{O}$ and since they are of finite extent, the resulting correlation function factorizes:

$$\begin{aligned} g_{\bar a_2,a_1}(\theta_2,\theta_1) \quad \sim \quad & \langle\text{vac}| \int dx_2\, e^{-ip_2x_2} e^{2\pi i\omega\Theta(-x_1)} \left(\partial_t \tilde\Psi_{\bar a_2}(x_2,0)^{(out)} - iE_2 \tilde\Psi_{\bar a_2}(x_2,0)^{(out)}\right) \times \\ & \times \int dx_1\, e^{ip_1x_1} \left(\partial_t \tilde\Psi^\dagger_{a_1}(x_1,0)^{(in)} + iE_1 \tilde\Psi^\dagger_{a_1}(x_1,0)^{(in)}\right) |\text{vac}\rangle \times \\ & \langle\text{vac}|\mathcal{O}|\text{vac}\rangle\ . \end{aligned}$$

The operations involved in factorizing are as follows: we first bring the operator at x_1 just to the left of $\mathcal{O}$ inside the vacuum expectation value in (3.5), which brings a factor of semi-locality since the extrapolating fields are for from the point $x = 0$, are of finite extent and have a well-defined $U(1)$ charge; then we factorize the correlation function as above, using the fact that the extrapolating fields have finite extent. Then, we can use translation invariance and the change of variable $x_2 \mapsto x_2 + x_1$ in order to find

$$\begin{aligned} g_{\bar a_2,a_1}(\theta_2,\theta_1) \quad \sim \quad & \int dx_1 e^{ix_1(p_1-p_2)} e^{2\pi i\omega\Theta(-x_1)} \int dx_2\, e^{-ip_2x_2} \langle\text{vac}| \left(\partial_t \tilde\Psi_{\bar a_2}(x_2,0)^{(out)} - iE_2 \tilde\Psi_{\bar a_2}(x_2,0)^{(out)}\right) \times \\ & \times \left(\partial_t \tilde\Psi^\dagger_{a_1}(0,0)^{(in)} + iE_1 \tilde\Psi^\dagger_{a_1}(0,0)^{(in)}\right) |\text{vac}\rangle\ \langle\text{vac}|\mathcal{O}|\text{vac}\rangle\ . \end{aligned}$$

The integral over x_1 can be evaluated using the distributional identities

$$\int dx\, e^{ipx}\, \text{sign}(x) = 2i\underline{\text{P}}\left(\frac{1}{p}\right)\ , \quad \int dx\, e^{ipx} = 2\pi\delta(p) \tag{3.6}$$

where $\underline{\text{P}}$ means *principal value* (that is, under integration, we cut a region of length ϵ symmetricaly distributed around the pole at $p = 0$ and evaluate the limit $\epsilon \to 0$ after integrating). This gives

$$\int dx_1 e^{ix_1(p_1-p_2)} e^{2\pi i\omega\Theta(-x_1)} = \left(1 - e^{2\pi i\omega}\right) i\underline{\text{P}}\left(\frac{1}{p_1 - p_2}\right) + \left(1 + e^{2\pi i\omega}\right) \pi\delta(p_1 - p_2)\ .$$

The integral over x_2 can be evaluated by using the explicit expression for the extrapolating fields. A shorter way to evaluate it at $\theta_1 = \theta_2$ (which is the only point we need) is to realize that we could have done the same calculation with $\mathcal{O}$ being just the identity operator $\mathbf{1}$, with $\omega = 0$, and that this should give $2\pi\delta(\theta_1 - \theta_2)\delta_{a_1,\bar a_2}$, hence

$$\int dx_2\, e^{-ip_2x_2} \langle\text{vac}| \left(\partial_t \tilde\Psi_{\bar a_2}(x_2,0)^{(out)} - iE_2 \tilde\Psi_{\bar a_2}(x_2,0)^{(out)}\right) \left(\partial_t \tilde\Psi^\dagger_{a_1}(0,0)^{(in)} + iE_2 \tilde\Psi^\dagger_{a_1}(0,0)^{(in)}\right) |\text{vac}\rangle = E_2 \delta_{a_1,\bar a_2}\ .$$

Puting all that together, we find

$$g_{\bar a_2,a_1}(\theta_2,\theta_1) \sim \left[\left(1 - e^{2\pi i\omega}\right) i\underline{\text{P}}\left(\frac{1}{\theta_1 - \theta_2}\right) + \left(1 + e^{2\pi i\omega}\right) \pi\delta(\theta_1 - \theta_2)\right] C_{a_1,a_2}$$

where we used the fact that $\delta_{a_1,\bar a_2}$ just gives the matrix element C_{a_1,a_2} of the conjugation matrix.

Now, the function $g_{\bar a_2,a_1}(\theta_2,\theta_1)$ is simply related to the form factor $F^{\mathcal{O}}_{a_1,a_2}(\theta_1,\theta_2)$ by crossing symmetry: $F^{\mathcal{O}}_{a_1,a_2}(\theta_1,\theta_2) = g_{\bar a_2,a_1}(\theta_2 - i\pi,\theta_1)$ where we analytically continue only the part $\left(1 - e^{2\pi i\omega}\right) i\underline{\text{P}}\left(\frac{1}{\theta_1-\theta_2}\right)$. This gives the kinematical residue equation in the 2-particle case.

One can try to do the many-particle case in a similar way still for non-integrable models, by considering the matrix element

$$g_{\bar{a}_n,a_1,a_2,\ldots,a_{n-1}}(\theta_n,\theta_1,\theta_2,\ldots,\theta_{n-1}) = \langle\text{vac}|A_{\bar{a}_n}(\theta_n)^{(out)}\mathcal{O}A^\dagger_{a_1}(\theta_1)^{(in)}A^\dagger_{a_2}(\theta_2)^{(in)}\cdots A^\dagger_{a_{n-1}}(\theta_{n-1})^{(in)}|\text{vac}\rangle \tag{3.7}$$

with $\theta_1 > \cdots > \theta_{n-1}$ and writing the asymptotic state operators in terms of extrapolating fields integrated over their positions $x_1,\ldots,x_n$. There is a subtlety, however, that lies in using the identities (3.6). They should really be understood as conditionally convergent integral, made convergent by giving a small x-dependent imaginary part to p, positive for $x>0$ and negative for $x<0$. This means that when we look at the region where x_n is very positive, we will really be considering the matrix element

$$x_n \gg m^{-1} \quad : \quad g_{\bar{a}_n,a_1,a_2,\ldots,a_{n-1}}(\theta_n - i0^+,\theta_1,\theta_2,\theta_{n-1})$$

whereas when we look at the region where both x_n is very negative, we will be considering the matrix element

$$x_n \ll -m^{-1} \quad : \quad g_{\bar{a}_n,a_1,a_2,\ldots,a_{n-1}}(\theta_n + i0^+,\theta_1,\theta_2,\theta_{n-1})\ .$$

The first matrix element is related by crossing symmetry to the form factor $F^{\mathcal{O}}_{a_1,\ldots,a_n}(\theta_1,\ldots,\theta_n)$ for $\theta_n > \theta_{n-1}$ and the second is related to $F^{\mathcal{O}}_{a_n,a_1,\ldots,a_{n-1}}(\theta_n,\theta_1,\ldots,\theta_{n-1})$ for $\theta_n < \theta_1$. Hence, the two regions of x_n will give us contributions to poles of these two different functions of θ_n at different points: the first at $\theta_{n-1}+i\pi$, the second at $\theta_1 - i\pi$. The second can also be seen as a pole at $\theta_1 + i\pi$ of the analytical continuation of the form factor $F^{\mathcal{O}}_{a_1,\ldots,a_n}(\theta_1,\ldots,\theta_n)$ from the region $\theta_n > \theta_{n-1}$, under the appropriate analyticity assumption (included in the maximal analyticity assumption). It is really from this and the fact that $\theta_{n-1}=\theta_1$ in the two-particle case that we obtained the two-particle result above.

The residue of these two contributions to poles can be evaluated as above for the two-particle case. We get a $n-2$-particle form factor, times an appropriate sign and semi-locality factor. These can be evaluated as follows. In the first case (x_n positive), we will have to bring the right-most operators $\tilde{\Psi}^{in}_{a_{n-1}}(x_{n-1})$ almost all the way to the left, just next to $\tilde{\Psi}^{out}_{\bar{a}_n}(x_n)$, on its right, before doing the factorisation. If both $\mathcal{O}$ and the particle $\bar{a}_n$ are fermionic, then the particle a_{n-1} will have to be fermionic as well and the operators $\tilde{\Psi}_{a_{n-1}}(x_{n-1})$ will cross an odd number of fermionic extrapolating fields as well as the operator $\mathcal{O}$, so that no sign will be gained in this process. A little thought on similar lines shows that no matter if $\mathcal{O}$ or a_n are bosonic or fermionic, no sign will be gained. In the second case, (x_n negative) the extrapolating field associated to a_1, which will have to be of the same statistics as $\bar{a}_n$, will only have to cross the operator $\mathcal{O}$, so that we will have a factor $(-1)^{f_{\mathcal{O}}f_{\Psi_{a_n}}}$. Moreover, by arguments similar to those above for the two-particle case, we will have a factor of semi-locality only for x_n negative, that is, in the second case.

We still do not have the full information about the poles, because the two contributions are for poles at different points. But in integrable models, we can also calculate the part of the x_n and

x_{n-1} integrals for very negative x_n and x_{n-1} in the matrix element $g_{\bar{a}_n,a_1,a_2,\ldots,a_{n-1}}(\theta_n,\theta_1,\theta_2,\ldots,\theta_{n-1})$ with $\theta_1 > \ldots > \theta_{n-1} > \theta_n$ written as integrations over extrapolating fields from (3.7). Let us divide this matrix element into two part: one that contains the x_{n-1}-integral from $-\infty$ up to some value X, the other that goes from X to ∞:

$$g_{\bar{a}_n,a_1,\ldots,a_{n-1}}(\theta_n,\theta_1,\ldots,\theta_{n-1}) = g^{(-)}_{\bar{a}_n,a_1,\ldots,a_{n-1}}(\theta_n,\theta_1,\ldots,\theta_{n-1}) + g^{(+)}_{\bar{a}_n,a_1,\ldots,a_{n-1}}(\theta_n,\theta_1,\ldots,\theta_{n-1}) \ .$$

The latter part gives rise to a contribution to the pole at $\theta_n = \theta_{n-1}$ that we already know how to calculate, but the former part also gives a contribution, which we do not know yet how to calculate. However, we know how to write this matrix element as an analytical continuation of the matrix element $g_{\bar{a}_n,a_{n-1},a_1,\ldots,a_{n-2}}(\theta_n,\theta_{n-1},\theta_1,\ldots,\theta_{n-2})$ with $\theta_{n-1} > \theta_1 > \cdots > \theta_{n-2}$, using multiple applications of generalized Watson's theorem, symbolically:

$$g_{\bar{a}_n,a_1,\ldots,a_{n-1}}(\theta_n,\theta_1,\ldots,\theta_{n-1}) = \left(\prod S\right) g_{\bar{a}_n,a_{n-1},a_1,\ldots,a_{n-2}}(\theta_n,\theta_{n-1},\theta_1,\ldots,\theta_{n-2}) \ .$$

This analytical continuation does not exactly hold separately for the positive-x_{n-1} and the negative-x_{n-1} parts; for instance:

$$g^{(-)}_{\bar{a}_n,a_1,\ldots,a_{n-1}}(\theta_n,\theta_1,\ldots,\theta_{n-1}) = \left(\prod S\right) g^{(-)}_{\bar{a}_n,a_{n-1},a_1,\ldots,a_{n-2}}(\theta_n,\theta_{n-1},\theta_1,\ldots,\theta_{n-2}) + \Delta$$

where Δ is a correction. However, we can evaluate the residue of the pole of the first term on the right-hand side at $\theta_n = \theta_{n-1}$ and we find that it is independent of X. Hence, since the correction Δ vanishes as $X \to \infty$, this gives the full residue of that pole of the left-hand side for all finite X. The contributions at x_n, x_{n-1} very positive and at x_n, x_{n-1} very negative are the only two contributions to the pole at $\theta_n = \theta_{n-1}$ of $g_{\bar{a}_n,a_1,a_2,\ldots,a_{n-1}}(\theta_n,\theta_1,\theta_2,\ldots,\theta_{n-1})$, hence we have the full residue of the pole at $\theta_n = \theta_{n-1} + i\pi$ of the form factor $F^{\mathcal{O}}_{a_1,\ldots,a_n}(\theta_1,\ldots,\theta_n)$, and we have shown point 5.

Note that a similar argument also holds for the part of $g_{\bar{a}_n,a_1,a_2,\ldots,a_{n-1}}(\theta_n,\theta_1,\theta_2,\ldots,\theta_{n-1})$ supported at $\theta_n = \theta_{n-1}$. In general, we can write

$$g_{\bar{a}_n,a_1,a_2,\ldots,a_{n-1}}(\theta_n,\theta_1,\theta_2,\ldots,\theta_{n-1}) = \left(\frac{-iR}{\theta_n - \theta_{n-1} + i0^+} + 2\pi\delta(\theta_n - \theta_{n-1})\right) F^{\mathcal{O}}_{a_1,\ldots,a_{n-2}}(\theta_1,\ldots,\theta_{n-2}) + \ldots$$

where $\ldots$ are convergent at $\theta_n = \theta_{n-1}$ and R is as calculated above. This should be understood using

$$\frac{1}{\theta + i0^+} = \underline{\mathrm{P}}\left(\frac{1}{\theta}\right) - i\pi\delta(\theta) \ .$$