Quantum Field Theory – Part 2

Rainer F. Hauser
rainer.hauser@gmail.com

June 1, 2021

Abstract

Quantum Field Theory combines Special Relativity and Quantum Mechanics and was initially developed to quantize the electromagnetic field. It later turned out that all fundamental interactions are necessarily described by Quantum Field Theory. Tobias Osborne from the Leibniz Universität Hannover held two times 18 lectures in 2016 and 2017, respectively, covering this topic. His lectures were available on YouTube at the time this transcript has been assembled and may as "Quantum Field Theory" and "Advanced Quantum Field Theory" still be available today.

1 Path Integrals

1.1 Derivation and Definition of Path Integrals

The goal is to explain the standard model of particle physics although this is not completely possible here. Instead, the building blocks of the standard model are introduced. The idea is not to explain every term in the Lagrangian of the standard model but just the main concepts the standard model interactions are built on. This means that in addition to the quantum field theory of bosons and fermions also the quantum field theory of abelian and non-abelian gauge bosons (quantum electrodynamics and Yang–Mills theory) is needed. This allows together with an understanding of the symmetry breaking mechanisms to look at the Lagrangian of the standard model and point at each term recognizing where it comes from and what the corresponding interactions describe.

As a first step path integrals are introduced and perturbation theory is reviewed from this side where path integrals are in most cases restricted here to Gaussian path integrals. The beauty of path integrals is that expressing concepts such as Wick's theorem or Feynman rules becomes much simpler compared to expressing those concepts in canonical quantization.

Suppose that there is a quantum system with a Hilbert space $\mathfrak{H}$ and a Hamiltonian $\boldsymbol{H}$ and that one can integrate the Schrödinger equation $\boldsymbol{U}(t) = e^{-i\,t\,\boldsymbol{H}}$ to get the propagator such that one knows everything about this quantum system. The objective is to get a representation of this propagator in terms of more classical data. It is not always possible to solve the Schrödinger equation but there is an interesting trick for approximating the propagator. It is based on

$$\boldsymbol{U}(t) = e^{-i\,t\,\boldsymbol{H}} = \left(e^{-\frac{i\,t}{m}\,\boldsymbol{H}}\right)^m = \lim_{m\to\infty}\left(\mathbf{1} - \frac{i\,t}{m}\,\boldsymbol{H}\right)^m = \lim_{m\to\infty}\boldsymbol{X}^m$$

and gives an alternative representation to the original propagator. If one cannot solve the Schrödinger equation one might still know how to go from $t = 0$ to $t = \varepsilon$ because the quantities building $\boldsymbol{X}$ are known. Note also that $\boldsymbol{X}$ is a matrix which is basically the identity. This is step one of the path integral.

Suppose that $|j\rangle$ is a basis for $\mathfrak{H}$ where it is assumed that $\mathfrak{H}$ is finite dimensional such that everything converges. The transition amplitude to go from an initial state $|\phi_{\text{initial}}\rangle$ to a final state $|\phi_{\text{final}}\rangle$ is

$$\langle\phi_{\text{final}}|\boldsymbol{U}(t)|\phi_{\text{initial}}\rangle = \langle\phi_{\text{final}}|e^{-i\,t\,\boldsymbol{H}}|\phi_{\text{initial}}\rangle = \langle\phi_{\text{final}}|e^{-\frac{i\,t}{m}\,\boldsymbol{H}}\,e^{-\frac{i\,t}{m}\,\boldsymbol{H}}\,\ldots\,e^{-\frac{i\,t}{m}\boldsymbol{H}}|\phi_{\text{initial}}\rangle$$

and the probability is the square of this amplitude. One can obviously introduce identities between two terms to get

$$\begin{aligned}
&\langle\phi_{\text{final}}|\boldsymbol{U}(t)|\phi_{\text{initial}}\rangle \\
&= \langle\phi_{\text{final}}|e^{-\frac{it}{m}\boldsymbol{H}}\textstyle\sum_{j_{m-1}}|j_{m-1}\rangle\langle j_{m-1}|\,e^{-\frac{it}{m}\boldsymbol{H}}\sum_{j_{m-2}}|j_{m-2}\rangle\langle j_{m-2}|\ldots\sum_{j_1}|j_1\rangle\langle j_1|\,e^{-\frac{it}{m}\boldsymbol{H}}|\phi_{\text{initial}}\rangle \\
&= \sum_{j_1\ldots j_{m-1}}\langle\phi_{\text{final}}|e^{-\frac{it}{m}\boldsymbol{H}}|j_{m-1}\rangle\langle j_{m-1}|\,e^{-\frac{it}{m}\boldsymbol{H}}|j_{m-2}\rangle\langle j_{m-2}|\ldots|j_1\rangle\langle j_1|\,e^{-\frac{it}{m}\boldsymbol{H}}|\phi_{\text{initial}}\rangle \\
&= \sum_{j_1\ldots j_{m-1}}F\big(\text{path}\,(j_1,j_2,...,j_{m-1})\big)
\end{aligned}$$

and this is a sum over some function F of a path.

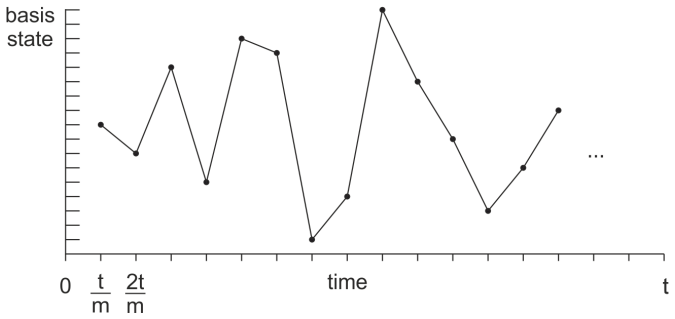


Figure 1: Visualization of a path

A path is $(j_1,...,j_{m-1})$ with $j_k \in \mathbb{N}$ and consists of $m-1$ numbers. It can be visualized as in the example shown in figure 1. Thus,

$$\langle\phi_{\text{final}}|\boldsymbol{U}(t)|\phi_{\text{initial}}\rangle = \sum_{j_1\ldots j_{m-1}}F(j_1,j_2,...,j_{m-1})$$

and the transition amplitude to go from the initial state to the final state is equal to the sum over all paths which are functions of tuples of natural numbers. This identity allows to employ a sequence of tricks and simplifications. The question is what this function F of the paths is and for the answer one only has to know

$$\langle j|\left(\mathbf{1}-\frac{i\,t}{m}\boldsymbol{H}\right)|k\rangle \cong e^{\frac{i\,t}{m}\mathcal{L}(j,k)}$$

where $\mathcal{L}$ will turn out to be the Lagrangian. The left side is a number depending on j and k and one can write numbers as exponentials. Thus, the transition amplitude can be written as

$$\langle\phi_{\text{final}}|\boldsymbol{U}(t)|\phi_{\text{initial}}\rangle = \sum_{j_1\ldots j_{m-1}}e^{\frac{i\,t}{m}\left(\sum_{l=2}^{m-1}\mathcal{L}(j_l,j_{l-1})\right)}$$

and this is as far one can get on the path integral without assuming something more about the system.

The path integral is interesting because of several reasons:

1. It allows to calculate quantum quantities via classical solutions. One can use all the experience in solving classical systems as a calculational tool to determine some quantum transition amplitude via the saddle point approximation applied to highly oscillatory integrals.
2. It can be used to build completely different approximation schemes for the transition amplitude by using Monte Carlo sampling. This is still the only known way to calculate non-perturbative results about the standard model.

1.2 Application to General Non-Relativistic Quantum Mechanical Systems

In the introduction of path integrals not much has been specified about the system. Just a Hilbert space and a Hamiltonian were assumed but nothing more. Given a quantization one can certainly write the

propagator as a sum over paths of some scalar quantity that depends on the path. In order to study a system in more detail it is supposed that there is a classical system with a Hamiltonian $H(\underline{q},\underline{p})$ involving generalized coordinates $\underline{q} = \{q^j\}$ and conjugate momenta $\underline{p} = \{p^j\}$. One can turn around the path integral to guess a quantum Hamiltonian via

$$U(\underline{q}_a,\underline{q}_b;T) = \langle \underline{q}_b|U(T)|\underline{q}_a\rangle = \langle \underline{q}_b|e^{-i\,\boldsymbol{H}\,T}|\underline{q}_a\rangle$$

and ask what is $\boldsymbol{H}$ in this equation. In other words, one turns the logic around, starts with the classical Hamiltonian and tries to find via this path integral a quantum Hamiltonian that corresponds in some natural way to the given classical Hamiltonian.

Thus, one proceeds as before with inserting complete relation as identities

$$\mathbf{1} = \left(\prod_j \int dq_k^j\right) |\underline{q}_k\rangle\,\langle \underline{q}_k|$$

which look differently than above because the system has a continuous set of degrees of freedom. This is the completeness relation for the basis of position eigenstates to be inserted into

$$\langle \underline{q}_b|U(T)|\underline{q}_a\rangle = \sum \langle \underline{q}_m|e^{-i\,\varepsilon\,\boldsymbol{H}}|\underline{q}_{k_{m-1}}\rangle\,\langle \underline{q}_{k_{m-1}}|...|\underline{q}_{k_1}\rangle\,\langle \underline{q}_{k_1}|e^{-i\,\varepsilon\,\boldsymbol{H}}|\underline{q}_0\rangle$$

with $\underline{q}_0 = \underline{q}_a$, $\underline{q}_m = \underline{q}_b$ and $\varepsilon = \frac{T}{m}$.

If the quantum Hamiltonian is a function of the quantum position operators only but not of the quantum momenta operators such that $\boldsymbol{H} = f(\{\boldsymbol{q}^j\}) = f(\underline{\boldsymbol{q}})$ then since $|q_k\rangle$ are energy eigenstates

$$\langle \underline{q}_{k+1}|f(\underline{\boldsymbol{q}})|\underline{q}_k\rangle = f(\underline{q}_k)\prod_j \delta(q_k^j - q_{k-1}^j)$$

relates classical and quantum quantities because $f(\underline{\boldsymbol{q}})$ is an operator and $f(\underline{q}_k)$ is a function of classical numbers $\underline{q}_k$. In order to introduce also momentum quantities one can use the identity

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} dp\, e^{i\,p\cdot q} = \delta(q)$$

for the Dirac delta function to get the momentum representation

$$\langle \underline{q}_{k+1}|f(\underline{\boldsymbol{q}})|\underline{q}_k\rangle = f\left(\frac{\underline{q}_{k+1}+\underline{q}_k}{2}\right)\left(\prod_j \int \frac{dp_k^j}{2\pi}\right) e^{i\sum_j p_k^j(q_{k+1}^j - q_k^j)}$$

where $\frac{\underline{q}_{k+1}+\underline{q}_k}{2}$ is $\underline{q}_k$ because of the delta functions $\delta(q_k^j - q_{k-1}^j)$.

If the quantum Hamiltonian is a function of the quantum momentum operators but not of the quantum position operators such that $\boldsymbol{H} = g(\{\boldsymbol{p}^j\}) = g(\underline{\boldsymbol{p}})$ then one needs to calculate

$$\begin{aligned}\langle \underline{q}_{k+1}|g(\underline{\boldsymbol{p}})|\underline{q}_k\rangle &= \langle \underline{q}_{k+1}|g(\underline{\boldsymbol{p}})\left(\prod_j \int dp_k^j\,|\underline{p}_k\rangle\,\langle \underline{p}_k|\right)|\underline{q}_k\rangle = \langle \underline{q}_{k+1}|g(\underline{\boldsymbol{p}})\left(\prod_j \int dp_k^j\,|\underline{p}_k\rangle\,\langle \underline{p}_k|\,\underline{q}_k\rangle\right)\\ &= \left(\prod_j \int \frac{dp_k^j}{2\pi}\right) g(\underline{p}_k)\, e^{i\sum_j p_k^j(q_{k+1}^j - q_k^j)}\end{aligned}$$

where initially the states are in the position eigenbasis and the operator is a momentum operator. By inserting the completeness relation in the momentum eigenbasis a Fourier phase $\langle \underline{p}_k|\underline{q}_k\rangle$ is introduced.

In interesting systems the Hamiltonian is usually a function of the quantum position operators and the quantum momentum operators such that $\boldsymbol{H} = f(\underline{\boldsymbol{q}}) + g(\underline{\boldsymbol{p}})$ which is the general case in non-relativistic quantum mechanics. Using

$$e^{-i\,\varepsilon\,\boldsymbol{H}} \approx \mathbf{1} - i\,\varepsilon\,\boldsymbol{H} = \mathbf{1} - i\,\varepsilon\,\big(f(\underline{\boldsymbol{q}}) + g(\underline{\boldsymbol{p}})\big)$$

up to first order $O(\varepsilon)$ and using the above two results gives

$$\langle \underline{q}_{k+1}|e^{-i\,\varepsilon\,\boldsymbol{H}}|\underline{q}_k\rangle = \left(\prod_j \int \frac{dp_k^j}{2\pi}\right) e^{-i\,\varepsilon\,H(\frac{q_{k+1}+q_k}{2},\underline{p}_k)}\; e^{i\sum_j p_k^j(q_{k+1}^j-q_k^j)}$$

for the transition amplitude.

Bringing all together results in

$$U(\underline{q}_a,\underline{q}_b;T) = \left(\prod_{j,k}\int dq_k^j \int \frac{dp_k^j}{2\pi}\right) e^{i\,\sum_k\left(\sum_j p_k^j(q_{k+1}^j-q_k^j)-\varepsilon\,H(\frac{q_{k+1}+q_k}{2},\underline{p}_k)\right)} \tag{1.1}$$

and the beauty of the expression on the right-hand side is that there are no operators but only classical variables. This is the equation (without the i) that is used for Monte Carlo sampling, and it is rigorous. It is also a useful representation of the propagator for multiple remarkable reasons. As already mentioned the Hamiltonian is classical and not an operator such that it can be used to define a quantum system $\boldsymbol{H}$. A second reason is that the saddle point method gives an approximation scheme for U because it is the integral of a highly oscillating exponential and the critical points correspond to the classical path. A last reason is that Monte Carlo sampling leads to an approximation of the transition amplitude.

However, there is an ε and a sum in the resulting expression (1.1) such that the limes $\varepsilon \to 0$ or equivalently $m \to \infty$ for the continuous limit

$$U(\underline{q}_a,\underline{q}_b;T) = \left(\int \mathcal{D}\underline{q}\int \mathcal{D}\underline{p}\right) e^{i\,\int_0^T dt\left(\sum_j p^j\dot{q}^j\right)-H(\underline{q},\underline{p})\big)} \tag{1.2}$$

seems to make sense. (Note that the integrand in the exponent is the Lagrangian.) The $\dot{q}$ comes from $(q_{k+1}^j - q_k^j)$ multiplied by $\frac{\varepsilon}{\varepsilon}$. The problematic terms are the integrals between the first brackets because one has taken a finite number of integrals and takes now the limit of this number of integrals going to infinity. The symbol $\mathcal{D}$ stands for this limit. A finite number of integrals is comprehensible, but the question is what the limit is when the number of integrals becomes infinite. Without the i in the exponent this is a Wiener measure or a measure on Brownian motions but that only works without an imaginary term in the exponent. With the i this is not an integral and there is no space to integrate over such that the result is no longer rigorous. Going in figure 1 to larger and larger m the result is highly oscillatory and the paths are almost never continuous. In the sense of effective theories with a cutoff the equation (1.2) is just a symbol for equation (1.1) supposed to be applied in a lattice.

1.3 Calculating Continuous Path Integrals

To work with the transition amplitude $U(\underline{q}_a,\underline{q}_b;T)$ given by (1.2) one has to discretize it first by changing $\underline{q}(t) \to \underline{q}_k$ and by changing

$$\left(\prod_j \int \mathcal{D}q^j(t)\int \mathcal{D}p^j(t)\right) e^{i\,\int_0^T dt\,L(\underline{q},\dot{\underline{q}})} \quad \overset{\text{discretize}}{\to} \quad \left(\prod_{j,k}\int dq_k^j\int\frac{dp_k^j}{2\pi}\right) e^{i\,\sum_k\left(\sum_j p_k^j(q_{k+1}^j-q_k^j)-\varepsilon\,H\right)}$$

to get (1.1) because one can usually not work out the continuous path integral directly. (One can avoid the discretizing step only with a lot of experience.) Using the discretized path integral one gets a discrete path illustrated in figure 2 as an answer $I(\varepsilon)$ depending on ε.

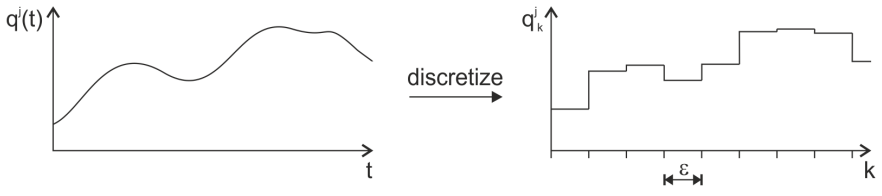


Figure 2: Discretization of a path

The final step is taking the limes $\lim_{\varepsilon\to 0} I(\varepsilon)$ if possible. One may have to deal at this point with infinities one encounters.

One should not think of the path integral as an integral. An important source of confusion in physics (and other sciences) is the use of wrong names. In equation (1.2) not only the term "integral" is used but there are even symbols for integrals. This is misleading because there is no underlying measure justifying this nomenclature. This equation is not an integral but a recipe.

As an example the classical Hamiltonian

$$H = \frac{p^2}{2m} + V(q)$$

is used and later it is assumed that $V(q)$ is quadratic such that this Hamiltonian represents a harmonic oscillator. To apply the above recipe to this Hamiltonian one has to solve Gaussian integrals such as

$$\int \frac{dp_k}{2\pi}\, e^{i\left(p_k(q_{k+1}-q_k)-\varepsilon\frac{p_k^2}{2m}\right)} = \frac{1}{C(\varepsilon)}\, e^{i\frac{m}{2\varepsilon}(q_{k+1}-q_k)^2} \qquad\qquad C(\varepsilon) = \sqrt{\frac{2\pi\,\varepsilon}{-i\,m}}$$

and the transition amplitude becomes

$$U(q_a, q_b; T) = \left(\prod_k \int \frac{dq_k}{C(\varepsilon)}\right) e^{i\sum_k\left[\frac{m}{2}\frac{(q_{k+1}-q_k)^2}{\varepsilon} - \varepsilon\, V\left(\frac{q_{k+1}+q_k)}{2}\right)\right]}$$

if one does not know more about $V(q)$. One can write this as

$$\lim_{\varepsilon\to 0} U(q_a, q_b; T) = \int \mathcal{D}q(t)\, e^{i\,S(q(t))} \qquad\qquad S(q(t)) = \int_0^T dt\left(\frac{m}{2}\sum_j (\dot{q}^j)^2 - V(q)\right)$$

where $S(q(t))$ is the action. If the potential is $V(q) = \frac{1}{2}m\,\omega\,q^2$ the full integral can be solved because the potential part is also a Gaussian integral.

1.4 Path Integrals for Relativistic Scalar Fields

The Lagrangian density $\mathcal{L}$ and the Hamiltonian H for a scalar field ϕ are

$$\mathcal{L} = \frac{1}{2}(\partial_\mu\phi)^2 - V(\phi) \qquad\qquad H = \int d^3\vec{x}\left[\frac{1}{2}\pi^2 + \frac{1}{2}(\nabla\phi)^2 + V(\phi)\right] \tag{1.3}$$

classically. Given the development of path integrals so far, it may be possible to do a path integral for the classical scalar field with the classical Hamiltonian and one gets out a quantum field at the end. Inserting the Hamiltonian blindly into (1.2) as one might try results in

$$\langle\phi_b|\, e^{-i\,\boldsymbol{H}\,T}\,|\phi_b\rangle = \left(\int \mathcal{D}\phi \int \mathcal{D}\pi\right) e^{i\int_0^T d^4X(\pi\,\dot{\phi} - H(\phi)} \tag{1.4}$$

with $\phi(0,\vec{x}) = \phi_a(X)$ and $\phi(T,\vec{x}) = \phi_b(X)$. Note that this defines a quantum field with $\boldsymbol{H}$.

Following the above recipe for calculating path integrals by discretizing the system, evaluating the integrals and taking the limit $\varepsilon \to 0$ allows to make sense of (1.4). The discretizing step turns spacetime into a grid $\phi(X) \to \phi(X_j) = q^j$ of width ε, integrals into sums $\int d^3\vec{x} \to \varepsilon^3 \sum_{j\in\mathbb{Z}}$, derivatives into differences

$$\nabla_\mu\phi(X) \to \frac{\phi(X_j + \varepsilon_\mu) - \phi(X_j)}{|\varepsilon_\mu|} \qquad\qquad \varepsilon_\mu \in \varepsilon\left\{\begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\0\\1\end{pmatrix}\right\}$$

and the potential into its values $V(\phi(X)) \to V(\phi(X_j))$ evaluated at the grid points. However, discretizing not only space but also time causes problems with the conjugate momenta π and only space can be discretized such that $\phi(t,\vec{x}) \to \phi(t,\vec{x}_j) = q^j(t)$. The Lagrangian is

$$L = \int d^3\vec{x}\,\mathcal{L} \overset{\text{discretize}}{=} \varepsilon^3 \sum_j \frac{1}{2}(\dot{\phi}_j)^2 + \ldots$$

where only the given discretized term matters. The conjugate momentum for the discretized theory is

$$\pi_j = \frac{\partial L}{\partial \dot{q}^j} = \varepsilon^3 \, \dot{q}^j$$

because $\dot{q}^j = \dot{\phi}_j = \dot{\phi}(t, \vec{x}_j)$. Thus, π_j picks up ε^3 which is there for dimensional reasons. If one forgets it one gets the wrong discretization, and this is emphasized here because it is tempting to forget it when just looking at the Hamiltonian H in(1.3). The discretized Hamiltonian for the field becomes

$$H_{\text{disc}} = \varepsilon^3 \sum_j \varepsilon^{-3} \, \pi_j^2 + \frac{1}{2} \left(\frac{q_{j+\varepsilon} - q_j}{\varepsilon} \right)^2 + V(q_j)$$

and one can do the non-relativistic path integral.

To summarize, the discretization step gives for the scalar field a lattice system in non-relativistic form and H becomes H_{disc} with a kinetic energy term and a potential energy term. The other two steps mean calculating Gaussian integrals to evaluate the discretized path integral and – as discussed below – renormalization for the continuum limit. The most important case is a potential in some quadratic form

$$V(\underline{q}) = \frac{1}{2} \, \underline{q}^T \, \underline{\underline{A}} \, \underline{q}$$

where $\underline{\underline{A}}$ is a matrix, and this will correspond to the Klein-Gordon theory. The discretized Klein-Gordon field results in such a potential and this means that one will get Gaussian integrals when evaluating it.

1.5 Gaussian Integrals

The simplest Gaussian integral is

$$I = \int_{-\infty}^{\infty} dx \, e^{-x^2} = \sqrt{\pi}$$

as the beautiful proof shows

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} dx \, e^{-x^2} \right) \left(\int_{-\infty}^{\infty} dy \, e^{-y^2} \right) = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \, e^{-(x^2+y^2)} \qquad (x,y) \to (r,\theta) \\ &= \int_0^{\infty} dr \, r \int_0^{2\pi} d\theta \, e^{-r^2} = 2\pi \int_0^{\infty} dr \, r \, e^{-r^2} = 2\pi \int_0^{\infty} \frac{d}{dr} \left(-\frac{e^{-r^2}}{2} \right) dr = \pi \end{aligned}$$

using $-\frac{e^{-\infty}}{2} = 0$ and $-(-\frac{e^0}{2}) = \frac{1}{2}$. However, the Gaussian integrals one will encounter in the following are not that easy and typically have the form

$$\int_{-\infty}^{\infty} dx \, e^{-\frac{1}{2}ax^2 + Jx} = \left(\frac{2\pi}{a} \right)^{\frac{1}{2}} e^{\frac{J^2}{2a}} \qquad \int_{-\infty}^{\infty} dx \, e^{i\,ax^2 + i\,Jx} = \left(\frac{2\pi\, i}{a} \right)^{\frac{1}{2}} e^{\frac{-i\,J^2}{2a}} \tag{1.5}$$

with a quadratic term and a linear term in the exponent where a and J can be complex numbers as the second example shows. (The constant J is the generating function for the Gaussian.) These two integrals are more than just examples because every single result in the path integral approach to quantum field theory will reduce to these two integrals.

In practice the moments

$$\langle x^n \rangle \equiv \frac{\int_{-\infty}^{\infty} dx \, x^n \, e^{-\frac{1}{2}ax^2}}{\int_{-\infty}^{\infty} dx \, e^{-\frac{1}{2}ax^2}} \tag{1.6}$$

are needed which are directly related to correlation functions. The Gaussian in the numerator of (1.6) is turned into a probability distribution by dividing it by the denominator. Because $\langle x^n \rangle$ is zero in case n is odd, one only has to determine it for even $n = 2m$ and gets

$$\langle x^{2m} \rangle = \frac{1}{a^m} \cdot (2m-1) \cdot (2m-3) \cdot \ldots \cdot 5 \cdot 3 \cdot 1 = \frac{(2m-1)!!}{a^m} \tag{1.7}$$

with the following trick

$$\langle x^{2m}\rangle = \left(\frac{d}{dJ}\right)^{2m} \frac{\int_{-\infty}^{\infty} dx\, e^{-\frac{1}{2}ax^2+Jx}}{\int_{-\infty}^{\infty} dx\, e^{-\frac{1}{2}ax^2}}\bigg|_{J=0} = \left(\frac{d}{dJ}\right)^{2m} \frac{\left(\frac{2\pi}{a}\right)^{\frac{1}{2}} e^{\frac{J^2}{2a}}}{\left(\frac{2\pi}{a}\right)^{\frac{1}{2}}}\bigg|_{J=0} = \left(\frac{d}{dJ}\right)^{2m} e^{\frac{J^2}{2a}}\bigg|_{J=0}$$

using the left Gaussian integral in (1.5). The result in (1.7) follows by induction. Note that the double factorial $(2m-1)!!$ is the number of ways of joining $2m$ points in pairs.

One can extend Gaussian integrals to multiple variables. With the vector $\underline{x} \in \mathbb{R}^n$ and the $n \times n$ symmetric real matrix $\underline{\underline{A}}$ the Gaussian distribution

$$I(\underline{\underline{A}}, \underline{J}) = \int_{-\infty}^{\infty} dx_1 \int_{-\infty}^{\infty} dx_2 ... \int_{-\infty}^{\infty} dx_n\, e^{-\underline{x}^T \underline{\underline{A}}\, \underline{x} + \underline{J}^T \underline{x}}$$

can be determined using the fact that there exists for any symmetric real matrix some orthogonal matrix $\underline{\underline{O}}$ such that $\underline{\underline{O}}^T \underline{\underline{D}}\, \underline{\underline{O}} = \underline{\underline{A}}$ where $\underline{\underline{D}}$ is diagonal with the eigenvalues λ_j. Defining $\underline{y} = \underline{\underline{O}}\, \underline{x}$ gives

$$I(\underline{\underline{A}}, \underline{0}) = \int_{-\infty}^{\infty} dy_1 ... \int_{-\infty}^{\infty} dy_n\, e^{-\underline{y}^T \underline{\underline{D}}\, \underline{y}} = \prod_{j=1}^{n} \int_{-\infty}^{\infty} dy_j\, e^{-y_j^2 \lambda_j} = \prod_{j=1}^{n} \sqrt{\frac{\pi}{\lambda_j}} = \sqrt{\frac{\pi^n}{\det \underline{\underline{A}}}}$$

for $\underline{J} = \underline{0}$ and

$$I(\underline{\underline{A}}, \underline{J}) = \sqrt{\frac{\pi^n}{\det \underline{\underline{A}}}}\, e^{\underline{J}^T \underline{\underline{A}}^{-1} \underline{J}} = \left(\frac{\pi^n}{\det \underline{\underline{A}}}\right)^{\frac{1}{2}} e^{\underline{J}^T \underline{\underline{A}}^{-1} \underline{J}} \tag{1.8}$$

for arbitrary $\underline{J}$. Knowing this generating function for the many-variable Gaussian probability distribution will be enough to do perturbative quantum field theory for all scalar fields.

1.6 Correlation Functions

The 2-point correlation function of the many-variable Gaussian distribution (1.8) is

$$\langle x_j x_k \rangle \equiv \frac{\int_{-\infty}^{\infty} dx_1 ... \int_{-\infty}^{\infty} dx_n\, e^{-\frac{1}{2}\underline{x}^T \underline{\underline{A}}\, \underline{x}}\, x_j\, x_k}{\int_{-\infty}^{\infty} dx_1 ... \int_{-\infty}^{\infty} dx_n\, e^{-\frac{1}{2}\underline{x}^T \underline{\underline{A}}\, \underline{x}}} = (\underline{\underline{A}}^{-1})_{jk} \tag{1.9}$$

for $j \neq k$. Note that differentiation with respect to J_j and J_k in

$$\langle x_j x_k \rangle = \frac{\frac{\partial^2}{\partial J_j\, \partial J_k} \int_{-\infty}^{\infty} dx_1 ... \int_{-\infty}^{\infty} dx_n\, e^{-\frac{1}{2}\underline{x}^T \underline{\underline{A}}\, \underline{x} + \underline{J}^T \underline{x}}}{\int_{-\infty}^{\infty} dx_1 ... \int_{-\infty}^{\infty} dx_n\, e^{-\frac{1}{2}\underline{x}^T \underline{\underline{A}}\, \underline{x} + \underline{J}^T \underline{x}}}\bigg|_{\underline{J}=\underline{0}}$$

produces a x_j and a x_k as needed.

Higher moments can be calculated similarly with more differentiations with respect to components of $\underline{J}$. The l-point correlation function for odd l is zero and for even l is

$$\begin{aligned}\langle x_{j_1} ... x_{j_l} \rangle &\equiv \frac{\int_{-\infty}^{\infty} dx_1 ... \int_{-\infty}^{\infty} dx_n\, e^{-\frac{1}{2}\underline{x}^T \underline{\underline{A}}\, \underline{x}}\, x_{j_1} ... x_{j_l}}{\int_{-\infty}^{\infty} dx_1 ... \int_{-\infty}^{\infty} dx_n\, e^{-\frac{1}{2}\underline{x}^T \underline{\underline{A}}\, \underline{x}}} \\ &\propto \sum_{\pi \in S_l} (A^{-1})_{j_{\pi^{-1}(1)} j_{\pi^{-1}(2)}} ... (A^{-1})_{j_{\pi^{-1}(l-1)} j_{\pi^{-1}(l)}}\end{aligned}$$

called Wick sum where S_l is the permutation group on l elements. (The proportionality comes from the redundant terms due to not yet eliminated overcounting.) The various elements $(\underline{\underline{A}}^{-1})_{jk}$ in this sum are the 2-point correlation functions (1.9), and this is extraordinary because once one has determined the 2-point correlation functions of a physical system by experiment or so then one knows all the higher order moments. That is the big advantage of Gaussian integrals. Note also that this result is just Wick's theorem in a different notation.

Because with $\underline{\underline{A}}$ also the inverse $\underline{\underline{A}}^{-1}$ is symmetric, the 2-point moments $(\underline{\underline{A}}^{-1})_{jk}$ and $(\underline{\underline{A}}^{-1})_{kj}$ are the same and should only be counted once. The 4-point correlation function $\langle x_{j_1} x_{j_2} x_{j_3} x_{j_4} \rangle$ as an example has $4! = 24$ permutations of the indices but only $\frac{24}{2!2!2!} = 3$ terms are unique and the symmetry factor is therefore 8. The result is

$$\begin{aligned}
\langle x_{j_1} x_{j_2} x_{j_3} x_{j_4} \rangle &= (A^{-1})_{j_1 j_2} (A^{-1})_{j_3 j_4} + (A^{-1})_{j_1 j_3} (A^{-1})_{j_2 j_4} + (A^{-1})_{j_1 j_4} (A^{-1})_{j_2 j_3} \\
&= \langle x_{j_1} x_{j_2} \rangle \langle x_{j_3} x_{j_4} \rangle + \langle x_{j_1} x_{j_3} \rangle \langle x_{j_2} x_{j_4} \rangle + \langle x_{j_1} x_{j_4} \rangle \langle x_{j_2} x_{j_3} \rangle
\end{aligned}$$

in three different notations.

The l-point correlation function is therefore

$$\langle x_{j_1} ... x_{j_l} \rangle = \sum_{\text{unique terms}} (A^{-1})_{j_{\pi^{-1}(1)} j_{\pi^{-1}(2)}} ... (A^{-1})_{j_{\pi^{-1}(l-1)} j_{\pi^{-1}(l)}} \tag{1.10}$$

where the unique terms are determined easiest by drawing the diagrams. In the case of 6 points, for example, there are $\frac{6!}{2!2!2!} = 90$ unique terms. Two important features have to be mentioned: Firstly, one can calculate all moments for a Gaussian distribution, and secondly, there is a diagrammatic calculus for them which will turn out to be the handling of Feynman diagrams. Thus, there is a direct connection between Feynman diagrams together with the whole perturbation calculus for quantum field theory and the Gaussian integrals.

A remaining question is what the matrix $\underline{\underline{A}}$ is in the case of perturbative quantum field theory. This question is discussed within non-relativistic quantum mechanics because a relativistic quantum field is nothing more than a non-relativistic quantum system with particular properties. With a quadratic potential V for a classical system the propagator for the system is

$$U(q_a, q_b; T) = \left(\prod_k \int \frac{dq_k}{C(\varepsilon)} \right) e^{\frac{i}{2} \underline{q}^T \underline{\underline{A}} \underline{q}}$$

where $\underline{q}^T \underline{\underline{A}} \underline{q}$ is

$$\underline{q}^T \underline{\underline{A}} \underline{q} = \sum_k \left(m \frac{(q_{k+1} - q_k)^2}{\varepsilon} - \varepsilon V \left(\frac{q_{k+1} + q_k)}{2} \right) \right)$$

with a kinetic energy term plus a potential energy term. The matrix $\underline{\underline{A}}$ for a chain of harmonic oscillators in one dimension is the kinetic energy matrix

$$\begin{pmatrix}
\frac{2m}{\varepsilon} & -\frac{m}{\varepsilon} & 0 & 0 & 0 & 0 & 0 & \cdots \\
-\frac{m}{\varepsilon} & \frac{2m}{\varepsilon} & -\frac{m}{\varepsilon} & 0 & 0 & 0 & 0 & \cdots \\
0 & -\frac{m}{\varepsilon} & \frac{2m}{\varepsilon} & -\frac{m}{\varepsilon} & 0 & 0 & 0 & \cdots \\
0 & 0 & -\frac{m}{\varepsilon} & \frac{2m}{\varepsilon} & -\frac{m}{\varepsilon} & 0 & 0 & \cdots \\
0 & 0 & 0 & -\frac{m}{\varepsilon} & \frac{2m}{\varepsilon} & -\frac{m}{\varepsilon} & 0 & \cdots \\
0 & 0 & 0 & 0 & -\frac{m}{\varepsilon} & \frac{2m}{\varepsilon} & -\frac{m}{\varepsilon} & \cdots \\
0 & 0 & 0 & 0 & 0 & -\frac{m}{\varepsilon} & \frac{2m}{\varepsilon} & \cdots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}$$

plus a potential energy matrix due to the quadratic potential V. Thus, the propagator is

$$U(q_a, q_b; T) = \text{infinite constant} \times \frac{1}{\sqrt{\det \underline{\underline{A}}}}$$

where the infinite constant is not a problem because one never measures the propagator. The operationally well-defined quantities are the l-point correlation functions and they are normalized such that the exact same infinite constant appears in the denominator and cancels. This means that the result for the correlation functions is a finite number.

To quantity

$$\langle x_{j_1}...x_{j_l}\rangle = \frac{\left(\prod_k \int \frac{dq_k}{C(\varepsilon)}\right) x_{j_1}...x_{j_l}\, e^{\frac{i}{2}\,\underline{q}^T\,\underline{\underline{A}}\,\underline{q}}}{\left(\prod_k \int \frac{dq_k}{C(\varepsilon)}\right) e^{\frac{i}{2}\,\underline{q}^T\,\underline{\underline{A}}\,\underline{q}}}$$

is finite as long as $\underline{\underline{A}}^{-1}$ exists, and this is a hint that it may be operationally well-defined. The question is how to interpret it in terms of the original quantum system because the right-hand side is a purely classical Gaussian integral.

Before discussing $\langle x_{j_1}...x_{j_l}\rangle$ further the matrix $\underline{\underline{A}}$ is discussed. It is a tridiagonal matrix such that its components with periodic boundary conditions are $f(j-k)$ modulo n for some function f where n is the number of oscillators in the chain. The matrix $\underline{\underline{Q}}$ built as

$$Q_{jk} = \frac{1}{\sqrt{n}}\, e^{2\pi i\, jk}$$

has the three properties $\underline{\underline{Q}}^\dagger\underline{\underline{Q}} = \underline{\underline{1}}$, $\underline{\underline{Q}}^4 = \underline{\underline{1}}$ and $\underline{\underline{Q}}^\dagger\underline{\underline{A}}\,\underline{\underline{Q}} = \underline{\underline{D}}$ where the first and the last are of interest here (and where $\underline{\underline{1}} = \mathbb{I}_{n\times n}$). The first property means that $\underline{\underline{Q}}$ is unitary, and the last property means that $\underline{\underline{Q}}$ diagonalizes $\underline{\underline{A}}$. Thus, this allows to determine

$$\det(\underline{\underline{A}}) = \prod_{j=1}^{n} \lambda_j(\underline{\underline{A}})$$

as the product of eigenvalues λ_j. Note also that matrix $\underline{\underline{Q}}$ acts as a discrete Fourier transform.

1.7 Connection Between Correlation Functions and Observable Quantities

One step is still missing because there is not yet an interpretation for the l-point correlation function $\langle x_{j_1}...x_{j_l}\rangle$. Its value is finite but being finite does not imply that it also has an operational meaning and is therefore useful.

To make sense of the path integral expression

$$\int \mathcal{D}\phi\, \phi(X_1)\phi(X_2)\, e^{i\int_{-T}^{T} d^4X\, \mathcal{L}(\phi)}$$

one has to discretize $\phi(X) \to \phi(X_j) = q_j$, evaluate and take the continuous limit as stated above. (With experience one can directly reason about this path integral expression without going through these three steps explicitly.) The boundary condition are $\phi(-T,\vec{x}) = \phi_a(\vec{x})$ and $\phi(T,\vec{x}) = \phi_b(\vec{x})$. The path integral measure on the left-hand side of

$$\int \mathcal{D}\phi = \int \mathcal{D}\phi_1(\vec{x}) \int \mathcal{D}\phi_2(\vec{x}) \int \mathcal{D}\phi(X)$$

where $\phi(X_1^0,\vec{x}_1) = \phi_1(\vec{x}_1)$ and $\phi(X_2^0,\vec{x}_2) = \phi_2(\vec{x}_2)$ can be replaced by the one on the right-hand side. This is like saying that integrating over a function is the same as integrating over a function where two points are fixed and then integrating over all values of those two points. Using this gives

$$\int \mathcal{D}\phi_1 \int \mathcal{D}\phi_2\, \phi_1(\vec{x}_1)\phi_2(\vec{x}_2)\, \langle\phi_b|e^{-i\,\boldsymbol{H}(T-X_2^0)}|\phi_2\rangle\, \langle\phi_2|e^{-i\,\boldsymbol{H}(X_2^0-X_1^0)}|\phi_1\rangle\, \langle\phi_1|e^{-i\,\boldsymbol{H}(X_1^0+T)}|\phi_a\rangle$$

where $\int \mathcal{D}\phi(X)$ has just been resolved. In the Schrödinger picture is $\boldsymbol{\phi}_{\mathrm{S}}(\vec{x})\,|\phi_1\rangle = \phi_1(\vec{x}_1)\,|\phi_1\rangle$ because a classical field configuration is an eigenstate of the field operator in the Schrödinger picture, and there is a completeness relation $\int \mathcal{D}\phi_1(\vec{x})\,|\phi_1\rangle\,\langle\phi_1| = \mathbb{I}$. This allows to put field operators into this equation. The result is for $X_2^0 > X_1^0$

$$\int \mathcal{D}\phi_2 \int \mathcal{D}\phi_1\, \langle\phi_b|e^{-i\,\boldsymbol{H}(T-X_2^0)}\boldsymbol{\phi}_{\mathrm{S}}(\vec{x})|\phi_2\rangle\, \langle\phi_2|e^{-i\,\boldsymbol{H}(X_2^0-X_1^0)}\boldsymbol{\phi}_{\mathrm{S}}(\vec{x})|\phi_1\rangle\, \langle\phi_1|e^{-i\,\boldsymbol{H}(X_1^0+T)}|\phi_a\rangle$$

and this is now also for $X_2^0 < X_1^0$

$$\langle\phi_b|e^{-i\,\boldsymbol{H}\,T}\mathcal{T}\left\{\boldsymbol{\phi}_{\rm H}(X_1)\boldsymbol{\phi}_{\rm H}(X_2)\right\}e^{-i\,\boldsymbol{H}\,T}|\phi_a\rangle$$

with the field operators in the Heisenberg picture. Taking the limes $T\to\infty$ leads to the expression

$$\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}_{\rm H}(X_1)\boldsymbol{\phi}_{\rm H}(X_2)\right\}|\Omega\rangle = \lim_{T\to\infty(1-i\varepsilon)}\frac{\int\mathcal{D}\phi\,\phi(X_1)\phi(X_2)\,e^{i\int_{-T}^{T}d^4X\,\mathcal{L}}}{\int\mathcal{D}\phi\,e^{i\int_{-T}^{T}d^4X\,\mathcal{L}}} \tag{1.11}$$

and this shows how to connect these path integrals on the right-hand side as classical objects that sums over classical data multiplied by a phase to the quantum observable on the right-hand side. Furthermore, the right-hand side can be evaluated when $\mathcal{L}$ is quadratic.

2 Functional Quantization of the Scalar Field

2.1 The Free Scalar Field in Position Space

Quantization of fields using the path integrals is much more efficient than using canonical quantization. The action of the classical Klein-Gordon field is

$$S_0 = \int d^4X\mathcal{L}_0 = \int d^4X\left(\frac{1}{2}(\partial_\mu\phi)^2 - \frac{1}{2}m^2\phi^2\right) \tag{2.1}$$

without interactions. One can apply the three steps for working with path integrals. Discretization is a way of introducing a cutoff Λ because it reduces the infinite number of degrees of freedom to a finite number.

In the first step one puts the field on a lattice but one also has to cut off infrared divergences because there is still an infinite number of degrees of freedom as also the integers go from minus infinity to plus infinity. The lattice is therefore compactified onto a torus of size L since there is no Gaussian measure for an infinite-dimensional Hilbert space. Thus, Minkowski spacetime $\mathbb{M}_{1,3}$ is mapped to $(\mathbb{Z}/N\mathbb{Z})^4$ where ε is the lattice spacing and $N = L/\varepsilon$ is supposed to be a rather big number. The field $\phi(X)$ is replaced with a list of N^4 numbers $\phi(X_j) = q_j$ where the field is sampled at $X_j \in \frac{L}{N}(\mathbb{Z}/N\mathbb{Z})^4$, and there are therefore N^4 classical position variables. (State of the art Monte Carlo calculations with supercomputers work in boxes of a side L about the proton radius and $N \approx 100$.)

The derivative and the integral are discretized by

$$\partial_\mu\phi(X) \approx \frac{\phi(X_j+\varepsilon e^\mu)-\phi(X_j)}{\varepsilon} \qquad\qquad \int d^4X \approx \varepsilon^4\sum_{j\in(\mathbb{Z}/N\mathbb{Z})^4}$$

where e^μ is a unit vector in the μ-direction. The path integral quantization recipe gives

$$\langle\phi_{\rm f}|U|\phi_{\rm i}\rangle \propto \int\mathcal{D}\phi\,e^{i\,S}$$

for the action S such that the last thing to be discretized is

$$\int\mathcal{D}\phi \equiv \prod_j\int d\phi(X_j) = \prod_j\int dq_j$$

assuming, of course, $\varepsilon > 0$.

2.2 The Free Scalar Field in Momentum Space

Moving from real space to momentum space simplifies the calculations because one knows how to diagonalize the Hamiltonian for the Klein-Gordon field and how to find the canonical degrees of freedom which are decoupled for the above action. The Fourier transformed classical field is

$$\phi(X_j) = \frac{1}{V}\sum_n e^{-i\,K_n\cdot X_j}\,\hat\phi(K_n) \qquad V = L^4 \qquad K_n^\mu = \frac{2\pi\,n^\mu}{L} \qquad n^\mu\in\mathbb{Z}/N\mathbb{Z} \qquad |K^\mu| < \frac{\pi}{\varepsilon}$$

where $\hat{\phi}$ is the field in momentum space. In the following, the distinction between ϕ and $\hat{\phi}$ will be resolved differently. If the argument of $\phi(.)$ is X, Y or Z the field is specified in spacetime, and if the argument is K, L or M the field is Fourier transformed to momentum space. Note that $\phi(K)$ is complex even if $\phi(X)$ is real and that $\phi(-K) = \phi^*(K)$ if $\phi(X)$ is real. The independent variables are $\operatorname{Re}\phi(K)$ and $\operatorname{Im}\phi(K)$ for $K_n^0 > 0$.

Since the Fourier transform is a unitary, the Jacobian is 1 and the path integral measure is

$$\int \mathcal{D}\phi = \prod_{n|_{K_n^0>0}} \int d\operatorname{Re}\phi(K_n)\, d\operatorname{Im}\phi(K_n)$$

in momentum space. The discretized action is

$$S_0 \approx -\frac{1}{V} \sum_{K_n^0>0} (m^2 - K_n^2) \left[\big(\operatorname{Re}\phi(K_n)\big)^2 + \big(\operatorname{Im}\phi(K_n)\big)^2 \right]$$

using a Kronecker delta $\frac{1}{n}\sum_{j=0}^{n-1} e^{\frac{2\pi\, i\, jk}{n}} = \delta_{k,0}$. Note that this action is decoupled in terms of the two independent variables $\operatorname{Re}\phi(K)$ and $\operatorname{Im}\phi(K)$. This gives the Gaussian integral

$$\begin{aligned} I_0 = \int \mathcal{D}\phi\, e^{i\,S_0} &= \left(\prod_{K_n^0>0} \int d\operatorname{Re}\phi(K_n)\, \operatorname{Im}\phi(K_n) \right) e^{-\frac{i}{V}\sum_{K_n^0>0}(m^2-K_n^2)|\phi(K_n)|^2} \\ &= \prod_{K_n^0>0} \sqrt{\frac{-i\,\pi\,V}{m^2-K_n^2}} \sqrt{\frac{-i\,\pi\,V}{m^2-K_n^2}} = \prod_{K_n} \sqrt{\frac{-i\,\pi\,V}{m^2-K_n^2}} \end{aligned}$$

for the path integral of the Klein-Gordon field on the lattice grid. Going to the continuum limit results in infinity because of V. Thus, this quantity is not operationally well-defined but since I_0 is only proportional and not equal to $\langle\phi_{\mathrm{f}}|U|\phi_{\mathrm{i}}\rangle$ this may turn out to be no problem.

A heuristic way to remember this result uses

$$\int d\underline{x}\, e^{-\underline{x}^T \underline{\underline{A}}\, \underline{x}} \propto \frac{1}{\sqrt{\det \underline{\underline{A}}}}$$

for the Gaussian integral and

$$\int \mathcal{D}\phi\, e^{i\,S} \simeq \int \mathcal{D}\phi\, e^{\frac{1}{2}\int d^4X\, \phi(X)(-\partial^2-m^2)\phi(X)} \qquad \int d^4X\, \phi(X)(-\partial^2-m^2)\phi(X) \approx \underline{x}^T \underline{\underline{A}}\, \underline{x}$$

for the path integral of the Klein-Gordon field. With the similarities $\phi(X) \sim \underline{x}$ and $(-\partial^2 - m^2) \sim \underline{\underline{A}}$ as a kind of continuous versions one can boldly extrapolate

$$\int \mathcal{D}\phi\, e^{i\,S} \propto \frac{1}{\sqrt{\det(-\partial^2-m^2)}}$$

where one can make sense of $\det\big(-\partial^2 - m^2\big)$ to a certain extent. Therefore, one could have omitted all the steps from discretization to the continuum limit. However, one has to know as a physicist when one can trust bold extrapolations and when not. If one has gone often enough through the three steps of discretization, evaluation and continuum limit, one will feel intuitively whether a result like this is to be trusted or not.

Because I_0 is infinite it should cancel for operationally well-defined quantities, and it does indeed. One can calculate the 2-point function (1.11)

$$\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2)\right\}|\Omega\rangle = \lim_{T\to\infty(1-i\varepsilon)} \frac{\int \mathcal{D}\phi\, \phi(X_1)\phi(X_2)\, e^{i\,S}}{\int \mathcal{D}\phi\, e^{i\,S}} = \lim_{T\to\infty(1-i\varepsilon)} \frac{(*)}{(**)}$$

by substituting the discretized Fourier sum

$$\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2) = \frac{1}{V^2} \sum_m e^{-i\,K_m\cdot X_1} \phi(K_m) \sum_l e^{-i\,K_l\cdot X_2} \phi(K_l)$$

into the discretized numerator $(*)$. This gives the mess

$$(*) = \frac{1}{V^2}\sum_{m,l} e^{-i(K_m\cdot X_1 + K_l\cdot X_2)} \left(\prod_{K_n^0>0}\int d\,\mathrm{Re}\,\phi(K_n)\,d\,\mathrm{Im}\,\phi(K_n)\right)\times$$
$$\big(\mathrm{Re}\,\phi(K_m) + i\ \mathrm{Im}\,\phi(K_m)\big)\big(\mathrm{Re}\,\phi(K_l) + i\ \mathrm{Im}\,\phi(K_l)\big) e^{-\frac{i}{V}\sum_{K_n^0>0}(m^2-K_n^2)((\mathrm{Re}\,\phi(K_n))^2+(\mathrm{Im}\,\phi(K_n))^2)}$$

but this mess is not too bad because many terms cancel as this is a bunch of independent Gaussian integrals multiplied by quadratic terms. Two integrals with odd integrands are zero and only two integrals remain. Furthermore, if $m \neq l$ or if $K_l = K_M$ then the integrals also vanish, and the only integrals not vanishing are the ones with $K_l = -K_m$. The result for the numerator after some steps is

$$(*) = \frac{1}{V^2}\sum_m e^{-i\,K_m\cdot(X_1-X_2)}\left(\prod_{K_n^0>0}\frac{-i\,\pi\,V}{m^2-K_n^2}\right)\frac{-i\,V}{m^2-K_n^2-i\varepsilon} = \frac{1}{V^2}\sum_m e^{-i\,K_m\cdot(X_1-X_2)} I_0 \frac{-i\,V}{m^2-K_n^2-i\varepsilon}$$

such that with $(**) = I_0$ the 2-point correlation function $\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2)\right\}|\Omega\rangle$ becomes

$$\frac{(*)}{(**)} = \frac{-i}{V}\sum_n \frac{e^{-i\,K_n\cdot(X_1-X_2)}}{m^2-K_n^2-i\varepsilon} \overset{V\to\infty}{=} \int\frac{d^4K}{(2\pi)^4}\,\frac{ie^{-i\,K\cdot(X_1-X_2)}}{-m^2+K_n^2+i\varepsilon}$$

and this is the propagator $D(X_1 - X_2)$.

With $\underline{\underline{A}} \sim (-\partial^2 - m^2)$ and $(A^{-1})_{jk} \sim \frac{1}{(-\partial^2-m^2)_{X_1X_2}} = D(X_1 - X_2)$ the expected similarity

$$\underline{\underline{A}}\,\underline{\underline{A}}^{-1} = \underline{\underline{1}} \qquad\sim\qquad (-\partial^2 - m^2)D(X_1 - X_2) = \delta^{(4)}(X_1 - X_2)$$

turns out to be correct. These heuristics allow to extrapolate

$$\frac{\int\mathcal{D}\phi\,\phi(X_1)\phi(X_2)\,e^{i\,S}}{\int\mathcal{D}\phi\,e^{i\,S}} = \frac{[-\partial^2-m^2]^{-\frac{1}{2}}\,D(X_1-X_2)}{[-\partial^2-m^2]^{-\frac{1}{2}}} = D(X_1-X_2)$$

and this gives the same result. This analogy between multivariable Gaussian integrals and continuous functional integrals works very well.

Applying this heuristic approach to 4-point correlation functions as the next higher order since 3-point correlation functions are zero due to their odd integrand gives a path integral

$$\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2)\boldsymbol{\phi}(X_3)\boldsymbol{\phi}(X_4)\right\}|\Omega\rangle = \lim_{T\to\infty(1-i\varepsilon)}\frac{\int\mathcal{D}\phi\,\phi(X_1)\phi(X_2)\phi(X_3)\phi(X_4)\,e^{i\,S}}{\int\mathcal{D}\phi\,e^{i\,S}}$$

where most of the terms in $(\mathrm{Re}\,\phi(K_m)+i\,\mathrm{Im}\,\phi(K_m)...(\mathrm{Re}\,\phi(K_l)+i\,\mathrm{Im}\,\phi(K_l)$ cancel. The remaining terms are like $K_l = -K_m$ and $K_q = -K_p$. With Wick's theorem and the limit $V \to \infty$ the result is

$$D_\mathrm{F}(X_1-X_2)D_\mathrm{F}(X_3-X_4) + D_\mathrm{F}(X_1-X_3)D_\mathrm{F}(X_2-X_4) + D_\mathrm{F}(X_1-X_4)D_\mathrm{F}(X_2-X_3)$$

and the path integral prescription gives precisely the same answer as the canonical quantization.

2.3 The Interacting Scalar Field

One of the remaining questions is whether one can get all of perturbative quantum field theory from the path integral formalism. The action for the ϕ^4-theory is

$$S = S_0 + S_\mathrm{int} = S_0 + \frac{i\,\lambda}{4!}\int d^4X\,\phi^4(X) \tag{2.2}$$

with the action of the free scalar field S_0 from (2.1). The 2-point correlation function for S is

$$\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2)\right\}|\Omega\rangle = \lim_{T\to\infty(1-i\varepsilon)}\frac{\int\mathcal{D}\phi\,\phi(X_1)\phi(X_2)\,e^{i\,(S_0+S_\mathrm{int})}}{\int\mathcal{D}\phi\,e^{i\,(S_0+S_\mathrm{int})}}$$
$$= \lim_{T\to\infty(1-i\varepsilon)}\frac{\int\mathcal{D}\phi\,\phi(X_1)\phi(X_2)\,e^{i\,S_0}(1+S_\mathrm{int}+\frac{S_\mathrm{int}^2}{2!}+...)}{\int\mathcal{D}\phi\,e^{i\,S_0}(1+S_\mathrm{int}+\frac{S_\mathrm{int}^2}{2!}+...)}$$

for a small λ such that perturbative quantum field theory with Taylor expansion can be applied. Inserting S_{int} from (2.2) gives

$$\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2)\right\}|\Omega\rangle = \lim_{T\to\infty(1-i\varepsilon)} \frac{\int \mathcal{D}\phi\, \phi(X_1)\phi(X_2)\, e^{i\,S_0}(1+\frac{i\lambda}{4!}\int d^4X\, \phi^4(X)+...)}{\int \mathcal{D}\phi\, e^{i\,S_0}(1+\frac{i\lambda}{4!}\int d^4X\, \phi^4(X)+...)}$$

with terms of the form

$$\frac{\lambda^m}{(4!)^m}\int d^4Z_1... \int d^4Z_m \int \mathcal{D}\phi\, \phi(X_1)\phi(X_2)\phi^4(Z_1)...\phi^4(Z_m)\, e^{i\,S_0)} =$$
$$\frac{\lambda^m}{(4!)^m}\int d^4Z_1... \int d^4Z_m \,\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2)\boldsymbol{\phi}^4(Z_1)...\boldsymbol{\phi}^4(Z_m)\right\}|\Omega\rangle$$

where field operators are in the Heisenberg picture. With the path integration formalism one gets the same terms and therefore the same Feynman diagrams as with canonical quantization.

2.4 Functional Derivatives and Generating Functionals

As shown above the Klein-Gordon operator $(-\partial^2 - m^2)$ is analogous to $\underline{\underline{A}}$ and the Feynman propagator $D_{\text{F}}(X_1 - X_2)$ is analogous to $\underline{\underline{A}}^{-1}$ in (1.9). This analogy can be made clearer and more precise. There is a way to carry out the n-point functions in the continuum of the path integrals using a formalism called functional derivatives and generating functions.

Given a many-variable Gaussian integral one can set up the generating functional

$$Z[\underline{J}] = \int dx_1... \int dx_n\, e^{-\frac{1}{2}\,\underline{x}^T\underline{\underline{A}}\,\underline{x}-\underline{J}^T\underline{x}} = e^{\frac{1}{2}\,\underline{J}^T\underline{\underline{A}}^{-1}\,\underline{J}}$$

and by differentiating this generating functional at $\underline{J} = \underline{0}$ one can bring down as many x as needed for the computation of the n-point function. This approach can be generalized.

The basics of functional derivatives which belong to the calculus of variations are here introduced by examples. The functional derivative is defined by

$$\frac{\delta}{\delta J(X)}J(Y) = \delta^{(4)}(X-Y) \tag{2.3}$$

where $\frac{\delta}{\delta J(X)}$ is the functional derivative. The result of the functional derivative applied to a function is in this case a generalized function. This is like the continuum analog of the standard partial derivative

$$\frac{\delta}{\delta J(X)}J(Y) = \delta^{(4)}(X-Y) \qquad\qquad \frac{\partial}{\partial x_k}x_j = \delta_{jk}$$

or in more detail

$x \in \mathbb{R}$	$j \in \mathbb{Z}$
$J(x) \in C(\mathbb{R})$	$J_j \in L_2(\mathbb{Z})$
$\frac{\delta}{\delta J(x)}F\big(J(y)\big)$	$\frac{\partial}{\partial J_j}f(J_1, J_2, ...)$

where the left side shows the continuum analog and the right side the corresponding discrete version. (The discretizing of $J(x)$ usually gives x_j but to make the distinction between the continuum and discrete side clear the variables J_j have been used instead of x_j.) One follows the usual rules for derivatives. The way how one works with functional derivatives is illustrated by two examples.

In the first example

$$\frac{\delta}{\delta J(X)}\,\exp\left[i\int d^4Y\, J(Y)\,\phi(Y)\right]$$

the exponential function results in a number depending on $J(X)$ such that this is something the functional derivative can process. This gives

$$\begin{aligned}\frac{\delta}{\delta J(X)}\,\exp\left[i\int d^4Y\,J(Y)\,\phi(Y)\right] &= i\,\exp\left[i\int d^4Y\,J(Y)\,\phi(Y)\right]\,\frac{\delta}{\delta J(X)}\left[i\int d^4Y\,J(Y)\,\phi(Y)\right]\\ &= i\,\exp\left[i\int d^4Y\,J(Y)\,\phi(Y)\right]\int d^4Y\,\delta^{(4)}(X-Y)\,\phi(Y)\\ &= i\,\phi(X)\,\exp\left[i\int d^4Y\,J(Y)\,\phi(Y)\right]\end{aligned}$$

using the chain rule.

The second example presents the functional derivative of a delta function. If one moves the functional derivative past the integral in

$$\frac{\delta}{\delta J(X)}\int d^4Y\,\partial_\mu J(Y)\,V^\mu(Y)$$

the result is the derivative of a delta function. In this example one can integrate by parts first to get

$$\begin{aligned}\frac{\delta}{\delta J(X)}\int d^4Y\,\partial_\mu J(Y)\,V^\mu(Y) &= \frac{\delta}{\delta J(X)}\left(-\int d^4Y\,J(Y)\,\partial_\mu V^\mu(Y) + \text{boundary term}\right)\\ &= -\partial_\mu V^\mu(X) + \text{boundary term}\end{aligned}$$

where the boundary term is almost always zero except for some topological theories.

The generating functional defined as

$$Z[J] = \lim_{T\to\infty(1-i\varepsilon)}\int \mathcal{D}\phi\,e^{i(S+J(x)\,\phi(x))} \tag{2.4}$$

is certainly useful because correlation functions $\langle\Omega|\mathcal{T}\boldsymbol{\phi}(X)\,\boldsymbol{\phi}(Y)|\Omega\rangle$ are directly related to derivatives

$$\langle\Omega|\mathcal{T}\boldsymbol{\phi}(X)\,\boldsymbol{\phi}(Y)|\Omega\rangle = -\frac{\dfrac{\delta}{\delta J(X)}\dfrac{\delta}{\delta J(Y)}\,Z[J]\Big|_{J=0}}{Z[J]\big|_{J=0}}$$

of this object. If one can compute $Z[J]$ as a function of J then one gets all the n-point functions for the field theory by functional derivatives evaluated at $J=0$ without the three steps of discretizing, evaluating and taking the continuum limit. It is not always possible to evaluate $Z[J]$ because the action contains non-linear terms, but if one can evaluate it one has solved the theory.

In free theories the action S_0 is quadratic. The exponent in the generating functional is

$$i(S_0 + J\,\phi) = i\int d^4\left[\frac{1}{2}\phi(X)\,(-\partial^2 - m^2 + i\varepsilon)\,\phi(X) + J(X)\,\phi(X)\right]$$

and becomes by completing the square with the new variable $\phi'(X) = \phi(X) - i\int d^4Y D_{\mathrm{F}}(X-Y)\,J(Y)$

$$\begin{aligned}i(S_0 + J\,\phi) = &\int d^4X\left[\frac{1}{2}\phi'(X)\,(-\partial^2 - m^2 + i\varepsilon)\,\phi'(X)\right]\\ &- i\int d^4X\int d^4Y\,\frac{1}{2}\,J(X)\,(-i\,D_{\mathrm{F}}(X-Y))\,J(Y)\end{aligned}$$

where the shift from ϕ to ϕ' is analogous to the shift from $\underline{x}$ to $\underline{x}' = \underline{x} - \underline{\underline{A}}^{-1}\underline{J}$. The Jacobian is 1 because it is just a shift. The generating functional is

$$Z[J] = Z_0\,\exp\left[-\frac{1}{2}\int d^4X\int d^4Y\big(J(X)\,D_{\mathrm{F}}(X-Y)\,J(Y)\big)\right]$$

where Z_0 is

$$Z_0 = \int \mathcal{D}\phi' \, \exp\left[i\,\phi'(-\partial^2 - m^2 + i\varepsilon)\phi'\right]$$

and does not depend on J. One never needs Z_0 because it cancels

$$\begin{aligned}\langle\Omega|\mathcal{T}\boldsymbol{\phi}(X)\,\boldsymbol{\phi}(Y)|\Omega\rangle &= -\frac{\dfrac{\delta}{\delta J(X)}\dfrac{\delta}{\delta J(Y)}\, Z_0\, \exp\left[-\frac{1}{2}\int d^4X \int d^4Y \big(J(X)\, D_{\mathrm{F}}(X-Y)\, J(Y)\big)\right]\Big|_{J=0}}{Z_0} \\ &= \frac{\delta}{\delta J(X)}\frac{\delta}{\delta J(Y)}\, \exp\left[-\frac{1}{2}\int d^4X \int d^4Y \big(J(X)\, D_{\mathrm{F}}(X-Y)\, J(Y)\big)\right]\Big|_{J=0}\end{aligned}$$

in actual calculations.

The 4-point correlation function for the Klein-Gordon theory as an example gives

$$\begin{aligned}\langle 0|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)\boldsymbol{\phi}(X_2)\boldsymbol{\phi}(X_3)\boldsymbol{\phi}(X_4)\right\}|0\rangle &= \frac{\dfrac{\delta}{\delta J(X_1)}\dfrac{\delta}{\delta J(X_2)}\dfrac{\delta}{\delta J(X_3)}\dfrac{\delta}{\delta J(X_4)}\, Z[J]\Big|_{J=0}}{Z[0]} \\ &= \frac{\delta}{\delta J(X_1)}\frac{\delta}{\delta J(X_2)}\frac{\delta}{\delta J(X_3)}\left[\left(-\int d^4Z\, J(Z)D(Z-X_4)\right) e^{-\frac{1}{2}\int d^4X \int d^4Y\, J(X)D(X-Y)J(Y)}\right]\Big|_{J=0} \\ &= \frac{\delta}{\delta J(X_1)}\frac{\delta}{\delta J(X_2)}\left[\left(-D(X_3-X_4) + \int d^4Z_1\, J(Z_1)D(Z_1-X_3)\int d^4Z_2\, J(Z_2)D(Z_2-X_4)\right) e^{\cdots}\right]\Big|_{J=0} \\ &= \frac{\delta}{\delta J(X_1)}\Bigg[\Bigg(D(X_3-X_4)\int d^4Z_1\, J(Z_1)D(Z_1-X_2) + D(X_2-X_4)\int d^4Z_2\, J(Z_2)D(Z_2-X_3) \\ &\qquad + D(X_2-X_3)\int d^4Z_3\, J(Z_3)D(Z_3-X_4) + O(J^2)\Bigg) e^{-\frac{1}{2}\int d^4X \int d^4Y\, J(X)D(X-Y)J(Y)}\Bigg]\Big|_{J=0} \\ &= D(X_3-X_4)D(X_1-X_2) + D(X_2-X_4)D(X_1-X_3) + D(X_2-X_3)D(X_1-X_4)\end{aligned}$$

as expected. Thus, the combination of generating functionals and functional derivatives turned out to be a very efficient tool to calculate correlation functions.

The correlation function for the scalar field with interactions in the ϕ^4-theory is

$$\langle\Omega|\mathcal{T}\left\{\boldsymbol{\phi}(X_1)...\boldsymbol{\phi}(X_n)\right\}|\Omega\rangle = \lim_{T\to\infty(1-i\varepsilon)} \frac{\int \mathcal{D}\phi\, e^{i\,S_0 + i\,S_{\mathrm{int}}}\, \phi(X_1)...\phi(X_n)}{\int \mathcal{D}\phi\, e^{i\,S_0 + i\,S_{\mathrm{int}}}}$$

with $i\,S_{\mathrm{int}} = \int d^4X\, \frac{\lambda}{4!}\phi^4(X)$. In this case, the generating functional is used after the exponent has been expanded into a Taylor series. The result is also the same as for the canonical quantization. The generating functional is just an elegant tool to gain the n-point correlation functions for free theories but with interactions it is not as useful.

3 Functional Quantization of the Dirac Field

3.1 Properties of Classical Fermions

In nature there are not only bosons but also fermions. If one wants to use path integals for fermions one has to make sense of (1.11) because on the left-hand side there are quantum objects where all the quantum operators are well defined for fermions but the question is what the right-hand side means. For the bosons the right-hand side is an integral over all classical field configurations. Assuming the same means that the path integral for fermions is a sum over all field configurations of a classical fermion field whatever that may be.

One might guess that the correlation functions for fermions can be determined using

$$\langle\Omega|\mathcal{T}\left\{\boldsymbol{\psi}(X)\boldsymbol{\psi}(Y)\right\}|\Omega\rangle \stackrel{?}{=} \lim_{T\to\infty(1-i\varepsilon)} \frac{\int \mathcal{D}\psi\, e^{i\,S}\, \psi(X)\psi(Y)}{\int \mathcal{D}\psi\, e^{i\,S}}$$

where the right side is the path integral of something classical. Thus, the question is what a classical fermion might be. The left-hand side of this equation is operationally well-defined. However, one should not think about the right-hand side operationally but similarly to complex number. A complex number is like a mathematical phantasm and one cannot see them in the daily experience. One can either say that the equation $x^2 = -1$ has no solution or one can introduce one and see how far one comes with it. Complex numbers are a kind of shortcut and one should see the right-hand side of the above equation similarly.

One can try to build a classical object satisfying the two properties

$$\psi(X)^2 = 0 \qquad\qquad \psi(X)\psi(Y) = -\psi(Y)\psi(X)$$

required for a classical fermionic field. This gives something similar to complex number that is useful for doing calculations. These objects are called anticommuting number or Grassmann numbers.

3.2 Functions of Grassmann Numbers

If V is an n-dimensional vector space with basis $\theta_a \in V$ for $a = 1, ..., n$ then every vector $v \in V$ can be represented in the form of a linear combination

$$v = \sum_{a=1}^{n} v_a \, \theta_a$$

of the basis vectors. One can build a bigger vector space $\mathcal{G}_n(V)$ from V where the vectors θ are the generators. This bigger vector space should be equipped with a multiplication denoted by concatenation such as $\theta_a\theta_b\theta_c$ and extended linearly. If one would stop here one would have $\mathcal{S}^\infty(V) = \text{span}\{\theta_a, \theta_a\theta_b, \theta_a\theta_b\theta_c, ...\}$ as the linear span of concatenations of these symbols. This would be an infinite-dimensional vector space if there are no constraints on the product. However, the product should obey the relations

$$\theta_a^2 = \theta_a\theta_a = 0 \;\; \forall\, a \qquad\qquad \theta_a\theta_b = -\theta_b\theta_a \;\; \forall\, a, b$$

where the relation on the left follows from the one on the right but is explicitly written to be emphasized. The relation on the right can also be written as $\theta_a\theta_b + \theta_b\theta_a = 0$. Imposing these relations on $\mathcal{S}^\infty(V)$ gives $\mathcal{G}_n(V) = \mathcal{S}^\infty(V)/\sim$ with dimension $\dim(\mathcal{G}_n(V)) = 2^n$. This is a space of differential forms when the vector space is a tangent space.

The so-called monomial basis of $\mathcal{G}_n(V)$ with 2^n basis vectors is

$$\{1, \theta_a, \theta_a\theta_b, \theta_a\theta_b\theta_c, ...\} \qquad\qquad 1 \le a < b < c ... \le n$$

or more detailed

$$\{1, \theta_1, \theta_2, ..., \theta_n, \theta_1\theta_2, \theta_1\theta_3, ..., \theta_{n-1}\theta_n, \theta_1\theta_2\theta_3, ..., \theta_1\theta_2\theta_n, \theta_1\theta_3\theta_4, ..., \theta_1\theta_2\theta_3\theta_4, ...\}$$

and a general element $f \in \mathcal{G}_n(V)$ is a linear combination of these basis vectors

$$f = \alpha\, 1 + \sum_{p=0}^{n} \sum_{1 \le j_1 < ... < j_p \le n} \alpha_{j_1 ... j_p} \, \theta_{j_1} ... \theta_{j_p} \qquad\qquad \alpha, ... \in \mathbb{C}$$

where 1 is added as a basis vector for the concatenation of zero θ_j. This additional basis vector 1 satisfies the relation $1\,\theta_a = \theta_a$ for all a. The objects θ_a are anticommuting numbers or variables.

The next step is to introduce functions of Grassmann variables. (Trying to introduce $f \in C^\infty(\mathbb{R})$ with Grassmann variables, for example, and to adjoin $f(\theta_j)$ as an object that anticommutes with everything results in an uncountable number of objects and is not what one expects.) Functions are supposed to have some structure. One would like to start with linear functions, and linear functions should map elements of the vector space $\mathcal{G}_n(V)$ to elements of the vector space $\mathcal{G}_n(V)$. Thus, linear functions are represented by $2^n \times 2^n$ matrices.

Non-linear functions are more difficult but one can define them by analogy with functional calculus. One takes some operator and defines some function of that operator by diagonalizing a matrix and then

applying the function to the eigenvalues. If the matrix $\underline{\underline{M}}$ is diagonalizable then $f(\underline{\underline{M}}) = \underline{\underline{S}}^{-1} f(\underline{\underline{D}}) \underline{\underline{S}}$ where $\underline{\underline{S}}^{-1} \underline{\underline{D}} \underline{\underline{S}} = \underline{\underline{M}}$. One strategy is to work with Taylor series and another strategy is to first represent $\mathcal{G}_n(V)$ as matrices and then use functional calculus.

Using the second strategy means to find a representation $\pi : \mathcal{G}_n(V) \to M_d(\mathbb{C})$ of Grassmann numbers as matrices with the anticommutation relations $\{\pi(\theta_j), \pi(\theta_k)\} = 0$ for all j, k. One example of such a representation could be called Jordan-Wigner representation, and it starts from the Hilbert space $\mathfrak{H} = \mathbb{C}^{2^n}$ and the two Pauli operators

$$\sigma^+ = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \qquad\qquad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

with the properties

$$\{\sigma^+, \sigma^z\} = 0 \qquad\qquad (\sigma^+)^2 = 0$$

such that

$$\begin{aligned} \pi(\theta_1) &= \sigma_1^+ \otimes \mathbb{I}_2 \otimes \mathbb{I}_3 \otimes ... \otimes \mathbb{I}_{n-1} \otimes \mathbb{I}_n \\ \pi(\theta_2) &= \sigma_1^z \otimes \sigma_2^+ \otimes \mathbb{I}_3 \otimes ... \otimes \mathbb{I}_{n-1} \otimes \mathbb{I}_n \\ &\dots \\ \pi(\theta_n) &= \sigma_1^z \otimes \sigma_2^z \otimes \sigma_3^z \otimes ... \otimes \sigma_{n-1}^z \otimes \sigma_n^+ \end{aligned}$$

is a representation of $\mathcal{G}_n(V)$.

This representation allows to define what functions $F \in C^\infty(\mathbb{R}, \mathbb{R})$ there are with Grassmann variables, and the single variable case $n = 1$ and $\mathcal{G}_1(V) \cong \{a + b\,\theta \,|\, a, b \in \mathbb{C}\}$ should satisfy

$$F(\pi(\ )) = S^{-1} F(D)\, S$$

to be consistent with the functional calculus on the representation π. One can do that by writing out the Taylor series and by replacing x by θ

$$F(x) = \sum_{j=0}^{\infty} \frac{F^{(j)}(0)\, x^j}{j!} \qquad \to \qquad F(\theta) = \sum_{j=0}^{\infty} \frac{F^{(j)}(0)\, \theta^j}{j!} = F^{(0)}(0) + F^{(1)}(0)\, \theta$$

because $\theta^n = 0$ for $n \geq 2$. Note that many functions which are different for ordinary variables are the same for Grassmann variables as the example

$$F(x) = \sin(x) = (x - \frac{x^2}{2!} + ...) \to F(\theta) = \theta \qquad\qquad F(x) = x + x^3 \to F(\theta) = \theta$$

shows. The many variable case with $F \in C^\infty(\mathbb{R}^n, \mathbb{R})$ can also be handled with Taylor series such that

$$F(\theta_1, ..., \theta_m) = F^{(0)}(0, ..., 0) + \sum_{j=1}^{m} \theta_j \frac{\partial}{\partial x_j} F(0, ..., 0) + ...$$

is the definition for multivariable functions. There are higher order terms for functions with two or more variables as the two examples $F(x, y) \in C^\infty(\mathbb{R}^2, \mathbb{R})$

$$\begin{aligned} F(\theta_1, \theta_2) &= \exp(-\lambda\, \theta_1\, \theta_2) \\ &= 1 - \lambda\, \theta_1\, \theta_2 + \lambda^2\, \theta_1^2\, \theta_2^2 + ... \\ &= 1 - \lambda\, \theta_1\, \theta_2 \end{aligned} \qquad\qquad \begin{aligned} G(\theta_1, \theta_2) &= \exp(-\lambda_1\, \theta_1)\, \exp(-\lambda_2\, \theta_2) \\ &= (1 - \lambda_1\, \theta_1)(1 - \lambda_2\, \theta_2) \\ &= 1 - \lambda_1\, \theta_1 - \lambda_2\, \theta_2 + \lambda_1\, \lambda_2\, \theta_1\, \theta_2 \end{aligned}$$

for $\lambda, \lambda_1, \lambda_2 \in \mathbb{R}$ illustrate. All non-linear functions collapse to linear functions in the single-variable case but higher order terms such as $\theta_1\, \theta_2$ remain for multiple variables.

To introduce complex Grassmann numbers one defines

$$\theta = \frac{\theta_1 + i\, \theta_2}{\sqrt{2}} \qquad\qquad \theta^* = \frac{\theta_1 - i\, \theta_2}{\sqrt{2}}$$

for $\theta_1, \theta_2 \in \mathcal{G}_2(V)$. The two objects θ and θ^* are elements of $\mathcal{G}_2(V)$. By defining the operation * as

$$\theta_j^* = \frac{\theta_{j_1} - i\, \theta_{j_2}}{\sqrt{2}} \qquad\qquad \theta_{j_1}, \theta_{j_2} \in \mathcal{G}_{2n}(V)$$

the definition of complex Grassmann numbers has been extended to the many variable case.

3.3 Calculus with Grassmann Numbers

Firstly, an operation is introduced that can be called a derivative. The derivative is a derivation in the sense of a linear map $\partial_{\theta_j} : \mathcal{G}_n(V) \to \mathcal{G}_n(V)$ obeying the two conditions

$$\partial_{\theta_j}(\theta_k) = \delta_{j,k} \qquad \partial_{\theta_j}(\theta_{k_1}...\theta_{k_p}) = \delta_{j,k_1}\theta_{k_2}...\theta_{k_p} - \delta_{j,k_2}\theta_{k_1}\theta_{k_3}...\theta_{k_p} + ... + (-1)^p\delta_{j,k_p}\theta_{k_1}\theta_{k_2}...\theta_{k_{p-1}}$$

where the rule on the right side is a kind of product rule. Thus, $\partial_{\theta_1}(\theta_1\theta_2) = \theta_2$ as with ordinary numbers but $\partial_{\theta_2}(\theta_1\theta_2) = \partial_{\theta_2}(-\theta_2\theta_1) = -\partial_{\theta_2}(\theta_2\theta_1) = -\theta_1$ due to the anticommutativity. The rule is to bring first the θ_j in ∂_{θ_j} to the front and then differentiate. This definition of the derivative extended by linearity to functions obeys the product rule and the chain rule.

Secondly, the definite integral should take some Grassmann numbers and return a complex number. It should obey shift invariance $\theta \to \theta + \eta$ in analogy to the integral from minus infinity to plus infinity and the map $\mathcal{G}_n(V) \to \mathbb{C}$ should be linear. The only definition for the case of a single variable which is consistent with these requirements is

$$\int (a + b\theta)\, d\theta = b \qquad\qquad \forall\, a, b \in \mathbb{C}$$

where the limits have been dropped because they do not make sense anyway. This leads to the surprising identity

$$\int (a + b\theta)\, d\theta = \partial_\theta(a + b\theta)$$

and integrals are the same as derivatives for Grassmann numbers in the case of single variables. This is also true for the case of many variables (consistent with the shift invariance)

$$\int d\theta_n\, d\theta_{n-1}... d\theta_1 \Big(f_0 + \sum_{j_1<...<j_p} f_p(j_1, ..., j_p)\, \theta_{j_1}...\theta_{j_p}\Big) = f_n(1, ..., n)$$

where $f_n(1, ..., n)$ is the highest term in the Taylor series. This definition behaves correctly under changes of variables. As an example the complex Gaussian

$$\int d\theta^* d\theta \left(e^{-\lambda\, \theta^*\theta}\right) = \int d\theta^* d\theta (1 - \lambda\, \theta^*\theta) = \partial_{\theta^*}\partial_\theta(1 - \lambda\, \theta^*\theta) = \lambda$$

is used where θ^* and θ are simply independent variables.

The many variable Gaussian integral

$$\int d\theta_1^* d\theta_1 ... d\theta_n^* d\theta_n\, e^{-\sum_{j,k} \theta_j^* B_{jk} \theta_k}$$

is elaborated as a more advanced example. Note that the order in which the integrations take place is not important as long as it is done in pairs of $d\theta_j^* d\theta_j$ with $d\theta_j$ before $d\theta_j^*$. The matrix $\underline{\underline{B}}$ satisfies $\underline{\underline{B}}^\dagger = \underline{\underline{B}}$ and is therefore Hermitian. The first step is a change of variables $\theta'_j = U_{jk}\theta_k$ to make $\underline{\underline{B}}$ diagonal where $\underline{\underline{U}}$ is unitary and $\underline{\underline{U}}^\dagger \underline{\underline{B}}\, \underline{\underline{U}} = \underline{\underline{D}}$. This gives

$$\begin{aligned}\theta'_1\theta'_2...\theta'_n &= \sum_{k_1,...,k_n} (U_{1k_1}U_{2k_2}...U_{nk_n})\theta_{k_1}\theta_{k_2}...\theta_{k_n} \\ &= \sum_{\pi\in S_n} \big(U_{1\pi(1)}U_{2\pi(2)}...U_{n\pi(n)}\, \mathrm{sgn}(\pi)\big)\theta_1\theta_2...\theta_n = \det\big(\underline{\underline{U}}\big)\theta_1\theta_2...\theta_n\end{aligned}$$

with $k_j = \pi(j)$. Thus, the Gaussian integral becomes

$$\begin{aligned}I &= \int d\theta_1^* d\theta_1 ... d\theta_n^* d\theta_n\, e^{-\sum_{j,k} \theta_j^* B_{jk} \theta_k} = \int d\theta'^*_1 d\theta'_1 ... d\theta'^*_n d\theta'_n\, e^{-\sum_{j,k} \theta'^*_j (UBU^\dagger)_{jk} \theta'_k} \det\big(\underline{\underline{U}}\big) \det\Big(\underline{\underline{U}}^\dagger\Big) \\ &= \int d\theta'^*_1 d\theta'_1 ... d\theta'^*_n d\theta'_n\, e^{-\lambda_1\, \theta'^*_1\theta'_1} \cdot ... \cdot e^{-\lambda_n\, \theta'^*_n\theta'_n} = \lambda_1\lambda_2...\lambda_n = \det\big(\underline{\underline{B}}\big)\end{aligned}$$

where λ_j are the eigenvalues of $\underline{\underline{B}}$. This result is astonishing because of the fact that the result for the ordinary Gaussian integral is $I = (\det(\underline{\underline{B}}))^{-1}$. The determinant of $\underline{\underline{B}}$ is a nasty non-linear function of $\underline{\underline{B}}$ making many calculations in quantum field theory difficult.

The generating function for the Gaussian integral is

$$Z[\underline{J}] = \int d\theta_1^* d\theta_1 ... d\theta_n^* d\theta_n \, e^{-\underline{\theta}^\dagger \underline{\underline{B}}\, \underline{\theta} + \underline{J}^\dagger \underline{\theta} + \underline{\theta}^\dagger \underline{J}} = e^{\underline{J}^\dagger \underline{\underline{B}}\, \underline{J}} \tag{3.1}$$

where $\underline{J}$ is a vector of Grassmann numbers. This results allows to calculate all moments such as

$$\int d\theta_1^* d\theta_1 ... d\theta_n^* d\theta_n \, \theta_j \theta_k^* \, e^{\underline{\theta}^\dagger \underline{\underline{B}}\, \underline{\theta}} = \det(\underline{\underline{B}}) B_{jk}$$

just like the classical moments.

One can now mix ordinary numbers and Grassmann numbers. If one defines

$$\underline{\Phi} = \begin{pmatrix} x_1 \\ \dots \\ x_n \\ \theta_1 \\ \dots \\ \theta_n \end{pmatrix} \qquad\qquad \underline{\underline{M}} = \begin{pmatrix} \underline{\underline{A}} & \underline{\underline{0}} \\ \underline{\underline{0}} & \underline{\underline{A}} \end{pmatrix}$$

where $\underline{\underline{A}}$ is an $n \times n$ matrix and $\underline{\underline{M}}$ therefore a $2n \times 2n$ matrix then the double Gaussian integral

$$\int d\underline{\Phi}\, e^{-\underline{\Phi}^\dagger \underline{\underline{M}}\, \underline{\Phi}} = \int dx_1 ... dx_n dx_1^* dx_n^* \int d\theta_1^* d\theta_1 ... d\theta_n^* d\theta_n \, e^{-\underline{\Phi}^\dagger \underline{\underline{M}}\, \underline{\Phi}} = \det(\underline{\underline{A}}) \cdot \det(\underline{\underline{A}})^{-1} = 1$$

gets a very simple result. This is the basis for the supersymmetric method.

3.4 The Classical Dirac Field

Mathematics with Grassmann numbers can be used to build an object similar to a path integral but for the Dirac field. In the generating function (3.1) for a Gaussian integral

$$Z[\underline{J}] = \left(\prod_{j=1}^{n} \int d\theta_j^* d\theta_j \right) e^{-\sum_{j,k} \theta_j^* B_{jk} \theta_k + \sum_j (J_j^* \theta_j + J_j \theta_j^*)} = e^{-\sum_{j,k} J_j^* (B^{-1})_{jk} J_k}$$

with n complex Grassmann numbers and Hermitian $\underline{\underline{B}}$ the anticommutation relations

$$\{\theta_j, \theta_k^*\} = \{\theta_j, \theta_k\} = \{\theta_j^*, \theta_k^*\} = \{J_j, \theta_k\} = \{J_j^*, \theta_k\} = \{J_j, J_k^*\} = 0$$

and so on are satisfied. The function $Z[\underline{J}]$ takes a vector $\underline{J}$ of Grassmann numbers as anticommuting input and produces commuting numbers as output. Expectation values of observables should never be Grassmann numbers.

The Dirac field $\psi(X)$ has four components and transforms as $\psi(X) \to M(\Lambda)\psi(\Lambda^{-1}X)$ where

$$M = \exp\left(-\frac{i}{2} \omega_{\mu\nu} S^{\mu\nu} \right) \qquad\qquad S^{\mu\nu} = \frac{i}{4}[\gamma^\mu, \gamma^\nu] \qquad\qquad \{\gamma^\mu, \gamma^\nu\} = 2\,\eta^{\mu\nu}$$

is the representation of the Lorentz group, and the Dirac equation is $(i\partial^\mu - m)\psi = 0$ from the Lagrangian density $\mathcal{L} = \bar{\psi}(i\,\partial\!\!\!/ - m)\psi$ where $\bar{\psi} = \psi^\dagger \gamma^0$ and $A\!\!\!/ = A_\mu \gamma^\mu$. The quantities ψ in the quantum Dirac field correspond to operators that anticommute and the classical limit should therefore be something that anticommutes too.

When building a quantum theory for the Dirac field using the approach of canonical quantization one looks at the single-particle theory, turns the quantities ψ into field operators, comes up with some algebra for the field operators to obey and looks for representations of this algebra. An alternative way is to

imagine that one has some classical Dirac field which one can quantize by the path integral. There are many advantages doing it this way. Assuming that the classical Dirac field is based on ordinary numbers for its values fails as the bosonic quantization of the Dirac field does. The values have to be Grassmann numbers.

The question is how one can make fields whose values are Grassmann numbers. The correct way is via sheaf theory and ringed spaces but the alternative way via discretization on a lattice works as well. Minkowski spacetime with $X \in \mathbb{M}_{1,3}$ is discretized on a torus with $X_j = \varepsilon\, j$ for $j \in \frac{L}{N}(\mathbb{Z}/N\mathbb{Z})^4$ as in the bosonic case above.

This also discretizes the classical field $\psi(X)$ as $\psi_j = \psi(X_j) = \psi(\varepsilon\, j)$ and as $\psi_j^\dagger = \psi^\dagger(\varepsilon\, j)$. Because the Dirac field ψ has four components ψ_a with $a \in \{1,2,3,4\}$ the actual values of the discretized field are in $[\psi_j]_a$ and the values of these quantities are Grassmann numbers. This defines the classical Dirac field not via sheaf theory but discretized on the lattice as the list

$$\big[[\psi_j]_a, [\psi_j^\dagger]_a\big] \qquad\qquad j \in \frac{L}{N}(\mathbb{Z}/N\mathbb{Z})^4 \qquad\qquad a \in \{1,2,3,4\}$$

of $8 \cdot N^4$ Grassmann numbers. That is all what is needed to apply path integrals. The continuous classical Dirac field is therefore the limit for $\varepsilon \to 0$, $N \to \infty$ and $L \to \infty$. In momentum space one has to truncate the description of the classical field to a finite number of Fourier modes to discretize the field and make the Fourier coefficients Grassmann numbers.

3.5 The Quantum Dirac Field via the Path Integral Formalism

In order to apply the path integral formalism the continuous Lagrangian density $\mathcal{L} = \bar\psi(i\,\partial\!\!\!/ - m)\psi$ must be discretized. One way to do that gives

$$\mathcal{L}(\psi_j, \psi_j^*) = \sum_{j \in (\mathbb{Z}/N\mathbb{Z})^4} i\,\bar\psi_j\,\gamma^\mu \frac{\psi_{j+\hat\mu} - \psi_j}{\varepsilon} - m\,\bar\psi_j\psi_j$$

where $\hat\mu$ is a unit vector in direction μ. This allows to define the 2-point function

$$\langle 0|\mathcal{T}\left\{\boldsymbol{\psi}(X)\bar{\boldsymbol{\psi}}(Y)\right\}|0\rangle = \lim_{T\to\infty(1-i\varepsilon)} \frac{\int \mathcal{D}\bar\psi\mathcal{D}\psi\, \psi(X)\bar\psi(Y)\, e^{i\int_{-T}^{T} d^4X\,\mathcal{L}}}{\int \mathcal{D}\bar\psi\mathcal{D}\psi\, e^{i\int_{-T}^{T} d^4X\,\mathcal{L}}} \tag{3.2}$$

in analogy to (1.11). The result is

$$\langle 0|\mathcal{T}\left\{\boldsymbol{\psi}(X)\bar{\boldsymbol{\psi}}(Y)\right\}|0\rangle = S_{\mathrm{F}}(X-Y) = \int \frac{d^4K}{(2\pi)^4}\, \frac{i\, e^{i\,K\cdot(X-Y)}}{K\!\!\!/ - m + i\varepsilon} \tag{3.3}$$

after applying the recipe for path integrals with discretization, evaluation and continuum limit.

There is a subtlety here called the fermion doubling problem. If one defines fermions on a lattice and takes the continuum limit then one gets extra fermions called doublers that do not go away in the continuum limit. This comes from the dispersion relation and is a topological result. The dispersion relation for fermions on a ring has to cross the real axis twice. One does not see these doublers in nature. By doing the continuum limit not at the end but during the integral process one does not encounter this problem.

The generating functional for the Dirac field is defined as

$$Z[J(X), \bar J(Y)] = \int \mathcal{D}\bar\psi\,\mathcal{D}\psi\, \exp\!\left(i\int d^4X\,\big(\bar\psi(i\,\partial\!\!\!/ - m)\psi + \bar J\psi + J\bar\psi\big)\right)$$

and all the correlation functions can be obtained from a Gaussian integral using the generating functional. The quantities $J(X)$ and $\bar J(Y)$ are Grassmann-valued source fields. By completing the square and some calculations one gets

$$Z[J,\bar J] = Z_0\, \exp\left[-\int d^4X \int d^4Y \big(\bar J(X)\, S_{\mathrm{F}}(X-Y)\, J(Y)\big)\right]$$

where $Z_0 = Z[0, \bar{0}]$.

The difficulty of Wick's theorem for fermions in the canonical quantization is keeping track of the minus signs. Here the proof is easy due to the rules for Grassmann derivatives

$$\frac{d}{d\eta}\theta\,\eta = -\theta\frac{d}{d\eta}\eta = -\theta$$

where θ and η are Grassmann numbers. This rule takes care of all the minus signs and permutations. The n-point function becomes

$$\langle 0|\mathcal{T}\left\{\boldsymbol{\psi}^{(\alpha_1)}(X_1)...\boldsymbol{\psi}^{(\alpha_n)}(X_n)\right\}|0\rangle = Z_0^{-1}\left(i\,(-1)^{(\alpha_1+1)}\frac{\delta}{\delta J^{(\alpha_1)}}\right)...\left(i\,(-1)^{(\alpha_n+1)}\frac{\delta}{\delta J^{(\alpha_n)}}\right)Z[J,\bar{J}]$$

where the notation

$$\psi^{(\alpha)}(X) = \begin{cases}\psi(X) & \text{for } \alpha = 0 \text{ (modulo 2)}\\ \bar{\psi}(X) & \text{for } \alpha = 1 \text{ (modulo 2)}\end{cases}$$

is used because there is no easy way to specify that an object can either be ψ or $\bar{\psi}$. The 2-point correlation function is

$$\langle 0|\mathcal{T}\left\{\boldsymbol{\psi}(X)\bar{\boldsymbol{\psi}}(Y)\right\}|0\rangle = S_{\mathrm{F}}(X-Y)$$

as to be expected.

3.6 Interacting Quantum Scalar Fields and Quantum Dirac Fields

One can see the path integrals as a kind of Wick's theorem generator. One inputs a Gaussian theory and it outputs the correct Wick's theorem. For interactions one puts in some non-linear terms for the Lagrangian, uses Taylor expansion, encounters higher and higher n-point functions and can use the path integral to apply Wick's theorem. The Feynman rules are a way to collect together the patterns one detects when applying Wick's theorem to a Taylor series. Once one knows the Feynman rules one can forget the path integral and just use the Feynman rules. The path integral is good for efficiently guessing Feynman rules.

Interactions between bosons and fermions are demonstrated with the theory of massive quantum electrodynamics. Because gauge theories have not yet been introduced quantum electrodynamics cannot be properly developed but there is something that looks very much like quantum electrodynamics

$$\mathcal{L} = \bar{\psi}\,(i\,\gamma^{\mu}(\partial_{\mu} - i\,e\,A_{\mu}) - m)\,\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\,\mu^2\,A_{\mu}A^{\mu}$$

and that is called massive quantum electrodynamics. There is the gauge field A consisting of boson fields and there is an action for this gauge field. The action is with $F^{\mu\nu} = \partial^{\mu}A^{\nu} - \partial^{\nu}A^{\mu}$ the Maxwell action $-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ which has a gauge invariance producing all kinds of interesting divergences in path integrals but one can get rid of all of them by adding a mass term $\frac{1}{2}\,\mu^2\,A_{\mu}A^{\mu}$ to the gauge field where $e \approx \frac{1}{137}$ is the fine structure constant, m is the electron mass and μ is the photon mass (and, independently, also the summation index). The photon is actually assumed to be massless within very small bounds. The massive quantum electrodynamics is a possible theory of reality but it is not the one physicists think that it is the correct one.

The theory has a Dirac field ψ and four boson fields A_{μ}. Path integral quantization gives

$$\langle 0|\mathcal{T}\left\{\boldsymbol{\psi}(X)\bar{\boldsymbol{\psi}}(Y)\right\}|0\rangle = \lim_{T\to\infty(1-i\varepsilon)}\frac{\int \mathcal{D}\bar{\psi}\mathcal{D}\psi\mathcal{D}A\,\psi(X)\,\bar{\psi}(Y)\,e^{i\,S}}{\int \mathcal{D}\bar{\psi}\mathcal{D}\psi\mathcal{D}A\,e^{i\,S}}$$

for a process where only fermions and no bosons are observed at the end. The action

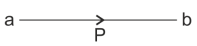
$$S = S_0 + S_{\mathrm{int}} \qquad S_0 = \int d^4X\,\bar{\psi}(i\,\partial\!\!\!/ - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}\,\mu^2\,A_{\mu}A^{\mu} \qquad S_{\mathrm{int}} = -i\,e\int d^4X\,\bar{\psi}A_{\mu}\gamma^{\mu}\psi$$

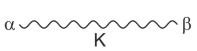
splits up into the easy Gaussian action S_0 for the free bosons and fermions and the difficult but small action S_{int} for the interaction between bosons and fermions.

The 2-point correlation function with S_{int} expanded as a Taylor series is

$$\langle 0|\mathcal{T}\left\{\psi(X)\bar{\psi}(Y)\right\}|0\rangle = \lim_{T\to\infty(1-i\varepsilon)} \frac{\int \mathcal{D}\bar{\psi}\mathcal{D}\psi\mathcal{D}A\,\psi(X)\,\bar{\psi}(Y)\,e^{i\,S_0}(1-i\,e\int d^4X\,\bar{\psi}A_\mu\gamma^\mu\psi+...)}{\int \mathcal{D}\bar{\psi}\mathcal{D}\psi\mathcal{D}A\,e^{i\,S_0}(1-i\,e\int d^4X\,\bar{\psi}A_\mu\gamma^\mu\psi+...)}$$

where the 0^{th}-order term is a Gaussian integral with the value $S_{\text{F}}(X-Y)$. This is an infinite series of n-point correlation functions for Gaussian path integrals in powers of e. One can see patterns and the different terms can be represented by Feynman diagrams obeying the Feynman rules:

1. Draw a line for ferminons 

and a line for bosons 

and associate to each fermion line $\left(\frac{1}{\not{P}-m+i\varepsilon}\right)_{ab}$
and to each boson line $\left(\frac{-i}{K^2-\mu^2+i\varepsilon}\right)\delta_{\alpha\beta}$
2. Associate to each vertex a factor $i\,e\,\gamma^\mu$
3. Enforce momentum conservation

$$(2\pi)^4\delta^{(4)}\left(\sum_{\text{in}} P - \sum_{\text{out}} Q\right)$$

4. Integrate over undetermined momenta
5. Amputate external lines
6. Incoming fermions get $\eta^a(P)$ and outgoing fermions get $\bar{\eta}^b(P)$
7. For each closed fermion loop multiply with a factor (-1)

Although the result of this is infinity it works but one has to change how one regards these fields and the remedy is called renormalization.

4 Renormalization

4.1 Progress in Physics

Physics starts with observations which lead to empirical data. In the beginning this is just a list of unstructured data such as numbers and this list gets longer and longer as time passes. The second step is to try to explain the data. With only a few data items the best explanation is the data itself but it becomes harder and harder to memorize this list of data as it grows bigger and bigger. Thus, an explanation is a form of data compression. However, any kind of data compression alone is not what one wants. Human beings want as a third step understanding, and there is an evolutionary reason for that because understanding helps avoid dangers and catch the prey. Understanding can mean various things. Having found a Hamiltonian may mean understanding, having defined a neural network or even a list of data. In physics there is a strong prejudice towards understanding the universe via a Hamiltonian or via a Lagrangian. Today the universe is seen as based on quantum processes because this turned out to be the best way to explain all the data collected so far. Understanding is important but there is a fourth step because physicists want to make predictions based on a hypothesis. This is a key part of physics because it allows to explain what might happen if one makes a new observation. One can never confirm a hypothesis but only reject it. At this point one has to go back to step one with the observations.

This iterative process of explaining, understanding and rejecting makes the list of models which explain everything increasingly smaller and smaller as illustrated in figure 3. As physicists make better and better observations they are left with fewer and fewer explanations which are consistent with all the observations so far. However, humans do not want to reject hypotheses but confirm hypotheses. As time passes and there are less and less explanations consistent with the observations one starts to develop

prejudices towards a certain explanation on the basis of simplicity. There is always a guaranteed complete explanation for all collected data. It is the list of all observations but it is not simple and one does not gain understanding from it. Simplicity is very hard to quantify but it is obvious that people want something simple. For physicists the simplest explanation is a Hamiltonian with a Hilbert space where one integrates the Schrödinger equation.

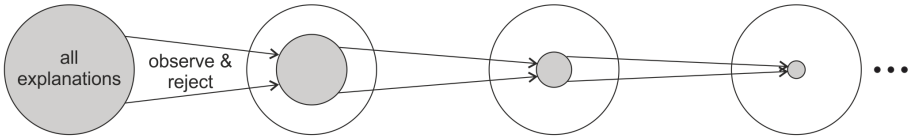


Figure 3: Progress in physics reducing the possible explanations by rejection

In the process of reducing the space of possible explanations in figure 3 physicists are focusing on certain explanations because they seem simpler in their eyes, and those are models in form of Hamiltonians $H(z_1, ..., z_n)$ depending on a finite number of unknown parameters $z_j \in \mathbb{R}$. The Hamiltonian can be a classical quantity H or a quantum operator $\boldsymbol{H}$ along with a Hilbert space $\mathfrak{H}$.

The Hamiltonian and the Hilbert space are actually $\boldsymbol{H}_\Lambda(z_1, ..., z_n)$ and $\mathfrak{H}_\Lambda$ where Λ is a list of degrees of freedom one wants to explain. It is not known whether the number of degrees of freedom for the universe is finite or infinite. There is some evidence that it is finite and some evidence that it is infinite but that is the beauty of physics that it can explain with a model like this an infinite list of degrees of freedom with a finite number of parameters. However, since only a finite number of experiments have been performed so far, every explanation pertains to a finite number of degrees of freedom.

As stated above one collects observational data during the first step and this results in expectation values $\langle A_j \rangle = \alpha_j \pm \delta\alpha_j$ for $j = 1, ..., m$ with some errors or uncertainties $\delta\alpha_j$. It is assumed that one can pinpoint the expectation values such that $\delta\alpha_j = 0$. In the second step one takes all the models $\boldsymbol{H}_\Lambda(z_1, ..., z_n)$ not yet rejected and sees what they predict for $\langle A_j \rangle$. If the parameters $z_1, ..., z_n$ do not lead to the found answer α_j the model is rejected. Thus, there is a map

$$f_j(z_1, ..., z_n; \Lambda) = \langle A_j \rangle \tag{4.1}$$

in the end where $f_j(z_1, ..., z_n; \Lambda)$ are exact solutions. This map is usually horribly many-to-one because multiple sets of parameters $z_1, ..., z_n$ and Λ provide an explanation for the same observations, and this map is therefore not invertible.

It is a fact of life that there are multiple explanations for all the observations made so far. The best model today is linearized gravity minimally coupled to the standard model with some cutoff but there are many other models explaining all observed data. However, if one demands that the explanation is as simple as possible with a given notion of what simplicity means then there remains only one model although this will change over time. There is a notion of simplicity available in physics for good one hundred years where one model is simpler than another model if it requires fewer parameters to explain all observations and applies to more degrees of freedom. Ideally, physicists would like a model that requires zero parameters and applies to all the degrees of freedom that are required to describe the universe. That is what string theory aims for.

A further question is how one interprets these parameters $z_1, ..., z_n$. They are the coupling constants and are not necessarily directly measurable such that they may not be operationally well-defined. There are quantities in physics that can be infinity. Inverse temperature, for example, is infinity if the universe is in the ground state. This happens all the time in quantum field theory but it does not lead to operational problems because the predictions will always be finite (and real numbers).

4.2 Infinities in Quantum Field Theories

The above listed steps for making progress in physics can be carried out in the context of quantum field theory but one encounters many infinities here. Some infinities can be shifted away without operational consequences as, for example, infinite ground-state energy outside of the context of general relativity.

Other infinities such as the amputate diagrams take a bit more work but are also justifiable. However, there are also infinite predicted quantities. The scattering cross-sections as an example of this kind is examined in the following within the context of the ϕ^4-theory although everything can be generalized to any other quantum field theory.

To summarize the process, one makes a list of predictions $f_j(z_1, ..., z_n; \Lambda)$ and compares them with the experimental observations $\langle A_j \rangle = \alpha_j$. Predictions must be finite and real, and it is possible that infinite parameters z_j give finite predictions.

The Lagrangian of the ϕ^4-theory is

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \phi)^2 - \frac{1}{2} m^2 \phi^2 - \frac{\lambda}{4!} \phi^4$$

and leads to many infinities. Thus, it looks like this model must be rejected because computations above for the scattering cross-section, for example, have given $f_j(m, \lambda) = \infty$. These computations have only been approximate and it may be that the approximation scheme has been wrong. However, there are so many independent calculations by now that the general consensus is that the problem is not in the approximations. It looks more like the models have not been interpreted adequately. The infinities have come out of $f_j(m, \lambda)$ for fixed $m \neq 0$ and $\lambda \neq 0$.

However, there is also Λ as the list of all degrees of freedom to be explained. So far this list has been all momentum modes and that is very ambitious. Thus, one can restrict this list to a number which is far beyond any value ever measured and that is what is called imposing a cutoff $|\Lambda| < \infty$. This approach has several problems. One is that this cutoff is completely arbitrary and another one is that the predictions $f_j(z_1, ..., z_n; \Lambda) = \langle A_j \rangle$ in (4.1) may depend on the chosen cutoff.

There is one last trick left and that comes from the fact that the parameters z_j are not directly observable and can depend on the chosen cutoff. If one can invert $f_j(z_1, ..., z_n; \Lambda)$ to get z_j not as the one true coupling constant but as one possible coupling constant $z_j = z_j(\Lambda)$ that is allowed to depend on Λ then this is sufficient as long as the predictions are consistent with all the observations made so far. The model $\boldsymbol{H}_\Lambda$ allows therefore predictions up to the cutoff Λ.

Note that the measured mass of the Higgs-boson is $125.10 \pm 0.14\,\text{GeV}$ but the coupling constant m called mass is a coupling constant and not its mass. The quantity m in the Lagrangian density $\mathcal{L}$ is equal to the mass of the Higgs-boson only in the low-energy limit but is not directly measurable as a coupling constant. It is therefore very unfortunate that this coupling constant is labeled m because it makes one think that it has an operational meaning.

Even if the parameters z_j are allowed to depend on the cutoff Λ it is far from obvious that it is possible to keep all observable quantities independent of this cutoff. A theory which allows

$$f_j(z_1(\Lambda), z_2(\Lambda), ..., z_n(\Lambda); \Lambda) = \langle A_j \rangle \qquad \forall \Lambda \text{ and fixed } n \tag{4.2}$$

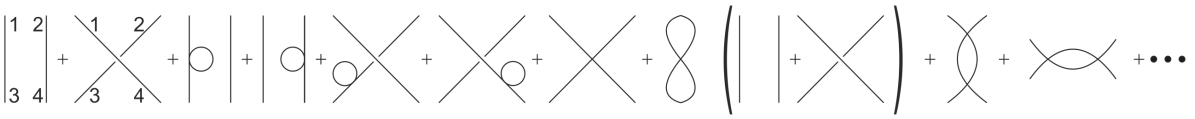
is called renormalizable. In other word, a theory that allows for a fixed n to absorb the cutoff Λ into redefining the parameters $z_1, ..., z_n$ such that the predictions are matching the observed data $\langle A_j \rangle$ is a renormalizable theory.

To do this for the ϕ^4-theory the focus is on the scattering amplitude. As an astonishing fact it is generally the case for quantum field theories that focusing on the scattering amplitude and possibly one or two more operationally well-defined quantities one can already prove that some theories obey the property of being renormalizable.

4.3 Renormalized Scattering Amplitude

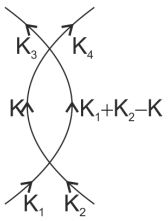
The scattering amplitude for the ϕ^4-theory in the case of $|\Lambda| = \infty$

$\langle P_1 P_2 | \mathsf{S} | P_3 P_4 \rangle =$



is the sum of all possible Feynman diagrams, and a Feynman diagram is just an integral. One computes all the integrals represented by these diagrams and gets the result infinity. One can get rid of some of the infinities by redefining parameters such as the ground-state energy.

The first infinity that cannot be eliminated for $|\Lambda| = \infty$ by a simple redefinition of E_0 and m_0 or by rescaling λ is the Feynman diagrams in the figure on the right side shown in momentum space. It has one undefined momentum K and the corresponding integral is



$$I = \frac{1}{2}(-i\lambda)^2 \int \frac{d^4K}{(2\pi)^4} \frac{i}{K^2 - m^2 + i\varepsilon} \frac{i}{(K_1 + K_2 - K)^2 - m^2 + i\varepsilon}$$

where the denominator scales as K^4 and the integral is over d^4K such that the divergence is basically logarithmic.

The cutoff selected here is a momentum cutoff $|K| < K_c \in \mathbb{R}$, and the integral I becomes

$$I = \frac{1}{2}(-i\lambda)^2 i^2 \int_\Lambda \frac{d^4K}{(2\pi)^4} \frac{1}{K^2 - m^2 + i\varepsilon} \frac{1}{(K_1 + K_2 - K)^2 - m^2 + i\varepsilon} = 2\,i\,C\,\log\left(\frac{K_c^2}{(K_1 + K_2)^2}\right)$$

where C is some constant one can work out but this is not needed here. The goal is therefore to adjust the parameters $z_1 = m$ and $z_2 = \lambda$ to eliminate this logarithmic divergence. The scattering amplitude to order $O(\lambda^2)$ is

$$\mathcal{M}(K_c) = -i\lambda + i\,C\,\lambda^2 \left[\log\left(\frac{K_c^2}{(K_1 + K_2)^2}\right) + \log\left(\frac{K_c^2}{(K_1 - K_3)^2}\right) + \log\left(\frac{K_c^2}{(K_1 - K_4)^2}\right)\right]$$

and depends on K_c.

If λ is assumed to be one true value set at the creation of the universe, the only possible conclusion is that this quantum field theory has to be rejected. However, if one allows $z_2 = z_2(K_c) = \lambda(K_c)$ then there is a solution. One cannot directly observe λ because one observes $\lambda + \lambda^2 + ...$ Thus, the question is how to change λ such that the scattering amplitude no longer depends on K_c and is equal to the experimental value. To determine $\lambda(K_c)$ one assumes $\mathcal{M}(K_c, \lambda) = \mathcal{M}_{\text{exp}}$ where $\mathcal{M}_{\text{exp}}$ has been measured experimentally. If λ follows K_c according to the differential equation

$$K_c \frac{d\lambda}{dK_c} = 6C\,\lambda^2 + O(\lambda^3)$$

then it is possible to compensate the dependence on K_c with the changing λ. This is legal because $z_2 = \lambda$ has no direct operational meaning. Thus, as long as λ depends on the momentum cutoff K_c in a fashion that obeys this differential equation one can explain all measured scattering amplitudes once and for all.

This is just one example of how renormalization works, but the question remains what about other divergences coming from three-particle or four-particle scattering and all the other experiments one can perform. The amazing result is that one can eliminate all the infinities by only rescaling m and λ. That is an astonishing result.

4.4 Renormalizability of Quantum Field Theories

The renormalizability of the ϕ^4-theory has been studied and proven by Bogoljubow, Parasiuk, Hepp and Zimmermann (known as the BPHZ theorem). The main points of this result is presented here. It involves an enormous amount of combinatorics.

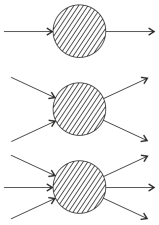
As introduced above renormalizability means a family of cutoff theories determined by a finite number of free parameters $\boldsymbol{H}(z_1, ..., z_n; \Lambda)$ such that for all observable quantities $\boldsymbol{A}_\alpha$ (self-adjoined operators) the expectation values $\langle \boldsymbol{A}_\alpha \rangle (z_1 ... z_n; \Lambda)$ depending on the parameters $z_1, ..., z_n$ and the cutoff Λ match the result of the experiments α by allowing the parameters $z_j(\Lambda)$ to depend on the cutoff Λ. Thus, by redefining $z_j = z_j(\Lambda)$ one can still match experimentally determinable quantities for any choice of Λ.

The weak form of renormalizability requires that by redefining $z_j = z_j(\Lambda)$ the observable quantities $\langle \boldsymbol{A}_\alpha \rangle_{z_1(\Lambda) ... z_n(\Lambda)}$ match experimentally determinable quantities only to order $\frac{1}{K_c}$. This means that observable data does not have to match exactly but only up to corrections that decay as this cutoff is taken

to infinity. That is the best one can get from an actual quantum field theory such as the ϕ^4-theory. Thus, there will be some observable quantities that depend on the cutoff but they depend on it in such a way that for a larger cutoff the dependence gets smaller.

The ϕ^4-theory is determined by a Hamiltonian that depends on three parameters $\boldsymbol{H}(m, \lambda, z; \Lambda)$ where z is the field strength renormalization, and the question is whether the ϕ^4-theory is renormalizable. For $\lambda = 0$ and $z = 1$ this is the free Klein-Gordon field and this is renormalizable. The challenge is to figure out whether the interacting ϕ^4-theory is renormalizable. The BPHZ theorem does not achieve this goal because all the arguments are perturbative. Even if the perturbation theory for the ϕ^4-theory does not depend on the cutoff Λ that does not mean that the ϕ^4-theory does not depend on the cutoff Λ. The full proof that a quantum field theory does not depend on the cutoff has only been done in 1+1 dimensions. In higher dimension the proof has only been performed on the level of perturbation theory. However, that is so strong and matches the experiments so well that the ϕ^4-theory can be called renormalizable.

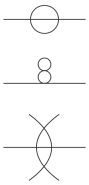
A Feynman diagram with B_{E} external lines is said to have a superficial degree of divergence D if it diverges as K_c^D. Logarithmic divergence $\log K_c$ has a degree of divergence $D = 0$. The superficial degree of divergence measures the badness of the divergence. A theorem states that $D = 4 - B_{\mathrm{E}}$ where 4 is the dimension of spacetime. The Feynman diagram on the right side has above been shown to have logarithmic divergence $D = 0$ in accordance with this theorem.



For any Feynman diagram with $B_{\mathrm{E}} = 2$ the propagator is

$$\int \frac{d^4K}{(2\pi)^4} \frac{i}{K^2 - m^2 + i\varepsilon} \sim K_c^2$$

and $D = 2$. For diagrams with $B_{\mathrm{E}} = 4$ the dependence is $\sim \log(K_c)$ and $D = 0$. For $B_{\mathrm{E}} = 6$ the dependence is $\sim \frac{1}{K_c^2}$ and $D = -2$. Thus, for processes with six or more particles the predicted values only weakly depend on the cutoff because the larger K_c the smaller the dependence. (Odd values B_{E} are not possible.) Note that the degree of divergence is in all cases independent of the internals of the shaded area in the figure on the left side. To summarize, at a very superficial level the only diagrams that are likely to depend on the cutoff in an observable way are the diagrams with two or four external lines.



Further numbers for Feynman diagrams which are important for the proof of this theorem are the number of internal lines B_{I}, the number of vertices V and the number of loops L (equal to the number of undetermined momenta). These numbers for the three examples on the right side are $B_{\mathrm{E}} = 2$, $B_{\mathrm{I}} = 3$, $V = 2$, $L = 2$ and $D = 2$ for the example at the top as well as for the example in the middle, and $B_{\mathrm{E}} = 6$, $B_{\mathrm{I}} = 5$, $V = 4$, $L = 2$ and $D = -2$ for the example at the bottom.

The proof for the theorem stating $D = 4 - B_{\mathrm{E}}$ is only outlined. A first observation is that the number L of loops corresponds directly to the number of $\int d^4K/(2\pi)^4$ integrals to be evaluated. These integrals correspond to the number of undetermined momenta and they are also responsible for the divergences. Without loop integrals there would not be any infinities and these integrals are like multiplying by K_c^4. It looks like there are B_{I} such integrals but momentum conservation reduces the total number of loop integrals such that $L = B_{\mathrm{I}} - (V - 1)$. Each vertex has four lines, each internal line connects two vertices and each external line has one vertex such that $4V = B_{\mathrm{E}} + 2B_{\mathrm{I}}$. For each loop there is a $\int d^4K/(2\pi)^4 \sim K_c^4$ and for each line there is a propagator $i/(K^2 - m^2 + i\varepsilon) \sim K_c^{-2}$ such that $D = 4L - 2B_{\mathrm{I}}$. Bringing all together gives

$$D = 4L - 2B_{\mathrm{I}} = 4\big(B_{\mathrm{I}} - (V - 1)\big) - 2B_{\mathrm{I}} = 2B_{\mathrm{I}} - 4V + 4 = 2B_{\mathrm{I}} - (B_{\mathrm{E}} + 2B_{\mathrm{I}}) + 4 = 4 - B_{\mathrm{E}}$$

as stated by the theorem.

The word divergence is not used here meaning that something goes to infinity but means that an observable quantity seems to depend on the cutoff. In other words, a diagram is divergent if it leads to an observable quantity depending on the cutoff in an observable way. The infinities only come when the cutoff is made infinite.

If one looks at the Lagrangian of the ϕ^4-theory

$$\mathcal{L}(z_1, z_2, z_3) = \frac{1}{2}\left(z^2(\partial_\mu\phi)^2 - z^2\, m^2\, \phi^2\right) - \frac{z^4\, \lambda}{4!}\phi^4$$

then it depends on three parameters $z_1 = m$, $z_2 = \lambda$ and $z_3 = z$ where z is the field strength renormalization. Splitting perturbation theory into a physical and a renormalized part gives

$$\mathcal{L} = \mathcal{L}_{\text{phys}} + \text{counterterms} = \frac{1}{2}\left((\partial_\mu \phi)^2 - m^2_{\text{phys}}\,\phi^2\right) - \frac{\lambda_{\text{phys}}}{4!}\phi^4 + A(\partial_\mu \phi)^2 + B\phi^2 + C\phi^4$$

for the Lagrangian where A, B, C are just parameters one can shift to eliminate dependencies on K_c and have no operational meaning. One can interpret this Lagrangian as a free non-interacting part $\mathcal{L}_{\text{phys}}$ and some interaction terms. The counterterms A, B, C are small because one knows that putting in the physical mass and coupling constant delivers more or less the correct result matching low-energy experiments. One should think of them as additional interactions and they are determined interactively. Their purpose is to ensure that physically observable quantities do not depend on K_c.

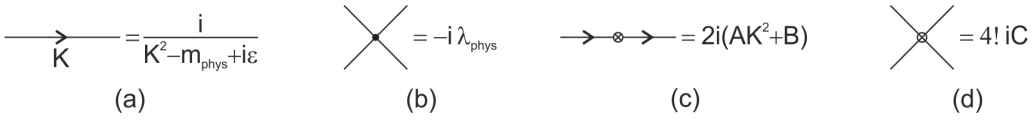


Figure 4: Feynman rules for the ϕ^4-theory with the counterterms

The Feynman rules shown in figure 4 (a) and (b) are the same as for the ϕ^4-theory but the rules in (c) and (d) with a new kind of vertex due to the counterterms are added. One can determine A, B, C iteratively but this is very hard combinatorial work. Suppose A, B, C have been determined at order λ^N_{phys} as A_N, B_N, C_N then one determines A_{N+1}, B_{N+1}, C_{N+1} requiring on one hand the propagator $\longrightarrow\!\!\otimes\!\!\longrightarrow$ in (c) to have a pole exactly at m_{phys} with residue 1 and requiring on the other hand the scattering amplitude in (d) to be $-i\,\lambda_{\text{phys}}$ at $O(\lambda^{N+1}_{\text{phys}})$. All the other diagrams do not depend on the cutoff. This determines the parameters A, B, C and shows that the ϕ^4-theory is renormalizable. This is just a very rough outline of the proof of the BPHZ theorem.

The characteristic of a non-renormalizable theory is that one cannot shift a finite number of parameters and leave all the predictions independent of the cutoff. It therefore would require an infinite number of parameters to ensure that operationally well-defined quantities do not depend on K_c.

Non-renormalizable theories appear in renormalized perturbation theory in the following way. Starting with

$$\mathcal{L} = \mathcal{L}^0_{\text{phys}} + \mathcal{L}^{\text{int}}_{\text{phys}}(\lambda) + \text{counterterms}$$

where the counterterms are a finite list one calculates to order $O(\lambda^N_{\text{phys}})$ and needs counterterms to eliminate dependences on K_c. For each N one can always find a finite number of counterterms but when going to higher orders one needs more and more counterterms. That is the signature of a non-renormalizable theory. Counterterms correspond to parameters one has to fix by experiments. Thus, if one needs an infinite number of parameters to fit by experiment before one can start making predictions then this theory is not useful as a theory of everything. There are several non-renormalizable theories and the most canonical one is gravity. If one writes down linearized gravity – that is a wave equation – and adds an interaction term coming from perturbative corrections then this theory is not renormalizable in this sense. If one does not take linearized gravity but full gravity it may be renormalizable but this is very hard to determine.

5 Classical Gauge Theories

5.1 Abelian Gauge Symmetry

The electromagnetic field as well as the SU(2) and SU(3) gauge bosons are examples of gauge fields. The steps for the development of gauge theories are the same as for other theories. One looks for symmetries specified as a group $\mathcal{G}$ and for projective unitary representations of $\mathcal{G}$ which are local. This extra constraint of locality is imposed on the symmetries because interactions are supposed to be local. The last step is quantization. The difference to other theories is that gauge theories are symmetric under a local gauge

group $\mathcal{G}$ in addition to the other symmetries such as the symmetry under the Poincaré group assumed for all relativistic field theories. The gauge group acts independently at all locations in spacetime.

The group $\mathrm{U}(1) = \{e^{i\,\theta} \,|\, \theta \in [0, 2\pi)\}$ as an example gives the gauge group $\mathcal{G} = \{g : \mathbb{M}_{1,3} \to \mathrm{U}(1)\}$. Group $\mathcal{G}$ is, in words, the set of all assignments of elements of U(1) to points in spacetime, and an element $g \in \mathcal{G}$ assigns an element of U(1) to every point in spacetime. An element $g \in \mathcal{G}$ acts on a Dirac field $\psi(X)$, for example, as

$$\psi(X) \overset{g\in\mathcal{G}}{\to} \pi(g(X))\psi(X) = e^{i\,\alpha(X)}\,\psi(X)$$

with $\alpha(X) \in [0, 2\pi)$ for each $X \in \mathbb{M}_{1,3}$ where π is a representation of the gauge group (and independently the number $\pi \approx 3.14$). In the case of U(1) the field is multiplied with a phase. All the local gauge groups presented in the following have some continuous Lie group in the place of U(1) as in this example.

Because the gauge group acts independently for all X the question arises which theories are invariant under the Poincaré group and the gauge group. To find such classical theories one starts with a Lagrangian density $\mathcal{L}$ and determines the terms that are invariant under the gauge symmetry. The terms $\bar{\psi}\psi$ and $(\bar{\psi}\gamma^\mu\psi)^2$ are invariant under U(1) but $\bar{\psi}\,\partial\!\!\!/\,\psi$ is not. The term $\bar{\psi}\,\partial\!\!\!/\,\psi$ is not even well-defined because

$$\partial_\mu\psi = \lim_{\varepsilon\to 0}\frac{\psi(X+\varepsilon n^\mu) - \psi(X)}{\varepsilon} \qquad \overset{g\in\mathcal{G}}{\to} \qquad g\,(\partial_\mu\psi) = \lim_{\varepsilon\to 0}\frac{e^{i\,\alpha(X+\varepsilon n^\mu)}\,\psi(X+\varepsilon n^\mu) - e^{i\,\alpha(X)}\,\psi(X)}{\varepsilon}$$

contains $\alpha(X+\varepsilon n^\mu)$ and $\alpha(X)$ and they are independent such that the limes is not defined and does not exist. This is serious because Lagrangian densities with only terms like $\bar{\psi}\psi$ and $(\bar{\psi}\gamma^\mu\psi)^2$ are not very interesting and nothing ever happens in such theories. One single field transforming this way under the gauge group cannot be made dynamic. The only way out is adding more fields.

The additional fields also transform under $\mathcal{G}$. (They play a similar role as catalysts in chemistry because sometimes there is no thermodynamical path for a reaction and a catalyst helps to initiate the reaction but is not consumed by it.) One way in physics to satisfy a constraint is adding auxiliary objects and this can also be done to impose the gauge symmetry.

The problem with the derivative is that one tries to compare the field at different locations in spacetime but the field is allowed to transform independently. The direct way of comparing does not work. However, there is another possibility to compare $\psi(X)$ and $\psi(Y)$ by introducing an object called parallel transporter $U(Y, X) \in \mathrm{U}(1)$ for all $X, Y \in \mathbb{M}_{1,3}$. (The parallel transporter $U_\gamma(Y, X)$ depends on the path γ chosen between X and Y but this fact is ignored for the moment.) In order to compare $\psi(X)$ and $\psi(Y)$ in a way invariant under the gauge symmetry one demands

$$U(Y, X) \overset{g\in\mathcal{G}}{\to} e^{i\,\alpha(Y)}\,U(Y, X)\,e^{-i\,\alpha(X)} \tag{5.1}$$

for every $g \in \mathcal{G}$.

One defines that $\psi(X)$ at spacetime location X is parallel transported to spacetime location Y to be

$$U(Y, X)\,\psi(X)$$

such that this object transforms as

$$U(Y, X)\,\psi(X) \overset{g\in\mathcal{G}}{\to} e^{i\,\alpha(Y)}\,U(Y, X)\,e^{-i\,\alpha(X)}\,e^{i\,\alpha(X)}\,\psi(X) = e^{i\,\alpha(Y)}\,U(Y, X)\,\psi(X)$$

under a gauge transformation. Thus, the object $U(Y, X)\,\psi(X)$ transforms exactly as an object at spacetime location Y but the question is whether such an object exists. (Mathematically, this parallel transporter belongs to the theory of fiber bundles or principal bundles.)

5.2 Abelian Covariant Derivative

The object $U(Y, X)$ cannot be a field because it needs two spacetime locations. It is also slightly non-local but one can specify this parallel transporter using local data to allow building local theories which are invariant under this gauge group. To achieve this a different derivative is needed.

The covariant derivative is defined as

$$D_\mu\psi(X) = \lim_{\varepsilon\to 0} \frac{\psi(X+\varepsilon\, n^\mu) - U(X+\varepsilon\, n^\mu, X)\psi(X)}{\varepsilon} \tag{5.2}$$

and this is a sound definition because $\psi(X+\varepsilon\, n^\mu)$ and $U(X+\varepsilon\, n^\mu, X)\psi(X)$ transform the same way under any $g \in \mathcal{G}$. This is like a derivative but it is not the derivative because the definition uses additional data. However, to give a theory dynamics it is not needed to use the derivative but just something acting like a derivative by comparing objects at neighboring locations.

Assuming that $U(X+\varepsilon\, n^\mu, X)$ is differentiable (near X) then the Taylor expansion gives

$$\begin{aligned} U(X+\varepsilon\, n^\mu, X) = U(X,X) + \varepsilon\, n^\mu\, \partial_\mu U(X,X) + ... &= 1 + \varepsilon\, n^\mu \big(-i\, e\, A_\mu(X)\big) + ... \\ &= 1 - i\, e\, \varepsilon\, n^\mu\, A_\mu(X) + ... \end{aligned}$$

where $U(X,X) = 1$ transforms in the right way and $\partial_\mu U(X,X) = -i\, e\, A_\mu(X)$ is just a definition. Note that the gauge theory developed so far is electrodynamics where e is the fine structure constant and A_μ is traditionally defined with the factor $-i\, e$.

The defined quantity $A_\mu(X)$ cannot be arbitrary because $U(Y,X)$ behaves in a certain way under the gauge transformation. Note that $A_\mu(X)$ looks local despite the non-locality of $U(Y,X)$. The constraint on $A_\mu(X)$ is

$$e^{i\,\alpha(X+\varepsilon\, n^\mu)}\, U(X+\varepsilon\, n^\mu, X)\, e^{-i\,\alpha(X)} = 1 - i\, e\, \varepsilon\, n^\mu\, A_\mu(X) + i\, n^\mu\, \partial_\mu\alpha(X) + ...$$

and the auxiliary gauge field $A_\mu(X)$ as a helper field must transform as

$$A_\mu(X) \overset{g\in\mathcal{G}}{\to} A_\mu(X) - \frac{1}{e}\, \partial_\mu\alpha(X) \tag{5.3}$$

under the gauge transformation. This is an infinitesimal way to define the parallel transport in (5.1) but it requires that $\alpha(X)$ is differentiable and can no longer be completely arbitrary. Note that the local gauge field $A_\mu(X)$ is slightly non-local because of $\frac{1}{e}\, \partial_\mu\alpha(X)$.

Inserting (5.3) into the definition of the covariant derivative (5.2) gives

$$D_\mu\psi(X) = \partial_\mu\psi(X) + i\, e\, A_\mu(X)\, \psi(X) \tag{5.4}$$

as can be shown by substituting the transformation law for $A_\mu(X)$ into the above Taylor expansion. The covariant derivative as defined in (5.4) is a completely local object because all parts only depend on X. Further, the covariant derivative transforms correctly

$$D_\mu\psi(X) \overset{g\in\mathcal{G}}{\to} e^{i\,\alpha(X)} D_\mu\psi(X)$$

under the gauge transformation.

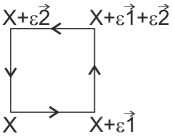
5.3 Dynamics of the Abelian Gauge Field

The task is now to find a field A_μ that transforms according to (5.3) because this transformation law is just a definition but does not ensure the existence of such a field. One can go ahead and build Lagrangians $\mathcal{L} = \bar\psi \not{D}\psi - m\bar\psi\psi + ...$ by replacing $\not\partial \to \not{D}$. If one wants to explain everything in the universe one has also to explain what $A_\mu(X)$ is. It is a vector field and should have some reality. In order to quantize it later one has to give it some dynamics because up to now it is just a static helper field.

To give $A_\mu(X)$ some dynamics one sees that in

$$U(X+\varepsilon n, X) \approx \exp\Big(-i\, e\, \varepsilon n^\mu A_\mu(X + \frac{\varepsilon}{2} n) + O(\varepsilon^3)\Big)$$

both sides give the same Taylor series. One can use the parallel transporter to build a plaquette operator $U_\square$ which is gauge invariant. On the round trip in the figure on the right side one uses the parallel transporter to move around in steps of ε to the right



in direction $\vec{1}$, up in direction $\vec{2}$, to the left and down again. The starting and the ending point must transform the same way because they are the same point but it is possible that the result of this round trip is not one because there is this field A_μ that can vary in different ways along this round trip path. It actually depends on the curvature of A_μ but it does not depend on the gauge transformation. Multiplying these four parallel transporters gives

$$\begin{aligned} U_\square &= U(X, X+\varepsilon\vec{2})U(X+\varepsilon\vec{2}, X+\varepsilon(\vec{1}+\vec{2}))U(X+\varepsilon(\vec{1}+\vec{2}), X+\varepsilon\vec{1})U(X+\varepsilon\vec{1}, X) \\ &= \exp\left\{-i\,\varepsilon\,e\left[-A_2(X+\frac{\varepsilon}{2}\vec{2}) - A_1(X+\frac{\varepsilon}{2}\vec{1}+\varepsilon\vec{2}) + A_2(X+\varepsilon\vec{1}+\frac{\varepsilon}{2}\vec{2}) + A_1(X+\frac{\varepsilon}{2}\vec{1})\right] + O(\varepsilon^3)\right\} \\ &= 1 - i\,\varepsilon^2\,e\,[\partial_1 A_2 - \partial_2 A_1] + O(\varepsilon^3) \end{aligned}$$

after expansion in powers of ε where terms with ε cancel and only terms with ε^2 remain. This transforms trivially under the gauge transformations and therefore $\partial_1 A_2 - \partial_2 A_1$ transforms trivially under $g \in \mathcal{G}$ as well. The selection of the two direction $\vec{1}$ and $\vec{2}$ was arbitrary.

Thus, one can define the curvature tensor $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$ which is locally gauge invariant under $\mathcal{G}$. The fact that this object is gauge invariant and has some derivatives means that A has some dynamics. This is a 2-tensor and is invariant under Lorentz transformations but it is not invariant under Poincaré transformations because it has spacetime indices. The quantity $F^{\mu\nu}F_{\mu\nu}$ is a Lorentz scalar and is invariant under Poincaré and gauge transformations. It can be put into a Lagrangian where it appears traditionally as $-\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$. One can therefore build a Lagrangian density

$$\mathcal{L} = \bar{\psi}(\not{D} - m)\psi - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} \tag{5.5}$$

and this is the Lagrangian for quantum electrodynamics. As a classical theory this is the electromagnetic field minimally coupled to the Dirac field. One could use other covariant derivatives but the resulting theories would not be renormalizable.

The tensor $F^{\mu\nu}$ can be derived via an alternative and shorter path but the derivation using the plaquette operator will later be useful when simulating quantum field theories on a lattice. Because D_μ is gauge invariant also the commutator $[D_\mu, D_\nu]\psi \to e^{i\,\alpha(X)}[D_\mu, D_\nu]\psi$ is gauge invariant but unlike the ordinary derivative $[\partial_\mu, \partial_\nu] = 0$ the commutator $[D_\mu, D_\nu]$ is not zero. It is

$$\begin{aligned} [D_\mu, D_\nu]\psi &= [\partial_\mu, \partial_\nu]\psi + i\,e\,([\partial_\mu, A_\nu] - [\partial_\nu, A_\mu])\,\psi - e^2[A_\mu, A_\nu]\psi \\ &= i\,e\,([\partial_\mu, A_\nu] - [\partial_\nu, A_\mu])\,\psi = i\,e\,(\partial_\mu A_\nu - \partial_\nu A_\mu)\psi \end{aligned}$$

because A_μ and A_ν commute as numbers. This means that the commutator of the covariant derivative is the curvature tensor $[D_\mu, D_\nu] = i\,e\,F_{\mu\nu}$.

To summarize, a theory has been built which is invariant under the Poincaré group as a global symmetry and the gauge as a local symmetry. To allow dynamics for the Dirac field $\psi(X)$ additional degrees of freedom have been added in form of the helper field $A_\mu(X)$ giving a different representation of the local gauge group. This leads to a new kind of derivative D allowing dynamic terms for $\psi(X)$ in the Lagrangian density $\mathcal{L}$ such as $\bar{\psi}(\not{D})\psi$. The quantity $F^{\mu\nu}F_{\mu\nu}$ in $\mathcal{L}$ describes the dynamics of the helper field $A_\mu(X)$. Because of $e^{i\,\alpha(X)} \cdot e^{i\,\alpha(Y)} = e^{i\,\alpha(Y)} \cdot e^{i\,\alpha(X)}$ the result is an abelian gauge theory. As mentioned above it corresponds to classical electrodynamics.

5.4 Non-Abelian Gauge Symmetry

The non-abelian gauge theory of the SU(2) gauge group is used here as an example for what is called a Yang-Mills theory when quantized. Thus, the theory is supposed to be invariant under transformations $V(X) \in \mathrm{SU}(2)$ where $X \in \mathbb{M}_{1,3}$ and $V(X)$ is a 2×2 matrix of the form and with the constraints

$$V(X) = \begin{pmatrix} v_{00}(X) & v_{01}(X) \\ v_{10}(X) & v_{11}(X) \end{pmatrix} \qquad \sum_{j,k=0}^{1} |v_{jk}|^2 = 2 \qquad V^\dagger V = \mathbb{I} \qquad \det(V) = 1$$

in the chosen representation. One has to make a choice for how these matrices which are defined independently for each spacetime location act on the field content. One introduces two independent spinor

fields $\psi_0(X)$ and $\psi_1(X)$ and defines

$$\psi_j(X) = \sum_{k=0}^{1} v_{jk}(X)\,\psi_k(X) \tag{5.6}$$

as the way the matrices $V(X)$ act. (However, there are many other ways this group can act because one could have chosen boson fields and so on.) The non-abelian gauge theory to be found is supposed to be invariant under (5.6).

One can think of the two fermion fields as a single doublet field $\Psi(X)$ in form of a 8×1 vector transforming under Poincaré transformations like two spinors. The local gauge group $\mathcal{G}$ acts as

$$\Psi(X) = \begin{pmatrix}\big(\psi_0(X)\big)\\ \big(\psi_1(X)\big)\end{pmatrix} \overset{V\in\mathcal{G}}{\to} \begin{pmatrix} v_{00}(X)\cdot\mathbb{I}_{4\times4} & v_{01}(X)\cdot\mathbb{I}_{4\times4}\\ v_{10}(X)\cdot\mathbb{I}_{4\times4} & v_{11}(X)\cdot\mathbb{I}_{4\times4}\end{pmatrix}\begin{pmatrix}\big(\psi_0(X)\big)\\ \big(\psi_1(X)\big)\end{pmatrix}$$

on this doublet. The invariant terms are

$$\bar\Psi\Psi = \sum_{j=0}^{1}\bar\psi\psi \qquad\qquad (\bar\Psi\Gamma^\mu\Psi)^2 \text{ where } \Gamma^\mu = \begin{pmatrix}\gamma^\mu & 0\\ 0 & \gamma^\mu\end{pmatrix}$$

and none of them contains derivatives for the same reason as for the abelian theory.

5.5 Non-Abelian Covariant Derivative

Thus, one defines a covariant derivative by introducing a parallel transporter $U(Y,X)\in\mathrm{SU}(2)$ as

$$U(Y,X) \overset{V\in\mathcal{G}}{\to} V(Y)\,U(Y,X)\,V^\dagger(X) \tag{5.7}$$

where $U(Y,X)$ is not a number as in the abelian case but a matrix. The covariant derivative is defined similarly to the abelian case as

$$n^\mu\,D_\mu\Psi = \lim_{\varepsilon\to0}\frac{1}{\varepsilon}\Big[\Psi(X+\varepsilon\,n) - U(X+\varepsilon\,n,X)\Psi(X)\Big] \tag{5.8}$$

where n is a unit 4-vector. One does not need to know the parallel transporter for any two locations but only for infinitesimally close neighbors and that is local information. As in the abelian case, one should further assume that $U(Y,X)$ is differentiable.

Any unitary 2×2 matrix U as an element of the fundamental representation of SU(2) can be written as

$$U = e^{i\,\underline{\underline{A}}} \qquad\qquad \underline{\underline{A}}^\dagger = \underline{\underline{A}} \qquad\qquad \mathrm{tr}\big(\underline{\underline{A}}\big) = 0$$

where $\underline{\underline{A}}$ is a 2×2 Hermitian matrix with trace zero. One could continue in a coordinate-free way without choosing a basis but as soon as one tries to perform computations on a computer one needs to use a basis. A basis for Hermitian and traceless 2×2 matrices are the Pauli matrices. Thus, the matrix $\underline{\underline{A}}(X)$ can be written as

$$\underline{\underline{A}} = \sum_{j=1}^{3} A^j\frac{\sigma^j}{2} \qquad \sigma^1 = \frac{1}{2}\begin{pmatrix}0&1\\1&0\end{pmatrix} \qquad \sigma^2 = \frac{1}{2}\begin{pmatrix}0&-i\\i&0\end{pmatrix} \qquad \sigma^3 = \frac{1}{2}\begin{pmatrix}1&0\\0&-1\end{pmatrix}$$

built from a commonly used set of Pauli matrices obeying the commutation relations $[\sigma^j,\sigma^k] = i\,\varepsilon^{jk}_{\ \ l}\,\sigma^l$. Thus, to specify $\underline{\underline{A}}$ one needs three real numbers A^j.

To define the covariant derivative only the parallel transporter $U(X+\varepsilon\,n,X)$ is needed. It can be written as

$$U(X+\varepsilon\,n,X) = e^{i\,\underline{\underline{B}}(X,n,\varepsilon)} \approx \mathbb{I} + i\,g\,\varepsilon\,n^\mu\sum_{j=1}^{3}A^j_\mu(X)\frac{\sigma^j}{2} + O(\varepsilon^2)$$

where $\underline{\underline{B}}$ has the given Taylor expansion to first order. (There is a summation over μ according to the summation convention.) The quantity g is a constant introduced for later use. The covariant derivative defined in (5.8) becomes

$$D_\mu = \partial_\mu - i\,g\sum_{j=1}^{3} A_\mu^j(X)\frac{\sigma^j}{2} \tag{5.9}$$

similarly to (5.4) for the abelian case but with a 8×8 matrix $\sum_{j=1}^{3} A_\mu^j(X)\frac{\sigma^j}{2}$ acting on two 4×1 spinors instead of a scalar A_μ. The coefficients $A_\mu^j(X)$ are not arbitrary numbers but have to give a representation of the local gauge group and that is determined by

$$\begin{aligned} U(X+\varepsilon\,n,X) &\overset{V\in\mathcal{G}}{\to} V(X+\varepsilon\,n)\,U(X+\varepsilon\,n,X)\,V^\dagger(X) \\ &\approx V(X+\varepsilon\,n)\,\mathbb{I} + i\,g\,\varepsilon\,n^\mu\sum_{j=1}^{3} A_\mu^j(X)\frac{\sigma^j}{2}\,V^\dagger(X) \end{aligned}$$

approximated to order $O(\varepsilon)$. Therefore, the coefficients $A_\mu^j(X)$ must transform as

$$\sum_{j=1}^{3} A_\mu^j(X)\frac{\sigma^j}{2} \overset{V\in\mathcal{G}}{\to} V(X)\left(\sum_{j=1}^{3} A_\mu^j(X)\frac{\sigma^j}{2} + \frac{i}{2}\partial_\mu\right)V^\dagger(X)$$

under the local gauge group as can be shown using $V(X+\varepsilon\,n)\,V^\dagger(X) = [(1+\varepsilon\,n^\mu\,\partial_\mu)V(X)]V^\dagger(X)+O(\varepsilon^2)$.

To compute $V(X)\,\partial_\mu V^\dagger(X)$ in the infinitesimal case with $V(X) = e^{i\sum\alpha^j(X)\frac{\sigma^j}{2}}$ for small numbers $\alpha^j(X)$ one gets

$$V(X)\,\partial_\mu V^\dagger(X) \approx \left(\mathbb{I} + i\sum_{j=1}^{3}\alpha^j(X)\frac{\sigma^j}{2}\right)\,\partial_\mu\left(\mathbb{I} - i\sum_{j=1}^{3}\alpha^j(X)\frac{\sigma^j}{2}\right) = -i\sum_{j=1}^{3}\frac{\partial_\mu\alpha^j(X)}{\partial X^\mu}\frac{\sigma^j}{2} + O(\alpha^2)$$

such that

$$\sum_{j=1}^{3} A_\mu^j(X)\frac{\sigma^j}{2} \overset{V\in\mathcal{G}}{\to} \sum_{j=1}^{3} A_\mu^j(X)\frac{\sigma^j}{2} + \frac{1}{g}\sum_{j=1}^{3}\left(\partial_\mu\alpha^j(X)\right)\frac{\sigma^j}{2} + i\sum_{k,l=1}^{3}\left[\alpha^k(X)\frac{\sigma^k}{2}, A_\mu^l(X)\frac{\sigma^l}{2}\right] \tag{5.10}$$

where the commutator term appears because the Pauli matrices do not commute.

One can see now what the infinitesimal local gauge transformations do to the covariant derivative $D_\mu\Psi$. The transformation law for $\Psi(X)$ is

$$\Psi(X) \overset{V\in\mathcal{G}}{\to} \left(\mathbb{I} + i\sum_{j=1}^{3}\alpha^j(X)\frac{\sigma^j}{2}\right)\Psi(X)$$

and the covariant derivative of the doublet spinor field becomes

$$D_\mu\Psi \overset{V\in\mathcal{G}}{\to} \left(\partial_\mu - i\,g\sum_{j=1}^{3} A_\mu^j\frac{\sigma^j}{2} - i\sum_{j=1}^{3}(\partial_\mu\alpha^j)\frac{\sigma^j}{2} + g\sum_{j,k=1}^{3}\left[\alpha^k\frac{\sigma^k}{2}, A_\mu^j\frac{\sigma^j}{2}\right]\right)\left(\mathbb{I} + i\sum_{j=1}^{3}\alpha^j\frac{\sigma^j}{2}\right)\Psi$$

to order $O(\alpha)$ by inserting (5.10). Thus, the transformation law is

$$D_\mu\Psi(X) \overset{V\in\mathcal{G}}{\to} \left(\mathbb{I} + i\sum_{j=1}^{3}\alpha^j(X)\frac{\sigma^j}{2}\right)D_\mu\Psi(X) = V(X)\,D_\mu\Psi(X)$$

at least for infinitesimal gauge transformations. For a big gauge transformation one gets the right side by exponentiating the left side using $(1+\frac{x}{n})^n = e^x$ for large n.

5.6 Dynamics of the Non-Abelian Gauge Field

Also in the non-abelian case one has to add dynamics to the gauge field $A^j_\mu(X)$. The commutator $[D_\mu, D_\nu]$ of the covariant derivative

$$[D_\mu, D_\nu]\Psi(X) \overset{V\in\mathcal{G}}{\to} V(X)\,[D_\mu, D_\nu]\Psi(X)$$

is an invariant term involving derivatives of $A^j_\mu(X)$ and is the curvature of the SU(2) fiber bundle. It is

$$[D_\mu, D_\nu] = i\,g\sum_{j=1}^{3} F^j_{\mu\nu}\frac{\sigma^j}{2}$$

with

$$\sum_{j=1}^{3} F^j_{\mu\nu}\frac{\sigma^j}{2} = \sum_{j=1}^{3}\left(\partial_\mu A^j_\nu\frac{\sigma^j}{2} - \partial_\nu A^j_\mu\frac{\sigma^j}{2}\right) - i\,g\sum_{j,k=1}^{3}\left[A^k_\mu\frac{\sigma^k}{2}, A^j_\mu\frac{\sigma^j}{2}\right]$$

where the first term which is linear in the A^j_μ looks similar to the abelian case but the second term which is quadratic in the A^j_μ is new.

The curvature tensor transforms as

$$\sum_{j=1}^{3} F^j_{\mu\nu}\frac{\sigma^j}{2} \overset{V\in\mathcal{G}}{\to} V(X)\left(\sum_{j=1}^{3} F^j_{\mu\nu}\frac{\sigma^j}{2}\right)V^\dagger(X)$$

such that one can build a Lorentz scalar

$$\mathrm{tr}\left(\left(\sum_{j=1}^{3} F^j_{\mu\nu}\frac{\sigma^j}{2}\right)\left(\sum_{k=1}^{3} F^k_{\mu\nu}\frac{\sigma^k}{2}\right)\right) = \frac{1}{8}\sum_{j=1}^{3} F^j_{\mu\nu}\,F^{\mu\nu\,j}$$

invariant under the gauge group and use it to build a Lagrangian density

$$\mathcal{L} = \bar\Psi(i\,\not{D} - m)\Psi - \frac{1}{4}\sum_{j=1}^{3} F^j_{\mu\nu}\,F^{\mu\nu\,j} \tag{5.11}$$

which is invariant under global Poincaré and local gauge transformations.

In electrodynamics as the abelian case the term $F^2_{\mu\nu} = (\partial_\mu A_\nu - \partial_\nu A_\mu)$ is a wave function corresponding to Maxwell's equations and describes the gauge bosons alone without interactions with fermions. The gauge field appears only quadratic and this means that the gauge bosons in the abelian case do not interact with each other. The Lagrangian density in the non-abelian case on the other hand has cubic and quartic terms of the gauge field due to the commutator term in $\sum F^j_{\mu\nu}\,F^{\mu\nu\,j}$ even in the absence of fermions. This makes already the classical non-abelian theory extraordinarily non-trivial and leads to interactions between the gauge bosons. For $g = 0$ there are also wave functions as in the abelian case but with $g \neq 0$ the theory gets highly non-linear equations of motion. This will cause many ambiguities when one tries to find a quantum system modeling this classical limit.

6 Quantization of Gauge Theories

6.1 General Issues Concerning Quantization of Gauge Theories

The goal is to find a quantum system invariant under the Poincaré group and the local gauge group $\mathcal{G}$ which has as its effective classical limit the theory with the Lagrangian density (5.5) for the abelian gauge group U(1) or (5.11) for the non-abelien group SU(2). Invariant under a group means finding a representation. There are two problems specifically for the quantization of gauge theories:

1. The classical theory is non-linear. For gauge theories it is hard to find a quadratic system $\mathcal{L}_0$ as the starting point for perturbations because interactions are already built into these theories.
2. There is a lot of symmetry involved because there is the global symmetry from the Poincaré group and there is also one symmetry per spacetime location from the local gauge symmetry. This leads to extra divergences when quantized.

Apart from canonical quantization there are two approaches for quantization. The first one is the path integral quantization $\int \mathcal{D}A\,\mathcal{D}\psi\,\mathcal{D}\bar{\psi}\,e^{iS}$. It is good for high energy scattering but not very good for low energy processes such as ground state correlation functions because of the non-linearity of the theory. The second one puts the theory on a lattice and is much closer to working with a cutoff. The advantage is that non-linearity becomes easy but the disadvantage is that the Poincaré invariance gets lost. The two approaches have not been proven to be equivalent. The path integral approach is not yet mathematically rigorous while the lattice approach is mathematically rigorous but getting back the Poincaré invariance when the lattice spacing goes to zero is hard work and has not yet been successfully done. Both approaches are shown in the following but only the non-abelien case SU(2) is quantized because the abelian case U(1) can easily be deduced from it.

6.2 Gauge Fixing

The two gauge configurations $A^a_\mu \frac{\sigma^a}{2}$ and $A'^a_\mu \frac{\sigma^a}{2} = A^a_\mu \frac{\sigma^a}{2} + \frac{1}{g}(\partial_\mu \alpha^a)\frac{\sigma^a}{2} + i[\alpha^b \frac{\sigma^b}{2}, \alpha^c \frac{\sigma^c}{2}]$ (with implicit summations over a, b, c) where one is a gauge transformation away from the other are gauge equivalent. Thus, if one integrates over all gauge configurations $A^a_\mu \frac{\sigma^a}{2}$ for the path integral $\int \mathcal{D}A$ then there is an infinite number of different configurations in terms of numbers but the action S is invariant such that one sums up the same e^{iS} an infinite number of times. That is how gauge theories lead to extra divergences.

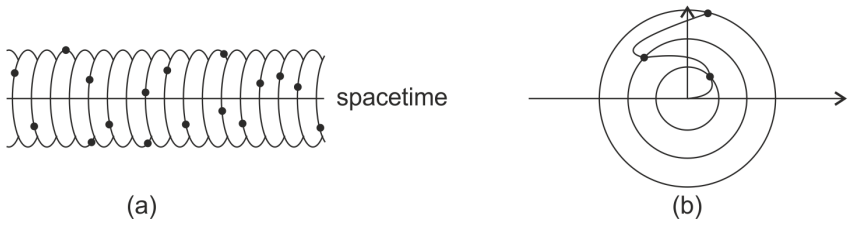
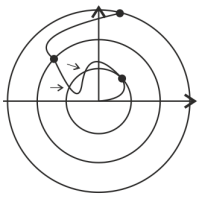


Figure 5: Schematic pictures of the space of all A^a_μ

The situation is illustrated in figure 5 schematically although it is difficult to show a sphere $\mathbb{S}^3$ for SU(2) attached to each point in spacetime $\mathbb{M}_{1,3}$. In (a) spacetime is reduced to discretized $\mathbb{M}_{0,1}$ and SU(2) has become a circle and therefore visualizes more the abelian case U(1). At every point in spacetime SU(2) acts independently and multiplies by a phase. This twists the field around at every point. One should think of A^a_μ (three numbers) for every spacetime location as a choice of a vector on a sphere. There are equivalence classes $[A^a_\mu]$ of A^a_μ defined by $\mathcal{G}$. Thus, these extra divergences come from integrating over all elements of an equivalence class instead of integrating over one element selected per equivalence class. Because visualizing this situation even in $\mathbb{M}_{0,1}$ is hard, another illustration shown in (b) is commonly used here. In analogy with rotations SO(2) in two dimensions a single point becomes an equivalence class shown as a circle. If a theory is invariant under SO(2) then every point on a circle is equivalent and one chooses a representative A^a_μ from $[A^a_\mu]$ for the path integral. This selection process is called fixing the gauge and allows to do the path integral on a reduced configuration space.



The gauge fixing condition is some equation $G(A^a_\mu) = 0$. An examples used in the following is $G(A^a_\mu) = \partial^\mu A^a_\mu - \omega^a = 0$. Gauge fixing is known from electrodynamics where the Lorenz or the Coulomb gauge, for example, are used. A good gauge fixing condition should pick out precisely one representative per equivalence class. However, it can happen as in the figure on the right side marked with little arrows that the gauge choice turns back and selects more than one representative of an equivalence class. This is called the Gribov ambiguity and it happens for the Coulomb gauge which works for U(1) but has this problem when used for SU(2). If one can find such a function G one can separate out the overcounting in the path integral such that the path integral is split into two

pieces where one piece integrates over exactly each representatives and the other integral goes over the equivalence classes. Once one knows that one got the integral over the equivalence classes one can pull it out and throw it away. It is not obvious why one can throw away a whole integral but this is still the process of guessing a quantum theory. Thus, if the resulting theory is healthy the weird steps not entirely justified do not matter and the path integral just helps guessing the right theory.

As a common trick one inserts the number 1 written in a clever way which is here represented as

$$1 = \int \mathcal{D}\alpha\, \delta(G(A^\alpha))\, \det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right)$$

with

$$(A^\alpha)^a_\mu = A^a_\mu + \frac{1}{g}\partial_\mu \alpha^a + f^{abc} A^b_\mu \alpha^c = A^a_\mu + \frac{1}{g} D_\mu \alpha^a \tag{6.1}$$

where the integral is over all gauge transformations α and f^{abc} from $[\sigma^a/2, \sigma^b/2] = i\, f^{abc}\sigma^c/2$ are the structure constants. The two δ have different meaning. The one in $\delta(G(A^\alpha))$ is the functional Dirac delta and the one in $\delta G(A^\alpha)/\delta\alpha$ is the functional derivative. This path integral makes more sense if it is written as the limit of many integrals

$$1 = \left(\prod_{j=1}^{n} \int da_j\right) \delta^{(n)}\big(g(\underline{a})\big)\, \det\left(\frac{\partial g_j}{\partial a_k}\right)$$

where $\delta^{(n)}$ is the n-fold Dirac function. The discretized quantity A^α becomes the list $\underline{a}$ with n numbers a_j. Under a change of variable this gets the form

$$1 = \left(\prod_{j=1}^{n} \int db_j\right) \delta^{(n)}\big(\underline{b}\big)$$

where the determinant before the change of variables is the Jacobian. Thus, if one defines the path integral version of this equation as the limit $n \to \infty$ of the discretized equation then the number 1 can indeed be written in this form.

Inserting it into the path integral for the non-abelian theory gives

$$\int \mathcal{D}\alpha \int \mathcal{D}A \int \mathcal{D}\psi \mathcal{D}\bar{\psi}\, \delta(G(A^\alpha))\, \det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right) e^{iS}$$

where the determinant is called Faddeev-Popov determinant. The term $\delta(G(A^\alpha))$ is very promising because it is zero unless the A^α happens to obey the gauge fixing condition and therefore select one element per equivalence class. Since one integrates over all gauge transformations α and then over all A^α the term $\delta(G(A^\alpha))$ selects a representative for each equivalence class such that the integral $\int \mathcal{D}\alpha$ representing the overcounting gets decoupled and can be thrown away. That is the strategy but the price one has to pay for this is the Faddeev-Popov determinant.

The next question is what is $\delta G(A^\alpha)/\delta\alpha$. Choosing the gauge fixing condition $G(A^a_\mu) = \partial^\mu A^a_\mu - \omega^a = 0$ results in

$$\det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right) = \det\left(\frac{1}{g}\partial^\mu D_\mu\right)$$

by evaluating the gauge fixing condition with (6.1). The partial derivative of a covariant derivative is an operator (a matrix with infinite dimensions) and one takes the determinant of it. The next step uses the fact that the path integral over Grassmann numbers gives the determinant. This leads to the trick

$$\det\left(\frac{1}{g}\partial^\mu D_\mu\right) = \int \mathcal{D}c \int \mathcal{D}\bar{c}\, \exp\left(\frac{i}{g}\int d^4X\, \bar{c}(-\partial^\mu D_\mu)c\right)$$

where c and $\bar{c}$ are auxiliary Grassmann-valued fields called ghosts. Note that these auxiliary field are not spinor fields but scalar fields with spin 0 and probably do not correspond to anything physical. These

ghosts will appear everywhere in the following expressions but expressions for observations cannot depend on them and they must disappear.

The path integral becomes

$$\int \mathcal{D}\alpha \int \mathcal{D}A \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\,\delta(G(A^\alpha))\,\det\left(\frac{\delta G(A^\alpha)}{\delta\alpha}\right) e^{iS} =$$
$$\int \mathcal{D}\alpha \int \mathcal{D}A \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\,\delta(\partial^\mu A^a_\mu - \omega^a)\det\left(\frac{1}{g}\partial^\mu D_\mu\right) e^{iS} =$$
$$\int \mathcal{D}\alpha \int \mathcal{D}A \int \mathcal{D}c\mathcal{D}\bar{c} \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\,\delta(\partial^\mu A^a_\mu - \omega^a)\, e^{i\left(S+\frac{1}{g}\int d^4X\,\bar{c}(-\partial^\mu D_\mu)c\right)} =$$
$$\int \mathcal{D}\omega\, e^{-i\int d^4X\frac{\xi(\omega^a)^2}{2}} \int \mathcal{D}\alpha \int \mathcal{D}A \int \mathcal{D}c\mathcal{D}\bar{c} \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\,\delta(\partial^\mu A^a_\mu - \omega^a)\, e^{i\left(S+\frac{1}{g}\int d^4X\,\bar{c}(-\partial^\mu D_\mu)c\right)} =$$
$$N(\xi)\int \mathcal{D}\alpha \int \mathcal{D}A \int \mathcal{D}c\mathcal{D}\bar{c} \int \mathcal{D}\psi\mathcal{D}\bar{\psi}\, e^{i\int d^4X\mathcal{L}}$$

where ω^a is an arbitrary gauge choice just selecting a different element from the equivalence class and has been used to eliminate $\delta(\partial^\mu A^a_\mu - \omega^a)$. The quantity $N(\xi)$ is a number depending on ξ and ξ can be selected such that $\xi \in [0,1]$. With the resulting Lagrangian density for the Yang-Mills theory

$$\mathcal{L} = \bar{\psi}(i\not{D} - m)\psi - \frac{1}{4}(F^a_{\mu\nu})^2 + \frac{1}{2}\,\xi\,(\partial^\mu A^a_\mu)^2 + \frac{1}{g}\,\bar{c}^a(-\partial^\mu D^{ab}_\mu)c^b \tag{6.2}$$

the path integral no longer depends on α. Thus, the path integral becomes $\infty \cdot N(\xi)\int ...$ because of the integral over α. The correlation functions are ratios such as $\int \mathcal{D}...\bar{\psi}(X)\psi(Y)e^{iS}$ over $\int \mathcal{D}...e^{iS}$ and the terms $\infty \cdot N(\xi)$ always cancel. The resulting path integral with this Lagrangian density is just a normal looking path integral.

The remaining big questions are whether this path integral defines a quantum theory invariant under Poincaré and local gauge transformations at all, whether the ghosts vanish from processes, and whether the Lagrangian density (6.2) is renormalizable. If the ghosts do not vanish a Dirac particle could transform into a ghost particle and this event has never been observed. However, these questions can only be touched here but cannot be addressed as they would deserve. The question of renormalizability has been resolved by Gerardus 't Hooft. The question of the ghosts becomes clearer when looking at Feynman diagrams. With respect to the first question it is just assumed that the resulting theory is a quantum theory and that it is invariant under Poincaré and local gauge transformations.

6.3 Path Integral Quantization of Non-Abelian Gauge Theories

The result so far after the gauge has been fixed is the transition amplitude defined as

$$\langle\Phi_\mathrm{f}|U|\Phi_\mathrm{i}\rangle = \int \mathcal{D}A \int \mathcal{D}\psi\mathcal{D}\bar{\psi} \int \mathcal{D}c\mathcal{D}\bar{c}\, e^{iS}$$

where the right side of the equation is still classical and the action is $S = \int d^4X\mathcal{L}$ with the Lagrangian density in (6.2)

$$\mathcal{L} = -\frac{1}{4}\left(\partial_\mu A^a_\nu - \partial_\nu A^a_\mu\right)^2 + \frac{1}{2}\,\xi\,(\partial^\mu A^a_\mu)^2 + \bar{\psi}(i\not{\partial} - m)\psi + \frac{1}{g}\,\bar{c}^a(-\partial^\mu\partial_\mu)c^a + \mathcal{L}_\mathrm{int}(g) = \mathcal{L}_0 + \mathcal{L}_\mathrm{int}(g)$$

where all the dependencies on g including those from the covariant derivatives have been moved into the interaction term. This is therefore a Lagrangian density for a free theory plus an interaction term and one can use perturbation theory assuming g is small (although g is actually not really small). The result of the Taylor expansion is a set of Feynman rules based on the assumption that the theory defined perturbatively by evaluating this path integral is a valid quantum theory.

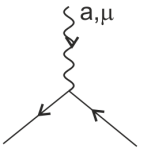
The theory contains fermion fields $\boldsymbol{\psi}_{j\alpha}(X)$ with two indices because of the two spinors with four components in the doublets, three vector bosons $\boldsymbol{A}^a_\mu(X)$ with $a = 1,2,3$ and the ghost fields c^a. One does not have an intuition for what the ghost fields are but knows how to calculate the integrals.

The Feynman rules for the propagators from $g = 0$ and the vertices from $g \neq 0$ are

$$= \langle \boldsymbol{\psi}_{j\alpha}(X)\, \bar{\boldsymbol{\psi}}_{k\beta}(Y)\rangle \quad = \int \frac{d^4K}{(2\pi)^4} \left(\frac{i}{\not{K} - m}\right)_{\alpha\beta} \delta_{jk}\, e^{-iK\cdot(X-Y)}$$

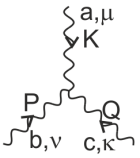
$$= \langle \boldsymbol{A}^a_\mu(X)\, \boldsymbol{A}^b_\nu(Y)\rangle \quad = \int \frac{d^4K}{(2\pi)^4} \left(\eta_{\mu\nu} - \frac{(1-\xi)K_\mu K_\nu}{K^2}\right) \delta^{ab}\, \frac{e^{-iK\cdot(X-Y)}}{K^2 + i\varepsilon}$$

$$= \langle \boldsymbol{c}^a(X)\, \bar{\boldsymbol{c}}^b(Y)\rangle \quad = \int \frac{d^4K}{(2\pi)^4} \frac{i}{K^2}\, \delta^{ab}\, e^{-iK\cdot(X-Y)}$$



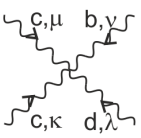
$$= i\, g\, \gamma^\mu \frac{\sigma^a}{2}$$

from the covariant derivative



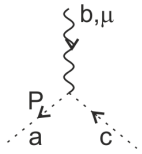
$$= g\, f^{abc}\, [\eta^{\mu\nu}(K-P)^\kappa + \eta^{\nu\kappa}(P-Q) + \eta^{\kappa\mu}(Q-K)^\nu]$$

from the cubic terms



$$\begin{aligned} = -i\, g \Big[& f^{abe} f^{cde} \left(\eta^{\mu\kappa}\eta^{\nu\lambda} - \eta^{\mu\lambda}\eta^{\nu\kappa}\right) \\ & + f^{ace} f^{bde} \left(\eta^{\mu\nu}\eta^{\kappa\lambda} - \eta^{\mu\lambda}\eta^{\nu\kappa}\right) \\ & + f^{ade} f^{bce} \left(\eta^{\mu\nu}\eta^{\kappa\lambda} - \eta^{\mu\kappa}\eta^{\nu\lambda}\right) \Big] \end{aligned}$$

from the quartic terms



$$= g\, f^{abc}\, P^\mu$$

from the ghosts

plus the symmetry factors, the signs and the momentum conservation.

One question is whether one can interpret this theory as a quantum theory and whether it defines $\boldsymbol{A}^a$, $\boldsymbol{\psi}$ and $\boldsymbol{c}$ as quantum fields. It would not define a quantum theory if propagating in time, for example, would not be unitary. One can test that by making sure that the correlation functions are symmetric under the Poincaré group such that probability is conserved. The existence of negative norm states $\langle\Phi|\Phi\rangle < 0$ for some $|\Phi\rangle$ after building the Fock space is another possibility for failure because this is not a Hilbert space. However, it is possible that the inner product is positive for all physical states and there is a subsector of this linear space with negative norm. This is what happens here and the theory can still be an effective quantum theory of the physical states. It can be shown that this theory is indeed a quantum theory and the fields are quantum fields.

6.4 Dimensional Regularization

Another question is whether the theory is renormalizable because some of the Feynman diagrams evaluate to infinity. It is renormalizable as proved by Gerard 't Hooft and Martinus Veltman using dimensional regularization. A cutoff is always admitting ignorance about how the theory behaves at high momenta or other degrees of freedom one cannot observe. Different cutoffs reveal different hypotheses about reality. There is a strong believe that physics is Lorentz invariant and a sharp momentum cutoff breaks this invariance. Cutoffs can also break gauge invariance. Dimensional regularization is a way of rendering the integral in a Feynman diagram expansion finite but in a way such that Lorentz and local gauge invariance is preserved.

The dimensional regularization is introduce via an example that can be generalized. In theories such as the ϕ^4-theory one encounters diverging loop integrals. This is a clear indication that one has made a too

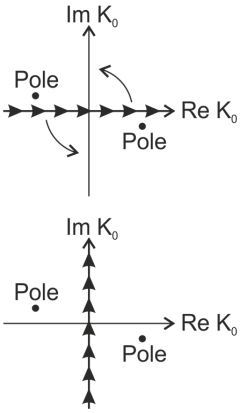
strong hypothesis about how the degrees of freedom of this theory behave. A typical integral of this kind is

$$I_n(m) = \int \frac{d^D K}{(2\pi)^D} \frac{1}{(K^2 - m^2 + i\varepsilon)^n}$$

where D is the dimension and n is an integer basically telling how many undetermined momenta there are. Up to now the dimension has always be $D = 4$. With the cutoff Λ the term $d^D K$ contributes Λ^D and the second term contributes $1/\Lambda^{2n}$ giving together Λ^D/Λ^{2n}. Thus, there are no problems in low dimensions or, in other word, the lower the dimension the less painful these divergences are.

One can separate this integral into an integral over the temporal component K_0 and an integral over the remaining spatial components $\vec{k}$ of momentum K

$$I_n(m) = \int \frac{d^{D-1}\vec{k}}{(2\pi)^D} \left(\int_{-\infty}^{\infty} dK_0 \frac{1}{(K_0^2 - \vec{k}^2 - m^2 + i\varepsilon)^n} \right)$$



with poles at the two places $K_0 = \pm\sqrt{\vec{k}^2 + m^2} + i\varepsilon$. As long as one avoids those poles one should be able to move that contour around to make this integral a bit nicer. Initially one integrates over the $\mathrm{Re}\, K_0$-axis but can use Cauchy's integral theorem and rotate the contour to the $\mathrm{Im}\, K_0$-axis as illustrated in the figure on the right side. The integral is zero when the contour is closed as long as there are no poles enclosed. The quantity $K_0^2 - \vec{k}^2$ looks like K^2 and therefore like the radius of an object in a higher-dimensional sphere except for the minus sign. The reason for moving the contour is $K_0 \to i\, K_\mathrm{E}$ such that the integral becomes

$$I_n(m) = -i \int \frac{d^{D-1}\vec{k}}{(2\pi)^D} \left(\int_{-\infty}^{\infty} \frac{dK_\mathrm{E}}{(K_\mathrm{E}^2 + \omega^2)^n} \right)$$

with $\omega^2 = \vec{k}^2 + m^2$. After reabsorbing K_E into K the result is an integral in Euclidean space with dimension D and can be resolved by going to spherical polar coordinates with $r = |K|$

$$I_n^\mathrm{E}(m) = -i \int_\mathrm{Euclidean} \frac{d^D K}{(2\pi)^D} \frac{1}{(K^2 + m^2)^n} = \frac{-i}{(2\pi)^D} \int d\Omega_D \int_0^\infty dr\, \frac{r^{D-1}}{(r^2 + m^2)^n}$$

where Ω_D is the surface of the sphere. With

$$\int d\Omega_D = \frac{(2\pi)^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)}$$

where Γ is a gamma function the integral becomes

$$I_n^\mathrm{E}(m) = \frac{-i}{(2\pi)^D} \frac{(2\pi)^{\frac{D}{2}}}{\Gamma\left(\frac{D}{2}\right)} \int_0^\infty dr\, \frac{r^{D-1}}{(r^2 + m^2)^n} = \frac{m^{D-2n}}{(2\pi)^{\frac{D}{2}} \Gamma\left(\frac{D}{2}\right)} \int_0^1 dx\, x^{n - \frac{D}{2} - 1} (1 - x)^{\frac{D}{2} - 1}$$

with the substitution $x = \frac{m^2}{r^2 + m^2}$, and the remaining integral is a beta function defined in terms of gamma functions. The final result is

$$I_n^\mathrm{E}(m) = \frac{m^{D-2n}}{(2\pi)^{\frac{D}{2}} \Gamma\left(\frac{D}{2}\right)} \frac{\Gamma(n - \frac{D}{2}) \Gamma(\frac{D}{2})}{\Gamma(n)} = \frac{m^{D-2n}}{(2\pi)^{\frac{D}{2}}} \frac{\Gamma(n - \frac{D}{2})}{\Gamma(n)}$$

with the explicit divergence of the gamma function as a generalization of the factorial. There are poles for $n = 0, -1, -2, ...$ if D is an integer and the number n is given such that it cannot be changed.

Thus, the idea behind dimensional regularization is to regard D as a parameter in the Feynman diagram expansion that can be shifted slightly. It has no operational meaning to talk about D not being an integer but that does not matter because at the point where one has a Taylor series of an integral depending on the parameter D one can think of that as a hypothesis or as a cutoff. The hypothesis is $D = 4 - \varepsilon$ where $\varepsilon > 0$ and one can study how the integrals diverge with $\varepsilon \to 0$. The nice property of dimensional

regularization is that it does not break any symmetries such as the one from the Poincaré group or the local gauge symmetry.

The integral $I_2^{\mathrm{E}}(m)$ as an example becomes for $D = 4 - \varepsilon$

$$I_2^{\mathrm{E}}(m) = \frac{m^{-\varepsilon}}{(2\pi)^{2-\frac{\varepsilon}{2}}}\,\Gamma\left(\frac{\varepsilon}{2}\right) = \frac{1-\varepsilon\,\log(m)}{(2\pi)^2\left(1-\frac{\varepsilon}{2}\,\log(2\pi)\right)}\left(\frac{2}{\varepsilon}-\gamma\right)+O(\varepsilon^2)$$

where γ is the Euler-Mascheroni constant. The divergence comes from $\frac{2}{\varepsilon}$ when $\varepsilon \to 0$ and the cutoff at $|K| < \Lambda$ discussed above should be compared with $\frac{1}{\varepsilon}$.

Thus, $\varepsilon > 0$ renders every term in the Feynman diagram expansion finite but depending on ε and one has to add counterterms to the Lagrangian such that one can adjust the coupling constant to absorb the dependence. Note that the quantity D is a parameter and not the dimension of spacetime. Therefore, one does not renormalize the dimension of spacetime in this approach. This is not different from what has been done for the Feynman diagram on the right side with the hard cutoff $|K| < \Lambda$. The hard cutoff is easy to interpret operationally because it just means no momenta larger than the cutoff. In dimensional regularization it much harder to understand what it means to introduce the cutoff. The operational interpretation is therefore not clear.

6.5 Renormalization of Green's Functions

As introduced above a quantum theory $\boldsymbol{H}(z_1,...,z_n;\Lambda)$ is renormalizable if it leads to finite predictions for all operationally well-defined observables $\boldsymbol{A}_a$ and the expectation values are

$$\langle \boldsymbol{A}_a \rangle\,(z_1,...,z_n;K_c) = f_a(z_1,...,z_n;K_c) = \alpha_a^{\mathrm{obs}}$$

for all values of K_c by allowing the parameters $z_j(K_c)$ to depend on K_c. The fixed quantities α_a^{obs} are the experimentally observed values, and the difference between Λ and K_c is that Λ is the list of all degrees of freedom and K_c is a convenient number such as the length of this list specifying the number of degrees of freedom. The ϕ^4-theory is renormalizable as discussed above and the non-abelian gauge theory (Yang-Mills theory) is also renormalizable.

A question not yet fully answered is how the parameters $z_j(K_c)$ depend on K_c. One introduces a specific set of quantities that play the role of the $\boldsymbol{A}_a$ and these are the Green's functions

$$G^{(m)}(X_1,...,X_m;K_c) = \langle\Omega|\mathcal{T}\boldsymbol{\phi}(X_1)...\boldsymbol{\phi}(X_m)|\Omega\rangle$$

where the ground state $|\Omega\rangle = |\Omega(z_1,...,z_n;K_c)\rangle$ and the dynamics

$$\boldsymbol{\phi}(X) = e^{-i\,\boldsymbol{H}(z_1,...,z_n;K_c)\,t}\,\boldsymbol{\phi}(0,\vec{x})\,e^{i\,\boldsymbol{H}(z_1,...,z_n;K_c)\,t}$$

of the field operators depend on the cutoff K_c. The Green's function is not directly observable but it is near enough, and all observables can be expressed as functions of them if one has determined them. A quantum theory is therefore renormalizable if the Green's functions do not depend on the cutoff.

The cutoff K_c is usually a continuous parameter and may just be the maximum momentum $K_c = |P_{\max}|$ or similar. Continuous parameters can be differentiated using

$$F(x) = \int_0^x \frac{dF}{dy}dy - F(0)$$

and it is often easier to think about changing something than calculating it. One can do this also here, work out the Green's function for some value K_c and then change K_c infinitesimally. Thus, the Green's function changes with

$$dG^{(m)} = \frac{\partial G^{(m)}}{\partial K_c}\delta K_c + \frac{\partial G^{(m)}}{\partial z_j}\delta z_j$$

by changing $K_c \to K_c + \delta K_c$ and $z_j \to z_j + \delta z_j$ a little bit. It is assumed that $dG^{(m)} = 0$ because the Green's functions are not supposed to depend on K_c. If K_c changes to $K_c + \delta K_c$ and if $z_j(K_c)$ are chosen to fix $G^{(m)}$ then $dG^{(m)} = 0$ should be satisfied.

The equation becomes

$$K_c\frac{\partial G^{(m)}}{\partial K_c}+K_c\frac{dz_j}{dK_c}\frac{\partial G^{(m)}}{\partial z_j}=0 \qquad \beta_j\big(z_j(K_c)\big)=K_c\frac{dz_j}{dK_c} \qquad \left[K_c\frac{\partial}{\partial K_c}+\beta_j\big(z_j(K_c)\big)\frac{\partial}{\partial z_j}\right]G^{(m)}=0$$

and can be expressed with a beta function. (Multiplying by K_c has been done because of a variable change to exponential coordinates. It is usually easier to work with exponential cutoffs.) If one knows $\beta_j\big(z_j(K_c)\big)$ then one knows $z_j(K_c)$ because

$$\frac{dz_j}{dK_c}=\frac{1}{K_c}\beta_j\big(z_j(K_c)\big)$$

can be integrated.

As an aside, when one fixes $G^{(m)}$ to match the observations α^{obs} for massive theories one usually demands that the 2-point function $\langle\Omega|\mathcal{T}\boldsymbol{\phi}(P)\,\boldsymbol{\phi}(-P)|\Omega\rangle$ has a pole at m_{phys}. This is a problem when a theory is massless because it gives divergences. Thus, for massless theories one insists that $\langle\Omega|\mathcal{T}\boldsymbol{\phi}(P)\,\boldsymbol{\phi}(-P)|\Omega\rangle$ has a pole at $P^2=-K_c^2$ with residue 1. This leads to finite quantities for $K_c\to\infty$.

6.6 Beta Functions for Quantum Theories

The beta function for the ϕ^4-theory is

$$\beta(\lambda)=\frac{3\lambda^2}{16\pi^2}+O(\lambda^3) \tag{6.3}$$

and shows how the λ parameter should shift as one changes the cutoff. For quantum electrodynamics as a abelian gauge theory the beta function is

$$\beta(e)=\frac{e^3}{12\pi^2}+O(e^4) \tag{6.4}$$

where e is the coupling constant.

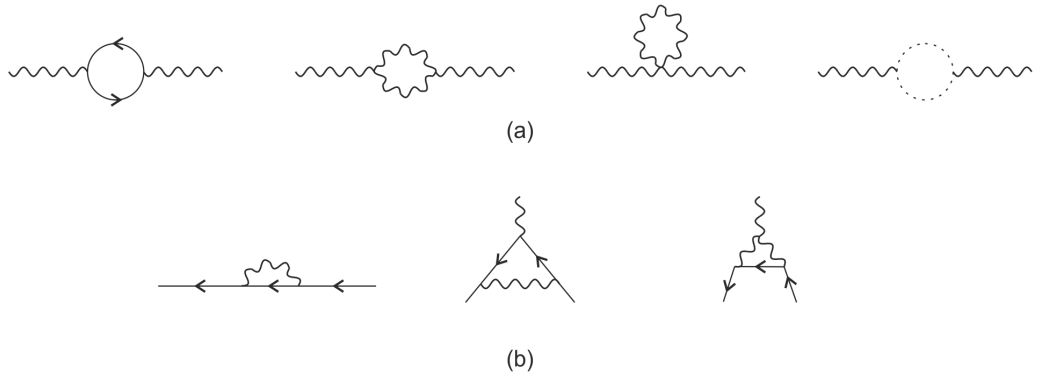


Figure 6: Diverging Feynman diagrams for the Yang-Mills theory

Non-abelian gauge theories have one free parameter g whose running can be worked out by analyzing the divergences of the diagrams presented in figure 6 and summing them together. The diagrams under (a) show propagation processes of gauge boson going to gauge boson where the left most diagram involves ghosts. Note that the gauge boson has no mass. These diagrams do not completely specify the dependence of the coupling constant g on the cutoff and one also has to study the fermion propagations in (b). Only after doing all these diagrams one can find out how the counterterms depend on the cutoff.

If one uses dimensional regularization for each of these processes and insists that all these integrals become finite then that tells how the counterterms in the theory have to change and the change of the counterterms, in turn, tells how the coupling constants should change. There is only the coupling constant g in the non-abelian gauge theories.

The beta function for the SU(N) local gauge group is

$$\beta(g) = \frac{-g^3}{(4\pi)^2}\left[\frac{11}{3}N - \frac{2}{3}n_f\right] \tag{6.5}$$

where n_f is the number of fermion families. If the number of fermion families is small the beta function $\beta(g)$ is negative and g gets smaller as K_c gets larger. That is amazing because even if g for low energies is enormously big it becomes very small as the theory goes to higher and higher energy processes such that one can use perturbation theory. This is called asymptotic freedom which allows to work nearly with the free theory.

6.7 Lattice Formulation of Non-Abelian Gauge Theories

Thus, perturbation theory cannot be used for low energy processes of non-abelian gauge theories but the low energy sector is the sector that pertains to the everyday life. Another approach is needed and the lattice formulation for low energy physics of non-abelian gauge theories going back to work of Kenneth Wilson published 1974 turned out to be very useful.

His idea was to use a lattice regulator. It is very challenging to put a quantum gauge theory on a lattice because of the local gauge invariance. The gauge group acts in a slightly non-local way as (5.10) or

$$A^j_\mu \frac{\sigma^j}{2} \to A^j_\mu \frac{\sigma^j}{2} + \frac{1}{g}(\partial_\mu \alpha^j)\frac{\sigma^j}{2} + i\left[\alpha^k \frac{\sigma^k}{2}, \alpha^l \frac{\sigma^l}{2}\right]$$

written with implicit summations. The slightly non-local term $\partial_\mu \alpha^j$ cannot be discretized as easily as $A^j_\mu(\varepsilon \underline{n})$ on the lattice. Wilsons solution was to use the parallel transporter and not the derivative directly. The fundamental degrees of freedom for a lattice approximation $\varepsilon\,\mathbb{Z}^4$ of a gauge theory is not the gauge field but the little parallel transporters

$$U(j, j+\hat{e}^\mu) \in \mathrm{SU}(2) \qquad \hat{e}^0 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix}, \hat{e}^1 = \begin{pmatrix}0\\1\\0\\0\end{pmatrix}, \hat{e}^2 = \begin{pmatrix}0\\0\\1\\0\end{pmatrix}, \hat{e}^3 = \begin{pmatrix}0\\0\\0\\1\end{pmatrix}$$

as shown in figure 7 reduced to time and one spatial dimension. A parallel transporter leads from one cell to the next, and one can build as in $U(j, j+2\hat{e}^\mu) = U(j+\hat{e}^\mu, j+2\hat{e}^\mu)\,U(j, j+\hat{e}^\mu)$ bigger transporters by multiplying. These parallel transporters are unitary 2×2 matrices because they are elements of SU(2), and Wilson's proposal is that in classical lattice gauge theory the fundamental degrees of freedom are lists of unitary 2×2 matrices, one per link of the lattice. This approach can be implemented on a computer but at Wilson's time only 4^4 lattice sites were possible while it is about 100^4 sites today.

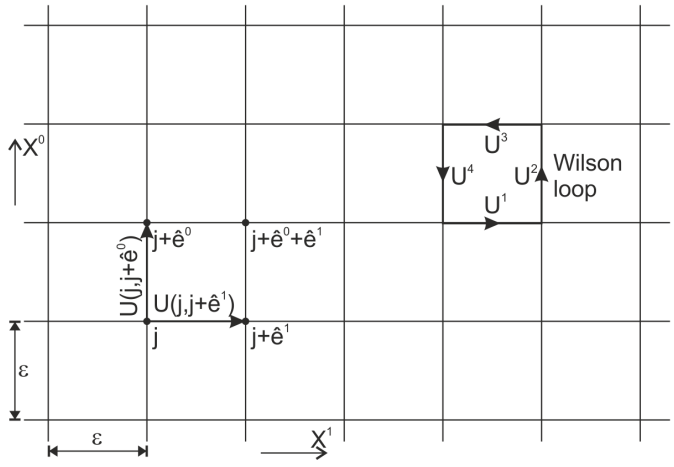


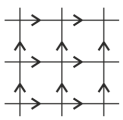
Figure 7: The parallel transporters on the lattice and the Wilson loop

To quantize the gauge field on the lattice Wilson's idea was to build an action and sample the path integral $\int \mathcal{D}U\, e^{-iS[U]}$ after a Wick rotation to get $\int \mathcal{D}U\, e^{-S[U]}$. One randomly generates 2×2 matrices U according to the probability distribution of the discretized path integral and takes the average. The action $S[U]$ of Wilson is a sum over so-called Wilson loops $S[U] = \mathrm{tr}(U_4 \cdot U_3 \cdot U_2 \cdot U_1)$. The Wilson loop illustrated in figure 7 with the parallel transporters is gauge invariant. The reason for the use of the plaquettes forming Wilson loops in the action is the covariant derivative. It gives dynamics to the degrees of freedom in non-abelian gauge theories and arises from the small plaquette limit. The remarkable result is that the outcome of the Monte Carlo simulations performed on a computer match the predictions for low energy processes.

This approach is extraordinarily successful and is the best way known to get non-perturbative answers from quantum field theories. However, it has some awkward downsides. One problem is that the calculations of the processes are done in imaginary time via the Wick rotation. This is roughly equivalent to do statistical mechanics with these gauge theories and therefore talking about processes at some temperature rather than talking about real time processes. It therefore does not work well for scattering processes, for example. Another problem is that this is not obviously a quantum theory because it is not clear what would guarantee that this theory gives rise to observations coming from a quantum theory. The quantum nature of the theory has been thrown out with the path integral.

6.8 Hamiltonian Lattice Gauge Theory

There is an alternative way to model gauge theories on a lattice in the Hamiltonian picture where one knows for sure that one has a quantum theory. The downside is that it is not so easy to perform on a normal computer but would be suitable for quantum computers. This approach has been proposed in 1975 by Kogut and Susskind and they argued that it yields the Yang-Mills theory as lattice spacing goes to zero but this has not been proven yet. All the predictions of this theory are finite and the theory is therefore based on a cutoff $K_c = \frac{1}{a}$ where a is the lattice spacing.



The figure on the right side shows some links of the lattice, and one 2×2 unitary matrix $U \in \mathrm{SU}(2)$ is associated per link as a parallel transporter. So far, there is no difference to Wilson's approach but in the Wilson's theory the degrees of freedom were classical. The way of quantizing this for the approach by Kogut and Susskind is to use wave functions. The quantum theory associates to each link e a wave function $\psi : \mathrm{SU}(2) \to \mathbb{C}$ with $\psi \in L^2(\mathrm{SU}(2))$. These square integrable wave functions build a Hilbert space $\mathfrak{H}_e$. The total Hilbert space $\mathfrak{H}_E$ is

$$\mathfrak{H}_E = \bigotimes_{e \in E} \mathfrak{H}_e = \bigotimes_{e \in E} L^2(\mathrm{SU}(2))$$

which is the tensor product over all Hilbert spaces of the links e in the lattice E.

To introduce the Hamiltonian for this lattice theory which should become the Hamiltonian for the Yang-Mills theory when the spacing goes to zero, some operations on the Hilbert space $L^2(\mathrm{SU}(2))$ are needed. States in $L^2(\mathrm{SU}(2))$ are defined as

$$|\psi\rangle = \int_{\mathrm{SU}(2)} dU\, \psi(U)\, |U\rangle$$

via position basis where $|U\rangle$ is a position eigenvector on SU(2). Because SU(2) is isomorphic to the sphere $\mathbb{S}^3$ one can see these wave functions as wave functions on the sphere $\mathbb{S}^3$ such that the mental path from wave functions on spheres leads via spherical harmonics and angular momentum to the hydrogen atom and this is all one needs to understand the Hilbert space $\mathfrak{H}_e$. Thus, $|U\rangle$ is defined by three coordinates, and the position basis is defined with $\langle U|V\rangle = \delta(U - V)$ using the Dirac delta function although this is not strictly correct.

The two operators L_U and R_U on this Hilbert space both $L^2(\mathrm{SU}(2)) \to L^2(\mathrm{SU}(2))$ are defined as

$$L_U\, |\psi\rangle \equiv \int dV\, \psi(V)\, |UV\rangle \qquad\qquad R_U\, |\psi\rangle \equiv \int dV\, \psi(V)\, |VU^\dagger\rangle$$

and are called left and right multiplication where the commutation relations are $[L_U, R_U] = 0$. They are unitary and form a representation of SU(2) with $L_U^\dagger L_U = R_U^\dagger R_U = \mathbb{I}$ and $L_{UV} = L_U L_V$. Note that UV

and $VU^\dagger$ are also elements of SU(2) because two unitary 2×2 matrices multiplied give also a unitary 2×2 matrix. These operations are the analog of the shift operator e^{ixp} in Schrödinger's mechanics where p is the momentum operator. Therefore these two operators give the momentum operators needed for the dynamics via differentiation because one way to define them is to say that a momentum operator is the thing one gets from the shift when making a smaller and smaller shift.

With a representation of a group on a space one can use representation theory to decompose this space into irreducible representation of the group. The theorem needed to understand how the infinite-dimensional Hilbert space $L^2(\mathrm{SU}(2))$ breaks up into a direct sum over irreducible representations of SU(2) is the Peter-Weyl theorem. It states applied to $L^2(\mathrm{SU}(2))$ that

$$\mathfrak{H}_e = L^2(\mathrm{SU}(2)) \cong \bigotimes_{l\in\frac{1}{2}\mathbb{Z}^+} V_l \otimes V_l^* \qquad\qquad V_l \cong \mathbb{C}^{2l+1}$$

where V_l is the vector space furnishing the irreducible representation of SU(2) of spin or angular momentum l and $\frac{1}{2}\mathbb{Z}^+ = \{0, \frac{1}{2}, 1, \frac{3}{2}, 2, ...\}$. SU(2) acts on the vector spaces V_l as $\Pi_l(\mathrm{SU}(2)) : V_l \to V_l$.

There is a quick way to determine the matrices Π_l for any spin where one needs to know only the fundamental representation (the spin-$\frac{1}{2}$ 2×2 matrix representation) of SU(2) and tensor products. To get the spin-l representation one takes the tensor product $V_{\frac{1}{2}} \otimes V_{\frac{1}{2}} \otimes ... \otimes V_{\frac{1}{2}}$ n times where $n = 2l$, and this is the space

$$V_l \subseteq V_{\frac{1}{2}} \otimes V_{\frac{1}{2}} \otimes ... \otimes V_{\frac{1}{2}} \cong \mathbb{C}^2 \otimes \mathbb{C}^2 \otimes ... \otimes \mathbb{C}^2 = \mathbb{C}^{2^n}$$

of n qbits $|0\rangle$ or $|1\rangle$ where V_l can be built as a subspace. The dimension of V_l is $n+1$ and the dimension of the tensor product is 2^n. In the next step one builds the set

$$|w_{\frac{n}{2}}\rangle = |11...11\rangle$$
$$|w_{\frac{n}{2}-1}\rangle = \frac{1}{\sqrt{n}}\big(\,|11...10\rangle + |11...01\rangle + ... + |10...11\rangle + |01...11\rangle\,\big)$$
$$\dots$$
$$|w_{\frac{n}{2}-k}\rangle = \frac{1}{\sqrt{\binom{n}{k}}}\big(\text{all possible combinations } |xx...xx\rangle \text{ with } k \text{ values } x=0 \text{ and } n-k \text{ values } x=1\big)$$
$$\dots$$
$$|w_{-\frac{n}{2}}\rangle = |00...00\rangle$$

of $n+1 = 2l+1$ orthonormal vectors in this big space $\mathbb{C}^{2^n}$. The matrix elements of $\Pi_l(U)$ are given by

$$\Big[\Pi_l(U)\Big]_{jk} = \langle w_j \,|\, U \otimes ... \otimes U \,|\, w_k\rangle \tag{6.6}$$

with n copies of U in the expectation value and with $j, k \in \{-\frac{n}{2}, ..., \frac{n}{2}\}$. This is the spin-$l$ representation of SU(2). (This works fine for low spin representations but is not very efficient for larger values of n such as $n = 1000$.) This method give $\Pi_0(U) = 1$ and $[\Pi_{\frac{1}{2}}(U)]_{jk} = [U]_{jk}$ for $l = 0$ and $l = \frac{1}{2}$, respectively. For $l = 1$ the orthonormal vectors are

$$|w_1\rangle = |11\rangle \qquad\qquad |w_0\rangle = \frac{1}{\sqrt{2}}(|01\rangle + |10\rangle) \qquad\qquad |w_{-1}\rangle = |00\rangle$$

and the matrix elements of the 3×3 matrix $\Pi_1(U)$ are

$$\big[\Pi_1(U)\big]_{11} = \langle 11|U\otimes U|11\rangle = \langle 1|U|1\rangle\,\langle 1|U|1\rangle = d^2$$
$$\big[\Pi_1(U)\big]_{00} = \frac{1}{2}\left(\langle 10|U\otimes U|10\rangle + \langle 01|U\otimes U|10\rangle + \langle 10|U\otimes U|01\rangle + \langle 01|U\otimes U|01\rangle\right) = ad + bc$$

with $\Pi_{\frac{1}{2}}(U) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Knowing the Peter-Weyl theorem and this trick one can therefore work out how $L^2(\mathrm{SU}(2))$ decomposes and how SU(2) acts on it, and one gets

$$L^2(\mathrm{SU}(2)) \cong \bigotimes_{l\in\frac{1}{2}\mathbb{Z}^+} \left(\mathbb{C}^{2l+1} \otimes \mathbb{C}^{2l+1}\right) \qquad\qquad L_U \cong \bigotimes_{l\in\frac{1}{2}\mathbb{Z}^+} \Pi_l(U) \otimes \mathbb{I}$$

with the matrix elements (6.6). This is infinite-dimensional but if one truncates the angular momentum then it is finite-dimensional. Also with the Peter-Weyl theorem the matrix element $[\Pi_l(.)]_{jk}$ is interpreted as a function $[\Pi_l(.)]_{jk}$ from SU(2) to the complex numbers such that, for example, $[\Pi_{\frac{1}{2}}(.)]_{11} = d$. These functions are

$$t^l_{jk}(U) \equiv [\Pi_l(U)]_{jk} : \mathrm{SU}(2) \to \mathbb{C}$$

and satisfy $t^l_{jk} \in L^2(\mathrm{SU}(2))$, and they form an orthogonal basis for $L^2(\mathrm{SU}(2))$ where $l \in \{0, \frac{1}{2}, 1, \frac{3}{2}, 2, ...\}$ and $-l \leq j, k \leq l$. Thus, every state $|\psi\rangle \in L^2(\mathrm{SU}(2))$ can be expressed as a linear combination of these functions

$$|\psi\rangle = \sum_l \sum_{j,k} \psi^l_{jk} \; |j\rangle_l \, |k\rangle_l$$

with $|j\rangle_l \, |k\rangle_l \equiv \sqrt{2l+1} \, |t^l_{jk}\rangle$. The basis is countable, orthogonal and normalized. The inner product in $L^2(\mathrm{SU}(2))$ is

$$(\psi, \phi) = \int dU \, \psi^*(U) \phi(U)$$

where dU is a Haar measure. The basis is normalized as

$$(l^l_{jk}, t^{l'}_{j'k'}) = \delta^{ll'} \, \delta_{jj'} \, \delta_{kk'} \frac{1}{2l+1}$$

shows. This gives a basis for the total Hilbert space $\mathfrak{H}_E$ containing wave functions $\Psi : \mathcal{C} \to \mathbb{C}$

$$|\Psi\rangle = \sum_{l_1 l_2 ...} \sum_{j_1 j_2 ...} \sum_{k_1 k_2 ...} \Psi^{l_1 l_2 ...}_{j_1 j_2 ... k_1 k_2 ...} \, |j_1\rangle_{l_1} \, |k_1\rangle_{l_1} \, |j_2\rangle_{l_2} \, |k_2\rangle_{l_2} \, ...$$

in $\mathfrak{H}_E$ where $\mathcal{C} = \mathrm{SU}(2) \times ... \times \mathrm{SU}(2)$ with one SU(2) per link e.

6.9 Dynamics for the Hamiltonian Lattice Gauge Theory

To add dynamics to this theory some observables on this Hilbert space $\mathfrak{H}_E$ are needed. If $U \in \mathrm{SU}(2)$ then $U = e^{c_j \, \tau^j}$ where $c_j \, \tau^j$ is an element of the Lie algebra of SU(2) consisting of three real numbers c_j and the three matrices τ^j. These matrices τ^j are the Pauli matrices multiplied by $\frac{i}{2}$ and satisfy the commutation relations $[\tau^j, \tau^k] = -2 \, \varepsilon^{jk}_{\;\; l} \, \tau^l$.

One can get the Pauli matrices by differentiation

$$\frac{d}{ds} e^{s \, \tau^j} \Big|_{s=0} = \tau^j$$

and can use this property to define the anti-Hermitian observables

$$\frac{d}{ds} L_{e^{s \, \tau^j}} \Big|_{s=0} = \boldsymbol{l}^j_L \qquad\qquad \frac{d}{ds} R_{e^{s \, \tau^j}} \Big|_{s=0} = \boldsymbol{l}^j_R$$

called left and right angular momentum. The position observable is the operator $L^2(\mathrm{SU}(2)) \to L^2(\mathrm{SU}(2))$ defined as

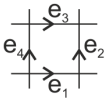
$$\boldsymbol{U}_{jk} \, |U\rangle = \big[\Pi_{\frac{1}{2}}(U)\big]_{jk} \, |U\rangle$$

when acting on position eigenstates $|U\rangle$. From

$$|\psi\rangle = \int dU \, \psi(U) \, |U\rangle \qquad\qquad \boldsymbol{U}_{jk} \, |\psi\rangle = \int dU \, \psi(U) \big[\Pi_{\frac{1}{2}}(U)\big]_{jk} \, |U\rangle = \int dU \, \psi(U) t^{\frac{1}{2}}_{jk}(U) \, |U\rangle$$

one can see that the position operator takes a wave function and multiplies it as expected by the position. Thus, there are operators for position and momentum and one can define the Hamiltonian as the kinetic energy plus the potential energy.

To do so the 4-dimensional plaquette operator



$$\begin{aligned}\boldsymbol{U}_\square : L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2)) \\ \to L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2))\end{aligned}$$

is defined via

$$\boldsymbol{M}_{\square,j,k} = \sum_{k_1,k_2,k_3} \boldsymbol{U}_{jk_1} \otimes \boldsymbol{U}_{k_1k_2} \otimes \boldsymbol{U}^\dagger_{k_2k_3} \otimes \boldsymbol{U}^\dagger_{k_3k} \qquad\qquad \mathrm{tr}(\boldsymbol{U}_\square) = \sum_k \boldsymbol{M}_{\square,k,k}$$

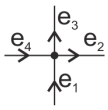
where $\boldsymbol{U}_{jk_1}$ is the U-observable for link e_1, $\boldsymbol{U}_{k_1k_2}$ for e_2, $\boldsymbol{U}_{k_2k_3}$ for e_3, $\boldsymbol{U}_{k_3k}$ for e_4 in the little figure on the right side. The object $\mathrm{tr}(\boldsymbol{U}_\square)$ is an operator and not a number that looks like a trace as an abuse of notation. It acts on $L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2)) \otimes L^2(\mathrm{SU}(2))$.

The Kogut-Susskind Hamiltonian is defined as

$$\boldsymbol{H} = -\frac{g}{2a} \sum_{e\in E} \boldsymbol{l}^j_L(e)\boldsymbol{l}^j_L(e) + \frac{1}{2g^2a} \sum_\square \big(\,\mathrm{tr}(\boldsymbol{U}_\square) + \text{Hermitian conjugates}\big) \tag{6.7}$$

where the first sum goes over all left angular momentum operators acting on link e squared and the second sum goes over all plaquettes. The angular momentum operators are anti-Hermitian and give a negative Hermitian operator when squared. The quantity g is the coupling constant. The theory with this $\boldsymbol{H}$ has a huge group of gauge symmetries.

No boundary conditions have been specified but the approach works for any boundary condition. The lattice may be infinite or finite. It may have periodic boundary conditions, closed boundary conditions or open boundary conditions such that the plaquette are not defined at the boundary.



As stated above the model with this Hamiltonian $\boldsymbol{H}$ has a huge group of gauge symmetries. The question is where the local gauge group is in this theory. The local gauge group acts on stars as shown in the figure on the left side. A vertex v in this lattice has four links that touch it and one can define the unitary operator

$$M_x(v) = R_x(e_1) \otimes L_x(e_2) \otimes L_x(e_3) \otimes R_x(e_4)$$

where $x \in \mathrm{SU}(2)$. These operators obey the commutation relations

$$[M_x(v), M_y(w)] = 0 \qquad\qquad [M_x(v), \boldsymbol{H}] = 0$$

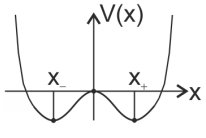
for all vertices v and w and all $x, y \in \mathrm{SU}(2)$. There is an independent copy of the action of SU(2) at every vertex leading to a huge number of symmetries for this Hamiltonian. That is how the local gauge symmetry manifests itself in this lattice model.

This model did not receive very much attention because Wilson's original formulation of gauge theory on a lattice turned out to be more useful for programs (especially for Monte Carlo simulations) and there are some open questions about this model. Wilson's formulation is fairly obviously a discretization of a gauge theory. However, the Kogut-Susskind formulation looks like a promising candidate for next generation simulation on quantum computers because quantum computers are not very suitable for Monte Carlo simulations but can naturally simulate local lattice models.

7 Spontaneous Symmetry Breaking and the Higgs Mechanism

7.1 Discrete Examples of Spontaneous Symmetry Breaking

The first example for spontaneous symmetry breaking is a classical particle in a double well potential $V(x)$ shown in the figure on the right side with the Hamiltonian



$$H = \frac{p^2}{2m} + V(x)$$

and the two ground states $x_\pm$. This system has a $\mathbb{Z}_2$ symmetry because one gets the same dynamics for $x \to -x$. However, the ground state of the system is degenerate and does not respect this symmetry. The particle is either at x_+ or x_- and there is not one unique ground state. Each ground state breaks this symmetry but the symmetry is not lost because the Hamiltonian still has this symmetry. The symmetry has created this degeneracy. The ground states individually are not symmetric but the manifold of ground states is.

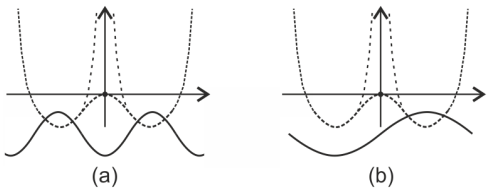


Figure 8: Ground state $\psi_0(x)$ and first excited state $\psi_1(x)$ in the quantum double well potential

The quantum version of this system is not exhibiting this spontaneous symmetry breaking as figure 8 shows. The ground state $\psi_0(x)$ in (a) is symmetric and does not break the symmetry at all. The first excited state $\psi_1(x)$ in (b) is separated by a small gap related to the barrier depth. This is a case where the classical system exhibits symmetry breaking but the quantum system is symmetric under $x \to -x$.

The next example is the Ising model from statistical physics. It describes the dynamics of ferromagnetic materials classically with spins $s_j = \pm 1$ either pointing up or down arranged as a grid. The Hamiltonian is

$$H = -\sum_{\langle j,k \rangle} s_j s_k$$

where antialigned spins $s_j \neq s_k$ are penalized but aligned spins $s_j = s_k$ are not penalized. There are two obvious ground states corresponding to the cases where all spins point up or all spins point down, and H has a $\mathbb{Z}_2$ symmetry acting on the spins $s_j \to -s_j$. There are two regimes for the thermal state of this system. Looking at the Gibbs state of the Hamiltonian

$$\rho = \frac{e^{-\beta H}}{Z}$$

one can see that it is $\mathbb{Z}_2$ symmetric because of H. Small fluctuations will lead to a breaking of symmetry depending on the temperature T. There is a critical temperature below which the system is ordered. For β going to infinity $(\frac{1}{2} - \varepsilon)$ of the spins point up and $(\frac{1}{2} + \varepsilon)$ point down (plus some corrections). Small external fluctuations will force the system into either all up or all down and doing so break the symmetry. Above the critical temperature the system is disordered.

A $\mathbb{Z}_2$ symmetry breaking example in classical field theory is based on the Lagrangian density

$$\mathcal{L} = \frac{1}{2} (\partial_\mu \phi) (\partial^\mu \phi) + \frac{1}{2}\mu^2\phi^2 - \frac{\lambda}{4!}\phi^4$$

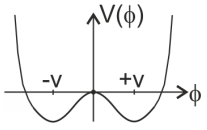
which looks like the Lagrangian density of the ϕ^4-theory but has a negative mass term $\frac{1}{2}\mu^2\phi^2$ instead of the positive mass term $\frac{1}{2}m^2\phi^2$ of the ϕ^4-theory. The Hamiltonian is

$$H = \int d^3\vec{x} \left[\frac{1}{2}\pi^2 + \frac{1}{2}(\nabla\phi)^2 - \frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4!}\phi^4 \right]$$

and one can ask what classical configuration has the smallest energy. The minimum for $V(\phi)$ is

$$V(\phi) = -\frac{1}{2}\mu^2\phi^2 + \frac{\lambda}{4!}\phi^4 \qquad\qquad \frac{\partial V(\phi)}{\partial \phi} = -\mu^2\phi + \frac{\lambda}{6}\phi^3 = 0$$

with three solutions $\phi = 0$, $\phi = \pm\sqrt{6\mu^2/\lambda} = \pm v$ where the first one is not a minimum energy solution. Illustrated in the figure on the right side the other two solutions show the same double well potential as the first example but in this case the quantum



version also exhibits spontaneous symmetry breaking. The difference is that in the first example the first excited state was separated from the ground state by a small gap but this is not the case here because there is an infinite number of these double wells. A fluctuation will push the system in one of the two states.

The transverse Ising model is a quantum example of spontaneous symmetry breaking. Its Hamiltonian

$$\boldsymbol{H} = -\sum_{j\in\mathbb{Z}} \sigma_j^x \sigma_{j+1}^x + h \sum_{j\in\mathbb{Z}} \sigma_j^z$$

with

$$\sigma^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad\qquad \sigma^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

is a model pertaining to a 1-dimensional lattice of quantum spins. The first term in the Hamiltonian sums over neighbors and the second term represents the magnetic field. If the two neighbors both point in the same direction then it is not penalized but if the point in opposite directions then it is penalized. There are two ground states

$$|\Omega_+\rangle = |+\rangle\,|+\rangle\,...\,|+\rangle \qquad\qquad |\Omega_-\rangle = |-\rangle\,|-\rangle\,...\,|-\rangle$$

with

$$|+\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + |1\rangle\right) \qquad\qquad |-\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle - |1\rangle\right)$$

for the magnetic field strength $h = 0$. The Hamiltonian has $\mathbb{Z}_2$ symmetry because $\Phi = ...\sigma_{j-1}^z \sigma_j^z \sigma_{j+1}^z ...$, $[\Phi, \boldsymbol{H}] = 0$ and $\Phi^2 = \mathbb{I}$. Saying that there are two ground states for $h = 0$ is not quite right. The correct statement is that the ground eigenspace is 2-fold degenerate. In quantum mechanics one is allowed to take superpositions of states such that one can use

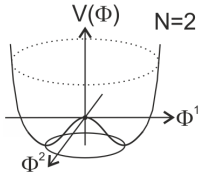
$$\frac{1}{\sqrt{2}}\left(|\Omega_+\rangle + |\Omega_-\rangle\right) \qquad\qquad \frac{1}{\sqrt{2}}\left(|\Omega_+\rangle - |\Omega_-\rangle\right)$$

as the basis instead of $|\Omega_+\rangle$ and $|\Omega_-\rangle$ but one never sees these ground states in experiments because of the Schrödinger cat states. It is a massive superposition of all spins pointing to the left and all spins pointing to the right. The tiniest decoherence in form of interactions with the environment will immediately destroy this superposition and prefer the system to enter one of the two ground states $|\Omega_+\rangle$ and $|\Omega_-\rangle$. Thus, the system will exhibit a broken symmetry because the ground states observed in the lab are either $|\Omega_+\rangle$ or $|\Omega_-\rangle$ and neither of the two states is symmetric under Φ.

7.2 Continuous Examples of Spontaneous Symmetry Breaking

In the above examples the symmetry is discrete but there are also examples with continuous symmetries. The inverted pendulum is an easy classical model. The pendulum fixed at the bottom on a flat surface going straight up is not in the state with lowest energy. It is symmetric under rotations in this metastable state but it easily falls down.

The linear σ-model is another example of a continuous symmetry that gets broken. It is a reasonable model of pions and has the Lagrangian density



$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \phi^j\right)^2 + \frac{1}{2}\mu^2 (\phi^j)^2 - \frac{\lambda}{4!}\left((\phi^j)^2\right)^2$$

with N scalar fields labeled $j = 1, ..., N$. The dynamics are invariant under transformations $\phi^j \to [O]^j_{\ k}\, \phi^k$ for any $N \times N$ orthogonal matrix O. Thus, the orthogonal matrices $\mathrm{O}(N)$ build a continuous group of symmetries. The potential visualized in the figure on the right side for $N = 2$ is

$$V(\phi) = -\frac{1}{2}\mu^2 (\phi^j)^2 + \frac{\lambda}{4!}\left((\phi^j)^2\right)^2$$

and the lowest energy configuration satisfy $(\phi_0^j)^2 = \frac{\mu^2}{\lambda}$. There are several solution such as

$$\underline{\phi}_0 = \begin{pmatrix} \frac{\mu}{\sqrt{\lambda}} \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \qquad\qquad \underline{\phi}_0 = \begin{pmatrix} 0 \\ \frac{\mu}{\sqrt{\lambda}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and the potential is called the bottom of the wine bottle potential or the Mexican hat potential. There is not just a finite number of minima but all the continuous points on the circle shown in the figure are minima. Thus, any configuration on this circle is a lowest energy configuration for this field. Each ground state breaks symmetry but the manifold of all ground states has the symmetry.

The ground state of a field is boring because of no dynamics but the low energy dynamics of the system with spontaneous symmetry breaking is interesting. The question is how this system behaves a little bit away from the lowest energy configuration.

In the case of $\mathbb{Z}_2$ symmetry a particle in the double well slightly away from the lowest energy configuration oscillates and does not have enough energy to traverse the barrier or tunnel through it. Effectively the low energy dynamics are that of a harmonic oscillator with some corrections. In particular it looks as if the particle has a mass because this effective harmonic oscillator has some width and that width is proportional to the mass. This effective mass corresponds to the restoring force.

In the case of the continuous symmetry one chooses new coordinates

$$\underline{\phi}_0 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ v \end{pmatrix} \qquad\qquad v = \frac{\mu}{\sqrt{\lambda}}$$

and defines shifted classical fields π^k and σ as

$$\underline{\phi}(X) = \left(\pi^k(X), v + \sigma(X)\right)$$

with $k = 1, ..., N-1$. The Lagrangian density is

$$\mathcal{L} = \frac{1}{2}\left(\partial_\mu \pi^k\right)^2 + \frac{1}{2}\left(\partial_\mu \sigma\right)^2 - \frac{1}{2}\left(2\mu^2\sigma^2 - \sqrt{\lambda}\mu\sigma^3 - \sqrt{\lambda}\mu(\pi^k)^2\sigma - \frac{\lambda}{4}\sigma^4 - \frac{\lambda}{2}(\pi^k)^2\sigma^2 - \frac{\lambda}{4}\left((\pi^k)^2\right)^2\right)$$

in terms of these shifted fields. The is the Lagrangian density of $N-1$ massless fields π^k and one massive field σ with some interactions between them.

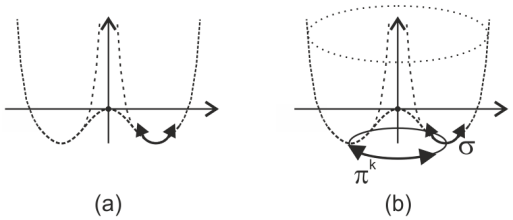


Figure 9: The low energy dynamics for the double well and the Mexican hat potential

The different fields in the Mexican hat potential are illustrated in figure 9 for $N = 2$. The field σ in (b) corresponds to the situation (a) in the double well potential and gets its mass due to the restoring force. A fluctuation perpendicular to the circle of ground states costs energy leading to an effective mass but moving around the circle of lowest energy costs nothing such that the fields π^k correspond to effective massless degrees of freedom. For $N > 2$ there is always one direction that costs energy (the σ field) and $N-1$ directions tangentially to the manifold of the minima (the π^k fields).

7.3 Goldstone's Theorem

This is not just an accident of the linear σ-model but the consequence of a much more general result due to Jeffrey Goldstone. In the O(N) linear σ-model there are $\binom{N}{2}$ (the dimension of the rotation group O(N)) independent continuous symmetries and after symmetry breaking there are $\binom{N-1}{2}$ remaining symmetries corresponding to the rotation group O($N-1$) of the fields π^k. However, there is no obvious symmetry between the σ and the π^k fields. Thus, the number of broken symmetries is $N-1=\binom{N}{2}-\binom{N-1}{2}$ and the number of massless fields is the same as the number of broken symmetries. The Goldstone theorem states that for every broken symmetry there is a corresponding massless bosonic particle called Goldstone mode or Goldstone boson.

To prove this theorem one can consider a theory with boson fields ϕ^a with $a=1,2,...$ where the Lagrangian density is assumed to have the general form $\mathcal{L}=(\text{derivatives})-V(\phi^a)$. The goal is to find the potential minima, the classical configurations, analyze the dynamics of fluctuations around these minima and prove that per broken symmetry these classical configurations give rise to a massless boson in the effective theory of the low energy dynamics. To do so

$$\frac{\partial V}{\partial \phi^a}\Big|_{\phi^a=\phi_0^a}=0 \qquad\qquad V(\phi^a)=V(\phi_0^a)+\frac{1}{2}(\phi^a-\phi_0^a)(\phi^b-\phi_0^b)\left(\frac{\partial^2 V}{\partial\phi^a\,\partial\phi^b}\right)_{\phi=\phi_0}$$

gives the minima of the potential and the Taylor expansion of the potential close to the minima. (Because of the left equation there is no first-order term in the right equation.) The Hessian matrix $\underline{\underline{m^2}}$ defined as $[m^2]_{ab}=\partial^2 V/(\partial\phi^a\,\partial\phi^b)$ is symmetric and the real eigenvalues give masses for the effective fields. Because $\underline{\underline{m^2}}$ is symmetric it can be diagonalized by an orthogonal matrix $\underline{\underline{m^2}}=\underline{\underline{O}}^T\,\underline{\underline{d}}\,\underline{\underline{O}}$ such that one can define the fields as $\pi^a=[O]_b\phi^b$ and the Lagrangian density becomes $\mathcal{L}=(\text{derivatives})-d_a^2(\pi^a)^2+...$ and these terms are mass terms.

The next step of the proof is to show that these mass terms can be zero or, in other words, that every continuous symmetry leads to a zero eigenvalue. A general symmetry has the form

$$\phi^a\to\phi^a+\alpha\Delta^a(\phi)$$

where α is infinitesimal and Δ^a is a function shifting the field. This symmetry is a symmetry of V because derivatives vanish. Thus, the symmetry for the potential with its expansion to first order

$$V(\phi^a)=V(\phi^a+\alpha\Delta^a(\phi)) \qquad\qquad \Delta^a(\phi)\frac{\partial}{\partial\phi^a}V(\phi)=0$$

show that the directional derivative of the potential in the direction $\Delta^a(\phi)$ is zero. Differentiating the expansion to first order with respect to ϕ^b gives

$$\frac{\partial\Delta^a(\phi)}{\partial\phi^b}\,\frac{\partial V(\phi)}{\partial\phi^a}+\Delta^a(\phi)\,\frac{\partial^2 V(\phi)}{\partial\phi^a\,\partial\phi^b}=0$$

but when $\phi^a=\phi_0^a$ then the first term vanishes and the result is

$$\sum_a\Delta^a(\phi)\,\frac{\partial^2 V(\phi)}{\partial\phi^a\,\partial\phi^b}\Big|_{\phi^a=\phi_0^a}=\sum_a\Delta^a(\phi_0^a)\,[m^2]_{ab}=0$$

with explicit summation. This can be written as $\underline{\Delta}^T\,\underline{\underline{m^2}}=0$ where $\underline{\Delta}$ is therefore a zero eigenvector of the matrix $\underline{\underline{m^2}}$. One finally has to argue that $\Delta^a(\phi^a)$ are linearly independent for each continuous symmetry but that follows from the definition of independent continuous symmetry.

7.4 The Higgs Mechanism for an Abelian Model

The question is here what happens when one combines Goldstone's theorem which pertains the symmetry breaking with local gauge invariance. In the SU(2) gauge theory

$$\mathcal{L}=F^{\mu\nu}F_{\mu\nu}+\bar{\psi}(i\not{D}-m)\psi$$

there is the puzzle that the gauge bosons (the auxiliary fields introduced to give the theory the local gauge symmetry) have no explicit mass term. In the weak interactions there are massive gauge bosons but if one adds a mass term $\frac{1}{2}m^2 A^j_\mu A^j_\mu$ to the Lagrangian density then it breaks gauge invariance. The reason is that these gauge fields transform according to (5.3) for the abelian case and (5.10) for the non-abelian case but the added mass term does not transform correctly. Thus, massive gauge bosons and gauge invariance cannot coexist.

Whenever there are constraints in physics, adding new degrees of freedom can resolve them. One cannot expect solutions, for example, when N equations with M unknowns are given where $N \gg M$. Overconstrained problems usually have no solution and adding unknows may help. This is exactly what happens here with the Higgs mechanism. Every new degree of freedom in high energy physics is a particle (field). Thus, with adding a degree of freedom one makes a prediction, and predictions can be tested.

As a toy model the abelian case gives valuable insight. This model is a theory of a complex scalar field $\phi = \phi_1 + i\,\phi_2$ coupled to a U(1) gauge field A_μ. The Lagrangian density is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu} + |D_\mu\phi|^2 - V(\phi)$$

where $D_\mu = \partial_\mu + i\,e\,A_\mu$ is the covariant derivative. This theory is invariant under local gauge transformations such that

$$A_\mu(X) \to A_\mu(X) + \frac{1}{e}\,\partial_\mu\alpha(X) \qquad\qquad \phi(X) \to e^{i\,\alpha(X)}\,\phi(X)$$

must be satisfied. Because $e^{i\,\alpha(X)}$ is just a phase $\in \mathbb{S}^1$ the term $|D_\mu\phi|^2$ in $\mathcal{L}$ is invariant under this gauge transformation. The potential

$$V(\phi) = -\mu^2\phi^*\phi + \frac{\lambda}{2}\,(\phi^*\phi)^2$$

with $\mu^2 > 0$ is a Mexican hat potential similar to the above examples.

The steps are the same as for spontaneous symmetry breaking and start with finding ϕ_0 which minimizes $V(\phi)$. There is a whole family of minima multiplied by a phase $e^{i\,\varphi}$

$$e^{i\,\varphi}\,\sqrt{\frac{\mu^2}{\lambda}} \qquad\qquad \phi_0 = \sqrt{\frac{\mu^2}{\lambda}}$$

but here the minimum ϕ_0 with $\varphi = 0$ is selected. The field expanded around this $\phi = \phi_0$ is

$$\phi(X) = \phi_0 + \frac{1}{\sqrt{2}}\,(\phi_1(X) + i\,\phi_2(X))$$

and allows to study what is going to happen for small fluctuations $\phi_1(X)$ and $\phi_2(X)$ around this steady state. The potential becomes

$$V(\phi) \approx -\frac{1}{2\lambda}\mu^4 + \frac{1}{2}2\,\mu^2\phi_1^2 + O(\phi_j^3)$$

and is independent of the imaginary ϕ_2 as expected. The Lagrangian density is

$$\mathcal{L} \approx \frac{1}{2}|\partial_\mu\phi_1|^2 + \frac{1}{2}|\partial_\mu\phi_2|^2 - V(\phi) + ... \approx \frac{1}{2}|\partial_\mu\phi_1|^2 + \frac{1}{2}|\partial_\mu\phi_2|^2 + \frac{1}{2\lambda}\mu^4 - \,\mu^2\phi_1^2 + ... + O(\phi_j^3)$$

around ϕ_0. One sees that there is an effectively massive particle ϕ_1 and an effectively massless particle ϕ_2. The covariant derivative contains the kinetic energy of the two fields in the first two terms and is

$$|D_\mu\phi|^2 = \frac{1}{2}(\partial_\mu\phi_1)^2 + \frac{1}{2}(\partial_\mu\phi_2)^2 + \sqrt{2}\,e\,\phi_0\,A_\mu\,\partial^\mu\phi_2 + e^2\,\phi_0^2\,A_\mu A^\mu + ...$$

where the term $\Delta\mathcal{L} = e^2\,\phi_0^2\,A_\mu A^\mu$ is interesting. It contains e and ϕ_0 as constants and $A_\mu A^\mu$ looks like the mass term. Thus, the term $\Delta\mathcal{L} = \frac{1}{2}\,m_A^2\,A_\mu A^\mu$ is the bosonic mass term. The originally massless gauge bosons behave around ϕ_0 for all practical purposes as though they have a mass. The whole Lagrangian density is still gauge invariant, and this is a classical theory for the low-energy limit.

For the quantum version of this theory one can use path integral quantization on $\mathcal{L} = \mathcal{L}_0 + \mathcal{L}_{\text{int}}$ with

$$\mathcal{L}_0 = \frac{1}{2}(\partial_\mu \phi_1)^2 - \mu^2 \phi_1^2 + \frac{1}{2}(\partial_\mu \phi_2)^2 - \frac{1}{4} F_{\mu\nu} F^{\mu\nu}$$

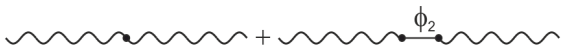
where the constant term $\frac{1}{2\lambda}\mu^4$ has been shifted away. The remaining terms belong to $\mathcal{L}_{\text{int}}$ interpreted as a perturbation. The term $\sqrt{2}\, e\, \phi_0\, A_\mu\, \partial^\mu \phi_2$ is an interaction between the gauge field and the ϕ_2 field. It corresponds to the vertex

$$= i\sqrt{2}\, e\, \phi_0 (-i K^\mu) = m_A K^\mu$$

in momentum space. There are also cubic and quartic terms in $\mathcal{L}_{\text{int}}$ leading to other Feynman diagrams.

The mass m_A of $\boldsymbol{A}_\mu$ as an operationally well-defined and experimentally observable quantity appears at the quantum level in poles of the Green's functions as in

$=$



$$= i\, m_A^2\, \eta^{\mu\nu} + (m_A K^\mu) \frac{i}{K^2} (m_A K^\nu) \sim i\, m_A^2$$

where a gauge boson turns in the second term into a ϕ_2 and back into a gauge boson.

7.5 The Higgs Mechanism for a Non-Abelian Model

A set of scalar fields $\underline{\phi}$ transforming under a general continuous group $\mathcal{G}$ as

$$\underline{\phi} = \begin{pmatrix} \phi_1 \\ \vdots \end{pmatrix} \qquad\qquad \phi_j \overset{g\in\mathcal{G}}{\to} (\mathbb{I} + i\, \alpha^a\, t^a)_{jk}\ \phi_k$$

is given with $g = \exp(i\, \alpha^a\, t^a)$ (or $\pi(g) = \exp(i\, \alpha^a\, t^a)$ to stress that it is a representation of the group) where it is easiest when the matrices t^a are purely imaginary. They span a Lie algebra of this group and their dimensions depend on the representation but note that the symmetry is global. To absorb the imaginary i and to simplify the notation the matrices $T^a = i\, t^a$ are introduced.

A gauge theory based on $\mathcal{G}$ is built by promoting it to a local symmetry $\underline{\phi} \to \pi(g(X))\, \underline{\phi}$ and replacing partial derivatives ∂_μ with covariant derivatives $D_\mu \underline{\phi} = (\partial_\mu + g\, A_\mu^a\, T^a)\underline{\phi}$. Squaring the covariant derivative and dividing it by 2 gives

$$\frac{1}{2}(D_\mu \underline{\phi})^2 = \frac{1}{2}\left((\partial_\mu + g\, A_\mu^a\, T^a)_{jk} \phi_k\right)^2 = \frac{1}{2}(\partial_\mu \phi_j)^2 + g\, A_\mu^a (\partial_\mu \phi_j\, T_{jk}^a\, \phi_k) + \frac{1}{2}\, g^2\, A_\mu^a A^{b\mu} (T^a\phi)_j (T^b\phi)_j$$

which is the kinetic energy term. Combining it with a version of the Mexican hat potential

$$V(\underline{\phi}) = -\nu \left|\underline{\phi}\right|^2 + \lambda \left|\underline{\phi}\right|^4$$

where $\underline{\phi}$ has classical minima at $\underline{\phi}_0$ allows to expand around $\underline{\phi}_0$ as $\underline{\phi}(X) = \underline{\phi}_0(X) + \varepsilon \underline{\phi}_1(X)$ with fluctuations $\varepsilon \underline{\phi}_1$ up to order ε.

The Lagrangian density certainly has the kinetic energy term and the potential energy term but may have other terms which are not important here. The kinetic energy contains a term

$$\Delta\mathcal{L} = \frac{1}{2}(m^2)_{ab}\, A_\mu^a A^{b\mu}$$

after substituting the first-order expansion $\underline{\phi}(X) = \underline{\phi}_0(X) + \varepsilon \underline{\phi}_1(X)$ into it. The matrix $(m^2)_{ab}$ is given by $(T^a\phi)_j (T^b\phi)_j$ in the term $\frac{1}{2}\, g^2\, A_\mu^a A^{b\mu} (T^a\phi)_j (T^b\phi)_j$ applied to $\underline{\phi}_0$ such that $-(m^2)_{ab} = g^2\, (T^a\phi_0)_j (T^b\phi_0)_j$. The matrix $-(m^2)_{ab}$ is a positive semidefinite matrix where the indices a and b run from 1 to $\dim(\mathcal{G})$. Because it depends on the group it can therefore be a big matrix.

For $\mathcal{G} = \mathrm{SU}(2)$ the index a runs from 1 to 3 for the fundamental representation and

$$T^1 = i\,\sigma^x \qquad T^2 = i\,\sigma^y \qquad T^3 = i\,\sigma^z \qquad \underline{\phi} = \begin{pmatrix}\phi_1\\ \phi_2\end{pmatrix} \qquad \underline{\phi}_0 = \begin{pmatrix}1\\ 0\end{pmatrix}$$

are the matrices T^a as well as $\underline{\phi}$ and $\underline{\phi}_0$ which must be 2-dimensional vectors because T^a are 2×2-matrices. Thus, the values needed to build $(m^2)_{ab}$ are

$$T^1\phi_0 = i\begin{pmatrix}0\\ 1\end{pmatrix} \qquad T^2\phi_0 = -\begin{pmatrix}0\\ 1\end{pmatrix} \qquad T^3\phi_0 = i\begin{pmatrix}1\\ 0\end{pmatrix}$$

and the matrix $(m^2)_{ab}$ becomes

$$m^2 = g^2\begin{pmatrix}-1 & -i & 0\\ -i & -1 & 0\\ 0 & 0 & -1\end{pmatrix}$$

which is multiplied by -1 supposed to be positive semidefinite. If the representation t is purely imaginary there are no problems.

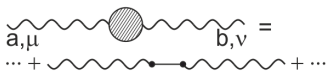
When building a theory one may take another representation than the fundamental presentation. In this case the matrices T^a would become bigger matrices. Using the fundamental representation all $T^a\underline{\phi}_0$ gave contributions but it could be in another representation that a T^a annihilates $\underline{\phi}_0$. If a generator T^a leaves the vacuum invariant then $T^a\underline{\phi}_0 = 0$ and there will be massless bosons. It is possible by choosing a representation where one gets zero masses in form of zero eigenvalues. This is important because in the Glashow-Weinberg-Salam theory of weak interactions there are gauge bosons with a mass and others without a mass.

To summarize, given $\underline{\phi}$ whose components are scalar fields ϕ_1 to ϕ_d which transform under the local gauge group $\mathcal{G}$ as $\phi_j \to (\mathbb{I} + i\,\alpha^a\,t^a)_{jk}\;\phi_k$ one can add a potential $\mathcal{L}_0 \to \mathcal{L} + V(\underline{\phi})$ to the Lagrangian density of the local gauge theory of bosons. If the components of the classical minimal configuration $\underline{\phi}_0$ is v for $j = d$ and 0 for all $j \neq d$, one term in the covariant derivative and the matrix $\underline{\underline{m^2}}$ are

$$\left|D_\mu\underline{\phi}\right|^2 = ... + \frac{1}{2}\,g^2\,(\underline{0}\;v)\,T^aT^b\begin{pmatrix}\underline{0}\\ v\end{pmatrix}A^a_\mu A^{b\mu} + ... \qquad (m^2)_{ab} = g^2\,(\underline{0}\;v)\,T^aT^b\begin{pmatrix}\underline{0}\\ v\end{pmatrix} \tag{7.1}$$

when expanding $\underline{\phi} = \underline{\phi}_0 + \varepsilon\,\underline{\phi}_1$ close to the classical minima. The mass matrix is positive or negative semidefinite depending on the metric. Because of $A^a_\mu A^{b\mu} = A^a_0A^{b0} - A^a_1A^{b1} - A^a_2A^{b2} - A^a_3A^{b3}$ the only non-negative term is the longitudinal term $A^a_0A^{b0}$. Thus, if the degrees of freedom of this theory are such that the gauge bosons have only transverse components of momentum as the photon does there is a legitimate mass term.

The way to check if the gauge bosons are transverse also in the quantum theory is to study the vacuum polarization process in the figure on the right side where the diagram shown cancels off the longitudinal contribution. The vacuum polarization process is



$$i\,(m^2)_{ab}\left(\eta^{\mu\nu} - \frac{K^\mu\,K^\nu}{K^2}\right)$$

and is therefore transverse. For SU(2) the mass matrix is proportional to the identity such that all gauge bosons have the same mass.

7.6 Theory of the Weak Interactions

The Glashow-Weinberg-Salam theory of the weak interactions is only sketched here. It describes SU(2) gauge bosons and electromagnetic field interactions combined in one theory. The local gauge group is $\mathrm{SU}(2) \times \mathrm{U}(1)$ and one of the questions is how it is possible to give some gauge bosons mass and others no mass. Though it has not been proven that photons as the gauge bosons of electromagnetism are massless

but they are for sure extraordinarily light. The gauge bosons W and Z of the weak interactions on the other hand have a mass.

The local gauge group acts on a boson field as

$$\underline{\phi} \to e^{i\,\alpha^a \tau^a}\, e^{i\,\frac{\beta}{2}} \underline{\phi} \tag{7.2}$$

where ϕ is a doublet, $\tau^a = \frac{\sigma^a}{2}$ are the Pauli matrices and β is a scalar. The U(1) part of the local gauge symmetry gives the boson field a charge of $\frac{\beta}{2}$. In the next step one adds a potential $V(\underline{\phi})$ to the Lagrangian density so that $\underline{\phi} = \underline{\phi}_0 + \underline{\phi}_1$ with the classical minimum

$$\underline{\phi}_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}$$

and expands the covariant derivative around $\underline{\phi}_0$ as above. The covariant derivative is

$$D_\mu \underline{\phi} = \left(\partial_\mu - i\, g\, A^a_\mu \tau^a - i\, \frac{1}{2}\, g'\, B_\mu \right) \underline{\phi}$$

and the connection gauge field is

$$\begin{pmatrix} A^1_\mu \\ A^2_\mu \\ A^3_\mu \\ B_\mu \end{pmatrix}$$

for the $\mathrm{SU}(2) \times \mathrm{U}(1)$ local gauge group in the fundamental representation (7.2). The Lagrangian density of this theory contains terms like $\left| D_\mu \underline{\phi} \right|^2$, $F_{\mu\nu} F^{\mu\nu}$, fermionic terms and the potential $V(\underline{\phi})$.

Going through the steps discussed above shows how the masses for the gauge bosons are generated. Expansion around $\underline{\phi}_0$ leads to a mass term

$$\frac{1}{2} \frac{v^2}{4} \left[g^2 (A^1_\mu)^2 + g^2 (A^2_\mu)^2 + (-g\, A^3_\mu + g' B_\mu)^2 \right]$$

from $\left| D_\mu \underline{\phi} \right|^2$. Thus, the gauge bosons A^1 and A^2 behave as if they have a mass. Similarly, the particular linear combination of A^3 and B in the last term also has a mass. The fourth degree of freedom is massless. It is convenient to give these gauge fields names as

$$\begin{aligned} W^\pm_\mu &= \frac{1}{\sqrt{2}} \left(A^1_\mu \mp i A^2_\mu \right) \\ Z^0_\mu &= \frac{1}{\sqrt{g^2 + g'^2}} \left(g A^3_\mu - g' B_\mu \right) \\ A_\mu &= \frac{1}{\sqrt{g^2 + g'^2}} \left(g' A^3_\mu + g B_\mu \right) \end{aligned} \tag{7.3}$$

where the two $W^\pm$ bosons have mass $\frac{1}{2}\, g\, v$, the Z^0 boson has mass $\frac{1}{2} \sqrt{g^2 + g'^2}\, v$, and the gauge boson A corresponding to the photon has mass 0.

These gauge bosons couple in different ways to fermions with different chirality. The kinetic energy of massless fermions $\bar{\psi}(i \not{\partial}) \psi$ decouples into

$$\bar{\psi}(i \not{\partial}) \psi = \bar{\psi}_{\mathrm{L}} (i \not{\partial}) \psi_{\mathrm{L}} + \bar{\psi}_{\mathrm{R}} (i \not{\partial}) \psi_{\mathrm{R}}$$

because this is how the gamma matrices work. This changes to

$$\bar{\psi}(i \not{D}) \psi = \bar{\psi}_{\mathrm{L}} (i \not{D}) \psi_{\mathrm{L}} + \bar{\psi}_{\mathrm{R}} (i \not{D}) \psi_{\mathrm{R}}$$

for the gauge theory of the weak interactions. One observes that the gauge bosons couple differently to the left-handed and the right-handed fermions. There is a degree of freedom not yet exploited and this is

that one can choose different representations of the same gauge group. This leads to different connection fields and also to different coupling to the boson and the fermion fields. The covariant derivative contains the gauge bosons which intern are described by the representation of the gauge group one is using. One can write down a locally gauge invariant theory which couples differently to left-handed and right-handed fields by choosing different representations of SU(2) in $\bar{\psi}_{\mathrm{L}}(i\not{D})\psi_{\mathrm{L}}$ and $\bar{\psi}_{\mathrm{R}}(i\not{D})\psi_{\mathrm{R}}$. That is as it seems the way nature has chosen to be.

There are many details such as how to deal with these different representations for the left-handed and the right-handed fermions which have not been covered. Also checking whether the symmetries at the classical level are maintained at the quantum level is a difficult task left out and quantum chromodynamics has been completely left aside. However, the objective was to outline the various stages of the discovery of the standard model of particle physics.

7.7 Some Closing Remarks

The best effective theory for all the experiments performed up to now is the standard model of particle physics together with linearized quantum gravity where linearized quantum gravity uses the metric

$$g^{\mu\nu} = \eta^{\mu\nu} + \delta g^{\mu\nu} \tag{7.4}$$

and perturbation theory but is not renormalizable. One takes linear quantum gravity, puts in some cutoff Λ and just ensures that predictions match normal gravity up to some experimental threshold. This gives a perfectly working effective theory of gravity.

However, one cannot get rid of the cutoff, and physicists believe that this combined effective theory of the standard model and linearized quantum gravity is not fundamental because black holes exist and cannot be described by it. Linearized quantum gravity tells a lot about low-energy gravitational processes, about scattering of gravitons and so on but as soon as one asks questions about high-energy processes near the cutoff as black holes require it collapses.

A last question asked here is why it is so hard to make quantum field theory rigorous despite attempts over decades to define quantum field theory axiomatically and otherwise rigorously. A surprising answer is that perturbation theory works so well – and it is the only tool available. It is actually not the only tool but it is the standard tool. One knows that perturbation theory is wrong and the reason is illustrated with an example.

In $0+1$ dimensions the action $S = x^2 + g\,x^4$ leads to the Gaussian looking path integral

$$Z = \int_{-\infty}^{\infty} dx\, e^{-x^2 - g\,x^4}$$

where one assumes that g is small and one uses perturbation theory. One gets lots of terms going to higher powers of g and the radius of convergence is 0. Looking at the integral then it should converge. With $g = 0$ it is a Gaussian integral and converges. With $g > 0$ the minus sign in front of the interaction term $-g\,x^4$ should make it more convergent but perturbation theory cannot make use of this information. A more complicated path integral

$$Z = \int \mathcal{D}\phi\, e^{-S}$$

with a Wick rotation has this structure and one thinks that it is beautifully convergent but one cannot calculate it. If one has a system where one thinks that it is well behaved and the only tool one has is not well behaved then that makes things difficult. Thus, quantum field theory is hard to make rigorous because perturbation theory as the standard tool physicists use is not the one that proves that the integral works.