Physics 504, Lectures 25-26
April 29 and May 4, 2010

1 Local Symmetry and Gauge Fields

Electromagnetism is the basic interaction which controls physics on all scales from atomic up to planetary, with myriad applications of great technological and scientific importance, and we have spent a year looking at some of these applications, all based on the simple foundational equations, Maxwell's Equations and the Lorentz force. But Electromagnetism also has clues to the elegant structure of local symmetry, and generalizing this symmetry from the very simple (and not obvious) symmetry of electromagnetism to larger symmetry groups is the basis of all of our modern theories of fundamental physics. That is, it leads the way to understand the theories which describe subatomic physics, and also somewhat in the direction of gravity! The clue electromagnetism gives towards this basic principle is in the gauge invariance, which says that substituting $A_\mu \rightarrow A_\mu - \partial_\mu \Lambda$ makes no change in the physics.

To understand what can come of this, or rather what to think is the origin of this invariance, we need to discuss several topics from outside classical electromagnetism.

- Lattice approach to field theories
- Internal field variables
- Global internal symmetries
- Quantum mechanics of a charged particle
- How to make global symmetries local

This we lead us to minimal substitution, covariant derivates, and understanding the electromagnetic field strengths as a form of curvature. It will also lead to non-Abelian gauge theories, which are the basis of QCD, our current model of the strong interactions, and Glashow-Salam-Weinberg $SU(2)\times U(1)$ electro-weak theory, which together form the *standard model.*

1.1 Lattice approach to field theories

When we discussed the lagrangian approach to electromagnetism, I assumed you were already familiar with the continuum formulation of dynamics of a field. Most of you had seen me develop this from the discrete dynamics where the degrees of freedom are defined on a lattice. These variables could be displacements, as in describing deformations of a solid, or they could be internal degrees of freedom, that is, not describing spatial degrees of freedom, as for example, the electric and magnetic fields (but, as we shall see, these require a more sophisticated latticization.)

The Lagrangian on the lattice whose continuum limit gives us a field theory depends on the degrees of freedom $\phi_{\vec{s}}$ defined on each site $\vec{s}$, and on differences of $\phi_{\vec{s}}$ on neighboring sites. Were there no such difference terms each site would have dynamics independent of its neighbors, and would not be a field theory. These difference terms become, in the continuum limit, the terms in $\mathcal{L}$ involving $\vec{\nabla}\phi$.

For simplicity of insight, we generally take our lattice to be cubic, with lattice sites given by a triple of integers times some lattice spacing. Time starts out being treated as a continuum, but for relativistic treatments we may discretize time as well (though then we are no longer doing lagrangian mechanics). In addition to the sites on the lattice, we may discuss links, joining two neighboring sites, and plaquettes, which are squares of four neighboring sites with the links between them.

1.2 Internal field variables

Let us begin by discussing possible degrees of freedom of a system, associated with points on a lattice or, in the continuum limit, with each point in space. This is how we have discussed the electromagnetic fields but not how we have discussed particles, where we have focussed on their motion through space. In today's discussion we will not be discussing motion through ordinary space, but rather degrees of freedom at fixed points in space.

The degrees of freedom ϕ take values in some domain. This could be the positive reals (*e.g.* temperature), the reals, the complex numbers (wave function ψ in quantum mechanics), a finite set (spin 1/2 ising model), or a vector, such as the electric field (again, not quite right).

Perhaps the most intuitive situation is a lattice of points upon which there are fixed objects with spin. So at each site j there is a dynamical variable[1] $\vec{S}_j$, and the spins can interact with each other, say with a Hamiltonian $H = J\sum_{nn} \vec{S}_i \cdot \vec{S}_j$, where the sum is over nearest neighbor sites. Notice that there is no coupling between the spins and vectors in ordinary space (no $(\vec{r}_j - \vec{r}_i) \cdot (\vec{S}_j - \vec{S}_i)$, for example), so the space in which the spins live can be considered completely independant of ordinary space, and in fact need not be a three-dimensional space at all. But as long as there are possible rotations in the space, there will be a symmetry in the theory, because if each spin is rotated by some orthogonal matrix $\mathbf{R}$, $\vec{S}_j \to \mathbf{R}\vec{S}_j$, each term $\vec{S}_i \cdot \vec{S}_j$ in the Hamiltonian in unchanged and a solution of the equations of motion will be transformed into another solution.

The same considerations hold if the dynamical degrees of freedom S_j are fields, taking independent (though not uncorrelated) values at each spatial point. That can be considered a continuum limit of the lattice version. Here the nearest-neighbor type coupling is likely to be replaced by a gradient, say $\sum_{\mu\alpha} \partial_\mu S_\alpha \partial^\mu S_\alpha$, which would keep our internal space and real space uncoupled. If the internal space is $\mathbb{R}^3$, then we might have $(\vec{\nabla} \cdot \vec{S})^2$ terms, and then space and spin would be coupled with an $\vec{L} \cdot \vec{S}$ term.

1.3 Global internal symmetries

If there is no coupling of the spins with space, the symmetry under rotation in spin space is called an *internal symmetry.* A famous example from nuclear physics is isotopic spin symmetry, where protons and neutrons are a doublet, treated like the two spin states of a spin 1/2 particle, but with the rotations having nothing to do with ordinary space, but occurring in isotopic spin space. Strong interactions are invariant or symmetric under such rotations. The lagrangian density depends on the nucleon wave function Ψ which has two complex components at every space-time point, and it is invariant if the wavefunction undergoes an isospin rotation $\Psi(x^\mu) \to e^{i\vec{\omega}\cdot\vec{I}}\Psi(x^\mu)$, where the three components of $\vec{I}$ are pauli spin matrices acting on the doublet Ψ. The isospin rotation has to be the same at all points in spacetime for this to be a symmetry. Thus it is a *global* symmetry.

[1] We are interested here in discussing the direction of $\vec{S}$, not the possible dynamics of its magnitude, so assume that $|\vec{S}_j|$ is fixed.

1.4 Quantum mechanics of a charged particle

One way to understand quantum mechanics is to think of momentum and energy are being differential operators $\vec{p} \to -i\hbar\vec{\nabla}$ and $E \to i\partial/\partial t$. Then a free nonrelativistic particle with $E = \vec{p}^{\,2}/2m$ is described by a wave function

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi$$

which is Schrödinger's equation for a free particle. Notice that the field ψ is, of necessity, complex. Also note that if $\psi(x^\mu)$ is a solution, so is $e^{i\lambda}\psi(x^\mu)$, as long as λ is a constant.

As we know from Ms. Noether, a continuous symmetry transformation of the fields gives rise to a conserved current. For our particle, the classical $\vec{J} = \rho\vec{v}$ becomes $-i\hbar q\psi^*\vec{\nabla}\psi$. Noether gives us a general procedure starting from the lagrangian density. A phase change gives $\Delta\phi = i\lambda\phi$, the current $J^\mu = -\frac{\partial\mathcal{L}}{\partial\partial_\mu\phi}\Delta\phi$ is just like the kinetic energy part of the lagrangian density with one derivative left out. For example, the Dirac Lagrangian $\mathcal{L} = i\hbar\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi$ gives a conserved current $J^\mu = \bar{\psi}\gamma^\mu\psi$, while for the Klein-Gordon lagrangian for a charged scalar field, $\mathcal{L} = \hbar^2(\partial_\mu\phi)(\partial^\mu\phi^*) - m^2\phi^*\phi$ gives $J_\mu = \frac{i\hbar}{2m}\left(\phi^*\partial_\mu\phi - \phi\partial_\mu\phi^*\right)$.

If the particle is charged and in the presence of an external field $A^\mu(x^\nu)$, this interaction can be incorporated by "minimal substitution", which is to say that $\vec{p} \to \vec{p} - q\vec{A}/c$, $E \to E - q\Phi$, so for a non-relativistic particle

$$i\hbar\frac{\partial\psi}{\partial t} = \left(q\Phi - \frac{\hbar^2}{2m}\left(\vec{\nabla} + iq\vec{A}/\hbar c\right)^2\right)\psi.$$

We recognize the $q\Phi$ as the potential energy term for the charged particle in an electrostatic field, though the $\vec{A}$ term, giving the interaction with the magnetic field (and a velocity dependent force) is not so familiar. If we had derived the Hamiltonian for a non-relativistic particle from the Lagrangian including the $-qU_\alpha A^\alpha$ interaction with an external field, we would have found this $(\vec{P} - qA/c)^2$ term as well, as I understand Schnetzer did do in 502. The important point, however, is that this equation does not remain invariant under gauge transformation $A_\mu \to A_\mu - \partial_\mu\Lambda$. As A_μ is entering only as an addition to the derivative operators, this gauge transformation's only effect is to add a piece $iq\partial_\mu\Lambda/c$ to each derivative operator, or $\partial_\mu \to e^{-iq\Lambda/c}\partial_\mu e^{iq\Lambda/c}$, so if ψ satisfies the equation in the original gauge, $\psi' = e^{-iq\Lambda/c}\psi$ saitisfies the equation in the transformed gauge. That is

$$\begin{aligned} A_\mu &\rightarrow A_\mu + \partial_\mu \Lambda \\ & \qquad (\Phi \rightarrow \Phi - \frac{1}{c}\frac{\partial \Lambda}{\partial t}, \qquad \vec{A} \rightarrow \vec{A} + \vec{\nabla}\Lambda) \\ \psi &\rightarrow e^{-iq\Lambda/c}\psi \end{aligned}$$

is an invariance of the theory, and is the correct form of a gauge transformation, which affects not only Φ and $\vec{A}$, but also the phase of the wave function.

Thus the gauge invariance is related to a symmetry of the field ψ under change of phase. Now a complex wave function can be viewed as a two-component wave function taking values in a two dimensional real space. What happens if we do a rotation in this two-dimensional space? That is equivalent to multiplying the wave function by a phase,

$$\psi(\vec{x}) \rightarrow e^{i\alpha}\psi(\vec{x}).$$

or we may think of this as a change in basis used to describe the components of this vector in this two dimensional space. If the wave function is given a constant phase shift everywhere in spacetime, this corresponds to the same change in basis at all points, and there is no change in any physics, and we are using a global symmetry. But we see here that we have a much stronger symmetry — we can change the phase independently at each point in spacetime, provided we simultaneously make the corresponding gauge change in the vector potential.

Clearly the various equations of motion are invariant under this, and the Hamiltonian and currents are unchanged. So we have a symmetry of rotations in a two dimensional space. Note there is no inherent meaning in distinguishing the real part of the wave function.

Now without the gauge fields we have a symmetry which only holds if the **same** phase is applied at all points, which was also true for our lattice of spins. Such a symmetry is called a *global symmetry.* If we use different basis vectors to describe the spins at different points, the value of $\sum_\alpha S_{i\,\alpha} S_{j\,\alpha}$ will depend on our choices of basis for i and j, and not be invariant. But our electromagnetic field permits a much larger, *local* symmetry.

Indeed, in a relativistic theory, why should what we choose at one point in space depend on what we chose at another? Can we make a theory which is invariant under independent choices of the coordinates at each point? Electromagnetism does this for us for our complex field, but what about for spins?

1.5 Latticizing Quantum Field Theory

In Quantum Field theory, perturbative approaches starting with free particles work well if the coupling is weak, although even in that case there are complications arising from intermediate states which correspond to very short distances between interactions, known as ultraviolet divergences. One way to attempt to understand these issues is to consider a discretized theory of degrees of freedom on a lattice, and then investigate the continuum limit. This gives insight into regularization (avoiding the infinities from ultraviolet divergence) and renormalization, both in elementary particle theories and in phase transitions in condensed matter.

We are going to consider field theory on a lattice for a different reason — because it helps to clarify the fundamental idea of gauge fields.

2 Symmetry

Consider a theory which involves a set of N real fields $\phi_i(x^\mu)$ which have an internal symmetry group[2] $\mathcal{G}$ under which they transform with a representation M, so that a particular symmetry transformation $G \in \mathcal{G}$ acts on the ϕ fields by

$$G : \phi_i(x) \mapsto \phi'_i(x) = \sum_j M_{ij}(G)\phi_j(x). \tag{1}$$

If it is a symmetry, the Lagrangian must be invariant. If the kinetic term is of the usual form, $\frac{1}{2}\sum_{\mu,i}\partial_\mu\phi_i\partial^\mu\phi_i$, invariance requires that M is an orthogonal matrix $\sum_k M_{ki}M_{kj} = \delta_{ij}$. That condition also insures the invariance of the mass term $-\frac{1}{2}m^2\sum_i\phi_i^2$, and of any other "potential" term $V(\sum_i\phi_i^2)$ depending only on the "length" of ϕ. Provided V has that form, we see that the theory should be invariant under all the orthogonal transformations (1).

[2]The notation is not completely standard. Many books would use G for the group, $\mathcal{G}$ for the Lie algebra of the group, and g for an element of G. Because we are going to use g as the analogue of the fundamental charge, I am using G for a group element, $\mathcal{G}$ for the group, and $\mathfrak{G}$ for the Lie algebra, elements of which will be called $\mathcal{A}$.

We see that the individual components ϕ_i are only projections along the unit vectors of an arbitrary orthonormal basis of $\mathbb{R}^N$, and do not have separate intrinsic physical meanings. Alternatively, V might not be invariant under all of O(N), but only under the subgroup[3] $\mathcal{G}$. For example, one important group is the SU(3) of colors which act on each triplet (in color) of quarks. Replacing the 3 complex quark fields by 6 real fields, the kinetic term would be invariant under the group O(6) $\sim$ SU(4), but the interaction terms are only invariant under the subgroup SU(3).

So we are going to be considering a symmetry group[4] $\mathcal{G}$ which has generators L_b which form a basis of the "Lie algebra" $\mathfrak{G}$ of the group[5]. As we saw for the Lorentz group, the Lie algebra for SO(N) is the set of antisymmetric real $N \times N$ matrices, with $\frac{1}{2}N(N-1)$ independent generators $\tilde{L}$, or, for physicists, $\frac{1}{2}N(N-1)$ purely imaginary antisymmetric $N \times N$ matrices. For SU(3), the generators may be thought of as traceless hermitean 3×3 matrices.

2.1 Discretization

How might we approximate the continuum theory on a lattice? Instead of $\phi_i(\mathbf{x})$ defined for all values of $\mathbf{x} \in \mathbb{R}^4$, we might have $\phi_i(\vec{n})$ discrete variables defined only for integer values $\vec{n} \in \mathbb{Z}^4$, representing a lattice in space-time with lattice spacing a, with $\mathbf{x}^\mu = an^\mu$. The mass term in the action

$$-\frac{1}{2}\int d^4x \sum_i \phi_i^2(\mathbf{x}) \to -\frac{1}{2}a^4 \sum_{\vec{n}\in\mathbb{Z}^4}\sum_i \phi_i^2(\vec{n}).$$

[3]More precisely, the image of $\mathcal{G}$ under the representation $M : \mathcal{G} \to N \times N$ matrices is a subgroup of O(n).

[4]We will only consider connected groups which are either Abelian or semisimple, or products of such groups.

[5]Here is what we will need to know about groups and Lie algebras: The algebra can be represented by generators L_a which satisfy $[L_a, L_b] = i\sum_k c_{ab}{}^k L_k$, with $c_{ab}{}^k$ real numbers known as the *structure constants* of the group. These give a bilinear *Killing form* $\beta : \mathfrak{G} \times \mathfrak{G} \to \mathbb{R}$ given by $\beta(L_i, L_j) = -\sum_{ab} c_{ai}{}^b c_{bj}{}^a$. As this is a real symmetric matrix, it can be diagonalized. For compact, semisimple groups, all the eigenvalues are positive, the L_i's can be scaled, so that $\beta_{ij} = 2\delta_{ij}$. Then the basis has be chosen such that $\sum_k L_k^2$ is a casimir operator, commuting with each of the L_a's, and it can be shown that the structure constants are totally antisymmetric. This should be familiar for the rotation group, and is explained in more detail in "Lightning review of group" and "Notes on Representations, the Adjoint rep, the Killing form, and antisymmetry of $c_{ij}{}^k$".

For the kinetic energy term we need to replace a derivative by a finite difference. The simplest substitution is to replace

$$\partial_\mu \phi_i(\mathbf{x}) \to \frac{1}{a}\left(\phi_i(\vec{n} + \vec{\Delta}_\mu) - \phi_i(\vec{n})\right),$$

where $\vec{\Delta}_\mu$ is 1 in the μ direction and 0 in the others. Here the relation of x^μ and $\vec{n}$ is $x^\nu = an^\nu + \frac{1}{2}a\delta^\nu_\mu$, representing most accurately the x in the middle of the two lattice points. If we expand out the squares of the differences, we get terms which look just like the mass terms, but also nearest neighbor couplings $\sum_i \phi_i(\vec{n} + \vec{\Delta}_\mu)\phi_i(\vec{n})$.

Each of these contributions to the action is still invariant under the transformation (1), providing we use the same group transformation at every point in space-time. This is called a global symmetry.

In a relativistic field theory, all information is local, because information can only travel at the speed of light. So we might ask, if the theory is unchanged by a group action at one point, why should that depend on having the same transformation at every other point? In other words, could we have a **local** symmetry, in which equation (1) holds with the group element varying from one point of space-time to another? The mass terms and other terms in $V(\phi)$ only depend on one point, so they don't care whether M varies, and they are invariant under such transformations. But the nearest-neighbor coupling

$$\sum_i \phi_i(\vec{n} + \vec{\Delta}_\mu)\phi_i(\vec{n}) \to M_{ik}(G(\vec{n} + \vec{\Delta}_\mu))M_{ij}(G(\vec{n}))\,\phi_k(\vec{n} + \vec{\Delta}_\mu)\phi_j(\vec{n})$$

is not invariant because

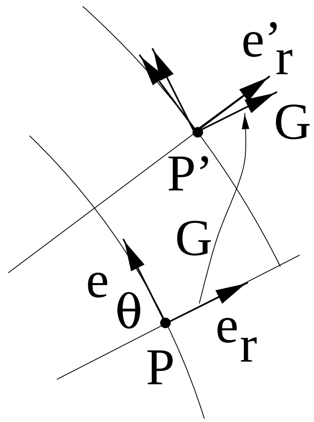
$$M^{-1}(G(\vec{n} + \vec{\Delta}_\mu))M(G(\vec{n})) \neq 1$$

if the G's (and hence the M's) vary from point to point.

2.2 Parallel Transport

The problem is that we have a term in the Lagrangian that is a function of how ϕ changes from point to point, but we measure that change by how much the components change. That is only correct if the basis for comparing the ϕ's does not change. We must have a way to measure change from point to point, but before we can subtract one ϕ vector from another at a different point, we must "parallel transport" it to that new point. That is, for each link between neighboring points, we must have a rule for parallel transporting ϕ fields from one end of the link to the other. The change in the field $\boldsymbol{\phi} = \sum_a \phi^a \hat{e}_a$ as we go from point A to point B is equal to $\Delta\boldsymbol{\phi} = \sum_a (\phi^a_B - \phi^a_A)\hat{e}_a$ *only* if we can assume that the basis vectors don't change, $\hat{e}^A_a = \hat{e}^B_a$. If we allow for the possibility that the basis we have chosen at the point $\vec{n} + \vec{\Delta}_\mu$ differs by a group element G from that which corresponds to parallel transport from $\vec{n}$, we get a more elaborate definition of $\Delta\boldsymbol{\phi}$.

As an example, it might help to think of an ordinary vector in the plane, expressed in polar coordinates. Consider the unit basis vectors $\vec{e}_r$ and $\vec{e}_\theta$ at the point P. If we transport $\vec{e}_r$ to the point P' while keeping it parallel to what it was, we arrive at the vector labelled Ge_r, which is not the same as the unit radial vector e'_r at the point P'.



Note that if we have a vector $\vec{V}\,' = V'_r\vec{e}\,'_r + V'_\theta\vec{e}\,'_\theta$ at P' which is unchanged (parallel transported) from the vector $\vec{V} = V_r\vec{e}_r + V_\theta\vec{e}_\theta$ at P, we **do not have** $V'_r = V_r$.

Now in our example we had an *a priori* rule for what parallel transport means, but if we are to allow local gauge invariance, this rule becomes a new degree of freedom. This dynamical variable is actually one element of the symmetry group (and therefore perhaps several degrees of freedom, $\frac{1}{2}N(N-1)$ for $SO(N)$, the orthogonal transformations in N dimensions), for each point on the lattice and each direction we might parallel transport $\boldsymbol{\phi}$. We can then build a theory with a local symmetry, but at the expense of introducing a lot of new degrees of freedom.

The theory that emerges from these consideration is a **gauge field theory**. Its degrees of freedom include not only the "matter fields" at each site of the lattice, but also "gauge fields" on each link between nearest neighbors. The matter fields live in a vector space which transforms linearly as a representation[6] of the "gauge group" $\mathcal{G}$. The gauge fields live in the group itself,

[6] To a physicist, the vector space in which the matter fields live is called the representation, but what mathematicians call a representation consists of the matrices M_{ij}, or more accurately the mapping from elements of the group into matrices, $G \mapsto M(G)$. What we call a representation they call a module.

at least in the lattice field theory, but may alternately be considered to take values in the Lie algebra of generators of the group, especially if we are to take the continuum limit of the lattice.

2.3 Covariant Derivative

When a group element G acts on a vector $\vec{V} = \sum_i V_i \hat{e}_i$ which transforms under a representation M, the components of the new vector are multiplied by the matrix:

$$G : \vec{V} \to \vec{V}\,' = \sum_{ij} M_{ij}(G) V_j \hat{e}_i, \quad \text{so } V'_i = \sum_j M_{ij}(G) V_j.$$

So if G parallel transports $\vec{\phi}(\vec{n})$ from $\vec{n}$ to $\vec{n} + \vec{\Delta}_\mu$, and if we subtract this from $\vec{\phi}(\vec{n} + \vec{\Delta}_\mu)$ to get the change in ϕ, we have

$$\Delta\phi = \sum_i \left[\phi_i(\vec{n} + \vec{\Delta}_\mu) - \sum_j M_{ij}(G)\phi_j(\vec{n}) \right] \hat{e}_i.$$

If the fields are slowly varying over the distance of one lattice spacing, which is necessary if we are to consider the lattice an approximation to the continuum, we can approximate

$$\phi_i(\vec{n} + \vec{\Delta}_\mu) \approx \phi_i(\vec{n}) + a\partial_\mu \phi_i.$$

We can also assume that the group transformation that parallel transports by one lattice spacing is close to the identity, and that the Lie algebra element which generates it should be proportional to the lattice spacing a. Thus we may write $G = e^{iag\mathcal{A}}$, $M(G) = M(e^{iag\mathcal{A}}) \approx 1 + iagM(\mathcal{A})$, where $\mathcal{A}$ is an element in the Lie algebra $\mathfrak{G}$ of the gauge group $\mathcal{G}$. [We have added a parameter g which will turn out to be the fundamental charge, in order to get conventionally defined $\mathcal{A}$ fields, although sometimes that is not done, and the scale for measuring $\mathcal{A}$ is the natural one for the group.] Then we find, to first order in the lattice spacing a,

$$\Delta\phi_i = a\left(\partial_\mu \phi_i - ig \sum_j M_{ij}(\mathcal{A})\phi_j \right).$$

In the continuum limit, we define $1/a$ times this to be the **covariant derivative**, but first I must say a few words about the gauge field $\mathcal{A}$. First, as there is a different value on each link, and in the continuum limit there are four[7] links radiating from each point, we need to be defining four fields $\mathcal{A}_\mu(\mathbf{x})$. Also, each $\mathcal{A}_\mu$ is not a single field, in general, but an element of the Lie algebra, which is a vector space. The Lie algebra for the rotation group, for example, is parameterized by a vector with three components, $\vec{\omega}$. Rotations themselves are not a gauge group, but one possible gauge group to consider is the $SU(2)$ of the electro-weak theory, which is isomorphic[8] to the rotation group. One usually uses L_i to represent a basis vector of the Lie algebra vector space, so the gauge field can be expanded as

$$\mathcal{A}_\mu(\mathbf{x}) = \sum_b A_\mu^{(b)}(\mathbf{x}) L_b.$$

This brings us to the definition of the **covariant derivative**:

$$(D_\mu\phi)_j = \partial_\mu\phi_j - ig\sum_{kb} A_\mu^{(b)} M_{jk}(L_b)\phi_k\,; \qquad D_\mu\phi = \partial_\mu\phi - igA_\mu^{(b)} M(L_b)\phi,$$

where on the right we have written the expression with implied summations on matrix and vector indices and implied multiplication.

2.4 Gauge Transformations

What does this have to do with local symmetry? We saw that the transformation (1), where we let G vary with x, is a symmetry for the lattice terms involving only a single site, but *not* for the kinetic term, $(\partial\phi)^2$, which involves cross terms such as $\sum_i \phi_i(\vec{n}+\vec{\Delta}_\mu)\phi_i(\vec{n})$. These couple neighboring points, and are not invariant. But with our improved definition of $(\Delta\phi)$, the cross terms now have the form

$$\phi(\vec{n}+\vec{\Delta}_\mu)\cdot M(G_L)\cdot\phi(\vec{n}),$$

where G_L is the group transformation associated with the link $(\vec{n},\vec{n}+\vec{\Delta}_\mu)$ that implements a parallel transport.

We can now ask what happens under the transformation in a different way. If we think of the gauge transformation $G(\mathbf{x})$ in the passive language

[7] Actually there are eight, as there are forward and backwards links in each direction. But the "backwards" ones can be thought of as belonging to "previous" sites.

[8] Not exactly: the Lie algebra of $SU(2)$ is the same as the Lie algebra of the three dimensional rotation group $SO(3)$, but the actual groups differ, as is discussed when considering how spinors transform under rotations of 2π.

as a change in the basis elements for the matter fields, we realize that they will also effect the rule for doing parallel transport. If G_L was the group transformation on the basis which did a parallel transport from site p to site q, with link L going from p to q, then after a change of basis by G_p at p and one by G_q at q, the way to parallel transport the new basis at p must be $G_L' = G_q G_L G_p^{-1}$. So we now define the gauge transformation Λ, which is specified by a group element at each lattice site

$$\Lambda : \begin{cases} \phi(x_p) \to M(G_p)\cdot\phi(x_p) \\ \phi(x_q) \to M(G_q)\cdot\phi(x_q) \\ \quad G_L \to G_q G_L G_p^{-1} \end{cases}$$

This gauge transformation is a **local symmetry** of the gauge field theory. Let's verify that this is an invariance of the nearest neighbor term:

$$\begin{aligned} \phi(x_q)\cdot M(G_L)\cdot\phi(x_p) &= \phi_i(x_q)M_{ij}(G_L)\phi_j(x_p) \\ &\to M_{ik}(G_q)\phi_k(x_q)M_{ij}(G_qG_LG_p^{-1})M_{j\ell}(G_p)\phi_\ell(x_p) \\ &\quad = \phi_k(x_q)M^{-1}_{ki}(G_q)M_{ij}(G_qG_LG_p^{-1})M_{j\ell}(G_p)\phi_\ell(x_p) \\ &\quad = \phi_k(x_q)M_{k\ell}(G_L)\phi_\ell(x_p) = \phi(x_q)\cdot M(G_L)\cdot\phi(x_p), \end{aligned}$$

where we have used the orthogonality of $M(G_q)$ and the fact that the M's are a representation, and therefore $M^{-1}_{ki}(G_q)M_{ij}(G_qG_LG_p^{-1})M_{j\ell}(G_p) = M_{k\ell}(G_L)$.

In a continuum field theory, we consider only local gauge transformations where the group element varies differentially in the continuum limit. We may think of Λ as given by a Lie-algebra valued scalar field $\lambda(\mathbf{x}) = \sum_b \lambda^{(b)}(\mathbf{x})L_b$. Then the matter fields transform as

$$\phi(\mathbf{x}) \to \phi'(\mathbf{x}) = e^{i\sum_b \lambda^{(b)}(\mathbf{x})M(L_b)}\phi(\mathbf{x}),$$

while the gauge field itself transforms by

$$A_\mu^{(b)}(\mathbf{x}) \to A_\mu'^{(b)}(\mathbf{x}),$$

with

$$e^{iagA_\mu'^{(b)}(\mathbf{x})} = e^{i\lambda(\mathbf{x}+\frac{1}{2}a\Delta_\mu)}e^{iagA_\mu^{(b)}(\mathbf{x})L_b}e^{-i\lambda(\mathbf{x}-\frac{1}{2}a\Delta_\mu)}. \tag{2}$$

We have placed x at the middle of the link. We now expand to first order in the lattice spacing, remembering that $\lambda(\mathbf{x})$ and $\partial_\mu\lambda(\mathbf{x})$ may not commute.

So we will expand the exponential rather than λ. Approximating

$$\begin{aligned} e^{iag\mathcal{A}_\mu} &\rightarrow 1 + iag\mathcal{A}_\mu, \\ e^{i\lambda(\mathbf{x}\pm\frac{1}{2}a\Delta\mu)} &\rightarrow e^{i\lambda(\mathbf{x})} \pm \frac{1}{2}a\partial_\mu\left[e^{i\lambda(\mathbf{x})}\right], \end{aligned}$$

and plugging these into (2), we get

$$\begin{aligned} 1 + iag\mathcal{A}'_\mu &= \left(e^{i\lambda} + \frac{1}{2}a\partial_\mu e^{i\lambda}\right)(1 + iag\mathcal{A}_\mu)\left(e^{-i\lambda} - \frac{1}{2}a\partial_\mu e^{-i\lambda}\right) \\ &= 1 + iage^{i\lambda}\mathcal{A}_\mu e^{-i\lambda} + \frac{1}{2}a\left(\partial_\mu e^{i\lambda}\right)e^{-i\lambda} - \frac{1}{2}ae^{i\lambda}\left(\partial_\mu e^{-i\lambda}\right) \end{aligned}$$

Note from $\partial_\mu\left(e^{i\lambda}e^{-i\lambda}\right) = 0$ that the third and fourth terms are equal, so we can drop the third and double the fourth, to get

$$\begin{aligned} \mathcal{A}'_\mu &= e^{i\lambda}\mathcal{A}_\mu e^{-i\lambda} + \frac{i}{g}e^{i\lambda}\partial_\mu e^{-i\lambda} \\ &= e^{i\lambda}\left(\mathcal{A}_\mu + \frac{i}{g}\partial_\mu\right)e^{-i\lambda} \end{aligned}$$

Let us now ask how this is related to the gauge transformations we know from Maxwell's theory, which look less complicated. Electromagnetism is a gauge field, but one with a very simple gauge group, that of rotations about a single fixed axis[9]. The group consists of $\mathcal{G} = \{e^{i\theta L_1}\}$ and the Lie algebra has only one generator, L_1, and is therefore isomorphic to the real line $\mathbb{R}$, and the single structure constant ${c_{11}}^1$ is zero (a counterexample to assuming that the Killing form can always be set to $2 \times \mathbb{1}$). The rotations act on charged fields, which are usually represented by complex fields Φ but in our treatment here are represented by a doublet of real fields, $(\phi_1, \phi_2) = (\text{ Re }\Phi, \text{ Im }\Phi)$. The transformation

$$\phi \rightarrow \phi' = \begin{pmatrix} \text{Re }\Phi' \\ \text{Im }\Phi' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \text{Re }\Phi \\ \text{Im }\Phi \end{pmatrix}$$

gives $\Phi' = e^{i\theta}\Phi$, so the gauge transformations are **local** changes in phase of the charged fields. The gauge transformations of fields themselves is vastly

[9] These are not rotations in real space, but in some abstract space of field configurations. For QED that abstract space was represented by complex numbers, and the rotation is simply multiplication by $e^{i\theta}$ for a real phase θ.

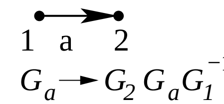
simplified by the fact that all the terms commute, so

$$A'_\mu = e^{i\lambda}\left(A_\mu + \frac{i}{g}\partial_\mu\right)e^{-i\lambda} = A_\mu + g^{-1}\partial_\mu\lambda.$$

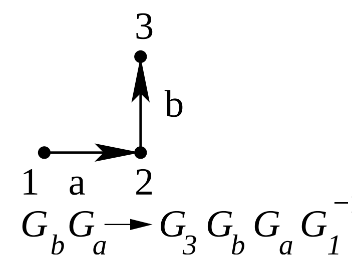
But this simplicity only holds for an **Abelian** group, one where all the generators commute, which is not enough when we wish to consider the gauge theories of the electroweak and strong interactions.

2.5 Pure Gauge Terms in $\mathcal{L}$

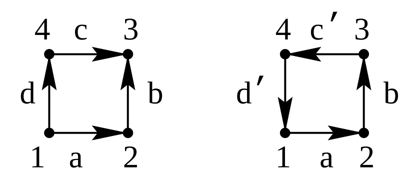
We now know how the kinetic terms for charged fields are modified by the presence of an external gauge field, but we have not yet discussed the terms which propagate the gauge fields themselves. We need these terms in the Lagrangian to be invariant under gauge transformations. In particular this means that they cannot depend only on a single link, because we can always make a gauge transformation $G_1 = G_a$ which resets the group element for a single link to 1, so there would be no dependence on the field. In fact, the simplest way to get rid of the gauge dependance of $G_a = e^{iag\mathcal{A}_x(\mathbf{x}_a)}$ on $G_2 = e^{i\lambda(\mathbf{x}_2)}$ is to premultiply it by G_b,



$$G_bG_a \rightarrow G_3G_bG_2^{-1}G_2G_aG_1^{-1} = G_3G_bG_aG_1^{-1}.$$



There is still a gauge dependence on the endpoints of the path, however, so the best thing to do is close the path. To do so, we are traversing some links backwards from the way they were defined, but from that definition in terms of parallel transport it is clear that the group element associated with taking a link backwards is the inverse of the element taken going forwards. So the group element associated with the closed path on the right (which is called a plaquette) is $G_P = G_d^{-1}G_c^{-1}G_bG_a$, which transforms under gauge transformations as

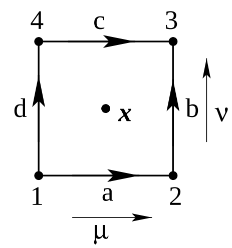


$$G_P \rightarrow G'_P = \left(G_4G_dG_1^{-1}\right)^{-1}\left(G_3G_cG_4^{-1}\right)^{-1}G_3G_bG_2^{-1}G_2G_aG_1^{-1}$$

$$\begin{aligned} &= G_1 G_d^{-1} G_4^{-1} G_4 G_c^{-1} G_3^{-1} G_3 G_b G_2^{-1} G_2 G_a G_1^{-1} \\ &= G_1 G_d^{-1} G_c^{-1} G_b G_a G_1^{-1} \\ &= G_1 G_P G_1^{-1}. \end{aligned}$$

So the plaquette group element is not invariant but it does have a simpler and more restricted variation. In the continuum limit we expect each link's group element to be near the identity and also to have G_c differ from G_a by something proportional to the lattice spacing, so G_P should be close to the identity, the difference considered a generator in the Lie algebra. The Killing form acting on that generator will provide us with an invariant. Let us define $\mathcal{F}_{\mu\nu} = -ia^{-2}g^{-1}(G_P - 1)$ to be the field-strength tensor, where μ and ν are the directions of links a and b respectively. Let us take $\mathbf{x}$ in the center of the placquette. Expanding each link to order $\mathcal{O}(a^2)$

$$\begin{aligned} G_a &\approx 1 + iag\mathcal{A}_\mu(\mathbf{x} - \frac{1}{2}a\Delta_\nu) - \frac{1}{2}a^2 g^2 \mathcal{A}_\mu^2(\mathbf{x} - \frac{1}{2}a\Delta_\nu) \\ &\approx 1 + iag\mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}ia^2 g \partial_\nu \mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}a^2 g^2 \mathcal{A}_\mu^2(\mathbf{x}) \\ G_c^{-1} &\approx 1 - iag\mathcal{A}_\mu(\mathbf{x} + \frac{1}{2}a\Delta_\nu) - \frac{1}{2}a^2 g^2 \mathcal{A}_\mu^2(\mathbf{x} + \frac{1}{2}a\Delta_\nu) \\ &\approx 1 - iag\mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}ia^2 g \partial_\nu \mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}a^2 g^2 \mathcal{A}_\mu^2(\mathbf{x}), \end{aligned}$$



we have, to second order[10] in a,

$$\begin{aligned} G_P = &\left(1 - iag\mathcal{A}_\nu(\mathbf{x}) + \frac{1}{2}ia^2 g \partial_\mu \mathcal{A}_\nu(\mathbf{x}) - \frac{1}{2}a^2 g^2 \mathcal{A}_\nu^2(\mathbf{x})\right) \\ &\left(1 - iag\mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}ia^2 g \partial_\nu \mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}a^2 g^2 \mathcal{A}_\mu^2(\mathbf{x})\right) \\ &\left(1 + iag\mathcal{A}_\nu(\mathbf{x}) + \frac{1}{2}ia^2 g \partial_\mu \mathcal{A}_\nu(\mathbf{x}) - \frac{1}{2}a^2 g^2 \mathcal{A}_\nu^2(\mathbf{x})\right) \\ &\left(1 + iag\mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}ia^2 g \partial_\nu \mathcal{A}_\mu(\mathbf{x}) - \frac{1}{2}a^2 g^2 \mathcal{A}_\mu^2(\mathbf{x})\right) \\ = &\ 1 + a^2 g \left\{ g \left[\mathcal{A}_\mu(\mathbf{x}), \mathcal{A}_\nu(\mathbf{x})\right] + i\partial_\mu \mathcal{A}_\nu(\mathbf{x}) - i\partial_\nu \mathcal{A}_\mu(\mathbf{x}) \right\} \end{aligned}$$

Thus

$$\mathcal{F}_{\mu\nu}(\mathbf{x}) = \partial_\mu \mathcal{A}_\nu(\mathbf{x}) - \partial_\nu \mathcal{A}_\mu(\mathbf{x}) - ig\left[\mathcal{A}_\mu(\mathbf{x}), \mathcal{A}_\nu(\mathbf{x})\right].$$

[10] Note that the terms in $\mathcal{A}_\mu^2(\mathbf{x})$ cancel, and only the commutator, not the product, of L_a's is left.

Note that $\mathcal{F}_{\mu\nu}$ is

- a Lie-algebra valued field, $\mathcal{F}_{\mu\nu}(\mathbf{x}) = \sum_b F_{\mu\nu}^{(b)}(\mathbf{x}) L_b$.
- An antisymmetric tensor, $\mathcal{F}_{\mu\nu}(\mathbf{x}) = -\mathcal{F}_{\nu\mu}(\mathbf{x})$.
- Because the Lie algebra is defined in terms of the structure constants, ${c_{ab}}^d$ by

  $$[L_a, L_b] = i{c_{ab}}^d L_d,$$

  the field-strength tensor may also be written

  $$F_{\mu\nu}^{(d)} = \partial_\mu A_\nu^{(d)} - \partial_\nu A_\mu^{(d)} + g{c_{ab}}^d A_\mu^{(a)} A_\nu^{(b)}.$$

Before we turn to the Lagrangian, let me point out a crucial relationship between the covariant derivatives and the field-strength. If we take the commutator of covariant derivatives

$$D_\mu = \partial_\mu - igA_\mu^{(b)} L_b$$

at the same point but in different directions,

$$\begin{aligned} [D_\mu, D_\nu] &= [\partial_\mu - ig\mathcal{A}_\mu, \partial_\nu - ig\mathcal{A}_\nu] = -ig\partial_\mu \mathcal{A}_\nu - g^2 \mathcal{A}_\mu \mathcal{A}_\nu - (\mu \leftrightarrow \nu) \\ &= -g^2 \left[\mathcal{A}_\mu, \mathcal{A}_\nu\right] - ig\partial_\mu \mathcal{A}_\nu + ig\partial_\nu \mathcal{A}_\mu \\ &= -ig\mathcal{F}_{\mu\nu}. \end{aligned}$$

Notice that although the covariant derivative is in part a differential operator, the commutator has neither first or second derivatives left over to act on whatever appears to the right. It does need to be interpreted, however, as specifying a representation matrix that will act on whatever is to the right.

Now consider adding to the Lagrangian a term proportional to the Killing form evaluated on $\mathcal{F}$, twice, $\beta(\mathcal{F}_{\mu\nu}, \mathcal{F}^{\mu\nu}) = 2\sum_b F_{\mu\nu}^{(b)} F^{(b)\,\mu\nu}$. I have assumed the generators L_a have been normalized so that the Killing form $\beta(L_a, L_b) = 2\delta_{ab}$, and the stucture constants are totally antisymmetric[11]. We know that under a gauge transformation $\mathcal{F}_{\mu\nu} \to e^{i\lambda}\mathcal{F}_{\mu\nu}e^{-i\lambda}$. If λ is infinitesimal, $\mathcal{F}_{\mu\nu} \to \mathcal{F}_{\mu\nu} + i[\lambda, \mathcal{F}_{\mu\nu}] = \left\{F_{\mu\nu}^{(d)} - \lambda^{(a)} F_{\mu\nu}^{(b)} {c_{ab}}^d\right\} L_d$, so

$$\delta\beta(\mathcal{F}_{\mu\nu}, \mathcal{F}^{\mu\nu}) = 2\beta(\delta\mathcal{F}_{\mu\nu}, \mathcal{F}^{\mu\nu}) = 2 \times (-2)\lambda^{(a)} F_{\mu\nu}^{(b)} {c_{ab}}^d F^{(d)\,\mu\nu} = 0$$

[11] See groups.pdf and adjnote.pdf in the Supplementary Notes section of the Lecture Notes page of the course website.

where the expression vanishes because ${c_{ab}}^d$ is antisymmetric under interchange of b and d but $F^{(b)}_{\mu\nu} F^{(d)\,\mu\nu}$ is symmetric under the same interchange (and we are summing on b and d). As $\beta(\mathcal{F}_{\mu\nu}, \mathcal{F}^{\mu\nu})$ doesn't change to first order under infinitesimal transformations, it also doesn't change under the finite transformations they generate.

2.6 Lagrangian for the Gauge Fields

We choose the normalization of the A fields so that the pure gauge term in the Lagrangian density is $-\frac{1}{4} F^{(b)}_{\mu\nu} F^{(b)\,\mu\nu}$. Suppose we also have Dirac matter fields transforming under a representation $t^b_{ij} = M_{ij}(L_b)$ of the group, and perhaps some scalar fields as well, transforming under a (possibly) different representation $\bar{t}^b_{ij} = \bar{M}_{ij}(L_b)$, where the bars here only represent a different representation, not any kind of conjugation. The gauge fields come into the matter terms in the Lagrangian because, in order to maintain local gauge invariance, all derivatives need to be replaced by covariant derivatives. Thus the potential terms for matter fields in $\mathcal{L}$ will not be involved in the equations of motion of the gauge fields, and we need only look at

$$\begin{aligned} \mathcal{L} &= -\frac{1}{4} F^{(b)}_{\mu\nu} F^{(b)\,\mu\nu} + i\bar{\psi}\gamma^\mu \left(\partial_\mu - igA^{(b)}_\mu t^b\right)\psi \\ &\quad + \frac{1}{2}\left[\left(\partial_\mu - igA^{(b)}_\mu \bar{t}^b\right)\phi\right]^T \left[\left(\partial^\mu - igA^{(b)\,\mu}\bar{t}^b\right)\phi\right]. \end{aligned}$$

The theory we have just defined, the gauge theory based on a non-Abelian Lie group, is known as Yang-Mills theory.