

CONSTRAINED HAMILTONIAN SYSTEMS

Introduction - Second Class Constraints - First Class Constraints - Electrodynamics - Observables - Canonical Gauge Fixing¹

Introduction

The Hamiltonian formulation of gauge theories is due to Dirac, who described it brilliantly in the first two chapters of *Lectures on Quantum Mechanics*. These notes are intended as a supplement to Dirac's lectures. As is customary, the general theory will be presented for theories whose phase spaces are finite dimensional. However, it turns out that most of the naturally occurring examples of Dirac's theory arise in relativistic mechanics, which means—among other things—that our examples will be field theories. I will begin with a comparison of scalar field theory and the field theory of electromagnetism, to see what the problems are. The Klein-Gordon equation is

$$\square\varphi(x) = m^2\varphi(x) . \quad (1)$$

This is a hyperbolic equation. To solve it in a general situation, one first specifies the initial data

$$\varphi(0, \mathbf{x}) \quad \dot{\varphi}(0, \mathbf{x}) \quad (2)$$

on some spacelike hypersurface defined by $t = 0$. This can be done arbitrarily, subject only to some smoothness requirements. Then—as shown by Cauchy, Kowalewskaya, Hadamard and others—the Klein-Gordon equation suffices to determine $\varphi(x)$ everywhere in space-time. Essentially, this happens because the equation can be used to compute

$$\ddot{\varphi} = \ddot{\varphi}(\varphi, \dot{\varphi}) . \quad (3)$$

By iteration, we can compute the time derivatives to all orders, and construct the function everywhere under the assumption that it is analytic. Further analysis reveals that the assumption of analytic data can be dispensed with, and indeed one can prove that smooth initial data determine a unique solution. Moreover the latter depends continuously on these data. It is however non-trivial to go beyond analytic data—a corresponding analysis of the Laplace equation would give a different result.

The Hamiltonian formulation of the Klein-Gordon equation is closely tied to the initial value problem, and poses no problems. In phase space, the variables that are used to describe the theory are just φ itself, together with

$$\pi \equiv \dot{\varphi} . \quad (4)$$

¹These notes, from a 1995 relativity course, were gently modified in 2012.

The Poisson brackets and the Hamiltonian are

$$\{\varphi(x), \pi(y)\} = \delta(x, y) \quad (5)$$

$$H = \frac{1}{2} \int d^3x (\pi^2 + \partial_a \varphi \partial_a \varphi + m^2 \varphi^2) . \quad (6)$$

It is easy to check that Hamilton's equations

$$\dot{\varphi} = \{\varphi, H\} \quad \dot{\pi} = \{\pi, H\} \quad (7)$$

are equivalent to the original Klein-Gordon equation.

As an aside, we introduce a piece of notation that is often used in both classical and quantum field theory, namely smearing with test functions $f(x), g(x), \dots$. If we define

$$\varphi[f] \equiv \int d^3x f(x) \varphi(x) \quad (8)$$

and so on, we can evidently rewrite the Poisson brackets as

$$\{\varphi[f], \pi[g]\} = f[g] = g[f] . \quad (9)$$

This is often a convenient notation, especially when one wants to keep track of partial integrations in a calculation. Also, both in classical and quantum field theory there are many formal calculations that may or may not be valid, depending on—in the classical case—the behaviour of the fields at spatial infinity. The use of test functions is convenient for keeping such problems under control as well.

Next we look at Maxwell's equations:

$$\square A_\alpha - \partial_\alpha \partial \cdot A = 0 . \quad (10)$$

By inspection, we see that there will be problems. The obvious guess, that one can specify A_α and \dot{A}_α on a spacelike hypersurface, and then use the equation of motion to construct A_α everywhere, is simply wrong. First, we observe that the time component of Maxwell's equations reads

$$\partial_\alpha (\dot{A}_\alpha - \partial_\alpha A_t) = 0 . \quad (11)$$

This equation has to hold on every spacelike hypersurface, which shows that the proposed initial data *cannot* be arbitrarily specified. Second, there is no equation for the second time derivative of A_t , i.e. the function

$$\ddot{A}_t = \ddot{A}_t(A_\alpha, \dot{A}_\alpha)$$

is simply not there, so that we can not set up a power series solution for A_t , given Maxwell's equations and the initial data. Third, it is well known that given a solution

$A_\alpha(x)$ to Maxwell's equations, we can construct an infinite number of other solutions related to the first by

$$A'_\alpha(x, t) = A_\alpha(x, t) + \partial_\alpha \lambda(x, t) , \quad (12)$$

where $\lambda(x)$ is an arbitrary function. This function may be chosen to vanish on the hypersurface where we are trying to specify initial data, so that we have a proof that no choice of initial data is capable of giving a unique solution of the equations.

At first sight, these problems appear to be rather different. The first problem says that equations are somehow overconstrained, the second and third that they are nevertheless underdetermined. In the Hamiltonian formulation of Maxwell's equations, these problems will have to be faced squarely. Dirac's theory of constrained Hamiltonian systems reveals that in fact these problems are directly connected to each other, and moreover that they are not fatal.

Second Class Constraints

Before we come to the Hamiltonian formulation of gauge theories such as electrodynamics, we will discuss Dirac's approach to the symplectic geometry of phase space. In elementary books on analytical mechanics, it is shown that the Euler-Lagrange equations from the Lagrangian

$$L = L(q, \dot{q}) \quad (13)$$

are equivalent to Hamilton's equations

$$\dot{q} = \frac{\partial H}{\partial p} = \{q, H\} \quad \dot{p} = -\frac{\partial H}{\partial q} = \{p, H\} , \quad (14)$$

where the Poisson brackets, for two arbitrary functions f and g on phase space, are

$$\{f, g\} = \frac{\partial f}{\partial q} \frac{\partial g}{\partial p} - \frac{\partial f}{\partial p} \frac{\partial g}{\partial q} \quad (15)$$

and the Hamiltonian $H = H(q, p)$ is related to L through a Legendre transformation, so that

$$S = \int dt L(q, \dot{q}) = \int dt \dot{q}p - H(q, p) . \quad (16)$$

In the derivation of this result, it is assumed that the equation

$$p \equiv \frac{\partial L}{\partial \dot{q}} = p(q, \dot{q}) \quad (17)$$

is invertible, so that we can solve it for $\dot{q} = \dot{q}(q, p)$. There are, however, many choices of L for which this assumption fails. As a fairly general example, consider the action

$$S = \int dt \theta_\mu(z) \dot{z}^\mu - H(z) . \quad (18)$$

The definition of the canonical momentum that was used above now gives

$$p_\mu = \theta_\mu(z) , \quad (19)$$

which evidently can not be solved for \dot{z}^μ - in fact \dot{z}^μ does not even occur in the formula. Nevertheless the equations of motion, like any differential equation that comes from extremizing an action, can be written in Hamiltonian form. It is just that the form of Hamilton's equations is now somewhat more general than what we had above.

We will look at this problem from two different points of view, first directly, and then using Dirac's theory of second class constraints. First the direct approach: If we vary our action, we find that the equations of motion can be written in the form

$$\omega_{\mu\nu} \dot{z}^\nu = \partial_\mu H(z) , \quad (20)$$

where we have defined the anti-symmetric tensor

$$\omega_{\mu\nu} = \partial_\mu \theta_\nu - \partial_\nu \theta_\mu . \quad (21)$$

This is a tensor on phase space. We now assume that it possesses an inverse,

$$\omega^{\mu\sigma} \omega_{\sigma\nu} = \delta_\nu^\mu . \quad (22)$$

Since it is a well known fact that anti-symmetric matrices can be inverted only if they have an even number of rows and columns, this means that we assume that there is an even number of z 's, or in other words that the phase space is even dimensional. With this assumption made, we can write the equations of motion in a form which provides a natural generalization of Hamilton's equations of motion as given above, namely

$$\dot{z}^\mu = \omega^{\mu\nu} \partial_\nu H(z) = \{z^\mu, H(z)\} , \quad (23)$$

where a generalization of the Poisson bracket has been introduced, to wit:

$$\{f(z), g(z)\} = \partial_\mu f(z) \omega^{\mu\nu} \partial_\nu g(z) \quad \Rightarrow \quad \{z^\mu, z^\nu\} = \omega^{\mu\nu} . \quad (24)$$

To see why this provides a natural generalization of what we had before, split the phase space coordinate z into q 's and p 's according to

$$z^\mu = \begin{pmatrix} q_i \\ p_i \end{pmatrix} . \quad (25)$$

(Note that we are now using the fact that phase space is even dimensional.) Then what was called "the Hamiltonian form of a set of differential equations" above is just given by the special case that

$$\omega^{\mu\nu} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} . \quad (26)$$

We have now - by a slight extension of the meaning of the words - succeeded in casting our equations of motion in Hamiltonian form. This is an opportunity to present the general definition of a Hamiltonian system, which is the pinnacle of classical physics. It has three ingredients, and reads as follows:

Definition: A Hamiltonian system consists of

- 1) A phase space, which is a differentiable manifold of even dimension.
- 2) A closed non-degenerate two-form $\omega_{\mu\nu}(z)$ defined on phase space.
- 3) A function $H(z)$ on phase space. By assumption, the time evolution of the Hamiltonian system is defined by the vector field

$$\dot{z}^\mu = \omega^{\mu\nu} \partial_\nu H(z) . \quad (27)$$

A “closed non-degenerate two-form” means the following: A covariant tensor is a two-form if it has rank two and is anti-symmetric. The anti-symmetry guarantees that the Poisson brackets are anti-symmetric. The two-form is non-degenerate if its inverse exists, which is needed to ensure that the Poisson brackets exist. The two-form is closed if

$$\partial_{[\mu} \omega_{\nu\sigma]} = 0. \quad (28)$$

This condition guarantees that the Poisson brackets obey the Jacobi identity. (The last statement is not quite obvious, and should be checked by the Gentle Reader.) Concerning point 3 in the definition, we observe that any function on phase space can be associated with a vector field by means of the symplectic two-form. The symplectic form plays a role which - in Arnold’s phrase - “is analogous to, but refreshingly different from” the role played by the metric tensor in Riemannian geometry, and it is therefore customary to talk about “symplectic geometry”. We see that the elementary definition of Hamiltonian systems differs from the general definition in that it requires the symplectic two-form to take the “flat” form given (for its inverse) in eq. (26). There is a theorem, due to Darboux, which says that locally in phase space one can always find a coordinate system in which $\omega_{\mu\nu}$ takes this form.

We now turn to Dirac’s way of dealing with general symplectic structures. Suppose that, when we perform the Legendre transformation of some Lagrangian, we encounter a set of constraints

$$\Phi_m = \Phi_m(q, p) = 0 . \quad (29)$$

It will still be possible to carry through the Legendre transformation and to compute $H = H(q, p)$, so at first we simply ignore the constraints, and postulate the ordinary Poisson brackets

$$\{q_i, p_j\} = \delta_{ij} \quad \{q_i, q_j\} = \{p_i, p_j\} = 0 \quad (30)$$

—even though they are inconsistent with the constraints, as is evident in the particular case we studied earlier:

$$\delta_\nu^\mu = \{z^\mu, p_\nu\} = \{z^\mu, \theta_\nu(z)\} = 0 .$$

We call these Poisson brackets “the naive Poisson brackets”. Now Dirac gives a simple recipe for how to use the naive brackets to construct an improved Poisson bracket which is consistent with the constraints. This improved bracket is known in the literature as the Dirac bracket, and it is defined, for an arbitrary pair of phase space functions, as follows:

$$\{f, g\}^* \equiv \{f, g\} - \{f, \Phi_m\} C_{mn}^{-1} \{\Phi_n, g\} , \quad (31)$$

where we employ the inverse of the constraint matrix

$$C_{mn} \equiv \{\Phi_m, \Phi_n\} . \quad (32)$$

(Note that we now assume that C_{mn} can be inverted.) It is easy to see that the Dirac bracket of a constraint with an arbitrary phase space function vanishes by construction, so that if we use the Dirac bracket the kind of inconsistency that plagued the naive brackets can not occur. This is of course the point of the construction.

In the example that we studied earlier, we see that

$$\{\Phi_\mu, \Phi_\nu\} = \{P_\mu - \theta_\mu(z), P_\nu - \theta_\nu(z)\} = \partial_\mu \theta_\nu(z) - \partial_\nu \theta_\mu(z) = \omega_{\mu\nu} . \quad (33)$$

It follows that the Dirac bracket $\{z^\mu, z^\nu\}^*$ agrees precisely with the “curved” Poisson bracket that was used in the direct approach to the definition of the Hamiltonian system.

Dirac’s definition of the constrained Hamiltonian system defined by a given Lagrangian now reads as follows: There is a “naive” phase space, spanned by the q ’s and p ’s that were introduced in the Legendre transformation of the action. The physical phase space of the model is the submanifold of the naive phase space defined by the constraint equations. There is a “flat” symplectic structure in the naive phase space, which induces a non-trivial symplectic structure on the physical phase space. This is precisely given by the Dirac bracket. Finally, the Hamiltonian is the one that is obtained from the Legendre transformation.

First Class Constraints

In the previous section we had to assume that a certain matrix—the symplectic two-form in the first case and the constraint matrix in the second—has an inverse. But this might fail for a constraint surface. It could behave like a null surface in space-time, for which the induced metric is degenerate. When the induced symplectic form on the constraint

surface is degenerate we are dealing with gauge theories, or in Dirac's terminology with constrained systems with first class constraints. It is in the treatment of this case that the Master's Book is truly original.

Dirac approaches every constrained system in the same fashion: We start from a Lagrangian $L(q, \dot{q})$, derive the canonical momenta, postulate the naive Poisson brackets, and compute the Hamiltonian. For simplicity, we assume that no second class constraints occur, or if they do, that they have been dealt with already and the naive brackets replaced with Dirac brackets. There remain a set of constraints

$$\Psi_m \approx 0 . \quad (34)$$

The wavy equality sign here denotes "weak equality". It means that two things are equal, and we will keep track of this, but for the time being it will be ignored: In any solution of the equations of motion, the constraints do indeed vanish, but Dirac's idea is to ignore them while the canonical formalism is being set up. So "weak equalities" may not be used inside Poisson brackets. It is of course essential that

$$\dot{\Psi}_m \equiv \{\Psi_m, H\} \approx 0 . \quad (35)$$

If the right hand side does not vanish as a consequence of the constraints already obtained, one declares it to be zero, whatever it is. In this way one can obtain new constraints (called secondary). Also their time derivatives have to vanish, so by repeating this procedure one may obtain additional (tertiary) constraints, and so on. Eventually this procedure stops, because the right hand side vanishes as a consequence of the constraints already obtained. At this point we have all the constraints, say M of them. One then computes their Poisson bracket algebra. Since we assumed that no second class constraints are present, the result is of the general form

$$\{\Psi_m, \Psi_n\} = U_{mn}{}^r \Psi_r \approx 0 , \quad (36)$$

where the "structure functions" $U_{mn}{}^r$ may depend on the q 's and p 's. We also define an "extended" Hamiltonian, using some Lagrange multipliers, as follows:

$$H_{ext} = H + \lambda_m \Psi_m \approx H . \quad (37)$$

At each step of the algorithm we have to change $H \rightarrow H_{ext}$ before checking consistency through eq. (35).

We can summarize the results from Dirac's treatment of a system with first class constraints only as follows: The phase space action is

$$S = \int dt \dot{q}p - H - \lambda_m \Psi_m . \quad (38)$$

The constraints together with the Hamiltonian obey the Poisson bracket algebra

$$\{\Psi_m, \Psi_n\} = U_{mn}{}^r \Psi_r \quad \{\Psi_m, H\} = V_m{}^r \Psi_r . \quad (39)$$

The equations of motion are

$$\dot{q} = \{q, H\} + \lambda_m \{q, \Psi_m\} \quad (40)$$

$$\dot{p} = \{p, H\} + \lambda_m \{p, \Psi_m\} \quad (41)$$

$$0 = \Psi_m(q, p) . \quad (42)$$

and—a crucial point—there are no equations of motion for the λ_m 's. As a result, since the λ_m 's do enter into the equations of motion for the q 's and p 's, the time evolution of the latter is partly arbitrary. This seems unsatisfactory, but it is precisely the situation that we encountered for Maxwell's equations, and we will soon see that it all makes perfect sense.

Let us go back to our assumption that the constraint matrix is weakly vanishing. In general, what may happen is that

$$\{\Psi_m, \Psi_n\} \approx C_{mn} \neq 0 , \quad (43)$$

where C_{mn} is a matrix whose rank is less than or equal to M , the number of constraints. If the rank is less than M but greater than zero, say equal to R , this means that *both* second and first class constraints are present, and it becomes necessary to isolate a set of R second class constraints and use them to define the appropriate Dirac brackets. Then there will remain $M - R$ first class constraints to be dealt with according to the procedure that we are about to describe. Dirac's book contains an easy-to-follow recipe for how to do this kind of things. With experience, one learns of various shortcuts, which are easy to remember but complicated to describe. I will make only one comment on this here: Suppose that all the constraints are second class, and suppose that we insist that eq. (35) holds. Then nonsense will result, unless we take appropriate precautions. To be precise, what one has to do is to change the Hamiltonian to the form given in eq. (37) before the calculation is made. Then what the calculation (35) gives is just an equation for the Lagrange multiplier that was introduced in eq. (37). So, remember that Lagrange multipliers that multiply second class constraints will be determined by the consistency conditions, while those that multiply first class constraints are left arbitrary. Also remember the overriding rule that if there is trouble of any kind, an unthinking application of the recipe in the Master's Book, without shortcuts, will give you the right answer.

Electrodynamics

At this point an example is usually helpful. So consider the action for electrodynamics, with a mass term for the photon added:

$$S = \int d^4x \mathcal{L} = - \int d^4x \left[\frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} + \frac{1}{2} m^2 A_\alpha A^\alpha \right]. \quad (44)$$

The canonical momenta are

$$E^a = \frac{\partial \mathcal{L}}{\partial \dot{A}_a} = F_{ta} \quad \pi^t = \frac{\partial \mathcal{L}}{\partial \dot{A}_t} = 0. \quad (45)$$

It is straightforward to compute the Hamiltonian; if partial integrations are allowed and the magnetic field is defined as usual, we find that

$$\begin{aligned} H &= \int d^3x \left[\dot{A}_a E^a - \mathcal{L} \right] = \\ &= \int d^3x \left[\frac{1}{2} (E^a E^a + B^a B^a) - A_t \partial_a E^a + \frac{m^2}{2} (A_a A_a - A_t A_t) \right]. \end{aligned} \quad (46)$$

Now from eq. (45) we evidently have the primary constraint

$$\pi^t \approx 0. \quad (47)$$

Consistency then implies the secondary constraint

$$0 \approx \dot{\pi}^t = \{\pi^t, H\} = \partial_a E^a + m^2 A_t. \quad (48)$$

Following custom, we refer to the constraint

$$\mathcal{G} \equiv \partial_a E^a + m^2 A_t \approx 0 \quad (49)$$

as ‘‘Gauss’ law’’. It is of course recognizable as the time component of the Lagrangian field equation

$$\partial_\beta F^{\beta\alpha} - m^2 A^\alpha = 0. \quad (50)$$

Before we investigate whether there are any tertiary constraints, we compute the algebra of the constraints that we have already collected, and find that

$$\{\mathcal{G}(x), \pi(y)\} = m^2 \delta(x, y). \quad (51)$$

The conclusion here depends critically on m^2 . If $m^2 = 0$ the constraints are first class, if not they are second class. Let us deal with the latter case first. The Dirac brackets are easily computed, and include

$$\{A_a(x), E^b(y)\}^* = \{A_a(x), E^b(y)\} = \delta_a^b \delta(x, y). \quad (52)$$

Once the naive brackets have been replaced by Dirac brackets we can set the constraints ‘‘strongly’’ to zero (i.e. we replace the wavy equality signs with ordinary ones), which in

this case means that we can express A_t as a definite function of A_a and E^b , which remain as the “true” degrees of freedom of the system. The time components simply drop out from the Hamiltonian formulation. In this way we arrive at a consistent Hamiltonian system described by the Poisson brackets (52) and the Hamiltonian

$$H = H[A, E] = \frac{1}{2} \int d^3x \left[E^a E^a + B^a B^a + m^2 A_a A_a + \frac{1}{m^2} (\partial \cdot E)^2 \right]. \quad (53)$$

This is all. By the way we note that the Hamiltonian is now manifestly bounded from below, something which was not evident from the start. Since all the constraints have been solved for, we can also count the number of degrees of freedom in the theory, i.e. one half of the dimension of the phase space, and find this number to be $3 \cdot \infty^3$, or three per point in space.

The massless case is subtler. Gauss’ law now says that the electric field is divergenceless. It is easy to check that there are no tertiary constraints, so that we have reached the position that was reached for a general Hamiltonian system with first class constraints in the previous section. Thus, the phase space action is

$$S = \int d^4x \left[\dot{A}_a E^a - \frac{1}{2} (E^a E^a + B^a B^a) - \lambda \partial_a E^a \right]. \quad (54)$$

Here I have renamed $\lambda \equiv -A_t$ and then quietly dropped the time components from the phase space; I could have included them in the action and added a Lagrange multiplier to impose the constraint that the time component of the momentum is weakly zero, but since this leaves the time evolution of A_t entirely arbitrary the content of the resulting equations would have been exactly the same as the content of the equations that we derive from the action (54), viz.

$$\dot{A}_a = E^a - \partial_a \lambda \quad (55)$$

$$\dot{E}^a = \partial_b F^{ba} \quad (56)$$

$$0 = \partial_a E^a. \quad (57)$$

This is the situation encountered in the introduction; there is a $3 \cdot \infty^3$ -dimensional phase space spanned by $A_a(x)$, $E^b(y)$, but the time evolution of the vector potential is arbitrary because the evolution equation includes an unspecified gauge transformation. Moreover the initial data are subjected to a constraint, or more precisely $1 \cdot \infty^3$ constraints (one at each point in space). The reason why Maxwell’s equations do not determine the time component of the vector potential is revealed to be that it enters the action as a Lagrange multiplier which enforces the constraint. Concerning the physical interpretation of this formalism, the crucial idea is familiar from elementary electrodynamics: Changes of “gauge”, i.e. changes in the vector potential of the form

$$A'(x) = A(x) + \partial_a \Lambda(x) , \quad (58)$$

do not correspond to any changes in the state of the physical electromagnetic field. The fields $A'(x)$ and $A(x)$ are related by a gauge transformation and describe the same physics. As we will see, there is a similar resolution to the general problem of giving physical interpretation to an arbitrary Hamiltonian system having first class constraints in its phase space.

Observables

Our understanding of systems with second class constraints is complete. For a Hamiltonian system without constraints, the fundamental assumption made in the interpretation of the theory is that there is *a one-to-one correspondence between the points in phase space and the physical states of the system being described*; also the phase space is equipped with two structures, a symplectic two-form and a Hamiltonian function, which together define the time evolution of the system. If second class constraints occur, there is a constraint hypersurface embedded in the “naive” phase space, defined by the constraints $\Phi(q, p) = 0$. Because of the embedding, the symplectic form on the naive phase space induces a definite symplectic form on the constraint surface, and the interpretation of the whole structure is that there is a one-to-one correspondence between the points on the constraint surface and the physical states of the system. There is no real distinction between the constraint surface on the one hand and the phase space of an unconstrained system on the other—the naive phase space is just a convenient crutch used in setting up the description.

We have left the physical interpretation of Hamiltonian systems with first class constraints floating in the air, and it is time to fix it. The picture of a system with second class constraints is still relevant, since again there is a naive phase space and a constraint surface, and the “physical points” in phase space have to lie on the latter. The essential point in Dirac’s theory is now that the one-to-one correspondence between points on the constraint surface and physical states of the system is given up—the latter correspond to equivalence classes of points on the constraint surface. To see this, we recall that the constraint surface is defined by a set of constraints

$$\Psi_m(z) \approx 0 . \quad (59)$$

Now let us concentrate our attention on the fact that these functions are given, rather than on the fact that they vanish in a solution of the equations. The symplectic two-form can be used to associate vector fields not only to the Hamiltonian, but to any phase space function, and in particular to the constraint functions; so we introduce the vector fields

$$\xi_m^\mu \equiv \omega^{\mu\nu} \partial_\nu \Psi_m . \quad (60)$$

These vector fields give rise to curves which have the property that if they start out from a point on the constraint surface they stay on the constraint surface. (This can be proved

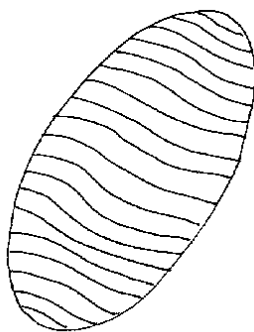


Figure 1: The constraint surface partitioned into gauge orbits (schematically).

using the Poisson bracket algebra of the constraints. The constraints are said to be “in involution” with each other and with the Hamiltonian, and if you carry through the proof you may recognize it from the theory of integrable systems, where the same argument occurs in a different context.) Let us agree to call each such curve a “gauge orbit”. If there are M constraints altogether, the vector fields give rise to M dimensional surface elements at every point, and it can be proved that they are integrable, and define an M dimensional surface embedded in the constraint surface. (This follows from the algebra of the constraints and Frobenius’ theorem.) These surfaces are sometimes called “gauge flats”, although they are often loosely referred to as gauge orbits as well. We can now draw a picture of the situation, which shows the constraint surface embedded in the naive phase space, and foliated by the gauge orbits. The symplectic form on the naive phase space again induces a two-form ω^* on the constraint surface, but it is not immediately useful since it is degenerate—its null eigenvectors are precisely the vectors ξ^μ that point along the gauge orbits.

The crucial idea is that the constraint surface is divided into equivalence classes of points. Each equivalence class consists of all the points that lie on a given gauge flat. The physical interpretation of the construction is now given by the assumption that there is *a one-to-one correspondence between the gauge flats and the physical states of the system being described*. Thus,

$$\text{Physical phase space} \equiv \frac{\text{Constraint surface}}{\text{Gauge transformations}} . \quad (61)$$

A physical observable is a function that takes a definite value if the physical system assumes a definite state, hence in the formalism an observable is a function of the constraint surface that takes the same value at all the points on a given gauge orbit. Thus, if \mathcal{O} is an observable, we must have that

$$\mathcal{L}_{\xi_m} \mathcal{O} = 0 \quad \Leftrightarrow \quad \{\mathcal{O}, \Psi_m\} \approx 0 . \quad (62)$$

In words, \mathcal{O} is a gauge invariant function on the naive phase space.

This definition provides some guidance in the physical interpretation of a theory; how

stringent the guidance is varies from theory to theory. In electrodynamics, the electric and magnetic fields provide examples of observables. This is no longer true in non-abelian Yang-Mills theory, where the canonical variables carry an extra index supplied by the adjoint representation of some compact Lie group. We have the canonical Poisson bracket

$$\{A_{ai}(x), E_j^b(y)\} = \delta_b^a \delta_{ij} \delta(x, y) . \quad (63)$$

We also have the first class constraints

$$\mathcal{G}_i[\lambda_i] = \int d^3x \lambda_i (\partial_a E_i^a + f_{ijk} A_{aj} E_k^a) \approx 0 . \quad (64)$$

Here f_{ijk} are the structure constants of the group, $\lambda_i(x)$ are (fairly) arbitrary functions, and I used the convenient idea of presenting the constraints smeared with test functions. One checks that the constraints generate the gauge transformations

$$\delta A_{ai} = -\partial_a \lambda_i - f_{ijk} A_{aj} \lambda_k , \quad \delta E_i^a = -f_{ijk} E_j^a \lambda_k . \quad (65)$$

The electric field is no longer gauge invariant, but transforms like a vector in the internal representation space. It remains possible to write down observables by inspection, such as the energy density.

In general relativity, the situation is very subtle since there the role of the canonical Hamiltonian is entirely taken over by constraints; the time evolution itself is a gauge transformation, and indeed people still disagree on the questions raised by this circumstance. More than that, quarrels erupt and friendships are broken up when this issue is discussed—much like what happens with the meaning of entropy, probability, and such things.

Canonical Gauge Fixing

In looking over the picture of a system with first class constraints, an idea suggests itself, namely that one could define a physical phase space by simply picking one point on each gauge orbit (or gauge flat, if there are many constraints), and call the resulting set of points the physical phase space. In the end, one would then arrive at an ordinary Hamiltonian system without constraints on the degrees of freedom that remain in the theory.

Within limits, such a procedure can indeed be carried through, and is called “canonical gauge fixing”. The first limit is that it may be impossible to do this globally, all over the constraint surface. Problems arise if the latter is a non-trivial fiber bundle of some sort. Even when this problem does not occur, there may be problems with non-locality in space—physical space, not phase space. We will show, using electrodynamics as an example, how canonical gauge fixing necessarily leads to the appearance of non-local differential operators or, which is the same thing, that a description making use of gauge

degrees of freedom is necessary if electromagnetism is to be described in a way that is manifestly local in space-time.

The recipe for how to perform canonical gauge fixing is given in Dirac’s book. Suppose for simplicity that we are interested in fixing the gauge pertaining to one first class constraint Ψ only. Suppose that we find (by inspection, trial-and-error, or whatever) a condition

$$\chi(q, p) = 0 , \tag{66}$$

such that the matrix

$$\begin{pmatrix} 0 & \{\chi, \Psi\} \\ \{\Psi, \chi\} & 0 \end{pmatrix} \tag{67}$$

is invertible. Then we can treat χ and Ψ as a pair of second class constraints, solve them, compute the Dirac brackets for the remaining degrees of freedom, and in this way we arrive at a description of our model as an unconstrained system with fewer degrees of freedom than was used from the start. To be precise, we see that the naive phase space has two “extra” degrees of freedom, compared to the gauge fixed version.

Of course there is a detail that has to be checked before one concludes that the gauge has been successfully fixed, namely that every gauge flat contains one and only one point which obeys the condition $\chi = 0$. Let us see, using electrodynamics as an example, how the question can be analyzed. Our proposed gauge condition is the well known Coulomb gauge,

$$\chi(x) = \partial \cdot A = 0 . \tag{68}$$

One finds

$$\{\chi(x), \mathcal{G}(y)\} = -\Delta\delta(x, y) , \tag{69}$$

where Δ is the Laplacian. To show that the Coulomb gauge condition is a good gauge choice the first thing to do is to check that the constraint matrix (67) is invertible. It is clear that this requires the Laplacian to be invertible, so that we can write an inverse constraint matrix—in a meaningful way—as

$$C^{-1}(x, y) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \frac{1}{\Delta}\delta(x, y) . \tag{70}$$

Exactly the same question comes up when we check the second requirement on the gauge choice, that every gauge flat should contain one and only one point that obeys the gauge condition. Indeed, suppose that the vector potential A_a does obey the Coulomb gauge condition. Then all the other vector potentials A'_a on the same gauge flat obey

$$\partial \cdot A'_a(x) = \partial_a(A_a(x) + \partial_a\Lambda(x)) = \Delta\Lambda(x) . \tag{71}$$

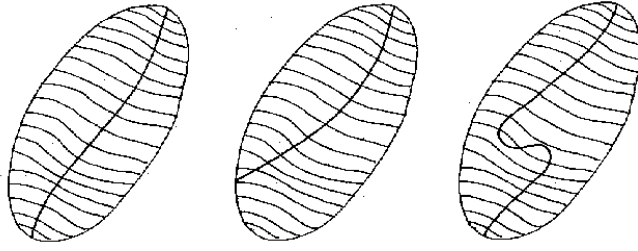


Figure 2: A good gauge choice, and two bad ones.

We need to show that the left hand side vanishes if and only if $\Lambda(x)$ vanishes. So the question comes down to whether the Laplace equation

$$\Delta\Lambda(x) = 0 \tag{72}$$

has a unique solution ($= 0$) for Λ . It must be stressed that this question can not be answered unless our so far careless description of electrodynamics is supplemented with a certain amount of additional information. The extra information concerns the behaviour of the fields at large distances, and—for a relativist—the behaviour of space itself at large distances. If space is flat and Euclidean and if all the fields (and Λ , whose behaviour at infinity is tied up with that of the vector potential) vanish sufficiently fast at infinity, then the Laplace equation has a unique solution, and therefore there can be at most one vector potential obeying the Coulomb condition on every gauge flat. Some restrictions on the large distance behaviour of the fields are in fact necessary to ensure that the Hamiltonian formalism is well defined in the first place—obviously the integrals that define the Hamiltonian must converge, and there are other requirements too. But it is worth observing that if we solve these problems with periodic boundary conditions, so that space is a torus, then the inverse of the Laplacian is not well defined, and the Coulomb gauge as it stands is not acceptable.

An extension of the argument shows that—given suitable boundary conditions—there is one vector potential on every gauge flat that obeys the Coulomb condition, and the appearance of the Laplacian in a denominator in eq. (70) can be justified. The constraint and the gauge condition together can now be regarded as a pair of second class constraints, and we can solve for the independent physical degrees of freedom, which are the “transverse parts” of the vector potential,

$$A_a^T \equiv \left(\delta_{ab} - \frac{\partial_a \partial_b}{\Delta}\right) A_b, \tag{73}$$

and similarly for the electric field. It is straightforward to compute the Dirac brackets of these degrees of freedom, but it should be noted that these brackets then involve the inverse of the Laplacian, which is a non-local operator. It is in this sense that manifest locality is lost once the Coulomb gauge has been fixed. (Although it remains true that propagation of influences from a point is confined to the interior of its light cone.)

There are other conditions that may be employed to fix the gauge in electrodynamics, such as the axial gauge

$$A_3(x) = 0 \tag{74}$$

and many others. The familiar and manifestly covariant Lorenz gauge

$$\partial \cdot A - \dot{A}_t = 0 \tag{75}$$

on the other hand belongs to a different kettle of fish, as is evident from the fact that it involves the Lagrange multiplier A_t . It is not a canonical gauge fixing condition at all; its treatment belongs to a part of the theory of constrained Hamiltonian system that we will not discuss, namely the part marked “BRST” by its practitioners.

Which will serve as a reminder that this sketch of the theory is incomplete.

Exercises

- Show how the assumption that the symplectic form is closed implies that the Poisson brackets obey the Jacobi identity.
- Explain in technical terms why Dirac’s construction gives a symplectic form which is “induced” from the symplectic form in the naive phase space through the embedding of the constraint surface. Make the explanation as general as you can.
- Show how the induced symplectic form is degenerate when first class constraints are present.
- Show that the gauge orbits are confined to the constraint surface.
- Show that the gauge flats are integrable.

A coherent account of these questions will be appreciated.

Recommended reading:

The only indispensable reference is

P.A.M. Dirac, *Lectures on Quantum Mechanics*, Belfer Graduate School of Science,

New York 1964.

The geometrical point of view on constrained phase spaces is described for instance in

A. Ashtekar and M. Stillerman, Geometric quantization of constrained systems, J. Math. Phys. 27 (1986) 1319.

The modern “BRST” approach to first class constraints is summarized for instance in

R. Marnelius, Introduction to the quantization of general gauge theories, Acta Phys. Pol. B13 (1982) 669.

For a brief discussion of the initial value problem, see

R. M. Wald, *General Relativity*, Chicago UP, 1984.