1993 (?)

Term project for Math 189 by Dimitri Shlyakhtenko

Generating Functions of Canonical Transformations

In this project I would like to look back at some of the treatments of mechanical systems in classical texts and see how these treatments look like in modern-day terminology and from a modern point of view. This project is based on three texts, Gantmakher's *Lektsii po Analiticheskoj Mekhanike* (in Russian), Goldstein's *Classical Mechanics* and Whittaker's *Treatise on the Analytical Dynamics of Particles and Rigid Bodies.* Some (and admittedly, most) ideas on modern treatments come from Marsden and Ratiu, *An Introduction to Mechanics and Symmetry*, volume 1.

I would like to begin by setting up the stage. Most of what will be done is done locally, so in most cases, unless otherwise is specified, mechanical systems are represented by linear spaces. I tried to uniformly use the letter q for spatial variables and p for momenta; these are used without explanation. These notes aside, we consider the phase space $\mathbb{R}^n \times \mathbb{R}^{*n}$, with the canonical symplectic structure given by the form $\Omega = dq^i \wedge dp_i = -d\Theta$. In classical literature, especially Whittacker, this form is not considered a whole lot; this is probably due to the fact that the differential geometry machinery was not developed well enough at the time. Gantmakher has a one-page treatment of the symplectic form, but he is reluctant to let it depend on the position; rather, he considers it as a constant matrix. Now, given a Hamiltonian function H, we may consider a Hamiltonian system of equations, given by, in modern notation the vector field X_H, defined by the property that

$$\Omega(X, v) = dH \cdot v$$

where v is any tangent vector. In classical picture, as I mentioned, the form Ω is considered constant and (since everyone uses p and q as the coordinates) has the form

$$\begin{bmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{bmatrix}$$

Hence if $\gamma(t) = (p(t), q(t))$ is an integral curve of the vector field X_H above, we have the equations

$$\dot{q}^i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q^i}$$

which are called the Hamilton canonical equations. In these considerations, H was a function on the phase space, i.e., of p and q. However, in some cases it is interesting to consider a time-dependent Hamiltonian system. Our machinery is already setup for that: we enlarge the configuration space by multiplying with another copy of the real numbers, and consider a new symplectic form, given by

$$dH \wedge \pi_2^*(dt) + \pi_1^*(\Omega)$$

where π_1 denotes the projection onto the original configuration space and π_2 denotes the projection onto the "time" part. The form dt is a *fixed* form on the time part. One of the fundamental properties of Hamiltonian systems is that their flow preserves the canonical form. Our immediate goal is to demonstrate that the classical texts already have this notion, and in fact "secretly" consider the same canonical forms as we did. To do this, let us look at what Gantmakher called the Poincaré-Cartan invariant, or the first-order invariant. He considers a one-dimensional contour (i.e., boundary of a two-dimensional region) that is perpendicular to the time direction (i.e., a contour in the "original" phase space before the extension) and computes the integral of the form

$$p_i dq^i - H dt$$

over this contour. He then states and proves (in coordinates) that this is actually an invariant of the translation of this contour along the flow of the system (i.e., along the flow lines of the physical system as time progresses). But by applying the Stokes' theorem to this integral, one quickly realizes that this integral is actually an integral over a two- dimensional contour, of the negative of the extended canonical form (i.e., the form $\Omega + dH \wedge dt$.) Hence the fact that the original integral is conserved over all possible contours is equivalent (by choosing smaller and smaller two-dimensional contours) to insisting that the canonical form be preserved by the flow. The bottom line is that we can use the coordinate calculations in classical books that rely on this "first-order invariant" without the fear of being off.

Let us now consider the question of mappings between symplectic spaces. The appropriate mappings in this category are, of course, the canonical or symplectic transformations, that is, maps that preserve the canonical form (and hence all of the structure of a symplectic manifold.) In disguise, the classical definition is exactly the same: a system of differential equations in the Hamilton canonical form is transferred into a system in Hamilton canonical form. Since both times the systems are written by using the canonical form in canonical coordinates, this just says that the symplectic form is mapped into itself. However, from such a definition it is hard to produce any results; so Gantmakher considers the "first order invariant" again and shows what this property of the transformation means for this invariant. Consider a symplectic map ϕ between spaces M and N. Suppose H is an arbitrary Hamiltonian on M

and γ is an integral curve of the corresponding Hamiltonian system on M. Then, since the system on M gets mapped into a Hamiltonian system on N, γ remains an integral curve of the associated system on N. Now compute the first-order invariant on M. By invariance of integration, it is the same as computing it on N, so that the invariant on M gets pushed forward into a certain first-order "object" (meaning a closed two-form in modern terminology) that is preserved by the flow of a certain Hamiltonian system on N. Gantmakher next gives a coordinate proof that the only invariants conserved by the flow of a system differ from the "first-order invariant" by a constant multiple. The proof Gantmakher provides works only for two-dimensional manifolds, and occupies two pages. Restated, this result says that whenever a certain closed two-form ω is preserved by the flow of every Hamiltonian (with respect to the canonical form) system, then it has to be a multiple of the canonical form. In two dimensions (from the modern point of view) this result is completely obvious; for given a canonical form Ω, and a two-form Ω', they differ by a multiple, a certain function f. The invariance under the flow means that

$$L_X \Omega' = 0$$

where X is any Hamiltonian vector field for Ω. But then

$$0 = L_X f\Omega = f L_X \Omega + X \cdot f\Omega = X \cdot f\Omega$$

since Ω is invariant under the flow of X. Thus $X \cdot f = 0$ and so f is a constant.

We now consider an alternate description of canonical transformations, via generating functions. In the book of Goldstein, it is pointed out that canonical transformations can be defined by the means of generating functions. Goldstein (with little motivation) gives four generic types of generating function, for which the transformation laws become:

- **Type 1.** The transformation laws are:

$$p_i = -\frac{\partial S}{\partial q^i} \quad P_i = \frac{\partial S}{\partial Q^i}$$

- **Type 2.** The transformation laws are:

$$p_i = \frac{\partial S}{\partial q^i} \quad Q^i = \frac{\partial S}{\partial P_i}$$

- **Type 3.** The transformation laws are:

$$P_i = -\frac{\partial S}{\partial Q^i} \quad q^i = -\frac{\partial S}{\partial p_i}$$

- **Type 4.** The transformation laws are:

$$q^i = -\frac{\partial S}{\partial p_i} \quad Q^i = \frac{\partial S}{\partial P_i}$$

The aim of this paper is to find modern-language explanations for these formulas. We follow the treatment of Marsden and Ratiu.

The general setup is as follows. Suppose that the manifolds T^*M andT^*N are endowed with the usual symplectic structure, and f is a map between the two. Consider the submanifold $\Gamma \subset T^*M \times T^*N$, the graph of f. Let π_1, π_2 denote the projections onto M and N. Finally, if Ω_1, Ω_2 are the symplectic forms on M, N, let

$$\Omega = \pi_1^*\Omega_1 - \pi_2^*\Omega_2$$

be a symplectic form on the product. Notice that if i is the inclusion map from Γ into the product, then $i^*\Omega = 0$ iff f is symplectic. Indeed, if f is symplectic, then $\Omega_1 - f^*\Omega_2 = 0$; so $0 = \pi_1^*\Omega_1 - \pi_1^* f^*\Omega_2 = \pi_1^*\Omega_1 - \pi_2^*\Omega_2$. The other direction follows similarly. Note also that there always is a form Θ, such that $\Omega = d\Theta$; possible candidates include forms that have local expressions $-p_i dq^i + P_i dQ^i$, $-p_i dq^i + Q^i dP_i$, etc. Thus $i^*\Theta$ is closed on Γ, and hence is (locally) a $-dS$ for some function S. We call such a function S a generating function of the transformation f.

Suppose (q,p) and (Q,P) are charts on T^*M and T^*N. Then it is possible that the map f is such that (q,Q), (q,P), (p,Q) and (p,P) are coordinates on Γ. These choices of coordinates correspond to the four types of generating functions that Goldstein considers. Indeed, if (q,Q) are coordinates on Γ, and we let $\Theta = P_i dQ^i - p_i dq^i$, then the equation for S becomes

$$dS = \frac{\partial S}{\partial q^i} dq^i + \frac{\partial S}{\partial Q^i} dQ^i = p_i dq^i - P_i dQ^i$$

so that

$$p_i = \frac{\partial S}{\partial q^i} \quad -P_i = \frac{\partial S}{\partial Q^i}$$

i.e., S is a function of the first type.

Now suppose that (q, P) form a coordinate system on Γ. Let $\Theta = -p_i dq^i - Q^i dP_i$. Then the equations for S become:

$$dS = \frac{\partial S}{\partial q^i} dq^i + \frac{\partial S}{\partial P_i} dP_i = p_i dq^i + Q^i dP_i$$

so that

$$p_i = \frac{\partial S}{\partial q^i} \quad Q^i = \frac{\partial S}{\partial P_i}$$

i.e., S is a function of the second type.

Now assume that (p, Q) form a coordinate system on Γ. This is just the previous case with p and Q switched. Note that this switching changes the sign of Θ, so that $\Theta = q^i dp_i + P_i dQ^i$. We get:

$$dS = \frac{\partial S}{\partial p_i} dp_i + \frac{\partial S}{\partial Q^i} dQ^i = -q^i dp_i - P_i dQ^i$$

and hence

$$P_i = -\frac{\partial S}{\partial Q^i} \quad q^i = -\frac{\partial S}{\partial p_i}$$

i.e., S is a function of the third type.

Finally, if (p, P) is a chart on Γ, we have, for $\Theta = q^i dp_i - Q^i dP_i$ that

$$dS = \frac{\partial S}{\partial p_i} dp_i + \frac{\partial S}{\partial P_i} dP_i = -q^i dp_i + Q^i dP_i$$

and we get

$$q^i = -\frac{\partial S}{\partial p_i} \quad Q^i = \frac{\partial S}{\partial P_i}$$

i.e., S is a function of the fourth type.

Note that we again assumed that the generating functions (and Haminltonians) are time-independent. For the time-dependent case, one can simply replace Θ everywhere by $\Theta - Hdt$ and get an extra equation that shows how the Hamiltonian changes. In the one case that we shall require later, namely, the functions of type 1, we get the following equation:

$$dS = \frac{\partial S}{\partial p_i} dp_i + \frac{\partial S}{\partial Q^i} dQ^i + \frac{\partial S}{\partial t} = -q^i dp_i - P_i dQ^i - (H - \overline{H})dt$$

so that the Hamiltonian changes according to

$$\overline{H} = H + \frac{\partial S}{\partial t}$$

Note that we used the same dt when pulling back the extended forms. Thus can be justified by saying that there is no canonical choice of the forms dt and we can therefore always choose them in such a way that they are the same when pulled to the product of the domain and range. Some authors (esp. Gantmakher) believe that altering time is incorrect and we should simply put a certain constant in front of one of the dt's so that the equations become

$$\overline{H} = cH + \frac{\partial S}{\partial t}$$

Then the constant c is called the valency of the transformation and those transformations that have $c = 1$ are called univalent. For our purposes, we can always slow down or speed up the time by this constant and forget of its existence.

Before proceeding any further, let us consider an example that might provide some insight into the physical significance of generating functions. This example comes primarily from Whittacker. Consider a typical optical problem: light is propagating through a medium of variable density. We model light by considering wavefronts rather than corpuscles of light. Suppose $\Sigma(0)$ is the initial surface of the wavefront and $\Sigma(t)$ is the position at time t. Let $V(x, y, z, x', y', z')$ be a function of 6 variables that assigns to these 6 variables the (minimal) value of the parameter t at which the point $(x, y, z) \in \Sigma(s)$ and the point $(x', y', z') \in \Sigma(s + t)$ for some s, i.e., the time it takes for the light to get from the point (x, y, z) to the point (x', y', z'). Let $\mu(x, y, z)$ be the density of space (that is, a factor representing the speed of light at that point in space). Finally, let ξ, η, ζ denote the momenta conjugate to variables x,y,z [1], and ξ', η' and ζ' to be conjugate to x',y' and z'. In other words, we have a family of spaces M_t and a family of functions $\Phi_t : M_0 \to M_t$; the cotangent bundle of each space is endowed with the symplectic structure. Out of physical reasons, Whittacker derives the following equations linking V to the rest of the setup:

$$\frac{\partial V}{\partial x} = -\xi, \quad \frac{\partial V}{\partial y} = -\eta, \quad \frac{\partial V}{\partial z} = -\zeta$$

[1] μ plays a role in defining what this conjugacy is physically. One can let k,l,m be the direction cosines of the normals to the surface; then ξ, η, ζ will be μ times these. I am not inserting any more physical insight into this, since we shall later see a more general example like this, in which little physics is involved

and

$$\frac{\partial V}{\partial x'} = \xi', \quad \frac{\partial V}{\partial y'} = \eta', \quad \frac{\partial V}{\partial z'} = \zeta'$$

We recognize these equations as those defined by a generating function of type 1. Thus the "flow" Φ_t is symplectic. Note that all M_p are isomorphic as manifolds. Thus we can treat Φ_t as a one-parameter family of diffeomorphisms on $M_0 \cong \mathbb{R}^3$. Note also that the way Φ is defined — as the transformation of wavefronts — it follows that this family is actually a *group*. For, according to Huygens principle, the wavefront $\Sigma(s)$ is completely determined by the wavefront $\Sigma(t)$, for $t < s$; indeed, the former is obtained from the latter by looking at how waves emitted by individual points of $\Sigma(t)$ propagate in time. Thus the image of $\Sigma(0)$ under $\Phi_{(t+s)}$ is the same as the image of $\Sigma(0)$ under $\Phi_t \circ \Phi_s$. Thus Φ_t is the flow of a certain vector field; and from the fact that these transformations are symplectic, it actually follows that this vector field is Hamiltonian. We have thus shown two things: first, that light propagation is governed by a certain Hamiltonian system of equations; and second, that the generating function of the flow along the integral curves of this vector field for a certain time t has a physical interpretation: it is the time it takes for the light (particle, if you want) to travel from a certain point to a certain other point.

Note an interesting coincidence: the laws of optics can be stated saying that light will always take the path of least time length. The time it takes to get from one point to another is exactly our function V above. This suggests that there must be a link between the subject of generating functions and variational principle. We shall delay addressing this coincidence until we finish the discussion of the Hamilton-Jacobi equation.

Now that we considered transformations of symplectic manifolds, we can ask the following question: could we transform the whole Hamiltonian system into something completely trivial? One desirable system would have a zero Hamiltonian. So suppose that we have such a transformation and its generating function is of type 1. Then we should have that the resulting Hamiltonian $\overline{H} = 0$, so that

$$H + \frac{\partial S}{\partial t} = 0$$

Also, H is a function of q and p, and we already have an expression for p in terms of S:

$$p_i = \frac{\partial S}{\partial q^i}$$

Hence we get the equation, called the Hamilton-Jacobi equation:

$$H\left(q^1,\dots,q^n,\frac{\partial S}{\partial q^1},\dots,\frac{\partial S}{\partial q^n}\right)+\frac{\partial S}{\partial t}=0$$

If the transformation defined by S happens to be invertible, then it is the desired transformation, mapping our system into a system with the zero Hamiltonian. Since such a system is readily integrable, with solutions being some constant solutions $P=\alpha$ and $Q=\beta$, we have that the integral curves of the original system satisfy the relations

$$\frac{\partial S}{\partial q^i}=p_i,\quad \frac{\partial S}{\partial Q}=P$$

In fact, the opposite is true. If p and q are certain curves satisfying the equations above, then they are the transforms of the solutions of the trivial Hamiltonian system under the inverse of the transform defined by S and hence are solutions of the original system. This is the content of the Hamilton-Jacobi theorem.

We now return to the interesting coincidence we pointed out before. Recall that in the application to optics, the generating function turned out to be the time it takes for the light to pass from one point to another. This time was what we minimize in the Lagrangian approach to mechanics. Hence it is plausible to suppose that the generating function S of the canonical transformation given by the flow along a Hamiltonian vector field of a Hamiltonian corresponding to a Lagrangian L is the integral of the "action", i.e.,

$$S(q,Q,T,t)=\int_\gamma L\,dt$$

where γ is the flow path joining points q and Q. This definition only makes sense for a small period of time $T-t$, since γ may for example be a geodesic, and there are problems with picking the minimal geodesic between far-away points. But these problems aside, let us prove that such a function is indeed the desired generating function. We adopt the treatment in Whittacker.

To compute dW, we need to compute the variation of the Lagrangian over variations of paths in the direction of motion (this corresponds to the fact that our initial and final conditions are on the integral curves of the system). Thus we need to compute the difference in actions between the points a, b and points c, d, with a, c and b, d close together, and such that the integral curve through a passes b at time t, and the integral curve through c passes c at time δt_0 and d at time $t + \delta t_1$. We compute the difference in actions, for δt small:

$$\int_{cd} L dt - \int_{ab} L dt \approx L(b)\delta t_1 - L(a)\delta t_0 + \int_{ab} \sum \left(\frac{\partial L}{\partial q^i}\delta q^i + \frac{\partial L}{\partial \dot{q}^i}\delta \dot{q}^i \right) dt^2$$

The expression in the last integral is nothing else but

$$\sum \left(\frac{\partial L}{\partial \dot{q}^i}\delta \dot{q}^i + \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}^i}\right)\delta q^i \right)$$

by Lagrange's equations (which are satisfied, since we are following an integral curve of the system), so we get that the last integral is equal to

$$\int \frac{d}{dt}\left(\sum \frac{\partial L}{\partial \dot{q}^i}\delta q^i \right)$$

and so the difference becomes

$$\sum \frac{\partial L}{\partial \dot{q}^i}\delta q^i \Big|_a^b + L(b)\delta t_1 - L(a)\delta t_0$$

Note that here δq is a function of time that shows how far the two paths are apart. Hence at the endpoints, we get

$$b - d = \delta q + \delta \dot{q}^i \delta t$$

and similarly for the other two ends. Consequently, substituting this into the formula for the difference in action, and noting the evaluation at the endpoints, we get that the difference is

$$\sum \frac{\partial L}{\partial \dot{q}^i}\delta q^i(t) - \sum \frac{\partial L}{\partial \dot{q}^i}\delta q^i(0) + \left(L - \sum \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i \right)\delta t_0 - \left(L - \sum \frac{\partial L}{\partial \dot{q}^i}\dot{q}^i \right)\delta t_1$$

Note that the expression in () is the Hamiltonian; also, $\frac{\partial L}{\partial \dot{q}^i}$ is the momentum p. Thus the equations for the difference reads

$$p_i \delta q^i(t) - p_i \delta q^i(0) + H\delta t_0 - H\delta t_1$$

[2] $\approx$ means up to second-order terms.

To find $\frac{\partial W}{\partial q}$, where q denotes the initial conditions, we consider paths with b same as d, i.e., $\delta t_1 = 0$. Then we get, from the fact that

$$dW = \sum \frac{\partial L}{\partial q^i} dq^i + \frac{\partial L}{\partial t} dt$$

(keeping in mind that in this case,

$$dW = \sum \frac{\partial W}{\partial q^i} dq^i + \frac{\partial W}{\partial t} dt$$

since only q varies [3]) that

$$p_i = -\frac{\partial W}{\partial q^i}$$

and

$$H = -\frac{\partial W}{\partial t}$$

so that W satisfies the Hamilton-Jacobi equation. Note also that if we insist that the other endpoint stays fixed, we shall get that

$$P = \frac{\partial W}{\partial Q}$$

so that W is indeed the generator of the flow of the system for time t. Hence there is a complete analogy with the situation in optics. The function W was called by Hamilton the *Principal Function.* Before the Hamilton-Jacobi equation was known, Hamilton already knew of this way of describing the flow, that is, new the W generates the flow of the system (according to Gantmakher). However, to get the function one had to compute the integral $\int Ldt$ over an *integral curve* of the system; since once had to *solve* the system to get integral curves, there did not seem to be an effective way of finding what this W was. However, Jacobi realized that the function W satisfies the Hamilton-Jacobi equation, and that this equation completely characterizes it.

In conclusion, let us recall what was demonstrated. We saw how to view generating functions from a modern standpoint; we have even seen that the various types of generating function discussed in classical texts are actually manifestations of the very

[3] To get this equation, we first find all of the partials of W. We do this as follows: if, for example, $\frac{\partial W}{\partial q^i}$ is of interest, we set all δ's equal to zero, except for δq^i; then we divide through by δq^i and take the limit as it approaches zero.

same idea; we also saw the content of the Hamilton-Jacobi equation, in that it defines a generating function that generates the flow of the system. We also found another way of expressing this function, in terms of the Lagrangian of the system, a way that indicates on another link between the Lagrangian and Hamiltonian ides of mechanics. It is interesting to note at this stage that as we have seen a lot of advanced ideas are contained in classical texts, perhaps in disguise. One's heart fills with admiration for the classical authors, that with poor tools managed to penetrate very deeply into the matter of things.

Works Cited

1. Gantmaker, F. *Lektsii po Analiticheskoj mekhanike* (in Russian) Nauka, 1966

2. Goldstein, H. *Classical Mechanics* Addison-Wesley Pub. Co., 1980.

3. Marsden, J.E. and T.S. Ratiu *An Introduction to Mechanics and Symmetry* vol. 1, 1993

4. Whittaker, E. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge, 1927