The Report committee for David Matthew Rosen
Certifies that this is the approved version of the following report:

Geometric Mechanics

APPROVED BY

SUPERVISING COMMITTEE:

Rafael de la Llave, Supervisor

Oscar Gonzalez

Geometric Mechanics

by

David Matthew Rosen, B.S.

REPORT

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF ARTS

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2010

Acknowledgments

The author would like to acknowledge and thank Professor Rafael de la Llave for introducing him to geometric mechanics, and for his guidance all throughout the preparation of this report.

Geometric Mechanics

David Matthew Rosen, M.A.
The University of Texas at Austin, 2010

Supervisor: Rafael de la Llave

This report provides an introduction to geometric mechanics, which seeks to model the behavior of physical mechanical systems using differential geometric objects. In addition to its elegance as a method of representation, this formulation also admits the application of powerful analytical techniques from geometry as an aid to understanding these systems. In particular, it reveals the fundamental role that symplectic geometry plays in mechanics (something which is not at all obvious from the traditional Newtonian formulation), and in the case of systems exhibiting symmetry, leads to an elucidation of conservation and reduction laws which can be used to simplify the analysis of these systems.

The contribution here is primarily one of exposition. Geometric mechanics was developed as an aid to understanding physics, and we have endeavored throughout to highlight the physical principles at work behind the mathematical formalism. In particular, we show quite explicitly the entire

development of mechanics from first principles, beginning with Newton's laws of motion and culminating in the geometric reformulation of Lagrangian and Hamiltonian mechanics. Self-contained presentations of this entire range of material do not appear to be common in either the physics or the mathematics literature, but we feel very strongly that this is essential in order to understand how the more abstract mathematical developments that follow actually relate to the real world. We have also attempted to make many of the proofs contained herein more explicit than they appear in the standard references, both as an aid in understanding and simply to make them easier to follow, and several of them are original where we feel that their presentation in the literature was unacceptably opaque (this occurs primarily in the presentation of the geometric formulation of Lagrangian mechanics and the appendix on symplectic geometry).

Finally, we point out that the fields of geometric mechanics and symplectic geometry are vast, and one could not hope to get more than a fragmentary glimpse of them in a single work, which necessiates some parsimony in the presentation of material. The subject matter covered herein was chosen because it is of particular interest from an applied or engineering perspective in addition to its mathematical appeal.

Table of Contents

Chapter 1

Review of Classical Mechanics

A bit of background

Mechanics is the branch of physics concerned with the motion of physical bodies being acted upon by forces. Although the study of mechanics has its roots in antiquity, the modern (quantitative) synthesis originated in 1687 with the publication of Isaac Newton's *Philosophiæ Naturalis Principia Mathematica* (*Mathematical Principles of Natural Philosophy*), in which he presented his celebrated Three Laws of Motion. These laws define *Newtonian mechanics*, in which the time evolution of a mechanical system is obtained as the solution of an initial value problem corresponding to a system of second-order ordinary differential equations determined by the forces acting on the system.

While Newton's laws provide a complete description of mechanical systems, they are of limited utility when certain forces acting upon a system are difficult to determine directly (as is the case, for example, when attempting to calculate constraint forces acting upon a mechanical system). In order to overcome this difficulty, Joseph-Louis Lagrange introduced an equivalent reformulation of Newtonian mechanics, called *Lagrangian mechanics*, in 1788.

Whereas Newtonian mechanics models the time evolution of a system in terms of forces, Lagrangian mechanics describes a system through the use of a *Lagrangian*, a function incorporating the system's kinetic and potential energies. While formally equivalent to the Newtonian formulation of mechanics, Lagrangian mechanics has the advantage that the forces acting upon a system do not appear explicitly in calculation.

Similarly, in 1833, William Rowan Hamilton in turn introduced a reformulation of Lagrangian mechanics, called *Hamiltonian mechanics*. Whereas the system of ordinary differential equations obtained from the direct application of Newton's laws or Lagrange's equations of motion to a mechanical system is second-order in the system's m configuration variables, Hamiltonian mechanics introduces a set of m auxiliary variables to form a first-order system of ordinary differential equations in the combined $2m$ variables.

Finally, the most modern reformulation of mechanics is *geometric mechanics*. In contrast to Newtonian, Lagrangian, and Hamiltonian dynamics, in which mechanical systems are described in a functional-analytic context through differential equations together with constraints, geometric mechanics models mechanical systems through the use of differential geometric objects. In this formulation, the allowable configurations of the system are modeled as a manifold, called the *configuration space*, and Newtonian, Lagrangian, and Hamiltonian dynamics are encoded through the use of geometric objects associated to the configuration space (e.g., functions and vector fields on the tangent and cotangent bundles of the configuration space, etc.).

Our aim in this chapter is to provide a concise overview of Newtonian, Lagrangian, and Hamiltonian mechanics from the classical (i.e., functional-analytic) viewpoint, which serves as background and motivation for the formulation of geometric mechanics that occurs in the sequel. The exposition is at the level of a beginning to intermediate-level collegiate physics course, and may readily be skipped by readers already familiar with the material.

1.1 Newtonian mechanics

1.1.1 Preliminaries

Newtonian mechanics, as encapsulated in Newton's Laws, describes the motion of *particles*: massive objects (i.e., objects that have mass) that are localized at a point (have no spatial extent).

We think of these particles as existing in (some subset of) $\mathbb{R}^3$, so that given a choice of Cartesian coordinates (a *reference frame*), the location of the particle can be expressed as a 3-tuple $x = (x_1, x_2, x_3)$, called the particle's *position*. In general, a particle's position varies as a function of time, so that

$$x = x(t) = (x_1(t),\ x_2(t),\ x_3(t)).$$

We define the *velocity* of the particle, denoted v, to be the derivative of its position with respect to time:

$$v(t) = \dot{x}(t) = (\dot{x}_1(t),\ \dot{x}_2(t),\ \dot{x}_3(t))$$

(here we follow Newton's convention and use the dot notation to denote differentiation with respect to time). Similarly, we define a particle's *acceleration*, denoted a, to be the derivative of the particle's velocity with respect to time:

$$a(t) = \dot{v}(t) = \ddot{x}(t) = (\ddot{x}_1(t),\ \ddot{x}_2(t),\ \ddot{x}_3(t)).$$

The study of particle motion solely in terms of position, velocity, and acceleration (i.e., without reference to force or mass) is called *kinematics*. Newton's laws relate the kinematic behavior of particles to the forces acting on those particles.

1.1.2 Newton's Laws of Motion

Newtonian mechanics is described by Newton's famous Three Laws of Motion:

Lex Prima

> *Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus a viribus impressis cogitur statum illum mutare.*

Lex Secunda

> *Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.*

Lex Tertia

> *Actioni contrariam semper et æqualem esse reactionem: sive corporum duorum actiones in se mutuo semper esse æquales et in partes contrarias dirigi.*

Rendered into modern English, these laws read:

First Law

> Every body persists in its state of being at rest or of moving uniformly straight forward, except insofar as it is compelled to change its state by force impressed.

Second Law

> The alteration of motion is ever proportional to the motive force impressed; and is made in the direction of the right line in which that force is impressed.

Third Law

> To every action there is always an equal and opposite reaction: or the forces of two bodies on each other are always equal and are directed in opposite directions.

We briefly consider the modern formulation of each of these laws, and a few of their immediate consequences.

1.1.2.1 The First Law

The First Law is often referred to as the *law of inertia*, as it describes *inertial motion*; that is, motion in the absence of external forces. Specifically, it states that *in the absence of externally imposed forces, a particle's velocity is constant as a function of time*. More subtlely, the law of inertia also implicitly defines a distinguished set of reference frames, called *inertial frames*.

By way of example, consider an astronaut in free-fall in a spaceship in deep space (i.e., sufficiently far away from massive bodies that the effects of gravitation are negligible). One "natural" reference frame, from the astronaut's perspective, might be the one whose origin is at his head, with the positive x_3-axis emanating from the top of his head, the positive x_2-axis emanating from the front of his chest, and the positive x_1-axis chosen so as to obtain a right-handed Cartesian coordinate system.

Suppose now that the astronaut has a small ball in his right hand, which he releases. When the astronaut releases the ball, he will observe it to be motionless in this frame of reference (i.e., relative to himself). Since no forces act on the ball, this is in accordance with Newton's first law. However, suppose that after releasing the ball, the astronaut ignites the engine on his spaceship, causing the spaceship to accelerate. Supposing that the astronaut is standing on the floor of the spaceship, with the spaceship accelerating upwards (i.e., along the positive x_3-axis in the astronaut's frame of reference), the astronaut will observe the ball to be *accelerating downwards* (i.e., in the $-x_3$ direction), even though there are no forces acting on it! This clearly violates the first law.

Consequently, we conclude that the law of inertia does *not* hold in an arbitrary reference frame; we call the distinguished class of frames in which it *does* hold *inertial frames*.

As an aside, it may seem surprising that one of the three fundamental "laws" of motion does *not*, in fact, hold for an arbitrary selection of reference frame. The reason for this is that Newton posulated the existence of *absolute* time and space; equivalently, he postulated the existence of a distinguished, canonical choice of inertial reference frame against which all motion could be measured, and all of his laws of motion were written with respect to accelerations and velocities as measured in this distinguished reference frame, which he often identified with the frame defined by the background of "fixed" stars.

This was a controversial claim, even in his day. For denote this distinguished frame by $\mathcal{A}$, and let $\mathfrak{F}$ be another frame, whose axes moves irrotationally and with constant translational velocity $v_{\mathfrak{F}}$ relative to the origin of $\mathcal{A}$. Since the first law holds in the distinguished frame $\mathcal{A}$, a particle having no forces acting upon it will be observed to move with constant velocity in the frame $\mathcal{A}$, and therefore with constant velocity in the frame $\mathfrak{F}$. Since no forces act upon the particle, and it is observed to move with constant velocity in $\mathcal{F}$, the first law holds in $\mathcal{F}$, and therefore $\mathcal{F}$ is an inertial frame. This shows that *any reference frame that moves irrotationally and with uniform translational velocity relative to an inertial frame is also an inertial frame.* Consequently, there are in fact *infinitely many* inertial frames, each of which moves irrotationally and with constant velocity relative to all the others. Furthermore, in *all* of these frames, a particle upon which no forces are acting will be observed to move with constant velocity. Consequently, even assuming that there *does* exist a distinguished inertial reference frame $\mathcal{A}$, *it is impossible to determine empirically which reference frame this is.* This principle is referred to as *Galilean relativity*, and was described by Galileo in 1632 in his *Dialogue Concerning the Two Chief World Systems.*

Thus, even though Newton postulated the existence of absolute time and space (and hence also absolute rest and motion), it was only possible to directly measure the *relative motion* of objects. As a result of the inability to empirically determine a distinguished frame in which absolute motion can be measured, modern physics rejects the existence of such a system outright, instead replacing it with the (equivalent) infinite class of inertial frames, in any one of which Newton's laws hold. Consequently, modern formulations of Newton's laws often list the First Law as:

First Law (modern synthesis):

There exists a distinguished class of reference frames, called *inertial ref-*

erence frames, in which the Second and Third Laws hold. An inertial reference frame is any reference frame in which a particle upon which no external forces act is observed to move with constant velocity. Any reference frame that moves irrotationally and with constant velocity relative to an inertial frame is again an inertial frame.

1.1.2.2 The Second Law

The Second Law describes the effect that applied forces have on the kinematics of a particle. Given a particle m with velocity v as measured in an inertial frame, we define that particle's *momentum*, customarily denoted by p, to be the quantity $p = mv$. Newton's Second Law then states that a force F applied to a particle induces a change in that particle's momentum according to the equation:

$$F = \frac{dp}{dt}. \tag{1.1}$$

In the case in which the particle's mass is constant as a function of time (which will always be the case for our purposes), the above equation can be rewritten as:

$$F = \frac{dp}{dt} = \frac{d}{dt}(mv) = m\frac{d}{dt}(v) = ma. \tag{1.2}$$

Note that, thus far, we have been using the term "force" without having properly defined it, relying instead upon an intuitive understanding of what is meant by the term. One can also view equation (1.1) as providing an *operational definition* of force: namely, a *force is any physical interaction that causes a particle to undergo a change in momentum* (or equivalently, change in velocity if the particle's mass remains constant in time).

The observant reader may notice that we have seemingly constructed circular definitions: we have defined an inertial frame as one in which particles appear to move with constant velocity in the absence of external forces (so that the definition of inertial reference frames depends upon forces), but we have also defined forces in terms of their ability to induce a change in momentum as measured in an inertial frame (so that the definition of force depends upon the definition of inertial frames). Equivalently, this seems to suggest that we can "make" any frame into an inertial frame by simply postulating the existence of forces that induce the observed motion of particles within that frame in accordance with equation (1.1), contradicting our earlier assertion that *not* all reference frames are inertial. The resolution of this apparent dilemma is the observation that forces arise as a result of *physical interactions*.

For example, let us return to the astronaut observing the ball after he has ignited the engine on his spaceship. The astronaut observes the ball accelerating towards the floor of the spaceship, and can hypothesize the existence of a force that would account for such an acceleration in accordance with equation (1.2). However, this hypothesized force is a *fictitious force*: it does not correspond to the physical interaction of any other body with the astronaut's ball. In contrast, *real forces* are those that *are* due to physical interactions between bodies (e.g., contact forces, electromagnetism, gravitation, etc.). Thus, *fictitious forces are absent* (*respectively, present*) *in a given frame of reference precisely when that frame is inertial* (*respectively, noninertial*).

Of course, one could further argue that we have merely moved the goal-posts: we introduced a distinction between fictitious and real forces in order to resolve the ambiguity of whether a frame is inertial, but at the same time introduced the undefined concept of "physical interaction" in the discussion of fictitious forces. Put more succinctly, how can one determine if a force acting on a given body is real or fictitious (equivalently, can be traced to a "physical interaction")?

We feign no hypothesis: questions of what constitute "forces" and "inertial frames" are actually quite subtle. Indeed, the question of how one could empirically distinguish inertial from noninertial frames was a central thread of the investigation that led Einstein to formulate the theories of special and general relativity (using, in part, thought experiments that were very similar to our example of the astronaut and his ball). Similarly, the question of how to properly define the concept of "force" is one which is closely related to the philosophy of science, including what it means for a "natural law" (of which Newton's Three Laws are, of course, particular instances) to have meaningful content, as opposed to being merely a tautology.

Clearly, these considerations (although quite interesting!) are far beyond the scope of this paper (although for a more in-depth discussion along these lines, the reader is referred to Section 12–1 of [4]). Perhaps the best answer is simply that you know a “physical interaction” when you see it. Physical theories are constructed as idealizations meant to explain empirically observed phenomena, including those physical interactions that we subsequently model as “forces” in the Newtonian sense. Thus, while this seeming lack of rigor may be unsatisfying to the pure mathematician, it is an unavoidable consequence of the messy business of trying to construct a mathematical framework that will enable scientists to model the empirically observed behavior of natural phenomena.

Consequently, since the primary purpose of this paper is to present mathematical formulations of mechanics, rather than an in-depth discussion of the nature of physical law, henceforward we shall simply blithely *assert* the existence of an infinite set of inertial frames. All coordinates assigned to systems that we model will be assumed to come from an inertial frame, and all forces described will be assumed real (rather than fictitious).

1.1.2.3 The Third Law

Finally, Newton's Third Law asserts that all forces come in pairs. Concretely, if two particles P_1 and P_2 interact in such a way that P_1 exerts a force F_{12} on P_2, then the Third Law asserts that P_2 exerts a corresponding force

F_{21} on P_1, and that

$$F_{21} = -F_{12}.$$

1.1.2.4 Summary of the modern synthesis of Newton's Three Laws

For easy reference, we summarize the modern synthesis of Newton's Three Laws below.

First Law

There exists a distinguished class of reference frames, called *inertial reference frames*, in which the Second and Third Laws hold. An inertial reference frame is any reference frame in which a particle upon which no external forces act is observed to move with constant velocity.

Second Law

Let P be a particle of mass m, and let v be its velocity and $p = mv$ its momentum as measured in an inertial frame. Suppose that P is acted upon by a force F as measured in this frame. Then F induces a change in P's momentum according to

$$F = \frac{dp}{dt}.$$

Third Law

Let P_1 and P_2 be two particles that interact in such a way that P_1 exerts a force F_{12} on P_2. Then P_2 exerts a corresponding force F_{21} on P_1, and

$$F_{21} = -F_{12}.$$

Remark 1.1.1. Henceforward, unless otherwise noted, whenever we describe a system using coordinates and forces, these coordinates and forces are always written with respect to an inertial frame, so that the Second and Third Laws hold.

1.1.3 Basic concepts in Newtonian mechanics, part one

We briefly review a few concepts in Newtonian mechanics that are commonly employed in the analysis of physical systems.

1.1.3.1 Impulse

Let P be a particle that is acted upon by a time-varying force $F(t)$. We call the integral of $F(t)$ over the interval $[t_0, t_1]$ the *impulse* of this action, denoted by $I(t_0, t_1)$:

$$I(t_0, t_1) = \int_{t_0}^{t_1} F(t)\, dt.$$

Denoting P's momentum by p, and using the Second Law, we can rewrite this as:

$$I(t_0, t_1) = \int_{t_0}^{t_1} F(t)\, dt = \int_{t_0}^{t_1} \frac{dp}{dt}(t)\, dt = p(t_1) - p(t_0).$$

This statement is referred to as the Impulse-Momentum Theorem.

Theorem 1.1.1 (Impulse-Momentum Theorem). *Let P be a particle that is acted upon by a time-varying force $F(t)$ over the time interval $[t_0, t_1]$. Then the impulse of this action is equal to the net change in P's momentum $p(t)$*

over this interval:

$$I(t_0, t_1) = \int_{t_0}^{t_1} F(t)\, dt = p(t_1) - p(t_0).$$

1.1.3.2 Mechanical work

Suppose now that a particle P travels along a path γ and is acted upon by a force F, where F is regarded as a function of P's position. The (*mechanical*) *work* W done by the force F acting on P as P traverses γ is defined as:

$$W = \int_\gamma F(s) \cdot ds. \tag{1.3}$$

Note the work W is the integral of the dot-product of the force vector $F(s)$ and the infinitesimal displacement ds along the curve γ, hence is a scalar.

1.1.3.3 Kinetic energy

Let P be a particle of mass m, and let v be its velocity. We define the *kinetic energy* K of P to be:

$$K = \frac{1}{2} m \|v\|^2. \tag{1.4}$$

Since momentum is related to velocity according to $p = mv$, we can equivalently write equation (1.4) as

$$\begin{aligned} K &= \frac{1}{2} m \|v\|^2 \\ &= \frac{1}{2} m \left\| \frac{1}{m} p \right\|^2 \\ &= \frac{\|p\|^2}{2m}. \end{aligned}$$

Now, suppose that a particle P of mass m is acted upon by a *total* force F_{total} (that is, F_{total} is the vector sum of *all* of the forces acting on P), resulting in a time-dependent position given by $x\colon [t_0, t_1] \to \mathbb{R}^3$. Then by definition, the work done by the force F_{total} acting on P over the time interval $[t_0, t_1]$ is given by:

$$W = \int_x F_{total} \cdot ds = \int_{t_0}^{t_1} F_{total} \cdot \dot{x}(t)\, dt. \tag{1.5}$$

But

$$\dot{x} = v,$$

so that $\dot{x}\, dt = v\, dt$. Consequently, for a particle of constant mass m, equation (1.5) can be rewritten as:

$$\begin{aligned} W &= \int_{t_0}^{t_1} F_{total} \cdot v\, dt \\ &= \int_{t_0}^{t_1} \frac{dp}{dt} \cdot v\, dt \\ &= \int_{t_0}^{t_1} \frac{d}{dt}(mv) \cdot v\, dt \\ &= \int_{t_0}^{t_1} m\frac{dv}{dt} \cdot v\, dt \\ &= \int_{v_0}^{v_1} mv \cdot dv \end{aligned} \tag{1.6}$$

where we have used the Second Law to write the force F_{total} as the rate-of-change of the particle's momentum with respect to time, and where v_0 and v_1 denote the particle's velocities at times t_0 and t_1, respectively. Now

$$\|v\|^2 = v \cdot v,$$

so that

$$d\left(\|v\|^2\right) = d\,(v \cdot v) = v \cdot dv + dv \cdot v = 2(v \cdot dv) \tag{1.7}$$

by the product rule. Equations (1.6) and (1.7) together show that

$$\begin{aligned} W &= \int_{v_0}^{v_1} m\,d\left(\frac{\|v\|^2}{2}\right) \\ &= \frac{1}{2}m\|v_1\|^2 - \frac{1}{2}m\|v_0\|^2 \\ &= K_1 - K_0, \end{aligned} \tag{1.8}$$

where K_0 and K_1 denote the kinetic energy of the particle P at times t_0 and t_1, respectively. Equation (1.8) can be summarized in the Work-Energy Theorem.

Theorem 1.1.2 (Work-Energy Theorem). *Let P be a particle, and let F_{total} denote the vector sum of all forces acting upon P. Then the total work W done by F_{total} on P over a given interval in time is equal to the net change in P's kinetic energy over that interval:*

$$W = K_1 - K_0.$$

In particular, the kinetic energy K of a particle P at any time is precisely the amount of work needed to accelerate that particle from rest to its instantaneous velocity at that time.

1.1.3.4 Conservative forces and potential energy

Let F denote a force acting on a particle P. It often happens in applications that the value of F at each moment in time depends only upon P's position. In that case, we can represent F as a vector field on $\mathbb{R}^3$; this vector field is then called a *force field*. If F has the additional property that it can be written as:

$$F = -\nabla V$$

for some scalar function $V\colon \mathbb{R}^3 \to \mathbb{R}$, then F is called a *conservative force field* or a *conservative force*, and the scalar function V is called the *potential energy* or *potential*.

Conservative forces have the property that the work they do on a particle P as it traverses a path γ depends only upon the path's endpoints, and not upon the path itself. For let F be a conservative force with potential V, fix $x_0, x_1 \in \mathbb{R}^3$, and let $\gamma\colon [t_0, t_1] \to \mathbb{R}^3$ be any path with $\gamma(t_0) = x_0$ and $\gamma(t_1) = x_1$. Then the work W done on a particle P as it traverses the path γ may be computed according to equation (1.3):

$$\begin{aligned} W &= \int_\gamma F \cdot ds \\ &= \int_{t_0}^{t_1} F(\gamma(t)) \cdot \dot{\gamma}(t)\, dt \\ &= \int_{t_0}^{t_1} -\nabla V(\gamma(t)) \cdot \dot{\gamma}(t)\, dt \\ &= -\int_{t_0}^{t_1} \nabla V(\gamma(t)) \cdot \dot{\gamma}(t)\, dt \\ &= -\left[V(\gamma(t))\right]_{t_0}^{t_1} \\ &= -\left[V(\gamma(t_1)) - V(\gamma(t_0))\right] \\ &= -\left[V(x_1) - V(x_0)\right] \end{aligned} \tag{1.9}$$

where we have used the Fundamental Theorem of Calculus in passing from line 4 to line 5. Notice that the path γ does *not* appear in the final line of equation (1.9); rather, the work done by F as the particle traverses γ depends only upon the endpoints x_0 and x_1, and is determined by the difference of the values of the potential function V at those points. In particular, this shows that *a conservative force does no net work on a particle when it traverses a closed path*.

More generally, let $\{P_i \mid 1 \leq i \leq n\}$ be a set of particles; then we can represent the configuration of this system (i.e., the position of each particle P_i) by a vector $x \in \mathbb{R}^{3n}$ by concatenating the positions of each of the individual particles (and similarly for the system's velocity, acceleration, momentum, etc.). Suppose now that we have a *set* of forces $\{F_j \mid 1 \leq j \leq m\}$ that act on the system, and that each force F_j depends only upon the configuration x of the system. Then each force can be modeled as a vector $F_j(x) \in \mathbb{R}^{3n}$, again by concatenating the 3-vectors representing the force exerted upon each individual particle in the system. By analogy with the single-particle case, we call such a set of forces $\{F_j\}$ *conservative* if there exists a function $V\colon \mathbb{R}^{3n} \to \mathbb{R}$, called the *potential* (*energy*), such that

$$\sum_{j=1}^{m} F_j(x) = -\nabla V(x).$$

The same argument as was used in equation (1.9) then shows that the work done by this set of forces along any path depends only on the endpoints of that path, and that the net work done on the system around any closed curve is zero.

1.1.3.5 Center of mass and conservation of momentum

Let $\{P_i \mid 1 \leq i \leq n\}$ be a set of particles with constant masses, and let m_i, x_i, v_i, and a_i denote the mass, position, velocity, and acceleration of the particle P_i for all $1 \leq i \leq n$. Regarding each 3-tuple x_i as a vector, we define the *center of mass* of this system of particles, denoted x_{CM}, as

$$x_{CM} = \frac{1}{M}\sum_{i=1}^{n} m_i x_i, \tag{1.10}$$

where

$$M = \sum_{i=1}^{n} m_i$$

is the total mass of the system of particles.

Similarly, let v_{CM} and a_{CM} denote the velocity and acceleration of the center of mass. Multiplying both sides of equation (1.10) by M and differentiating with respect to time produces

$$Mv_{CM} = \sum_{i=1}^{n} m_i v_i = \sum_{i=1}^{n} p_i, \tag{1.11}$$

where p_i denotes the momentum of the ith particle. Differentiating again produces:

$$Ma_{CM} = \sum_{i=1}^{n} m_i a_i = \sum_{i=1}^{n} F_i, \tag{1.12}$$

where F_i denotes the total force acting on each particle. We can decompose the total force F_i acting on each particle P_i as

$$F_i = F_i^{int} + F_i^{ext}$$

where F_i^{int} is the total force on P_i due to interactions with the other particles in the system, and F_i^{ext} is the remaining force (that is, F_i^{ext} is the sum of external forces acting on the particle P_i). Thus, we can rewrite equation (1.12) as

$$Ma_{CM} = \sum_{i=1}^{n} F_i^{int} + \sum_{i=1}^{n} F_i^{ext}. \tag{1.13}$$

Now, let F_{ij} be the force on particle P_i due to its interaction with particle P_j. Then the total force F_i^{int} acting on particle P_i is

$$F_i^{int} = \sum_{j=1}^{n} F_{ij}.$$

Substitution into equation (1.13) gives

$$\begin{aligned} Ma_{CM} &= \sum_{i=1}^{n} F_i^{int} + \sum_{i=1}^{n} F_i^{ext} \\ &= \left(\sum_{i=1}^{n} \sum_{j=1}^{n} F_{ij} \right) + \sum_{i=1}^{n} F_i^{ext} \end{aligned} \tag{1.14}$$

However, by Newton's Third Law, $F_{ij} = -F_{ji}$ for all $1 \leq i, j \leq n$; consequently, the terms in the left-hand summation above cancel pairwise, and equation (1.14) reduces to

$$Ma_{CM} = \sum_{i=1}^{n} F_i^{ext}. \tag{1.15}$$

Now equation (1.11) shows that the total momentum of the system of particles is equal to the product of the velocity of the center of mass and the total mass of the system; similarly, equation (1.15) shows that the product of the acceleration of the center of mass with the total mass of the system is equal to the sum of external forces acting on the system of particles. Conceptually, we regard this as asserting that *the system of particles behaves as though the entire mass of the system were concentrated into a single particle located at the center of mass of the system.*

As a special case of the above, suppose that there are no external forces acting on the system. Then

$$Ma_{CM} = M\dot{v}_{CM} = 0;$$

consequently, the quantity

$$Mv_{CM} = \sum_{i=1}^{n} m_i v_i = \sum_{i=1}^{n} p_i$$

in equation (1.11) is constant. This property is called *conservation of momentum.*

Theorem 1.1.3 (Conservation of momentum). *Let* $\{P_i \mid 1 \le i \le n\}$ *be a set of particles, and let* p_i *be the momentum of* P_i *for all* $1 \le i \le n$*. If there are no external forces acting on these particles* (*i.e., the only forces acting on them are their mutual interactions*)*, then the total momentum*

$$P_{total} = \sum_{i=1}^{n} p_i$$

is constant.

1.1.3.6 Conservation of energy

In addition to conservation of momentum, there is another important conservation law: the *conservation of energy*. Suppose that the only forces acting on a particle P are conservative. Let K denote P's kinetic energy, let F denote the sum of forces acting upon P, and let V denote the potential corresponding to F. Let F act on P over some interval $[t_0, t_1]$ of time, and let $K_1 = K(t_1)$, $K_0 = K(t_0)$, $V_1 = V(t_1)$, and $V_0 = V(t_0)$. By equation (1.8), the total work W done by F on P over $[t_0, t_1]$ is equal to the change in P's kinetic energy:

$$W = K_1 - K_0.$$

However, equation (1.9) also shows that the change in potential over the same period of time is

$$W = V_0 - V_1.$$

Substitution for the work W then shows that

$$K_1 + V_1 = K_0 + V_0,$$

which proves the following.

Theorem 1.1.4 (Conservation of energy). *Let P be a particle, and suppose that the only forces acting upon P are conservative. Let K denote P's kinetic energy, let F denote the sum of forces acting upon P, and let V denote the potential corresponding to F. Then the total energy*

$$E = K + V$$

is constant.

1.1.4 Rigid bodies and rotational dynamics

As written, Newton's Laws describe the motion of *particles*, massive objects that have no spatial extent, and that are completely described by their position in space. Of course, no macroscopic real-world objects are *actually* particles, but in many applications, it suffices to approximate them as such. This might be the case, for example, when the spatial extent of these objects is small relative to the other lengths appearing in the problem (as is the case, for example, when modeling a spacecraft's orbit), when the object has a high degree of symmetry, or when the orientation of an object is not considered important within the context of the analysis being performed.

However, oftentimes one *would* like to be able to model both the position *and* the orientation of an object; in this case, Newton's laws, as written, are insufficient. However, by modeling objects as aggregations of particles, one can use Newton's laws to derive equations governing the rotational dynamics of *rigid bodies*: objects that have some spatial extent, a fixed geometric configuration (i.e., do not deform in response to externally applied forces), and a fixed distribution of mass.

1.1.4.1 Angular position, velocity, and acceleration

When a rigid body rotates about a fixed axis A, it is generally convenient to fix a coordinate system whose x_3-axis coincides with the axis of rotation; in that case, as the body rotates, every point in that body traces out a circle in a plane lying parallel to the x_1x_2-axis. If we fix a particular point P in the body, we can define the body's *angular position about* A, denoted θ, to be the angular displacement of the right-handed rotation about the positive x_3-axis sending the positive x_1-axis onto the line segment connecting the origin to the projection of P into the x_1x_2-plane. Analogously to the case of particle kinematics, we can then define the body's *angular velocity about* A, denoted ω, according to

$$\omega = \dot{\theta},$$

and the *angular acceleration about* A, denoted by α, as

$$\alpha = \dot{\omega} = \ddot{\theta}.$$

Remark 1.1.2. We shall always use radian angle measure.

It will be convenient for us in the future to be able to specify the axis of rotation and the magnitudes of the angular velocity and acceleration simulaneously. To that end, we can replace the *scalar* quantities ω and α with *vector* quantities, where the *direction* of the vector specifies the axis of rotation, and the *magnitude* of the vector specifies the value of the angular velocity/acceleration about that axis, as defined above.

1.1.4.2 Torque and moments of inertia for rotating point masses

Consider the following experiment: a particle P of mass m is attached to one end of an idealized massless rod of length r, and the rod's other end is fixed to the origin in such a way that the rod can rotate freely about that end in the x_1x_2-plane; then P is constrained to lie on a circle of radius r in the x_1x_2-plane, centered on the origin. Let R denote the vector whose initial point is the origin and whose terminal point is P, so that $r = \|R\|$ (see Figure 1.1).

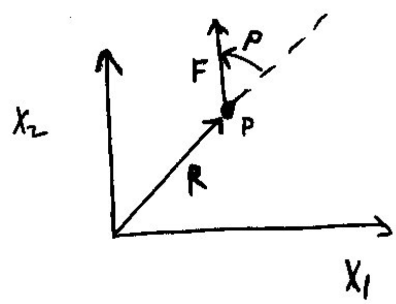


Figure 1.1: Demonstration of torque.

Now consider applying some force F in the x_1x_2-plane to the particle P, and suppose that the counterclockwise angle from the vector R to F is ρ. Then F decomposes as the sum of a force F_{rad} acting in the radial direction (i.e., collinearly with the rod) and a force F_{tan} acting tangentially; these forces are given by:

$$F_{rad} = \|F\| \cos(\rho)\hat{R}$$

$$F_{tan} = \|F\| \sin(\rho)(\hat{x}_3 \times \hat{R})$$

where here $\hat{R}$ denotes the unit vector in the direction of R, $\hat{x}_3$ denotes the unit vector in the direction of the positive x_3-axis (out of the page), and "$\times$" denotes the usual vector product operation in $\mathbb{R}^3$.

We now make a few observations. Since the rod is assumed rigid and fixed to the origin, the radial force F_{rad} acting on P is balanced by a radial force of either tension or compression exerted on P by the rod. Consequently, the application of a radially-directed force does *not* induce a a rotation of the rod-mass system about the x_3-axis.

On the other hand, the tangential force F_{tan} induces an instantaneous tangential acceleration a_{tan} given by

$$a_{tan} = \frac{1}{m} F_{tan}$$

in accordance with Newton's Second Law (equation (1.2)). Since we use radian angle measure, this corresponds to an instantaneous counterclockwise angular acceleration of

$$\alpha = \hat{R} \times \left(\frac{1}{r} \cdot a_{tan} \right)$$

about the x_3-axis. Equivalently:

$$mr\alpha = \hat{R} \times F_{tan}$$

or

$$\begin{aligned} mr^2\alpha &= r\hat{R} \times F_{tan} \\ &= R \times F_{tan}. \end{aligned} \tag{1.16}$$

Since

$$R \times F_{rad} = 0$$

as F_{rad} and R are collinear, then

$$R \times F = R \times (F_{rad} + F_{tan}) = R \times F_{rad} + R \times F_{tan} = R \times F_{tan},$$

and therefore equation (1.16) can be rewritten as:

$$mr^2\alpha = R \times F. \tag{1.17}$$

Equation (1.17) is therefore a measure of the ability of the force F to induce an angular acceleration of the mass-rod system about the x_3-axis.

With this as motivation, we make the following definitions. Fix an axis of rotation A, and let P be a particle of mass m. Let r denote the vector perpendicular to A originating on A and terminating at P's position. Let F be a force applied to P, and let F_{rot} be the component of F lying in the plane perpendicular to A. Then the *torque about* A corresponding to F, denoted τ, is given by:

$$\tau = r \times F_{rot}. \tag{1.18}$$

Similarly, fix an axis of rotation A, and let P be a particle of mass m. As before, let r be the unique displacement vector perpendicular to A originating on A and terminating at the particle P. Then the *moment of interia of* P *about* A, denoted I, is given by

$$I = m\|r\|^2. \tag{1.19}$$

With these definitions,

$$\begin{aligned}\tau &= r \times F_{rot} \\ &= r \times ma \\ &= r \times m(a_{rad} + a_{tan}) \\ &= r \times ma_{rad} + r \times ma_{tan} \\ &= r \times ma_{tan}\end{aligned} \tag{1.20}$$

where a is the acceleration of P due to the application of F_{rot}, a_{tan} denotes the component of a perpendicular to r, and a_{rad} denotes the component collinear with r. Since we measure angles in radians, then the instantaneous angular acceleration α of P about A satisfies:

$$\|r\|\alpha = \hat{r} \times a_{tan},$$

so that

$$m\|r\|^2\alpha = m\|r\|\hat{r} \times a_{tan}$$

or

$$I\alpha = r \times ma_{tan}. \tag{1.21}$$

Equations (1.20) and (1.21) show that

$$\tau = I\alpha; \tag{1.22}$$

this describes how torque, angular acceleration, and moments of inertia are related, and is the analogue of Newton's Second Law in the case of rotational dynamics.

1.1.4.3 Torques and moments of inertia for generic rigid bodies

Given a generic rigid body with a fixed mass distribution rotating about a fixed axis, one can compute its rotational dynamics by treating it as an aggregation of particles and applying the results of the previous section.

Specifically, let B be a rigid body with a fixed mass distribution $\rho(x)$, and suppose that B rotates about a fixed axis A. For each point $x \in B$, let $r(x)$ denote the unique vector perpendicular to the axis of rotation A originating on A and terminating at the point x. Then one can compute the *moment of inertia of* B *about* A according to

$$I = \int_B \rho \|r\|^2 \, dV. \tag{1.23}$$

Note that in equation (1.23), we implicitly regard B as being composed of infinitesimally small particles of volume dV, each of whose masses dm is then given by $\rho \, dV$, and then "sum" (via integration) the moments of inertia of each individual particle to obtain the total moment of inertia of B.

Now, suppose that B rotates about the axis A with some angular acceleration α. Since B is rigid, each infinitesimal particle of B accelerates with the same angular acceleration α. By equation (1.22), this corresponds to the application of a torque

$$d\tau = \alpha m \|r\|^2 \, dm = \alpha \rho \|r\|^2 \, dV$$

to each such particle. Summing these torques over all of the infinitesimal particles in B produces the total torque acting on B:

$$\begin{aligned}\int_B d\tau &= \int_B \alpha\rho\|r\|^2\, dV \\ \tau &= \alpha \int_B \rho\|r\|^2\, dV \\ &= I\alpha.\end{aligned}$$

Therefore, we see that $\tau = I\alpha$ also holds in the case of extended rigid bodies.

As an example, consider computing the moment of inertia of a rod of length l and mass m rotating about an axis passing through one of its ends, assuming that it has a uniform mass distribution. Since its mass distribution is uniform, its linear mass density λ is given by $\lambda = m/l$. Regarding the axis of rotation as the origin of a coordinate system extending lengthwise down the rod, its moment of inertia about that axis can then be computed according to equation (1.23) as:

$$\begin{aligned}I_{end} &= \int_0^l x^2\, dm \\ &= \int_0^l x^2\lambda\, dx \\ &= \frac{1}{3}\lambda l^3 \\ &= \frac{1}{3}ml^2.\end{aligned} \tag{1.24}$$

On the other hand, if we consider the axis of rotation to be the center of the rod, the corresponding computation is

$$\begin{aligned} I_{center} &= \int_{-l/2}^{l/2} x^2 \lambda \, dx \\ &= \frac{1}{12} \lambda l^3 \\ &= \frac{1}{12} m l^2. \end{aligned} \tag{1.25}$$

1.1.4.4 Centers of mass for generic rigid bodies

Similarly, given a rigid body B with a fixed mass distribution ρ, we can compute its center of mass as:

$$x_{CM} = \frac{1}{M} \int_B \rho x \, dV,$$

where

$$M = \int_B \rho \, dV$$

is the total mass of B.

1.1.5 Basic concepts in Newtonian dynamics, continued

Torque and moment of inertia play the same role in the study of rotational motion that force and mass do in translational motion. Similarly, the concepts of momentum and kinetic energy also have straightforward analogues in the context of rotational motion.

1.1.5.1 Angular momentum

Fix an axis of rotation A, and let P be a particle of mass m. Let a force F act on P, let F_{rot} be the component lying in the plane perpendicular to A, and let p_{rot} be the component of P's momentum lying in the plane perpendicular to A. Then by Newton's Second Law:

$$F_{rot} = \frac{dp_{rot}}{dt}.$$

Let r be the vector perpendicular to A originating on A and terminating on the particle P. Taking the cross product of both sides of the above equation with the vector r produces:

$$r \times F_{rot} = r \times \frac{dp_{rot}}{dt},$$

or

$$\tau = r \times \frac{dp_{rot}}{dt}. \tag{1.26}$$

Now

$$\frac{dr}{dt} = v_{rot}$$

by definition, where v_{rot} is the component of P's velocity lying in a plane perpendicular to A; therefore, v_{rot} and p_{rot} are parallel. Consequently,

$$\frac{dr}{dt} \times p_{rot} = 0.$$

Thus, we can write equation (1.26) as

$$\begin{aligned} \tau &= r \times \frac{dp_{rot}}{dt} + \frac{dr}{dt} \times p_{rot} \\ &= \frac{d}{dt}\left[r \times p_{rot}\right]. \end{aligned} \tag{1.27}$$

We define the quantity $r \times p_{rot}$ appearing on the right-hand side of equation (1.27) to be the *angular momentum of P about A*, denoted L. Now

$$\begin{aligned} r \times p_{rot} &= r \times mv_{rot} \\ &= r \times m(v_{tan} + v_{rad}) \\ &= r \times mv_{tan} + r \times mv_{rad} \\ &= r \times mv_{tan}, \end{aligned} \tag{1.28}$$

and

$$\|r\|\omega = \hat{r} \times v_{tan},$$

since we use radian angle measure. Therefore

$$m\|r\|^2\omega = m\|r\|\hat{r} \times v_{tan},$$

or

$$I\omega = r \times mv_{tan},$$

so that equation (1.28) can be rewritten as:

$$L = I\omega. \tag{1.29}$$

This equation provides an alternative definition of angular momentum in terms of moment of inertia and angular velocity, and is the rotational analogue of $p = mv$.

Equation (1.27) provides the relation between angular momentum and applied torque:

$$\tau = \frac{dL}{dt}.$$

i.e., *torque is equal to the rate of change of angular momentum.*

As in the case of translational momentum, this leads to the following conservation law.

Theorem 1.1.5 (Conservation of angular momentum). *In the absence of externally applied torques, the total angular momentum of a system is constant.*

1.1.5.2 Rotational kinetic energy

Finally, we can also compute the kinetic energy associated with a body's rotational motion. Let a rigid body B with a fixed mass distribution ρ rotate about a fixed axis A. Given a point $P \in B$, let $r(P)$ denote the vector perpendicular to A originating on A and terminating on P. Since B is rigid, then each point of B rotates about A with the same instantaneous angular velocity ω. Given an individual particle $P \in B$, this corresponds to an instantaneous tangential speed of

$$\|v_P\| = \|r(P)\| \cdot \|\omega\|.$$

Consequently, the instantaneous kinetic energy of particle P's motion is

$$\frac{1}{2} m_P \|v_P\|^2 = \frac{1}{2} \rho(P) dV \|r(P)\|^2 \|\omega\|^2.$$

Thus, the total kinetic energy associated with B's rotation is:

$$\begin{aligned} T &= \int_B \frac{1}{2} \rho \, dV \|r\|^2 \|\omega\|^2 \\ &= \frac{1}{2} \|\omega\|^2 \int_B \rho \|r\|^2 \, dV \\ &= \frac{1}{2} I \|\omega\|^2, \end{aligned} \tag{1.30}$$

where I is B's moment of inertia about A. Again, this is the rotational analogue of the computation of kinetic energy for translational motion of a particle.

1.1.6 Examples

Given a description of the initial configuration and velocity of any physical system, and knowledge of the forces acting upon it, Newton's laws completely describe the future evolution of the system in time. Indeed, according to Newton's Second Law, the initial configuration and velocity of the system together with a description of the forces acting upon it form an initial value problem for a second-order system of ordinary differential equations; the solution of this initial-value problem is precisely the curve representing the future time evolution of the system.

1.1.6.1 Firing a cannon

As an example, suppose that we wish to determine the path of a cannonball of mass m shot out of a cannon at time $t = 0$ at an angle of θ radians above the horizontal with an initial speed v. Neglecting the effects of air resistance, the only force acting upon the cannonball is the force of gravity, which acts straight downwards with a magnitude of mg. If we set up a coordinate system such that x_1 denotes the horizontal distance from the cannon and x_2 the vertical distance, then applying Newton's Second Law produces produces:

$$\begin{cases} m\ddot{x}_1(t) = 0 \\ m\ddot{x}_2(t) = -mg. \end{cases}$$

Integrating twice and applying the initial conditions

$$\begin{cases} \dot{x}_1(0) = v\cos(\theta) \\ \dot{x}_2(0) = v\sin(\theta) \\ x_1(0) = 0 \\ x_2(0) = 0 \end{cases}$$

gives the solution

$$\begin{cases} x_1(t) = v\cos(\theta)t \\ x_2(t) = -\frac{1}{2}gt^2 + v\sin(\theta)t, \end{cases}$$

corresponding to the well-known observation that a ballistic projectile follows a parabolic path in the absence of air resistance.

1.1.6.2 Bead on a wire

While Newton's laws provide a complete description of all mechanical processes, they are oftentimes difficult or impossible to apply to a given problem as written. One class of problems in which this holds is the case of *constrained motion.*

For example, we might consider a bead traveling along a wire under the influence of gravity. Consider a circular track of radius R as shown in Figure 1.2.

In order to model the motion of the bead as a function of time using Newton's equations, we would need to be able to write down all of the forces acting on the bead. Clearly, the bead is constrained to travel along the wire;

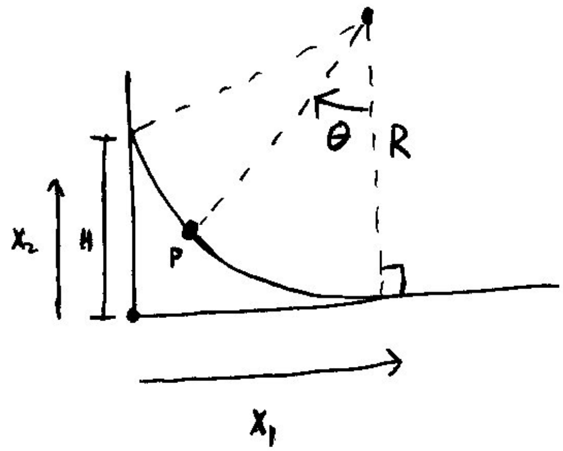


Figure 1.2: A bead on a circular track.

therefore, the wire must *exert a force* on the bead, called the *constraint force*, in order to keep it traveling along this path. However, the force exerted by the wire on the bead will, in general, depend upon the bead's position on the wire *and* the speed at which the bead is traveling. Thus, there is no obvious convenient way to write down the force of the wire acting on the bead (as was the case for the force of gravity acting on the canonball in the previous example).

However, we can make two key observations. First, we note that *the bead's velocity vector is always tangent to the wire* (since the bead is constrained to travel along the wire). Secondly, we also observe that the constraint force exerted by the wire on the bead (to keep the bead traveling along the wire) *always acts along the normal to the wire.* For if this were not the case, then some component of the constraint force would necessarily act tangentially to the wire, thereby accelerating the bead tangentially along the wire; this does not make physical sense. Since the constraint force and the bead's instantaneous velocity vector (hence infinitesimal displacement vector) are always orthogonal, we see that *the constraint force does no work on the bead.* Since the only other force acting on the bead is the force of gravity (which is conservative), *the total mechanical energy of the bead is conserved.*

We can compute the total mechanical energy E of the bead as the sum of its kinetic and gravitational potential energies (K and V, respectively):

$$\begin{aligned} K &= \frac{1}{2}m\|v\|^2, \\ V &= mgx_2. \end{aligned} \tag{1.31}$$

We also observe that since the bead is constrained to lie along the track, we can completely characterize its motion using only the angle θ as shown in Figure 1.2. From the geometry of the track, we see that

$$\begin{aligned} \|v\| &= R\dot{\theta} \\ x_2 &= R(1 - \cos(\theta)); \end{aligned} \tag{1.32}$$

substituting these values into equation (1.31) shows that

$$E = K + V = \frac{1}{2}mR^2\dot{\theta}^2 + mgR(1 - \cos(\theta)). \tag{1.33}$$

This is a somewhat complicated-looking first-order differential equation. However, we can actually simplify it somewhat by differentiating again with respect to time:

$$\begin{aligned}\frac{d}{dt}E &= \frac{d}{dt}\left[\frac{1}{2}mR^2\dot{\theta}^2 + mgR(1-\cos(\theta))\right]\\ 0 &= mR^2\dot{\theta}\ddot{\theta} + mgR\sin(\theta)\dot{\theta} \\ &= \dot{\theta}\left(mR^2\ddot{\theta} + mgR\sin(\theta)\right)\end{aligned} \tag{1.34}$$

(here we have used conservation of energy to show that the time derivative of the left-hand side is zero). Since the angular velocity of the system is obviously not identically zero, the quantity in parentheses on the right-hand side of the above equation *must* be identically zero, and therefore (dividing through by the constant $mR^2 \neq 0$ and rearranging):

$$\ddot{\theta} = -\frac{g}{R}\sin(\theta). \tag{1.35}$$

The fact that the bead is initially at rest then implies that

$$\dot{\theta}(0) = 0, \tag{1.36}$$

and the initial configuration of the system shows that

$$\theta(0) = \arccos\left(1 - \frac{H}{R}\right). \tag{1.37}$$

Thus, equations (1.35), (1.36), and (1.37) together define the second-order initial value problem:

$$\begin{cases}\ddot{\theta} = -\frac{g}{R}\sin(\theta)\\ \dot{\theta}(0) = 0\\ \theta(0) = \arccos\left(1-\frac{H}{R}\right)\end{cases} \tag{1.38}$$

whose solution $\theta(t)$ completely describes the motion of the bead according to

$$\begin{aligned}x_1(t) &= R\left(\sin(\theta(0)) - \sin(\theta(t))\right)\\ x_2(t) &= R\left(1 - \cos(\theta(t))\right).\end{aligned} \tag{1.39}$$

Observe that at no point in these computations did we compute *any* of the actual forces acting on the bead. Instead, we introduced an auxiliary variable θ whose definition implicitly encoded the fact that the bead was constrained to lie on the circular track, wrote the system's total energy as a function of θ, and then used conservation of energy to construct an initial value problem (1.38) whose solution can be mapped back onto the variables of interest (1.39).

It is a general theme in physics that conservation laws like the conservation of energy and momentum provide powerful tools with which to simplify the analysis of complex systems.

1.2 Lagrangian mechanics

In the previous section we introduced the Newtonian formulation of mechanics, in which the time evolution of a system is described in terms of the forces acting on that system. In this section, we introduce *Lagrangian mechanics*, a reformulation of Newtonian mechanics that describes the evolution of a system in terms of its kinetic and potential energies. While Lagrangian mechanics is formally equivalent to Newtonian mechanics, it has two significant advantages: it more readily allows the introduction of kinematic constraints, and it is, in general, much easier to compute kinetic and potential energies than it is to compute the forces acting on a system, as the energies can often be obtained directly from a description of the system's geometry.

1.2.1 Lagrangian mechanics for unconstrained conservative systems

Consider a system of particles $\{P_i \mid 1 \leq i \leq n\}$ being acted upon by a set of conservative forces whose sum is F. Denote the mass, position, velocity, and momentum of P_i by m_i, x_i, v_i, and p_i, respectively. Similarly, denote by x the concatenation of the n 3-vectors x_i, so that x gives the configuration of the entire system:

$$x = (x_1, \ldots, x_n) \in \mathbb{R}^{3n}.$$

Then the total kinetic energy K of the system is given by:

$$\begin{aligned} K &= \sum_{i=1}^{n} \frac{1}{2} m_i \|\dot{x}_i\|^2 \\ &= \frac{1}{2} \sum_{i=1}^{n} m_i \left((\dot{x}_i)_1^2 + (\dot{x}_i)_2^2 + (\dot{x}_i)_3^2 \right) \end{aligned}$$

so that

$$\frac{\partial K}{\partial \dot{x}_i} = m_i \left((\dot{x}_i)_1,\ (\dot{x}_i)_2,\ (\dot{x}_i)_3 \right) = m_i v_i = p_i. \tag{1.40}$$

Since the total force F acting on the system is conservative, it can be described by a potential energy function $V \colon \mathbb{R}^{3n} \to \mathbb{R}$, and the force acting on the ith particle P_i is given by

$$F_i = -\frac{\partial V}{\partial x_i}. \tag{1.41}$$

Observe that equation (1.40) gives P_i's momentum, and equation (1.41) gives the total force acting on P_i. By Newton's Second Law:

$$F_i = \frac{dp_i}{dt},$$

so

$$-\frac{\partial V}{\partial x_i} = \frac{d}{dt}\left(\frac{\partial K}{\partial \dot{x}_i}\right), \tag{1.42}$$

which implies that

$$-\frac{\partial V}{\partial x} = \frac{d}{dt}\left(\frac{\partial K}{\partial \dot{x}}\right) \tag{1.43}$$

(simply by equating coordinates over each of the n particles in equation (1.42)).

Thus, equation (1.43) gives a version of Newton's Second Law in terms of the system's kinetic and potential energies. We can encode this data in a single function $\mathfrak{L}$, called the *Lagrangian*:

$$\mathfrak{L}(x, \dot{x}) = K(\dot{x}) - V(x).$$

Then

$$-\frac{\partial V}{\partial x} = \frac{\partial \mathfrak{L}}{\partial x}, \quad \frac{\partial K}{\partial \dot{x}} = \frac{\partial \mathfrak{L}}{\partial \dot{x}},$$

so that equation (1.43) can be restated as

$$\frac{\partial \mathfrak{L}}{\partial x} = \frac{d}{dt}\left(\frac{\partial \mathfrak{L}}{\partial \dot{x}}\right). \tag{1.44}$$

This leads to the Lagrangian formulation of mechanics, as stated below.

Theorem 1.2.1 (Lagrange's equations for unconstrained conservative systems). *Let* $\{P_i \mid 1 \leq i \leq n\}$ *be a system of particles acted upon by conservative forces with corresponding potential function* V*, let* K *denote the system's kinetic energy, and define the Lagrangian*

$$\mathfrak{L}(x, \dot{x}) = K(x, \dot{x}) - V(x). \tag{1.45}$$

Then the time evolution $x(t)$ of the system is the solution to Lagrange's equations of motion:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right) = \frac{\partial \mathcal{L}}{\partial x}. \tag{1.46}$$

1.2.2 Lagrangian mechanics for conservative systems with holonomic constraints

Thus far, we have shown how to reformulate Newton's Laws in terms of kinetic and potential energies, but it is not clear what, if anything, is to be gained from so doing. One of the advantages of the Lagrangian formulation is that it enables one to incorporate kinematic constraints on mechanical systems in an elegant way through the use of *generalized coordinates*.

1.2.2.1 Holonomic constraints and generalized coordinates

Consider now a system of n particles $\{P_i \mid 1 \leq i \leq n\}$ acted upon by a conservative force F, and suppose that the system is subject to *holonomic constraints*: constraints on the allowable configuration of the system that can be written as:

$$c_k(x_1, \ldots, x_n) = 0 \tag{1.47}$$

for smooth functions $c_k \colon \mathbb{R}^{3n} \to \mathbb{R}$. In many cases (indeed, in "generic" physical systems) the zero locus of a set of z constraints of the form (1.47) is a manifold of dimension $m = 3n - z$, and can be parametrized by a set of m

generalized coordinates through (possibly time-dependent) functions:

$$
\begin{aligned}
x_1 &= x_1(t, q_1, \ldots, q_m) \\
x_2 &= x_2(t, q_1, \ldots, q_m) \\
&\vdots \\
x_n &= x_n(t, q_1, \ldots, q_m).
\end{aligned}
\tag{1.48}
$$

1.2.2.2 Derivation of Lagrange's equations for conservative systems with holonomic constraints

We shall now derive a version of Lagrangian mechanics for the use of generalized coordinates. Let the mass, position, velocity, and acceleration of P_i be denoted by m_i, x_i, v_i, and a_i, respectively, and let $Q \subseteq \mathbb{R}^{3n}$ denote the manifold determined by the holonomic constraints (1.47). We define a *virtual displacement* δx to be an element of the tangent bundle $TQ \subseteq T\mathbb{R}^{3n}$; equivalently, a virtual displacement δx can be regarded as a differential dx that is consistent with the holonomic constraints (1.47).

At each moment in time, each particle P_i has some set of forces acting on it, whose sum we shall denote F_i^{total}. We can decompose this sum of forces into a component F_i corresponding to the *conservative force* acting on the system and a component R_i corresponding to the *constraint forces* acting to keep the system confined to the manifold defined by the holonomic constraints (1.47):

$$
F_i^{total} = F_i + R_i. \tag{1.49}
$$

Now, by Newton's Second Law,

$$F_i^{total} = m_i a_i. \tag{1.50}$$

Given any virtual displacement δx of the system, let δx_i be the component of that displacement corresponding to particle P_i's motion. Then

$$(F_i^{total} - m_i a_i) \cdot \delta x_i = 0$$

by equation (1.50). Substituting equation (1.49) into the above equation produces:

$$(F_i - a_i m_i) \cdot \delta x_i + R_i \cdot \delta x_i = 0. \tag{1.51}$$

The important observation here is that *constraint forces do no work on virtual displacements*; for constraint forces act to keep the system of particles confined to the manifold Q (i.e., they only act *orthogonally* to Q), while virtual displacements, by definition, are always *tangent* to Q. Consequently, $R_i \cdot \delta x_i = 0$ always, and therefore equation (1.51) can be reduced to

$$(F_i - a_i m_i) \cdot \delta x_i = 0; \tag{1.52}$$

this is the version of Newton's Second Law that we will need for the following derivation. Note that equation (1.52) does *not* imply that $F_i - m_i a_i = 0$; indeed, the deviation from zero is precisely the contribution due to constraint forces acting on P_i. Rather, (1.52) states that the *component of P_i's acceleration tangent to the configuration space Q is due to the conservative force F_i.*

Now, given a displacement δx, we define the *virtual work* δW_i done by the force F_i on particle P_i over this displacement as:

$$\delta W_i = F_i \cdot \delta x_i.$$

Consequently, the total virtual work δW done on the system is

$$\delta W = \sum_{i=1}^{n} \delta W_i = \sum_{i=1}^{n} F_i \cdot \delta x_i. \tag{1.53}$$

Since the functions (1.48) define a *parametrization*, then any virtual displacement δx_i can be written as:

$$\delta x_i = \sum_{j=1}^{m} \frac{\partial x_i}{\partial q_j}\, \delta q_j \tag{1.54}$$

for some virtual displacements δq_j. Consequently, equation (1.53) can be rewritten as:

$$\begin{aligned} \delta W &= \sum_{i=1}^{n} F_i \cdot \sum_{j=1}^{m} \frac{\partial x_i}{\partial q_j}\, \delta q_j \\ &= \sum_{j=1}^{m} \left(\sum_{i=1}^{n} F_i \cdot \frac{\partial x_i}{\partial q_j} \right) \delta q_j \\ &= \sum_{j=1}^{m} Q_j\, \delta q_j, \end{aligned} \tag{1.55}$$

where

$$Q_j = \sum_{i=1}^{n} F_i \cdot \frac{\partial x_i}{\partial q_j} \tag{1.56}$$

is the *generalized force* corresponding to the generalized coordinate q_j. From equations (1.52) and (1.53) we obtain:

$$\delta W = \sum_{i=1}^{n} F_i \cdot \delta x_i = \sum_{i=1}^{n} m_i a_i \cdot \delta x_i,$$

and substitution for δW using (1.55) and for the virtual displacement δx using (1.54) produces:

$$\begin{aligned}\sum_{j=1}^{m} Q_j \,\delta q_j &= \sum_{i=1}^{n} m_i a_i \cdot \sum_{j=1}^{m} \frac{\partial x_i}{\partial q_j}\,\delta q_j \\ &= \sum_{j=1}^{m}\left(\sum_{i=1}^{n} m_i a_i \cdot \frac{\partial x_i}{\partial q_j}\right)\,\delta q_j.\end{aligned} \tag{1.57}$$

To finish the derivation of Lagrange's equations in generalized coordinates, we wish to show that the quantity in parentheses in the above equation can be expressed in terms of derivatives of the kinetic energy K:

$$\begin{aligned}K &= \frac{1}{2}\sum_{i=1}^{n} m_i \|v_i\|^2 \\ &= \frac{1}{2}\sum_{i=1}^{n} m_i (v_i \cdot v_i).\end{aligned} \tag{1.58}$$

Differentiating the above equation with respect to q_j produces:

$$\frac{\partial K}{\partial q_j} = \sum_{i=1}^{n} m_i v_i \cdot \frac{\partial v_i}{\partial q_j}. \tag{1.59}$$

Similarly, differentiating (1.58) with respect to $\dot{q}_j$ produces:

$$\begin{aligned}\frac{\partial K}{\partial \dot{q}_j} &= \sum_{i=1}^{n} m_i v_i \cdot \frac{\partial v_i}{\partial \dot{q}_j} \\ &= \sum_{i=1}^{n} m_i v_i \cdot \frac{\partial \dot{x}_i}{\partial \dot{q}_j}.\end{aligned} \tag{1.60}$$

The next step in the derivation is somewhat technical, and relies upon an identity known as "the cancellation of dots", namely, that:

$$\frac{\partial \dot{x}_i}{\partial \dot{q}_j} = \frac{\partial x_i}{\partial q_j}. \tag{1.61}$$

To prove this, differentiate the parametrization (1.48) with respect to time:

$$\dot{x}_i = \frac{\partial x_i}{\partial t} + \sum_{j=1}^{m} \frac{\partial x_i}{\partial q_j}\dot{q}_j. \tag{1.62}$$

Since $v_i = \dot{x}_i$ by definition, equation (1.62) shows that

$$\frac{\partial v_i}{\partial \dot{q}_j} = \frac{\partial x_i}{\partial q_j}. \tag{1.63}$$

Back-substitution for v_i in equation (1.63) proves the identity (1.61).

Using the identity (1.61) in the expression (1.60) for the kinetic energy produces:

$$\frac{\partial K}{\partial \dot{q}_j} = \sum_{i=1}^{n} m_i v_i \cdot \frac{\partial x_i}{\partial q_j}$$

and taking the time derivative gives:

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_j}\right) = \sum_{i=1}^{n} m_i a_i \cdot \frac{\partial x_i}{\partial q_j} + m_i v_i \cdot \frac{d}{dt}\left(\frac{\partial x_i}{\partial q_j}\right). \tag{1.64}$$

Now

$$\frac{d}{dt}\left(\frac{\partial x_i}{\partial q_j}\right) = \frac{\partial^2 x_i}{\partial q_j\,\partial t} + \sum_{k=1}^{m} \frac{\partial^2 x_i}{\partial q_j\,\partial q_k}\dot{q}_k,$$

and differentiating (1.62) with respect to q_j likewise produces:

$$\frac{\partial \dot{x}_i}{\partial q_j} = \frac{\partial^2 x_i}{\partial t\,\partial q_j} + \sum_{k=1}^{m} \frac{\partial^2 x_i}{\partial q_k\,\partial q_j}\dot{q}_k,$$

so that

$$\frac{d}{dt}\left(\frac{\partial x_i}{\partial q_j}\right) = \frac{\partial \dot{x}_i}{\partial q_j} = \frac{\partial v_i}{\partial q_j}. \tag{1.65}$$

Consequently, substituting (1.65) into (1.64) shows that:

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_j}\right) = \sum_{i=1}^{n} m_i a_i \cdot \frac{\partial x_i}{\partial q_j} + m_i v_i \cdot \frac{\partial v_i}{\partial q_j}. \tag{1.66}$$

Equations (1.59) and (1.66) together show that

$$\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_j}\right) - \frac{\partial K}{\partial q_j} = \sum_{i=1}^{n} m_i a_i \cdot \frac{\partial x_i}{\partial q_j}. \tag{1.67}$$

Substitution of (1.67) into (1.57) shows that

$$\sum_{i=1}^{j} Q_j \, \delta q_j = \sum_{j=1}^{m} \left(\frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_j}\right) - \frac{\partial K}{\partial q_j}\right) \delta q_j. \tag{1.68}$$

Since each of the virtual displacements δq_j is linearly independent (by the independence of the coordinates q_j), equation (1.68) implies that

$$Q_j = \frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_j}\right) - \frac{\partial K}{\partial q_j} \tag{1.69}$$

for all $1 \leq j \leq m$. Thus, equation (1.69) shows how to compute the generalized forces in terms of the derivatives of the system's kinetic energy.

Note that, so far, we have not yet used the hypothesis that the forces F_i acting on the system are *conservative*; that this is so implies that there exists a potential energy function $V(x) = V(x(t, q)) = V(t, q)$ such that

$$F_i = -\frac{\partial V}{\partial x_i}. \tag{1.70}$$

Thus,

$$\begin{aligned} \frac{\partial V}{\partial q_j} &= \sum_{i=1}^{n} \frac{\partial V}{\partial x_i} \cdot \frac{\partial x_i}{\partial q_j} \\ &= -\sum_{i=1}^{n} F_i \cdot \frac{\partial x_i}{\partial q_j} \end{aligned} \tag{1.71}$$

and therefore

$$Q_j = -\frac{\partial V}{\partial q_j} \tag{1.72}$$

by the definition (1.56). Now, form the Lagrangian

$$\mathcal{L}(t, q, \dot{q}) = K(t, q, \dot{q}) - V(t, q) \tag{1.73}$$

as a function of the generalized coordinates q_j. Then equations (1.69) and (1.72) show that

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\right) - \frac{\partial \mathcal{L}}{\partial q_j} &= \frac{d}{dt}\left(\frac{\partial}{\partial \dot{q}_j}\left[K(t, q, \dot{q}) - V(t, q)\right]\right) - \frac{\partial}{\partial q_j}\left(K(t, q, \dot{q}) - V(t, q)\right) \\
&= \frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_j}\right) - \frac{\partial K}{\partial q_j} + \frac{\partial V}{\partial q_j} \\
&= Q_j - Q_j \\
&= 0
\end{aligned} \tag{1.74}$$

for all $1 \leq j \leq m$ (note that equations (1.73) and (1.74) are simply the generalizations of (1.45) and (1.46) to generalized coordinates). This proves the following result.

1.2.2.3 Lagrange's equations for conservative systems with holonomic constraints

Theorem 1.2.2 (Lagrange's equations for conservative systems with holonomic constraints). *Let* $\{P_i \mid 1 \leq i \leq n\}$ *be a system of particles with corresponding positions* $x_i \in \mathbb{R}^3$. *Suppose that the configuration of this system is subject to the holonomic constraints*

$$c_k(x_1, \ldots, x_n) = 0 \tag{1.75}$$

for functions $c_k\colon \mathbb{R}^{3n} \to \mathbb{R}$, and that the configuration space Q of the system determined by the constraints (1.75) *is a manifold described by the generalized coordinates $q_1, \ldots, q_m$ through the relations*

$$\begin{aligned} x_1 &= x_1(t, q_1, \ldots, q_m) \\ x_2 &= x_2(t, q_1, \ldots, q_m) \\ &\vdots \\ x_n &= x_n(t, q_1, \ldots, q_m). \end{aligned}$$

Suppose further that the system $\{P_i\}$ is acted upon by a set of conservative forces with corresponding potential energy function $V(t,q) = V(x(t,q))$. Let $K(t, q, \dot{q}) = K(\dot{x}(t, q, \dot{q}))$ denote the system's kinetic energy, and define the Lagrangian

$$\mathcal{L}(t, q, \dot{q}) = K(t, q, \dot{q}) - V(t, q). \tag{1.76}$$

Then the time evolution $q(t)$ of the system's configuration is the solution of Lagrange's equations:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\right) = \frac{\partial \mathcal{L}}{\partial q_j} \tag{1.77}$$

for all $1 \leq j \leq m$.

Remark 1.2.1. The same techniques as were used in Sections 1.1.4.3 and 1.1.5.2 show that Lagrange's equations (1.77) also describe the time evolution of mechanical systems composed of rigid bodies.

1.2.3 Lagrangian mechanics for holonomically constrained systems with nonconservative external forces

In the derivation of Lagrange's equations (1.46) and (1.77), we relied heavily upon the fact that the Lagrangian formulation of mechanics essentially encodes the requirement of conservation of energy. Consequently, conservative systems have a particularly nice representation within the Lagrangian framework: namely, they are described *completely* by their kinetic and potential energies. However, we shall often want to model systems that are *not* conservative. In this subsection, we shall extend Lagrange's equations (1.77) to incorporate nonconservative external time-varying forces.

1.2.3.1 Derivation of Lagrange's equations for holonomically constrained systems with nonconservative external forces

We now consider a system of n particles $\{P_i \mid 1 \leq i \leq n\}$ subject to holonomic constraints as in (1.47) with corresponding generalized coordinates $q_1, \ldots, q_m$. We suppose that this system is subject to a conservative force $F^{con}(x)$ with corresponding potential energy function $V(x)$ *and* a (possibly) time-varying nonconservative force $F^{ext}(t, x, \dot{x})$. Fortunately, much of the analysis of Section 1.2.2.2 still applies, so we will simply retrace the argument given there by referencing only those parts that require modification.

As before, we can write the total force F_i^{total} acting on each particle P_i

as a sum of forces:

$$F_i^{total} = F_i^{con} + F_i^{ext} + R_i = m_i a_i,$$

and given an arbitrary virtual displacement δx_i for particle P_i , we can compute the corresponding virtual work

$$\begin{aligned}\delta W_i &= F^{total} \cdot \delta x_i \\ &= F_i^{con} \cdot \delta x_i + F_i^{ext} \cdot \delta x_i + R_i \cdot \delta x_i \\ &= m_i a_i \cdot \delta x_i.\end{aligned}$$

Again, the observation that constraint forces do no work then implies that

$$F_i^{con} \cdot \delta x_i + F_i^{ext} \cdot \delta x_i = m_i a_i \cdot \delta x_i.$$

Defining the generalized forces

$$\begin{aligned}Q_j^{con} &= \sum_{i=1}^{n} F_i^{con} \cdot \frac{\partial x_i}{\partial q_j} = F^{con} \cdot \frac{\partial x}{\partial q_j} \\ Q_j^{ext} &= \sum_{i=1}^{n} F_i^{ext} \cdot \frac{\partial x_i}{\partial q_j} = F^{ext} \cdot \frac{\partial x}{\partial q_j}\end{aligned}$$

as in equation (1.56) and using the equality (1.54), equation (1.57) becomes:

$$\sum_{j=1}^{m} \left(Q_j^{con} + Q_j^{ext}\right) \delta q_j = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} m_i a_i \cdot \frac{\partial x_i}{\partial q_j} \right) \delta q_j. \tag{1.78}$$

Consequently, equation (1.68) together with the linear independence of the virtual displacements δq_j shows that

$$Q_j^{con} + Q_j^{ext} = \frac{d}{dt} \left(\frac{\partial K}{\partial \dot{q}_j} \right) - \frac{\partial K}{\partial q_j}, \tag{1.79}$$

analogous with equation (1.69). Now, equation (1.70) becomes

$$F_i^{con} = -\frac{\partial V}{\partial x_i}$$

so (1.71) and (1.72) imply that

$$Q_j^{con} = -\frac{\partial V}{\partial q_j}. \tag{1.80}$$

Consequently, forming the Lagrangian

$$\mathcal{L}(t, q, \dot{q}) = K(t, q, \dot{q}) - V(t, q),$$

equation (1.74) becomes:

$$\begin{aligned}
\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\right) - \frac{\partial \mathcal{L}}{\partial q_j} &= \frac{d}{dt}\left(\frac{\partial}{\partial \dot{q}_j}\left[K(t, q, \dot{q}) - V(t, q)\right]\right) - \frac{\partial}{\partial q_j}\left(K(t, q, \dot{q}) - V(t, q)\right) \\
&= \frac{d}{dt}\left(\frac{\partial K}{\partial \dot{q}_j}\right) - \frac{\partial K}{\partial q_j} + \frac{\partial V}{\partial q_j} \\
&= Q_j^{con} + Q_j^{ext} - Q_j^{con} \\
&= Q_j^{ext}
\end{aligned}$$

by equations (1.79) and (1.80). Thus, we have the following generalization of Theorem 1.2.2.

1.2.3.2 Lagrange's equations for holonomically constrained systems with nonconservative external forces

Theorem 1.2.3 (Lagrange's equations for holonomically constrained systems with nonconservative external forces). *Let $\{P_i \mid 1 \leq i \leq n\}$ be a system of particles with corresponding positions $x_i \in \mathbb{R}^3$. Suppose that the configuration of this system is subject to the holonomic constraints*

$$c_k(x_1, \ldots, x_n) = 0 \tag{1.81}$$

for functions $c_k\colon \mathbb{R}^{3n} \to \mathbb{R}$*, and that it is acted upon by a conservative force with corresponding potential energy function* $V(x)$ *and a nonconservative, time-varying external force* $F^{ext}(t, x, \dot{x})$*. Suppose further that the configuration space* Q *of the system determined by the constraints* (1.81) *is a manifold described by the generalized coordinates* $q_1, \ldots, q_m$ *through the relations*

$$\begin{aligned} x_1 &= x_1(t, q_1, \ldots, q_m) \\ x_2 &= x_2(t, q_1, \ldots, q_m) \\ &\vdots \\ x_n &= x_n(t, q_1, \ldots, q_m). \end{aligned}$$

Let

$$V(t, q) = V(x(t, q))$$

denote the system's potential energy and

$$K(t, q, \dot{q}) = K(\dot{x}(t, q, \dot{q}))$$

its kinetic energy, let F_i^{ext} *denote the nonconservative force exerted on the particle* P_i*, and define the generalized nonconservative force* Q_j^{ext} *corresponding to the generalized coordinate* q_j *via:*

$$Q_j^{ext}(t, q, \dot{q}) = \sum_{j=1}^{m} F_i^{ext}\left(t, x(t, q), \dot{x}(t, q, \dot{q})\right) \cdot \frac{\partial x_i}{\partial q_j}(t, q). \tag{1.82}$$

Finally, define the Lagrangian

$$\mathcal{L}(t, q, \dot{q}) = K(t, q, \dot{q}) - V(t, q). \tag{1.83}$$

Then the time evolution $q(t)$ of the system's configuration is the solution of Lagrange's equations:

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_j}\right) = \frac{\partial \mathcal{L}}{\partial q_j} + Q_j^{ext} \tag{1.84}$$

for all $1 \leq j \leq m$.

1.2.4 Examples

1.2.4.1 Return of bead on a wire

As an application of Theorem 1.2.2, let us return to the example of the bead on a wire from Section 1.1.6.2. In this case, the configuration space is a 1-dimensional submanifold of $\mathbb{R}$, since the bead is constrained to move tangentially along the wire. Adopting the coordinate θ as the generalized coordinate for this system as in equation (1.39), we found the kinetic and gravitational potential energies of the system were given by:

$$\begin{aligned} K &= \frac{1}{2}mR^2\dot{\theta}^2 \\ V &= mgR(1-\cos(\theta)). \end{aligned}$$

Consequently, the Lagrangian for this system is

$$\mathcal{L}(t,\theta,\dot{\theta}) = \frac{1}{2}mR^2\dot{\theta}^2 - mgR(1-\cos(\theta)), \tag{1.85}$$

and Lagrange's equations (1.77) become:

$$\begin{aligned} 0 &= \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}}\right) - \frac{\partial \mathcal{L}}{\partial \theta} \\ &= \frac{d}{dt}\left(mR^2\dot{\theta}\right) - (-mgR\sin(\theta)) \\ &= mR^2\ddot{\theta} + mgR\sin(\theta). \end{aligned}$$

Dividing through by $mR^2 \neq 0$ and rearranging gives

$$\ddot{\theta} = -\frac{g}{R}\sin(\theta),$$

which agrees with the previous result (1.35) obtained by the *ad hoc* application of conservation of energy.

1.2.4.2 The double compound pendulum

As another example, consider the double compound pendulum shown in Figure 1.3. We assume that this pendulum is composed of two linked rods R_1 and R_2 of lengths l_1 and l_2 with uniformly distribued masses m_1 and m_2, respectively. To describe the system, we adopt as the generalized coordinates the angles θ_1 and θ_2 as shown in the diagram. Since the only external force acting on this system is the force of gravity (which is conservative), we can again analyze this system via Theorem Theorem 1.2.2.

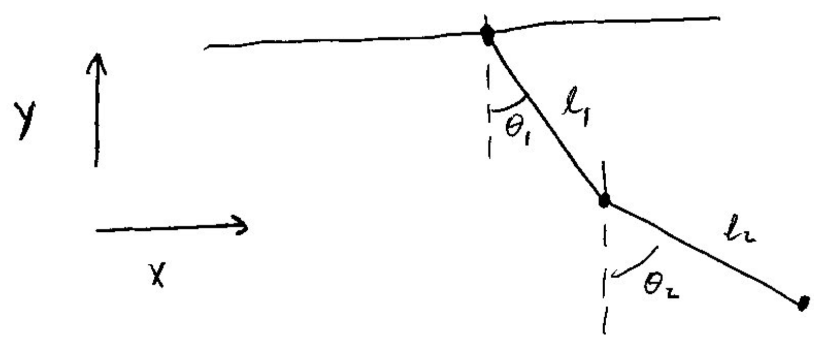


Figure 1.3: The double compound pendulum.

We begin by computing the centers of mass of the two rods; since they have uniform mass distributions by assumption, symmetry implies that their centers of mass coincide with their geometric centroids (i.e., their midpoints). Letting (x_1^{CM}, y_1^{CM}) denote the coordinates of R_1's center of mass and (x_2^{CM}, y_2^{CM}) denote the coordinates of R_2's center of mass, we compute

$$\begin{aligned} x_1^{CM} &= \frac{1}{2} l_1 \sin(\theta_1) \\ y_1^{CM} &= -\frac{1}{2} l_1 \cos(\theta_1) \\ x_2^{CM} &= l_1 \sin(\theta_1) + \frac{1}{2} l_2 \sin(\theta_2) \\ y_2^{CM} &= -l_1 \cos(\theta_1) - \frac{1}{2} l_2 \cos(\theta_2). \end{aligned} \tag{1.86}$$

Now, the gravitational potential energies of the two rods can be computed by measuring the vertical positions of their centers of mass. If we assign zero gravitational potential to the anchor point for the pendulum, then these energies are given by

$$\begin{aligned} V_1 &= m_1 \cdot g \cdot y_1^{CM} = -\frac{1}{2} m_1 g l_1 \cos\theta_1 \\ V_2 &= m_2 \cdot g \cdot y_2^{CM} = -m_2 g \left(l_1 \cos\theta_1 + \frac{1}{2} l_2 \cos\theta_2 \right). \end{aligned}$$

Thus, the total gravitational potential energy of the system is given by:

$$\begin{aligned} V &= V_1 + V_2 \\ &= -\frac{1}{2} m_1 g l_1 \cos\theta_1 - m_2 g \left(l_1 \cos\theta_1 + \frac{1}{2} l_2 \cos\theta_2 \right). \end{aligned}$$

Now we compute the kinetic energy of each of the rods. R_1 is straightforward: since this rod is anchored to the pendulum's pivot point, we can consider it as a rigid body in rotation about the pivot point, and equations (1.30) and (1.24) show that its total kinetic energy is given by:

$$K_1 = \frac{1}{2} I_{end} \omega_1^2 = \frac{1}{2} \left(\frac{1}{3} m_1 l_1^2 \right) \dot{\theta}_1^2 = \frac{1}{6} m_1 l_1^2 \dot{\theta}_1^2.$$

The second rod is more complicated: in general, it will have some rotational kinetic energy, but also some *translational* kinetic energy associated with the fact that the *entire* rod can move through space irrotationally (i.e., with θ_2 held constant). In order to simplify the computation, we partition R_2's kinetic energy into two parts: its *translational* kinetic energy K_2^{trans}, corresponding to the translational motion of its center of mass, and its *rotational* kinetic energy K_2^{rot}, corresponding to its rotation *about* its center of mass. Now

$$K_2^{trans} = \frac{1}{2} m_2 \left\| v_2^{CM} \right\|^2, \tag{1.87}$$

where $v_2^{CM} = (\dot{x}_2^{CM}, \dot{y}_2^{CM})$ is the velocity of R_2's center of mass. Differentiating (1.86) gives

$$\begin{aligned} \dot{x}_2^{CM} &= l_1 \cos(\theta_1)\dot{\theta}_1 + \frac{1}{2} l_2 \cos(\theta_2)\dot{\theta}_2 \\ \dot{y}_2^{CM} &= l_1 \sin(\theta_1)\dot{\theta}_1 + \frac{1}{2} l_2 \sin(\theta_2)\dot{\theta}_2. \end{aligned}$$

Thus,

$$\begin{aligned} \left\| v_2^{CM} \right\|^2 &= \left(\dot{x}_2^{CM} \right)^2 + \left(\dot{y}_2^{CM} \right)^2 \\ &= \left(l_1^2 \cos^2(\theta_1)\dot{\theta}_1^2 + l_1 l_2 \cos(\theta_1)\cos(\theta_2)\dot{\theta}_1\dot{\theta}_2 + \frac{1}{4} l_2^2 \cos^2(\theta_2)\dot{\theta}_2^2 \right) \\ &\quad + \left(l_1^2 \sin^2(\theta_1)\dot{\theta}_1^2 + l_1 l_2 \sin(\theta_1)\sin(\theta_2)\dot{\theta}_1\dot{\theta}_2 + \frac{1}{4} l_2^2 \sin^2(\theta_2)\dot{\theta}_2^2 \right) \\ &= l_1^2\dot{\theta}_1^2 + l_1 l_2 \dot{\theta}_1\dot{\theta}_2 \cos(\theta_1 - \theta_2) + \frac{1}{4} l_2^2 \dot{\theta}_2^2, \end{aligned}$$

so that (1.87) becomes

$$K_2^{trans} = \frac{1}{2} m_2 \left(l_1^2\dot{\theta}_1^2 + l_1 l_2 \dot{\theta}_1\dot{\theta}_2 \cos(\theta_1 - \theta_2) + \frac{1}{4} l_2^2 \dot{\theta}_2^2 \right).$$

Finally, we can compute R_2's rotational kinetic energy about its center of mass using equations (1.30) and (1.25):

$$K_2^{rot} = \frac{1}{2} I_{center} \omega_2^2 = \frac{1}{2}\left(\frac{1}{12} m_2 l_2^2\right)\dot{\theta}_2^2 = \frac{1}{24} m_2 l_2^2 \dot{\theta}_2^2.$$

Thus, the Lagrangian for this system is

$$\begin{aligned}
\mathcal{L} &= K - V \\
&= K_1^{rot} + K_2^{trans} + K_2^{rot} - (V_1 + V_2) \\
&= \frac{1}{6} m_1 l_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 \left(l_1^2 \dot{\theta}_1^2 + l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \cos(\theta_1 - \theta_2) + \frac{1}{4} l_2^2 \dot{\theta}_2^2 \right) + \frac{1}{24} m_2 l_2^2 \dot{\theta}_2^2 \\
&\quad + \frac{1}{2} m_1 g l_1 \cos\theta_1 + m_2 g \left(l_1 \cos\theta_1 + \frac{1}{2} l_2 \cos\theta_2 \right),
\end{aligned}$$

and by grouping like terms, we finally obtain:

$$\begin{aligned}
\mathcal{L} = &\left(\frac{1}{6} m_1 l_1^2 + \frac{1}{2} m_2 l_1^2\right)\dot{\theta}_1^2 + \frac{1}{2} m_2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \cos(\theta_1 - \theta_2) + \frac{1}{6} m_2 l_2^2 \dot{\theta}_2^2 \\
&+ \left(\frac{1}{2} m_1 g l_1 + m_2 g l_1\right)\cos\theta_1 + \frac{1}{2} m_2 g l_2 \cos\theta_2.
\end{aligned} \tag{1.88}$$

Thus, to compute this system's time evolution, we need only apply Lagrange's equations (1.77). Differentiating (1.88), we find:

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial \theta_1} &= -\frac{1}{2} m_2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \sin(\theta_1 - \theta_2) - \left(\frac{1}{2} m_1 g l_1 + m_2 g l_1\right)\sin(\theta_1) \\
\frac{\partial \mathcal{L}}{\partial \theta_2} &= \frac{1}{2} m_2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \sin(\theta_1 - \theta_2) - \frac{1}{2} m_2 g l_2 \sin(\theta_2) \\
\frac{\partial \mathcal{L}}{\partial \dot{\theta}_1} &= \left(\frac{1}{3} m_1 l_1^2 + m_2 l_1^2\right)\dot{\theta}_1 + \frac{1}{2} m_2 l_1 l_2 \dot{\theta}_2 \cos(\theta_1 - \theta_2) \\
\frac{\partial \mathcal{L}}{\partial \dot{\theta}_2} &= \frac{1}{2} m_2 l_1 l_2 \dot{\theta}_1 \cos(\theta_1 - \theta_2) + \frac{1}{3} m_2 l_2^2 \dot{\theta}_2,
\end{aligned} \tag{1.89}$$

and differentiating the final two lines of (1.89) with respect to time gives:

$$\begin{aligned}\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}_1}\right) &= \left(\frac{1}{3}m_1 l_1^2 + m_2 l_1^2\right)\ddot{\theta}_1 + \frac{1}{2}m_2 l_1 l_2 \cos(\theta_1 - \theta_2)\ddot{\theta}_2 \\ &\quad - \frac{1}{2}m_2 l_1 l_2 \dot{\theta}_2 \sin(\theta_1 - \theta_2)(\dot{\theta}_1 - \dot{\theta}_2), \\ \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}_2}\right) &= \frac{1}{2}m_2 l_1 l_2 \cos(\theta_1 - \theta_2)\ddot{\theta}_1 - \frac{1}{2}m_2 l_1 l_2 \dot{\theta}_1 \sin(\theta_1 - \theta_2)(\dot{\theta}_1 - \dot{\theta}_2) \\ &\quad + \frac{1}{3}m_2 l_2^2 \ddot{\theta}_2.\end{aligned} \tag{1.90}$$

Substitution of (1.89) and (1.90) into Lagrange's equations (1.77) gives the system of equations:

$$\begin{aligned}0 &= \left(\frac{1}{3}m_1 l_1^2 + m_2 l_1^2\right)\ddot{\theta}_1 + \frac{1}{2}m_2 l_1 l_2 \cos(\theta_1 - \theta_2)\ddot{\theta}_2 \\ &\quad - \frac{1}{2}m_2 l_1 l_2 \dot{\theta}_2 \sin(\theta_1 - \theta_2)(\dot{\theta}_1 - \dot{\theta}_2) + \frac{1}{2}m_2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \sin(\theta_1 - \theta_2) \\ &\quad + \left(\frac{1}{2}m_1 g l_1 + m_2 g l_1\right)\sin(\theta_1), \\ 0 &= \frac{1}{2}m_2 l_1 l_2 \cos(\theta_1 - \theta_2)\ddot{\theta}_1 - \frac{1}{2}m_2 l_1 l_2 \dot{\theta}_1 \sin(\theta_1 - \theta_2)(\dot{\theta}_1 - \dot{\theta}_2) \\ &\quad + \frac{1}{3}m_2 l_2^2 \ddot{\theta}_2 - \frac{1}{2}m_2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \sin(\theta_1 - \theta_2) + \frac{1}{2}m_2 g l_2 \sin(\theta_2)\end{aligned}$$

which is a linear system in $\ddot{\theta}_1$ and $\ddot{\theta}_2$. Solving this system produces:

$$\begin{aligned}\ddot{\theta}_1 &= C_1 \Big(4m_2 g \sin(\theta_1) + 2m_2 l_2 \sin(\theta_1 - \theta_2)\dot{\theta}_2^2 \\ &\quad + 3m_2 l_1 \cos(\theta_1 - \theta_2)\sin(\theta_1 - \theta_2)\dot{\theta}_1^2 \\ &\quad - 3m_2 g \cos(\theta_1 - \theta_2)\sin(\theta_2) + 2m_1 g \sin(\theta_1)) \\ \ddot{\theta}_2 &= C_2 \left(6m_2 g \cos(\theta_1 - \theta_2)\sin(\theta_1) + 3m_1 g \cos(\theta_1 - \theta_2)\sin(\theta_1)\right. \\ &\quad + 3m_2 l_2 \cos(\theta_1 - \theta_2)\sin(\theta_1 - \theta_2)\dot{\theta}_2^2 \\ &\quad + 2m_1 l_1 \sin(\theta_1 - \theta_2)\dot{\theta}_1^2 + 6m_2 l_1 \sin(\theta_1 - \theta_2)\dot{\theta}_1^2 \\ &\quad \left. - 2m_1 g \sin(\theta_2) - 6m_2 g \sin(\theta_2)\right)\end{aligned} \tag{1.91}$$

for

$$C_1 = \frac{3}{9m_2 l_1 \cos^2(\theta_1 - \theta_2) - 4m_1 l_1 - 12m_2 l_1},$$
$$C_2 = \frac{3}{9m_2 l_2 \cos^2(\theta_1 - \theta_2) - 4m_1 l_2 - 12m_2 l_2}$$

which, together with the initial configuration and velocity, determines the corresponding initial value problem for this system.

While the derivation of equations (1.91) using the Lagrangian formulation may seem laborious, it is in fact far less laborious than the direct application of Newton's laws. Indeed, the real power of the Lagrangian formulation is that, generally speaking, there are closed-form formulae that one can use to compute the kinetic and potential energies of a system directly from a description of that system's geometry, as we did in this example. The derivation of the system's equations of motion via the application of Lagrange's equations then simply reduces to "turning the crank", i.e., performing the indicated differentiation, and then solving the resulting set of equations.

1.2.5 Hamilton's Principle

Lagrange's equations of motion (1.77) are a particular instance of a set of second-order partial differential equations called the *Euler-Lagrange equations*. These equations describe sufficient conditions for a mapping $u \in C^2(\overline{\Omega}, \mathbb{R}^m)$ to be a critical point of a functional of the form

$$\mathcal{F}[v] = \int_\Omega F(s, v(s), Dv(s))\, ds \tag{1.92}$$

amongst all functions in $C^2(\overline{\Omega}, \mathbb{R}^m)$ satisfying certain boundary conditions (see Appendix A for an introduction to problems of this type and a derivation of the Euler-Lagrange equations). In the case where $\overline{\Omega} = [t_0, t_1]$, the Euler-Lagrange equations state that if $x(t)$ is a critical point of the functional (1.92), then x satisfies the Euler-Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) - \frac{\partial F}{\partial x} = 0.$$

Consequently, Lagrange's equations of motion can be recast in a variational form as a statement about the critical points of the *action functional*, a functional of the form (1.92) defined on mechanical systems. Specifically, given a mechanical system with generalized coordinates $q_1, \ldots, q_m$, the action functional $\mathcal{A}$ is defined to be

$$\mathcal{A}[u] = \int_{t_0}^{t_1} \mathcal{L}(t, u(t), \dot{u}(t))\, dt$$

where $u \in C^2([t_0, t_1], \mathbb{R}^n)$ and has prescribed endpoints $u(t_0) = a$ and $u(t_1) = b$, and $\mathcal{L}(t, q, \dot{q}(t))$ is the corresponding Lagrangian (1.76) of the system. This leads to *Hamilton's principle*.

Theorem 1.2.4 (Hamilton's principle). *Consider a mechanical system described by generalized coordinates $q_1, \ldots, q_m$ and with corresponding Lagrangian $\mathcal{L}(t, q, \dot{q})$ as in equation* (1.76)*. If the system's configuration at time t_0 and t_1 is a and b, respectively, then the time evolution $q(t)$ of the system over the interval $[t_0, t_1]$ is the critical point of the action functional*

$$\mathcal{A}[u] = \int_{t_0}^{t_1} \mathcal{L}(t, u(t), \dot{u}(t))\, dt$$

where the variation is taken over all $u \in C^2([a,b],\mathbb{R}^m)$ *satisfying* $u(t_0) = a$ *and* $u(t_1) = b$*.*

1.3 Hamiltonian mechanics

Finally, we introduce *Hamiltonian mechanics*. The primary distinction between Lagrangian and Hamiltonian mechanics lies in how the system of ordinary differential equations describing the future time evolution of the configuration is represented.

When one forms Lagrange's equations of motion (1.77), the system of ordinary differential equations so obtained is second-order in the generalized coordinates $q_1, \ldots, q_m$. To obtain the Hamiltonian formulation, we use the Legendre tranformation to introduce an auxiliary set of variables $p_1, \ldots, p_m$, and then represent the system as a first-order system of ordinary differential equations in the $2m$ variables $q_1, \ldots, q_m, p_1, \ldots, p_m.$.

Specifically, let $q_1, \ldots, q_m$ denote the generalized coordinates for a mechanical system, and let $\mathcal{L}(t, q, \dot{q})$ denote that system's Lagrangian. We introduce an auxiliary set of m variables, called *generalized momenta*, according to

$$p_j = \frac{\partial \mathcal{L}}{\partial \dot{q}_j}. \tag{1.93}$$

Supposing that the system of equations (1.93) can be solved for the velocity $\dot{q}$

as a function of the time t, position q, and the generalized momentum p, we then define the *Hamiltonian* $\mathcal{H}$ as

$$\mathcal{H}(t, q, p) = \sum_{j=1}^{m} p_j \dot{q}_j(t, q, p) - \mathcal{L}(t, q, \dot{q}(t, q, p)). \tag{1.94}$$

Now since $\mathcal{H}$ is a function of q, p, and t, its total differential is

$$d\mathcal{H} = \sum_{j=1}^{m} \left(\frac{\partial \mathcal{H}}{\partial q_j} \, dq_j + \frac{\partial \mathcal{H}}{\partial p_j} \, dp_j \right) + \frac{\partial \mathcal{H}}{\partial t} \, dt. \tag{1.95}$$

On the other hand, directly taking the total differential of the right-hand side of (1.94) gives

$$d\mathcal{H} = \sum_{j=1}^{m} \left(\dot{q}_j \, dp_j + p_j \, d\dot{q}_j - \frac{\partial \mathcal{L}}{\partial q_j} \, dq_j - \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \, d\dot{q}_j \right) - \frac{\partial \mathcal{L}}{\partial t} \, dt, \tag{1.96}$$

and substitution of the definition (1.93) into (1.96) produces

$$\begin{aligned} d\mathcal{H} &= \sum_{j=1}^{m} \left(\dot{q}_j \, dp_j + \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \, d\dot{q}_j - \frac{\partial \mathcal{L}}{\partial q_j} \, dq_j - \frac{\partial \mathcal{L}}{\partial \dot{q}_j} \, d\dot{q}_j \right) - \frac{\partial \mathcal{L}}{\partial t} \, dt \\ &= \sum_{j=1}^{m} \left(\dot{q}_j \, dp_j - \frac{\partial \mathcal{L}}{\partial q_j} \, dq_j \right) - \frac{\partial \mathcal{L}}{\partial t} \, dt. \end{aligned} \tag{1.97}$$

Equating like terms in equations (1.95) and (1.97) shows that

$$\dot{q}_j = \frac{\partial \mathcal{H}}{\partial p_j}, \quad \frac{\partial \mathcal{H}}{\partial q_j} = -\frac{\partial \mathcal{L}}{\partial q_j}, \quad \frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t}. \tag{1.98}$$

Now Lagrange's equations of motion (1.77) show that

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial q} &= \frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \\ &= \dot{p}, \end{aligned}$$

and consequently,

$$\dot{p}_j = -\frac{\partial \mathcal{H}}{\partial q_j}$$

by virtue of (1.98). Thus, we have proved the following.

Theorem 1.3.1 (Hamilton's canonical equations). *Let $q = (q_1, \ldots, q_m)$ be the generalized coordinates for a conservative mechanical system with Lagrangian $\mathcal{L}(t, q, \dot{q})$. Define the generalized momenta $p = (p_1, \ldots, p_m)$ by*

$$p = \frac{\partial \mathcal{L}}{\partial \dot{q}}, \tag{1.99}$$

and supposing that the system of equations (1.99) *can be solved for the velocity $\dot{q}$ as a function of the time t, position q, and the generalized momentum p, define the corresponding Hamiltonian $\mathcal{H}$ for this system by*

$$\mathcal{H}(t, q, p) = p \cdot \dot{q}(t, q, p) - \mathcal{L}(t, q, \dot{q}(t, q, p)). \tag{1.100}$$

Then the time evolution $q(t)$, $p(t)$ of the configuration and momenta can be obtained as the solution of the system of first-order ordinary differential equations

$$\dot{p} = -\frac{\partial \mathcal{H}}{\partial q}, \quad \dot{q} = \frac{\partial \mathcal{H}}{\partial p}, \quad \frac{\partial \mathcal{H}}{\partial t} = -\frac{\partial \mathcal{L}}{\partial t}. \tag{1.101}$$

Remark 1.3.1. The equations (1.101) describing the time evolution of the augmented state $(q_1(t), \ldots, q_m(t), p_1(t), \ldots, p_m(t))$ are called *Hamilton's canonical equations*, or simply *the canonical equations*.

Example: bead on a wire part 3

In Section 1.2.4.1, we computed the Lagrangian for the bead on a wire in equation (1.85):

$$\mathcal{L}(t, \theta, \dot{\theta}) = \frac{1}{2} m R^2 \dot{\theta}^2 - mgR(1 - \cos(\theta)). \tag{1.102}$$

Differentiating the Lagrangian with respect to $\dot{\theta}$ produces:

$$p = \frac{\partial \mathcal{L}}{\partial \dot{\theta}} = mR^2\dot{\theta},$$

so that

$$\dot{\theta} = \frac{p}{mR^2}. \tag{1.103}$$

Now the Hamiltonian for this system is

$$\begin{aligned}\mathcal{H} &= p \cdot \dot{\theta} - \mathcal{L}(t, q, \dot{\theta}) \\ &= p \cdot \dot{\theta} - \frac{1}{2}mR^2\dot{\theta}^2 + mgR(1 - \cos(\theta))\end{aligned}$$

and substituting for $\dot{\theta}$ using (1.103) gives:

$$\begin{aligned}\mathcal{H} &= p \cdot \dot{\theta} - \frac{1}{2}mR^2\dot{\theta}^2 + mgR(1 - \cos(\theta)) \\ &= p\left(\frac{p}{mR^2}\right) - \frac{1}{2}mR^2\left(\frac{p}{mR^2}\right)^2 + mgR(1 - \cos(\theta)) \\ &= \frac{p^2}{2mR^2} + mgR(1 - \cos(\theta)).\end{aligned} \tag{1.104}$$

Thus, appyling Hamilton's equations (1.101) shows that

$$\begin{aligned}\dot{\theta} &= \frac{\partial \mathcal{H}}{\partial p} = \frac{p}{mR^2}, \\ \dot{p} &= -\frac{\partial \mathcal{H}}{\partial \theta} = -mgR\sin(\theta).\end{aligned} \tag{1.105}$$

Observe that equation (1.105) is really just an integrated form of equation (1.35). For by differentiating the first equation in (1.105), we obtain

$$\ddot{\theta} = \frac{\dot{p}}{mR^2},$$

and then substitution for $\dot{p}$ using the second equation in (1.105) produces

$$\ddot{\theta} = -\frac{g}{R}\sin(\theta),$$

which is (1.35).

Chapter 2

Geometric Mechanics

Thus far, we have studied mechanics by describing the allowable configurations of a mechanical system as a subset of some Euclidean space, this subset being determined by some number of constraints on the system's configuration variables. However, as we have seen, the subsets of Euclidean space so determined are usually manifolds in their own right. Consequently, it seems more natural and elegant to attempt a description of these systems solely in terms of their configuration spaces when viewed as abstract manifolds.

In this chapter, we develop *geometric mechanics*, a formulation of mechanics that recasts the Lagrangian and Hamiltonian formulations from Chapter 1 in terms of geometric objects associated to the system's configuration manifold. In addition to being a much more natural mathematical representation of mechanical systems, this formulation also admits the application of powerful analytical techniques from differential geometry to aid in the analysis of these systems. This chapter assumes familiarity with smooth manifold topology and geometry at the level of [14] and [8], respectively.

2.1 Symplectic geometry and Poisson manifolds

We begin our study of geometric mechanics with a word on symplectic geometry and Poisson manifolds. This may seem something of a *non sequitur* at first glance, but as it turns out, symplectic manifolds are absolutely fundamental to the study of mechanics – indeed, the realization that these two fields are intimately related was what prompted the initial interest in symplectic geometry. For convenience, we have compiled some elementary facts about Poisson and symplectic manifolds in Appendix B.

Recall that a *Poisson manifold* is a pair $(P, \{,\})$, where P is a smooth manifold and $\{,\}$ is a binary operation

$$\{\cdot,\cdot\}\colon C^\infty(P) \times C^\infty(P) \to C^\infty(P)$$

(called the *Poisson bracket*) satisfying the following properties:

1. Bilinearity.

2. Skew-symmetry.

3. The Jacobi identity:

$$\{\{f, g\}, h\} + \{\{h, f\}, g\} + \{\{g, h\}, f\} = 0.$$

4. The Leibniz rule:

$$\{fg, h\} = f\{g, h\} + g\{f, h\}.$$

By virtue of property 4, given any $H \in C^\infty(P)$, the mapping

$$f \mapsto \{f, H\}$$

is a derivation on $C^\infty(P)$, and consequently, by the identification of the tangent space $T_p(P)$ with the set of derivations on the equivalence class $\tilde{F}_p$ of germs of smooth functions at p, we can associate to H a unique vector field $X_H \in \mathfrak{X}(P)$, called the *Hamiltonian vector field*, with the property that

$$X_H(f) = \{f, H\}$$

for all $f \in C^\infty(P)$.

Similarly, recall that a *symplectic manifold* is a pair (M, ω), where M is a (necessarily even-dimensional) smooth manifold and ω is a *symplectic form* (a smooth, closed, nondegenerate 2-form) on M. Since ω is nondegenerate, it induces an isomorphism

$$\omega_\flat \colon \mathfrak{X}(M) \to \Omega^1(M)$$
$$\omega_\flat(X) = i_X \omega,$$

where here i_X denotes contraction with the vector field X. Given a vector field $X \in \mathfrak{X}(M)$, we write $X^\flat$ for $\omega_\flat(X)$, and given any $\theta \in \Omega^1(M)$, we write $\theta^\sharp$ to denote the inverse of θ under $\omega_\flat$. Defining

$$X_f \in \mathfrak{X}(M)$$
$$X_f = (df)^\sharp$$

for all $f \in C^\infty(M)$, the binary operation defined by

$$\{\cdot,\cdot\} \colon C^\infty(M) \times C^\infty(M) \to C^\infty(M)$$
$$\{f, g\} = \omega(X_f, X_g),$$

is a Poisson bracket on M, and X_f is the Hamiltonian vector field corresponding to f (see Section B.3.2 in Appendix B for the details of these computations). Furthermore, Darboux's Theorem (Theorem B.2.5 in Appendix B) provides, around each point $m \in M$, a *symplectic chart*, an open set U with *canonical coordinate functions* $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ on which the symplectic form ω can be written as

$$\omega|_U = \sum_{i=1}^{n} dq_i \wedge dp_i.$$

Likewise, on these symplectic charts, the Hamiltonian vector field X_H corresponding to a smooth function $H \in C^\infty(M)$ can be written in the form

$$X_H = \sum_{i=1}^{n} \frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i}\frac{\partial}{\partial p_i} \tag{2.1}$$

(Theorem B.3.15 in Appendix B). Now here is the crucial observation: *Equation* (2.1) *is precisely Hamilton's canonical equations* (1.101) *when H is the Hamiltonian!* So symplectic manifolds come with Hamiltonian dynamics built in: given any symplectic manifold (M, ω) and any $H \in C^\infty(M)$, there exists a unique Hamiltonian vector field $X_H \in \mathfrak{X}(M)$ whose integral curves are the solutions of Hamilton's equations when H is the Hamiltonian!

This is certainly a remarkable property, but perhaps it is not as useful as it might at first appear: perhaps symplectic manifolds are sufficiently "rare" that they do not come up in practice often enough to be useful. As it turns out, quite the opposite is true: symplectic manifolds are in fact quite common in practice. Indeed, one elementary yet remarkable fact (Theorem B.2.7 in Appendix B) is the following: for any smooth manifold M, the cotangent bundle T^*M can be equipped with a *canonically defined* symplectic form, called (appropriately) *the canonical 2-form* Ω, and the natural charts on T^*M induced by the charts on M are symplectic charts for Ω. So not only can every cotangent bundle T^*M be made into a symplectic manifold, but this can be done in a canonical way! This will turn out to be crucially important for our study of mechanics.

In summary, equation (2.1) shows that symplectic geometry is somehow the "right" setting for the study of Hamiltonian mechanics. Furthermore, the close connection between Hamiltonian and Lagrangian mechanics (at least under sufficient regularity conditions) suggests that symplectic geometry will be discreetly lurking in the background throughout our entire study of geometric mechanics. Indeed, this turns out to be the case.

2.2 The fiber derivative

Before continuing on to the study of geometric mechanics proper, we pause here to introduce a mathematical gadget that will be of considerable use in the sequel: the *fiber derivative.*

Definition 2.2.1 (Fiber derivative). Let Q be a manifold and $\psi\colon TQ \to \mathbb{R}$ a smooth real-valued function on the tangent bundle TQ. The mapping $\mathbb{F}\psi\colon TQ \to T^*Q$ which is defined fiberwise by

$$\begin{aligned} \mathbb{F}\psi|_{(TQ)_x}\colon (TQ)_x &\to (T^*Q)_x \\ (\mathbb{F}\psi(v))\,(w) &= \frac{d}{dt}\left[\psi(v+tw)\right]\big|_{t=0} \end{aligned} \tag{2.2}$$

where $x \in Q$ and $v, w \in (TQ)_x$ (i.e., v and w lie in the same fiber of TQ), is called the *fiber derivative*.

Let $x \in Q$, let $(q_1, \ldots, q_m)$ be a set of local coordinates on Q on a neighborhood of x, and (in a slight abuse of notation) let $(q_1, \ldots, q_m, \dot{q}_1, \ldots, \dot{q}_m)$ and $(q_1, \ldots, q_m, p_1, \ldots, p_m)$ be the corresponding set of natural coordinates on TQ and T^*Q, respectively. Fix $v, w \in T_x(Q)$, where

$$w = \sum_{i=1}^{n} c_i \frac{\partial}{\partial q_i} \in T_x(Q). \tag{2.3}$$

Then by definition (2.2),

$$\begin{aligned} (\mathbb{F}\psi(v))(w) &= \frac{d}{dt}\left[\psi(v+tw)\right]\big|_{t=0} \\ &= \sum_{i=1}^{n} \frac{\partial\psi}{\partial q_i}\bigg|_v \frac{\partial(q_i(v+tw))}{\partial t}\bigg|_{t=0} + \sum_{i=1}^{n} \frac{\partial\psi}{\partial \dot{q}_i}\bigg|_v \frac{\partial(\dot{q}_i(v+tw))}{\partial t}\bigg|_{t=0}. \end{aligned} \tag{2.4}$$

Now since $v, w \in T_x(Q)$ by hypothesis, then

$$\frac{\partial(q_i(v+tw))}{\partial t}\bigg|_{t=0} = 0$$

identically, since the summation $v + tw \in T_x(Q)$ as well. Similarly,

$$\left.\frac{\partial(\dot{q}_i(v + tw))}{\partial t}\right|_{t=0} = c_i$$

by equation (2.3) and the fact that $(q_1, \ldots, q_m, \dot{q}_1, \ldots, \dot{q}_m)$ is a natural coordinate system on TQ. Consequently, (2.4) reduces to

$$(\mathbb{F}\psi(v))(w) = \sum_{i=1}^{n} \frac{\partial \psi}{\partial \dot{q}_i}(v)\, c_i,$$

and therefore, we can write $\mathbb{F}\psi(v) \in (T_x(Q))^*$ as

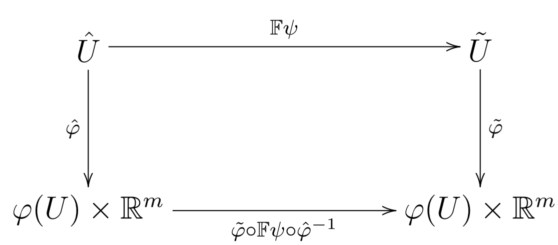
$$\mathbb{F}\psi(v) = \sum_{i=1}^{n} \frac{\partial \psi}{\partial \dot{q}_i}(v)\, dq_i$$

(again using (2.3)). This proves the following.

Theorem 2.2.1 (The fiber derivative in local coordinates). *Let Q be a manifold and $\psi\colon TQ \to \mathbb{R}$ a smooth real-valued function on its tangent bundle. The fiber derivative $\mathbb{F}\psi\colon TQ \to T^*Q$ is a smooth bundle map over id_Q. Furthermore, if (U, φ) is a chart on Q about x, $(\hat{U}, \hat{\varphi})$ is the corresponding natural chart on TQ with natural coordinates $(q_1, \ldots, q_m, \dot{q}_1, \ldots, \dot{q}_m)$, and $(\tilde{U}, \tilde{\varphi})$ is the corresponding natural chart on T^*Q with natural coordinates $(q_1, \ldots, q_n, p_1, \ldots, p_n)$, then $\mathbb{F}\psi$ can be written with respect to these coordinate charts as:*

$$\mathbb{F}\psi(v) = \sum_{i=1}^{n} \frac{\partial \psi}{\partial \dot{q}_i}(v)\, dq_i \in (T^*Q)_x \tag{2.5}$$

where $v \in (TQ)_x$. Equivalently, given the commutative diagram:



the induced map $\tilde{\varphi} \circ \mathbb{F}\psi \circ \hat{\varphi}^{-1}$ *satisfies*

$$(\tilde{\varphi} \circ \mathbb{F}\psi \circ \hat{\varphi}^{-1})(r, \dot{r}) = \left(r, \frac{\partial \psi}{\partial \dot{q}} \circ \hat{\varphi}^{-1}(r, \dot{r}) \right). \tag{2.6}$$

Remark 2.2.1. This motivates the use of the term "fiber derivative" in describing the map $\mathbb{F}\psi$. As equations (2.5) and (2.6) show, intrinsically, $\mathbb{F}\psi$ differentiates the function ψ along the fibers of TQ.

One can define a corresponding notion of the fiber derivative for a function $\psi \colon T^*Q \to \mathbb{R}$ on the cotangent bundle.

Definition 2.2.2 (Fiber derivative). Let Q be a manifold and $\psi \colon T^*Q \to \mathbb{R}$ a smooth real-valued function on T^*Q. The mapping $\mathbb{F}\psi \colon T^*Q \to TQ$ which is defined fiberwise by

$$\begin{aligned} \mathbb{F}\psi|_{(T^*Q)_x} &\colon (T^*Q)_x \to (TQ)_x \\ (\mathbb{F}\psi(\theta))(\omega) &= \frac{d}{dt}\left[\psi(\theta + t\omega)\right]\Big|_{t=0} \end{aligned} \tag{2.7}$$

where $x \in Q$ and $\theta, \omega \in (T^*Q)_x$ (i.e., θ and ω lie in the same fiber of T^*Q), is called the *fiber derivative.*

Remark 2.2.2. Note that in definition (2.7) we are implicitly using the reflexivity of finite-dimensional vector spaces to make the identification

$$((T_x(Q))^*)^* \cong T_x(Q).$$

More precisely,

$$(\mathbb{F}\psi(\theta))(\omega) = \frac{d}{dt}\left[\psi(\theta + t\omega)\right]\Big|_{t=0} \in ((T^*Q)_x)^* = ((T_x(Q))^*)^*,$$

but using the canonical isomorphism

$$\varphi\colon T_x(Q) \to ((T_x(Q))^*)^*$$
$$\varphi(v) = E_v,$$

where E_v is the *evaluation function*

$$E_v\colon (T_x(Q))^* \to \mathbb{R}$$
$$E_v(\theta) = \theta(v),$$

we can regard $\mathbb{F}\psi$ as a map into $T_x(Q)$.

Of course, with this definition in hand, the analogue of Theorem 2.2.1 obviously holds. Finally, we make a few definitions.

Definition 2.2.3 (Regular, hyperregular functions). Let Q be a manifold. A smooth function $\psi\colon TQ \to \mathbb{R}$ is called *regular* if its fiber derivative $\mathbb{F}\psi\colon TQ \to T^*Q$ is a local diffeomorphism, and *hyperregular* if $\mathbb{F}\psi$ is a diffeomorphism.

Theorem 2.2.2 (Necessary and sufficient conditions for $\mathbb{F}\psi$ to be a local diffeomorphism). *Let Q be a manifold and $\psi\colon TQ \to \mathbb{R}$ a smooth function. Let (U, φ) be a coordinate chart on Q and $(\hat{U}, \hat{\varphi})$ the corresponding natural chart on TQ with natural coordinates $(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n)$. Then $\mathbb{F}\psi$ is a local diffeomorphism on $\hat{U}$ if and only if the $n \times n$ matrix*

$$\left(\frac{\partial^2 \psi}{\partial \dot{q}_i \partial \dot{q}_j}\right) \tag{2.8}$$

is invertible.

Proof. Let $(\tilde{U}, \tilde{\varphi})$ be the corresponding natural chart on T^*Q with natural coordinates $(q_1, \ldots, q_n, p_1, \ldots, p_n)$. Then equation (2.6) shows that

$$d\mathbb{F}\psi = \begin{pmatrix} I & 0 \\ \frac{\partial^2 \psi}{\partial \dot{q}_i \partial q_j} & \frac{\partial^2 \psi}{\partial \dot{q}_i \partial \dot{q}_j} \end{pmatrix}$$

so that $d\mathbb{F}\psi$ is invertible if and only if (2.8) is invertible. Since $\mathbb{F}\psi$ is a local diffeomorphism if and only if $d\mathbb{F}\psi$ is invertible by the Inverse Function Theorem, the result follows. □

2.3 Lagrangian mechanics on manifolds

In this section we describe how to represent the Lagrangian formulation of mechanics from Chapter 1 using differential geometric objects, following the general outline in [10].

2.3.1 Configuration manifolds, tangent bundles, and the Lagrangian

Definition 2.3.1 (Configuration space). Given a mechanical system, the corresponding *configuration space* or *configuration manifold*, denoted Q, is the smooth manifold whose points correspond to the admissible configurations of the system.

For example, the configuration space of the bead on a wire from Section 1.1.6.2 is a connected submanifold of $\mathbb{R}$, since the bead's position in space can be uniquely characterized by its displacement along the wire. Similarly, the configuration of the double compound pendulum from Section 1.2.4.2 is determined by specifying the two angles θ_1 and θ_2, which we can regard as corresponding to points on the circle S^1; consequently, we see that this system's configuration space is the 2-torus $\mathbb{T}^2$.

When viewed from this context, we see that a system of generalized coordinates $(q_1, \ldots, q_m)$ as defined in Section 1.2.2.1 is really just a system of coordinates on the manifold Q. Restricting to coordinate systems which are time-invariant, we can then make the following definition.

Definition 2.3.2 (Generalized coordinates). Given a mechanical system with configuration space Q, a *system of generalized coordinates* $(q_1, \ldots, q_m)$ for this system is a local coordinate system on Q.

So we see that there there is a very natural way of describing the possible configurations of mechanical systems in terms of smooth manifolds. What about the dynamics?

Recall that the Lagrangian $\mathcal{L}$ of a system was defined in (1.76) in terms of the system's potential energy (which depends solely on its configuration) and its kinetic energy (which depends on its configuration and velocity when written as a function of generalized coordinates). Now the velocity of the system is simply the time rate-of-change of its configuration, which we can regard as a tangent vector to Q. Thus, the value of the Lagrangian is determined by specifying the system's configuration, i.e., a point $q \in Q$, together with a tangent vector $\dot{q} \in T_q(Q)$. But the pair $(q, \dot{q})$ is just a point in the tangent bundle TQ, so we see that the Lagrangian is also quite naturally a real-valued function on TQ.

Definition 2.3.3 (Lagrangian). Let Q be the configuration space for a mechanical system. A *Lagrangian* $\mathfrak{L}$ is any real-valued function $\mathfrak{L}\colon TQ \to \mathbb{R}$ on Q's tangent bundle TQ.

So now it remains only to construct a version of Lagrange's equations (1.77) using $\mathfrak{L}$. We will achieve this through the introduction of a pair of forms on the tangent bundle TQ which are the analogues of the canonical 1- and 2-form on the cotangent bundle.

2.3.2 The Lagrangian forms

We begin with the following definitions.

Definition 2.3.4 (Lagrangian 1-form, Lagrangian 2-form). Let Q be the configuration manifold of a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$ and let Θ and Ω denote the canonical 1- and 2-forms on the cotangent bundle T^*Q. Then the *Lagrangian 1-form* $\Theta_L \in \Omega^1(TQ)$ is the 1-form defined by

$$\Theta_L = (\mathbb{F}\mathfrak{L})^*\Theta. \tag{2.9}$$

Similarly, the *Lagrangian 2-form* $\Omega_L \in \Omega^2(TQ)$ is defined by

$$\Omega_L = (\mathbb{F}\mathfrak{L})^*\Omega. \tag{2.10}$$

Theorem 2.3.1 (Lagrangian forms in local coordinates). *Let Q be the configuration manifold of a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$. Let (U, φ) be a chart on Q with coordinate functions $(q_1, \dots, q_n)$, and let $(\hat{U}, \hat{\varphi})$ be the corresponding natural chart on TQ with natural coordinates $(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n)$. Then on $\hat{U}$, the Lagrangian 1-form Θ_L can be written as*

$$\Theta_L = \sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\, dq_i, \tag{2.11}$$

and the Lagrangian 2-form Ω_L can be written as

$$\Omega_L = \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j}\, dq_i \wedge dq_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j}\, dq_i \wedge d\dot{q}_j. \tag{2.12}$$

Proof. Let $(\tilde{U}, \tilde{\varphi})$ denote the corresponding natural chart on the cotangent bundle T^*Q with natural coordinates $(q_1, \dots, q_n, p_1, \dots, p_n)$. By Theorem B.2.7, the canonical 1-form $\Theta \in \Omega^1(T^*Q)$ can be written locally on $\tilde{U}$ as

$$\Theta = \sum_{i=1}^{n} p_i\, dq_i,$$

and by Theorem 2.2.1, the map $\mathbb{F}\mathfrak{L}$ can be written locally with respect to the natural coordinate systems on $\hat{U}$ and $\tilde{U}$ as

$$\mathbb{F}\mathfrak{L}(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n) = \left(q_1, \dots, q_n, \frac{\partial \mathfrak{L}}{\partial \dot{q}_1}, \dots, \frac{\partial \mathfrak{L}}{\partial \dot{q}_n} \right).$$

Consequently, the pullback $\Theta_L = (\mathbb{F}\mathfrak{L})^*\Theta$ can be written locally on $\hat{U}$ as

$$\begin{aligned}
(\mathbb{F}\mathfrak{L})^*\Theta &= (\mathbb{F}\mathfrak{L})^* \left(\sum_{i=1}^{n} p_i \, dq_i \right) \\
&= \sum_{i=1}^{n} (\mathbb{F}\mathfrak{L})^* p_i \, (\mathbb{F}\mathfrak{L})^* dq_i \\
&= \sum_{i=1}^{n} p_i \circ (\mathbb{F}\mathfrak{L}) \, dq_i \\
&= \sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i} \, dq_i.
\end{aligned}$$

Similarly, let $\Omega \in \Omega^2(T^*Q)$ be the canonical 2-form on T^*Q. Then the commutativity of the exterior derivative with pullbacks implies that

$$\Omega_L = (\mathbb{F}\mathfrak{L})^*\Omega = (\mathbb{F}\mathfrak{L})^*(-d\Theta) = -d(\mathbb{F}\mathfrak{L})^*\Theta = -d\Theta_L,$$

so the local coordinate expression for Ω_L can be computed by exterior differentiating the local coordinate expression for Θ_L:

$$\begin{aligned}
\Omega_L &= -d\Theta_L \\
&= -d \left(\sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i} \, dq_i \right) \\
&= -\sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \, dq_j + \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \, d\dot{q}_j \right) \wedge dq_i \qquad (2.13) \\
&= \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \, dq_i \wedge dq_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \, dq_i \wedge d\dot{q}_j.
\end{aligned}$$

□

Corollary 2.3.2 (Necessary and sufficient conditions for Ω_L to be symplectic). *Let Q be the configuration manifold of a mechanical system with Lagrangian*

$\mathfrak{L}\colon TQ \to \mathbb{R}$. Then the Lagrangian 2-form $\Omega_L \in \Omega^2(TQ)$ is a symplectic form if and only if $\mathfrak{L}$ is regular.

Proof. We first observe that Ω_L is obviously a closed 2-form, so it is symplectic if and only if it is nondegenerate. Let (U, φ) be a coordinate chart on Q and $(\hat{U}, \hat{\varphi})$ the corresponding natural chart on TQ with natural coordinates $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. Then we may represent Ω_L on $\hat{U}$ by a $2n \times 2n$ matrix W as in equation (B.3) in Appendix B; by equation (2.12), this matrix is

$$W = \begin{pmatrix} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} - \frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j} & \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \\ -\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} & 0 \end{pmatrix}. \tag{2.14}$$

Thus, Theorem B.1.3 implies that Ω_L is nondegenerate precisely when the $n \times n$ matrix

$$\left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \right)$$

is invertible. But Theorem 2.2.2 shows that this is precisely the condition for $\mathfrak{L}$ to be regular. □

Remark 2.3.1. We shall be concerned primarily with regular Lagrangians. Note that while this condition *does* impose a restriction on the set of Lagrangians that we will consider, one might expect (arguing heuristically) that it should not be *too* terribly restrictive; for observe that the invertibility of the matrix (2.8) corresponds exactly with the nonsingularity of Lagrange's equations of motion (1.77) when viewed as a linear system in the variables $\ddot{q}_1, \ldots, \ddot{q}_n$.

2.3.3 Lagrangian and second order vector fields

Definition 2.3.5 (Energy function). Let Q be the configuration manifold of a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$. Then the corresponding *energy function* $E\colon TQ \to \mathbb{R}$ is defined by

$$\begin{aligned} E\colon TQ \to& \mathbb{R} \\ E(v) = (\mathbb{F}\mathfrak{L}(v))&(v) - \mathfrak{L}(v). \end{aligned} \tag{2.15}$$

Theorem 2.3.3 (The energy function in local coordinates). *Let Q be the configuration manifold of a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$, and let $E\colon TQ \to \mathbb{R}$ be the corresponding energy function. Given a chart (U, φ) on Q, let $(\hat{U}, \hat{\varphi})$ be the corresponding natural chart on TQ with natural coordinate functions $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. Then on $\hat{U}$, E can be written as*

$$E(q, \dot{q}) = \sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}(q, \dot{q})\, \dot{q}_i - \mathfrak{L}(q, \dot{q}). \tag{2.16}$$

Proof. Any element $v \in \hat{U}$ can be written with respect to the natural coordinates on $\hat{U}$ as

$$v = \sum_{i=1}^{n} c_i \frac{\partial}{\partial q_i} \tag{2.17}$$

for some constants $c_i \in \mathbb{R}$. Letting $(\tilde{U}, \tilde{\varphi})$ denote the corresponding natural chart on T^*Q with natural coordinate functions $(q_1, \ldots, q_n, p_1, \ldots, p_n)$, Theorem 2.2.1 shows that

$$\begin{aligned}
(\mathbb{F}\mathfrak{L}(v))(v) &= \left(\sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}(v)\, dq_i\right)(v) \\
&= \left(\sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}(v)\, dq_i\right)\left(\sum_{i=1}^{n} c_i \frac{\partial}{\partial q_i}\right) \\
&= \sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}(v)\, c_i \\
&= \sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}(v)\, \dot{q}_i(v)
\end{aligned}$$

by equation (2.17). Thus, we can write E locally on $\hat{U}$ as (2.16). $\square$

Definition 2.3.6 (Lagrangian vector field)*.* Let Q be the configuration manifold for a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$, corresponding energy function $E\colon TQ \to \mathbb{R}$, and Lagrangian 2-form $\Omega_L \in \Omega^2(TQ)$. A vector field $X \in \mathfrak{X}(TQ)$ is called *Lagrangian* if

$$i_X \Omega_L = dE. \tag{2.18}$$

Equation (2.18) is referred to as the *Lagrangian condition.*

Remark 2.3.2*.* Note that if $\mathfrak{L}$ is a regular Lagrangian, then Ω_L is symplectic (Corollary 2.3.2), and the Lagrangian condition is precisely the statement that X is the Hamiltonian vector field X_E!

Theorem 2.3.4 (Lagrangian vector fields in local coordinates)*. Let Q be the configuration manifold for a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$ and corresponding energy function $E\colon TQ \to \mathbb{R}$, and let $X \in \mathfrak{X}(TQ)$ be a vec-*

tor field. Let (U, φ) *be a chart on the manifold* Q*, and let* $(\hat{U}, \hat{\varphi})$ *be the corresponding natural chart on* TQ *with natural coordinates* $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$*. Then on* $\hat{U}$*,* X *can be written as*

$$X(v) = \sum_{i=1}^{n} a_i(v) \left.\frac{\partial}{\partial q_i}\right|_v + \sum_{i=1}^{n} b_i(v) \left.\frac{\partial}{\partial \dot{q}_i}\right|_v$$

for smooth functions $a_i, b_i \in C^\infty(\hat{U})$*, and* $X|_{\hat{U}}$ *is Lagrangian if and only if*

$$\sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j} \dot{q}_j - \frac{\partial \mathfrak{L}}{\partial q_i} = \sum_{j=1}^{n} \left(\frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j} a_j - \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} a_j - \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_j \right) \tag{2.19}$$

and

$$\sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \dot{q}_j = \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} a_j \tag{2.20}$$

for all $1 \leq i \leq n$*. Equivalently, defining the* $n \times 1$ *column matrices*

$$A = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad C = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad D = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix},$$

where

$$c_i = \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j} \dot{q}_j - \frac{\partial \mathfrak{L}}{\partial q_i}, \quad d_i = \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \dot{q}_j$$

for all $1 \leq i \leq n$*,* $X|_{\hat{U}}$ *is Lagrangian if and only if*

$$\begin{pmatrix} C \\ D \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j} - \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} & -\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \\ \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} & 0 \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} \tag{2.21}$$

on all of $\hat{U}$*.*

Proof. By equation (2.16) of Theorem 2.3.3, E can be written on $\hat{U}$ as

$$E = \sum_{i=1}^{n} \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}(q, \dot{q})\, \dot{q}_i - \mathfrak{L}(q, \dot{q}),$$

and therefore its exterior derivative is

$$
\begin{aligned}
dE &= \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial q_j}\left(\frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\right)\dot{q}_i + \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\frac{\partial}{\partial q_j}\left(\dot{q}_i\right)\right)\, dq_j \\
&\quad + \sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial}{\partial \dot{q}_j}\left(\frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\right)\dot{q}_i + \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\frac{\partial}{\partial \dot{q}_j}\left(\dot{q}_i\right)\right)\, d\dot{q}_j \\
&\quad - \sum_{i=1}^{n}\left(\frac{\partial \mathfrak{L}}{\partial q_i}\, dq_i + \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\, d\dot{q}_i\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j}\dot{q}_i\, dq_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j}\dot{q}_i\, d\dot{q}_j + \sum_{i=1}^{n}\frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\, d\dot{q}_i \\
&\quad - \sum_{i=1}^{n}\left(\frac{\partial \mathfrak{L}}{\partial q_i}\, dq_i + \frac{\partial \mathfrak{L}}{\partial \dot{q}_i}\, d\dot{q}_i\right) \\
&= \sum_{i,j=1}^{n}\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j}\dot{q}_i\, dq_j + \sum_{i,j=1}^{n}\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j}\dot{q}_i\, d\dot{q}_j - \sum_{j=1}^{n}\frac{\partial \mathfrak{L}}{\partial q_j}\, dq_j.
\end{aligned}
\tag{2.22}
$$

Similarly, using the local expression for the Lagrangian 2-form Ω_L given by equation (2.12) of Theorem 2.3.1 together with the local expression for X, we can obtain a local expression for the contraction $i_X\Omega_L$:

$$
\begin{aligned}
i_X\Omega_L &= \left(\sum_{i,j=1}^{n}\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j}\, dq_i \wedge dq_j\right)\left(\sum_{i=1}^{n} a_i\,\frac{\partial}{\partial q_i} + \sum_{i=1}^{n} b_i\,\frac{\partial}{\partial \dot{q}_i}\right) \\
&\quad + \sum_{i,j=1}^{n}\left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j}\, dq_i \wedge d\dot{q}_j\right)\left(\sum_{i=1}^{n} a_i\,\frac{\partial}{\partial q_i} + \sum_{i=1}^{n} b_i\,\frac{\partial}{\partial \dot{q}_i}\right) \\
&= \sum_{i,j=1}^{n}\left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j}a_i - \frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j}a_i\right)\, dq_j + \sum_{i,j=1}^{n}\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j}a_i\, d\dot{q}_j \\
&\quad - \sum_{i,j=1}^{n}\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j}b_i\, dq_j.
\end{aligned}
\tag{2.23}
$$

Now by definition, X is Lagrangian if and only if $i_X\Omega_L = dE$. Equating the coefficients of the 1-forms dq_j and $d\dot{q}_j$ for all $1 \le j \le n$ in equations (2.22) and (2.23) and interchanging the indices i and j shows that X is Lagrangian if and only if the functions a_i and b_i are solutions of the simultaneous system of equations determined by (2.19) and (2.20) on all of $\hat{U}$. This system is equivalent to the matrix equation (2.21), which proves the result. □

Recall that the solutions of Lagrange's equations are solutions of a second-order system of ODE's in the variables $q_1, \ldots, q_n$. The geometric analogue of solutions of this type are *base integral curves* of *second order vector fields.*

Definition 2.3.7 (Second order vector field, base integral curve). Let Q be a manifold. A vector field $X \in \mathfrak{X}(TQ)$ on the tangent bundle $\pi\colon TQ \to Q$ is called a *second order vector field* if

$$d\pi \circ X = \mathrm{id}_{TQ}\,. \tag{2.24}$$

If X is a second order vector field and $c(t)$ is an integral curve of X, its projection $\pi \circ c(t)$ onto the base manifold Q is called a *base integral curve of* X.

Let Q be a manifold, X a second order vector field on TQ, (U, φ) a coordinate chart on Q, and $(\hat{U}, \hat{\varphi})$ the corresponding natural chart on TQ with natural coordinate functions $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. Since X is a vector field on TQ, then letting $v = (q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n) \in \hat{U}$, X can be locally written on $\hat{U}$ as

$$X(v) = \sum_{i=1}^{n} a_i(v)\, \frac{\partial}{\partial q_i}\bigg|_v + \sum_{i=1}^{n} b_i(v)\, \frac{\partial}{\partial \dot{q}_i}\bigg|_v \tag{2.25}$$

for some smooth functions $a_i, b_i \in C^\infty(\hat{U})$. The second order condition (2.24) then shows that

$$\begin{aligned} v = (q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n) &= d\pi(X(v)) \\ &= d\pi \left(\sum_{i=1}^n a_i(v) \left.\frac{\partial}{\partial q_i}\right|_v + \sum_{i=1}^n b_i(v) \left.\frac{\partial}{\partial \dot{q}_i}\right|_v \right) \\ &= \sum_{i=1}^n a_i(v) \left.\frac{\partial}{\partial q_i}\right|_{\pi(v)}. \end{aligned} \tag{2.26}$$

Equating coordinates in (2.26) shows that

$$\dot{q}_i(v) = a_i(v)$$

for all $1 \leq i \leq n$, so that (2.25) can be rewritten as:

$$X(v) = \sum_{i=1}^n \dot{q}_i(v) \left.\frac{\partial}{\partial q_i}\right|_v + \sum_{i=1}^n b_i(v) \left.\frac{\partial}{\partial \dot{q}_i}\right|_v \tag{2.27}$$

for some smooth functions $b_i \in C^\infty(\hat{U})$. Conversely, any vector field $X \in \mathfrak{X}(TQ)$ that has the local form (2.27) is obviously second order.

Now let $c(t)$ be an integral curve of X on $\hat{U}$, and write $c(t)$ in local coordinates as

$$c(t) = (q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t)).$$

Temporarily denoting differentiation with respect to t by $'$, the fact that $c(t)$ is an integral curve of X means that $c'(t) = X(c(t))$ for all t. Now

$$c'(t) = \sum_{i=1}^n q_i'(t) \frac{\partial}{\partial q_i} + \sum_{i=1}^n \dot{q}_i'(t) \frac{\partial}{\partial \dot{q}_i}, \tag{2.28}$$

and

$$X(c(t)) = \sum_{i=1}^{n} \dot{q}_i(t)\frac{\partial}{\partial q_i} + \sum_{i=1}^{n} b_i(c(t))\frac{\partial}{\partial \dot{q}_i} \tag{2.29}$$

by (2.27), so equating coordinates in (2.28) and (2.29) shows that

$$\begin{aligned} \dot{q}_i'(t) =& b_i(q(t), \dot{q}(t)) \\ q_i'(t) =& \dot{q}(t). \end{aligned} \tag{2.30}$$

Thus, the coordinate functions $(q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t))$ satsify the second-order system of ODE's given by (2.30). Since the coordinate functions for the corresponding base integral curve are given by $\pi \circ c(t) = (q_1(t), \ldots, q_n(t))$, this justifies our earlier assertion that base integral curves of second order vector fields are the geometric analogue of solutions of second order ODEs. We summarize these results below, as they will be useful for future computations.

Theorem 2.3.5 (Second order vector fields and integral curves). *Let Q be a manifold and $X \in \mathfrak{X}(TQ)$ a vector field on TQ. Let (U, φ) be a chart on Q, and let $(\hat{U}, \hat{\varphi})$ be the corresponding natural chart on TQ with natural coordinates $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. Then X is second order if and only if there exist smooth functions $b_i \in C^\infty(\hat{U})$ such that X can be written as*

$$X(v) = \sum_{i=1}^{n} \dot{q}_i(v) \left.\frac{\partial}{\partial q_i}\right|_v + \sum_{i=1}^{n} b_i(v) \left.\frac{\partial}{\partial \dot{q}_i}\right|_v$$

for all $v \in \hat{U}$.

Furthermore, if X is second order and $c(t)$ is an integral curve of X on $\hat{U}$ with coordinate functions

$$c(t) = (q(t), \dot{q}(t)) = (q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t)),$$

then

$$\frac{d}{dt}q_i(t) = \dot{q}_i(t),$$

the corresponding base integral curve $\bar{c}(t) = \pi \circ c(t)$ *is given by*

$$\bar{c}(t) = q(t),$$

and $q(t) = (q_1(t), \ldots, q_n(t))$ *is a solution of the second order system of ordinary differential equations*

$$\ddot{q}(t) = b(q(t), \dot{q}(t)). \tag{2.31}$$

Remark 2.3.3. Note that any curve $c(t) = (q_1(t), \ldots, q_n(t))$ on the base manifold Q has a unique corresponding curve

$$\hat{c}(t) = (q_1(t), \ldots, q_n(t), q_1'(t), \ldots, q_n'(t))$$

on the tangent bundle TQ which gives the velocity of $c(t)$, and the vector field X along $\hat{c}(t)$ determined by

$$X(\hat{c}(t)) = \hat{c}'(t)$$

is a second order vector field. Thus, *every curve on the base manifold Q is the base integral curve of a second order vector field on TQ*. Thus, for the purposes of mechanics, we are only interested in integral curves of *second order* Lagrangian vector fields, as these are the only vector fields arising from curves on the base manifold Q (i.e., these are the only vector fields on TQ associated to *system trajectories*).

2.3.4 Lagrange's equations for conservative mechanical systems

With these results in hand, we are now ready to translate Lagrange's equations to the manifold setting.

Theorem 2.3.6 (Lagrange's equations for conservative mechanical systems). *Let Q be the configuration manifold for a conservative mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$ and corresponding energy function $E\colon TQ \to \mathbb{R}$. Then the following hold:*

1. *If $X \in \mathfrak{X}(TQ)$ is a second order Lagrangian vector field, then the base integral curves of X on Q are solutions of Lagrange's equations. More precisely, given any chart (U, φ) on Q, let $(\hat{U}, \hat{\varphi})$ be the corresponding natural chart on TQ with natural coordinates $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. If $c(t)$ is any integral curve of X with coordinate functions*

$$c(t) = (q(t), \dot{q}(t)) = (q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t)),$$

then

$$\frac{d}{dt}q(t) = \dot{q}(t),$$

$c(t)$'s base integral curve $\bar{c}(t) = \pi \circ c(t)$ is given by

$$\bar{c}(t) = q(t),$$

and $q(t)$ satisfies Lagrange's equations:

$$\frac{d}{dt}\left(\frac{\partial \mathfrak{L}}{\partial \dot{q}}\right) = \frac{\partial \mathfrak{L}}{\partial q}. \tag{2.32}$$

2. *If $\mathfrak{L}$ is regular, then the Lagrangian 2-form $\Omega_L \in \Omega^2(TQ)$ is symplectic, the Hamiltonian vector field X_E corresponding to the energy function E is second order and is the unique Lagrangian vector field on TQ, and the base integral curves of X_E are the unique solutions of Lagrange's equations. Furthermore, defining the $n \times 1$ column matrix*

$$D = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix},$$

where

$$d_i = \frac{\partial \mathfrak{L}}{\partial q_i} - \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_j,$$

the integral curves $c(t) = (q(t), \dot{q}(t))$ of X_E on $\hat{U}$ are the solutions of the second-order system of ordinary differential equations given by

$$\ddot{q}(t) = \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \right)^{-1} D. \tag{2.33}$$

Proof. **Claim 1:** Let (U, φ) be a chart on Q and let $(\hat{U}, \hat{\varphi})$ be the corresponding natural chart on TQ with natural coordinates $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. Write $X|_{\hat{U}}$ as

$$X = \sum_{i=1}^{n} a_i \frac{\partial}{\partial q_i} + \sum_{i=1}^{n} b_i \frac{\partial}{\partial \dot{q}_i} \tag{2.34}$$

for smooth functions $a_i, b_i \in C^\infty(\hat{U})$. If X is second order and Lagrangian, Theorem 2.3.5 shows that $a_i = \dot{q}_i$ for all $1 \le i \le n$, so equation (2.20) holds identically and equation (2.19) reduces to

$$\sum_{j=1}^{n} \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_j + \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_j \right) - \frac{\partial \mathfrak{L}}{\partial q_i} = 0. \tag{2.35}$$

Now let $c(t)$ be an integral curve of X, and write $c(t)$ in coordinates as

$$c(t) = (q(t), \dot{q}(t)) = (q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t)).$$

Then equation (2.31) of Theorem 2.3.5 together with equation (2.35) show that

$$\sum_{j=1}^{n} \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \ddot{q}_j + \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_j \right) - \frac{\partial \mathfrak{L}}{\partial q_i} = 0. \tag{2.36}$$

But equation (2.36) is precisely the expanded form of Lagrange's equations (2.32). Since the coordinates of the base integral curve $\bar{c}(t) = \pi \circ c(t)$ are given by

$$\bar{c}(t) = q(t),$$

this shows that these coordinate functions are solutions of (2.32), as claimed.

Claim 2: Observe that if $\mathfrak{L}$ is regular, Corollary 2.3.2 shows that the Lagrangian 2-form Θ_L is symplectic, and therefore the Lagrangian condition (2.18) is equivalent to the statement that $X = X_E$, the unique Hamiltonian vector field on TQ corresponding to the energy function E. By Theorem 2.3.5, X_E is second order if and only if $a_i = \dot{q}_i$ in equation (2.34) for all $1 \leq i \leq n$. Now since $\mathfrak{L}$ is regular, Theorem 2.2.2 shows that the matrix in equation (2.21) is invertible, and therefore has a unique solution, which we have already established is X_E. So to prove that X_E is second order, it suffices to show that the system defined by (2.21) remains consistent when the substitution $a_i = \dot{q}_i$ is made.

Equation (2.21) is equivalent to the simultaneous system of equations given by (2.19) and (2.20). Equation (2.20) is obviously satisfied when $a_i = \dot{q}_i$, and equation (2.19) can be rearranged to give:

$$\sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_j = \frac{\partial \mathfrak{L}}{\partial q_i} - \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_j \tag{2.37}$$

for all $1 \leq i \leq n$. Defining the matrices

$$B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad D = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix}$$

where

$$d_i = \frac{\partial \mathfrak{L}}{\partial q_i} - \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_j,$$

equation (2.37) can be written as the matrix equation

$$D = \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \right) B \tag{2.38}$$

and the regularity of the Lagrangian $\mathfrak{L}$ together with a final application of Theorem 2.2.2 shows that (2.38) can still be solved, hence is consistent. Thus, X_E must be second order, and equations (2.31) and (2.38) together prove (2.33).

Finally, to show that the integral curves of X_E are the *only* solutions of Lagrange's equations, simply observe that if $c(t) = (q(t), \dot{q}(t))$ is any solution of Lagrange's equations, then $q(t)$ satisfies (2.36), and the regularity of $\mathfrak{L}$ implies that (2.36) can be solved to yield (2.33), showing that $c(t)$ is in fact an integral curve of X_E. □

2.3.5 Lagrange's equations with external forces

As was done in Section 1.2.3, we would like to extend Lagrange's equations to admit nonconservative external forces.

2.3.5.1 Force fields and time-varying forms

We begin by describing a geometric way to represent external forces acting on a system.

Recall (see the remark following Theorem 2.3.5) that in the geometric Lagrangian formulation, system trajectories are obtained as base integral curves of second order vector fields. Let Q be the configuration manifold of a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$, and suppose that X is a second order Lagrangian vector field with corresponding base integral curve $\bar{c}(t)$. Let (U, φ) be a chart on Q and $(\hat{U}, \hat{\varphi})$ the corresponding natural chart on TQ with natural coordinates $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. We saw in equation (2.36) in the proof of Theorem 2.3.6 that, under these assumptions, equation (2.19) reduced to Lagrange's equations of motion in the case of a conservative system. Now equation (2.19) was itself obtained by requiring that X be a Lagrangian vector field, i.e., by requiring that $i_X\Omega_L = dE$, or equivalently, that

$$dE - i_X\Omega_L = 0.$$

Using equations (2.22) and (2.23), we can write $dE - i_X\Omega_L$ locally (after applying the second order hypothesis on X) as

$$
\begin{aligned}
dE - i_X \Omega_L = & \left(\sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_i \, dq_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \dot{q}_i \, d\dot{q}_j - \sum_{j=1}^{n} \frac{\partial \mathfrak{L}}{\partial q_j} \, dq_j \right) \\
& - \sum_{i,j=1}^{n} \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_i - \frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j} \dot{q}_i \right) \, dq_j \\
& - \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \dot{q}_i \, d\dot{q}_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_i \, dq_j \\
= & \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_i \, dq_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial q_i \partial \dot{q}_j} \dot{q}_i \, dq_j - \sum_{j=1}^{n} \frac{\partial \mathfrak{L}}{\partial q_j} \, dq_j .
\end{aligned}
\tag{2.39}
$$

Now recall (Theorem 2.3.5) that the parameter b_i is the second derivative with respect to time of the ith coordinate function $q_i(t)$ of the base integral curve $\bar{c}(t)$; making this substitution in (2.39) gives:

$$
dE - i_X \Omega_L = \sum_{j=1}^{n} \left(\sum_{i=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_j \partial \dot{q}_i} \ddot{q}_i + \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_j \partial q_i} \dot{q}_i - \frac{\partial \mathfrak{L}}{\partial q_j} \right) dq_j . \tag{2.40}
$$

But now, comparing the right-hand side of (2.40) to Lagrange's equations with external forces (1.84), we see that the coefficient of each of the 1-forms dq_j is nothing more than the component of the generalized external force Q_j^{ext} acting along the jth coordinate direction!

Motivated by this observation, and we make the following definition.

Definition 2.3.8 (Lagrangian force 1-form). Let Q be the configuration manifold of a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$, corresponding energy function $E\colon TQ \to \mathbb{R}$, and Lagrangian 2-form Ω_L. Given a vector field

$X \in \mathfrak{X}(TQ)$, the 1-form $\Phi_L(X) \in \Omega^1(TQ)$ defined by

$$\Phi_L(X) = dE - i_X \Omega_L \tag{2.41}$$

is called the *Lagrangian force* 1*-form*.

Equation (2.40) suggests that when $X \in \mathfrak{X}(TQ)$ is second-order, the Lagrangian force 1-form $\Phi_L(X)$ defined in (2.41) represents the external generalized forces that must act on the system in order to produce the system trajectories that are the base integral curves of X.

This suggests that, in order to determine the trajectories of a Lagrangian system being acted upon by an external force, we should equate $\Phi_L(X)$ with something that represents the external forces acting on the system, and then determine which, if any, second order vector fields on TQ satisfy that equality; equation (2.40) then suggets that the base integral curves of such a vector field give the system trajectories. Thus, we would like to develop a general method for representing generalized forces acting on a mechanical system. Furthermore, we would like to admit representations of forces that depend upon the generalized coordinates and their velocities, and that are allowed to vary with time. Thus, we are very naturally led to consider time-varying 1-forms on the tangent bundle.

Definition 2.3.9 (Time-varying 1-form)*.* Let M be a smooth manifold. A *time-*

varying 1-*form* ω *on* M is a smooth mapping $\omega\colon I \times M \to T^*M$ that satisfies

$$\pi \circ \omega(t, m) = m \tag{2.42}$$

for all $(t, m) \in I \times M$, where $I \subseteq \mathbb{R}$ is a connected submanifold of $\mathbb{R}$ and $\pi\colon T^*M \to M$ is the canonical projection map. Defining the map

$$\omega(t)\colon M \to T^*M$$

$$\omega(t) = \omega(t, \cdot)$$

for each $t \in I$, (2.42) shows that $\pi \circ \omega(t) = \mathrm{id}_M$ for all $t \in I$, so that $\omega(t) \in \Omega^1(M)$ for all $t \in I$.

We will construct representations of external forces with the aid of the following theorem.

Theorem 2.3.7 (Force fields and external force 1-forms). *Let* M *be a manifold,* $I \subseteq \mathbb{R}$ *a connected submanifold of* $\mathbb{R}$*, and* $F\colon I \times TM \to T^*M$ *a smooth mapping such that the following diagram commutes:*

$$\begin{array}{ccc} I \times TM & \xrightarrow{\quad F \quad} & T^*M \\ & {\scriptstyle \Pi} \searrow \quad \swarrow {\scriptstyle \pi} & \\ & M & \end{array} \tag{2.43}$$

where

$$\Pi\colon I \times TM \to M$$

$$\Pi(t, v) = \pi(v)$$

and $\pi\colon TM \to M$, $\pi\colon T^*M \to M$ *are the canonical projection maps. Then* F *induces a time-varying* 1*-form* $\omega(t) \in \Omega^1(TQ)$ *according to*

$$\omega(t,v)(\xi) = F(t,v)(d\pi_v(\xi)) \tag{2.44}$$

for all $(t,v) \in I \times TM$ *and* $\xi \in T_v(TM)$.

Furthermore, let (U,φ) *be a chart on* M*, and let* $(\hat{U},\hat{\varphi})$ *and* $(\tilde{U},\tilde{\varphi})$ *be the corresponding natural charts on* TM *and* T^*M *with corresponding natural coordinates* $(q_1,\ldots,q_n,\dot{q}_1,\ldots,\dot{q}_n)$ *and* $(q_1,\ldots,q_n,p_1,\ldots,p_n)$*, respectively. Then* F *can be locally written on* $\hat{U}$ *as*

$$F(t,q,\dot{q}) = (q_1,\ldots,q_n,F_1(t,q,\dot{q}),\ldots,F_n(t,q,\dot{q})) \tag{2.45}$$

for smooth functions $F_i \in C^\infty(I \times \hat{U})$*, and* ω *can be written on* $\hat{U}$ *as*

$$\omega(t,q,\dot{q}) = \sum_{i=1}^{n} F_i(t,q,\dot{q})\, dq_i. \tag{2.46}$$

Proof. Fix $x \in M$, let $v \in (TM)_x = T_x(M)$, and let $\xi \in T_v(TM)$. The fact that the diagram (2.43) is commutative means that

$$\pi \circ F(t,v) = \Pi(t,v) = \pi(v) = x \tag{2.47}$$

so that $F(t,v) \in (T^*M)_x = (T_x(M))^*$. Since $d\pi_v\colon T_v(TM) \to T_{\pi(v)}(M) = T_x(M)$, this shows that the pairing ω given in (2.44) is well-defined, hence defines a time-varying 1-form $\omega(t) \in \Omega^1(TM)$, as claimed. Smoothness follows immediately from the smoothness of the maps in (2.44).

Now equation (2.47) shows that $F(t,\cdot)$ maps the fiber $(TM)_x$ over x into the fiber $(T^*M)_x$ over x, and therefore F can be written locally on $\hat{U}$ as

$$F(t,q,\dot{q}) = (q_1,\ldots,q_n,F_1(t,q,\dot{q}),\ldots,F_n(t,q,\dot{q}))$$

with respect to the natural coordinate systems $(\hat{U},\hat{\varphi})$ and $(\tilde{U},\tilde{\varphi})$ on $\hat{U}$ and $\tilde{U}$, respectively, for smooth functions $F_i \in C^\infty(I\times\hat{U})$. Now fix $(t,v)=(t,q,\dot{q}) \in I\times\hat{U}$, and let $\xi \in T_v(TM)$. Then ξ can be written as

$$\xi = \sum_{i=1}^n a_i \frac{\partial}{\partial q_i} + \sum_{i=1}^n b_i \frac{\partial}{\partial \dot{q}_i}$$

for constants $a_i, b_i \in \mathbb{R}$, and equations (2.44) and (2.45) show that

$$\begin{aligned}
\omega(t,v)(\xi) &= F(t,q,\dot{q})(\xi)\\
&= \left(\sum_{i=1}^n F_i(t,q,\dot{q})\,dq_i\right)\left(d\pi_v\left(\sum_{i=1}^n a_i\frac{\partial}{\partial q_i} + \sum_{i=1}^n b_i\frac{\partial}{\partial \dot{q}_i}\right)\right)\\
&= \left(\sum_{i=1}^n F_i(t,q,\dot{q})\,dq_i\right)\left(\sum_{i=1}^n a_i\frac{\partial}{\partial q_i}\right)\\
&= \sum_{i=1}^n F_i(t,q,\dot{q})\,a_i\\
&= \sum_{i=1}^n F_i(t,q,\dot{q})\,dq_i(\xi),
\end{aligned}$$

which proves (2.46). □

If we regard the manifold M in Theorem 2.3.7 as the configuration manifold of a mechanical system, then the commutativity of the diagram (2.43) is equivalent to the statement that, at each $t \in I$, the mapping $F(t,\cdot)$ associates to each generalized position and velocity $(q,\dot{q})$ a corresponding generalized momentum (q, p) at the same point in the configuration manifold Q. Furthermore, the local coordinate expression for the induced time-varying 1-form ω shows that it depends in a simple way upon the map F, and that it is quite explicitly a function of time and the generalized position and velocity. Thus, we can regard the map F as an intrinsic geometric representation of a force field, and therefore make the following definitions.

Definition 2.3.10 (Force field, external force 1-form)*.* If Q is the configuration manifold of a mechanical system and $I \subseteq \mathbb{R}$ is a connected submanifold of $\mathbb{R}$, a map $F\colon I \times TQ \to T^*Q$ satisfying (2.43) is called a *force field*, and the induced time-varying 1-form $\omega(t) \in \Omega^1(TQ)$ is called an *external force* 1*-form*.

2.3.5.2 Time-varying second order vector fields

Now that we have a method of expressing external forces using geometric objects (i.e., using the external force 1-form $\omega(t)$) we would like to be able to solve the equation

$$\Phi_L(X) = \omega(t) \tag{2.48}$$

for second order vector fields X. However, since we allow the external force 1-form $\omega(t)$ to vary in time, it seems unlikely that we should be able to solve (2.48) for vector fields X that do not also depend upon time. Thus, we must generalize our notion of second order vector fields to admit time-dependence.

Definition 2.3.11 (Time-varying vector field)*.* Let M be a smooth manifold and $I \subseteq \mathbb{R}$ a connected submanifold of $\mathbb{R}$. A *time-varying vector field on M*

is a smooth mapping $X: I \times M \to TM$ satisfying

$$\pi \circ X(t, m) = m \tag{2.49}$$

for all $(t, m) \in I \times M$, where $\pi: TM \to M$ is the canonical projection map. For each fixed $t \in I$, define

$$X(t): M \to TM$$
$$X(t) = X(t, \cdot).$$

Then equation (2.49) shows that $\pi \circ X(t) = \mathrm{id}_M$, so that $X(t)$ is a vector field on M for each $t \in I$.

Definition 2.3.12 (Time-varying second order vector field)*.* Let Q be a manifold and $I \subseteq \mathbb{R}$ a connected submanifold of $\mathbb{R}$. A *time-varying second order vector field* is a time-varying vector field $X: I \times TQ \to TTQ$ satisfying

$$d\pi \circ X(t, v) = v \tag{2.50}$$

for all $(t, v) \in I \times TQ$, where $\pi: TQ \to Q$ is the canonical projection map. For each fixed $t \in I$, equation (2.50) shows that $X(t)$ satisfies

$$d\pi \circ X(t) = \mathrm{id}_{TQ},$$

i.e., $X(t)$ is a second order vector field for each $t \in I$.

With these definitions, we have the following generalization of Theorem 2.3.5

Theorem 2.3.8 (Time-varying second order vector fields and integral curves). *Let Q be a manifold, $I \subseteq \mathbb{R}$ a connected submanifold of $\mathbb{R}$, and $X\colon I \times TQ \to TTQ$ a time-varying vector field on TQ. Let (U, φ) be a chart on Q, and let $(\hat{U}, \hat{\varphi})$ be the corresponding natural chart on TQ with natural coordinates $(q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n)$. Then X is second order if and only if there exist smooth functions $b_i \in C^\infty(I \times \hat{U})$ such that X can be written as*

$$X(t, v) = \sum_{i=1}^{n} \dot{q}_i(v) \left.\frac{\partial}{\partial q_i}\right|_v + \sum_{i=1}^{n} b_i(t, v) \left.\frac{\partial}{\partial \dot{q}_i}\right|_v$$

for all $v \in \hat{U}$.

Furthermore, if X is second order and $c(t)$ is an integral curve of X on $\hat{U}$ with coordinate functions

$$c(t) = (q(t), \dot{q}(t)) = (q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t)),$$

then

$$\frac{d}{dt} q_i(t) = \dot{q}_i(t),$$

the corresponding base integral curve $\bar{c}(t) = \pi \circ c(t)$ is given by

$$\bar{c}(t) = q(t),$$

and $q(t) = (q_1(t), \ldots, q_n(t))$ is a solution of the second order system of ordinary differential equations

$$\ddot{q}(t) = b(t, q(t), \dot{q}(t)). \tag{2.51}$$

Proof. This is proved in the same manner as Theorem 2.3.5. Letting

$$X(t,v) = \sum_{i=1}^{n} a_i(t,v)\frac{\partial}{\partial q_i}\bigg|_v + \sum_{i=1}^{n} b_i(t,v)\frac{\partial}{\partial \dot{q}_i}\bigg|_v$$

for smooth functions $a_i, b_i \in C^\infty(I \times \hat{U})$ and defining

$$v = (q_1, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_n) \in \hat{U}, \tag{2.52}$$

one computes

$$\begin{aligned} d\pi(X(t,v)) &= d\pi\left(\sum_{i=1}^{n} a_i(t,v)\frac{\partial}{\partial q_i}\bigg|_v + \sum_{i=1}^{n} b_i(t,v)\frac{\partial}{\partial \dot{q}_i}\bigg|_v\right) \\ &= \sum_{i=1}^{n} a_i(t,v)\frac{\partial}{\partial q_i}\bigg|_{\pi(v)}; \end{aligned} \tag{2.53}$$

comparing (2.52) and (2.53) shows that $d\pi(X(t,v)) = v$ (so that X is second order) if and only if

$$a_i(t,v) = \dot{q}_i(v)$$

for all $1 \le i \le n$.

Now let $c(t)$ be an integral curve of X, with coordinate functions

$$c(t) = c(q(t), \dot{q}(t)) = (q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t)).$$

Denoting differentiation with respect to t by $'$, the fact that c is an integral curve shows that

$$c'(t) = (q_1'(t), \ldots, q_n'(t), \dot{q}_1'(t), \ldots, \dot{q}_n'(t)) = X(t, c(t)) = \sum_{i=1}^{n} \dot{q}_i(t)\frac{\partial}{\partial q_i} + \sum_{i=1}^{n} b_i(t, q(t), \dot{q}(t))\frac{\partial}{\partial \dot{q}_i}$$

and by equating coordinates, we obtain

$$\frac{d}{dt}q(t) = \dot{q}(t),$$
$$\frac{d}{dt}\dot{q}(t) = b(t, q(t), \dot{q}(t)).$$

Thus, the coordinate functions $q(t)$ are solutions of the second order system of differential equations:

$$\ddot{q}(t) = b(t, q(t), \dot{q}(t)).$$

Since the base integral curve $\bar{c}(t)$ of $c(t)$ is given by

$$\bar{c}(t) = \pi \circ c(t) = q(t),$$

this proves the result. □

2.3.5.3 Lagrange's equations with external forces

Theorem 2.3.9 (Lagrange's equations with external forces). *Let Q be the configuration manifold of a mechanical system with Lagrangian $\mathfrak{L}\colon TQ \to \mathbb{R}$, and denote by $E\colon TQ \to \mathbb{R}$ and $\Omega_L \in \Omega^2(TQ)$ the corresponding energy function and Lagrangian 2-form. Let $I \subseteq \mathbb{R}$ be a connected submanifold of $\mathbb{R}$, and let $F\colon I \times TQ \to T^*Q$ be a force field with corresponding external force 1-form $\omega(t)$. Then the following hold*

1. *If $X(t) \in \mathfrak{X}(TQ)$ is a time-varying second order vector field satisfying*

$$\Phi_L(X(t)) = \omega(t), \tag{2.54}$$

where $\Phi_L(X(t))$ *is the Lagrangian force* 1*-form defined in equation* (2.41)*, then the base integral curves of* X *are solutions of Lagrange's equations with external force* F*.*

More precisely, let (U, φ) *be a chart on* Q *and let* $(\hat{U}, \hat{\varphi})$ *be the corresponding natural chart on* TQ *with natural coordinates* $(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n)$*. Then the external force* 1*-form* $\omega(t)$ *can be written in local coordinates on* $\hat{U}$ *as*

$$\omega(t, q, \dot{q}) = \sum_{i=1}^{n} F_i(t, q, \dot{q})\, dq_i$$

for smooth functions $F_i \in C^\infty(I \times \hat{U})$*, and if* $c(t)$ *is any integral curve of* X *with coordinate functions*

$$c(t) = (q(t), \dot{q}(t)) = (q_1(t), \dots, q_n(t), \dot{q}_1(t), \dots, \dot{q}_n(t)),$$

then

$$\frac{d}{dt} q(t) = \dot{q}(t),$$

$c(t)$*'s base integral curve* $\bar{c}(t) = \pi \circ c(t)$ *is given by*

$$\bar{c}(t) = q(t),$$

and $q(t)$ *satisfies Lagrange's equations:*

$$\frac{d}{dt}\left(\frac{\partial \mathfrak{L}}{\partial \dot{q}}\right) = \frac{\partial \mathfrak{L}}{\partial q} + F(t, q(t), \dot{q}(t)).$$

2. *If* $\mathfrak{L}$ *is regular, then there exists a unique time-varying vector field* $X(t) \in \mathfrak{X}(TQ)$ *satisfying* (2.54)*,* $X(t)$ *is second order, and the base integral*

curves of $X(t)$ are the unique solutions of Lagrange's equations with external force F. Furthermore, defining the $n \times 1$ column matrix

$$D = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix},$$

where

$$d_i = \frac{\partial \mathfrak{L}}{\partial q_i} + F_i(t, q, \dot{q}) - \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_j,$$

the integral curves $c(t) = (q(t), \dot{q}(t))$ of $X(t)$ on $\hat{U}$ are the solutions of the second-order system of ordinary differential equations given by

$$\ddot{q}(t) = \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \right)^{-1} D. \tag{2.55}$$

Proof. **Claim 1:** Write X locally on $\hat{U}$ as

$$X(t, v) = \sum_{i=1}^{n} a_i(t, v) \frac{\partial}{\partial q_i}\bigg|_v + \sum_{i=1}^{n} b_i(t, v) \frac{\partial}{\partial \dot{q}_i}\bigg|_v \tag{2.56}$$

for smooth functions $a_i, b_i \in C^\infty(I \times \hat{U})$. By equations (2.22) and (2.23), the Lagrangian force 1-form $\Phi_L(X(t))$ corresponding to the time-varying second order vector field $X(t)$ is

$$\begin{aligned} \Phi_L(X(t)) &= dE - i_{X(t)} \Omega_L \\ &= \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_i \, dq_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \dot{q}_i \, d\dot{q}_j - \sum_{j=1}^{n} \frac{\partial \mathfrak{L}}{\partial q_j} \, dq_j \\ &\quad - \sum_{i,j=1}^{n} \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} a_i - \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_j \partial q_i} a_i \right) dq_j - \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} a_i \, d\dot{q}_j \\ &\quad + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_i \, dq_j. \end{aligned} \tag{2.57}$$

Since X is second order by hypothesis, Theorem 2.3.8 shows that $a_i = \dot{q}_i$, and therefore equation (2.57) reduces to

$$\Phi_L(X(t)) = \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_i \, dq_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_j \partial q_i} \dot{q}_i \, dq_j - \sum_{j=1}^{n} \frac{\partial \mathfrak{L}}{\partial q_j} \, dq_j. \tag{2.58}$$

Using Theorem 2.3.7, we may write the external force 1-form $\omega(t)$ in local coordinates on $\hat{U}$ as

$$\omega(t, q, \dot{q}) = \sum_{i=1}^{n} F_i(t, q, \dot{q}) \, dq_i \tag{2.59}$$

for smooth functions $F_i \in C^\infty(I \times \hat{U})$, and since $\Phi_L(X(t)) = \omega(t)$ by hypothesis, equations (2.58) and (2.59) then show that

$$\sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} b_i \, dq_j + \sum_{i,j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_j \partial q_i} \dot{q}_i \, dq_j - \sum_{j=1}^{n} \frac{\partial \mathfrak{L}}{\partial q_j} \, dq_j = \sum_{j=1}^{n} F_j(t, q, \dot{q}) \, dq_j, \tag{2.60}$$

and equating each of the differentials dq_j then shows that

$$\sum_{i=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_j \partial \dot{q}_i} b_i + \sum_{i=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_j \partial q_i} \dot{q}_i - \frac{\partial \mathfrak{L}}{\partial q_j} = F_j(t, q, \dot{q}) \tag{2.61}$$

for all $1 \leq j \leq n$.

Now if $c(t)$ is an integral curve of $X(t)$ with coordinate functions

$$c(t) = (q(t), \dot{q}(t)) = (q_1(t), \ldots, q_n(t), \dot{q}_1(t), \ldots, \dot{q}_n(t)),$$

then Theorem 2.3.8 shows that

$$\begin{aligned} \frac{d}{dt} q(t) &= \dot{q}(t), \\ \frac{d}{dt} \dot{q}(t) &= b(t, q(t), \dot{q}(t)), \end{aligned}$$

and substitution into equation (2.61) gives

$$\sum_{i=1}^{n} \frac{\partial^2 \mathcal{L}}{\partial \dot{q}_j \partial \dot{q}_i} \ddot{q}_i + \sum_{i=1}^{n} \frac{\partial^2 \mathcal{L}}{\partial \dot{q}_j \partial q_i} \dot{q}_i - \frac{\partial \mathcal{L}}{\partial q_j} = F_j(t, q, \dot{q}), \tag{2.62}$$

for all $1 \leq j \leq n$, which is precisely the expanded form of Lagrange's equations with external forces (1.84).

Claim 2: If $\mathcal{L}$ is regular, then Corollary 2.3.2 shows that Ω_L is symplectic, and therefore there is a unique time-varying vector field $X(t) \in \mathfrak{X}(TQ)$ satisfying (2.54), namely, the vector field defined by $X(t) = (dE - \omega(t))^\sharp$. Thus, it remains only to show that $X(t)$ is second order. Writing $X(t)$ locally as in equation (2.56), Theorem 2.3.8 shows that $X(t)$ is second order if and only if $a_i = \dot{q}_i$ for all $1 \leq i \leq n$. Now since $X(t)$ is the *unique* solution of (2.54), then in order to show that $X(t)$ is second order, it suffices to show that (2.54) remains consistent when the substitution $a_i = \dot{q}_i$ is made. After making this substitution in local coordinates on $\hat{U}$, equation (2.54) is equivalent to (2.60), and equating each of the differentials dq_j gives the system of equations (2.61), which we must show is consistent. Observe that (2.61) can be rearranged to give

$$\sum_{i=1}^{n} \frac{\partial^2 \mathcal{L}}{\partial \dot{q}_j \partial \dot{q}_i} b_i = \frac{\partial \mathcal{L}}{\partial q_j} + F_j(t, q, \dot{q}) - \sum_{i=1}^{n} \frac{\partial^2 \mathcal{L}}{\partial \dot{q}_j \partial q_i} \dot{q}_i, \tag{2.63}$$

and that, defining the matrices

$$B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}, \quad D = \begin{pmatrix} d_1 \\ \vdots \\ d_n \end{pmatrix},$$

where

$$d_i = \frac{\partial \mathfrak{L}}{\partial q_i} + F_i(t, q, \dot{q}) - \sum_{j=1}^{n} \frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial q_j} \dot{q}_j,$$

equation (2.63) is then equivalent to the matrix equation

$$D = \left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \right) B. \tag{2.64}$$

Now since $\mathfrak{L}$ is regular, Theorem 2.2.2 shows that $\left(\frac{\partial^2 \mathfrak{L}}{\partial \dot{q}_i \partial \dot{q}_j} \right)$ is invertible, and therefore equation (2.64) can be solved. This shows that (2.54) remains consistent when $a_i = \dot{q}_i$ is assumed, proving that $X(t)$ is second order. A final application of Theorem 2.3.8 together with equation (2.64) then proves (2.55).

Finally, to see that the integral curves of $X(t)$ are the *only* solutions of Lagrange's equations when the external force F is applied, simply observe that if $c(t) = (q(t), \dot{q}(t))$ is any solution of (2.62), then the regularity of $\mathfrak{L}$ implies that (2.62) can be solved to give (2.55), so that $c(t)$ is an integral curve of $X(t)$. □

Remark 2.3.4. Observe that all throughout the preceding derivations, system trajectories on Q were obtained as base integral curves of second order vector fields, which were themselves obtained by pulling back the canonical symplectic 2-form Ω on T^*Q to the Lagrangian 2-form Ω_L on TQ by means of the fiber derivative $\mathbb{F}\mathfrak{L}$. Although Ω_L is not symplectic unless $\mathfrak{L}$ is regular, the fact that it still determines the system trajectories shows that even in the non-regular case, symplectic geometry still underpins Lagrangian mechanics.

2.4 Hamiltonian mechanics on manifolds

2.4.1 Hamiltonian mechanics derived from the Lagrangian

We now wish to derive a geometric formulation of Hamiltonian mechanics from the corresponding formulation of Lagrangian mechanics, as was done in Section 1.3.

2.4.1.1 Hyperregular Lagrangians and the Hamiltonian

Let Q be the configuration manifold for a mechanical system with Lagrangian $\mathcal{L}\colon TQ \to \mathbb{R}$ and corresponding energy function $E\colon TQ \to \mathbb{R}$. If $\mathcal{L}$ is hyperregular, we can precompose E with the inverse diffeomorphism $(\mathbb{F}\mathcal{L})^{-1}$ to obtain

$$\mathcal{H}\colon T^*Q \to \mathbb{R}$$
$$\mathcal{H}(\eta) = E \circ (\mathbb{F}\mathcal{L})^{-1}(\eta) = \eta\left((\mathbb{F}\mathcal{L})^{-1}(\eta)\right) - \mathcal{L}\left((\mathbb{F}\mathcal{L})^{-1}(\eta)\right).$$

Let $x \in Q$, let (U, φ) be a chart on a neighborhood U about x with coordinate functions $(q_1, \ldots, q_m)$, and let $(\hat{U}, \hat{\varphi})$ and $(\tilde{U}, \tilde{\varphi})$ be the corresponding natural charts on TQ and T^*Q with natural coordinate functions $(q_1, \ldots, q_m, \dot{q}_1, \ldots, \dot{q}_m)$ and $(q_1, \ldots, q_m, p_1, \ldots, p_m)$, respectively. By Theorem 2.3.3, the energy function E can be locally written on $\hat{U}$ as

$$E(q, \dot{q}) = \sum_{i=1}^{n} \frac{\partial \mathcal{L}}{\partial \dot{q}_i}(q, \dot{q})\, \dot{q}_i - \mathcal{L}(q, \dot{q}), \tag{2.65}$$

so by precomposing (2.65) with $(\mathbb{F}\mathcal{L})^{-1}$ we obtain an expression for $\mathcal{H}$ using the local coordinates $(q_1, \dots, q_m, \dot{q}_1, \dots, \dot{q}_m)$ on TQ:

$$\mathcal{H}(\eta) = \sum_{i=1}^{n} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \circ (\mathbb{F}\mathcal{L})^{-1}(\eta) \right) (\dot{q}_i \circ (\mathbb{F}\mathcal{L})^{-1}(\eta)) - \mathcal{L} \circ (\mathbb{F}\mathcal{L})^{-1}(\eta) \tag{2.66}$$

for all $\eta \in \tilde{U}$. Defining

$$\mathcal{L}(\eta) = \mathcal{L} \circ (\mathbb{F}\mathcal{L})^{-1}(\eta),$$

$$\dot{q}_i(\eta) = \dot{q}_i \circ (\mathbb{F}\mathcal{L})^{-1}(\eta),$$

equation (2.66) becomes:

$$\mathcal{H}(\eta) = \sum_{i=1}^{n} \left(\frac{\partial \mathcal{L}}{\partial \dot{q}_i} \circ (\mathbb{F}\mathcal{L})^{-1}(\eta) \right) \dot{q}_i(\eta) - \mathcal{L}(\eta). \tag{2.67}$$

Finally, observe that since $(q_1, \dots, q_m, \dot{q}_1, \dots, \dot{q}_m)$ and $(q_1, \dots, q_m, p_1, \dots, p_m)$ are dual coordinate systems by construction, setting $\xi = (\mathbb{F}\mathcal{L})^{-1}(\eta)$, equation (2.6) in Theorem 2.2.1 implies that

$$p_i(\eta) = p_i \circ (\mathbb{F}\mathcal{L})(\xi) = \frac{\partial \mathcal{L}}{\partial \dot{q}_i}(\xi) = \frac{\partial \mathcal{L}}{\partial \dot{q}_i} \circ (\mathbb{F}\mathcal{L})^{-1}(\eta)$$

so that (2.67) can be rewritten as

$$\mathcal{H}(\eta) = \sum_{i=1}^{n} p_i(\eta) \cdot \dot{q}_i(\eta) - \mathcal{L}(\eta),$$

which is precisely the Hamiltonian from (1.100). Consequently, we can make the following definition.

Definition 2.4.1 (Hamiltonian (on the cotangent bundle)). Let Q be the configuration manifold for a mechanical system with corresponding hyperregular Lagrangian $\mathcal{L} \colon TQ \to \mathbb{R}$. We define the corresponding *Hamiltonian* $\mathcal{H}$ to be

$$\begin{aligned} &\mathcal{H} \colon T^*Q \to \mathbb{R} \\ &\mathcal{H}(\eta) = \eta \left((\mathbb{F}\mathcal{L})^{-1}(\eta) \right) - \mathcal{L} \left((\mathbb{F}\mathcal{L})^{-1}(\eta) \right) \end{aligned} \tag{2.68}$$

for all $\eta \in T^*(Q)$.

Remark 2.4.1. We have shown, given a mechanical system with configuration manifold Q and corresponding hyperregular Lagrangian $\mathcal{L}\colon TQ \to \mathbb{R}$, how to construct the corresponding Hamiltonian on the (symplectic!) cotangent bundle T^*Q. Together with equation (2.1) in our introductory remarks on symplectic geometry in Section 2.1, this proves the following result.

2.4.1.2 Hamiltonian mechanics (symplectic geometric formulation)

Theorem 2.4.1 (Hamilton mechanics (symplectic geometric formulation)). *Let Q be the configuration manifold of a conservative mechanical system with corresponding Lagrangian $\mathcal{L}\colon TQ \to \mathbb{R}$. If $\mathcal{L}$ is hyperregular, define the corresponding Hamiltonian $\mathcal{H}$ according to*

$$\mathcal{H}\colon T^*Q \to \mathbb{R}$$
$$\mathcal{H}(\eta) = \eta\left((\mathbb{F}\mathcal{L})^{-1}(\eta)\right) - \mathcal{L}\left((\mathbb{F}\mathcal{L})^{-1}(\eta)\right)$$

for all $\eta \in T^*Q$, and give T^*Q a symplectic structure by equipping it with the canonical 2-form Ω. Then the time evolution $\eta(t) \in T^*Q$ of the system corresponding to an initial condition $\eta(0) = \eta_0 \in T^*Q$ is the integral curve of the Hamiltonian vector field $X_{\mathcal{H}}$ passing through η_0.

Definition 2.4.2 (Phase space). When the cotangent bundle T^*Q is thought of in the context of Hamiltonian mechanics as in Theorem 2.4.1, it is often referred to as the *phase space* of the system.

2.4.2 Abstract Hamiltonian mechanics

In Section 2.4.1 we showed, given the configuration manifold Q of a mechancial system with corresponding Lagrangian $\mathcal{L}\colon TQ \to \mathbb{R}$, how to derive the corresponding Hamiltonian $\mathcal{H}\colon T^*Q \to \mathbb{R}$ on the cotangent bundle T^*Q so as to obtain an equivalent representation of the system. However, in the same way that one *postulates* the existence of a Lagrangian on TQ, one could also simply *postulate* the existence of a Hamiltonian on T^*Q, and thereby immediately obtain the corresponding dynamics through the symplectic structure (i.e., through the Hamiltonian vector field $X_{\mathcal{H}}$).

One interesting consequence of this approach is the following. In the Lagrangian case, the system trajectories were recovered as the base integral curves of second order vector fields defined on the cotangent bundle TQ, which were themselves obtained by pulling back the canonical 2-form Ω from the cotangent bundle T^*Q to the tangent bundle TQ using the fiber derivative of the Lagrangian, $\mathbb{F}\mathcal{L}$. In short, the ability to induce Lagrangian dynamics on the tangent bundle TQ depended upon the existence of a mapping $P\colon TQ \to T^*Q$ that could be used to "translate" (at least an approximation of) the canonical symplectic structure on T^*Q to TQ by means of the pullback $P^*\Omega$. In contrast, since any symplectic manifold (M, ω) tautologically already has a symplectic structure, one immediately obtains *uniquely defined* dynamics corresponding to any function $H \in C^\infty(M)$ by forming the associated Hamiltonian vector field X_H. In this way, one can obtain a straightforward generalization of the notion of Hamiltonian mechanics to manifolds that are obviously not associated to any physical mechanical systems; all that is required is the ability to associate Hamiltonian vector fields to functions, i.e., a Poisson structure.

For example, let Σ be a compact orientable surface. By compactness, Σ is clearly *not* the phase space of any physical system, but it *does* have a symplectic structure provided by the volume form. Consequently, given any smooth function $H \in C^\infty(\Sigma)$, one could investigate the corresponding Hamiltonian dynamics on Σ defined by the Hamiltonian vector field X_H. In this way, one obtains an (abstract) nontrivial generalization of the notion of Hamiltonian dynamics to the entire class of Poisson manifolds(!).

2.5 Symmetry

In the previous sections, we showed how to reformulate Hamiltonian and Lagrangian mechanics entirely in terms of geometric objects. This is aesthetically very pleasing, as it highlights the geometric underpinnings of mechanics while simultaneously providing a coordinate free representation, which can be a great aid in analysis.

We would argue that these results are desirable as ends in themselves due to their purely aesthetic appeal; however, the real power of the geometric formulation is that, by recasting classical mechanics in terms of geometric objects, we can now also apply powerful analytical techniques from differential geometry and topology.

For example, the concept of symmetry is central in both mathematics and physics, both for its purely aesthetic appeal, and as an invaluable aid in analysis. One of the benefits of the geometric formulation of mechanics is that symmetry can be expressed very concretely through the use of geometric objects: namely, one thinks of symmetry in terms of Lie group actions on manifolds that preserve some geometric structure. In this section, we explore a few ways in which one can exploit the existence of symmetries to aid in the analysis of mechanial systems, as in [1–3, 10].

2.5.1 Noether's Theorem

We have already seen (e.g., in Section 1.1.6.2) that the existence of quantities that are conserved along the trajectories of a mechanical system can greatly aid in their analysis (indeed, the entire formulation of Lagrangian mechanics, hence also Hamiltonian mechanics, crucially depended upon the conservation of energy). Consequently, one of the most important results in mechanics is *Noether's Theorem*, which (loosely speaking) states that *the existence of a symmetry implies the existence of a corresponding conserved quantity.*

2.5.1.1 Fundamental vector fields

Definition 2.5.1 (Fundamental vector field). Given a manifold M and Lie group G with a smooth left-action $\mu\colon G \times M \to M$, fix $\xi \in \mathfrak{g} = \mathrm{Lie}(G)$ and define the mapping

$$\begin{aligned} \varphi_\xi &\colon \mathbb{R} \times M \to M \\ \varphi_\xi(t, m) &= \mu(\exp(t\xi), m) \end{aligned} \tag{2.69}$$

for all $t \in \mathbb{R}$ and $m \in M$. Then the mapping

$$\begin{aligned} \xi_M &\colon M \to TM \\ \xi_M(m) &= \frac{d}{dt}\left[\varphi_\xi(t, m)\right]\Big|_{t=0} \end{aligned} \tag{2.70}$$

defines a smooth vector field on M, called the *fundamental vector field* or the *infinitesimal generator* corresponding to ξ.

Observe that the infinitesimal generator ξ_M is so called because it encodes the "infinitesimal" action of the group G in the direction $\xi(e)$ on M at the point m. More precisely, for each $m \in M$, define the *orbit map*

$$\begin{aligned} \mu_m &\colon G \to M \\ \mu_m(g) &= \mu(g, m). \end{aligned}$$

Then equation (2.69) can equivalently be written as:

$$\varphi_\xi(t, m) = \mu_m(\exp(t\xi)). \tag{2.71}$$

For fixed $m \in M$, differentiating the right-hand side of (2.71) at $t = 0$ shows that

$$\xi_M(m) = d(\mu_m)_e(\xi(e)), \tag{2.72}$$

which is often taken as an alternative definition of the fundamental vector field (this is the case, for example, in [8]).

We now derive a few basic properties of fundamental vector fields that we will need in the sequel.

Definition 2.5.2 (Left-multiplication map). Let M be a manifold and G a Lie group with a smooth left-action $\mu\colon G \times M \to M$. Given any fixed $g \in G$, we define the *left-multiplication map* L_g to be the diffeomorphism

$$L_g\colon M \to M$$
$$L_g(m) = \mu(g, m).$$

Lemma 2.5.1. *Let G be a Lie group with a left-action $\mu\colon G \times M \to M$ on a smooth manifold M. Given any $\xi, \zeta \in \mathfrak{g}$ and any $m \in M$, the following hold:*

1. $(\mathrm{Ad}_g\,\xi)_M(m) = d(L_g)_{L_{g^{-1}}(m)}(\xi_M(L_{g^{-1}}(m)))$.

2. $[\xi_M, \zeta_M] = -[\xi, \zeta]_M$.

Proof. Claim 1 is just a direct computation using the properties of the adjoint representation. Using the definition (2.70), we have

$$\begin{aligned}(\mathrm{Ad}_g\,\xi)_M(m) &= \frac{d}{dt}\left[\mu\left(\exp(t\,\mathrm{Ad}_g\,\xi), m\right)\right]\Big|_{t=0} \\ &= \frac{d}{dt}\left[\mu\left(g\exp(t\xi)g^{-1}, m\right)\right]\Big|_{t=0} \\ &= \frac{d}{dt}\left[L_g \circ \mu\left(\exp(t\xi), L_{g^{-1}}(m)\right)\right]\Big|_{t=0} \\ &= d(L_g)_{L_{g^{-1}}(m)}(\xi_M(L_{g^{-1}}(m))).\end{aligned}$$

This is Claim 1. For Claim 2, let $g = \exp(t\zeta)$ in Claim 1:

$$(\mathrm{Ad}_{\exp(t\zeta)}\,\xi)_M(m) = d(L_{\exp(t\zeta)})_{L_{\exp(-t\zeta)}(m)}(\xi_M(L_{\exp(-t\zeta)}(m))). \tag{2.73}$$

Now $L_{\exp(-t\zeta)}$ is the flow associated with $-\zeta_M$ (by (2.69) and (2.70)), so the derivative of the right-hand side of (2.73) with respect to t, evaluated at $t = 0$, can be expressed in terms of the Lie derivative of ξ_M with respect to $-\zeta_M$:

$$\begin{aligned} \frac{d}{dt}\left[d(L_{\exp(t\zeta)})_{L_{\exp(-t\zeta)}(m)}(\xi_M(L_{\exp(-t\zeta)}(m)))\right]\Big|_{t=0} &= L_{-\zeta_M}\xi_M(m) \\ &= -[\zeta_M, \xi_M](m) \\ &= [\xi_M, \zeta_M](m). \end{aligned} \tag{2.74}$$

On the other hand, taking the derivative of the left-hand side of (2.73) with respect to t and evaluating at $t = 0$ gives:

$$\begin{aligned} \frac{d}{dt}\left[(\mathrm{Ad}_{\exp(t\zeta)}\,\xi)_M(m)\right] &= (\mathrm{ad}_\zeta\,\xi)_M(m) \\ &= [\zeta, \xi]_M(m) \\ &= -[\xi, \zeta]_M(m). \end{aligned} \tag{2.75}$$

Equations (2.73), (2.74), and (2.75) then imply Claim 2. □

Corollary 2.5.2. *Let M be a manifold and G a Lie group with a left-action $\mu\colon G \times M \to M$. Then the assignemnt*

$$\lambda\colon \mathfrak{g} \to \mathfrak{X}(M)$$
$$\lambda(\xi) = \xi_M$$

is a Lie algebra antihomomorphism.

The notion of *equivariance* will also be of considerable interest for us in the future.

Definition 2.5.3 (Equivariant mapping). Let M and N be manifolds and G a Lie group with left-actions $\mu_M\colon G \times M \to M$ and $\mu_N\colon G \times N \to N$. A smooth map $\varphi\colon M \to N$ is called *G-equivariant* if, for all $g \in G$, we have

$$L_g \circ \varphi = \varphi \circ L_g.$$

Equivalently, φ is G-equivariant if the following diagram commutes:

$$\begin{array}{ccc} M & \xrightarrow{\ \varphi\ } & N \\ {\scriptstyle L_g}\downarrow & & \downarrow{\scriptstyle L_g} \\ M & \xrightarrow[\ \varphi\ ]{} & N \end{array}$$

Theorem 2.5.3. *Let M and N be smooth manifolds, G a Lie group with smooth left-actions $\mu_M\colon G \times M \to M$ and $\mu_N\colon G \times N \to N$, and $\varphi\colon M \to N$ a smooth G-equivariant map. Then for any $\xi \in \mathfrak{g}$, the fundamental vector fields ξ_M and ξ_N are φ-related:*

$$d\varphi \circ \xi_M = \xi_N \circ \varphi.$$

Equivalently, the following diagram commutes:

$$\begin{array}{ccc} TM & \xrightarrow{\ d\varphi\ } & TN \\ {\scriptstyle \xi_M}\uparrow & & \uparrow{\scriptstyle \xi_N} \\ M & \xrightarrow[\ \varphi\ ]{} & N \end{array}$$

Proof. The G-equivariane of φ shows that

$$\varphi \circ L_{\exp(t\xi)} = L_{\exp(t\xi)} \circ \varphi$$

for all $\xi \in \mathfrak{g}$. Diffentiation with respect to t at $t = 0$ shows that

$$d\varphi \circ \left(\frac{d}{dt} L_{\exp(t\xi)} \bigg|_{t=0} \right) = \left(\frac{d}{dt} L_{\exp(t\xi)} \bigg|_{t=0} \right) \circ \varphi,$$

or

$$d\varphi \circ \xi_M = \xi_N \circ \varphi.$$

□

2.5.1.2 Momentum maps

Definition 2.5.4 (Momentum map). Let $(P, \{,\})$ be a Poisson manifold and G a Lie group with Poisson action $\mu\colon G \times P \to P$. A mapping $J\colon P \to \Omega^1_{l\,inv}(G)$ from P to the vector space of left-invariant forms on G is called a *momentum map* if the function defined by

$$\begin{aligned} J_\xi \colon P &\to \mathbb{R} \\ J_\xi(p) =& \langle J(p), \xi \rangle \end{aligned} \tag{2.76}$$

satisfies

$$X_{J_\xi} = \xi_P \tag{2.77}$$

for all $\xi \in \mathfrak{g}$. In words, we require that the Hamiltonian vector field on P induced by the pairing of the left-invariant form $J(p)$ with the left-invariant vector field ξ agree with the fundamental vector field ξ_P on P.

Definition 2.5.5 (The adjoint and coadjoint action). Recall that a Lie group G acts on itself on the left by conjugation:

$$\mu\colon G \times G \to G$$

$$\mu(g, x) = gxg^{-1}.$$

Denoting by κ_g the Lie group automorphism $\kappa_g = \mu(g, \cdot)$, this action induces a group homomorphism

$$\rho\colon G \to \mathrm{Inn}(G)$$

$$\rho(g) = \kappa_g.$$

Since e is a fixed point of the self-map κ_g for all $g \in G$, the differential

$$d(\kappa_g)_e\colon T_e(G) \to T_e(G)$$

is an automorphism of the tangent space to the identity for all g. Consequently, identifying $T_e(G)$ with $\mathfrak{g} = \mathrm{Lie}(G)$ in the usual way, we obtain the familiar *adjoint representation* of G on its Lie algebra $\mathfrak{g}$:

$$\mathrm{Ad}\colon G \to \mathrm{Aut}(\mathfrak{g})$$

$$\mathrm{Ad}(g) = d(\kappa_g)_e$$

(for notational convenience, we shall often write Ad_g for $\mathrm{Ad}(g)$). This action induces a corresponding representation Ad^* on the dual space $\mathfrak{g}^* = \Omega^1_{l\,inv}(G)$, called the *coadjoint action*, defined by

$$\mathrm{Ad}^*\colon G \to \mathrm{Aut}(\mathfrak{g}^*)$$

$$\mathrm{Ad}^*(g) = (\mathrm{Ad}(g^{-1}))^*$$

(here again, we shall write Ad_g^* for $\mathrm{Ad}^*(g)$).

Definition 2.5.6 (Ad^*-equivariant momentum maps). Given a Lie group G with Poisson left-action $\mu\colon G \times P \to P$ on a Poisson manifold $(P, \{,\})$, a momentum map $J\colon G \to \Omega^1_{l\,inv}(G)$ is called Ad^*-*equivariant* if, for all $g \in G$, the following diagram of maps commutes:

$$\begin{array}{ccc} P & \xrightarrow{\quad L_g \quad} & P \\ \Big\downarrow{\scriptstyle J} & & \Big\downarrow{\scriptstyle J} \\ \Omega^1_{l\,inv}(G) & \xrightarrow{\quad \mathrm{Ad}_g^* \quad} & \Omega^1_{l\,inv}(G) \end{array} \tag{2.78}$$

Remark 2.5.1. Note that the diagram (2.78) shows that an Ad^*-equivariant momentum map $J\colon P \to \Omega^1_{l\,inv}(G)$ is G-equivariant when $\Omega^1_{l\,inv}(G)$ is equipped with the coadjoint action (hence the nomenclature).

2.5.1.3 Noether's Theorem

One very elementary yet quite pleasing result is that symmetric Hamiltonians have symmetric flows.

Theorem 2.5.4. *Let $(P, \{,\})$ be a Poisson manifold, G a Lie group with Poisson left-action $\mu\colon G \times P \to P$, and $H \in C^\infty(P)$ a G-invariant function; that is,*

$$H \circ L_g = H$$

for all $g \in G$. Then the following hold:

1. *The Hamiltonian vector field X_H is L_g-related to itself for all $g \in G$:*

$$d(L_g) \circ X_H = X_H \circ L_g. \tag{2.79}$$

2. *The flow* $\varphi_t \colon P \to P$ *on* P *induced by* X_H *is* G*-equivariant for all* $g \in G$*:*

$$\varphi_t \circ L_g = L_g \circ \varphi_t.$$

Proof. Let $f \in C^\infty(P)$. Since μ is a Poisson action,

$$\{f, H\} \circ L_g = \{f \circ L_g, H \circ L_g\} \tag{2.80}$$

for any $g \in G$. Now given any $p \in P$

$$\begin{aligned}\{f \circ L_g, H \circ L_g\}(p) &= \{f \circ L_g, H\}(p) \\ &= (X_H)_p(f \circ L_g) \\ &= (df_{g\cdot p} \circ d(L_g)_p)(X_H(p)) \\ &= df_{g\cdot p}\left(d(L_g)_p \circ X_H(p)\right)\end{aligned} \tag{2.81}$$

by the G-invariance of H. On the other hand,

$$\{f, H\} \circ L_g(p) = (X_H)_{g\cdot p}(f) = df_{g\cdot p}(X_H \circ L_g(p)). \tag{2.82}$$

Equations (2.80), (2.81), and (2.82) show that

$$df_{g\cdot p}(d(L_g)_p \circ X_H(p)) = df_{g\cdot p}(X_H \circ L_g(p))$$

holds identically for all $p \in P$, $g \in G$, and $f \in C^\infty(P)$. Allowing f to vary over any complete set of coordinate functions on a neighborhood of every $p \in P$ proves (2.79). Claim 2 is then immediate from Claim 1. $\square$

We now turn to the main result of this section.

Theorem 2.5.5 (Noether's Theorem). *Let* $(P, \{,\})$ *be a Poisson manifold,* G *a Lie group with a Poisson left-action on* P*, and* $J\colon P \to \Omega^1_{l\,inv}(P)$ *a momentum map. If* $H \in C^\infty(P)$ *is* G*-invariant, then* J *is constant along the integral curves of the Hamiltonian vector field* X_H*.*

Proof. If H is G-invariant, then $H(\mu_m(g)) = H(m)$ for every $g \in G$ and $m \in M$. Differentiating this identity with respect to g at $g = e$ shows that

$$\begin{aligned}
0 &= dH_m(d(\mu_m)_e(\xi(e))) \\
&= dH_m(\xi_P(m)) \\
&= dH_m(X_{J_\xi}(m)) \\
&= (X_{J_\xi})_m(H) \\
&= \{H, J_\xi\}(m) \\
&= \{J_\xi, H\}(m)
\end{aligned}$$

for every $\xi \in \mathfrak{g}$ (here we have used the fact that J is a momentum map in passing from line 2 to 3, the definition of the Hamiltonian vector field X_{J_ξ} (B.40) in passing from 4 to 5, and the antisymmetry of the Poisson bracket in passing from line 5 to 6). The final line of the above sequence of equalities together with Theorem B.3.1 shows that each function J_ξ is constant along the integral curves of X_H for every $\xi \in \mathfrak{g}$, and therefore so is J. □

2.5.1.4 Construction of momentum maps

Of course, this is not particularly helpful unless we can *identify* a momentum map that will produce a conserved quantity. Fortunately, it turns out that it is not difficult to construct such maps quite explicitly on an important class of symplectic manifolds.

Theorem 2.5.6 (Construction of momentum maps). *Let* (M, ω) *be a symplectic manifold and* G *a Lie group with a symplectic left-action* $\mu\colon G \times M \to M$*. Suppose further that the symplectic form* ω *is exact, with* $\omega = -d\theta$ *and* θ G*-invariant. Then the mapping defined by*

$$J\colon M \to \Omega^1_{l\,inv}(G)$$

$$J(m)(\xi) = \theta_m(\xi_M(m))$$

for all $m \in M$ *and* $\xi \in \mathfrak{g}$ *is an* Ad^**-equivariant momentum map for this action.*

Proof. Fix some $\xi \in \mathfrak{g}$, and define

$$\varphi_t\colon M \to M$$

$$\varphi_t(m) = L_{\exp(t\xi)}(m)$$

for all $t \in \mathbb{R}$ and all $m \in M$. Then the G-invariance of θ implies that $\varphi_t^*\theta = \theta$ for all $t \in \mathbb{R}$. Thus, we have

$$0 = \lim_{t \to 0} \left[\frac{\varphi_t^*\theta - \theta}{t} \right] \Bigg|_{t=0} = L_{\xi_M}\theta$$

by definition of φ_t and by virtue of (2.70). Applying Cartan's formula then produces:

$$\begin{aligned} 0 &= i_{\xi_M} d\theta + d(i_{\xi_M}\theta) \\ &= -i_{\xi_M}\omega + dJ_\xi, \\ &= -(\xi_M)^\flat + dJ_\xi, \end{aligned}$$

where $\omega = -d\theta$ is the symplectic form on M. Applying the sharp map $\sharp$ to the final line of the above equation shows that

$$\xi_M = (dJ_\xi)^\sharp$$

which is equivalent to the statement

$$\xi_M = X_{J_\xi}$$

since the Poisson bracket $\{\cdot, \cdot\}$ on M is induced from ω (cf. Theorem B.3.6). This shows that J is a momentum map for this action.

To prove Ad*-equivariance, we must show that the diagram (2.78) is commutative, i.e., that

$$\begin{aligned} J(L_g(m))(\xi) &= (\mathrm{Ad}_g^* \circ J(m))(\xi) \\ &= J(m)(\mathrm{Ad}_{g^{-1}} \xi) \end{aligned}$$

for all $m \in M$ and $\xi \in \mathfrak{g}$. By definition, this is equivalent to the statement that

$$\theta_{gm}(\xi_M(gm)) = \theta_m\left((\mathrm{Ad}_{g^{-1}} \xi)_M(m)\right); \tag{2.83}$$

thus, proving (2.83) is out immediate goal.

Now, the G-invariance of θ shows that

$$(L_g^*\theta)_m(X_m) = \theta_{gm}(d(L_g)_m X_m) = \theta_m(X_m)$$

for all $m \in M$, $X_m \in T_m(M)$, and $g \in G$. Taking $X_m = \left((\mathrm{Ad}_{g^{-1}}\,\xi)_M\right)(m)$ in the above equation produces:

$$\theta_{gm}\left(d(L_g)_m\left((\mathrm{Ad}_{g^{-1}}\,\xi)_M\right)(m)\right) = \theta_m\left(\left((\mathrm{Ad}_{g^{-1}}\,\xi)_M\right)(m)\right). \tag{2.84}$$

Now since

$$d(L_g)_m\left((\mathrm{Ad}_{g^{-1}}\,\xi)_M\right)(m) = d(L_g)_m d(L_{g^{-1}})_{gm}(\xi_M(gm)) = \xi_M(gm)$$

by Claim 1 of Lemma 2.5.1, substitution in the left-hand side of (2.84) produces

$$\theta_{gm}(\xi_M(gm)) = \theta_m\left(\left((\mathrm{Ad}_{g^{-1}}\,\xi)_M\right)(m)\right)$$

which is the desired equality (2.83). □

Using Theorem 2.5.6, we can obtain a particularly useful result in the context of Hamiltonian mechanical systems whose configuration spaces have symmetry. Specifically, let G be a Lie group and $\mu\colon G\times Q \to Q$ a left-action of G on Q by diffeomorphisms. By Theorem B.2.10, every such diffeomorphism $g\colon Q \to Q$ induces a corresponding *symplectomorphism* $\tilde{g}\colon T^*Q \to T^*Q$, and we thereby obtain a lifted *symplectic* action $\tilde{\mu}\colon G\times T^*Q \to T^*Q$. Furthermore, Theorem B.2.10 also states that the action of G on T^*Q preserves the canonical 1-form Θ on T^*Q. Thus, Theorem 2.5.6 implies the following result.

Theorem 2.5.7 (Construction of momentum maps on the cotangent bundle). *Let Q be a smooth manifold and G a Lie group with left-action $\mu\colon G\times Q \to Q$.*

Denote the canonical 1*-form on* T^*Q *by* Θ*, and the canonical symplectic* 2*-form by* Ω*. Define*

$$J\colon T^*Q \to \Omega^1_{l\,inv}(G)$$
$$J(\alpha)(\xi) = \langle \Theta_\alpha,\, \xi_{T^*Q} \rangle$$

for $\alpha \in T^*Q$ *and* $\xi \in \mathfrak{g}$*. Then* J *is an* Ad^**-equivariant momentum map corresponding to the lifted symplectic action* $\tilde{\mu}\colon G \times T^*Q \to T^*Q$ *of* G *on the cotangent bundle* (T^*Q, Ω)*.*

2.5.2 Symplectic reduction

In the previous section, we demonstrated that one application of symmetry is the elucidation of constants of motion. A closely related idea is the notion of *reduction*, which we motivate with a few preliminary remarks.

If one can identify a conserved quantity of motion for a system, that conservation law in effect represents an additional *constraint* on the possible set of system trajectories. For example, when examining the bead on a wire (Section 1.1.6.2), we observed that the total mechanical energy of the system was a constant of motion, and then used this in the derivation of equation (1.34), which we subsequently regarded as a *constraint* on the allowable system trajectories (namely, any system trajectory must be one for which equation (1.34) is satisfied). As we saw before (e.g., in Section 1.2.2.1), the imposition of constraints on a mechanical system cuts out a subset of the original configuration space on which those constraints are satisfied. Thus (by virtue of Noether's Theorem), we can regard the identification of a symmetry of a system as a kind of constraint, and might therefore reasonably expect that the imposition of some kind of symmetry-based constraint would result in the identification of a strictly lower-dimensional configuration space for the system that still captures the system's behavior.

Now, given a generic symmetry of (i.e., Lie group action on) a manifold satisfying suitably "nice" conditions, one method of generating a new, lower-dimensional manifold from this group action is familiar: namely, taking the quotient of the original manifold by the group action. Since the symmetry constraints that we wish to impose are precisely of this type (i.e., derived from a Lie group action on a manifold), we might expect that our notion of reduction (i.e., with respect to dynamics) might be linked with the quotient operation on manifolds. As it turns out, this is indeed the case, and leads us to the notion of *symplectic reduction*, the subject of the current section.

2.5.2.1 Preliminaries

As was alluded to in the preliminary remarks, not all quotient operations induced by group actions produce well-defined manifolds; thus, we will customarily assume that the Lie group actions we are investigating satisfy certain "niceness" conditions, which we outline below.

Let (M, ω) be a symplectic manifold, G a Lie group with symplectic left-action $\mu\colon G \times M \to M$ on M, and $J\colon M \to \Omega^1_{l\,inv}(G)$ an Ad^*-equivariant momentum map for this action. For $\eta \in \Omega^1_{l\,inv}(G)$, denote by G_η the isotropy group

$$G_\eta = \{g \in G \mid \mathrm{Ad}^*(g)(\eta) = \eta\}.$$

This is a closed (abstract) subgroup of G, and therefore also a Lie subgroup (cf. [14]). Furthermore, since J is Ad^*-equivariant, then for $m \in J^{-1}(\eta)$ and $g \in G_\eta$, we have

$$J(g \cdot m) = \mathrm{Ad}^*(g)J(m) = \mathrm{Ad}^*(g)(\eta) = \eta.$$

Thus, the left-action of G on M restricts to a left-action of G_η on $J^{-1}(\eta)$, and therefore

$$M_\eta = J^{-1}(\eta)/G_\eta$$

is well-defined as a *topological* space (we have not yet shown it to be a manifold).

Now if η is a regular value of J, then standard techniques from differential topology (as in [6] and [14]) show that $J^{-1}(\eta)$ is a smooth submanifold of M with $\dim(J^{-1}(\eta)) = \dim M - \dim G$. Furthermore, if the isotropy group G_η acts freely and properly on $J^{-1}(\eta)$, then the quotient M_η will likewise be a manifold, and the canonical quotient projection map $\pi\colon J^{-1}(\eta) \to M_\eta$ a submersion. Thus, we will frequently require the following conditions to hold:

1. $\eta \in \Omega^1_{l\,inv}(G)$ is a regular value of J.

2. G_η acts freely and properly on $J^{-1}(\eta)$.

2.5.2.2 Symplectic reduction

The main result of this section is the following theorem.

Theorem 2.5.8 (Symplectic reduction). *Let (M, ω) be a symplectic manifold, G a Lie group with symplectic left-action $\mu\colon G \times M \to M$, and $J\colon M \to \Omega^1_{l\,inv}(G)$ an* Ad**-equivariant momentum map. Let $\eta \in \Omega^1_{l\,inv}(G)$ be a regular value of J, and suppose that the isotropy group G_η acts freely and properly on $J^{-1}(\eta)$. Then the quotient manifold $M_\eta = J^{-1}(\eta)/G_\eta$ has a unique symplectic form ω_η satisfying*

$$\pi_\eta^* \omega_\eta = i_\eta^* \omega, \tag{2.85}$$

where $\pi_\eta\colon J^{-1}(\eta) \to M_\eta$ is the canonical projection map and $i_\eta\colon J^{-1}(\eta) \hookrightarrow M$ the inclusion map.

Definition 2.5.7 (Reduced phase space). The quotient symplectic manifold (M_η, ω_η) defined in Theorem 2.5.8 is called the *reduced phase space*.

The proof of Theorem 2.5.8 depends upon the following lemma.

Lemma 2.5.9. *Assume the hypotheses of Theorem* 2.5.8*. Fix $m \in J^{-1}(\eta)$ and let $\bar{\mu}\colon G_\eta \times M \to M$ denote the restriction of the action μ to the subgroup G_η. Then the following hold*

1. $\operatorname{image}(d(\bar{\mu}_m)_e) = \operatorname{image}(d(\mu_m)_e) \cap T_m(J^{-1}(\eta))$.

2. $T_m(J^{-1}(\eta))$ *is the* ω*-orthogonal complement of* $\operatorname{image}(d(\mu_m)_e)$.

Proof. By (2.72), $\operatorname{image}(d(\mu_m)_e) = \{\xi_M(m) \mid \xi \in \mathfrak{g}\}$. Now the Ad^*-equivariance of J is equivalent to the G-equivariance of J when $\mathfrak{g}^*$ is equipped with the coadjoint action, and in that case, Theorem 2.5.3 shows that $dJ_m(\xi_M(m)) = \xi_{\mathfrak{g}^*}(\eta)$. Thus, $\xi_M(m) \in T_m(J^{-1}(\eta)) = \ker(dJ_m)$ if and only if $\xi_{\mathfrak{g}^*}(\eta) = 0$. Since $\mathfrak{g}^*$ is equipped with the coadjoint action, $\xi_{\mathfrak{g}^*}(\eta) = 0$ is again equivalent to the statement that

$$\left(\mathrm{Ad}_{\exp(-t\xi)}\right)^* (\eta) = \eta,$$

so that $\exp(-t\xi) \in G_\eta$, or $\xi \in \mathfrak{g}_\eta = \mathrm{Lie}(G_\eta)$, in which case $\xi_M(m) \in \operatorname{image}(d(\bar{\mu}_m)_e)$; this proves Claim 1.

For Claim 2, if $\xi \in \mathfrak{g}$ and $v \in T_m(M)$, then

$$\omega(\xi_M(m), v) = \omega(X_{J_\xi}, v) = \omega((dJ_\xi)^\sharp, v) = dJ_\xi(v)$$

since J is a momentum map, which shows that $v \in T_m(J^{-1}(\eta)) = \ker(dJ_m)$ if and only if $\omega(\xi_M(m), v) = 0$ for all $\xi \in \mathfrak{g}$. This is precisely the statement that v is in the ω-orthogonal complement of $\operatorname{image}(d(\mu_m)_e) = \{\xi_M(m) \mid \xi \in \mathfrak{g}\}$. $\square$

Proof of Theorem 2.5.8. For $v \in T_m(J^{-1}(\eta))$, let $[v] = d(\pi_\eta)_m(v)$ denote the corresponding tangent vector in $T_{\pi_\eta(m)}(M_\eta) = T_m(J^{-1}(\eta))/\operatorname{image}(d(\bar{\mu}_m)_e)$. The statement $\pi_\eta^*\omega_\eta = i_\mu^*\omega$ is precisely that

$$\omega_\eta([v], [w]) = \omega(v, w) \quad \text{for all } v, w \in T_m(J^{-1}(\eta))), \tag{2.86}$$

so we must show that this gives a unique and well-defined symplectic form. To that end, let $k_1, k_2 \in \text{image}(d(\bar{\mu}_m)_e)$; by bilinearity,

$$\omega(v + k_1, w + k_2) = \omega(v, w) + \omega(v, k_2) + \omega(k_1, w) + \omega(k_1, k_2). \qquad (2.87)$$

Now $k_1, k_2 \in \text{image}(d(\bar{\mu}_m)_e) = \text{image}(d(\mu_m)_e) \cap T_m(J^{-1}(\eta))$, and $\text{image}(d(\mu_m)_e)$ and $T_m(J^{-1}(\eta))$ are ω-orthogonal complements of one another (by Lemma 2.5.9). Therefore, k_1 and k_2 produce zero when paired with *any* vectors in $T_m(J^{-1}(\eta))$, so equation (2.87) reduces to

$$\omega(v + k_1, w + k_2) = \omega(v, w),$$

which shows that ω is constant on the equivalence classes of $T_{\pi_\eta(m)}(M_\eta) = T_m(J^{-1}(\eta))/\,\text{image}(d(\bar{\mu}_m)_e)$. Since π_η (hence also $d(\pi_\eta)_m$) is surjective, equation (2.86) gives a well-defined 2-form ω_η on the quotient M_η. Now we must show that ω_η is *symplectic*.

To show that ω is closed, observe that

$$\pi^*(d\omega_\eta) = d(\pi_\eta^* \omega_\eta) = d(i_\eta^* \omega) = i_\eta^*(d\omega) = 0.$$

Since $d\pi_\eta$ is surjective, the pullback map π_η^* is injective, and therefore $d\omega_\eta = 0$, so that ω_η is closed. Now fix some $m \in J^{-1}(\eta)$ and some $[v] \in T_{\pi_\eta(m)}(M_\eta)$, and suppose that

$$\omega_\eta([v], [w]) = 0 \qquad \text{for all } [w] \in T_{\pi_\eta(m)}(M_\eta).$$

By (2.86), this is equivalent to

$$\omega(v, w) = 0 \quad \text{for all } w \in T_m(J^{-1}(\eta))),$$

so that $v \in \text{image}(d(\mu_m)_e)$ by Claim 2 of Lemma 2.5.9. Claim 1 of Lemma 2.5.9 then shows that $v \in \text{image}(d(\bar{\mu}_m)_e)$, and therefore $[v] = 0$. Thus, $\omega([v], [w]) = 0$ for all $[w] \in T_{\pi_\eta(m)}(M_\eta)$ implies that $[v] = 0$, so that ω_η is nondegenerate, hence symplectic. □

2.5.2.3 Hamiltonian dynamics on the reduced phase space

Thus far, we have shown how to obtain a symplectic quotient when a Lie group acts symplectically on a manifold. However, this is of little use in our investigation of mechanics unless there is likewise some relation between the *dynamics* on the two manifolds. We begin by showing how Hamiltonian dynamics on a symplectic manifold induce corresponding Hamiltonian dynamics on the reduced phase space.

Theorem 2.5.10 (Hamiltonian dynamics on the reduced phase space). *Assume the hypotheses of Theorem* 2.5.8*, let $H \in C^\infty(M)$ be G-invariant, and let $\varphi_t \colon M \to M$ denote the flow along the Hamiltonian vector field X_H. Then the following hold:*

1. *$J^{-1}(\eta)$ is invariant under the flow φ_t.*

2. *The function*

$$\begin{aligned} \bar{H} \colon M_\eta &\to \mathbb{R} \\ \bar{H}([m]) &= H(m) \end{aligned} \tag{2.88}$$

is well-defined on the quotient $M_\eta = J^{-1}(\eta)/G_\eta$ *and satisfies*

$$\bar{H} \circ \pi_\eta = H \circ i_\eta, \tag{2.89}$$

where $\pi_\eta \colon J^{-1}(\eta) \to M_\eta$ *is the canonical quotient projection map and* $i_\eta \colon J^{-1}(\eta) \hookrightarrow M$ *is the inclusion map.*

3. *The Hamiltonian vector fields* $X_H \in \mathfrak{X}(J^{-1}(\eta))$ *and* $X_{\bar{H}} \in \mathfrak{X}(M_\eta)$ *are* π_η*-related:*

$$X_{\bar{H}} \circ \pi_\eta = d\pi_\eta \circ X_H. \tag{2.90}$$

4. *Let* $\bar{\varphi}_t \colon M_\eta \to M_\eta$ *denote the flow on* M_η *corresponding to* $X_{\bar{H}}$*. Then* $\bar{\varphi}_t \circ \pi_\eta = \pi_\eta \circ \varphi_t$*.*

Proof. The proof is actually quite straightforward when done in stages:

1. By Noether's Theorem (Theorem 2.5.5), J is a conserved quantity for the flow φ_t along the Hamiltonian vector field X_H. In particular, this implies that the flow φ_t maps the submanifold $J^{-1}(\eta)$ diffeomorphically onto itself.

2. The G-invariance of H implies *a fortiori* the G_η-invariance of H; consequently,

$$H(m) = H(g \cdot m)$$

for all $m \in J^{-1}(\eta)$ and $g \in G_\eta$; that is, H is constant on the G_η-cosets of $J^{-1}(\eta)$, which shows that the function $\bar{H}$ defined in (2.88) is well-defined on the quotient M_η. Equation (2.89) follows immediately (and also shows that $\bar{H}$ is actually *smooth*).

3. Let $m \in J^{-1}(\eta)$, $v \in T_m(J^{-1}(\eta))$, and $\omega_J = i_\eta^*\omega$. Then by equation (2.85) of Theorem 2.5.8,

$$\begin{aligned}(\omega_J)_m(X_H(m), v) &= (\pi_\eta^*\omega_\eta)_m(X_H(m), v)\\ &= (\omega_\eta)_{[m]}(d\pi_\eta \circ X_H(m), d\pi_\eta(v)) \qquad (2.91)\\ &= (\omega_\eta)_{[m]}(d\pi_\eta \circ X_H(m), [v]).\end{aligned}$$

At the same time,

$$(\omega_J)_m(X_H(m), v) = (\omega_J)_m((dH_m)^\sharp, v) = dH_m(v) \qquad (2.92)$$

by Theorems B.1.2 and B.3.6, and

$$\begin{aligned}dH_m(v) &= d\bar{H}_{[m]} \circ d\pi_\eta(v)\\ &= d\bar{H}_{[m]}([v])\\ &= (\omega_\eta)_{[m]}((d\bar{H}_{[m]})^\sharp, [v])\\ &= (\omega_\eta)_{[m]}(X_{\bar{H}}([m]), [v])\end{aligned} \qquad (2.93)$$

by (2.89). Equations (2.91), (2.92), and (2.93) show that

$$(\omega_\eta)_{[m]}(d\pi_\eta \circ X_H(m) - X_{\bar{H}}([m]), [v]) = 0.$$

Since $[v]$ ranges over all of $T_{\pi_\eta(m)}(M_\eta)$ as v ranges over all of $T_m(J^{-1}(\eta))$ and ω_η is nondegenerate, this proves (2.90).

4. Finally, Claim 4 is immediate from (2.90).

□

Definition 2.5.8 (Reduced Hamiltonian). The function $\bar{H}$ defined on the reduced phase space in (2.89) is called the *reduced Hamiltonian*.

2.5.2.4 Reconstruction of Hamiltonian dynamics on the total space

So now, given a symplectic manifold equipped with Hamiltonian dynamics and a symplectic Lie group action, we have shown how to produce a reduced phase space equipped with induced Hamiltonian dynamics. But again, in order for this to be useful as an aid in the analysis of mechanical systems, we should require that it not only be possible to induce *some* Hamiltonian dynamics on the reduced phase space, but that these dynamics should *encode the behavior of the original system on the total space.* Equivalently, we ought to show that the induced dynamics from Theorem 2.5.10 can be used to *reconstruct* the original dynamics on the total space. In this subsection, we provide an explicit algorithm for reconstructing the dynamics on the total space given those on the reduced phase space.

Assume the hypotheses of Theorem 2.5.10. Fix some initial condition $p_0 \in J^{-1}(\eta)$, and let $\tau(t)$ denote the integral curve of the Hamiltonian vector field $X_H \in \mathfrak{X}(J^{-1}(\eta))$ satisfying $\tau(0) = p_0$. Then $\tau(t)$ represents a trajectory of the system on the total space $J^{-1}(\eta)$, and we wish to show how to determine this trajetory given only the data on the reduced phase space.

Now Theorems 2.5.8 and 2.5.10 give a reduced phase space (M_η, ω_η) and a corresponding reduced Hamiltonian $\bar{H} \in C^\infty(M_\eta)$; consequently, we can compute the integral curve $[\gamma(t)]$ of the Hamiltonian vector field $X_{\bar{H}} \in \mathfrak{X}(M_\eta)$ satisfying $\pi_\eta(p_0) = [p_0] = [\gamma(0)]$. Now, the important observation here is Claim

4 of Theorem 2.5.10: namely, that *the integral curve* $[\gamma(t)]$ *is the projection of the integral curve* $\tau(t)$ *on the total space* $J^{-1}(\eta)$, i.e.,

$$[\gamma(t)] = \pi_\eta(\tau(t)). \tag{2.94}$$

Thus, fixing any smooth curve $\gamma(t)$ on $J^{-1}(\eta)$ that satisfies $[\gamma(t)] = \pi_\eta(\gamma(t))$ and $\gamma(0) = p_0$, by (2.94) it must be the case that

$$\tau(t) = \mu(g(t), \gamma(t))$$

for some function $g\colon [t_0, t_1] \to G_\eta$, so reconstructing $\tau(t)$ reduces to determining the unknown function g.

Now since $\tau(t)$ is the integral curve of X_H, then

$$\begin{aligned} X_H(\tau(t)) &= \tau'(t) \\ &= \frac{d}{dt}\left[\mu(g(t), \gamma(t))\right] \\ &= d(L_{g(t)})_{\gamma(t)}(\gamma'(t)) + d(\mu_{\gamma(t)})_{g(t)}((g'(t)). \end{aligned} \tag{2.95}$$

Now we wish to simplify the right-hand side of (2.95) somewhat. Given any $g \in G$ and $m \in M$, let $\xi_g \in T_g(G)$, and let $\xi = d(L_{g^{-1}})_g(\xi_g)$ be its left-translation to the tangent space $T_e(G)$. Then for $m \in M$, we have

$$\begin{aligned} d(\mu_m)_g(\xi_g) &= d(\mu_m)_g \circ d(L_g)_e(\xi) \\ &= d(\mu_m \circ L_g)_e(\xi) \\ &= d(L_g \circ \mu_m)_e(\xi) \\ &= d(L_g)_m \circ d(\mu_m)_e(\xi) \\ &= d(L_g)_m(\xi_M(m)) \\ &= d(L_g)_m\left(\left[d(L_{g^{-1}})_g(\xi_g)\right]_M(m)\right). \end{aligned} \tag{2.96}$$

Substitution of (2.96) into (2.95) shows that

$$
\begin{aligned}
X_H(\tau(t)) &= d(L_{g(t)})_{\gamma(t)}(\gamma'(t)) + d(L_{g(t)})_{\gamma(t)}\left(\left[d(L_{g^{-1}(t)})_{g(t)}(g'(t))\right]_M(\gamma(t))\right) \\
&= d(L_{g(t)})_{\gamma(t)}\left(\gamma'(t) + \left[d(L_{g^{-1}(t)})_{g(t)}(g'(t))\right]_M(\gamma(t))\right).
\end{aligned}
\tag{2.97}
$$

Now observe that since $\tau(t) = L_{g(t)}(\gamma(t))$ by definition, the L_g-relatedness of X_H to itself (Theorem 2.5.4) applied to equation (2.97) shows that

$$
X_H(\gamma(t)) = \gamma'(t) + \left[d(L_{g^{-1}(t)})_{g(t)}(g'(t))\right]_M(\gamma(t)), \tag{2.98}
$$

which is a single equation relating the unknown curve $g(t)$ to the known curve $\gamma(t)$. Now the point of rewriting (2.95) as (2.98) is that $d(L_{g^{-1}(t)})_{g(t)}(g'(t)) \in \mathfrak{g}$ always, as it is the left-translation of $g'(t) \in T_g(G)$ to $T_e(G)$. Thus, we can solve (2.98) for $g(t)$ by finding the curve $\xi(t) \in \mathfrak{g}$ satisfying

$$
(\xi(t))_M(\gamma(t)) = X_H(\gamma(t)) - \gamma'(t)
$$

(which is purely algebraic), and then requiring that the tangent vector $g'(t)$ be the pushforward of $\xi(t)$ by left-multiplication by $g(t)$; that is, we solve

$$
g'(t) = d(L_{g(t)})_e \xi(t)
$$

for $g(t)$. The reconstructed orbit $\tau(t)$ is then given by:

$$
\tau(t) = L_{g(t)}(\gamma(t)).
$$

We summarize this algorithm below.

Orbit Reconstruction Algorithm

Let (M, ω) be a sympletic manifold, G a Lie group with symplectic left-action $\mu\colon G \times M \to M$, $H \in C^\infty(M)$ a G-invariant smooth function with corresponding momentum map $J\colon M \to \Omega^1_{l\,inv}(M)$, and $\eta \in \Omega^1_{l\,inv}(M)$. Suppose that the hypotheses of Theorems 2.5.8 and 2.5.10 are satisfied, so that we have a reduced phase space (M_η, ω_η) and corresponding reduced Hamiltonian $\bar{H} \in C^\infty(M_\eta)$. Then given any $m \in J^{-1}(\eta)$, the trajectory $\tau(t)$ of X_H on $J^{-1}(\eta)$ satisfying $\tau(0) = m$ can be found according to the following algorithm:

1. Set $[m] = \pi(m)$, where $\pi\colon J^{-1}(\eta) \to M_\eta$ is the canonical quotient projection map, and compute the unique integral curve $[\gamma(t)]$ of $X_{\bar{H}}$ on M_η satisfying $[\gamma(0)] = [m]$.

2. Let $\gamma(t)$ be *any* smooth curve on $J^{-1}(\eta)$ satisfying $\pi(\gamma(t)) = [\gamma(t)]$ and $\gamma(0) = m$.

3. Solve the algebraic equation

   $$(\xi(t))_M(\gamma(t)) = X_H(\gamma(t)) - \gamma'(t)$$

   for $\xi(t) \in \mathfrak{g}$.

4. Solve the ODE

   $$g'(t) = d(L_{g(t)})_e \xi(t)$$

   for $g(t)$.

5. The orbit $\tau(t)$ on $J^{-1}(\eta)$ is then determined by

   $$\tau(t) = L_{g(t)}(\gamma(t)).$$

Remark 2.5.2. The fact that we can write down an explicit algorithm for determining an orbit on the total space $J^{-1}(\eta)$ solely in terms of the data on the reduced phase space means that the reduced phase space together with the reduced Hamiltonian $\bar{H}$ do, indeed, capture the dynamics of the total system, as we had hoped would be the case. This is illustrative of the analytical power that the geometric formulation of mechanics provides by appealing to the theory of manifolds and Lie group actions, but is by no means exhaustive. Again, the reader is referred to [1, 3, 10] for additional information and examples.

Appendices

Appendix A

The Calculus of Variations

The *calculus of variations* is the branch of mathematics concerned with extremizing (minimizing or maximizing) the value of *variational integrals*: functionals $\mathcal{F}\colon \mathcal{C} \to \mathbb{R}$ of the form

$$\mathcal{F}(u) = \int_\Omega F\left(x, u(x), Du(x)\right)\, dx \tag{A.1}$$

where $\Omega \subseteq \mathbb{R}^n$, $\mathcal{C}$ denotes a distinguished class of functions $u\colon \Omega \to \mathbb{R}^m$, and $F\colon \mathcal{D} \to \mathbb{R}$ is given, where $\mathcal{D}$ is sufficiently large so that the integral (A.1) makes sense for all $u \in \mathcal{C}$.

In this appendix, we derive one of the fundamental results from this field: the *Euler-Lagrange equations*. Under suitable regularity conditions on the function F, the Euler-Lagrange equations provide a second-order system of partial differential equations that any C^2 extremizer u of (A.1) must satisfy.

A.1 Variations

We begin our exploration of this topic by reasoning by analogy with the familiar case of extremizing a scalar-valued function on an open set $\Omega \subseteq \mathbb{R}^n$.

If $x \in \Omega$ extremizes a function $f \in C^1(\Omega, \mathbb{R})$, then necessarily

$$Df(x) = 0. \tag{A.2}$$

In order to distinguish a minimizer (respectively, maximizer) from other critical points, we have the additional necessary condition that the Hessian $D^2 f(x)$ be positive semidefinite (respectively, negative semidefinite), and (at least for a local minimizer) the sufficient condition that it be positive definite (respectively, negative definite).

In the present case, however, we do not have a function f of finitely many variables, but a *functional* $\mathcal{F} \colon \mathcal{C} \to \mathbb{R}$ whose domain is a *class of functions* $\mathcal{C}$. Motivated by the corresponding definitions for derivatives of functions whose domains are in real Euclidean spaces, we try to define a kind of partial derivative $\delta\mathcal{F}$ of $\mathcal{F}$, called a *variation*.

Definition A.1.1 (Variation). Let $\mathcal{F} \colon U \to \mathbb{R}$ be a functional defined on a subset U of a real linear space X. Let $u \in U$, $\xi \in X$, and suppose that there exists an $\epsilon > 0$ such that

$$\{u + t\xi \colon |t| < \epsilon\} \subseteq U.$$

Then we define the function

$$\Phi \colon (-\epsilon, \epsilon) \to \mathbb{R}$$

$$\Phi(t) = \mathcal{F}(u + t\xi),$$

and if $\Phi^{(n)}(0)$ exists, we define the *nth variation* $\delta^n \mathcal{F}(u, \xi)$ *of* $\mathcal{F}$ *at* u *in the direction of* ξ to be

$$\delta^n \mathcal{F}(u, \xi) = \Phi^{(n)}(0) \tag{A.3}$$

(here $\Phi^{(n)}$ denotes the nth derivative of Φ). We shall write $\delta \mathcal{F}(u, \xi)$ for $\delta^1 \mathcal{F}(u, \xi)$.

With this concept in hand, the same argument as is used in the proof of equation (A.2) in single-variable calculus proves the following.

Theorem A.1.1. *Let* $\mathcal{F} \colon U \to \mathbb{R}$ *be a functional defined on a subset* U *of a real linear space* X*. If* $u \in U$ *is an extremizer of* $\mathcal{F}$ *(i.e.,* u *minimizes or maximizes* $\mathcal{F}$*), and if* $\delta \mathcal{F}(u, \xi)$ *exists for some* $\xi \in X$*, then*

$$\delta \mathcal{F}(u, \xi) = 0. \tag{A.4}$$

Definition A.1.2 (Critical point, stationary point). Let $\mathcal{F} \colon U \to \mathbb{R}$ be a functional defined on a subset U of a real linear space X. Fix $u \in U$, and suppose that

$$\delta \mathcal{F}(u, \xi) = 0$$

for all ξ contained in a subset $V \subseteq X$. Then u is said to be a *critical point amongst variations in* V or a *stationary point amongst variations in* V.

In this terminology, Theorem A.1.1 states that in order for a vector u to be an extremizer for a functional $\mathcal{F}$, it is necessary that u be a critical point amongst all admissible variations of $\mathcal{F}$.

A.2 First variations of variational integrals

A.2.1 Notation and the standard hypotheses

Let $F\colon \mathcal{D} \to \mathbb{R}$, where $\mathcal{D} \subseteq \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn}$. We shall write the independent variables x, u, and p in coordinates as

$$x = (x^1, \ldots, x^\alpha, \ldots, x^n) \in \mathbb{R}^n,$$

$$u = (u^1, \ldots, u^i, \ldots, u^m) \in \mathbb{R}^m,$$

$$p = \begin{pmatrix} p_1^1 & \cdots & p_\alpha^1 & \cdots & p_n^1 \\ \vdots & & \vdots & & \vdots \\ p_1^i & \cdots & p_\alpha^i & \cdots & p_n^i \\ \vdots & & \vdots & & \vdots \\ p_1^m & \cdots & p_\alpha^m & \cdots & p_n^m \end{pmatrix} \in \mathbb{R}^{mn}.$$

Similarly, if $\varphi\colon \Omega \to \mathbb{R}^m$ is a function on a subset $\Omega \subseteq \mathbb{R}^n$, we shall write φ in coordinates as

$$\varphi(x) = (\varphi^1(x), \ldots, \varphi^i(x), \ldots, \varphi^m(x))$$

for scalar-valued functions $\varphi^i\colon \Omega \to \mathbb{R}$. We shall always use the convention that Greek indices run from 1 to n and Latin indices run from 1 to m. For convenience, we shall also make use of the *Einstein summation convention*, in which one sums over any repeated indices in a single term.

In the sequel, it will be necessary for us to distinguish carefully *composition of a partial derivative with another function* and *differentiation of a composition of functions*. To that end, we shall denote the partial derivative of a function F with respect to the variable x^α by F_{x^α}, and reserve the symbol D_{x^α} to denote partial differentiation of a *composition* of functions with respect to the variable x^α. Thus,

$$F_{p^i_\alpha}(x, u(x), Du(x)) = \frac{\partial F}{\partial p^i_\alpha} \circ (x, u(x), Du(x))$$

is the composition of the partial derivative $\frac{\partial F}{\partial p^i_\alpha}$ with the mapping $x \mapsto (x, u(x), Du(x))$, while

$$D_{x^\alpha} F_{p^i_\alpha}(x, u(x), Du(x)) = \frac{\partial}{\partial x^\alpha}\left[F_{p^i_\alpha}(x, u(x), Du(x))\right]$$

denotes the partial derivative of the *composition* $F_{p^i_\alpha}(x, u(x), Du(x))$.

Definition A.2.1 (Standard hypotheses). We shall frequently consider the case of functions $u \in C^1(\overline{\Omega}, \mathbb{R}^m)$ and $F \in C^1(\mathcal{D}, \mathbb{R})$, where $\Omega \subseteq \mathbb{R}^n$ is a bounded open set and $\mathcal{D} \subseteq \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn}$ is an open set with

$$\{(x, u(x), Du(x)) \mid x \in \overline{\Omega}\} \subseteq \mathcal{D}.$$

We shall refer to these conditions as the *standard hypotheses.*

A.2.2 A sufficient condition for the existence of the first variation

Theorem A.2.1 (A sufficient condition for the existence of the first variation). *Assume the standard hypotheses: let $\Omega \subseteq \mathbb{R}^n$ be a bounded open set, fix $u \in C^1(\overline{\Omega}, \mathbb{R}^m)$ and $F \in C^1(\mathcal{D}, \mathbb{R})$, where $\mathcal{D} \subseteq \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn}$ is open and*

$$\{(x, u(x), Du(x)) \mid x \in \overline{\Omega}\} \subseteq \mathcal{D},$$

and define the variational integral

$$\mathcal{F}(v) = \int_\Omega F(x, v(x), Dv(x))\, dx$$

as in (A.1). *Then for any* $\varphi \in C^1(\overline{\Omega}, \mathbb{R}^m)$, *the variation* $\delta\mathcal{F}(u,\varphi)$ *exists and is given by*

$$\begin{aligned}\delta\mathcal{F}(u,\varphi) &= \int_\Omega \left\{ \sum_{i=1}^m \frac{\partial F}{\partial u^i}(x,u(x),Du(x)) \cdot \varphi^i(x) \right. \\ &\qquad \left. + \sum_{\alpha=1}^n \frac{\partial F}{\partial p^i_\alpha}(x,u(x),Du(x)) \cdot \frac{\partial \varphi^i}{\partial x^\alpha}(x) \right\} \, dx \qquad \text{(A.5)} \\ &= \int_\Omega \left\{ F_{u^i}(x,u,Du)\varphi^i + F_{p^i_\alpha}(x,u,Du)\varphi^i_{x^\alpha} \right\} \, dx.\end{aligned}$$

Proof. Define

$$S = \{(x,u(x),Du(x)) \mid x \in \overline{\Omega}\}.$$

We wish to show that there exists some $\delta > 0$ such that

$$N = \{B_\delta(x,u(x),Du(x)) \mid x \in \overline{\Omega}\} \subseteq \mathcal{D}; \qquad \text{(A.6)}$$

that is, any point within a distance δ of S lies in $\mathcal{D}$. Since $\mathcal{D} \subseteq \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn}$ and is open, if $\partial\mathcal{D} = \emptyset$, $\mathcal{D} = \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn}$, and any choice of $\delta > 0$ will do. Thus, suppose to the contrary that $\partial\mathcal{D} \neq \emptyset$, and define the minimum distance function

$$d_{\partial\mathcal{D}} \colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn} \to \mathbb{R}$$
$$d_{\partial\mathcal{D}}(y) = \min_{z \in \partial\mathcal{D}} \|y - z\|_2$$

for all $z \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn}$, where $\|\cdot\|_2$ denotes the standard Euclidean distance. Note that $d_{\partial\mathcal{D}}$ is a well-defined function since $\partial\mathcal{D}$ is a closed set. For let $y \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{mn}$, and let $\overline{B_R(y)}$ be a closed ball of radius R centered on y

with R sufficiently large that $B = \overline{B_R(y)} \cap \partial\mathcal{D} \neq \emptyset$. Then

$$\inf_{z\in\partial\mathcal{D}} \|y - z\|_2 = \inf_{z\in B} \|y - z\|_2 = \min_{z\in B} \|y - z\|_2$$

since B is closed and bounded, hence compact. Now $d_{\partial\mathcal{D}}$ is a continuous function, and since S is the continuous image of the compact set $\overline{\Omega}$, then it too is compact, and consequently $d_{\partial\mathcal{D}}$ attains a minimum on S. We claim that

$$\delta = \min_{y\in S} d_{\partial\mathcal{D}}(y) > 0. \tag{A.7}$$

For suppose that $\delta = 0$. Then there exists some point $y \in S$ such that $d_{\partial\mathcal{D}}(y) = 0$, so that $y \in \partial\mathcal{D}$, contradicting the fact that $y \in S \subseteq \mathcal{D}$. This proves (A.7), which in turn proves (A.6).

Now, recall that the space $C^1(\overline{\Omega}, \mathbb{R}^m)$ is a Banach space when equipped with the norm

$$\|f\|_{C^1} = \|f\|_\infty + \sum_{i=1}^{m} \|\nabla f_i\|_\infty$$

(here we use $\|\cdot\|_\infty$ to denote the norm

$$\|f\|_\infty = \sup_{x\in\overline{\Omega}} \|f(x)\|_2$$

where $\|\cdot\|_2$ is the standard Euclidean distance). If $v \in C^1(\overline{\Omega}, \mathbb{R}^m)$ and

$\|u-v\|_{C^1} < \gamma$, then

$$\begin{aligned}\|(x,u(x),Du(x)) - (x,v(x),Dv(x))\| &= \|(0,(u-v)(x),D(u-v)(x))\| \\ &\leq \sqrt{\|u-v\|_\infty^2 + \|D(u-v)\|_\infty^2} \\ &\leq \sqrt{\|u-v\|_\infty^2 + \sum_{i=1}^{m} \|\nabla(u-v)_i\|_\infty^2} \\ &< \frac{\gamma}{\sqrt{m+1}}.\end{aligned} \tag{A.8}$$

Consequently, if $\|u-v\|_{C^1} < \delta\sqrt{m+1}$, then equations (A.6) and (A.8) show that

$$\{(x,v(x),Dv(x)) \mid x \in \overline{\Omega}\} \subseteq \mathcal{D},$$

and therefore, given *any* $\varphi \in C^1(\overline{\Omega}, \mathbb{R}^m)$, the function

$$\begin{gathered}\Phi\colon (-\epsilon,\epsilon) \to \mathbb{R} \\ \Phi(t) = \mathcal{F}(u+t\varphi)\end{gathered}$$

is defined for $0 < \epsilon < (\delta\sqrt{m+1})/\|\varphi\|_{C^1}$. Now

$$\Phi(t) = \mathcal{F}(u+t\varphi) = \int_\Omega G(t,x)\,dx \tag{A.9}$$

for

$$\begin{gathered}G\colon (-\epsilon,\epsilon)\times\overline{\Omega} \\ G(t,x) = F\left(x,\ u(x)+t\varphi(x),\ Du(x)+tD\varphi(x)\right).\end{gathered}$$

Furthermore, G is continuously differentiable since it is a composition of continuously differentiable functions, and

$$\frac{\partial G}{\partial t}(t,x) = \sum_{i=1}^{m} \frac{\partial F}{\partial u^i}(x,\ u(x)+t\varphi(x),\ Du(x)+tD\varphi(x)) \cdot \varphi^i(x) + \sum_{i=1}^{m}\sum_{\alpha=1}^{n} \frac{\partial F}{p^i_\alpha}(x,\ u+t\varphi(x),\ Du(x)+tD\varphi(x)) \cdot \frac{\partial \varphi^i}{\partial x^\alpha}(x)$$

by the Chain Rule. Since G is continuous, it is bounded on the compact subset $[-\epsilon/2, \epsilon/2] \times \overline{\Omega}$ by some constant M, and the constant function M is integrable over Ω since Ω is bounded. Consequently, the Leibniz integral rule for multivariable Lebesgue integrals applied to equation (A.9) asserts that

$$\begin{aligned} \Phi'(0) &= \int_\Omega \frac{\partial G}{\partial t}(0,x)\,dx \\ &= \int_\Omega \left\{ \sum_{i=1}^{m} \frac{\partial F}{\partial u^i}(x,u(x),Du(x)) \cdot \varphi^i(x) + \sum_{\alpha=1}^{n} \frac{\partial F}{\partial p^i_\alpha}(x,u(x),Du(x)) \cdot \frac{\partial \varphi^i}{\partial x^\alpha}(x) \right\} dx. \end{aligned}$$

Since $\delta\mathcal{F}(u,\varphi) = \Phi'(0)$ by definition, this proves the claim. □

A.2.3 The weak Euler-Lagrange equation and weak extremals

Assuming the standard hypotheses, Theorems A.1.1 and A.2.1 show that

$$\delta\mathcal{F}(u,\varphi) = 0$$

for all $\varphi \in C^1(\overline{\Omega}, \mathbb{R}^m)$ if u is an extremizer of $\mathcal{F}$. In particular, $\delta\mathcal{F}(u,\varphi) = 0$ for all $\varphi \in C_0^\infty(\Omega, \mathbb{R}^m)$, the set of compactly supported smooth functions on

Ω. By Theorem A.2.1, this is equivalent to

$$\int_\Omega \left\{ \sum_{i=1}^m \left(F_{u^i}(x,u,Du)\varphi^i + \sum_{\alpha=1}^m F_{p^i_\alpha}(x,u,Du)\varphi^i_{x^\alpha} \right) \right\} dx = 0 \qquad \text{(A.10)}$$

for all $\varphi \in C_0^\infty(\Omega, \mathbb{R}^m)$.

Definition A.2.2 (Weak Euler-Lagrange equation, weak extremal). Equation (A.10) is called the *weak Euler-Lagrange equation*, and functions $u \in C^1(\Omega, \mathbb{R}^m)$ satisfying it are called *weak extremals*. Equivalently, a weak extremal is a critical point of $\mathcal{F}$ amongst variations in $C_0^\infty(\Omega, \mathbb{R}^m)$.

A.3 The Euler-Lagrange Equations

A.3.1 The Fundamental Lemma of the Calculus of Variations

Lemma A.3.1 (The Fundamental Lemma of the Calculus of Variations). *Let $\Omega \subseteq \mathbb{R}^n$ be an open set, and suppose that $f \in C^0(\Omega, \mathbb{R})$ satisfies*

$$\int_\Omega f(x)\varphi(x)\,dx = 0 \text{ for all } \varphi \in C_0^\infty(\Omega, \mathbb{R}). \qquad \text{(A.11)}$$

Then $f(x) = 0$ for all $x \in \Omega$.

Proof. Suppose by way of contradiction that there exists some $x_0 \in \Omega$ for which $f(x_0) > 0$. The continuity of f implies the existence of a $\delta > 0$ such that $f > 0$ on all of $B_\delta(x_0)$, where δ is taken sufficiently small that $\overline{B_\delta(x_0)} \subset \Omega$. Define

$$\eta \colon \mathbb{R}^n \to \mathbb{R}$$
$$\eta(x) = \begin{cases} \exp\left(-\frac{1}{\delta^2 - \|x-x_0\|^2}\right), & x \in B_\delta(x_0), \\ 0, & x \notin B_\delta(x_0). \end{cases}$$

Then $\eta \in C_0^\infty(\Omega, \mathbb{R})$, $\eta > 0$ on $B_\delta(x_0)$, and $\eta = 0$ outside of $\overline{B_\delta(x_0)}$. Thus,

$$\int_\Omega f(x)\eta(x)\, dx$$

is the integral of a function that is strictly positive on $B_\delta(x_0)$ and identically zero on $\Omega \backslash \overline{B_\delta(x_0)}$, hence must be positive. But this contradicts (A.11). A similar argument shows that f cannot take on any negative values. Consequently, f must be identically zero. □

A.3.2 The Euler-Lagrange Equations

Theorem A.3.2 (The Euler-Lagrange equations). *Assume the standard hypotheses, and suppose that u is a weak extremal of the variational integral*

$$\mathcal{F}(v) = \int_\Omega F\left(x, v(x), Dv(x)\right)\, dx.$$

Suppose further that $F_p \in C^1(\mathcal{D}, \mathbb{R}^{mn})$ and $u \in C^2(\Omega, \mathbb{R}^m)$. Then $u(x)$ satisfies

$$D_{x^\alpha} F_{p^i_\alpha}(x, u(x), Du(x)) = F_{u^i}(x, u(x), Du(x)) \tag{A.12}$$

for all $1 \leq i \leq m$ and all $x \in \Omega$.

Proof. By hypothesis,

$$\int_\Omega \left\{ F_{u^i}(x, u, Du)\varphi^i + F_{p^i_\alpha}(x, u, Du)\varphi^i_{x^\alpha} \right\}\, dx = 0 \tag{A.13}$$

for all $\varphi \in C_0^\infty(\Omega, \mathbb{R}^m)$. Integrating the second term of the integrand of (A.13) by parts and using the fact that φ vanishes along $\partial\Omega$ due to compact support, we obtain:

$$\begin{aligned} 0 &= \int_\Omega \left\{ F_{u^i}(x,u,Du)\varphi^i - \left[D_{x^\alpha} F_{p^i_\alpha}(x,u,Du) \right] \varphi^i \right\} \, dx \\ &= \int_\Omega \left(F_{u^i}(x,u,Du) - D_{x^\alpha} F_{p^i_\alpha}(x,u,Du) \right) \varphi^i \, dx. \end{aligned} \tag{A.14}$$

Now by hypothesis, (A.14) must hold for *all* choices of $\varphi \in C_0^\infty(\Omega, \mathbb{R}^m)$; in particular, fixing a value of i, we can choose $\varphi = (\varphi^1, \ldots, \varphi^i, \ldots \varphi^m)$ such that $\varphi^j = 0$ for all $j \neq i$, so that all but the ith term of the integrand in (A.14) vanishes. Applying the Fundamental Lemma of the Calculus of Variations term-by-term to each of these m terms then gives (A.12). □

Definition A.3.1 (The Euler-Lagrange equations). The m partial differential equations (A.12) are called the *Euler-Lagrange equations*.

Appendix B

Symplectic Geometry

In this appendix, we provide a brief introduction to *symplectic geometry*, the study of smooth manifolds equipped with a nondegenerate closed 2-form. The geometric and topological properties of these manifolds turns out to be closely related to the Hamiltonian formulation of geometric mechanics; consequently, they feature prominently in our study.

B.1 Symplectic vector spaces

Definition B.1.1 (Bilinear form). Let V be a real vector space. A *bilinear form* on V is a mapping $\omega\colon V \times V \to \mathbb{R}$ that is linear in each coordinate:

$$\omega(ax + by, z) = a\omega(x, z) + b\omega(y, z),$$

$$\omega(z, ax + by) = a\omega(z, x) + b\omega(z, y)$$

for all $a, b \in \mathbb{R}$ and all $x, y, z \in V$. Equivalently, a bilinear form is a 2-tensor on V.

Definition B.1.2 (Nondegenerate bilinear form). A bilinear form ω on a finite-dimensional real vector space V is called *nondegenerate* if, given any $x \in V$, $\omega(x, y) = 0$ for all $y \in V$ implies that $x = 0$.

Theorem B.1.1. *Let V be a finite-dimensional vector space and ω a nondegenerate bilinear form on V. Then ω induces a linear isomorphism $\omega_\flat$ between V and its dual space V^*:*

$$\omega_\flat \colon V \to V^*$$
$$\omega_\flat(x) = \omega(x, \cdot).$$

Proof. First, observe that each of the maps $\omega_\flat(x) = \omega(x, \cdot)$ is a linear functional on V by the linearity of ω in the second coordinate, and that $\omega_\flat$ itself is also a linear map by the linearity of ω in the first coordinate. The nondegeneracy of ω shows that $\omega_\flat(x) = 0$ if and only if $x = 0$. Since V is finite-dimensional, V and V^* have the same dimension, and therefore $\omega_\flat$ is an injective linear mapping between two linear spaces of the same dimension, hence must be an isomorphism. □

Definition B.1.3 (Flat map, sharp map). The map $\omega_\flat$ in Theorem B.1.1 is called the *flat map*. Its inverse isomorphism $\omega_\flat^{-1} \colon V^* \to V$ is denoted $\omega_\sharp$, called the *sharp map*.

Remark B.1.1. We shall often write $X^\flat$ for $\omega_\flat(X)$ and $\alpha^\sharp$ for $\omega_\sharp(\alpha)$ when the bilinear form ω is understood.

Theorem B.1.2. *Let V be a finite-dimensional vector space and ω a nondegenerate bilinear form on V. Given any $X, Y \in V$ and any $\alpha \in V^*$, the following* *identities hold:*

$$X^\flat(Y) = \omega(X, Y),$$
$$\omega(\alpha^\sharp, X) = \alpha(X).$$

Proof. The first identity holds by definition. To see the second, simply observe that $Z = \alpha^\sharp$ is, by definition, the inverse of α under the flat map; that is, it satisfies

$$Z^\flat(X) = \alpha(X).$$

Since

$$Z^\flat(X) = \omega(Z, X) = \omega(\alpha^\sharp, X),$$

the result is proved. □

Given a finite-dimensional real linear space V and a basis $\{e_1, \ldots, e_n\}$ for V, let $x, y \in V$. The fact that $\{e_1, \ldots, e_n\}$ is a basis means that there exist uniquely determined constants $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}$ such that

$$x = \sum_{i=1}^{n} x_i e_i, \quad y = \sum_{j=1}^{n} y_j e_j, \tag{B.1}$$

and therefore

$$\begin{aligned} \omega(x, y) &= \omega\left(\sum_{i=1}^{n} x_i e_i, \sum_{i=1}^{n} y_i e_i\right) \\ &= \sum_{i=1}^{n} x_i \cdot \omega\left(e_i, \sum_{j=1}^{n} y_j e_j\right) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \cdot \omega(e_i, e_j). \end{aligned} \tag{B.2}$$

Equation (B.2) shows that ω is *completely determined by its values on the set of pairs* (e_i, e_j) for $1 \leq i, j \leq n$, and conversely, specifying a value for the pairings (e_i, e_j) uniquely determines a corresponding bilinear form on V. Consequently, to each such form ω we can associate an $n \times n$ matrix W, given by

$$W = (\omega_{ij})_{ij=1}^{n} = \begin{pmatrix} \omega_{11} & \cdots & \omega_{1j} & \cdots & \omega_{1n} \\ \vdots & & \vdots & & \vdots \\ \omega_{i1} & \cdots & \omega_{ij} & \cdots & \omega_{in} \\ \vdots & & \vdots & & \vdots \\ \omega_{n1} & \cdots & \omega_{nj} & \cdots & \omega_{nn} \end{pmatrix}, \qquad \text{(B.3)}$$

$$\omega_{ij} = \omega(e_i, e_j)$$

and equation (B.2) shows that, given a fixed basis $\{e_1, \ldots, e_n\}$, the correspondence $\omega \leftrightarrow W$ as given in (B.3) is a bijection between the set of bilinear forms on V and the set of $n \times n$ matrices.

Now, let $x, y \in V$ be written in coordinates as $n \times 1$ column vectors relative to the basis $\{e_1, \ldots, e_n\}$ as in (B.1), and let $u = Wy$. Then by the definition of matrix multiplication,

$$u_i = \sum_{j=1}^{n} \omega_{ij} y_j,$$

and therefore

$$\begin{aligned} x^t u &= \sum_{i=1}^{n} x_i u_i \\ &= \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} \omega_{ij} y_j \right) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \cdot \omega(e_i, e_j) \\ &= \omega(x, y), \end{aligned} \tag{B.4}$$

where the last equality follows from equation (B.2). But since $u = Wy$, equation (B.4) shows that, given two vectors $x, y \in V$, we can compute the value of $\omega(x, y)$ according to

$$\omega(x, y) = x^t W y, \tag{B.5}$$

where we regard x and y as $n \times 1$ column vectors written in coordinates relative to the basis $\{e_1, \ldots, e_n\}$, and W is as in (B.3).

Theorem B.1.3. *Let ω be a bilinear form on a finite-dimensional real vector space V with basis $\{e_1, \ldots, e_n\}$. Then ω is nondegenerate if and only if the corresponding matrix W in* (B.3) *is nonsingular.*

Proof. By definition, ω is nondegenerate if and only if, given any $x \neq 0$, there exists some $y \in V$ such that

$$\omega(x, y) = x^t W y \neq 0,$$

or

$$y^t W^t x \neq 0. \tag{B.6}$$

If W^t is nonsingular, then $W^t x \neq 0$ for all $x \neq 0$. Consequently, for any $x \neq 0$, there exists some $y \in V$ for which $y^t W^t x \neq 0$ (simply take $y = (0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 occurs at any nonzero coordinate in $W^t x$). Thus, ω is nondegenerate if W^t is nonsingular.

Conversely, if ω is nondegenerate, then given any $x \neq 0$, there must be some $y \in V$ such that $\omega(x, y) \neq 0$. In particular, $W^t x \neq 0$, otherwise equation (B.6) could not be satisfied. Thus, W^t is nonsingular when ω is nondegenerate.

The result now follows, since the transpose of a matrix is nonsingular if and only if the original matrix is nonsingular. □

Definition B.1.4 (Skew-symmetric form). A bilinear form ω on a real vector space V is called *skew-symmetric* if $\omega(x, y) = -\omega(y, x)$ for all $x, y \in V$.

Remark B.1.2. Equation (B.3) shows that ω is a skew-symmetric bilinear form if and only if its corresponding matrix W is skew-symmetric.

Theorem B.1.4. *Let ω be a skew-symmetric bilinear form on the finite-dimensional real vector space V. Then there exists a set of basis vectors $n_1, \ldots, n_d, x_1, \ldots, x_r, y_1, \ldots, y_r$ for V with the following properties:*

1. $\omega(n_i, v) = 0$ *for all* $1 \leq i \leq d$ *and all* $v \in V$.

2. $\omega(x_i, y_j) = \delta_{ij}$ *for all* $1 \leq i, j \leq r$.

3. $\omega(x_i, x_j) = \omega(y_i, y_j) = 0$ *for all* $1 \leq i, j \leq r$.

Consequently, the matrix W of ω relative to the ordered basis $\{x_1, \ldots, x_r, y_1, \ldots, y_r, n_1, \ldots, n_d\}$ is given by

$$W = \begin{pmatrix} 0 & I_r & 0 \\ -I_r & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{B.7}$$

where I_r denotes the $r \times r$ identity matrix. Equivalently, ω may be written as

$$\omega = \sum_{i=1}^{r} dx_i \wedge dy_i. \tag{B.8}$$

Proof. Define

$$N = \{n \in V \mid \omega(n, v) = 0 \text{ for all } v \in V\}.$$

The linearity of ω in its first coordinate shows that N is a subspace of V, and therefore it has a basis $n_1, \ldots, n_d$. Fix a subspace U_m of V complementary to N with dimension m, so that

$$V = N \oplus U_m.$$

Then ω is nondegenerate on U_m: for any vector $u \in U_m$, there exists some $v \in V$ such that $\omega(u, v) \neq 0$, and in fact $v \in U_m$. (If this were not the case, then $\omega(u, v) = 0$ for all $v \in V$, so that $u \in N$, a contradiction, since N and U_m are complementary in V. Similarly, if this particular v were in N, then $\omega(u, v) = 0$ automatically by the skew-symmetry of ω and the definition of N). We will show how to inductively build up a basis $x_1, \ldots, x_r, y_1, \ldots, y_r$ for U_m such that the combined set $n_1, \ldots, n_d, x_1, \ldots, x_r, y_1, \ldots, y_r$ is a basis for V with the desired properties.

Let x_1 be an arbitrary nonzero vector in U_m. By what we just showed, there exists some vector $y_1' \in U_m$ such that $\omega(x_1, y_1') \neq 0$. Setting

$$y_1 = \frac{1}{\omega(x_1, y_1')} y_1'$$

the linearity of ω in the second coordinate shows that

$$\omega(x_1, y_1) = \omega\left(x_1, \frac{1}{\omega(x_1, y_1')} y_1'\right) = \frac{1}{\omega(x_1, y_1')} \omega(x_1, y_1') = 1.$$

Futhermore, x_1 and y_1 must be linearly independent; for if not, then there exists some constant $c \in \mathbb{R}$ such that $y_1 = cx_1$, and therefore

$$\omega(x_1, y_1) = c\omega(x_1, x_1) = -c\omega(x_1, x_1) = 0$$

since ω is skew-symmetric, contradicting the fact that $\omega(x_1, y_1) = 1$. Thus, x_1 and y_1 are linearly independent vectors spanning some 2-dimensional subspace of U_m. Now define

$$S_1 = \operatorname{span}(x_1, y_1)$$

and

$$U_{m-2} = \{u \in U_m \mid \omega(u, s) = 0 \text{ for all } s \in S_1\}.$$

Again by the linearity of ω, U_{m-2} is a subspace of U_m, and we claim that

$$U_{m-2} \cap S_1 = \{0\}. \tag{B.9}$$

For suppose that $u \in U_{m-2} \cap S_1$. Then by definition, $u = cx_1 + dy_1$ for some $c, d \in \mathbb{R}$ since $S_1 = \operatorname{span}(x_1, y_1)$ and $u \in S_1$. On the other hand, since

$u \in U_{m-2}$ and $x_1 \in S_1$, then

$$\begin{aligned} 0 &= \omega(u, x_1) \\ &= \omega(cx_1 + dy_1, x_1) \\ &= c\omega(x_1, x_1) + d\omega(y_1, x_1) \\ &= -d \end{aligned}$$

by the skew-symmetry of ω, and similarly

$$\begin{aligned} 0 &= \omega(u, y_1) \\ &= \omega(cx_1 + dy_1, y_1) \\ &= c\omega(x_1, y_1) + d\omega(y_1, y_1) \\ &= c; \end{aligned}$$

consequently, we see that $u = 0x_1 + 0y_1 = 0$, as claimed, and (B.9) is proved.

Conversely, we claim that

$$U_m = U_{m-2} + S_1. \tag{B.10}$$

For given any $u \in U_m$, define

$$\begin{aligned} s &= \omega(u, y_1)x_1 - \omega(u, x_1)y_1, \\ r &= u - s. \end{aligned} \tag{B.11}$$

Then obviously $u = s + r$, and $s \in S_1$ by definition. To see that $r \in U_{m-2}$, simply observe that

$$\begin{aligned} \omega(r, x_1) &= \omega\left(u - \omega(u, y_1)x_1 + \omega(u, x_1)y_1, x_1\right) \\ &= \omega(u, x_1) - \omega(u, y_1)\omega(x_1, x_1) + \omega(u, x_1)\omega(y_1, x_1) \\ &= \omega(u, x_1) - \omega(u, x_1) \\ &= 0 \end{aligned}$$

and

$$\begin{aligned}\omega(r, y_1) &= \omega\left(u - \omega(u, y_1)x_1 + \omega(u, x_1)y_1, y_1\right) \\ &= \omega(u, y_1) - \omega(u, y_1)\omega(x_1, y_1) + \omega(u, x_1)\omega(y_1, y_1) \\ &= \omega(u, y_1) - \omega(u, y_1). \\ &= 0\end{aligned}$$

Since $\omega(r, x_1) = \omega(r, y_1) = 0$ and $S_1 = \text{span}(x_1, y_1)$, the linearity of ω in the second coordinate shows that $\omega(r, s) = 0$ for all $s \in S_1$, and therefore $r \in U_{m-2}$ by definition. This proves (B.10), and (B.9) and (B.10) together show that

$$U_m = S_1 \oplus U_{m-2} \tag{B.12}$$

Now $U_{m-2} \subset U_m$ so that every vector $u \in U_{m-2}$ has some vector u' with which it pairs to give a nonzero pairing $\omega(u, u')$, and we have already proved that $u' \in U_m$. Since $u \in U_{m-2}$ and $\omega(u, u') \neq 0$, then $u' \notin S_1$, and therefore we can compute r' and s' with $u' = r' + s'$ as in (B.11). Since

$$\omega(u, u') = \omega(u, r' + s') = \omega(u, r') + \omega(u, s') = \omega(u, r')$$

with $r' \in U_{m-2}$, this shows that every vector $u \in U_{m-2}$ has a nonzero pairing with another vector $v \in U_{m-2}$, i.e., ω is nondegenerate when restricted to the subspace U_{m-2}. Consequently, we can iterate our decomposition of U_m into S_1 and U_{m-2} by decomposing U_{m-2} into as

$$U_{m-2} = S_2 \oplus U_{m-4},$$

where

$$S_2 = \text{span}(x_2, y_2)$$

for vectors $x_2, y_2 \in U_{m-2}$ satisfying $\omega(x_2, y_2) = 1$. Since $x_2, y_2 \in S_2 \subseteq U_{m-2}$, then

$$\omega(x_1, x_2) = \omega(x_1, y_2) = \omega(x_2, y_1) = \omega(y_1, y_2) = 0$$

by definition of U_{m-2}. Iterating on this procedure, we can continue decomposing U_m as a decreasing sequence of nested subspaces

$$U_m \supseteq U_{m-2} \supseteq U_{m-4} \supseteq \cdots \supseteq U_{m-2r} \supseteq \cdots \tag{B.13}$$

and building up a collection of subspaces $S_i = \text{span}(x_i, y_i)$ for vectors $\{x_i, y_i\}$ satisfying the conditions given in the theorem. Thus, to complete the proof, we need only show that the sequence (B.13) terminates on the subspace $\{0\}$ (so that the set $\{x_i, y_i\}$ spans *all* of U_m, hence forms a basis for V when combined with the $n_1, \ldots, n_d$). But this is immediate. For suppose, by way of contradiction, that (B.13) does *not* exhaust U_m. Since the dimension of each subsequent subspace in the sequence decreases by 2 and U_m is finite dimensional, the only way that (B.13) could fail to exhaust U is if m is not an even integer. In that case, letting $r = \lfloor \frac{m}{2} \rfloor$, we have

$$\dim(U_{m-2r}) = U_1 = 1,$$

and we have already shown that ω must be nondegenerate on U_1. But this is a contradiction, for ω cannot be nondegenerate on a 1-dimensional subspace, as it is skew-symmetric. Consequently, m *must* be even, in which case (B.13) is guaranteed to exhaust U_m, and in so doing, produce a bases $\{x_i, y_i \mid 1 \leq i \leq r = \frac{m}{2}\}$ for U with the properties stated in the theorem. Taking the union of this basis with the basis $n_1, \ldots, n_d$ for N gives the desired basis for V. $\square$

Remark B.1.3. Note that the proof of Theorem B.1.4 actually gives a constructive algorithm for *finding* the basis whose existence is asserted.

Corollary B.1.5. *Let ω be a a skew-symmetric bilinear form on V. Then ω is nondegenerate if and only if V has even dimension $2n$ and ω^n is a volume form on V.*

Proof. Since ω is skew-symmetric, Theorem B.1.4 shows that there exist linearly independent vectors $x_1, \ldots, x_n, y_1, \ldots, y_n$ with dual functionals $dx_1, \ldots dx_n, dy_1, \ldots, dy_n$ such that ω can be written as

$$\omega = \sum_{i=1}^{n} dx_i \wedge dy_i. \tag{B.14}$$

If ω is nondegenerate, then equation (B.7) and Theorem B.1.3 show that $\dim(V) = 2n$. Furthermore, with ω written as (B.14) we have by induction that

$$\omega^k = \sum_{j_1, \ldots, j_k = 1}^{n} dx_{j_1} \wedge dy_{j_1} \wedge \cdots \wedge dx_{j_k} \wedge dy_{j_k},$$

so that

$$\omega^n = n!(-1)^{\frac{n}{2}} dx_1 \wedge \cdots dx_n \wedge dy_1 \wedge \cdots \wedge dy_n. \tag{B.15}$$

Conversely, if ω^n is a volume form, then (B.14), (B.15), and (B.7) show that ω must be nondegenerate. □

Definition B.1.5 (Symplectic form, vector space, map). Let V be a finite-dimensional real vector space. A *symplectic form* on V is a nondegenerate skew-symmetric bilinear form ω. The pair (V, ω) is called a *symplectic vector* *space*. If (V, ω) and (W, ρ) are two symplectic vector spaces, a linear map $T: V \to W$ is called *symplectic* if $T^*\rho = \omega$. Equivalently, T is called *symplectic* if $\rho(T(x), T(y)) = \omega(x, y)$ for all $x, y \in V$.

As an immediate consequence, we have the following.

Theorem B.1.6. *Let (V, ω) and (W, ρ) be $2n$-dimensional symplectic vector spaces. If $T: V \to W$ is a symplectic linear transformation, then T is an isomorphism.*

Proof. First, T must be injective; otherwise, some nonzero $x \in V$ would map to 0 under T, and consequently,

$$T^*\rho(x, y) = \rho(T(x), T(y)) = \rho(0, T(y)) = 0$$

for all $y \in V$ with $x \neq 0$, so that $T^*\rho$ is not nondegenerate. Since $T^*\rho = \omega$ *is* nondegenerate, T is injective. Since T is an injective linear map between spaces of the same dimension, it is an isomorphism. □

Theorem B.1.7. *Let (V, ω) be a symplectic vector space of dimension $2n$. The set*

$$Sp(2n) = \{T: V \to V \mid T \text{ is a symplectic linear map}\}$$

of all symplectic automorphisms of V forms a group, called the symplectic group. In fact, the symplectic group $Sp(2n)$ is Lie group of dimension $2n^2 + n$: fixing a basis for V as in Theorem B.1.4*, $Sp(2n)$ can be identified with the* *matrix subgroup:*

$$Sp(2n) = \{A \in SL_{2n}(\mathbb{R}) \mid A^t J A = J\} \subseteq SL_{2n}(\mathbb{R}) \tag{B.16}$$

where

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

and under this identification, its Lie algebra $\mathfrak{sp}(2n)$ *is given by:*

$$\mathfrak{sp}(2n) = \{A \in \mathbb{R}^{2n \times 2n} \mid JA + A^t J = 0\}.$$

Proof. We first prove that $Sp(2n)$ is a group. Theorem B.1.6 shows that every symplectic map $T\colon V \to V$ is an automorphism, and so $Sp(2n)$ is a subset of the automorphism group of V. Consequently, to prove that $Sp(2n)$ is a subgroup, it suffices to show that it is closed under inverses and products. To that end, let $T_1, T_2\colon V \to V$ be two symplectic automorphisms of V. Then

$$(T_1 \circ T_2)^* \omega = T_2^*(T_1^* \omega) = T_2^* \omega = \omega$$

since T_1 and T_2 are each individually symplectic (hence preserve ω). Similarly, to show that T_1^{-1} is also symplectic, observe that

$$(T_1^{-1})^* \omega = (T_1^{-1})^*(T_1^* \omega) = (T_1 \circ T_1^{-1})^* \omega = I^* \omega = \omega,$$

and therefore T_1^{-1} is symplectic whenever T_1 is. This shows that $Sp(2n)$ is a subgroup of $\mathrm{Aut}(V)$.

Now we show that $Sp(2n)$ is in fact a Lie group by explicitly identifying it with a matrix subgroup of $GL_{2n}(\mathbb{R})$. By Theorem B.1.4 and equation (B.5), given any symplectic form ω on a vector space of dimension $2n$, there exists a basis $\mathcal{B}$ for V such that, when two arbitrary vectors $x, y \in V$ are written in coordinates as $2n \times 1$ column vectors with respect to this basis, the value of the pairing $\omega(x, y)$ may be computed as

$$\omega(x, y) = x^t J y, \tag{B.17}$$

where

$$J = \begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix}$$

and I_n denotes the $n \times n$ identity matrix.

Now any endomorphism $T \colon V \to V$ can be written in coordinates as a matrix M with respect to $\mathcal{B}$. By definition, a linear mapping T is symplectic if and only if $\omega(T(x), T(y)) = \omega(x, y)$ for all $x, y \in V$; by equation (B.17), this is equivalent to the statement that

$$\begin{aligned} x^t J y &= (Mx)^t J (My) \\ &= x^t M^t J M y \end{aligned} \tag{B.18}$$

for all $x, y \in V$. But observe that, choosing x to be the $2n \times 1$ vector with 1 in its ith coordinate and zeros elsewhere and y to be the $2n \times 1$ vector with 1 in its jth coordinate and zeros elsewhere, the product $x^t(M^t J M)y$ picks out the ij-element of $(M^t J M)$, which must be equal to the ij-element of J by equation (B.18) if M represents a symplectic map. Consequently, we conclude that M is the matrix of a symplectic linear map with respect to $\mathcal{B}$ if and only if

$$M^t J M = J.$$

This shows that equation (B.16) realizes $Sp(2n)$ as a matrix subgroup of $GL_{2n}(\mathbb{R})$. Now we show that $Sp(2n)$ is actually contained in $SL_{2n}(\mathbb{R})$. Given any $T \in Sp(2n)$, we know that T preserves the symplectic form ω on V by definition:

$$T^*\omega = \omega.$$

Corollary B.1.5 proves that ω^n is a volume form, and the fact that the pullback mapping T^* is an algebra homomorphism on the de Rham cohomology algebra $H^*_{\mathrm{dR}}(V)$ shows that

$$\omega^n = T^*(\omega^n) = \det(T)\omega^n$$

so that $\det(T) = 1$. Therefore, the matrix representation A of T also has determinant 1, hence lies in $SL_{2n}(\mathbb{R}) \subseteq GL_{2n}(\mathbb{R})$.

Now, observe that, for *any* matrix $A \in \mathbb{R}^{2n \times 2n}$, the product A^tJA is skew-symmetric:

$$(A^tJA)^t = A^tJ^tA = -A^tJA$$

since $J^t = -J$ as J is skew-symmetric by definition. Let $Sk_{2n}(\mathbb{R})$ denote the real skew-symmetric matrices of dimension $2n \times 2n$. These matrices have $4n^2$ elements, but the condition of skew-symmetry means that the $2n$ elements along the diagonal must be zero, and the remaining $\frac{4n^2-2n}{2} = 2n^2 - n$ elements above the diagonal determine the corresponding elements below the diagonal by skew-symmetry. Conversely, specifying *any* choice of the $2n^2 - n$ elements above the diagonal produces a uniquely determined skew-symmetric matrix. Consequently, fixing the basis $\mathcal{B}$, we can identify the set $Sk_{2n}(\mathbb{R})$ with a $(2n^2 - n)$-dimensional submanifold of $\mathbb{R}^{2n\times 2n}$, the manifold of $2n \times 2n$ matrices, and via this identification, we also immeidately see that $T_M(Sk_{2n}(\mathbb{R})) = Sk_{2n}(\mathbb{R})$ at each point $M \in Sk_{2n}(\mathbb{R})$.

Now, define the mapping

$$f\colon \mathbb{R}^{2n\times 2n} \to Sk_{2n}(\mathbb{R})$$
$$f(A) = A^t JA.$$

Then $Sp(2n) = f^{-1}(J)$ by (B.16), so we can use f together with some differential topology to deduce some additional information about $Sp(2n)$. By definition:

$$\begin{aligned} df_A(B) &= \lim_{h\to 0} \frac{f(A+hB) - f(A)}{h} \\ &= \lim_{h\to 0} \frac{(A+hB)^t J(A+hB) - A^t JA}{h} \\ &= \lim_{h\to 0} \frac{A^t JA + hA^t JB + hB^t JA + h^2 B^t JB - A^t JA}{h} \\ &= A^t JB + B^t JA. \end{aligned} \tag{B.19}$$

We wish to show that, for an arbitrary $A \in Sp(2n) = f^{-1}(J)$, the derivative mapping $df_A\colon T_A(\mathbb{R}^{2n\times 2n}) \to T_{f(A)}(Sk_{2n}(\mathbb{R})) = Sk_{2n}(\mathbb{R})$ is surjective. To that end, let $C \in Sk_{2n}(\mathbb{R})$; we must show that there exists some $B \in T_A(\mathbb{R}^{2n\times 2n}) = \mathbb{R}^{2n\times 2n}$ such that

$$df_A(B) = A^t JB + B^t JA = C.$$

Since $A \in Sp(2n)$, A is invertible, and therefore so is A^t, and J is likewise invertible. Thus, we can choose

$$B = \frac{1}{2}(A^t J)^{-1} C,$$

and equation (B.19) shows that

$$\begin{aligned}
df_A(B) &= A^t J \left(\frac{1}{2}(A^t J)^{-1} C\right) + \left(\frac{1}{2}(A^t J)^{-1} C\right)^t JA \\
&= \frac{1}{2}C + \frac{1}{2}\left(J^{-1}(A^t)^{-1}C\right)^t JA \\
&= \frac{1}{2}C + \frac{1}{2}\left(J^t(A^{-1})^t C\right)^t JA \\
&= \frac{1}{2}C + \left(C^t A^{-1} J\right) JA \\
&= \frac{1}{2}C - \frac{1}{2}C^t \\
&= C.
\end{aligned}$$

Here we have used the fact that $J^t = -J = J^{-1}$ in passing from line 2 to 3, the fact that $J^2 = -I$ in passing from 4 to 5, and the fact that $C = \frac{1}{2}C - \frac{1}{2}C^t$ in passing from line 5 to 6 (since C is skew-symmetric). This shows that J is a regular value for f, and consequently, $f^{-1}(J) = Sp(2n)$ is a manifold of dimension $\dim(\mathbb{R}^{2n\times 2n}) - \dim(Sk_{2n}(\mathbb{R})) = 4n^2 - (2n^2 - n) = 2n^2 + n$.

In particular, we observe that since J is a regular value of f, then

$$\mathfrak{sp}(2n) = \mathrm{Lie}(Sp(2n)) \cong T_I(Sp(2n)) = \ker(df_I) = \{A \in \mathbb{R}^{2n\times 2n} \mid JA + A^t J = 0\}$$

by equation (B.19). □

Finally, we make a few observations concerning the possible eigenvalues of symplectic automorphisms. By Theorem B.1.7, if $T\colon V \to V$ is a symplectic automorphism of a symplectic vector space (V, ω), then there exists some basis for V such that the matrix A of T, when written with respect to this basis, satisfies $A^t J A = J$. Since A and J are both invertible, this is equivalent to $JAJ^{-1} = B$, where $B = (A^t)^{-1} = (A^{-1})^t$. Now the eigenvalues of T are the roots of the characteristic polynomial $P_A(\lambda)$, which we can write as $P_A(\lambda) = \det(\lambda I - A)$. Since J has determinant 1, and since $J^t = J^{-1} = -J$, we see that each of these matrices also has determinant 1 (since $\det(M^t) = \det(M)$ for any M). Thus, the multiplicativity of the determinant enables us to compute:

$$\begin{aligned}
P_A(\lambda) &= \det(\lambda I - A) \\
&= \det(J(\lambda I - A)J^{-1}) \\
&= \det(\lambda I - B) \\
&= \det(\lambda I - (A^{-1})^t) \\
&= \det((\lambda I - A^{-1})^t) \\
&= \det(\lambda I - A^{-1}) \\
&= \det(A^{-1}(\lambda A - I)) \\
&= \det(\lambda A - I) \\
&= \det\left(\lambda\left(A - \frac{1}{\lambda} I\right)\right) \\
&= \lambda^{2n} \det\left(A - \frac{1}{\lambda} I\right) \\
&= \lambda^{2n} P_A\left(\frac{1}{\lambda}\right).
\end{aligned} \tag{B.20}$$

Here we have used the multiplicativity of the determinant together with the fact that $\det(A^{-1}) = \frac{1}{\det(A)} = 1$ (since $\det(A) = 1$) in passing from line 7 to 8, and the fact that $\lambda \neq 0$ for any eigenvalue λ of A (since A is invertible) in passing from line 8 to 9. Furthermore, together with the final line of (B.20), this also shows that $\frac{1}{\lambda}$ is an eigenvalue of A whenever λ is. Similarly, since P_A has real coefficients, then its roots come in complex conjugate pairs, and therefore $P_A(\bar{\lambda}) = 0$ whenever $P(\lambda) = 0$. Putting all this together, we see that if λ is an eigenvalue of a symplectic linear map, then so are $\bar{\lambda}$, $1/\lambda$, and $1/\bar{\lambda}$.

Now suppose that λ_0 is an eigenvalue of A of multiplicity k. Then the characteristic polynomial P_A of A factors as $P_A(\lambda) = (\lambda - \lambda_0)^k Q(\lambda)$ for some other polynomial $Q(\lambda)$. By the final line of (B.20), this implies that

$$\lambda^{2n} P\left(\frac{1}{\lambda}\right) = (\lambda - \lambda_0)^k Q(\lambda) = (\lambda\lambda_0)^k \left(\frac{1}{\lambda_0} - \frac{1}{\lambda}\right)^k Q(\lambda),$$

so that

$$P\left(\frac{1}{\lambda}\right) = \left(\frac{1}{\lambda_0} - \frac{1}{\lambda}\right)^k \frac{\lambda_0^k}{\lambda^{2n-k}} Q(\lambda). \tag{B.21}$$

But $Q(\lambda)$ is a polynomial of degree $2n - k$ since λ_0 has multiplicity k by assumption, and therefore

$$\frac{\lambda_0^k}{\lambda^{2n-k}} Q(\lambda)$$

is a polynomial in $\frac{1}{\lambda}$. Thus, the right-hand side of (B.21) is a polynomial factorization of $P_A\left(\frac{1}{\lambda}\right)$ showing that $\frac{1}{\lambda_0}$ occurs as a root of P_A with multiplicity $l \geq k$. Reversing the roles of λ_0 and $1/\lambda_0$, we obtain the reverse inequality $l \leq k$, and therefore $l = k$. Finally, we observe that since $P_A(\lambda)$ has real coefficients, then the multiplicities of complex conjugate roots are equal. We encapsulate these facts in the following theorem.

Theorem B.1.8 (Symplectic eigenvalue theoerm). *Let (V, ω) be a symplectic vector space and $T \colon V \to V$ be a symplectic automorphism of V. If λ is an eigenvalue of multiplicity k, then $\bar{\lambda}$, $1/\lambda$, and $1/\bar{\lambda}$ are also eigenvalues of multiplicity k.*

B.2 Symplectic manifolds

Definition B.2.1 (Nondegenerate 2-forms). Let M be a manifold. A smooth 2-form $\omega \in \Omega^2(M)$ on M is called *nondegenerate* if $\omega_m \colon T_m(M) \times T_m(M) \to \mathbb{R}$ is a nondegenerate skew-symmetric bilinear form for each $m \in M$.

Theorem B.2.1. *Let M be a smooth manifold and $\omega \in \Omega^2(M)$ a nondegenerate 2-form on M. Then ω induces an isomorphism*

$$\omega_\flat \colon \mathfrak{X}(M) \to \Omega^1(M)$$

$$\omega_\flat(X) = i_X \omega$$

from the set of smooth vector fields on M to the set of smooth 1-forms on M via contraction.

Proof. This is an immediate consequence of Theorem B.1.1, since

$$(\omega_\flat)_m \colon T_m(M) \to (T_m(M))^* = \omega_m(X_m, \cdot)$$

acts as the flat map on each tangent space $T_m(M)$. □

Definition B.2.2 (Flat map, sharp map). As before, the map $\omega_\flat$ is referred to as the *flat map*, and its inverse $\omega_\flat^{-1}\colon \Omega^1(M) \to \mathfrak{X}(M)$, denoted $\omega_\sharp$, is called the *sharp map*.

Definition B.2.3 (Symplectic form). Let M be a smooth manifold. A *symplectic form* on M is a closed, nondegenerate 2-form $\omega \in \Omega^2(M)$.

Definition B.2.4 (Symplectic manifold). A *symplectic manifold* (M, ω) is a pair consisting of a smooth manifold M and a symplectic form $\omega \in \Omega^2(M)$.

Definition B.2.5 (Symplectic map). Let (M, ω) and (N, ρ) be two symplectic manifolds. A smooth map $\varphi\colon M \to N$ is called *symplectic* if $\varphi^*\rho = \omega$.

Definition B.2.6 (Symplectomorphism). Let (M, ω) and (N, ρ) be two symplectic manifolds. A symplectic diffeomorphism $\varphi\colon M \to N$ is called a *symplectomorphism*.

By virtue of these definitions and Corollary B.1.5, we have the following results.

Theorem B.2.2. *If (M, ω) and (N, ρ) are symplectic manifolds of the same dimension and $\varphi\colon M \to N$ is a symplectic map, then φ is a volume preserving local diffeomorphism.*

Theorem B.2.3. *Let (M, ω) be a symplectic manifold and $\varphi\colon M \to N$ a diffeomorphism. Then $(N, (\varphi^{-1})^*\omega)$ is a symplectic manifold, and φ is a symplectomorphism.*

Theorem B.2.4. *Let (M, ω) and (N, ρ) be symplectic manifolds, $f\colon M \to N$ a smooth map, and $\varphi\colon M \to M'$ and $\psi\colon N \to N'$ diffeomorphisms.*

$$\begin{array}{ccc} M & \xrightarrow{\quad f \quad} & N \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle\psi} \\ M' & \xrightarrow[\psi\circ f\circ\varphi^{-1}]{} & N' \end{array}$$

Then f is symplectic if and only if $\psi \circ f \circ \varphi^{-1}$ is a symplectic map between $(M', (\varphi^{-1})^*\omega)$ and $(N', (\psi^{-1})^*\rho)$.

Proof. Suppose that f is symplectic. Then

$$\begin{aligned} (\psi \circ f \circ \varphi^{-1})^*((\psi^{-1})^*\rho) &= ((\varphi^{-1})^* \circ f^* \circ \psi^*)(\psi^{-1})^*\rho \\ &= (\varphi^{-1})^* f^*\rho \\ &= (\varphi^{-1})^*\omega \end{aligned}$$

so that $\psi \circ f \circ \varphi^{-1}$ is a sympletic map between $(M', (\varphi^{-1})^*\omega)$ and $(N', (\psi^{-1})^*\rho)$. Conversely, if $\psi \circ f \circ \varphi^{-1}$ is symplectic, then

$$\begin{aligned} f^*\rho &= (\psi^{-1} \circ (\psi \circ f \circ \varphi^{-1}) \circ \varphi)^*\rho \\ &= \varphi^* \circ (\psi \circ f \circ \varphi^{-1})^* \circ (\psi^{-1})^*\rho \\ &= \varphi^*((\varphi^{-1})^*\omega) \\ &= \omega. \end{aligned}$$

□

B.2.1 Darboux's Theorem

Our immediate goal is to prove Darboux's Theorem, one of the fundamental results in symplectic geometry, which asserts the existence of a special kind of coordinate chart on a manifold equipped with a symplectic form.

Theorem B.2.5 (Darboux's Theorem). *Let M be a manifold and $\omega \in \Omega^2(M)$ a symplectic form on M. Then for each $m \in M$, there exists a coordinate chart (U, φ) about m with coordinate functions $x_1, \ldots, x_n, y_1, \ldots, y_n$ such that $\varphi(m) = 0$ and*

$$\omega|_U = \sum_{i=1}^{n} dx_i \wedge dy_i. \tag{B.22}$$

Proof. We first prove the result on $\mathbb{R}^{2n}$. To that end, let $\rho \in \Omega^2(\mathbb{R}^{2n})$ be a symplectic form. First, we observe that since ρ is a nondegenerate 2-form on $\mathbb{R}^{2n}$, ρ_0 is a nondegenerate skew-symmetic bilinear form on the tangent space $T_0(\mathbb{R}^{2n})$; consequently, Theorem B.1.4 gives a basis for $T_0(\mathbb{R}^{2n})$ for which ρ_0 is in the standard form (B.8). Regarding these basis vectors as the coordinate vectors corresponding to a system of linear coordinates on $\mathbb{R}^{2n}$, we thereby obtain a (globally defined!) linear coordinate system $x_1, \ldots, x_n, y_1, \ldots, y_n$ for which

$$\rho_0 = \sum_{i=1}^{n} dx_i|_0 \wedge dx_i|_0.$$

Now define the symplectic form

$$\omega = \sum_{i=1}^{n} dx_i \wedge dy_i \in \Omega^2(\mathbb{R}^{2n}). \tag{B.23}$$

We wish to show that there exists a local diffeomorphism $\psi\colon U \to V$, where $0 \in U, V \subseteq \mathbb{R}^{2n}$, $\psi(0) = 0$, and with the property that

$$\psi^*\omega = \rho. \tag{B.24}$$

We will show this using an argument known as *Moser's trick*: we will construct a deformation of ρ into ω, use this to induce a time-varying vector field X_t on $\mathbb{R}^{2n}$, and then show that the resulting (diffeomorphism that is the) flow along X_t from $t = 0$ to $t = 1$ is actually the diffeomorphism that we seek.

Define the two-form

$$\eta_t = t\omega + (1 - t)\rho \tag{B.25}$$

for $t \in [0, 1]$. Since $\omega_0 = \rho_0$ by our choice of coordinates, $\eta_0^t = \omega_0 = \rho_0$ for all $t \in [0, 1]$. Since nondegeneracy of two-forms is an open condition (it corresponds to the invertibility of the matrix W in (B.3)), the fact that η_t is nondegenerate at 0 for all $t \in [0, 1]$ implies that there is an open set $O_1 \subseteq [0, 1] \times \mathbb{R}^{2n}$ containing $[0, 1] \times \{0\}$ on which η_t is nondegenerate; equivalently, there exists an open set $O_2 \subseteq \mathbb{R}^{2n}$ on which η_t is nondegenerate for all $t \in [0, 1]$. Thus, when restricted to O_2, η_t gives a straight-line homotopy between the initial symplectic form ρ and the target standard symplectic form ω through a family of sympletic forms.

Let X_t denote our target time-varying vector field, and let ψ_t denote the transformations in the local 1-parameter group of X_t (i.e., ψ_t denotes the map sending each point to its time-t image when flowed along X_t, provided that this exists). We shall show that X_t can be uniquely well-defined by requiring that

$$\psi_t^* \eta_t = \rho. \tag{B.26}$$

Differentiating (B.26) with respect to time shows that

$$\begin{aligned} 0 &= \frac{d}{dt} \left[\psi_t^* \eta_t \right] \\ &= \psi_t^* \left(L_{X_t} \eta_t + \frac{d}{dt} \eta_t \right) \\ &= \psi_t^* \left(L_{X_t} \eta_t + \omega - \rho \right). \end{aligned} \tag{B.27}$$

Now ρ is closed by hypothesis, and ω is obviously closed by definition (B.23). Consequently, using Cartan's formula

$$L_{X_t} = i_{X_t} \circ d + d \circ i_{X_t}$$

shows that (B.27) can be reduced to

$$0 = \psi_t^* \left(d(i_{X_t} \eta_t) + \omega - \rho \right). \tag{B.28}$$

Now since ρ and ω are closed, they are also locally exact by the Poincaré Lemma. Thus, there exists some open set $E \subseteq O_2$ containing 0 and some smooth 1-form α such that $\rho - \omega = d\alpha$ on E; furthermore, we may (by adding a constant 1-form if necessary) assume without loss of generality that $\alpha_0 = 0$. Consequently, (B.28) can be rewritten as

$$0 = \psi_t^* \left(d(i_{X_t} \eta_t - \alpha) \right). \tag{B.29}$$

on E. Thus, we see that in order for (B.26) to hold, it suffices by (B.29) to require that

$$i_{X_t} \eta_t = \alpha. \tag{B.30}$$

But since η_t is nondegenerate on E, then we can obtain X_t by definining

$$X_t = (\eta_t)^\sharp(\alpha) \tag{B.31}$$

on E for all $t \in [0,1]$ by Theorem B.2.1.

Now we need only make a few observations. Let C be a compact subset of (the open set) E with nonempty interior that contains 0. Then $[0,1] \times C$ is compact, therefore X_t is bounded on this set. Consequently, by taking a sufficiently small open set $U \subset C$ containing 0, we can be assured that the flow of U along X_t is contained in E for all $[0,1]$ (i.e., we can be assured of its existence), and this flow map is a diffeomorphism for each $t \in [0,1]$. Consequently, by restricting the flow map ψ_t to this small open subset U, we thereby obtain a diffeomorphism

$$\psi_1|_U : U \to V$$

where $V = \psi_1(U)$ is the image of U along this flow. Moreover, since $\alpha_0 = 0$, then (B.31) shows that $X_t(0) = 0$ identically, so that 0 is a fixed point of the flow, and therefore $\psi_1(0) = 0$. Thus, by defining $\psi = \psi_1$ we see from equations (B.25) and (B.26) that (B.24) is satisfied, as desired. This completes the proof in the case of $\mathbb{R}^{2n}$.

The general case follows immediately from the previous case. Given any chart $\varphi\colon U \to V \subseteq \mathbb{R}^{2n}$ on a symplectic manifold (M, ω) with $\varphi(m) = 0$, simply construct $\psi\colon U' \to V'$ as above. Then the restriction of the composition of these charts to a suitably small neighborhood around m gives a chart on which ω has the required form. □

Definition B.2.7 (Symplectic chart, canonical coordinates). Given a symplectic manifold (M, ω) and some $m \in M$, a coordinate chart $\varphi\colon U \to \mathbb{R}^{2n}$ with coordinate functions $x_1, \ldots, x_n, y_1, \ldots, y_n$ for which

$$\omega = \sum_{i=1}^{n} dx_i \wedge dy_i$$

on U (whose existence is guaranteed by Darboux's Theorem) is called a *symplectic chart*, and the coordinates $x_1, \ldots, x_n, y_1, \ldots, y_n$ are called *canonical coordinates*.

Thus, given any two symplectic manifolds (M, ω) and (N, ρ) of dimension $\mathbb{R}^{2n}$ with $m \in M$ and $n \in N$, one can take a symplectic chart $\varphi\colon U \to \mathbb{R}^{2n}$ about m with $\varphi(m) = 0$ and a symplectic chart $\psi\colon V \to \mathbb{R}^{2n}$ about n with $\psi(n) = 0$, and the composition $\psi^{-1} \circ \varphi$ gives a local symplectomorphism from a neighborhood of m to a neighborhood of n with $\psi^{-1} \circ \varphi(m) = n$. This proves the following corollary.

Corollary B.2.6. *All symplectic manifolds of the same dimension are locally symplectomorphic.*

Remark B.2.1. Corollary B.2.6 can be interpreted as stating that the local geometry of symplectic manifolds is trivial (i.e., all of the interesting properties of a symplectic manifold are due to its *global* geometry and topology, not its local structure). This is in marked contrast, for example, to the case of Riemannian geometry, in which the curvature provides an obstruction to the existence of local isometries.

B.2.2 The cotangent bundle

Thus far, we have established some results on properties of symplectic manifolds without explicitly mentioning any. As it turns out, the cotangent bundle of any manifold can be equipped with a symplectic structure.

Definition B.2.8 (Natural coordinates on the tangent and cotangent bundles). Let M be a smooth n-manifold and (U, φ) a chart on M with coordinate functions $x_1, \ldots, x_n$. To each such chart on M, we can associate a chart $(\pi^{-1}(U), \hat{\varphi})$ on the tangent bundle TM of M, where $\pi\colon TM \to M$ denotes the canonical projection map, and $\hat{\varphi}$ is defined by

$$\hat{\varphi}\colon \pi^{-1}(U) \to \mathbb{R}^{2n}$$
$$\hat{\varphi}(v) = (x_1(\pi(v)),\ \ldots, x_n(\pi(v)),\ dx_1(v),\ \ldots,\ dx_n(v)).$$

Similarly, we obtain a chart $(\pi^{-1}(U), \tilde{\varphi})$ on the cotangent bundle T^*M of M, where $\pi\colon T^*M \to M$ denotes the canonical projection map, and $\tilde{\varphi}$ is defined by

$$\tilde{\varphi}\colon (\pi)^{-1}(U) \to M$$
$$\tilde{\varphi}(\tau) = \left(x_1(\pi(\tau)),\ \ldots, x_n(\pi(\tau)),\ \tau\left(\frac{\partial}{\partial x_1}\right),\ \tau\left(\frac{\partial}{\partial x_n}\right)\right).$$

We shall call such charts *natural charts*, and the corresponding set of coordinate functions $(q_1(\tau), \ldots, q_n(\tau), p_1(\tau), \ldots, p_n(\tau))$, *natural coordinates*.

Theorem B.2.7 (Canonical forms on the cotangent bundle). *Let B be an n-manifold, $M = T^*B$ its cotangent bundle, $\pi\colon M \to B$ the standard bundle projection map, and $d\pi\colon TM \to TB$ the induced map from the tangent bundle of M to the tangent bundle of B, all defined in the usual way. Define a 1-form $\Theta \in \Omega^1(M)$ according to*

$$\Theta_\alpha(v) = \alpha(d\pi_\alpha(v)) \tag{B.32}$$

for all $\alpha \in M$ and all $v \in T_\alpha(M)$. Then $\Omega = -d\Theta$ is a symplectic form on M, and given a set of local coordinates $x_1, \ldots, x_n$ on a neighborhood U in the base manifold B with corresponding natural coordinates $(q_1, \ldots, q_n, p_1, \ldots, p_n)$ on $\pi^{-1}(U) \subseteq M$, Θ and Ω can be written locally as

$$\Theta = \sum_{i=1}^{n} p_i \, dq_i,$$
$$\Omega = \sum_{i=1}^{n} dq_i \wedge dp_i.$$

Proof. The proof is straightforward; one just has to work through the definitions. First, observe that since $d\pi_\alpha\colon T_\alpha(M) \to T_{\pi(\alpha)}(B)$, then $d\pi_\alpha(v) \in T_{\pi(\alpha)}(B)$. Since $\alpha \in (T_{\pi(\alpha)}(B))^*$, the assignment defined in equation (B.32) makes sense, and is smooth since all of the maps in (B.32) are smooth; this shows that $\Theta \in \Omega^1(M)$.

Now let $\alpha \in \pi^{-1}(U)$ and $v \in T_\alpha(M)$. Then v can be written uniquely as

$$v = \sum_{i=1}^{n} \beta_i \frac{\partial}{\partial q_i} + \gamma_i \frac{\partial}{\partial p_i}, \tag{B.33}$$

and by definition of the natural coordinates on $\pi^{-1}(U)$, we have

$$d\pi_\alpha(v) = d\pi_\alpha \left(\sum_{i=1}^{n} \beta_i \frac{\partial}{\partial q_i} + \gamma_i \frac{\partial}{\partial p_i} \right) = \sum_{i=1}^{n} \beta_i \frac{\partial}{\partial x_i}.$$

Thus,

$$\begin{aligned} \Theta_\alpha(v) &= \alpha(d\pi_\alpha(v)) \\ &= \alpha \left(\sum_{i=1}^{n} \beta_i \frac{\partial}{\partial x_i} \right) \\ &= \sum_{i=1}^{n} \beta_i \, \alpha \left(\frac{\partial}{\partial x_i} \right) \\ &= \sum_{i=1}^{n} \beta_i \, p_i(\alpha) \\ &= \left(\sum_{i=1}^{n} p_i(\alpha) \, dq_i \right)(v) \end{aligned} \tag{B.34}$$

by (B.33). Thus, equation (B.34) shows that

$$\Theta|_{T^*U} = \sum_{i=1}^{n} p_i \, dq_i,$$

and consequently,

$$\begin{aligned} \Omega &= -d\Theta \\ &= -\sum_{i=1}^{n} dp_i \wedge dq_i \\ &= \sum_{i=1}^{n} dq_i \wedge dp_i, \end{aligned}$$

showing that Ω is a symplectic form on M. □

Corollary B.2.8. *Let M be a smooth manifold and T^*M its cotangent bundle. Then T^*M is a symplectic manifold when equipped with the form $\Omega \in \Omega^2(T^*M)$ from Theorem* B.2.7*, and in that case the natural charts on T^*M induced by the charts on M are symplectic charts.*

Definition B.2.9 (Canonical forms). The forms $\Theta \in \Omega^1(T^*M)$ and $\Omega \in \Omega^2(T^*M)$ in Theorem B.2.7 are called the *canonical* 1- *and* 2-*form*, respectively.

The canonical 1-form on the cotangent bundle can be characterized by the following result.

Theorem B.2.9. *Let M be a smooth manifold and T^*M its cotangent bundle. The canonical* 1*-form Θ on T^*M is the unique* 1*-form on T^*M with the property that*

$$\alpha^*\Theta = \alpha \tag{B.35}$$

for all $\alpha \in \Omega^1(M)$ (note that here we are quite explicitly using the fact each 1*-form $\alpha \in \Omega^1(M)$ is a section $\alpha\colon M \to T^*M$).*

Proof. By definition, for any $\alpha \in \Omega^1(M)$, $m \in M$, and $v \in T_m(M)$:

$$\begin{aligned}
(\alpha^*\Theta)_m(v) &= \Theta_{\alpha_m}(d(\alpha)_m(v)) \\
&= \alpha_m(d\pi_{\alpha_m}(d(\alpha)_m(v))) \\
&= \alpha_m(d(\pi \circ \alpha)_m(v)) \\
&= \alpha_m(v),
\end{aligned}$$

(here we are using the fact that α is a section of the cotangent bundle T^*M in passing from line 3 to 4). This proves that Θ satisfies (B.35).

To see that Θ is the *unique* 1-form on T^*M with this property, it suffices to observe that $\alpha(m)$ and $d(\alpha)_m(v)$ span all of $(T_m(M))^*$ and $(T_{\alpha(m)}(T^*M))^*$ over all choices of $\alpha \in \Omega^1(M)$ and $v \in T_m(M)$. $\square$

Remark B.2.2. In light of Theorem B.2.9, the canonical 1-form Θ is also referred to as the *tautological* 1*-form*.

Theorem B.2.10 (Pullback maps on the cotangent bundle)*. Let* $\varphi\colon M \to N$ *be a diffeomorphism between two smooth manifolds* M *and* N*. Then the induced pullback mapping*

$$\varphi^*\colon T^*N \to T^*M \tag{B.36}$$

is a fiber-preserving diffeomorphism of these cotangent bundles fitting into the commutative diagram

$$\begin{array}{ccc} T^*M & \xleftarrow{\ \varphi^*\ } & T^*N \\ \downarrow{\scriptstyle \pi_M} & & \downarrow{\scriptstyle \pi_N} \\ M & \xrightarrow{\ \varphi\ } & N \end{array} \tag{B.37}$$

where here $\pi_M\colon T^*M \to M$ *and* $\pi_N\colon T^*N \to N$ *denote the canonical projection maps on the cotangent bundles.*

Furthermore, if Θ_M and Θ_N denote the canonical 1-forms and Ω_M and Ω_N the canonical 2-forms on T^*M and T^*N, respectively, then

$$(\varphi^*)^*\Theta_M = \Theta_N$$

and consequently,

$$(\varphi^*)^*\Omega_M = \Omega_N.$$

Thus, a diffeomorphism $\varphi\colon M \to N$ between two smooth manifolds M and N induces a corresponding fiber-preserving symplectomorphism $\varphi^*\colon T^*N \to T^*M$ on their cotangent bundles.

Proof. Since $\varphi\colon M \to N$ is a diffeomorphism, the induced derivative map $d\varphi_m\colon T_m(M) \to T_{\varphi(m)}(N)$ is an isomorphism for all $m \in M$. Consequently, the dual map defined by

$$(d\varphi_m)^*\colon (T_{\varphi(m)}(N))^* \to (T_m(M))^*$$
$$(d\varphi_m)^*(\alpha)(v) = \alpha(d\varphi_m(v))$$

is likewise an isomorphism between each pair of cotangent spaces. Since this is precisely the restriction

$$\varphi^*\big|_{(T^*N)_{\varphi(m)}}\colon (T_{\varphi(m)}(N))^* \to (T_m(M))^*$$

of the pullback mapping φ^* to the fiber $(T^*N)_{\varphi(m)} = \pi_N^{-1}(\varphi(m))$, this shows that φ^* is a fiber-preserving diffeomorphism fitting into the diagram (B.37).

Now let Θ_M and Θ_N denote the canonical 1-forms on T^*M and T^*N. Fix $\alpha_n \in (T^*N)_n = (T_n(N))^*$ and let $v \in T_{\alpha_n}(T^*N)$. Then by definition,

$$\begin{aligned}
(\varphi^*)^*\Theta_M(v) &= (\Theta_M)_{\varphi^*(\alpha_n)}(d(\varphi^*)(v)) \\
&= \varphi^*(\alpha_n)\left(d\pi_M(d(\varphi^*)(v))\right) \\
&= \alpha_n\left(d\varphi\left(d\pi_M(d(\varphi^*)(v))\right)\right) \\
&= \alpha_n\left(\left(d\varphi \circ d\pi_M \circ d(\varphi^*)\right)(v)\right) \\
&= \alpha_n\left(d\left(\varphi \circ \pi_M \circ \varphi^*\right)(v)\right).
\end{aligned} \tag{B.38}$$

Now the commutativity of (B.37) implies that

$$\varphi \circ \pi_M \circ \varphi^* = \pi_N,$$

and therefore equation (B.38) reduces to:

$$\begin{aligned}
(\varphi^*)^*\Theta_M(v) &= \alpha_n\left(d\left(\varphi \circ \pi_M \circ \varphi^*\right)(v)\right) \\
&= \alpha_n\left(d\pi_N(v)\right) \\
&= \Theta_N(v).
\end{aligned}$$

Thus, $(\varphi^*)^*\Theta_M = \Theta_N$, and consequently

$$\begin{aligned}
(\varphi^*)^*\Omega_M &= (\varphi^*)^*(-d\Theta_M) \\
&= -d\left((\varphi^*)^*\Theta_M\right) \\
&= -d\Theta_N \\
&= \Omega_N,
\end{aligned}$$

so that φ^* is a symplectomorphism, as claimed. □

B.3 Poisson structures

So far, our coverage of symplectic geometry has (unsurprisingly) been focused on the investigation of symplectic forms. However, there is a closely related kind of structure that can be placed on smooth manifolds, called a *Poisson structure*.

B.3.1 Poisson manifolds

Definition B.3.1 (Poisson bracket, Poisson manifold)*.* Let P be a smooth manifold and let $C^\infty(P)$ denote the algebra of smooth real-valued functions on P. A *Poisson bracket* is a binary operation

$$\{\cdot,\cdot\}\colon C^\infty(P)\times C^\infty(P)\to C^\infty(P)$$

with the following properties:

1. Bilinearity.
2. Skew-symmetry.
3. The Jacobi identity:
$$\{\{f,g\},h\}+\{\{h,f\},g\}+\{\{g,h\},f\}=0.$$
4. The Leibniz rule:
$$\{fg,h\}=f\{g,h\}+g\{f,h\}.$$

A pair $(P, \{,\})$ consisting of a smooth manifold P together with a Poisson bracket $\{\cdot,\cdot\}$ is called a *Poisson manifold*.

Observe that properties 1, 2, and 3 show that $\{\cdot,\cdot\}$ makes $C^\infty(P)$ into a Lie algebra. Furthermore, property 4 shows that, for a fixed $g \in C^\infty(P)$, the mapping

$$f \mapsto \{f, g\} \tag{B.39}$$

is a derivation on $C^\infty(P)$. Via the identification of the tangent space $T_p(P)$ with the set of derivations on the equivalence class $\tilde{F}_p$ of germs at p [14, p. 12], we see that the assignment (B.39) defines a (uniquely determined!) *vector field* on P.

Definition B.3.2 (Hamiltonian vector field). Let $(P, \{,\})$ be a Poisson manifold. Given a function $H \in C^\infty(P)$, we define the *Hamiltonian vector field* X_H to be the unique vector field $X_H \in \mathfrak{X}(P)$ such that

$$X_H(f) = \{f, H\} \tag{B.40}$$

for all $f \in \mathcal{F}$.

Theorem B.3.1. *Let $(P, \{\})$ be a Poisson manifold. Let $H \in C^\infty(P)$, fix $p \in P$, let $\varphi(t)$ denote the integral curve of the Hamiltonian vector field X_H satisfying $\varphi(0) = p$, and given any $f \in C^\infty(P)$, let $F = f \circ \varphi(t)$ denote evaluation of the function f along the trajectory $\varphi(t)$. Then*

$$\dot{F}(t) = \{f, H\}(\varphi(t)).$$

Proof. Let $\varphi\colon (-\epsilon, \epsilon) \to P$ denote the integral curve along X_H through p, where $\varphi(0) = p$. Then for $t_0 \in (-\epsilon, \epsilon)$,

$$\begin{aligned}
\dot{F}(t_0) &= \frac{d}{dt}\left[f(\varphi(t))\right]\big|_{t_0} \\
&= df_{\varphi(t_0)}\left(d\varphi_{t_0}\left(\frac{\partial}{\partial t}\right)\right) \\
&= df_{\varphi_{t_0}}\left(X_H(\varphi(t_0))\right) \\
&= (X_H)_{\varphi(t_0)}(f) \\
&= \{f, H\}(\varphi(t_0)).
\end{aligned}$$

□

Definition B.3.3 (Poisson map). Let $(P_1, \{,\}_1)$ and $(P_2, \{,\}_2)$ be two Poisson manifolds and $\varphi\colon P_1 \to P_2$ a smooth map between them. Then φ is called *Poisson* if it preserves the Poisson bracket operations: for all $f, g \in C^\infty(P_2)$

$$\{f, g\}_2 \circ \varphi = \{f \circ \varphi, g \circ \varphi\}_1.$$

One important observation is that the local 1-parameter groups corresponding to flows along Hamiltonian vector fields consist of Poisson diffeomorphisms.

Theorem B.3.2. *Let $(P, \{,\})$ be a Poisson manifold, $H \in C^\infty(P)$, and let $\varphi\colon (-\epsilon, \epsilon) \times P \to P$ denote the flow along the Hamiltonian vector field X_H, where $\epsilon > 0$ is such that φ_t is well-defined for all $t \in (-\epsilon, \epsilon)$. Then*

$$\varphi_t\colon P \to P$$

$$\varphi_t(p) = \varphi(t, p)$$

is a Poisson map for each $t \in (-\epsilon, \epsilon)$.

Proof. Observe that

$$\begin{aligned}
\frac{d}{dt}\left(\{f,g\}\circ\varphi_t\right) &= d(\{f,g\})\circ\frac{d}{dt}(\varphi_t) \\
&= d(\{f,g\})(X_H) \\
&= X_H(\{f,g\}) \\
&= \{\{f,g\},H\} \\
&= \{f,\{g,H\}\}+\{\{f,H\},g\} \\
&= \left\{f,\frac{d}{dt}(g\circ\varphi_t)\right\}+\left\{\frac{d}{dt}(f\circ\varphi_t),g\right\},
\end{aligned}$$

where line 2 follows from line 1 since φ_t is the flow along X_H, line 4 follows from line 3 by the definition of the Hamiltonian vector field X_H, line 5 follows from line 4 by the Jacobi identity and the skew-symmetry of the Poisson bracket, and line 6 follows from line 5 due to Theorem B.3.1. But

$$\frac{d}{dt}\{f\circ\varphi_t,g\circ\varphi_t\}=\left\{f,\frac{d}{dt}(g\circ\varphi_t)\right\}+\left\{g,\frac{d}{dt}(f\circ\varphi_t)\right\}$$

by the bilinearity of the Poisson bracket, so that

$$\frac{d}{dt}\left(\{f,g\}\circ\varphi_t\right)=\frac{d}{dt}\{f\circ\varphi_t,g\circ\varphi_t\},$$

and therefore

$$\{f\circ\varphi_t,g\circ\varphi_t\}=\{f,g\}\circ\varphi_t.$$

Thus φ_t is Poisson, as claimed. □

More generally, we have the following result.

Theorem B.3.3. *Let $(P_1, \{,\}_1)$ and $(P_2, \{,\}_2)$ be two Poisson manifolds and $\varphi\colon P_1 \to P_2$ a smooth map. Then φ is Poisson if and only if, for all $H \in C^\infty(P_2)$, the Hamiltonian vector fields $X_{H\circ\varphi} \in \mathfrak{X}(P_1)$ and $X_H \in \mathfrak{X}(P_2)$ are φ-related:*

$$d\varphi \circ X_{H\circ\varphi} = X_H \circ \varphi. \tag{B.41}$$

Proof. First, suppose that φ is Poisson. Then by definition, for any function $f \in C^\infty(P_2)$,

$$\{f \circ \varphi, H \circ \varphi\}_1 = \{f, H\}_2 \circ \varphi.$$

Now

$$\{f, H\}_2 \circ \varphi = (X_H \circ \varphi)(f) = df(X_H \circ \varphi) \tag{B.42}$$

and

$$\begin{aligned} \{f \circ \varphi, H \circ \varphi\}_1 &= X_{H\circ\varphi}(f \circ \varphi) \\ &= d(f \circ \varphi)(X_{H\circ\varphi}) \\ &= df \circ (d\varphi \circ X_{H\circ\varphi}) \end{aligned} \tag{B.43}$$

by definition of the Hamiltonian vector fields X_H and $X_{H\circ\varphi}$. Since equations (B.42) and (B.43) must hold for *all* choices of $f \in C^\infty(P_2)$, letting f range over a set of coordinate functions on a neighborhood of each $p \in P$ (whose differentials span the cotangent spaces at each p) shows that (B.41) holds when φ is Poisson.

Conversely, suppose that (B.41) holds for all $H \in C^\infty(P_2)$. Then given any $G \in C^\infty(P_2)$ and any $x \in P_1$,

$$\begin{aligned}\{G \circ \varphi, H \circ \varphi\}_1(x) &= (X_{H\circ\varphi}(G \circ \varphi))(x)\\ &= d(G \circ \varphi)_x(X_{H\circ\varphi})\\ &= dG_{\varphi(x)}(d\varphi_x(X_{H\circ\varphi}))\\ &= dG_{\varphi(x)}(X_H \circ \varphi)\\ &= (X_H(G)) \circ \varphi(x)\\ &= \{G, H\}_2 \circ \varphi(x),\end{aligned}$$

which proves that φ is Poisson. □

Definition B.3.4 (Poisson action). Let $(P, \{,\})$ be a Poisson manifold and G a Lie group. If G acts on P by Poisson maps, the action is called a *Poisson action*.

B.3.2 Poisson structures induced from symplectic structures

B.3.2.1 The Poisson bracket on one-forms

Let (M, ω) be a symplectic manifold. The symplectic form ω on M can be used to define a binary operation on $\Omega^1(M)$, which we also call the *Poisson bracket*.

Definition B.3.5 (Poisson bracket (of 1-forms)). Let (M, ω) be a symplectic manifold. The *Poisson bracket* $\{\cdot,\cdot\}$ *on* $\Omega^1(M)$ is defined as:

$$\{\cdot,\cdot\}\colon \Omega^1(M) \times \Omega^1(M) \to \Omega^1(M)$$
$$\{\theta, \omega\} = -[\theta^\sharp, \omega^\sharp]^\flat,$$

where here $[\cdot,\cdot]$ denotes the Lie bracket of vector fields on M. Thus, we have the following commutative diagram:

$$\begin{array}{ccc} \mathfrak{X}(M)\times\mathfrak{X}(M) & \xrightarrow{-[\cdot,\cdot]} & \mathfrak{X}(M) \\ \Big\downarrow{\scriptstyle \flat\times\flat} & & \Big\downarrow{\scriptstyle \flat} \\ \Omega^1(M)\times\Omega^1(M) & \xrightarrow[\{\cdot,\cdot\}]{} & \Omega^1(M) \end{array} \tag{B.44}$$

The diagram (B.44) immediately implies the following result.

Theorem B.3.4. *On a symplectic manifold* (M,ω)*, the space of* 1*-forms* $\Omega^1(M)$ *becomes a Lie algebra when equipped with the Poisson bracket* $\{\cdot,\cdot\}$*.*

Theorem B.3.5. *Let* (M,ω) *be a symplectic manifold and* $\alpha,\beta\in\Omega^1(M)$*. Then*

$$\{\alpha,\beta\} = -L_{\alpha^\sharp}\beta + L_{\beta^\sharp}\alpha + d(i_{\alpha^\sharp}i_{\beta^\sharp}\omega).$$

Proof. Recall the identity

$$\begin{aligned} d\omega(X,Y,Z) &= L_X(\omega(Y,Z)) + L_Y(\omega(Z,X)) + L_Z(\omega(X,Y)) \\ &\quad - \omega([X,Y],Z) - \omega([Y,Z],X) - \omega([Z,X],Y). \end{aligned}$$

(cf. [14, p. 70]). Writing $X=\alpha^\sharp, Y=\beta^\sharp$, and using the fact that ω is closed and that $\omega(\alpha^\sharp,Z)=\alpha(Z)$ by Theorem B.1.2 produces:

$$\begin{aligned} 0 &= L_{\alpha^\sharp}(\beta(Z)) - L_{\beta^\sharp}(\alpha(Z)) - L_Z(i_{\alpha^\sharp}i_{\beta^\sharp}\omega) \\ &\quad + \{\alpha,\beta\}(Z) + \alpha(L_{\beta^\sharp}(Z)) - \beta(L_{\alpha^\sharp}(Z)). \end{aligned} \tag{B.45}$$

Now applying the identity

$$L_X(\omega)(X_1,\ldots,X_k) = L_X(\omega(X_1,\ldots,X_k)) - \sum_{i=1}^{k}\omega(X_1,\ldots,L_XX_i,\ldots,X_k)$$

(cf. [14, p. 70]) shows that (B.45) can be rewritten as:

$$0 = (L_{\alpha^\sharp}\beta)(Z) - (L_{\beta^\sharp}\alpha)(Z) - d(i_{\alpha^\sharp}i_{\beta^\sharp}\omega)(Z) + \{\alpha, \beta\}(Z).$$

□

B.3.2.2 The Poisson bracket induced from the symplectic form

Definition B.3.6 (The Poisson bracket on a symplectic manifold). Let (M, ω) be a symplectic manifold and $f, g \in C^\infty(M)$. Then we define the corresponding *Poisson bracket* on M by

$$\{f, g\} = \omega((df)^\sharp, (dg)^\sharp). \tag{B.46}$$

Theorem B.3.6. *Let (M, ω) be a symplectic manifold and $H \in C^\infty(M)$. Then the corresponding Hamiltonian vector field X_H on M is given by*

$$X_H = (dH)^\sharp. \tag{B.47}$$

Proof. By definition X_H is the unique vector field for which $X_H = \{f, H\}$. Now

$$df\,(X_H) = X_H(f) = \{f, H\} = \omega((df)^\sharp, (dH)^\sharp) = df(dH^\sharp).$$

Since this equation must hold identically for all $f \in C^\infty(M)$, the claim is proved. □

Note that, although we have defined an operation $\{\cdot, \cdot\}$ in terms of the symplectic form ω and *named* it a "Poisson bracket", we have not yet actually shown that it satisfies the requisite four axioms. For the purposes of this verification, we regard (B.47) as a definition.

Theorem B.3.7. *Let (M, ω) be a symplectic manifold and $f, g \in C^\infty(M)$. Then*

$$\{f, g\} = \omega(X_f, X_g) = -i_{X_f} i_{X_g} \omega = -L_{X_f}(g) = L_{X_g}(f).$$

Proof. By definitions and equation (B.47)

$$\{f, g\} = \omega((df)^\sharp, (dg)^\sharp) = \omega(X_f, X_g) = -\omega(X_g, X_f) = -i_{X_f} i_{X_g} \omega.$$

Similarly,

$$L_{X_g}(f) = X_g(f) = \{f, g\}.$$

This shows that

$$\{f, g\} = -i_{X_f} i_{X_g} \omega = L_{X_g}(f).$$

Swapping X_f and X_g above and using the antisymmetry of ω gives

$$-L_{X_f}(g) = \{f, g\},$$

the final equality. □

Corollary B.3.8. *As defined in this section, for fixed $g \in C^\infty(M)$, the mapping $f \mapsto \{f, g\}$ is a derivation.*

Proof. This is immediate from the fact that $\{\cdot, g\} = L_{X_g}$ by Theorem B.3.7 and the fact that the Lie derivative is a derivation. □

Theorem B.3.9. *Let (M, ω) be a symplectic manifold and $f, g \in C^\infty(M)$. Then $d\{f, g\} = \{df, dg\}$.*

Remark B.3.1. Note that this statement uses the Poisson bracket operation on 1-forms.

Proof. By Theorems B.3.5 and B.3.7 and equation (B.47)

$$\begin{aligned}\{df, dg\} &= -L_{X_f} dg + L_{X_g} df + d(i_{X_f} i_{X_g} \omega) \\ &= d(-L_{X_f} g + L_{X_g} f + i_{X_f} i_{X_g} \omega) \\ &= d\{f, g\} + d\{f, g\} - d\{f, g\} \\ &= d\{f, g\}.\end{aligned}$$

□

Theorem B.3.10. *The real vector space* $C^\infty(M)$*, when equipped with the bracket* $\{\cdot, \cdot\}$*, is a Lie algebra.*

Proof. The bracket operation is clearly bilinear and skew-symmetric by definition, so it remains only to show that it satisfies the Jacobi identity. By Theorem B.3.7:

$$\begin{aligned}\{f, \{g, h\}\} &= -L_{X_f}\{g, h\} = L_{X_f}(L_{X_g} h) = X_f(X_g(h)) \\ \{g, \{h, f\}\} &= -L_{X_g}(\{h, f\}) = -L_{X_g}(L_{X_f} h) = -X_g(X_f(h)) \qquad \text{(B.48)} \\ \{h, \{f, g\}\} &= L_{\{f,g\}} h.\end{aligned}$$

Now, observe that

$$X_{\{f,g\}} = (d\{f, g\})^\sharp = \{df, dg\}^\sharp = -[(df)^\sharp (dg)^\sharp] = -[X_f, X_g].$$

Consequently,

$$L_{\{f,g\}} h = L_{[X_g, X_f]} h = [X_g, X_f](h),$$

so equation (B.48) shows that

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = X_f(X_g(h)) - X_g(X_f(h)) + [X_g, X_f](h),$$

and the right-hand side of the above equation is identically 0 by definition of the Lie bracket on vector fields. □

Remark B.3.2. Corollary B.3.8 and Theorem B.3.10 show that the bracket operation $\{\cdot,\cdot\}$ defined in (B.46) is indeed a Poisson bracket.

Now we establish some useful results about Hamiltonian vector fields on symplectic manifolds.

Theorem B.3.11. *Let (M,ω) and (N,ρ) be symplectic manifolds and $f\colon M \to N$ a diffeomorphism. Then f is symplectic if and only if*

$$df(X_{h\circ f}) = X_h$$

for all $h \in C^\infty(N)$.

Proof. Let $Y \in \mathfrak{X}(M)$, and suppose that f is symplectic. Then

$$\begin{aligned}\rho\left(df(X_{h\circ f}), df(Y)\right) &= f^*\rho(X_{h\circ f}, Y)\\ &= \omega(X_{h\circ f}, Y)\\ &= \omega((d(h\circ f))^\sharp, Y)\\ &= d(h\circ f)(Y) = dh(df(Y))\\ &= \rho\left(X_h, df(Y)\right).\end{aligned} \tag{B.49}$$

Since f is a diffeomorphism, its derivative df maps onto the tangent space to N at every point in the image of f. Consequently, the nondegeneracy of ρ together with equation (B.49) shows that

$$df(X_{h\circ f}) = X_h,$$

as claimed.

Conversely, suppose now that $df(X_{h\circ f}) = X_h$ for all $h \in C^\infty(N)$. Then

$$i_{X_{h\circ f}}\omega = d(h \circ f) = df(dh) = f^*(i_{X_h}\rho).$$

Using the identity

$$f^*(i_{X_h}\rho) = i_{d(f^{-1})(X_h)}(f^*\rho)$$

together with the fact that $df(X_{h\circ f}) = X_h$ by hypothesis shows that

$$i_{X_{h\circ f}}\omega = i_{X_{h\circ f}} f^*\rho. \tag{B.50}$$

Since f is a diffeomorphism (hence induces a bijection $C^\infty(N) \to C^\infty(M)$ by precomposition), any vector X_m for $m \in M$ can be realized as the value at m of a smooth vector field of the form $X_{h\circ f}$ for some suitably chosen $h \in C^\infty(N)$. Consequently, the nondegeneracy of ρ (hence also of $f^*\rho$ since f is a diffeomorphism) and ω shows, via equation (B.50), that $\omega = f^*\rho$. □

Theorem B.3.12. *Let (M, ω) and (N, ρ) be symplectic manifolds and $f : M \to N$ a diffeomorphism. Then f is symplectic if and only if it preserves the induced Poisson brackets of functions:*

$$\{g \circ f, h \circ f\}_M = \{g, h\}_N \circ f$$

for all $g, h \in M$.

Proof. By Theorem B.3.7,

$$\{g, h\}_N \circ f = L_{d(f^{-1})(X_h)}(g \circ f) \tag{B.51}$$

and

$$\{g \circ f, h \circ f\}_M = L_{X_{h \circ f}}(g \circ f). \tag{B.52}$$

By Theorem B.3.11, the right-hand sides of equations (B.51) and (B.52) are equal if and only if f is symplectic. □

Corollary B.3.13. *Let (M, ω) be a sympletic manifold, $H \in C^\infty(M)$, and let $\varphi\colon (-\epsilon, \epsilon) \times M \to M$ denote the flow along the Hamiltonian vector field X_H, where $\epsilon > 0$ is such that φ_t is well-defined for all $t \in (-\epsilon, \epsilon)$. Then*

$$\varphi_t\colon M \to M$$
$$\varphi_t(m) = \varphi(t, m)$$

is a symplectomorphism for each $t \in (-\epsilon, \epsilon)$.

Proof. Immediate from Theorems B.3.2 and B.3.12. □

Theorem B.3.14 (Liouville's Theorem). *Let (M, ω) be a sympletic manifold, $H \in C^\infty(M)$, and let $\varphi\colon (-\epsilon, \epsilon) \times M \to M$ denote the flow along the Hamiltonian vector field X_H, where $\epsilon > 0$ is such that φ_t is well-defined for all $t \in (-\epsilon, \epsilon)$. Then*

$$\varphi_t\colon M \to M$$
$$\varphi_t(m) = \varphi(t, m)$$

is a volume-preserving diffeomorphism.

Proof. Since the volume form on M is a scalar multiple of ω^n, this follows immediately from Corollary B.3.13 together with the naturality of pullbacks. □

Finally, we briefly mention how to compute Poisson brackets and Hamiltonian vector fields in canonical coordinates.

Theorem B.3.15 (Hamiltonian vector fields in canonical coordinates). *Let* (M, ω) *be a symplectic* $2n$*-manifold and* (U, φ) *a symplectic chart. If* $f \in C^\infty(U)$, *then*

$$X_f = \sum_{i=1}^{n} \frac{\partial f}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial f}{\partial q_i} \frac{\partial}{\partial p_i}.$$

Proof. In a symplectic chart,

$$\omega = \sum_{i=1}^{n} dq_i \wedge dp_i$$

and therefore

$$(dq_i)^\sharp = -\frac{\partial}{\partial p_i}, \quad (dp_i)^\sharp = \frac{\partial}{\partial q_i}.$$

Consequently,

$$\begin{aligned} X_f &= (df)^\sharp \\ &= \left(\sum_{i=1}^{n} \frac{\partial f}{\partial q_i} \, dq_i + \frac{\partial f}{\partial p_i} \, dp_i \right)^\sharp \\ &= \sum_{i=1}^{n} \left(-\frac{\partial f}{\partial q_i} \frac{\partial}{\partial p_i} + \frac{\partial f}{\partial p_i} \frac{\partial}{\partial q_i} \right). \end{aligned}$$

□

Theorem B.3.16 (The Poisson bracket in canonical coordinates). *Let* (M, ω) *be a symplectic* $2n$*-manifold and* (U, φ) *a symplectic chart. Then the Poisson bracket* $\{\cdot, \cdot\}$ *can be expressed on* U *with respect to these local coordinates as*

$$\{f, g\} = \sum_{i=1}^{n} \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i}.$$

Proof. Since U is a symplectic chart, then by Theorem B.3.15,

$$\begin{aligned}
\{f, g\} &= \omega(X_f, X_g) \\
&= \left(\sum_{i=1}^{n} dq_i \wedge dp_i\right)\left(\sum_{i=1}^{n} \frac{\partial f}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial f}{\partial q_i}\frac{\partial}{\partial p_i}, \sum_{i=1}^{n} \frac{\partial g}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial g}{\partial q_i}\frac{\partial}{\partial p_i}\right) \\
&= \sum_{i=1}^{n} -\frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i} + \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i}.
\end{aligned}$$

□

Index

Bibliography

[1] Ralph Abraham and Jerrold E. Marsden. *Foundations of Mechanics.* Benjamin-Cummings, 1978.

[2] Rolf Berndt. *An Introduction to Symplectic Geometry.* American Mathematical Society, 2001.

[3] Anthony Bloch, John Baillieul, Peter Crouch, and Jerrold Marsden. *Nonholonomic Mechanics and Control.* Springer-Verlag New York, 2003.

[4] Richard P. Feynman. *The Feynman Lectures on Physics, Vol. 1.* California Institute of Technology, 1963.

[5] Mariano Giaquinta and Stefan Hildebrandt. *Calculus of Variations I.* Springer-Verlag Berlin Heidelberg, 2004.

[6] Victor Guillemin and Alan Pollack. *Differential Topology.* Prentice Hall, 1974.

[7] Jurgen Jost and Xianqing Li-Jost. *Calculus of Variations.* Cambridge University Press, 1998.

[8] Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of Differential Geometry, Vol. 1.* Wiley Classics Library, 1991.

[9] Erwin Kreyszig. *Introductory Functional Analysis with Applications.* John Wiley & Sons, 1978.

[10] Jerrold E. Marsden and Tudor S. Ratiu. *Introduction to Mechanics and Symmetry.* Springer-Verlag, 1994.

[11] E. N. Moore. *Theoretical Mechanics.* John Wiley & Sons, 1983.

[12] Raymond A. Serway and John W. Jewett. *Physics for Scientists and Engineers, 6th Edition, Vol. 1.* Thomson Brooks/Cole, 2004.

[13] D. ter Haar. *Elements of Hamiltonian Mechanics, Second Edition.* Pergamon, 1971.

[14] Frank W. Warner. *Foundations of Differentiable Manifolds and Lie Groups.* Springer, 1983.

Vita

David Matthew Rosen is a native of Torrance, California. After graduating from South High School in 2004, he attended the California Institute of Technology, where he studied mathematics and control systems engineering. In 2008, he received the degree of Bachelor of Science in Mathematics with a minor in Control and Dynamical Systems, and went to Texas to pursue graduate study in mathematics and control at the University of Texas at Austin.

Permanent address: 23001 Carlow Road
Torrance, California 90505

This report was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.