Math 428/609E: Mathematical Classical Mechanics
Lecture Notes

Lior Silberman

Dr. Lior Silberman, UBC Department of Mathematics
lior@Math.UBC.CA, http://www.math.ubc.ca/~lior

Compiled April 3, 2025.

Contents

Administrivia

- See syllabus, especially about problem sets.
- Textbooks
  - Mathematical point of view: [**2, 7, 3**]
  - Physical point of view: [**4, 6**]

0.1. Course plan (subject to revision)

	Physics	Mathematics
1	Kinematics	Coordinates, tangent vectors, implicit and inverse function theorems
2	Newtonian mechanics	ODE, cotangent vectors
3	Lagrangian mechanics	Calculus of variations, convexity, symmetry and conservation laws
4	Angular momentum	The rotation group
5	Hamiltonian mechanics	Manifolds, measures

CHAPTER 1

Review of Newtonian mechanics

1.1. Newton's laws

1.1.1. Mathematics: Elementary kinematics. Equip the vector space $\mathbb{R}^d$ with the standard inner product and associated distance function $|\underline{v}-\underline{v}'| = \sqrt{\sum_{i=1}^d \left(v_i' - v_i'\right)^2}$. *Euclidean space* $\mathbb{E}^d$ is the affine space modeled on $\mathbb{R}^d$, in other words a principal homogenous space for $\mathbb{R}^d$. For any $x \in \mathbb{E}^d$ and $\underline{v} \in \mathbb{R}^d$ we have the translate $x+\underline{v} \in \mathbb{E}^d$ and conversely given $x, x' \in \mathbb{E}^d$ there is a unique *displacement vector* $\underline{v} \in \mathbb{R}^d$ with $x' = x+\underline{v}$ in which case $d(x,x') = |\underline{v}|$ is called the *Euclidean distance* from x to x'.

Let $\gamma\colon I \to \mathbb{E}^d$ denote the path of a particle through $\mathbb{E}^d$. Then $\gamma(t+h)-\gamma(t)$ is a displacement (vector in $\mathbb{R}^d$) and if γ is differentiable we can talk about the *velocity vector* $\dot{\gamma}(t) \in \mathbb{R}^d$. In coordinates if $\gamma(t) = (x_1(t),\ldots,x_d(t))$ with respect to some orthonormal coordinate system then $\dot{\gamma}$ is the vector of derivatives of the functions $x_i(t)$. We can similarly define the *acceleration* $\ddot{\gamma}(t) = (\ddot{x}_i(t))_{i=1}^d$. If we have N particles moving in the space, their joint position is given by a point $\gamma(t) \in \mathbb{E}^{dN}$ and we can similarly talk about the velocity vector$\dot{\gamma}(t)$ and acceleration vector $\ddot{\gamma}(t)$, both in $\mathbb{R}^{dN}$.

See also [**2**, \S2D].

1.1.2. Physics: The second law.

AXIOM 1 (Newton's second law). *There is a function* $F\colon \mathbb{E}^{dN} \times \mathbb{R}^{dN} \times \mathbb{R} \to \left(\mathbb{R}^{dN}\right)^8$ *called the* force *so that the path* $\gamma = (x(t))_{t\in I}$ *is determined by the ODE*

$$M\ddot{x} = F(x,\dot{x};t)\,.$$

REMARK 2 (Physics). The force F represents (1) *interactions* between the particles; (2) any *external forces* on the system; and (3) any *constraint forces.*

REMARK 3 (Mathematics). Writing the differential equation in the form

$$\frac{d}{dt}\begin{pmatrix} x \\ \dot{x} \end{pmatrix} = \begin{pmatrix} \dot{x} \\ F(x,\dot{x};t) \end{pmatrix}$$

we see that is suffices to analyze equations of the form $\dot{y}(t) = F(y;t)$ for $y \in \mathbb{R}^n$.

EXAMPLE 4. Some standard systems include:

(1) The *free particle* $m\ddot{x} = 0$ ("Newton's first law")
(2) The Hookean spring, a.k.a. harmonic oscillator : $m\ddot{x} = -kx$.

(3) The *physical pendulum* $mL\ddot{\theta} = -mg\sin\theta$
(4) Pulley systems

REMARK 5. *Newton's first law* is the statement "there is a way to match physical space with $\mathbb{E}^d$ so that free particles

1.1.3. Mathematics: ODE[1].

DEFINITION 6. An *ordinary differential equation* is a pair (Ω, F) where $\Omega \subset \mathbb{E}^n \times \mathbb{E}^1$ is open, $F\colon \Omega \to \mathbb{R}^n$ is continuous. A *solution* to the differential equation is pair (I, γ) where $I \subset \mathbb{E}^1$ is an interval, and $\gamma \in C^1(I; \mathbb{E}^n)$ is a curve such that for all $t \in I$ we have $(\gamma(t), t) \in \Omega$ and -

$$\dot{\gamma}(t) = F(\gamma(t); t) .$$

DEFINITION 7. We say that F is *locally Lipschitz* if for every $(y_0, t_0) \in \Omega$ there is a number $L > 0$ and a neighbourhood $(y_0, t_0) \in U \subset \Omega$ such that for all $(y,t), (y',t) \in U$ we have $|F(y,t) - F(y',t)| \leq L|y - y'|$.

EXERCISE 8. If $F \in C^1$ it is locally Lipschitz.

Fix an ODE (Ω, F). The following argument is called "Picard iteration".

PROPOSITION 9 (Picard existence and uniqueness). *Suppose that* F locally Lipschi tz. Then *for every* initial condition $(y_0, t_0) \in \Omega$ *there is* $\varepsilon > 0$ *so that the equation has a unique solution on* $(t_0 - \varepsilon, t_0 + \varepsilon)$.

LEMMA 10. *There are* $\varepsilon, R, L, M > 0$ *such that:*
(1) $B = B(y_0, R) \times [t_0 - \varepsilon, t_0 + \varepsilon] \subset \Omega$.
(2) *For all* $(y,t) \in B$ *we have* $|F| \leq M$.
(3) *For all* $(y,t), (y',t) \in B$ *we have* $|F(y,t) - F(y',t)| \leq L|y - y'|$.
(4) *We have* $\varepsilon \leq \frac{R}{M+1}$ *and* $\varepsilon \leq \frac{1}{L+1}$.

PROOF. Let U be a neighbourhood of (y_0, t_0) on which we have a Lipschitz constant L. Since U is open, we can choose R and $\tilde{\varepsilon}$ are small enough so that the closed box $\tilde{B} = B(y_0, R) \times [t_0 - \tilde{\varepsilon}, t_0 + \varepsilon] \subset U$. Since F is continuous and $\tilde{B}$ is compact, there is M such that $|F| \leq M$ on $\tilde{B}$. Finally let $\varepsilon = \min\left\{\frac{R}{M+1}, \frac{1}{L+1}, \tilde{\varepsilon}\right\}$. Then $B = B(y_0, R) \times [t_0 - \varepsilon, t_0 + \varepsilon] \subset \tilde{B}$ giving (1), (2), and (3), and (4) holds by the choice of ε. □

LEMMA 11 (A-priori estimate). *Let* $I = [t_0 - \varepsilon, t_0 + \varepsilon]$ *and let* $\gamma \in C^1(I; \mathbb{E}^n)$ *be a solution to the ODE on* I *satisfying* $\gamma(t_0) = y_0$. *Then* $\gamma(t) \in B(y_0, R)$ *for all* $t \in I$.

PROOF. Suppose there is t such that $\gamma(t) \notin B(y_0, R)$. Wlog $t > t_0$ (reverse time and replace F with $-F$ otherwise). Let $t_1 = \inf\{t > t_0 : |\gamma(t) - y_0| \geq R\}$ where, since γ is continuous, $\gamma(t_1) \geq R$ and in particular

[1]For a historical survey see [**1**] which is Chapter 11 of [**5**]. The Existence and Uniqueness Theorem is due to Cauchy (~1821) in one dimension and where F is differentiable; the "Lipschitz condition" is due to Lipschiz, who treated $d > 1$ as well. Their proofs relied on what is today called the Euler Scheme, which we discuss in the homework. The proof given here is due to Picard; the older argument also gives the Peano Existence Theorem; see the Homework for that.

$t_1 > t_0$. By construction for all $s \in (t_0, t_1)$ we have $\gamma(s) \in B(y_0, R)$, so $(\gamma(s), s) \in B$ and $|F(\gamma(s), s)| \le M$. By the FTC

$$\begin{aligned} |\gamma(t_1) - y_0| = |\gamma(t_1) - \gamma(t_0)| &= \left| \int_{t_0}^{t} \dot{\gamma}(s) ds \right| \\ &= \left| \int_{t_0}^{t_1} F(\gamma(s), s)\, ds \right| \le M\,|t - t_0| \\ &\le M\varepsilon \le \frac{M}{M+1} R < R, \end{aligned}$$

contradicting the fact that $|\gamma(t_1) - y_0| = R$. □

PROOF OF PROPOSITION 9. Let $\varepsilon, R, L, M, B, I$ be as in Lemma 10. Let $C(I; \mathbb{E}^n)$ be the space of continuous functions $[t_0 - \varepsilon, t_0 + \varepsilon] = I \to \mathbb{E}^n$ equipped with the metric

$$d(\gamma, \gamma') = \left\| \gamma - \gamma' \right\|_\infty \overset{\text{def}}{=} \sup_{t \in I} \left| \gamma(t) - \gamma'(t) \right| .$$

Then (X, d) is a complete metric space. Let $X \subset C^1(I; \mathbb{E}^n)$ be the set of functions $\gamma \colon I \to \mathbb{E}^n$ such that $\gamma(t) \in B(y_0, R)$ for all t. Then X is a closed subset (if γ_n converge uniformly in $C(I; \mathbb{E}^n)$ then they converge pointwise and $B(y_0, R)$ is closed), hence a complete metric space in its own right. Given $\gamma \in X$ and $t \in I$ define

$$(G(\gamma))(t) = y_0 + \int_{t_0}^{t} F(\gamma(s), s)\, ds .$$

The integral is well-defined since γ and F are continuous and since by the Lemma for all $s \in I$ we have $(\gamma(s), s) \in B \subset \Omega$. We observe that this also means that $|F(\gamma(s), s)| \le M$ for all s. Now by the FTC the function $G(\gamma) \colon I \to \mathbb{E}^n$ is continuously differentiable. Moreover we have

$$\begin{aligned} |(G(\gamma))(t) - y_0| &= \left| \int_{t_0}^{t} F(\gamma(s), s)\, ds \right| \\ &\le \left| \int_{t_0}^{t} |F(\gamma(s), s)|\, ds \right| \\ &\le |t - t_0|\, M \le \varepsilon M \le \frac{M}{M+!} R < R . \end{aligned}$$

It follows that $G(\gamma)$ is a continuous function on I valued in $B(y_0,R)$, hence also an element of X. Finally let $\gamma,\gamma' \in X$. Then for each t we have

$$\begin{aligned}\left|(G(\gamma))(t)-\left(G(\gamma')\right)(t)\right| &= \left|\int_{t_0}^{t} F\left(\gamma(s),s\right)ds - \int_{t_0}^{t} F\left(\gamma'(s),s\right)ds\right| \\ &\le \left|\int_{t_0}^{t} \left|F\left(\gamma(s),s\right) - F\left(\gamma'(s),s\right)\right| ds\right| \\ &\le \left|\int_{t_0}^{t} L\left|\gamma(s)-\gamma'(s)\right| ds\right| \\ &\le L\varepsilon \sup_{s\in I}\left|\gamma(s)-\gamma'(s)\right| .\end{aligned}$$

Taking the supremum over t we get for $\rho = \frac{L}{L+1} < 1$ that

$$d\left(G(\gamma),G(\gamma')\right) \le \rho d\left(\gamma,\gamma'\right) .$$

By the Banach Fixed Point Theorem (contractive mapping principle) there is a unique $\gamma \in X$ such that $G(\gamma) = \gamma$, in other words such that for all t we have

$$\gamma(t) = y_0 + \int_{t_0}^{t} F\left(\gamma(s),s\right)ds .$$

Clearly $\gamma(t_0) = y_0$. In addition, by the Fundamental Theorem of Calculus γ is a differentiable function and, for all $t \in I$,

$$\dot\gamma(t) = F\left(\gamma(t),t)\right) .$$

Conversely, any solution defined on I belongs to X by Lemma 11 and is a fixed point for G, so the solution is unique. ☐

THEOREM 12 (Picard Existence and Uniqueness Theorem). *Given the Lipschitz ODE (Ω,F) and an initial condition (y_0,t_0):*

(1) *(Existence) There exist a solution γ of the ODE on some interval $I = (t_0-\varepsilon, t_0+\varepsilon)$.*
(2) *(Uniqueness) If (I,γ) and (I',γ') are two solutions, then $\gamma = \gamma'$ on $I \cap I'$.*
(3) *(Blowup) There exists a a solution $(I_{\max},\gamma_{\max})$ defined on an open interval such such any other solution is obtained by restricting $\gamma_{\max}$ to a subinterval of $I_{\max}$. Furthermore if $I_{\max} = (a,b)$ and either a or b is finite then $\gamma(t)$ "escapes" as $t \to a^+$ or $t \to b^-$ in the sense that for any compact set $K \subset \Omega$ there is $\delta > 0$ such that if $t < a+\delta$ or $t > b-\delta$ we have $(\gamma(t),t) \notin K$.*
(4) *(Autonomous equation) Suppose $F_0 \colon \Omega_0 \to \mathbb{R}^n$ for $\Omega_0 \subset \mathbb{E}^n$ and $F(y,t) = F_0(y)$ is independent of t. Then in the blowup we get that eventually $\gamma(t) \notin K$ for compact subsets of Ω_0.*

PROOF. The first claim is Proposition 9. For the second claim suppose first that I,I' are open and let $J = \{t \in I \cap I' \mid \gamma(t) = \gamma'(t)\}$. By assumption $\gamma(t_0) = \gamma'(t_0) = y_0$ so $t_0 \in J$. This set is closed since γ,γ' are continuous. To see that it is open let $t_1 \in J$. Applying Proposition 9 to the initial condition $(\gamma(t_1),t)$ we see that $\gamma = \gamma'$ on an interval containing t_1. By connectedness $J = I \cap I'$. Finally if an endpoint of I or

I' is contained in both intervals then it is a limit point of the intersection, and the solutions agree there by continuity.

Let $\mathcal{S}$ be the set of solutions γ defined on open intervals containing t_0. By (2), $\gamma_{\max} = \bigcup \mathcal{S}$ is a function and its domain I is a union of intervals containing t_0 hence an open interval. For any $t \in I$ there is some solution $\gamma \in \mathcal{S}$ defined at t hence on a neighbourhood of t and since $\gamma_{\max}$ agrees with γ on that interval $\gamma_{\max}$ is differentiable at t and is a solution. Given a compact set $K \subset \Omega$ for each $(y,t) \in K$ obtain $\varepsilon, R, L, M, B, I$ are in Lemma 10. By compactness we can cover K with finitely many boxes

$$B'_k = B(y_k, R_k/2) \times [t_k - \varepsilon_k/2, t_k + \varepsilon_k/2] \subset B_k = B(y_k, R_k/2) \times [t_k - \varepsilon_k, t_k + \varepsilon_k] \subset \Omega$$

such that F is bounded by M_k and has Lipschitz constant L_k on B_k. Let $M = \max_k M_k$, $L = \max_k L_k$, $R = \frac{1}{2} \min_k R_k$ and let $\varepsilon = \min\left\{\frac{R}{M+1}, \frac{1}{L+1}, \frac{1}{2}\varepsilon_k\right\}$. Then for each $(y,t) \in K$ the parameters ε, R, L, M work on $B(y,R) \times [t-\varepsilon, t+\varepsilon]$. It follows that any solution passing through (y,t) can be extended by at least time ε, contradicting the minimality of a and the maximality of b if the return time is close enough to a or b respectively.

When F is autonomous the bounds above are independent of t so the the time to live is uniform on compacta of Ω_0 and the same argument applies. □

1.2. Galilean group; spacetime

In principle the force can be anything (say as "external" forces). The *interactions* between particles, however, are more restricted. To see this we need to introduce the symmetry.

1.2.1. Mathematics. We extend our affine space $\mathbb{E}^d$ to *spacetime* $A^{d,1} = \mathbb{E}^d \times \mathbb{E}^1$, equipped with the projection $t \colon A^{d+1} \to \mathbb{E}^1$ on the last coordinate we call "time". A point in spacetime is called an *event*. Recall that we have already equipped $\mathbb{E}^d$ with the Euclidean metric. The subset $\mathbb{E}^d \times \{t\}$ is called a *timeslice*; two events in it are said to be *simultaneous*.

EXERCISE 13. $\mathrm{Isom}(\mathbb{E}^d)$ contains $\mathbb{R}^d$ acting by translations, which is a simply transitive subgroup. The point stabilizer is $\mathrm{O}(d)$ and thus have $\mathrm{Isom}(\mathbb{E}^d) \simeq \mathrm{O}(d) \ltimes \mathbb{R}^d$. In particular, $\mathrm{Isom}(\mathbb{E}^d) \subset \mathrm{Aff}(\mathbb{E}^d)$.

DEFINITION 14. A *Galilean transformation* is an invertible affine map between two spacetimes of the same dimension, which (1) preserves simultaneity; (2) restricts to an isometry on each timeslice. The *Galilean group* is the group of Galilean automorphisms of a single spacetime ("Galilean symmetries").

EXAMPLE 15. Fix $\underline{v} \in \mathbb{R}^d$. Then mapping $(x,t) \mapsto (x + \underline{v}t, t)$ ("uniform motion") is a Galilean symmetry. Similarly the mapping ("time translation") $(x,t) \mapsto (x, t+s)$.

EXERCISE 16. The Galilean group is a group; it is an appropriate semidirect product.

1.2.2. Physics.

AXIOM 17 (Galilean symmetry). *The laws of physics are invariant under the action of the Galilean group. In other words, when the force F only represents internal forces between the particles, it must be equivariant for the Galilean group.*

AXIOM 18. *The force on each particle is the vector sum of an external force on the particle, and interaction forces between pairs of particles.*

LEMMA 19. *Suppose F_{ij} only depends on x_i, x_j. Then F_{ij} is parallel to the displacement $x_j - x_i \in \mathbb{R}^d$.*

AXIOM 20 (Newton's third law). $F_{ij} = -F_{ji}$.

COROLLARY 21 (Conservation of total momentum). *Suppose there are no external forces. Then $\sum_j m_j v_j$ is constant.*

EXAMPLE 22. Two masses connected by a spring freely moving in 1d.

1.3. Energy and work

The inner product on $\mathbb{R}^d$ represented by a function $g\colon \mathbb{R}^d \to \left(\mathbb{R}^d\right)^*$ extends to an inner product $\oplus_{j=1}^N g_j$ on $\mathbb{R}^{dN}$.

DEFINITION 23. Associating to each particle a mass $m_j > 0$ we let $M = \bigoplus_{j=1}^d m_j g$ be the *mass matrix*. If the jth particle has velocity $\underline{v}_j \in \mathbb{R}^d$ we we call

$$T = \frac{1}{2}\langle M\underline{v}, \underline{v}\rangle = \frac{1}{2}\sum_j m_j \left\langle \underline{v}_j, \underline{v}_j \right\rangle = \frac{1}{2}\sum_j m_j \left|\underline{v}_j\right|^2$$

the *kinetic energy* of the system.

What is the time derivative of this quantity?

$$\begin{aligned}\frac{dT}{dt} &= \sum_j \left\langle m_j \underline{a}_j, \underline{v}_j \right\rangle \\ &= \langle M\underline{a}, \underline{v}\rangle \\ &= \langle F, \underline{v}\rangle \,.\end{aligned}$$

DEFINITION 24. The *work* done by the force when the system moves along the path is the integral

$$\int \langle F, d\gamma\rangle = \int_{t_0}^{t_1} \langle F, \underline{v}\rangle \, dt \,.$$

Note that $\langle F, \underline{v}\rangle$ is the sum of terms $\left\langle F_j, \underline{v}_j \right\rangle$ corresponding to the individual particles.

COROLLARY 25. *The change in kinetic energy is the total work done by the force.*

We now divide the forces into categories.

DEFINITION 26. A force on a single particule is *conservative* if locally $F = -dU$ for some function U of position called the *potential*.

EXERCISE 27. This is equivalent to $\oint F dx = 0$ for small loops (all loops if true globally).

- For a force between two particles get $F_{ij} = -F_{ji}$. This iscalled "Newton's Third law".

LEMMA 28. *For a conservative force we have* $\frac{dU}{dt} = -F\underline{v}$.

OBSERVATION 29. *Constraint forces do no work.*

PROOF. A constraint force acts dually to the level sets, whereas $\underline{v}$ is in the tangent space. □

We have shown

PROPOSITION 30 (Conservation of mechanical energy). *Suppose that all inter-particle forces are conservative, with total potential U, and let F_j be the external force on the jth particle. Let $E = T + U$. Then*

$$\frac{dE}{dt} = \sum_j \left\langle F_j, \underline{v}_j \right\rangle .$$

Adding to U the potential due to any conservative external forces gives the same result with F_j representing any non-conservative forces.

CHAPTER 2

Kinematics

We begin by developing the language to *describe* motion, material which will lead directly to *predicting* motion in Chapter 1.

Physics keywords: configuration space,

Mathematics keyword: inverse and implicit function theorems,

2.1. Configurations and configuration space

2.1.1. Physics.

DEFINITION 31 (Informal). A *mechanical system* consists of several point particles moving in some *ambient space* subject to *interactions* and *constraints.*

The ambient space will be Euclidean d-space, denoted $\mathbb{E}^d$. For the distinction between $\mathbb{E}^d$ and $\mathbb{R}^d$ see the problems on "Affine Algebra" in Problem Set 1, and also [**2**, \S2A].

DEFINITION 32. A *configuration* of the system is then a point $x \in \left(\mathbb{E}^d\right)^N$ satisfying the constraints. The *configuration space* of the system is the set X of all configurations.

- We will think of a particle moving on the round 2-sphere as moving on the surface $x^2+y^2+z^2=1$ in $\mathbb{E}^n$, but it is also possible to think of it as moving on the sphere directly.
- We will almost always have $d \in \{1,2,3\}$ but this is a contingent fact about our everyday experience, not a mathematical requirement.

EXAMPLE 33. Single free particle; single particle at the end of a Hookean spring; physical pendulum with massless rod (2d ; 3d); rope-and-pulley system;

- From the mathematical point of view one can dispense with this definition and just talk about the configuration manifold from the start, but that's not how physicists think,. More importantly we need to be able to construct the configuration space of a physical system.
- We will pretend that continuum systems (e.g. a solid rod) actually consist of a finite but large number of particles. As long as the rod is *rigid* this will not cause a problem since the configuration space will be finite-dimensional.
- Truly infinite-dimensional systems (e.g. fluid flow, or deformation of a plastic material) are the subject of *continuum mechanics* and outside the scope of this course.

2.1.2. Mathematics. Often the constraints are *holonomic* in that, locally at least, they can be written in the form[1] $\underline{F}(x) = \underline{0}$? in other words they concern the configuration of the system only ? e.g. the particle constrained to move on the sphere (say of radius $r(t)$), or ("rigid motion") where the distance between a pair of particles is fixed. Other examples of constraints include

- Hard boundaries, e.g. a bouncing ball restricted to the upper half-plane $y \geq 0$. Mathematically we can either work on a *manifold with boundary* or handle the situation at the boundary separately.
- Constraints on the *motion* (e.g. rolling without slipping), which we will develop later. we

Formally we fix an open set $\Omega \subset \mathbb{E}^{dN}$, a time interval I, and assume the constraints take the form $\underline{F} = \underline{0}$ for some continuously differentiable function $\underline{F} \colon \Omega \to \mathbb{R}^m$. We will generally assume $\underline{F}$ is *non-degenerate* in that $\operatorname{rank} d\underline{F} = m$, at least locally on X (i.e. possibly we express the constraints by different functions in different places).

EXAMPLE 34. In $\mathbb{E}^2$ suppose we have a massless slider moving along a horizontal wire. A point of mass m is attached a massless rigid rod of length L freely swinging from the slider. We want to say something like:

"Let the slider be at $(x_\mathrm{s}, y_\mathrm{s})$ and the mass be at (x, y), with constraints

$$\begin{cases} x_s \in [a,b] \\ y_s = 0 \\ (x - x_\mathrm{s})^2 + (y - y_\mathrm{s})^2 = L^2 . \end{cases} \quad "$$

Of course a better parametrization would be through the angle θ the rod makes with the vertical axis (say). We also want to say:

"Instead use as coordinates the location x_s of the slider and the angle θ of the mass. The slider is then at $(x_\mathrm{s}, 0)$ and the mass is at $(x_\mathrm{s} + L\sin\theta, -L\cos\theta)$".

Finally, we would like to say

"The potential energy of the system will then be $U = -mgL\cos\theta$. The kinetic energy will be

$$\frac{1}{2}m\left[\dot{x}^2 + \dot{y}^2\right] = \frac{1}{2}m\left[\dot{x}_\mathrm{s}^2 + 2L\cos\theta\dot{x}_s\dot{\theta} + L^2\dot{\theta}^2\right] .$$

Our goal is to make sense of all these statements. For this we need to understand what we mean by the *coordinates* x_s, y_s, x, y, θ, what we mean by $U(\theta)$ where U ought to be a function on X, what we mean by derivatives on configuration space and of coordinates, and so on. For technical reasons we will begin with the derivatives, and the discuss coordinates, coordinate systems, and parametrization.

[1]In general one should permit the constraints to depend on time, in which case the configuration space would be $X = X(t)$, and the analysis below should be extended. This may be developed in a later version of the notes or in a problem set.

REMARK 35. Observe that the angle θ is not really a *function* on configuration space ? if we go around a full circle we acquire a phase of 2π. We can define branches locally, but not globally. The same will apply to constraints written in terms of θ.

2.1.3. tangent vectors and derivatives.

REMARK 36. Recall that $\underline{F}\colon \mathbb{E}^n \to \mathbb{E}^m$ is *differentiable* at $x \in \mathbb{E}^n$ if there is a linear map $d\underline{F}_x\colon \mathbb{R}^n \to \mathbb{R}^m$ so that for $\underline{v} \in \mathbb{R}^n$, $\underline{F}(x+\underline{v}) = \underline{F}(x) + d\underline{F}_x(\underline{v}) + \underline{R}(x,\underline{v})$ where $\frac{|\underline{R}(x,\underline{v})|}{|\underline{v}|} \to 0$ as $|\underline{v}| \to 0$. Note that there is no dot product or metric here; just the notion of *displacement* on affine space.

EXAMPLE 37. When $m = 1$ note that dF_x is not a vector ? it is a linear map from $\mathbb{R}^n \to \mathbb{R}$, in other words a *linear functional*; in physics language a *dual vector* or a *covector*.

REMARK 38. If we wish to speak of the gradient *vector* $\vec{\nabla} F(x)$ we need a way to associate a vector to each linear functional. This is provided by an *inner product* (look up "Riesz Representation Theorem"), but note that the choice of inner product matters, and different inner products will produce different gradients for the same function.

Let $x, x' \in X$ be close to each other. We can write $x' = x + \varepsilon \underline{v}$ where $\underline{v}$ is some unit vector and ε is small. Then

$$\begin{aligned} 0 = \underline{F}(x') &= \underline{F}(x + \varepsilon \underline{v}) \\ &= \underline{F}(x) + \varepsilon d\underline{F}_x(\underline{v}) + o(\varepsilon) \\ &= \varepsilon d\underline{F}_x(\underline{v}) + o(\varepsilon)\,. \end{aligned}$$

so

$$d\underline{F}_x(\underline{v}) = o(1)\,.$$

By compactness as $x' \to x$ we will have the $\underline{v}$ converge to a point on the sphere satisfying $d\underline{F}_x(\underline{v}) = 0$, that is to $\underline{v} \in \operatorname{Ker} d\underline{F}_x$.

DEFINITION 39. For $x \in X(t)$ the *tangent space* is $T_x X(t) = \operatorname{Ker} d\underline{F}_x$.

NOTATION 40. We use Newton's dot to denote derivatives with respect to time. By $\dot{\gamma}(t)$ we mean the vector of partial derivatives in $\mathbb{R}^{dN}$, which is also the image of the standard basis vector of $T_t\mathbb{R}$ by the linear map $d\gamma_t$.

LEMMA 41. *Let I be an interval, and let $\gamma\colon I \to \mathbb{E}^{dN}$ be a differentiable curve. Suppose that $\gamma(t_0) \in X$ for some t_0. Then the image of γ lies in X iff $\dot{\gamma}(t) \in \operatorname{Ker} d\underline{F}_{\gamma(t)}$ for all $t \in I$.*

PROOF. PS1. □

2.2. Coordinates

2.2.1. Mathematics.

THEOREM 42 (Implicit function theorem). *Let $x \in \Omega$ and suppose that* $\operatorname{rank} d\underline{F}_x = m$. *Then we can choose some m coordinates of $\mathbb{E}^{dN}$ so that locally these coordinates are uniquely determined by the others. Furthermore this function has the expected derivative. If $\underline{F}$ is k times differentiable so is the function defined implicitly.*

COROLLARY 43 (Inverse function theorem). *When $m = dN$ we have an inverse with the inverse derivative.*

By the implicit function theorem we can, at least locally, *parametrize* configuration space as follows: we choose some set $\neg C \in \binom{[dN]}{m}$. Then any configuration $x = (x_i)_{i\in[dN]}$ is uniquely determined by $(x_i)_{i\in C}$.

EXAMPLE 44. Particle on incline parametrized by x or y coordinate. Same for particle on circle, note different coordinates at different points.

- No constraints in such a a parametrization.

It is often more convenient, however, to parametrize by something other than the Euclidean coordinate axes. The key observation is that the x_i are not functions "valued in X" but rather *functions on X*!

DEFINITION 45. A *coordinate* (in the physics literature "*generalized coordinate*") is a pair (U, q_α) where $U \subset X$ is an open set and $q_\alpha \colon U \to \mathbb{R}$ is any function. A *system of coordinates* or *coordinate patch* is a tuple $q = \left(U, \{q_\alpha\}_{i=1}^{\dim X}\right)$ of coordinates defined on the same neighbourhood such that q is continuously differentiable and dq_x is invertible on T_xX for $x \in U$.

What do we mean by dq? Note that by the implicit function theorem, if $x, x' \in X$ are close enough we have $x' = x + \underline{v} + \underline{e}$ where $\underline{v} \in T_xX$ and $|\underline{e}| = o(|x - x'|)$ as $x \to x'$. We can thus differentiate functions on X (without extending them to the ambient $\mathbb{E}^{dN}$) by asking whether $f(x') - f(x)$ is approximately linear in $\underline{v}$.

DEFINITION 46. Let $f \colon U \to \mathbb{E}^r$ where $U \subset X$ is open . We shall say f is *differentiable* at x if there is a linear map $df_x \colon T_xX \to \mathbb{R}^r$ such that $f(x') = f(x) + df_x(\underline{v}) + o(|x - x'|)$.

LEMMA 47 (Differentiation on X).
1. *Let $V \subset \mathbb{E}^{dN}$ be an open set, and let $f \colon V \to \mathbb{E}^r$ be a function differentiable at some $x \in U = V \cap X$. Then $f \restriction_U \colon U \to \mathbb{E}^r$ is differentiable in the sense above and $d(f\restriction_U)_x = (df_x)\restriction_{T_xX}$.*
2. *df_x is linear in f and satisfies the chain rule in both directions (i.e. for composition with $g \colon \mathbb{E}^s \to X$ and with $h \colon \mathbb{E}^r \to \mathbb{E}^t$). It therefore also satisfies the Leibnitz rule.*

PROOF. Exercise. □

EXAMPLE 48. The angle θ for a circle, e..g the pendulum. Let $S^1 \subset \mathbb{E}^2$ be the unit circle $\{x^2 + y^2 = 1\}$. On the open right semicircle we define $\theta = \arctan(y/x)$. We then conversely have the *parametrization* (inverse map) $\theta \mapsto (\cos\theta, \sin\theta)$. It's also possible to define θ on any arc not covering the whole circle.

REMARK 49. Since θ is only locally defined, we sometimes prefer to have the coordinates NOT be valued in $\mathbb{R}$? e.g. have θ valued in S^1. This requires some care, but has some advantages.

EXERCISE 50. Given a mechanical system, find a coordinate system, and then a *parametrization* of configuration space by the coordinates.

NOTATION 51. A (locally defined) function $f\colon \mathbb{R}^{\dim X} \to \mathbb{R}$ induces a function on configuration space by composing with the coordinates: when we write $f(q)$ we really mean $f \circ (q_\alpha)_\alpha$ (example: the potential energy $U(\theta) = -mgL\cos\theta$ from Example 34). Conversely if f is a function on X we can identify it with a function defined on the coordinates by composing with the inverse of q. Similarly a curve $\gamma\colon I \to X$ induces a coordinate curve via $q \circ \gamma$, and a coordinate curve induces a curve in configuration space by composition with q^{-1}.

EXAMPLE 52. Potential energy due to external gravity, or due to interaction between pairs (or groups) of particles.

REMARK 53. It is in fact useful to have the coordinates depend on time ? to have $q_\alpha\colon U \times I \to \mathbb{R}$ where I is some time interval. This is significant and will play a role in the sequel.

2.2.2. Physics: computation in coordinates. We first clarify something.

Warning: Let $f\colon X \to \mathbb{R}$ be some function (say potential energy). Then $f \circ q^{-1}$ is a function on coordinate space (i.e. if you plug in values for the coordinates you get the value of f at the corresponding point of X). Following the physics convention we will use the same letter for both functions, and leave it to the reader to figure out which we mean. In particular when we write $x_i = x_i(q;t)$ we might mean the standard coordinate functions on X coming by restriction from $\mathbb{E}^{dN}$, or the Euclidean coordinates as functions of the generalized coordinates.

- Suppose the system moves according to a curve $\gamma \subset X$. Then the coordinates change according to $q_\alpha(\gamma(t);t)$. We usually write $q_\alpha(t)$ for these functions; the vector $(q_\alpha(t))_\alpha \in \mathbb{R}^{\dim X}$ is the *coordinate curve*. We often compute these functions directly, and then see the implications in physics space by applying $q^{-1}(\cdot;t)$ to get points in X.
- In particular we will usually write the *equations of motion* as ODEs for $q_\alpha(t)$ and solve those.
- We often try to choose the coordinates q_α to make the expression for relevant functions or for the equations of motion simpler.

Let $\gamma\colon I \to X$ be a differentiable curve through configuration space. As we saw in Lemma 41, at every time t we have $\dot\gamma(t) \in T_{\gamma(t)}X$.

DEFINITION 54. We call $\dot\gamma$ the *velocity* of the path. This is a vector of the N velocities of the individual particles.

DEFINITION 55. Let $\langle\cdot,\cdot\rangle$ denote the inner product on $\mathbb{R}^d = T_x\mathbb{E}^d$ with associated norm $|\cdot|$; think of $\dot\gamma(t)$ as a collection of N vectors $v_j \in \mathbb{R}^d$, and suppose the jth particle has mass m_j. Then the *kinetic energy* of the particles is

$$K = \frac{1}{2}\sum_j m_j \left|\dot v_j\right|^2 .$$

Observe that is exactly the restriction of a positive-definite quadratic form from $T_{\gamma(t)}\mathbb{E}^{dN}$ to $T_{\gamma(t)}X$, hence again a positive-definite quadratic form.

Via the parametrization $x_i = x_i(q;t)$ we obtain a linear relation

$$\dot{x}_i = \sum_\alpha \frac{\partial x_i}{\partial q_\alpha}\dot{q}_\alpha$$

which allows us to change variables in K, writing K as a quadratic form in the $\dot{q}_\alpha$ instead. Against it is positive definite.

EXAMPLE 56 (Rotating frame). Suppose we have a particle moving in the plane, where we name points by their $\begin{pmatrix} x \\ y \end{pmatrix}$ coordinates. Let $R(t) = \begin{pmatrix} \cos(\omega t) & -\sin(\omega t) \\ \sin(\omega t) & \cos(\omega t) \end{pmatrix}$ (warning: this is the matrix of rotation by $-\omega t$), and consider the time-dependent coordinate system

$$\begin{pmatrix} X \\ Y \end{pmatrix} = R(t)\begin{pmatrix} x \\ y \end{pmatrix}$$

(so $X = X(x,y;t)$ and $Y = Y(x,y;t)$ as indicated. Conversely we have

$$\begin{pmatrix} x \\ y \end{pmatrix} = R(-t)\begin{pmatrix} X \\ Y \end{pmatrix}$$

so

$$\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = R(-t)\begin{pmatrix} \ddot{X} \\ \ddot{Y} \end{pmatrix} - 2\dot{R}(-t)\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} + \ddot{R}(-t)\begin{pmatrix} X \\ Y \end{pmatrix}$$

and

$$m\begin{pmatrix} \ddot{X} \\ \ddot{Y} \end{pmatrix} = R(t)m\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} + 2R(t)\dot{R}(-t)m\begin{pmatrix} \dot{X} \\ \dot{Y} \end{pmatrix} - R(t)\ddot{R}(-t)m\begin{pmatrix} X \\ Y \end{pmatrix}.$$

Now Newton's 2nd law reads $m\begin{pmatrix} \ddot{x} \\ \ddot{y} \end{pmatrix} = F(x,y;t)$ where we

SCHOLIUM 57. This definition clarifies that velocity is a *pointwise* notion: we need more structure to compare velocities at different points. For example, a particle moving around the circle has velocities tangent to the circle. To study $v(t+h) - v(t)$ we need a *connection*.

REMARK 58. In fact we could have started with *any* positive-definite quadratic form. If the ambient space is a Riemannian manifold (M,g) then on $\left(M^N, \bigoplus_{j=1}^N g\right)$ the kinetic energy is the quadratic form on the tangent space which, relative to $\bigoplus_{j=1}^N g$, is block-diagonal with eigenvalues m_j.

REMARK 59. As we shall see later, the most important fact is the *convexity* of K as a function on T_xX.

CHAPTER 3

Lagrangian mechanics

3.1. Introduction

3.1.1. Historical overview.

- Euler: the equations of motions of Newtonian mechanics can be written in a form that works for any coordinate system.
- Lagrange: even if there are constraints.
- Hamilton: these equations follow from a variational principle.

3.1.2. Plan.

(1) (Mathematics) Calculus of variations I
(2) (Physics) Hamilton's principle and the Euler?Lagrange equations; examples
(3) (Physics) conservation laws
(4) (Mathematics) Lagrange multipliers for variational problems
(5) (Physics) Constraint forces

REMARK 60. We will not *derive* the Euler?Lagrange equations (i.e. show that they are equivalent to Newtonian mechanics), which is essentially a tedious calculation.

3.2. Calculus of Variations

3.2.1. The problem; formal calculation. Fix a bounded domain $\Omega \subset \mathbb{R}^r$, and consider the problem of minimizing the expression

$$S = \int_\Omega L(u(t), du_t; t)\, dt$$

over the space of sufficiently nice functions $u\colon \bar{\Omega} \to \mathbb{R}^n$, subject to a boundary condition $u \restriction_{\partial\Omega} = g$ for some fixed g. Here $L\colon \mathbb{R}^n \times M_{n\times r} \times \Omega \to \mathbb{R}$ is some sufficiently nice functions.

EXAMPLE 61 (Brachistochrone; Johann Bernouli 1696 after Galileo). Given two points A, B in a vertical plane with A higher, find the curve $y = u(x)$ such that a mass sliding along the graph of u subject to gravity alone will reach B from A in the shortest time.

Align the y-axis vertically down, and suppose $A = (0,0)$ and $B = (x_B, y_B)$. When the mass is at $(x, u(x))$ it has velocity $\frac{1}{2}mv^2 = mgu(x)$ by conservation of energy, so $v(x) = \sqrt{2gu(x)}$. The length of the part of the curve from x to $x+dx$ is $\sqrt{1+u'(x)^2}dx$ so the time it takes to cover that segment is $\sqrt{\frac{1+u'^2}{u}}dx$. We therefore need to minimize

$$\int_0^{x_B}\sqrt{\frac{1+u'^2}{u}}dx$$

subject to $u(0)=0$; $u(x_B)=y_B$.

- Problem famously solved by Newton overnight after finding a challenge letter from Bernoulli returning from his work at the mint; he sent the solution anonymously to the Royal Academy by first post and Bernoulli famously remarked "I recognize the lion from his claw mark" (The Bernoullis took two weeks to solve the problem).

EXAMPLE 62 (Catenary). A chain of length L made from a material of constant density hangs from two points A,B in the vertical plane. It is known ("principle of virtual work") that the chain will hang so as to minimize total potential energy. What shape will it take?

EXAMPLE 63 (Minimal surface). Let $\Omega\subset\mathbb{R}^d$ be a plane, u any curve. Then $\int_\Omega\sqrt{1+|Du(x)|^2}dx$ is the area of the hypersurface $y=u(x)$ given by the graph of u in $\mathbb{R}^{d+1}$.

- Idea: Differentiate wrt u, set derivative to zero.

3.2.2. Differentiation in function spaces. Let V be a (real) vector space (example: the space of paths $\gamma\colon I\to\mathbb{R}^n$ in coordinate space satisfying some differentiability conditions. A function $f\colon V\to\mathbb{R}$ is called a *functional*. Let $x\in A$.

DEFINITION 64. We say that f is

(1) *differentiable* at x if V is a Banach space and there is a linear functional $\lambda=df_x\in V^*$ such that $f(x+v)=f(x)+\langle\lambda,v\rangle+o(\|v\|)$.
(2) *Gateaux differentiable* at x if V is topological and there is a linear functional $\lambda=df_x\in V^*$ such that $df_x(v)=\lim_{h\to0}\frac{f(x+hv)-f(x)}{h}$ holds for each $v\in V$.

EXAMPLE 65. If $V=\mathbb{R}^n$ the first notion is the usual derivative, and the second is the directional derivative (when linear).

REMARK 66. Can directly define the differentiability of functionals on manifolds of paths $\gamma\colon I\to X$ but we'll elide this point.

Since we will only be interested in critical points of f, we will consider a much weaker condition, where we only differentiate at a single direction, and only consider a subset of the possible directions. We will also concentrate on the case $r=1$ where the situation is considerably simpler (for functional-analytic reasons),

3.2.3. Formal calculation in coordinates. Let $I = [t_0, t_1]$ be a closed interval. Let $\Omega \subset \mathbb{R}^n$ be open *"coordinate patch". Fix a function ("Lagrangian") $L\colon T\Omega \times I \to \mathbb{R}$ as nice as needed (write this as $L = L(q, v; t)$, and let $\gamma\colon I \to \Omega$ be as nice as needed. The associated *action* is

$$S(\gamma) = \int_{t_0}^{t_1} L(\gamma, \dot{\gamma}; t)\, dt\,.$$

suppose γ minimizes S among reasonable curves with $\gamma(t_0) = a$, $\gamma(t_1) = b$.

To investigate this let $\eta(t) \in C_c^\infty(I^\circ)$ be a "bump function". Then

$$S(\gamma + \varepsilon\eta) = \int_{t_0}^{t_1} L(\gamma + \varepsilon\eta, \dot{\gamma} + \varepsilon\dot{\eta}; t)\, dt\,.$$

Taylor expansion gives (suppressing the dependence of L on time)

$$L(\gamma(t) + \varepsilon\eta(t), \dot{\gamma}(t) + \varepsilon\dot{\eta}(t); t) - L(\gamma(t), \dot{\gamma}(t); t) = \varepsilon \left\langle \frac{\partial L}{\partial q}_{(\gamma(t),\dot{\gamma}(t))}, \eta(t) \right\rangle + \varepsilon \left\langle \frac{\partial L}{\partial v}_{(\gamma(t),\dot{\gamma}(t))}, \dot{\eta}(t) \right\rangle + O(\varepsilon^2)\,.$$

Integrating this dt we get

$$S(\gamma + \varepsilon\eta) - S(\gamma) = \varepsilon \int_{t_0}^{t_1} \left[\left\langle \frac{\partial L}{\partial q}_{(\gamma(t),\dot{\gamma}(t))}, \eta(t) \right\rangle + \left\langle \frac{\partial L}{\partial v}_{(\gamma(t),\dot{\gamma}(t))}, \dot{\eta}(t) \right\rangle \right] dt + O(\varepsilon^2)\,.$$

Thus if γ is extremal (or even critical) for S we must have

$$\int_{t_0}^{t_1} \left[\left\langle \frac{\partial L}{\partial q}_{(\gamma(t),\dot{\gamma}(t))}, \eta(t) \right\rangle + \left\langle \frac{\partial L}{\partial v}_{(\gamma(t),\dot{\gamma}(t))}, \dot{\eta}(t) \right\rangle \right] dt = 0\,.$$

REMARK 67. If γ needs to be valued in some open domain we can always choose ε small enough to ensure that $\gamma + \varepsilon\eta$ remains in the domain.

We now integrate the second term by parts. Since η has compact support we have $\eta(t_0) = \eta(t_1) = 0$ so there are no boundary terms and we get

$$\int_{t_0}^{t_1} \left[\left\langle \frac{\partial L}{\partial q}_{(\gamma(t),\dot{\gamma}(t))}, \eta(t) \right\rangle - \left\langle \frac{d}{dt}\frac{\partial L}{\partial v}_{(\gamma(t),\dot{\gamma}(t))}, \eta(t) \right\rangle \right] dt = 0\,,$$

that is

$$\int_{t_0}^{t_1} \left\langle \frac{\partial L}{\partial q}_{(\gamma(t),\dot{\gamma}(t))} - \frac{d}{dt}\frac{\partial L}{\partial v}_{(\gamma(t),\dot{\gamma}(t))}, \eta(t) \right\rangle dt = 0\,.$$

Now if γ is extremal this must hold for *all* η. However if $\gamma \in C^1$ and $L \in C^2$ then the function $\frac{\partial L}{\partial x}_{(\gamma(t),\dot{\gamma}(t))} - \frac{d}{dt}\frac{\partial L}{\partial v}_{(\gamma(t),\dot{\gamma}(t))}$ is continuous; if it were nonzero somewhere we could choose η to make the integral nonzero. It follows ("Euler?Lagrange equation") that along the path we have

$$\frac{d}{dt}\frac{\partial L}{\partial v}(\gamma(t), \dot{\gamma}(t); t) = \frac{\partial L}{\partial q}(\gamma(t), \dot{\gamma}(t); t)\,.$$

By the chain rule we can also write this as

$$\left(\frac{\partial^2 L}{\partial v^2}(\gamma(t),\dot\gamma(t);t)\right)\ddot\gamma(t)+\left(\frac{\partial^2 L}{\partial v\partial x}(\gamma(t),\dot\gamma(t);t)\right)\dot\gamma(t)+\left(\frac{\partial^2 L}{\partial v\partial t}(\gamma(t),\dot\gamma(t);t)\right)-\frac{\partial L}{\partial q}(\gamma(t),\dot\gamma(t);t)=0\,,$$

which is visibly a second order ODE we can hope to solve.

- Classical approach: "indirect method", that is first solve this equation, then show that the solution is extremal.
- "Direct method": show a-priori that the action has minimizers and that they must satisfy the equation.

3.2.4. From informal to formal.

LEMMA 68. *Let $f \in C(a,b)$. Suppose that for all non-negative $\eta \in C_c^\infty(a,b)$ we have $\int_a^b f(t)\eta(t)dt \geq 0$. Then $f \geq 0$.*

PROOF. Suppose $f(t_0) < 0$. Then there is a small interval J about t_0 we have $f\restriction_J \leq -\delta$. Let η be any nonzero nonegative function supported on J. Then $\int_a^b$ □

LEMMA 69. *Let $f \in L^1(a,b)$. Suppose that for all non-negative $\eta \in C_c^\infty(a,b)$ we have $\int_a^b f(t)\eta(t)dt \geq 0$. Then $f \geq 0$ almost everywhere.*

PROOF. Let $d\mu = f dt$ be the measure on $[a,b]$ with density f wrt Lebesgue. Then $\mu(\eta) = \int_a^b f(t)\eta(t)dt$ and from the Riesz Representation Theorem we get that μ is a positive measure, so its Radon?Nykodim derivative f with respect to dt must be non-negative (if f we negative on a set of positive measure, the measure of that set would be negative). □

COROLLARY 70. *In either case, if the integrals always vanish so does the function.*

We have proved:

PROPOSITION 71. *Let $L \in C^2(T\Omega \times I \to \mathbb{R})$ and suppose $\gamma \in C^2(I;\Omega)$ is critical for $S = \int_{t_0}^{t_1} L dt$ given its endpoints. Then*

$$\frac{d}{dt}\frac{\partial L}{\partial v}(\gamma(t),\dot\gamma(t);t)=\frac{\partial L}{\partial q}(\gamma(t),\dot\gamma(t);t)\,.$$

Furthermore, suppose ("Ellipticity") that $\frac{\partial^2 L}{\partial v^2}$ is positive definite. We can then write the ODE in the form

$$\ddot\gamma=\left(\frac{\partial^2 L}{\partial v^2}\right)^{-1}\left[\frac{\partial L}{\partial q}(\gamma(t),\dot\gamma(t);t)-\left(\frac{\partial^2 L}{\partial v\partial x}(\gamma(t),\dot\gamma(t);t)\right)\dot\gamma(t)-\left(\frac{\partial^2 L}{\partial v\partial t}(\gamma(t),\dot\gamma(t);t)\right)\right],$$

which will have a unique solution for each initial condition $(\gamma(t_0),\dot\gamma(t_0))$.

- The ellipticity condition is exactly the positive definiteness of the mass matrix.

We would like to do two related things:

(1) Show that there actually exists a minimizer. We will concentrate on a particular class of Lagrangians
(2) Extend the class of acceptable paths γ.

DEFINITION 72. Say the Lagrangian is *standard* if it has the form

$$L(q,\dot{q};t) = \frac{1}{2}\langle M(t)\dot{q},\dot{q}\rangle - U(q,t)$$

where $M = M(t)$ is symmetric and satisfies $M \geq \mu$ for some constant $\mu > 0$, and U is continuous.

3.2.5. Existence of minimizers.

DEFINITION 73 (Sobolev space). For a sufficiently differentiable function $\gamma\colon I \to \mathbb{R}^n$ define

$$\|\gamma\|_{H^k}^2 = \sum_{i=0}^{k} \left\|\gamma^{(k)}\right\|_{L^2(I)}$$

and let $H^k(I;\mathbb{R}^n)$ be the completion of the space of smooth functions wrt this norm.

FACT 74. *This is the space of γ such that the kth distributional derivative is represented by an L^2-function.*

THEOREM 75 (Sobolev embedding). *The inclusion map $\left(C^k(I), \|\cdot\|_{H^k}\right) \to \left(C^{k-1}(I), \|\cdot\|_{C^{k-1}}\right)$ is compact.*

COROLLARY 76. *Let L be a standard Lagrangian. Then $\gamma \mapsto S(\gamma)$ is continuous with $\|\cdot\|_{H^1}$ and thus extends to a continuous function on $H^k(I)$.*

LEMMA 77. *Let $u\colon I \to \mathbb{R}^n$ be a differentiable function with $u(t_0) = u(t_1) = 0$. Then*

$$\int_{t_0}^{t_1} |u|^2\,dt \leq \left(\frac{t_1 - t_0}{2}\right)^2 \int_{t_0}^{t_1} |\dot{u}|^2\,dt\,.$$

PROOF. Wlog the interval is $[-\Delta, \Delta]$. Integrating by parts we have

$$\begin{aligned}\int_{\Delta}^{\Delta} |u|^2\,dt &= \left[t\,|u|^2\right]_{-\Delta}^{\Delta} - \int_{-\Delta}^{\Delta} tu\dot{u}dt \\ &\leq \Delta\left(\int_{\Delta}^{\Delta} |u|^2\,dt\right)^{1/2}\left(\int_{\Delta}^{\Delta} |\dot{u}|^2\,dt\right)^{1/2}.\end{aligned}$$

□

REMARK 78. This is not the optimal constant ? which is the smallest eigenvalue of the Dirichlet Laplacian.

LEMMA 79 (Coercivity). *Suppose $U(q) \leq A + C\,|q|^2$. For $\varepsilon > 0$ we can find $\delta > 0$ such that*

$$\int_{t_0}^{t_1+\delta} U(q,t) \leq B + \varepsilon \int_{t_0}^{t_1+\delta} |\dot{q}|^2\,dt\,.$$

PROOF. Let $\tilde{q}$ be a linear function of time interpolating $a = q(t_0), b = q(t_0+\delta)$ and let $u = q - \tilde{q}$. Then $\dot{u} = \dot{q} - \frac{b-a}{\delta}$ so $|\dot{u}|^2 \le 2|\dot{q}|^2 + 2\frac{|b-a|^2}{\delta^2}$

$$U(q) \le A + C|\tilde{q}+u|^2 \le A + 2C|\tilde{q}|^2 + 2C|u|^2 .$$

Now integrate on $[t_0, t_0+\delta]$. Get

$$\begin{aligned}\int_{t_0}^{t_0+\delta} U(q)dt &\le A\delta + 2C\delta\left(|a|^2+|b|^2\right) + \frac{1}{2}C\delta^2\int_{t_0}^{t_0+\delta}|\dot{u}|^2 dt \\ &\le A\delta + 2C\delta\left(|a|^2+|b|^2\right) + \frac{1}{2}C\delta|b-a|^2 + C\delta^2\int_{t_0}^{t_0+\delta}|\dot{q}|^2 dt .\end{aligned}$$

Now take δ small enough so that $C\delta^2 < \varepsilon$. □

COROLLARY 80 (Lower bound). *Suppose U grows at most quadratically and that the time interval is short enough (depending on the constants including the initial conditions) to get $\varepsilon < \mu$. Then*

(1) *$S(\gamma)$ are bounded below. In particular* $\inf_\gamma S(\gamma)$ *exists.*
(2) *Sublevel sets are bounded in H^1.*

Now let $\{\gamma_n\}_{n=1}^{\infty} \subset C^1(I)$ have $S(\gamma_n) \to \inf_\gamma S(\gamma)$. By the Sobolev embedding theorem we can pass to a subsequence so that γ_n converge uniformly to a continuous function γ_∞. By Banach?Alaoglu we can also assume that γ_n converge weakly, so that

3.3. The Euler–Lagrange equations

Let us now calculate with this formalism.

DEFINITION 81. We call $\frac{\partial L}{\partial v}$ the (generalized) *momentum*. In a particular coordinate system call $p_\alpha = \frac{\partial L}{\partial \dot{q}_\alpha}$ the *momentum* associated to the coordinate q_α (though of course this depends on the entire coordinate system). We call $\frac{\partial L}{\partial q}$ the (generalized) *force*. In particular coordinate system we get a generalized force $F_\alpha = \frac{\partial L}{\partial q_\alpha}$ associated to each coordinate, with *equation of motion*

$$\frac{d}{dt}p_\alpha = F_\alpha .$$

EXAMPLE 82. For a standard Lagrangian $L(x,v) = \frac{1}{2}\langle M(x,t)v, v\rangle - U(x,t)$ we have $p = M(x,t)\dot{x}$ and $F = -dU$.

3.3.1. Cyclic coordinates and conserved quantities.

Fix a coordinate system $q = (q_\alpha)_{\alpha=1}^n : X \to \Omega \subset \mathbb{R}^n$. The Euler?Lagrange equation take the form

$$\left\{ \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_\alpha}\right) = \frac{\partial L}{\partial q_\alpha} \quad 1 \le \alpha \le n .\right.$$

Here we think of L as a function on L via composition with q^{-1}.

DEFINITION 83. Call a coordinate q_β *cyclic* if $\frac{\partial L}{\partial q_\beta} = 0$, in other words if L does not depend on q_β explicitly (of course in the given coordinate system, that is when the other coordinates are the other q_α).

OBSERVATION 84. *If q_α is cyclic then $\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_\alpha}\right) = 0$. In other words, the associated generalized momentum $p_\alpha = \frac{\partial L}{\partial \dot{q}_\alpha} \colon TX \times \mathbb{R} \to \mathbb{R}$ is constant along the physical path.*

DEFINITION 85. We say that a quantity $f \colon TX \times \mathbb{R} \to \mathbb{R}$ is a *conserved quantity* in that situation, in other words if $\frac{d}{dt} f(\gamma(t), \dot{\gamma}(t); t) = 0$ along the physical path $\gamma(t)$.

EXAMPLE 86. Consider a particle moving in the plane with downward-pointing gravity, that is the Lagrangian

$$L = \frac{1}{2} m \left(\dot{x}^2 + \dot{y}^2\right) + mgy\,.$$

Clearly the x-coordinate is cyclic. Now retain the x-coordinate but switch to the coordinate system (x, z) where $z = x + y$. Then $y = z - x$ so we also have

$$L = \frac{1}{2} m \left(2\dot{x}^2 + \dot{z}^2 - 2\dot{z}\dot{x}\right) + mgz - mgx\,.$$

Now the x-coordinate is not cyclic ? showing that the notion of cyclicity depends on the coordinate system and not just on a single coordinate.

L itself is a function which we can differentiate along the physical path. We have

$$\begin{aligned}
\frac{d}{dt}\left(L(\gamma, \dot{\gamma}; t)\right) &= \frac{\partial L}{\partial x}\dot{\gamma} + \frac{\partial L}{\partial v}\ddot{\gamma} + \frac{\partial L}{\partial t} && \text{chain rule} \\
&= \frac{d}{dt}\left(\frac{\partial L}{\partial v}\right)\dot{\gamma} + \frac{\partial L}{\partial v}\frac{d}{dt}\left(\dot{\gamma}\right) + \frac{\partial L}{\partial t} && \text{equations of motion} \\
&= \frac{d}{dt}\left(\frac{\partial L}{\partial v}\dot{\gamma}\right) + \frac{\partial L}{\partial t} && \text{Leibnitz rule}\,.
\end{aligned}$$

Rearranging we obtain the *Beltrami identity*

$$\frac{d}{dt}\left(\frac{\partial L}{\partial v} v - L\right) = -\frac{\partial L}{\partial t}\,.$$

Here we interpret $\frac{\partial L}{\partial v} v - L$ as a function on $TX \times \mathbb{R}$, which is to be evaluated at $(\gamma(t), \dot{\gamma}(t); t)$ and then differentiated wrt t.

DEFINITION 87. The *energy of the system is $E = \frac{\partial L}{\partial v} v - L$.*

COROLLARY 88 (Conservation of energy). *Suppose $\frac{\partial L}{\partial t} = 0$, that is that L does not depend on time explicitly. Then E is a conserved quantity.*

- Observe that $E = C$ is a first-order ODE. With one degree of freedom that is the *first integral* of the equations of motion.

EXERCISE 89. Use this to solve the catenary and brachistochrone problems.

REMARK 90. We will discuss conserved quantities further in Section 3.4

3.3.2. Constraints. Paying debt.

3.3.3. Examples.

3.4. More on conserved quantities: symmetries and Noether's Theorem

3.4.1. Symmetries of configuration space.

DEFINITION 91. A *one-parameter group* is a smooth function $g\colon I \times X \to X$ (which we write $g_r(x)$ instead of $g(r,x)$) satisfying $g_0(x) = x$ and $g_{r+s} = g_r \circ g_s$.

OBSERVATION 92. $g_{-r} = g_r^{-1}$.

REMARK 93. This is a Lie group action of the Lie group $(\mathbb{R},+)$ on configuration space. Everything will make sense for an action of any (connected) Lie group, but the full theory of Lie groups is not necessary here, but just taking the one-parameter subgroups defined by generators of the Lie algebra.

EXAMPLE 94. In $\mathbb{E}^d$ fix a vector $v \in \mathbb{R}^d$ and let $g_r(x) = x + vt$. In $\mathbb{R}^2$ (i.e. fixing an origin) let $g_\theta \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$ be the rotation by θ.

We can differentiate g with respect to each variable separately. In particular g_r induces a map $TX \to TX$ (which we denote with the same symbol) via

$$g_r(x,v) = \left(g_r(x), \left(d_x g_r\right)(v)\right).$$

DEFINITION 95. We say that the one-parameter group $\{g_r\}_r$ is a *symmetry* of L or that L is *invariant* by the group if $L \circ g_r = L$ for all r.

EXAMPLE 96. Translation by a cyclic coordinate.

3.4.2. Noether's Theorem. For a fixed x, $r \mapsto g_r(x)$ is a differentiable curve. Write $g'(x) \in T_x X$ for its derivative at $x = 0$. This is a vector field on X.

LEMMA 97. *$g_r(x)$ are the integral curves of this vector field.*

PROOF. We have $g_{r+\varepsilon}(x) = g_\varepsilon\left(g_r(x)\right)$. It follows that $\frac{d}{dr} g_r(x) = g'\left(g_r(x)\right)$ so we have the (unique) solution to $\frac{dy}{dr} = g'(y)$. □

LEMMA 98. $\frac{d}{dr}\left(d_x g_r(\dot\gamma)\right) = \frac{\partial}{\partial r}\frac{\partial}{\partial t} g_r(\gamma(t)) = \frac{\partial}{\partial t}\frac{\partial}{\partial r} g_r\left(\gamma(t)\right)) = \frac{d}{dt} g'\left(g_r(x)\right)$.

THEOREM 99 (Noether; weak version). *Suppose that g_r is a symmetry. Then the quantity $\left\langle \frac{\partial L}{\partial v}, g'(x) \right\rangle$ is conserved.*

PROOF. By assumption we have $S(g_r \circ \gamma) = S(\gamma)$ for all r. We now differentiate this identity with respect to r, and use the Lemma:

$$\begin{aligned} 0 &= \frac{d}{dr} S(g_r \circ \gamma) \\ &= \frac{d}{dr} \int_{t_0}^{t_1} L\left(g_r(\gamma(t), \dot{\gamma}(t)); t\right) dt \\ &= \int_{t_0}^{t_1} \left[\left\langle \left. \frac{\partial L}{\partial x} \right|_{(g_r(\gamma(t),\dot{\gamma}(t));t)}, g'(g_r(\gamma(t))) \right\rangle + \left\langle \left. \frac{\partial L}{\partial v} \right|_{(g_r(\gamma(t),\dot{\gamma}(t));t)}, \frac{d}{dr}\left(\frac{d}{dt} g_r(\gamma(t)) \right) \right\rangle \right] dt \\ &= \int_{t_0}^{t_1} \left[\left\langle \left. \frac{\partial L}{\partial x} \right|_{(g_r(\gamma(t),\dot{\gamma}(t));t)}, g'(g_r(\gamma(t))) \right\rangle + \left\langle \left. \frac{\partial L}{\partial v} \right|_{(g_r(\gamma(t),\dot{\gamma}(t));t)}, \frac{d}{dt} g'(g_r(\gamma(t)) \right\rangle \right] dt\,. \end{aligned}$$

Finally setting $r = 0$ and integrating by parts gives the claim from

$$0 = \int_{t_0}^{t_1} \left\langle \frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial v}, g'(\gamma(t) \right\rangle dt + \left[\left\langle \frac{\partial L}{\partial v}, g'(\gamma(t)) \right\rangle \right]_{t=t_0}^{t=t_1}.$$

□

3.4.3. Total derivatives. Let $f\colon X \times \mathbb{R} \to \mathbb{R}$ be any function, and define formally the "total derivative" $\frac{df}{dt}\colon TX \times \mathbb{R} \to \mathbb{R}$ by

$$\frac{df}{dt}(x,v,t) = \left\langle \frac{\partial f}{\partial x}, v \right\rangle + \frac{\partial f}{\partial t}.$$

LEMMA 100. *Let γ be any path. Then $\frac{d}{dt} f(\gamma(t);t) = \frac{df}{dt}(\gamma(t), \dot{\gamma}(t); t)$. In particular*

$$\int_{t_0}^{t_1} \frac{df}{dt}(\gamma(t), \dot{\gamma}(t); t)\, dt = f(\gamma(t_1); t_1) - f(\gamma(t_0); t_0).$$

COROLLARY 101. *Let $\tilde{L} = L + \frac{df}{dt}$. Then for any path γ with endpoints a, b we have $\tilde{S}(\gamma) = S(\gamma) + f(b;t_1) - f(a;t_0)$ and in particular $S, \tilde{S}$ have the same critical points and the same Euler?Lagrange equations.*

EXAMPLE 102. Let $L = T - U$ be time independent, with conserved energy $E = T + U$. Then $\tilde{L} = T - U + \frac{1}{2}t^2$ has the same conserved quantity despite not being time independent.

We now generalize the previous discussion.

DEFINITION 103. A one-parameter group is a smooth family of smooth maps $g_r\colon X \times \mathbb{E}^1 \to X \times \mathbb{E}^1$ so that $g_0(x,t) = (x,t)$ and so that $g_{r+s} = g_r \circ g_s$.

We say that the one-parameter group is a *symmetry* of the Lagrangian L if $L \circ g_r - L$ is a total derivative for each r.

- Now write $\frac{d}{dr}\Big|_{r=0} g_r(x,t) = (g'(x,t), T'(x,t))$.

The following is a common generalization of the law of conservation of energy and the weak version of the theorem.

THEOREM 104 (Noether). *Suppose that $\{g_r\}_r$ is a symmetry. Then the quantity $\left\langle \frac{\partial L}{\partial v}, g'(x,t) \right\rangle - T'(x,t)\left(\frac{\partial L}{\partial v}v - L\right)$ is conserved.*

PROOF. Exercise. □

3.5. Rotations and angular momentum

3.5.1. Linear algebra. Equip $\mathbb{R}^d$ with its standard inner product and Euclidean metric, and let $\mathrm{O}(d)$ be the group of rigid motions fixing the origin.

LEMMA 105. *Each $g \in O(d)$ is linear and satisfies $g^*g = \mathrm{Id}$. Conversely $\mathrm{O}(d) = \{g \in M_d(\mathbb{R}) \mid g^*g = \mathrm{Id}\}$.*

LEMMA 106. *The (upper triangular part of the) constraints $g^*g = \mathrm{Id}$ are non-degenerate. The tangent space at the identity is $\mathfrak{so}(d) = \{X \in M_d(\mathbb{R}) : X^* + X = 0\}$.*

PROOF. Let $F(g) = g^*g$. Given a deformation $Y \in M_d(\mathbb{R})$ we have

$$\begin{aligned} F(g+Y) &= (g^* + Y^*)(g+Y) \\ &= g^*g + g^*Y + Y^*g + O(Y^2), \end{aligned}$$

so $dF_g(Y)$ is the symmetrization of g^*Y. Where g is invertible the image of this map is the space of symmetric matrices, which has the same dimension as the target space of F. □

COROLLARY 107. $\mathfrak{so}(d) = T_1\mathrm{O}(d) = \{X \in M_d(\mathbb{R}) \mid X^* + X = 0\}$; $T_g\mathrm{O}(d) = \left\{g^{-1}X \mid X \in \mathfrak{so}(d)\right\}$.

DEFINITION 108. We call elements $X \in \mathfrak{so}(d)$ *infinitesimal rotations.*

DEFINITION 109. The *matrix exponential* is given by $\exp X = \sum_{k=0}^{\infty} \frac{1}{k!} X^k$. The *matrix logarithm* is $\log g = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{1}{k} (g - \mathrm{Id})^k$.

LEMMA 110. log *converges in a small neighbourhood of the identity,* exp *converges for all X; for small enough g, X they are inverse to each other. If X, Y commute we have* $\exp(X+Y) = \exp X \exp Y$*; if g, h commute we have* $\log(gh) = \log g + \log h$.

LEMMA 111. *On the neighbourhood $V \subset \mathrm{O}(d)$ of absolute convergence* $\log\colon V \to \mathfrak{so}(d)$ *is a coordinate system with parametrization* exp.

PROOF. For $g \in \mathrm{O}(d)$ close enough to the identity we have $(\log g) + (\log g)^* = (\log g) + (\log g^*) = (\log g^*)(\log g) = \log(g^*g) = 0$ since g, g^* commute (they are inverse to each other). Conversely if $X^* = -X$ then X, X^*commute and $\exp(X)^* \exp(X) = \mathrm{Id}$.

We also remark that near the identity we have

$$\log\left(I+X+O(X^2)\right)=X+O(X^2)\,,$$

giving a different confirmation that log has full rank (and in fact its derivative is the identity). □

COROLLARY 112. *Each $X\in\mathfrak{so}(d)$ defines a one-parameter subgroup $g_r=\exp(rX)$.*

FACT 113. *Write $\mathbb{R}^d\simeq\left(\mathbb{R}^2\right)^{d/2}$ or $\left(\mathbb{R}^2\right)^{(d-1)/2}\oplus\mathbb{R}$ (orthogonal sum) depending on whether d is even or odd. For each $1\le i\le d/2$ let $X_i=\begin{pmatrix} & 1\\ -1 & \end{pmatrix}\in\mathfrak{so}(2)$ in the relevant coordinates. Then $\{\exp(r_iX_i)\}_{i\le d/2}$ is a maximal family of one-parameter subgroups; $\{\exp(\sum_{i=1}^r r_iX_i)\}$ is a maximal commutative subgroup ("maximal torus").*

3.5.2. Angular momentum. Begin with our standard Lagrangian

$$\frac{1}{2}\sum_{j=1}^{N}m_jv_j^2+U(x)\,.$$

Each $g\in\mathrm{O}(d)$ acts by matrix multiplication on the coordinate of each particle. We have $\frac{d}{dt}\left(gx_j\right)=gv_j$ so $\left|gv_j\right|^2=\left|v_j\right|^2$. Accordingly if $U(gx)=U(x)$ (g acting diagonally) we have a rotationally invariant Lagrangian. Now for each $X\in\mathfrak{so}(d)$ we obtain a one-parameter subgroup $g_r=\exp(rX)$ with $\left.\frac{d}{dr}\right|_{r=0}\exp(rX)x_j=Xx_j$. It follows that the quantity

$$\sum_{j=1}^{N}m_j\sum_{i=1}^{d}v_jXx_j$$

is conserved.

DEFINITION 114. Fix $x_0\in\mathbb{E}^d$. The *angular momentum* of a particle of mass m at position x moving at velocity $v\in\mathbb{R}^d$ is the linear functional $L\in\mathfrak{so}(d)'$ given by

$$L(X)=(x-x_0)^TXv\,.$$

EXERCISE 115. Using the basis $X_1=\begin{pmatrix} & & \\ & & 1\\ & -1 & \end{pmatrix}$, $X_1=\begin{pmatrix} & & -1\\ & & \\ 1 & & \end{pmatrix}$, $X_3=\begin{pmatrix} & 1 & \\ -1 & & 0\\ & 0 & \end{pmatrix}$ see that in 3d we recover the usual angular momentum.

COROLLARY 116. *If the potential is invariant under rotation, total angular momentum is conserved.*

3.5.3. More linear algebra. Better to think of X as a map $\mathbb{R}^n \to \mathbb{R}^{n*}$ with X^* the dual map $\mathbb{R}^n = \mathbb{R}^{n**} \to \mathbb{R}^{n*}$ required to equal $-X$. Note that this still forces $\langle Xu, u\rangle = \langle X^*u, u\rangle = -\langle Xu, u\rangle$.

LEMMA 117. *Let $u, v \in \mathbb{R}^d$ be vectors. Then the functional $X \mapsto uXv$ depends only on the plane spanned by u, v (modulu rescaling) and (if nonzero) conversely.*

PROOF. By antisymmetry $(au+bv)X(cu+dv) = (ad-bc)uXv$. Conversely let w be independent of u, v and let u^*, v^*, w^* be corresponding elements of a dual basis, $X = uw^* - wu^*$. Then $uXv = 0$ since both w^*, u^* vanish at v, On the other hand $uXw = u \neq 0$. □

PROPOSITION 118. *Given L we can find $2k \le d$ orthonormal vectors $\{u_i, v_i\}_{i=1}^k$ such that L is a linear combination of the functionals $u_i^T X v_i$.*

PROOF. Think of L as an antisymmetric matrix; find orthonormal eigenbasis invariant under complex conjugation, let u, v be the real and imaginary parts of an eigenvector. Alternatively apply Darboux's Theorem. □

- Multiparticle motion is more complicated.

3.5.4. Central potential. Suppose a single mass in $\mathbb{R}^d$ is moving in a central potential $U(x) = U(r)$ where $r = |x|$. At some particular time t either v, x are proportional to each other, and then we have a 1d problem, or they are not. By Lemma 117 and conservation of angular momentum the motion is restricted to the plane spanned by x, v. We therefore have the Lagrangian

$$\frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - U(r)\,.$$

We have two conserved quantities:

$$E = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) + U(r)$$
$$L = mr^2\dot{\theta}\,.$$

COROLLARY 119 (Kepler equal area law). *The angle is monotone; the area swept by the orbit between times t_0, t_1 is $\int_{t_0}^{t_1} r^2 d\theta = \frac{L}{m}(t_1 - t_2)$.*

Combining the two equations we get

$$E = \frac{1}{2}m\dot{r}^2 + \tilde{U}(r)$$

where $\tilde{U}(r) = U(r) + \frac{L^2}{2mr^2}$ is the "effective potential". This is a separable ODE, which (in theory) can be integrated to give $r = r(t)$. We can then determine the angle from $\dot{\theta} = \frac{L}{mr^2}$ and thus obtain the orbit.

EXAMPLE 120. If U blows up at zero slower than $\frac{1}{r^2}$ then we can't have $r \to 0$.

- Clearly the orbit is either *unbounded* (coming from infinity, to a a least radius and returning to infinity) or bounded (oscillating between $r_{\min}$, $r_{\max}$. These extrema determined by $\dot{r} = 0$, $E = \tilde{U}(r)$.
- Orbit *periodic* only if while going between extreme radii gives multiple of 2π. Note that

$$2\int_{r_{\min}}^{r_{\max}} \dot{\theta} dt = 2\int_{r_{\min}}^{r_{\max}} \frac{\dot{\theta}}{\dot{r}} dr = \frac{2L}{\sqrt{2m}} \int_{r_{\min}}^{r_{\max}} \frac{dr}{r^2 \sqrt{(E - \tilde{U}(r))}}$$

CHAPTER 4

Specific systems

4.1. Small oscillations

Let $L = \frac{1}{2}\langle M(x)v, v\rangle - U(x)$ with equation of motion

$$\frac{d}{dt}(M(x)v) = -dU\,.$$

In particular if $dU(x_0) = 0$ then $\gamma(t) \equiv x_0$ solves the equations of motion. Letting q denote the displacement $x - x_0$ we have to first order in q,

$$M(x_0)\ddot{q} \approx -H(x_0)q$$

where $H(x_0)$ is the Hessian of U at x_0, since $\left(\frac{d}{dt}M(q)\right)\dot{q} = dM \cdot \dot{q}^2$ is of second order. We are thus interested in solving the equation

$$M\ddot{q} = -Hq$$

where $q \in \mathbb{R}^n$ and M, H are symmetric positive-definite matrices (i.e. we are working near a potential *minimum*). Letting $y = \sqrt{M}q$ this takes the form

$$\ddot{y} = -\tilde{H}y$$

where $\tilde{H} = M^{-1/2}HM^{1/2}$. Suppose H is diagonable with eigenvectors ("normal modes") $\left\{q_j\right\}_{j=1}^n$, eigenvalues $\left\{\omega_j^2\right\}_{j=1}^n$. Then $\tilde{H}$ has same eigenvalues, but eigenvectors $M^{-1/2}q_j$. It follows that

$$y(t) = \Re\left(\sum_{j=1}^n A_j e^{i\omega_j t} + \sum_{j=1}^n B_j e^{-i\omega_j t}\right) M^{-1/2} q_j$$

and hence

$$q(t) = \Re\left(\sum_{j=1}^n A_j e^{i\omega_j t} + \sum_{j=1}^n B_j e^{-i\omega_j t}\right) M^{-1} q_j\,.$$

EXAMPLE 121. N equal masses connected by identical springs, say with x_0 pinned, q_j the displacement from equillibrium of the jth mass. Then $U = \frac{1}{2}k\sum_{j=1}^{N}\left(q_j - q_{j-1}\right)^2$ so

$$H = \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 1 \end{pmatrix}$$

4.2. Rigid body motion

4.2.1. Kinematics; the inertia tensor.

DEFINITION 122. A *rigid body* is a system of masses subjected to the (holonomic!) constrain that all pairwise distances remain fixed. We assume, moreover, that the masses are in *general position*.

LEMMA 123. *Let $A, B \subset \mathbb{E}^d$ be subsets, and let $f\colon A \to B$ be a distance preserving bijection (for all $a, a' \in A$ we have $|f(a) - f(a')| = |a - a'|$. Then there exist an isometry $\tilde{f} \in \mathrm{Isom}(\mathbb{E}^d)$ such that $f = \tilde{f}\restriction_A$. The map $\tilde{f}$ is unique iff A is in* general position *in that its affine hull is all of $\mathbb{E}^d$.*

COROLLARY 124. *The configuration of a rigid body relative to a reference embedding is determined by an element of the Euclidean group. In particular, a rigid body has $d + \frac{d(d-1)}{2}$ degrees of freedom regardless of its number of particles (6 if $d = 3$).*

Concretely, fix point O on the body; in a particular configuration we can identify the points P of the body with their displacements $a = P - O$ relative to O. Suppose we now rotate the body by $g \in \mathrm{SO}(d)$ and then translate it by x. The point P is then located at ("parametrization")

$$X = x + ga\,.$$

If the motion is along the path $x = x(t)$, $g = g(t)$ we have

$$V = \frac{dX}{dt} = \dot{x} + \dot{g}\cdot a = v + \dot{g}a$$

where $v = \dot{x}$is the velocity of O, and $\dot{g} \in T_g\mathrm{SO}(d)$.

DEFINITION 125. The *angular velocity* of the body is $\Omega = \dot{g}g^{-1} \in T_1\mathrm{SO}(d) = \mathfrak{so}(d)$. Equivalently $\dot{g} = \Omega g$.

In those terms

$$V = v + \Omega g a = v + \Omega x$$

LEMMA 126. *The angular velocity is independent of the choice of origin and initial orientation.*

PROOF. If we rotate the reference frame by h about O, the position of P is given by $a' = ha$, so

$$X = x + g'h^{-1}a\,.$$

We thus have $g' = gh$ and hence $\Omega' = \dot{g}'(gh)^{-1} = \dot{g}hh^{-1}g^{-1} = \Omega$.

If we replace O with O' then the absolute position of O' is $x' = x + gb$ where $b = O' - O$. The position of P is given by $a' = a - b$

$$X = x + ga' + gb = x' + g'a'$$

where $g' = g$. □

Let us now compute the kinetic energy of the entire rigid body. Writing $M = \sum_j m_j$ for the total mass we have

$$\begin{aligned} T &= \frac{1}{2}\sum_j m_j V_j^2 = \frac{1}{2}\sum_j m_j \left\langle v + \Omega g a_j, v + \Omega g a_j \right\rangle \\ &= \frac{1}{2}Mv^2 + \left\langle v, \Omega g \sum_j m_j a_j \right\rangle + \frac{1}{2}\sum_j m_j \left\langle \Omega g a_j, \Omega g a_j \right\rangle\,. \end{aligned}$$

In the second term, $\sum_j m_j a_j = M\bar{a}$ where $\bar{a} = \sum_j \frac{m_j}{M} a_j$ is the location of the *centre of mass* in body coordinates. We thus have

$$\left\langle v, \Omega g \sum_j m_j a_j \right\rangle = M \left\langle v, \Omega g \bar{a} \right\rangle = M \left\langle v, \Omega(\bar{X} - x) \right\rangle$$

where $\bar{X}$is the location of the centre of mass. For the second term let $r_j = g a_j$ be the vector from x to the position of the jth particle. Then that term s

$$\begin{aligned} \frac{1}{2}\sum_j m_j \left\langle \Omega r_j, \Omega r_j \right\rangle &= \frac{1}{2}\sum_j m_j r_j^T \Omega^T \Omega r_j \\ &= \frac{1}{2}\sum_j m_j \operatorname{Tr}\left(r_j r_j^T \Omega^T \Omega\right) \\ &= \frac{1}{2}\operatorname{Tr}\left(I\Omega^T\Omega\right) \end{aligned}$$

where $I = \sum_j m_j r_j r_j^T$ is the *tensor of inertia.*

REMARK 127. Note that if we define $I_0 = \sum_j m_j a_j a_j^T$ then $I = gI_0 g^T = gI_0 g^{-1}$, encoding the fact that the tensor of inertia rotates with the body. Alternatively we have

$$\begin{aligned} \frac{1}{2}\operatorname{Tr}\left(I\Omega^T\Omega\right) &= \frac{1}{2}\operatorname{Tr}\left(I_0 g^T \Omega^T \Omega g\right) \\ &= \frac{1}{2}\operatorname{Tr}\left(I_0 \dot{g}^T \dot{g}\right)\,. \end{aligned}$$

Naturally we choose O to be the centre of mass, in which case we have the kinetic term

$$T = \frac{1}{2}Mv^2 + \frac{1}{2}\operatorname{Tr}\left(I\Omega^T\Omega\right)$$

Assuming the forces depend only on the configuration, we then have a corresponding *angular momentum*

$$J = \operatorname{Tr}\left(I_0 v^T \dot{g} + I_0 \dot{g}^T v\right) = \frac{1}{2}\left(\dot{g}^T I_0 + I_0 \dot{g}^T\right),$$

thought of as a functional on $T_g\mathrm{SO}(d)$. Equivalently this reads ($\dot{g}^T = g^T\Omega^T$)

$$J = \frac{1}{2}\left(\Omega I + I\Omega\right) = I\Omega$$

as a functional on $\mathfrak{so}(d)$ (since $\mathfrak{so}(d)$ consists of antisymmetric matrices, both define the same functional).

EXAMPLE 128. Free body. The equation of motion is then conservation of angular momentum:

$$0 = \frac{dJ}{dt} = I\dot{\Omega} + \Omega J$$

In coordinates:

$$I_1\dot{\omega}_1 + (I_3 - I_2)\omega_2\omega_3 = 0$$

plus cyclic. Now give the body rotating around the third axis a small "displacement". We can ignore $\omega_1\omega_2$ so initially ω_3 remains constant and the result is a linear equation for (ω_1, ω_2).

4.2.2. Rotating reference frames.

CHAPTER 5

Hamiltonian mechanics

Another formulation of mechanics, with a different (but equivalent) state space and equations of motion. Key point: *convexity* of the quadratic kinetic term in the velocity.

5.1. The Legendre transformation

5.1.1. Convexity. Fix a real vector space V.

DEFINITION 129. A subset $C \subset V$ is *convex* if C contains the interval connecting any two points in it. If C is convex a function $f\colon C \to \mathbb{R}$ is *convex* if for all $x, y \in C$ and $t \in [0,1]$, $f((1-t)u+v) \le (1-t)f(u) + tf(v)$. Say that f is *strictly* convex if the inequality is strict whenever $u \neq v$ and $t \neq 0, 1$.

EXERCISE 130. f is convex iff its *epigraph* $\{(v,a) : f(v) \le a\} \subset V \times \mathbb{R}$ *is convex.*

We will allow our functions to take the value ∞, as long as they are not identically infinite. The *effective domain* of an extended convex function is the set $\{v \mid f(v) < \infty\}$. The convex function is *closed* if $\{v \mid f(v) \le a\}$ is closed for every a (this is equivalent to lower semicontinuity).

EXERCISE 131. For $I \subset \mathbb{R}$ let $f\colon I \to \mathbb{R}$ be convex.

(1) Then f has a (possibly infinite) right and left derivative at every interior point; if $u \le v$ then $f'_L(u) \le f'_R(u) \le f'_L(v) \le f'_R(v)$.
(2) There are at most countably many points where f' is not differentiable.
(3) If f is strictly convex then $f'_R(u) < f'_L(v)$ whenever $u < v$.

COROLLARY 132. *We can parametrize (almost all of) the points on the graph by the slopes of the tangent lines instead. This is a mapping $I \to \mathbb{R}^*$ given by $x \mapsto f'(x)$.*

DEFINITION 133. The *Fenchel or convex conjugate* (or the *Legendre Transform*) of the function f on $\mathbb{R}^n$ is the function f^* on $(\mathbb{R}^n)^*$ given by

$$f^*(p) = \sup_{x \in \mathbb{R}^n} \langle p, v \rangle - f(v).$$

EXAMPLE 134. The convex conjugate of

(1) $f(x) = \langle m, x \rangle + b$ is $f^*(p) = \begin{cases} b & p = m \\ \infty & p \neq m \end{cases}$
(2) For $1 < r < \infty$ $f(x) = \frac{1}{r}|x|^r$ is $\frac{1}{r^*}|p|^{r^*}$ where r^* is the dual exponent $\frac{1}{r} + \frac{1}{r^*} = 1$.

(3) Let $g\colon V\to V^*$ be a symmetric and positive -definite, and let $f(x)=\frac{1}{2}\langle gx,x\rangle+\langle m,x\rangle+b$. Then $f^*(p)=\frac{1}{2}\left\langle p-m,g^{-1}(p-m)\right\rangle+b$. Indeed given $x\in V$ let $h=g^{-1}(p-m)$. Then

$$\begin{aligned}\langle p,x\rangle-f(x)&=\langle p-m,x\rangle-\frac{1}{2}\langle gx,x\rangle-b\\&=-\frac{1}{2}\langle g(x-h),(x-h)\rangle+\frac{1}{2}\langle gh,h\rangle-b\end{aligned}$$

which is maximized at $x=h$.

LEMMA 135. *Let f be any function*

(1) *f^* is l.s.c. and convex.*
(2) *If $f\le g$ then $g^*\le f^*$.*
(3) *$\{(v,a)\,a\ge f^{**}(v)\}$ is the closed convex hull of $\{(v,a)\,a\ge f(v)\}$.*

PROOF. Given p,ε let v be such that $\langle p,v\rangle-f(v)\ge f^*(p)-\varepsilon$. Then for all p' we have

$$\begin{aligned}f^*(p')&\ge\left\langle p',v\right\rangle-f(v)=\langle p,v\rangle-f(v)+\left\langle p'-p,v\right\rangle\\&\ge f^*(p)-\varepsilon+\left\langle p'-p,v\right\rangle.\end{aligned}$$

It follows that for all p' which are enough to p we have $f^*(p')\ge f^*(p)-\varepsilon$, equivalently $f^*(p)\le f^*(p')+\varepsilon$. It follows that if $p_i\to p$ and $f(p_i)\le a$ then $f(p)\le a$ as well.

For any point $v\in V$ we have

$$\begin{aligned}\langle p,v\rangle-f(v)&\le f^*(p)\\\langle q,v\rangle-f(v)&\le f^*(q).\end{aligned}$$

It then follows that

$$\langle(1-t)p+tq,v\rangle-f(v)\le(1-t)f^*(p)+f^*(q).$$

Taking the supremum over v gives

$$f^*\left((1-t)p+tq\right)\le(1-t)f^*(p)+f^*(q).$$

□

THEOREM 136 (Fenschel–Moreau). *Let f be proper convex and lsc (equivalently closed convex). Then $f^{**}=f$.*

PROOF. A closed convex set (here the epigraph of f) is the intersection of the halfspaces containing it. □

Suppose now that f is strictly convex and differentiable. Then the function $u\mapsto\langle df_v,u\rangle-f(u)$ has a critical point at v. By strict convexity this is the unique maximum of the function, and we get $f^*(df_v)=\langle df_v,v\rangle-f(v)$.

5.1.2. From the Lagrangian to the Hamiltonian.

DEFINITION 137. *Phase space* also known as the *cotangent bundle* is $T^*X = \{(x,p) \mid x \in X,\, p \in (T_xX)^*\}$.

We now fix a Lagrangian $L\colon TX \times \mathbb{E}^1 \to \mathbb{R}$. We assume that $v \mapsto L(x,v;t)$ is strictly convex for fixed x,t.

DEFINITION 138. The *Hamiltonian* of the system is the function $H\colon T^*X \times \mathbb{E}^1 \to \mathbb{R}$ so that for fixed x,t we have $p \mapsto H(x,p;t)$ is the convex conjugate of $v \mapsto L(x,v;t)$.

To any *state of motion* $(x,v;t)$ we associate a *phase point* $(x,p;t)$ by $p = \frac{\partial L}{\partial v}\big|_{(x,v;t)}$. By the observation above we have

$$H(x,p;t) = \langle p,v\rangle - L(x,v,t)$$

whenever $p = p(x,v;t)$.

OBSERVATION 139. *When L is time-independent this is exactly the conserved energy from Beltrami's identity, but now thought of as a function of x,p rather than x,v.*

EXAMPLE 140. If $L(x,v) = \frac{1}{2}\langle M(x)v,v\rangle - U(x)$ then $p = M(x)v$ so $v = M(x)^{-1}p$ and

$$\begin{aligned} H(x,p) &= \langle p,v\rangle - L \\ &= \langle M(x)v,v\rangle - \frac{1}{2}\langle M(x)v,v\rangle + U(x) \\ &= \frac{1}{2}\left\langle p, M(x)^{-1}p\right\rangle + U(x)\,. \end{aligned}$$

From $L^{**} = L$ we also obtain that $\frac{\partial H}{\partial p} = v$.

5.2. The Hamiltonian flow

5.2.1. Hamilton's equations. From $\frac{\partial H}{\partial p} = v$ we see that along the physical path we have $\dot{x} = v = \frac{\partial H}{\partial p}$. What about $\dot{p}$? The Euler–Lagrange equation reads

$$\dot{p} = \frac{\partial L}{\partial x}$$

and we would like to understand what this means in terms of H. For this think of p and H, for the moment, as functions on $TX \times \mathbb{E}^1$. We then have

$$\begin{aligned} dH &= \langle dp,v\rangle + \langle p,dv\rangle - \left\langle \frac{\partial L}{\partial v}, dv\right\rangle - \left\langle \frac{\partial L}{\partial x}, dx\right\rangle - \frac{\partial L}{\partial t}dt \\ &= \langle dp,v\rangle - \left\langle \frac{\partial L}{\partial x}, dx\right\rangle - \frac{\partial L}{\partial t}dt\,. \end{aligned}$$

Now push everything forward to $T^*X \times \mathbb{E}^1$ by the identification. We see that

$$\begin{cases} \frac{\partial H}{\partial x} = -\frac{\partial L}{\partial x} \\ \frac{\partial H}{\partial p} = v \\ \frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t} \, . \end{cases} \tag{5.2.1}$$

It follows that, along the physical path

$$\begin{cases} \dot{x} = \frac{\partial H}{\partial p} \\ \dot{p} = -\frac{\partial H}{\partial x} \, . \end{cases}$$

DEFINITION 141. These are *Hamilton's equations*. The solution curves, that is the function $F_{s,t}\colon T^*X \to T^*X$ mapping the initial condition $(z,s) \in T^*X \times \mathbb{E}^1$ to the location of the solution at time t, is called the *Hamiltonian flow* (associated to the Hamiltonian $H \in C^\infty(T^*X)$.

- Clearly $F_{t,r} \circ F_{s,t} = F_{s,r}$.
- If H does not depend on time then $F_{s,t}$ only depends on $t-s$ and we simply write the flow as a one-parameter group F_{s-t}.

REMARK 142. **WARNING**. Equation (5.2.1) must be interpreted very carefully. On the LHS we take partial derivatives where x, p are fixed. On the right where x, v are fixed. Those are not the same!

One conclusion is that, along the physical path, we have

$$\dot{H} = \frac{\partial H}{\partial x}\dot{x} + \frac{\partial H}{\partial p}\dot{p} + \frac{\partial H}{\partial t} = \frac{\partial H}{\partial t} \, .$$

When the Lagrangian (and the Hamiltonian) does not depend explicitly on time we again obtain the Beltrami identity $\dot{H} = 0$.

EXAMPLE 143. Consider the harmonic oscillator $L = \frac{1}{2}mv^2 - \frac{1}{2}kx^2$. With Euler–Lagrange equation $m\ddot{x} = -kx$. Here $p = \frac{\partial L}{\partial v} = mv$, so $H = pv - L = \frac{1}{2}mv^2 + \frac{1}{2}kx^2 = \frac{1}{2m}p^2 + \frac{1}{2}kx^2$, and we get

$$= \begin{pmatrix} & p/m \\ -k & \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} \begin{cases} \dot{x} = \frac{p}{m} \\ \dot{p} = -kx \end{cases} \iff \frac{d}{dt}\begin{pmatrix} x \\ p \end{pmatrix} = \begin{pmatrix} & 1/m \\ -k & \end{pmatrix} \begin{pmatrix} x \\ p \end{pmatrix} .$$

We therefore have

$$\begin{pmatrix} x \\ p \end{pmatrix} = \exp\left(\begin{pmatrix} & 1/m \\ -k & \end{pmatrix} t \right) \begin{pmatrix} x_0 \\ p_0 \end{pmatrix} .$$

EXAMPLE 144. In the central potential $L = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - U(r)$ we have $p_r = m\dot{r}$ and $p_\theta = J = mr^2\theta$ (the angular momentum!) so

$$\begin{aligned} H &= p_r\dot{r} + p_\theta\dot{\theta} - L \\ &= \frac{1}{2}m\dot{r}^2 + \frac{1}{2}mr^2\dot{\theta}^2 + U(r) \\ &= \frac{1}{2m}p_r^2 + \frac{1}{2mr^2}p_\theta^2 + U(r). \end{aligned}$$

The equations of motion as therefore

$$\begin{cases} \dot{r} = \frac{1}{m}p_r \\ \dot{p}_r = -\frac{dU}{dr} - \frac{J^2}{mr^3} \\ \dot{\theta} = \frac{1}{mr^2}J \\ \dot{J} = 0 \end{cases}$$

We therefore see that J is a *conserved quantity* (nice to have it as a coordinate!), that $\theta = \theta_0 + \frac{J}{m}\int_0^t \frac{d\tau}{r^2}$ and that it remains to determine $r(t)$ from the first two equations.

5.2.2. The symplectic structure.

DEFINITION 145. An *observable* is a smooth function $A \in C^\infty(T^*X \times \mathbb{E}^1)$.

Let A be an observable. Then along the physical path we have

$$\begin{aligned} \dot{A} &= \frac{\partial A}{\partial x}\dot{x} + \frac{\partial A}{\partial p}\dot{p} + \frac{\partial A}{\partial t} = \frac{\partial A}{\partial x}\frac{\partial H}{\partial p} - \frac{\partial A}{\partial p}\frac{\partial H}{\partial x} + \frac{\partial A}{\partial t} \\ &= \omega(dA, dH) + \frac{\partial A}{\partial t} \end{aligned}$$

where $\omega_{(x,p)}$ is the bilinear form on $T^*_{(x,p)}T^*X$ with the matrix $J = \begin{pmatrix} & I_n \\ -I_n & \end{pmatrix}$. Note that ω is non-degenerate and antisymmetric. Since it is constant in those coordinates we also have $d\omega = 0$, that is ω is closed.

DEFINITION 146. A *symplectic manifold* is a pair (M, ω) where M is a manifold and ω is a non-degenerate closed 2-form on M.

EXAMPLE 147. Let X be any manifold, and let $M = T^*X$ be the cotangent bundle. Let $\pi\colon T^*X \to X$ be the natural projection, $d\pi\colon T_{(x,p)}M \to T_xX$ its derivative. Observe that $p \in (T_xx)^*$ so $d\pi^*_{(x,p)}(p) \in T^*_{(x,p)}M$. Setting $\theta_{(x,p)} = d\pi^*_{(x,p)}(p)$ gives a section of the cotangent bundle of M called the *tautological 1-form*. This is canonical: if $f\colon X \to Y$ is a diffeomorphism then $f^*\theta^Y = \theta^X$.Now let $\omega = -d\theta$ (exterior derivative). By duality we can think of ω also as a map $TM \to T^*M$ and also as a bilinear form on T^*T^*M as above.

DEFINITION 148. A coordinate system $\{q_i, p_i\}_{i=1}^n$ on M with respect ti which ω has the constant matrix J (i.e. $\omega = \sum_{i=1}^n dq_i \wedge dp_i$) is called a *canonical coordinate system*.

EXAMPLE 149. If $M = T^*X$ then $(q_i)_{i=1}^n$ can be any coordinate system on X, and (p_i) can be the associated momenta with respect to any Lagrangian.

THEOREM 150 (Darboux). *A symplectic manifold always has local canonical coordinate systems.*

DEFINITION 151. The *Poisson bracket* of the observables A, B is the observable $\{A, B\} = \omega(dB, dA)$. The *Hamiltonian vector field* corresponding to A is the vector field $X_A = \omega^{-1}(dA)$, equivalently $\langle V, dA\rangle = \omega(X_A, V)$ for every vector field V.

In this language Hamilton's equations read $\dot{z} = X_H(z)$ and for time-independent observables we have $\dot{A} = \{H, A\}$.

DEFINITION 152. A map $F\colon T^*X \to T^*X$ that preserves the symplectic form is called a *symplectomorphism* or a *canonical transformation.*

LEMMA 153. *The Hamiltonian flow is a symplectomorphism.*

PROOF. Direct calculation.

$$\begin{aligned}
\frac{d}{dt}(F_t^*\omega) &= F_t^*(L_{X_H}\omega) && \text{group property} \\
&= F_t^*(i_{X_H}d\omega + d(i_{X_H}\omega)) && \text{Cartan's formula} \\
&= F_t^*(0 + d(dH)) && d\omega = 0 \\
&= 0\,.
\end{aligned}$$

□

COROLLARY 154 (Liouville's Theorem). *The flow preserves volume.*

- We will discuss this further in Section 5.3.

5.3. Interlude: dynamics and ergodic theory

5.4. Integrable systems and the KAM theorem

Suppose that $\dim X = n$, so $M = T^*X$ is $2n$-dimensional, and fix a Hamiltonian $H \in C^\infty(M)$.

DEFINITION 155. We call the Hamiltonian *integrable* if there are further constants of the motion $\{J_i\}_{i=2}^n$ so that, together with $J_1 = H$ we have $\left\{J_i, J_j\right\} = 0$ and that, on a level set $M_{\underline{a}} = J^{-1}(\underline{a})$ the$\{dJ_i\}_{i=1}^n$ are linearly independent.

- Since J_i are constants of the motion, $M_{\underline{a}}$ is invariant by the Hamiltonian flow (in fact invariant by the n-dimensional group of flows generated by all the J_i!).
- By the implicit function theorem, $M_{\underline{a}} \subset M$ is a smooth n-dimensional submanifold.

- Since $\langle dJ_i, \omega^{-1}(dJ_j)\rangle$ is the derivative of J_i along the Hamiltonian flow associated to dJ_j we see that $\omega^*(dJ_i, dJ_j) = 0$ (or equivalently $\omega(X_i, X_j) = 0$ for the corresponding vector fields, which are tangent to $M_{\underline{a}}$). We say $M_{\underline{a}} \subset M$ is a *Lagrangian submanifold*.

ASSUMPTION 156. *Suppose $M_{\underline{a}}$ is compcat (e.g. because energy surfaces are compact).*

PROPOSITION 157. *Each connected component of $M_{\underline{a}}$ is a torus.*

PROOF. Let $\Phi^{(i)}_{t_i}$ be the 1-parameter group generated by J_i. Since these commute we can for $\underline{t} \in \mathbb{R}^n$ and $z \in C$ set

$$\Phi_{\underline{t}}(z) = \Phi^{(1)}_{t_1} \cdots \Phi^{(n)}_{t_n}(z)$$

and get an $\mathbb{R}^n$-action. Each orbit $\left\{\Phi_{\underline{t}}(z)\right\}_{\underline{t}\in\mathbb{R}^n}$ is open (by the inverse function theorem, since the vector fields are independent and C is n-dimensional), and the orbits are disjoint, so each orbit is closed, hence a connected component. In addition each orbit is compact, hence has the form $\mathbb{R}^n/\Lambda$ where Λ is a lattice in $\mathbb{R}^d$, that is is a torus. □

Now let $\{\gamma_i\}_{i=1}^n \subset \Lambda$ be a basis; on the torus $M_{\underline{a}}$ these are cycles which generate the homology. Let $I_i = \oint_{\gamma_i} p dq$. If γ_i' is another cycle on $M_{\underline{a}}$ homologous to γ_i then the integral remains the same, because if $\gamma_i' - \gamma_i = dA$ for a 2-cycle $A \subset M$ then by Stokes

$$\oint_{\gamma'-\gamma} p dq = -\int_A dp \wedge dq = -\int_A \omega = 0$$

since $\omega \equiv 0$ on $M_{\underline{a}}$. Now the H-flow on $M_{\underline{a}}$ maps the cycles to homologous cycles, and it follows that the I_i are constants of the motion.

ASSUMPTION 158. *Call the torus $M_{\underline{a}}$* non-degenerate *if the $\{I_i\}_{i=1}^n$ are independent.*

We can now replace the J_i with the I_i. Let $\{\theta_i\}$ be the conjugate positions. Then Hamilton's equations give $\frac{dI}{dt} = -\frac{\partial H}{\partial \theta}$, so H is independent of the θ_i and is only a function of the *action* variables I_i. It follows that $\frac{d\theta_i}{dt} = \frac{\partial H}{\partial I_i} = \omega_i(\underline{a})$, so θ_i are essentially angles.

THEOREM 159 (Kolmogorov–Arnol'd–Moser). *Suppose that the frequency vector $\underline{\omega}(\underline{a})$ is sufficiently irrational. Then under small perturbations of H the torus $M_{\underline{a}}$ deforms smoothly.*

5.5. Hamilton–Jacobi Theory

Fix an initial condition $x_0 \in X$ and set $S(x;t)$ as the action for the physical path that reaches $x \in X$ at time t starting from x_0. Suppose we vary the final endpoint x; then the physical path γ will change by $\delta\gamma$ and to first order we have

$$\begin{aligned} \delta S &= \int_{t_0}^{t} \left[\frac{\partial L}{\partial x}\delta\gamma + \frac{\partial L}{\partial v}\delta\dot{\gamma}\right] dt \\ &= \int_{t_0}^{t} \left[\frac{\partial L}{\partial x} - \frac{d}{dt}\frac{\partial L}{\partial v}\right]\delta\gamma dt + \left[\frac{\partial L}{\partial v}\delta\gamma\right]_{t_0}^{t_1}. \end{aligned}$$

Now $\delta\gamma(t_0) = 0$ (we fixed the initial condition) and $\delta\gamma(t) = \Delta x$, the change in the endpoint. Between the two the the physical path satisfies the Euler–Lagrange equations, and we get:

$$\frac{\partial S}{\partial x} = \frac{\partial L}{\partial v} = p\,.$$

Now let us work along the physical path. By definition we have

$$\frac{dS}{dt} = L\,.$$

By the chain rule we also have

$$\frac{dS}{dt} = \frac{\partial S}{\partial t} + \frac{\partial S}{\partial x}\frac{dx}{dt}\,.$$

It follows that

$$\frac{\partial S}{\partial t} = L - pv = -H\,.$$

Equivalently, we have

$$\frac{\partial S}{\partial t} + H(x,p;t) = 0\,.$$

But $p = \frac{\partial S}{\partial x}$, and we obtain a nonlinear first-order PDE

$$\frac{\partial S}{\partial t} + H\left(x, \frac{\partial S}{\partial x}; t\right) = 0\,.$$

This is the *Hamilton–Jacobi equation*. Solving this equation amounts to solving the equations of motion: suppose we can find a solution $S = S(x;t;\underline{a})$ where $\underline{\alpha}$ are constants of integration (functions of the initial condition x_0). There are nominally $n+1$ of them but one will be an overall additive constant since the equations only involve derivatives of S,and we ignore it.

Define $\beta_i = \frac{\partial S}{\partial \alpha_i}$ and let (α, β) be coordinates at the point (x, p) where $p = \frac{\partial S}{\partial x}$. This is a symplectomorphism (check!). Conversely solving for x, p as functions of (α, β) gives the solution of the system.

EXAMPLE 160. The harmonic oscillator. Here $H = \frac{p^2}{2m} + \frac{1}{2}kx^2$. We then need to solve the PDE

$$\frac{\partial S}{\partial t} + \frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + \frac{1}{2}kx^2 = 0\,,$$

Since the Hamiltonian does not depend explicitly on time we try an Ansatz $S = W(x) - \alpha t$. We then obtain the equation

$$\frac{1}{2m}(W')^2 + \frac{1}{2}kx^2 = a$$

and we can see that α is the total energy E. We then have

$$W' = \sqrt{2m\alpha - mkx^2}\,.$$

We therefore have

$$W = \int \sqrt{2m\alpha - mkx^2}dx$$

and

$$S = \int \sqrt{2m\alpha - mkx^2}dx - \alpha t\,.$$

We can compute the integral directly, but we really want the change of coordinates. We will have

$$\beta = \frac{\partial S}{\partial \alpha} = \sqrt{\frac{m}{2\alpha}} \int \frac{dx}{\sqrt{1 - \frac{kx^2}{2\alpha}}} - t = \frac{1}{\omega} \arcsin\left(\sqrt{\frac{k}{2\alpha}}x\right) - t$$

where $\omega = \sqrt{k/m}$. We thus obtain

$$x = \sqrt{\frac{2\alpha}{k}} \sin\left(\omega t + (\omega\beta)\right) = \sqrt{\frac{2\alpha}{k}} \sin\left(\omega t + \phi\right)\,.$$

We can also compute $p = \frac{\partial S}{\partial x} = \frac{dW}{dx} = \sqrt{2m\alpha - mkx^2} = \sqrt{2m\alpha}\cos\left(\omega t + \phi\right)$. Finally since $\frac{1}{2m}p^2 + \frac{1}{2}kx^2 = \alpha$ we see that α is indeed the energy (which can be computed from the initial conditions), and

$$\tan\phi = \sqrt{km}\frac{x_0}{p_0}\,.$$

CHAPTER 6

Connections to Quantum Mechanics

6.1. Introduction: The dictionary

Classical particles occupy *points* in phase space. Quantum mechanical particles are *smeared*. We will discuss the underlying mathematics. Let us start with a dictionary

	Classical	Quantum
State space	$M = T^*X$	$L^2(X)$
Observables	$a \in C^\infty(M)$	Selfadjoint linear operators $\hat{A}\colon L^2(X) \circlearrowleft$
Observable algebra	$\{a,b\}$	$\frac{1}{i\hbar}\left[\hat{A},\hat{B}\right] = \frac{1}{i\hbar}\left(\hat{A}\hat{B}-\hat{B}\hat{A}\right)$
Time evolution	$\frac{dz}{dt} = -\omega^{-1}(dH)$	$i\hbar\frac{d}{dt}\Psi = \hat{H}\Psi$
Flow group	$z(t) = \Phi_t(z)$	$\Psi(t) = U(t)\Psi(0)$, $\frac{d}{dt}U(t) = i\hbar U(t)\hat{H}$
Evolution of observables	$\frac{da}{dt} = \{H,a\}$	$\frac{dA}{dt} = -i\hbar\,[H,A]$ (***)
Ensemble	measure μ on M	Positive trace class Hermitian operator on $L^2(X)$ of trace 1.

Let's remember what time evolution of observables means:

- Classical: we use the initial condition z as the coordinate for all times, so we the observable $a_t = a\circ\Phi_t$ becomes time-dependent.
- Quantum: we use the initial state Ψ as the state for all times. Then the observables must evolve so that $\hat{A}(t)\Psi$ is the state so that at time t it would be $\hat{A}\Psi(t)$. In other words $U(t)\hat{A}(t)\Psi = \hat{A}U(t)\Psi$ and hence

  $$\hat{A}(t) = U(-t)\hat{A}U(t)\,.$$

  Then

  $$\begin{aligned}\frac{d\hat{A}(t)}{dt} &= -i\hbar U(-t)\hat{H}\hat{A}U(t) + i\hbar U(-t)\hat{A}\hat{H}U(t)\\ &= -i\hbar\left[\hat{H},\hat{A}(t)\right]\end{aligned}$$

Treating the Schrödinger equation as an ODE in the Hilbert space $L^2(X)$, we see that it is a constant-coefficient linear equation, so can be solved by separating variables via the spectral decomposition of $\hat{H}$. For example if $\{\phi_n\}_{n=0}^\infty \subset L^2(X)$ is an o.n.b. of eigenfunctions of $\hat{H}$ with $\hat{H}\phi_n = E_n\phi_n$ then

$$\Psi(t) = \sum_{n=0}^\infty a_n e^{\frac{E_n}{i\hbar}t}\phi_n$$

is a solution to the Schrödinger equation. It thus suffices to solve the eigenvalue equation, the so-called *time-independent* Schrödinger equation

$$\hat{H}\phi_n = E_n\phi_n .$$

- The term "quantum", that is "unit" of somethingm comes exactly from the discreteness of the spectrum of $\hat{H}$.

6.2. Canonical ("algebraic") quantization

6.2.1. Formalities. Fix $X = \mathbb{R}^d$, so that $T^*X = \left\{(x,p) \mid x \in \mathbb{R}^d, p \in (\mathbb{R}^d)^*\right\}$. To each observable we want to associate an operator.

- Classical: $\left\{x_i, p_j\right\} = \delta_{ij}$.
- Quantum: we want operators $\hat{x}_i$, $\hat{p}_j$ so that $\left[\hat{x}_i, \hat{p}_j\right] = i\hbar$.
  - Let $\hat{x}_j$ be multiplication by x
  - Let $\hat{p}_j$ be $-i\hbar\frac{\partial}{\partial x_j}$.
  - Then

$$\begin{aligned}\left(\hat{x}_i\hat{p}_j - \hat{p}_j\hat{x}_i\right)\Psi(x) &= -i\hbar\left(x_i\frac{\partial\Psi}{\partial x_j} - \frac{\partial(x_i\Psi)}{\partial x_j}\right) \\ &= i\hbar\frac{\partial x_i}{\partial x_j}\Psi \\ &= \left(i\hbar\delta_{ij}\right)\Psi .\end{aligned}$$

Dirac suggested that quantization should work by a map Op_h so that $\mathrm{Op}_h(\{a,b\}) = \frac{1}{i\hbar}\left[\mathrm{Op}_h(a), \mathrm{Op}_h(b)\right]$.

LEMMA 161. $\{ab,c\} = a\{b,c\} + \{a,c\}b$. $\left[\hat{A}\hat{B}, C\right] = \hat{A}\left[\hat{B},\hat{C}\right] + \left[\hat{A},\hat{C}\right]\hat{B}$.

COROLLARY 162. *Knowing $\{x,p\}$ and $[\hat{x},\hat{p}]$ determines the algebraic structure of the Poisson / commutator brackets for all commutative / noncommutative polynomials.*

THEOREM 163 (Groenewold 1946). *There is no map Op_h mapping polynomials in $\underline{x}, \underline{p}$ to noncommutative polynomials in $\hat{x}, \hat{p}$ which is: (1) unital; (2) linear; (3) respects brackets as above.*

- Concretely the problem is: should we have $\widehat{xp} = \hat{x}\hat{p}$,$\widehat{xp} = \hat{p}\hat{x}$, $\widehat{xp} = \frac{\hat{x}\hat{p}+\hat{p}\hat{x}}{2}$, or something else?

6.2.2. The Heisenberg group and algebra. Let $H = \left\{\begin{pmatrix}1 & a & c \\ & 1 & b \\ & & 1\end{pmatrix}\right\} \subset \mathrm{GL}_3(\mathbb{R})$. This closed subgroup is called the "Heisenberg group" for reasons that are about to become apparent.

The tangent space at the origin consists of the matrices $\left\{\begin{pmatrix} 0 & x & z \\ & 0 & y \\ & & 0 \end{pmatrix}\right\}$ and we can check that

$$\exp\begin{pmatrix} 0 & x & z \\ & 0 & y \\ & & 0 \end{pmatrix} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} + \begin{pmatrix} 0 & x & z \\ & 0 & y \\ & & 0 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 0 & x & z \\ & 0 & y \\ & & 0 \end{pmatrix}^2 = \begin{pmatrix} 1 & x & z+\frac{1}{2}xy \\ & 1 & y \\ & & 1 \end{pmatrix}.$$

Consider the infinitesimal generators $X = \begin{pmatrix} 0 & 1 & 0 \\ & 0 & 0 \\ & & 0 \end{pmatrix}$, $Y = \begin{pmatrix} 0 & 0 & 0 \\ & 0 & 1 \\ & & 0 \end{pmatrix}$. We have $[X,Y] = \begin{pmatrix} 0 & 1 & 0 \\ & 0 & 0 \\ & & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 & 0 \\ & 0 & 1 \\ & & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ & 0 & 1 \\ & & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 & 0 \\ & 0 & 0 \\ & & 0 \end{pmatrix} = Z = \begin{pmatrix} 0 & 0 & 1 \\ & 0 & 0 \\ & & 0 \end{pmatrix}$ where Z is central.

Now consider $L^2(\mathbb{R})$ and let the 1-parameter group $\left\{\begin{pmatrix} 1 & 0 & 0 \\ & 1 & b \\ & & 1 \end{pmatrix}\right\}$ act by $\left(\begin{pmatrix} 1 & 0 & 0 \\ & 1 & b \\ & & 1 \end{pmatrix} f\right)(x) = f(x+b)$. Thus Y acts by $Yf(x) = \frac{df}{dx}(x)$.

Let $\begin{pmatrix} 1 & a & 0 \\ & 1 & 0 \\ & & 1 \end{pmatrix}$ act as follows: given $f \in L^2(\mathbb{R})$ let $\hat{f}$ be its Fourier transform,

$$\hat{f}(k) = \int_{\mathbb{R}} f(x)e(-kx)dx.$$

Define

$$\widehat{\begin{pmatrix} 1 & a & 0 \\ & 1 & 0 \\ & & 1 \end{pmatrix} f}(k) = \hat{f}(k+\hbar a).$$

In other words

$$\widehat{\begin{pmatrix} 1 & a & 0 \\ & 1 & 0 \\ & & 1 \end{pmatrix} f}(k) = \widehat{f(x)e(a\hbar x)}$$

so

$$\begin{pmatrix} 1 & a & 0 \\ & 1 & 0 \\ & & 1 \end{pmatrix} f(x) = e(a\hbar x)f(x)..$$

Differentiating with respect to a at $a = 0$ we see that

$$Xf(x) = 2\pi i\hbar x f(x) = ihxf(x).$$

From the calculation above we get

$$[X,Y] = ih\left[x, \frac{d}{dx}\right] = -ih\,.$$

thus as long as we define $Zf = -ihf$, that is $\begin{pmatrix}1 & 0 & z \\ & 1 & 0 \\ & & 1\end{pmatrix} f(x) = e^{-ihz} f(x)$ we get an action of H.

THEOREM 164 (Stone–von Neumann). *For each $\lambda = e^{-ih} \in S^1$ there is a unique (up to isomorphism) irreducible action of H on a Hilbert space such that Z acts by the scalar $-ih$.*

DEFINITION 165. The associative algebra generated by X, Y subject to $[X,Y] = Z$ and $[Z,X] = [Z,Y] = 0$ is called the *Heisenberg algebra*.

6.2.3. Example: the harmonic oscillator. Consider $H = \frac{1}{2m}p^2 + \frac{1}{2}kx^2$. Define $a = \sqrt{\frac{\sqrt{km}}{2\hbar}}x + \frac{i}{\sqrt{2\hbar\sqrt{km}}}p$. Then $a^\dagger = \sqrt{\frac{\sqrt{km}}{2\hbar}}x - \frac{i}{\sqrt{2\hbar\sqrt{km}}}p$ so with $\omega = \sqrt{k/m}$ we have

$$\begin{aligned} N &= a^\dagger a \\ &= \left(\sqrt{\frac{\sqrt{km}}{2\hbar}}x - \frac{i}{\sqrt{2\hbar\sqrt{km}}}p\right)\left(\sqrt{\frac{\sqrt{km}}{2\hbar}}x + \frac{i}{\sqrt{2\hbar\sqrt{km}}}p\right) \\ &= \frac{\sqrt{km}}{2\hbar}x^2 + \frac{1}{2\hbar\sqrt{km}}p^2 + \frac{i}{2\hbar}(xp - px) \\ &= \frac{1}{\hbar\omega}H - \frac{1}{2}, \end{aligned}$$

that is

$$H = \hbar\omega\left(N + \frac{1}{2}\right).$$

For the same reason $aa^\dagger = \frac{1}{\hbar\omega}H + \frac{1}{2}$ so $\left[a, a^\dagger\right] = 1$. It follows that $[N,a] = \left[a^\dagger a, a\right] = \left[a^\dagger, a\right]a = -a$ and, taking adjoints, $\left[N, a^\dagger\right] = a^\dagger$.

Now suppose $N\phi = \lambda\phi$. Then

$$\begin{aligned} N(a\phi) &= (Na - aN)\phi + a(N\phi) \\ &= -a\phi + a\lambda\phi \\ &= (\lambda - 1)a\phi\,. \end{aligned}$$

It follows that if $a\phi \neq 0$ then $a\phi$ is an eigenvector with eigenvalue $\lambda - 1$. If $a\phi = 0$ then $N\phi = 0$ so it follows that if $\lambda \neq 0$ then $\lambda - 1$ is also an eigenvalue. But $\langle\phi, N\phi\rangle = \left\langle\phi, a^\dagger a\phi\right\rangle = \langle a\phi, a\phi\rangle \geq 0$ so the eigenvalues of N are non-negative.

It follows that the eigenvalues are the non-negative integers.

For the same reason $N\left(a^\dagger\phi\right) = (\lambda+1)a^\dagger\phi$, and this is never zero since $\left\langle a^\dagger\phi, a^\dagger\phi\right\rangle = \left\langle\phi, aa^\dagger\phi\right\rangle = \langle\phi,(N+1)\phi\rangle \geq \langle\phi,\phi\rangle$.

Let ϕ_0 be a vector such that $a\phi_0 = 0$ (equivalently, $N\phi_0 = 0$) and set $\phi_n = \left(a^\dagger\right)^n\phi_0$. Then

$$N\phi_n = n\phi_n\,.$$

Also, the subspace $\mathrm{Span}_{\mathbb{C}}\left\{\phi_n\right\}_{n=0}^{\infty}$ is an irreducible representation of the Heisenberg algebra: by induction $\left[a,\left(a^\dagger\right)^n\right] = n(a^\dagger)^{n-1}$ so $a\phi_n = n\phi_{n-1}$.

From this we also get $\langle\phi_n,\phi_n\rangle = \left\langle\phi_0, a^n(a^\dagger)^n\phi_0\right\rangle = n!\,\langle\phi_0,\phi_0\rangle$ so $\frac{1}{\sqrt{n!}}\phi_n$ are L^2-normalized.

Finally we see $H\phi_n = \hbar\omega(n+1)\phi_n$.

EXERCISE 166. In the representation on $L^2(\mathbb{R})$ we have

$$\phi_n(x) = 2^{-n/2}\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\exp\left(-\frac{m\omega x^2}{2\hbar}\right)H_n\left(\sqrt{\frac{m\omega}{\hbar}}x\right)\,.$$

6.2.4. Connection to classical mechanics. Instead of looking at the “deep quantum” eigenstates, consider a *Gaussian wave packet*

$$\Psi(x,0) = \frac{1}{\left(\pi\sigma^2\right)^{1/4}}\exp\left(\frac{i}{\hbar}p_0x - \frac{(x-q_0)^2}{2\sigma^2}\right)\,.$$

Let us try the Ansatz

$$\Psi(x,t) = \frac{1}{\left(\pi\sigma^2(t)\right)^{1/4}}\exp\left(i\theta(t) + \frac{i}{\hbar}p(t)x - \frac{(x-q(t))^2}{2\sigma^2(t)}\right)\,.$$

Then

$$\partial_x\Psi = \frac{1}{\left(\pi\sigma^2\right)^{1/4}}\exp\left(i\theta(t) + \frac{i}{\hbar}px - \frac{(x-q)^2}{2\sigma^2}\right)\cdot\left(\frac{i}{\hbar}p - \frac{x-q}{\sigma^2}\right)$$

so

$$\partial_x^2\Psi = \left[\left(\frac{i}{\hbar}p - \frac{x-q}{\sigma^2}\right)^2 - \frac{1}{\sigma^2}\right]\Psi\,.$$

It follows that

$$-\frac{\hbar^2}{2m}\partial_x^2\Psi + \frac{1}{2}kx^2\Psi = \left[-\frac{\hbar^2}{2m}\left(\frac{i}{\hbar}p - \frac{x-q}{\sigma^2}\right)^2 + \frac{\hbar^2}{2m\sigma^2} + \frac{1}{2}kx^2\right]\Psi\,.$$

On the other hand

$$i\hbar\frac{d}{dt}\Psi = i\hbar\left[-\frac{1}{2}\frac{\dot\sigma}{\sigma} + i\dot\theta + \frac{i}{\hbar}\dot px + \frac{(x-q)\dot q}{\sigma^2} + \frac{(x-q)^2}{\sigma^3}\dot\sigma\right]\Psi\,.$$

We therefore have a solution if

$$-\frac{\hbar^2}{2m}\left(\frac{i}{\hbar}p-\frac{x-q}{\sigma^2}\right)^2+\frac{\hbar^2}{2m\sigma^2}+\frac{1}{2}kx^2=-\frac{i\hbar}{2}\frac{\dot\sigma}{\sigma}-\hbar\dot\theta-\dot p x+i\hbar\frac{(x-q)\dot q}{\sigma^2}+i\hbar\frac{(x-q)^2}{\sigma^3}\dot\sigma\,.$$

Matching real and imaginary parts we get

$$\frac{p^2}{2m}-\frac{\hbar^2(x-q)^2}{2m\sigma^4}+\frac{\hbar^2}{2m\sigma^2}+\frac{1}{2}kx^2=-\dot p x-\hbar\dot\theta$$

and

$$\frac{p(x-q)}{m\sigma^2}=-\frac{\dot\sigma}{2\sigma}+\frac{(x-q)\dot q}{\sigma^2}+\frac{(x-q)^2}{\sigma^3}\dot\sigma\,.$$

Looking at the coefficient of x^2 we see $\dot\sigma=0$, and, dividing by $x-q$, that

$$\boxed{\dot q=\frac{p}{m}}.$$

Now in real part equation, the coefficient of x^2 is

$$\frac{1}{2}\left(k-\frac{\hbar^2}{m\sigma^4}\right)$$

so

$$\sigma=\left(\frac{\hbar^2}{mk}\right)^{1/4}.$$

The coefficient of x gives

$$\frac{\hbar^2 q}{m\sigma^4}=-\dot p$$

that is

$$\boxed{\dot p=-kq}.$$

The constant term reads

$$\frac{p^2}{2m}-\frac{\hbar^2q^2}{2m\sigma^4}+\frac{\hbar^2}{2m\sigma^2}=-\hbar\dot\theta\,,$$

so we can solve for the phase $\theta(t)$ which is largely irrelevant.

CONCLUSION 167. If we have $\sigma=\left(\frac{\hbar^2}{mk}\right)^{1/4}$ exactly then the Gaussian wave packet

$$\Psi(x,t)=\frac{1}{(\pi\sigma^2)^{1/4}}\exp\left(i\theta(t)+\frac{i}{\hbar}p(t)x-\frac{(x-q(t))^2}{2\sigma^2}\right)$$

keeps it shapes, with p,q evolving by Hamilton’s equations.

QUESTION 168. *What happens for general Hamiltonians?*

6.3. The Fourier transform and pseudodifferential operators

We construct a quantization map.

6.3.1. The fourier transform. $L^2(\mathbb{R}^n)$, $\mathcal{S}(\mathbb{R}^n)$, $H^k(\mathbb{R}^n)$
Classical Fourier transform: $\hat{f}(k) = \int_{\mathbb{R}^n} f(x)e(-kx)dx$.

(1) $f \in \mathcal{S}(\mathbb{R}^n)$ then
 (a) $\widehat{\partial_j f}(k) = \int_{\mathbb{R}^n} \partial_j f(x)e(-kx)dx = \left(2\pi i k_j\right)\int_{\mathbb{R}^n} f(x)e(-kx)dx = (2\pi i k_j)\hat{f}(k)$. Thus $\widehat{D^\alpha f}(k) = (2\pi i)^{|\alpha|}k^\alpha \hat{f}(k)$.
 (b) $\partial_j \hat{f}(k) = \widehat{(-2\pi i x)} f(k)$.
  - Thus $\hat{f} \in \mathcal{S}(\mathbb{R}^{n*})$.
 (c) $f(x) = \int_{\mathbb{R}^{n*}} \hat{f}(k)e(kx)dk$

(2) Smoothing gives Fourier inversion more generally.

(3) Let $\Phi(x) = \sum_\Lambda f(x+\delta\lambda)$. Then for $k \in \delta^{-1}\Lambda^*$ have

$$\hat{\Phi}(k) = \hat{f}(k)$$

and

$$\begin{aligned}\sum_{k\in\delta^{-1}\Lambda^*}\left|\hat{\Phi}(k)\right|^2 &= \frac{1}{\delta^n \operatorname{vol}(\mathbb{R}/\Lambda)}\int_{\mathbb{R}/\delta\Lambda}|\Phi(x)|^2\,dx\\ &= \frac{1}{\delta^n \operatorname{vol}(\mathbb{R}/\Lambda)}\sum_{\lambda,\lambda'\in\Lambda}\int_{\mathbb{R}/\delta\Lambda}\overline{\Phi(x+\delta\lambda)}\cdot\Phi(x+\delta\lambda')dx\,.\end{aligned}$$

Absolute convergence by Cauchy–Schwartz. Now take $\delta\to\infty$. Then $\operatorname{vol}(\mathbb{R}^{n*}/\delta^{-1}\Lambda^*)\sum_{k\in\delta^{-1}\Lambda^*}\left|\hat{\Phi}(k)\right|^2 \to \int_{\mathbb{R}^{n*}}\left|\hat{\Phi}(k)\right|^2 dk$, while on LHS only the summand with $\lambda=\lambda'=0$ survives. We conclude that

$$\int\left|\hat{\Phi}(k)\right|^2 dk = \int|\Phi(x)|^2\,dx\,.$$

(4) Extend FT by continuity to a unitary map $\mathcal{F}\colon L^2(\mathbb{R}^n)\to L^2(\mathbb{R}^{n*})$. By Fourier inversion this is an isometry.

(5) $f \in H^r$ if $\langle k\rangle^r \hat{f} \in L^2$.

Semiclassical Fourier transform: $\tilde{f}(p) = \hat{f}\left(\frac{p}{h}\right) = \int_{\mathbb{R}^n} f(x)e(-\frac{p}{h}x)dx = \int_{\mathbb{R}^n} f(x)\exp\left(\frac{ipx}{\hbar}\right)dx$. Then

$$\begin{aligned}f(x) &= \int \hat{f}(k)e(kx)dk\\ &= h^{-n}\int \tilde{f}(p)\,e\left(\frac{px}{h}\right)dp\,.\end{aligned}$$

Pancherel now reads

$$\|f\|_{L^2}^2 = h^{-n}\int\left|\tilde{f}(p)\right|^2 dp\,.$$

6.3.2. Quantization for Schwartz-class symbols. The position operator acts by M_x. Since the Fourier transform of $\widehat{\partial_j f}(k)$ is $\left(2\pi i k_j\right)\hat{f}(k)$ we see that $-i\hbar\partial_j$ is maps $\tilde{f}$ to $p\tilde{f}(p)$.

We can thus quantize $a(x,p) = a_1(x)a_2(p)$ by

$$\begin{aligned}\left(\mathrm{Op}_f\right)(x) &= h^{-n} a_1(x) \int a_2(p)\tilde{f}(p) e\left(\frac{px}{h}\right) dp \\ &= h^{-n} \int a(x,p)\tilde{f}(p) e\left(\frac{px}{h}\right) dp \\ &= h^{-n} \iint a(x,p) f(y) e\left(\frac{p(x-y)}{h}\right) dy dp\,.\end{aligned}$$

More generally we can set

$$\left(\mathrm{Op}_h^t f\right)(x) = h^{-n} \iint a\left((1-t)x + ty, p\right) f(y) e\left(\frac{p(x-y)}{h}\right) dy dp\,.$$

- $t = 0$ is the *standard* (Kohn–Nirenberg) *quantization* above
- $t = 1$ applies $a_1(x)$ and *then* $a_2(p)$ ("adjoint Kohn–Nirenberg quantization").
- $t = \frac{1}{2}$ is *Weyl symmetrization*:

$$\begin{aligned}\left\langle \mathrm{Op}_{\mathrm{W}(a) f, g} \right\rangle &= h^{-n} \iiint \overline{a\left(\frac{x+y}{2}, p\right)} g(x) \overline{f(y)} e\left(\frac{p(y-x)}{h}\right) dx dy dp \\ &= \left\langle f, \mathrm{Op}_{\mathrm{W}(\bar{a}) g} \right\rangle\end{aligned}$$

Standard quantization is easier to work with (take Fourier transform, multiply by a, take inverse Fourier transform), but as we see here Weyl symmetrization is better: if a is real, then $\mathrm{Op}_h^{\mathrm{W}}(a)$ is symmetric (and often selfadjoint).

EXERCISE 169. $\left(\mathrm{Op}_h^t a\right)^\dagger = \mathrm{Op}_h^{1-t}(\bar{a})$.

EXAMPLE 170. If $a = a(p)$ then $\mathrm{Op}_h^t(a)$ is the Fourier multiplier. In particular if $a(p) = p^\alpha$ then $\mathrm{Op}_h^t(a) = (i\hbar)^{|\alpha|} D^\alpha$.

THEOREM 171. *Suppose $a \in \mathcal{S}(M)$. Then $\mathrm{Op}_h^t(a)$ makes sense on very general f, for example it gives a continuous map $\mathcal{S}' \to \mathcal{S}$.*

PROOF. The operator has kernel

$$K(x,y) = h^{-n} \int_{\mathbb{R}^{n*}} a\left((1-t)x + ty, p\right) e\left(\frac{p(x-y)}{h}\right) dp\,.$$

opThis is a Schwartz function on $\left(\mathbb{R}^n\right)^2$ (exercise!), and the claim follows. □

PROPOSITION 172. *If $a = a(x)$ then $\mathrm{Op}_h^t(a)$ is just multiplication by a.*

PROOF. We checked this explicitly for $t = 0$. Next,

$$\begin{aligned}\frac{d}{dt}\left(\mathrm{Op}_h^t(a)f\right) &= h^{-n}\iint \langle \partial_x a\left((1-t)x+ty,p\right), y-x\rangle f(y) e\left(\frac{p(x-y)}{h}\right) dydp \\ &= -\frac{h}{2\pi i h^n}\iint \left\langle \partial_x a\left((1-t)x+ty,p\right), \partial_p e\left(\frac{p(x-y)}{h}\right)\right\rangle f(y)dydp \\ &= -\frac{h}{2\pi i h^n}\iint \mathrm{Div}_p\,\partial_x a\left((1-t)x+ty\right) f(y) e\left(\frac{p(x-y)}{h}\right) dydp \\ &= -\frac{h}{2\pi i h^n}\int \mathrm{Div}_p\left[e\left(\frac{px}{h}\right)\cdot\left(\partial_x a\left(\widetilde{(1-t)x+}ty\right) f(y)\right)(p)\right] dp = 0\,.\end{aligned}$$

□

LEMMA 173. *We can recover the symbol from the operator:* $\mathrm{Op}_(a)\left(e\left(\frac{p}{h}\cdot\right)\right) = a(x,p)e\left(\frac{px}{h}\right)$.

PROOF. By Fourier inversion we have

$$\begin{aligned}e\left(-\frac{px}{h}\right)\mathrm{Op}_(a)\left(e\left(\frac{p}{h}\cdot\right)\right)(x) &= h^{-n}\iint a\left(x,p'\right) e\left(\frac{p(y-x)}{h}\right) e\left(\frac{p'(x-y)}{h}\right) dydp'. \\ &= h^{-n}\iint a\left(x,p'\right) e\left(\frac{(p'-p)x}{h}\right)\left(-\frac{(p'-p)y}{h}\right) dydp \\ &= \int a\left(x,p'\right) e\left(\frac{(p'-p)x}{h}\right)\delta\left(p'-p\right) dp \\ &= a(x,p)\,.\end{aligned}$$

□

6.3.3. Composition. A key fact is that $\mathrm{Op}_{\mathrm{W}(a)}\mathrm{Op}_{\mathrm{W}(b)}$ is an operator of the same type: there is a classical observable $a\natural b$ such that $\mathrm{Op}_{\mathrm{W}(a)}\mathrm{Op}_{\mathrm{W}(b)=\mathrm{Op}_{\mathrm{W}(a\#b)}}$ (this is also true for the other quantizations).

THEOREM 174 (Zworski Theorem 4.12). *Let* $a,b\in\mathcal{S}$. *Then* $a\natural b = \sum_{k=0}^{N}\frac{i^h h^k}{k!}A(D)^k a(x,p)b(x',p') + O(h^{N+1})$. *Here*

$$A(D) = \omega\left(\begin{pmatrix}\partial_x\\ \partial_p\end{pmatrix},\begin{pmatrix}\partial_{x'}\\ \partial_{p'}\end{pmatrix}\right)\,.$$

COROLLARY 175. *We have*

(1) $a\#b = ab + \frac{h}{2i}\{a,b\} + O(h^2)$, *and hence* $\left[\mathrm{Op}_{W(a)},\mathrm{Op}_{W(b)}\right] = ih\,\mathrm{Op}_{W(\{a,b\})+O(h^3)}$.
   - *The error term* $O(h^3)$ *is special to Weyl quantization. For the standard quantization we'd get* $O(h^2)$.

(2) *If* a,b *have disjoint supports then* $a\#b = O(h^\infty)$.

6.3.4. More general observables: pseudodifferential operators. To go beyond $\mathcal{S}$ need some preparation.

Our definition of $\mathrm{Op}_{W(a)}$ would also work for a of *moderate growth* $\frac{\partial^\alpha}{\partial x^\alpha}\frac{\partial^\beta}{\partial p^\beta}a(x,p) \ll_{a,\alpha,\beta} \langle x\rangle^{O_{a,\alpha}(1)+r|\beta|}\langle p\rangle^{O_{a,\alpha,\beta}(1}$ (check the integration by parts arguments). We won't go that far.

On the other hand we will now let the symbols (=observables) depend explicitly on h.

DEFINITION 176. Say $a \in S^m$ if $\frac{\partial^\alpha}{\partial x^\alpha}\frac{\partial^\beta}{\partial \xi^\beta}a(x,p) \ll_{a,\alpha,\beta} \langle p\rangle^{m-|\beta|}$, $a \in S^m_\delta$ if $\frac{\partial^\alpha}{\partial x^\alpha}\frac{\partial^\beta}{\partial \xi^\beta}a(x,p) \ll_{a,\alpha,\beta} h^{-\delta|\alpha+\beta|}\langle p\rangle^{m-|\beta|}$. The implied constants should be uniform for $0 < h \le h_0 = h_0(a)$.

EXAMPLE 177. $\sum_{|\alpha|\le m} c_\alpha(x)p^\alpha$ with every derivative of c_α bounded.

Sometimes $a \sim \sum_{j=0}^\infty a_j h^j$ with $a_j \in S^m_\delta$ in the sense that $a = \sum_{j=0}^N a_j h^j + O_{S^m}(h^{N+1})$. In that case we call a_0 the *principal symbol* of a.

EXERCISE 178. (Borel Theorem) for any choice of a_j there is a_h.

PROPOSITION 179. *If $a \in S^m_\delta$ then $\mathrm{Op}_{W(a)\colon \mathcal{S}\to\mathcal{S}}$ is continuous, and similarly $\mathcal{S}' \to \mathcal{S}'$.*

LEMMA 180. *We can recover the symbol from the operator.*

PROOF. For $a \in \mathcal{S}(M)$ this is 173; in general use the density of $\mathcal{S}$in S^m_δ. □

LEMMA 181 (Pseudolocality). *Up to negligible terms,* $\mathrm{Op}(a)f(x)$ *only depends on f near x.*

PROOF. In the integral

$$\left(\mathrm{Op}_h^t f\right)(x) = h^{-n}\iint a\left((1-t)x+ty,p\right)f(y)e\left(\frac{p(x-y)}{h}\right)dydp\,.$$

Observe that $\left\langle x-y,-i\hbar\partial_p\right\rangle e\left(\frac{p(x-y)}{h}\right) = |x-y|^2 e\left(\frac{p(x-y)}{h}\right)$ so

$$Le\left(\frac{p(x-y)}{h}\right) = e\left(\frac{p(x-y)}{h}\right)$$

for $L = \frac{\left\langle x-y,-i\hbar\nabla_p\right\rangle}{|x-y|^2}$. We can apply this N times and integrate by parts N times to get

$$\left(\mathrm{Op}_h^t f\right)(x) \propto h^{-n}h^N\iint \frac{\left\langle x-y,\nabla_p\right\rangle^N}{|x-y|^N}a\left((1-t)x+ty,p\right)f(y)e\left(\frac{p(x-y)}{h}\right)dydp\,.$$

Now the contribution of the region $|x-y| \ge \varepsilon$ to the integral can be made $O(h^N)$ by noting that the more we differentiate, the more decay wrt p we get. □

PROPOSITION 182. *If $a \in S^m_\delta$ and $b \in S^{m'}_\delta$ then $a\sharp b \in S^{m_1+m_2}_\delta$.*

THEOREM 183. *Let $a \in S^0$. Then $\mathrm{Op}_{W(a)}$ is bounded on L^p for each $p \in (1,\infty)$, uniformly in h. If $a \in S^0_\delta$ the bound is mildly h-dependent..*

PROOF. Littlewood–Paley decomposition. □

COROLLARY 184. *If $a,b \in S^0_\delta$ then* $\mathrm{Op}_{W(ab)=\mathrm{Op}_{W(a)}\mathrm{Op}_{W(b)+O_{L^2\to L^2}\left(h^{1-2\delta}\right)}}$ *; if ab have disjoint supports (so $ab = 0$ classically) then* $\mathrm{Op}_{W(a)}\mathrm{Op}_{W(b)=O_{L^2\to L^2}(h^\infty)}$.

- Now $\mathrm{Op}_{W(a)}\mathrm{Op}_{W(b)=\mathrm{Op}_{W(ab)}}$

COROLLARY 185 ("Ehrenfest's Theorem"). *Let $\psi \in L^2(X)$ be a quantum state, $\psi(t) = U(t)\psi$ where $i\hbar\frac{d}{dt}U(t) = \hat{H}(t)U(t)$. Let $A = \mathrm{Op}_{W(a)}$. Then*

$$\begin{aligned}\frac{d}{dt}\left\langle \psi(t)\middle|\mathrm{Op}_{W(a)}\middle|\psi(t)\right\rangle &= \frac{1}{ih}\left\langle \psi(t)\middle|\left[\hat{H},A\right]\middle|\psi(t)\right\rangle \\ &= \left\langle \psi(t)\middle|\mathrm{Op}_{W\{H,a\}}\middle|\psi(t)\right\rangle + O(h)\,.\end{aligned}$$

- Physics interpretation: the expectation values morally satisfy Hamilton's equations.

6.4. Egorov's Theorem

6.4.1. Time-dependent version: fixed time. Let $H \in C^\infty(M\times\mathbb{E}^1)$ be a (possibly time-dependent) observable, which we suppose vanishes at all times outside some fixed bounded set (in particular $H(t) \in C^\infty_{\mathrm{c}}(M) \subset \mathcal{S}(M)$ and let Φ_t be the corresponding Hamiltonian flow. Let $\hat{H}(t) = \mathrm{Op}_{W(H(t))}$. One can prove that the time-dependent Schrödinger equation $i\hbar\frac{d}{dt}U(t) = U(t)\hat{H}(t)$ has a unique solution in unitary operators $U(t)$.

THEOREM 186 (Egorov I). *Let $a \in S^m$. Then on any fixed interval $[0,T]$ we have*

$$B(t) \overset{def}{=} U(t)^{-1}\,\mathrm{Op}_{W(a)U(t)=\mathrm{Op}_{W(b_t)}}$$

where $b_t = a\circ\Phi_t + O_S(h)$.

PROOF. (Taken from Zworski) $a\circ\Phi_t \in S(m)$ since the vector fields X_H are smooth of uniform compact support. Thus we can define$B_0(t) = \mathrm{Op}_{W(a\circ\Phi_t)}$. Differentiating under the integral sign,

$$\begin{aligned}i\hbar\frac{d}{dt}B_0(t) &= i\hbar\,\mathrm{Op}_{W\left(\frac{d}{dt}a\circ\Phi_t\right)} \\ &= i\hbar\,\mathrm{Op}_{W(\{H(t),a\circ\Phi_t\})} \\ &= \left[\hat{H}(t),B_0(t)\right] + E(t)\end{aligned}$$

where $E(t) = \mathrm{Op}_{W(e(t))}$ for $e(t) \in h^2\mathcal{S}$. Indeed, by Theorem 174, $a\sharp b$ is obtained by differentiating products of a,b and here $H(t)$ has compact support so $a\sharp b$ also has compact support and is hence Schwartz. It follows that $E(t)$ is $O(h^2)$ in operator norm.

It follows that

$$i\hbar\frac{d}{dt}\left(U(t)B_0(t)U(t)^{-1}\right) = U(t)\hat{H}(t)B_0(t)U(t)^{-1} + U(t)\left[\hat{H}(t),B_0(t)\right]U(t)^{-1} + U(t)E(t)U(t)^{-1} - U(t)\hat{H}(t)B_0(t)\hat{H}($$
$$= U(t)E(t)U(t)^{-1} = O(h^2)\,.$$

Integrating on $[0,t]$ for $t \in [0,T]$ we get

$$U(t)B_0(t)U^{-1}(t) - \mathrm{Op}_{\mathrm{W}(a)=O_{L^2\to L^@}(h)}\,.$$

More precisely we have

$$B(t) - B_0(t) = \frac{i}{\hbar}U(t)^{-1}\int_0^t U(s)E(s)U(s)^{-1}dsU(t)\,,$$

Further technical work now shows that the RHS is pseudodifferential with symbol of size $O_S(h)$. □

REMARK 187. If $a \sim \sum_j a_j h^j$ with $a_j \in S^m$ then the proof gives $b_t \sim a_0 \circ \Phi_t + \sum_{j=1}^{\infty} b_j h^j$ but it's harder to give explicit formulas for the b_j.

REMARK 188. The proof also applies if $H(t), a \in S$.

6.4.2. Time-independent version: the Ehrenfest time. Suppose now that $H \in S^0$ does not depend on t or h, and that H is bounded below. Let $M(R) = \{(x,p) \mid H(x,p) \le R\}$ (this is flow-invariant by conservation of energy) and set

$$\Gamma_R = \left\|\nabla^2 H\right\|_{L^\infty(M(R))}\,.$$

and the *Lyapunov exponent*

$$\tilde{\Gamma}_R = \lim_{t\to\infty}\frac{1}{t}\sup_{M(R)}\log|\nabla\Phi_t|$$

(the limit exists by subadditivity). Then all derivatives of Φ grow at most exponentially, with exponential rate at most $\tilde{\Gamma}_R + \varepsilon \le \Gamma_R + \varepsilon$.

THEOREM 189 (Egorov's Theorem up to the Ehrenfest time). *Let $a \in S$. Then for $\gamma > \tilde{\Gamma}_R$, $T > 0$, $\delta \in [0, \frac{1}{2})$ if*

$$|t| \le T + \frac{\delta}{\gamma}\log(h^{-1})$$

then

$$U(t)^{-1}\,\mathrm{Op}_{W(a)U(t)=\mathrm{Op}_{W(b_t)}}$$

for a symbol $b_t \in S^0_\delta$ with such that

$$b_t = a\circ\Phi_t + O_{S^0_\delta}(h^{2-3\delta})\,.$$

Bibliography

[1] Tom Archibald. Differential equations: a historical overview to circa 1900. In *A history of analysis*, volume 24 of *Hist. Math.*, pages 325–353. Amer. Math. Soc., Providence, RI, 2003.

[2] V. I. Arnol′ d. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, [1989?]. Translated from the 1974 Russian original by K. Vogtmann and A. Weinstein, Corrected reprint of the second (1989) edition.

[3] Richard H. Cushman and Larry M. Bates. *Global aspects of classical integrable systems*. Birkhäuser/Springer, Basel, second edition, 2015.

[4] Herbert Goldstein. *Classical mechanics*. Addison-Wesley Series in Physics. Addison-Wesley Publishing Co., Reading, MA, second edition, 1980.

[5] Hans Niels Jahnke, editor. *A history of analysis*, volume 24 of *History of Mathematics*. American Mathematical Society, Providence, RI; London Mathematical Society, London, 2003. Translated from the German.

[6] L. D. Landau and E. M. Lifshitz. *Mechanics*, volume Vol. 1 of *Course of Theoretical Physics*. Pergamon Press, Oxford; Addison-Wesley Publishing Co., Inc., Reading, MA, 1960. Translated from the Russian by J. B. Bell.

[7] Michael Spivak. *Physics for mathematicians—mechanics I*. Publish or Perish, Inc., Houston, TX, 2010.