Fermat's Principle and the Geometric Mechanics of Ray Optics

Summer School Lectures, Fields Institute, Toronto, July 2012

Darryl D Holm   Imperial College London   d.holm@ic.ac.uk
http://www.ma.ic.ac.uk/~dholm/

Texts for the course include:

Geometric Mechanics I: Dynamics and Symmetry, & II: Rotating, Translating and Rolling,
by DD Holm,
World Scientific: Imperial College Press, Singapore, Second edition (2011).
ISBN 978-1-84816-195-5 and ISBN 978-1-84816-155-9.

Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions,
by DD Holm,T Schmah and C Stoica.
Oxford University Press, (2009).
ISBN 978-0-19-921290-3

Introduction to Mechanics and Symmetry,
by J. E. Marsden and T. S. Ratiu
Texts in Applied Mathematics, Vol. 75. New York: Springer-Verlag (1994).

Contents

Of course, not all of these lectures could be given at a five-day summer school at Toronto.

To help select topics, students are asked at the end of each lecture to turn in a piece of paper including their **name, date and email address** on which you answer the following two questions *in sentences*.

(1) What was this lecture about?
(2) Write a question that you would like to see addressed in a subsequent lecture.

Subsequent lectures at the summer school will be selected material from these notes that will emphasize the questions and interests expressed by the students.

1 Mathematical setting

- The mathematical setting for geometric mechanics involves manifolds, (matrix) Lie groups and (maybe later) diffeomorphisms
  - Manifold $M \simeq_{loc} \mathbb{R}^n$ e.g., $n = 1$ (scalars), $n = m$ (m-vectors), $n = m \times m$ (matrices),
  - Motion equation on TM: $\dot{q}(t) = f(q) \Longrightarrow$ transformation theory (pullbacks and all that)
  - Hamilton's principle for Lagrangian $L : TM \to \mathbb{R}$ vector fields
    - ∗ Euler–Lagrange equations on T^*M
    - ∗ Hamilton's canonical equations on T^*M
    - ∗ Euler–Poincaré eqns on $T^*_e G \simeq \mathfrak{g}^*$ for reduced Lagrangian $\ell : \mathfrak{g} \to \mathbb{R}$, e.g., *rigid body*.

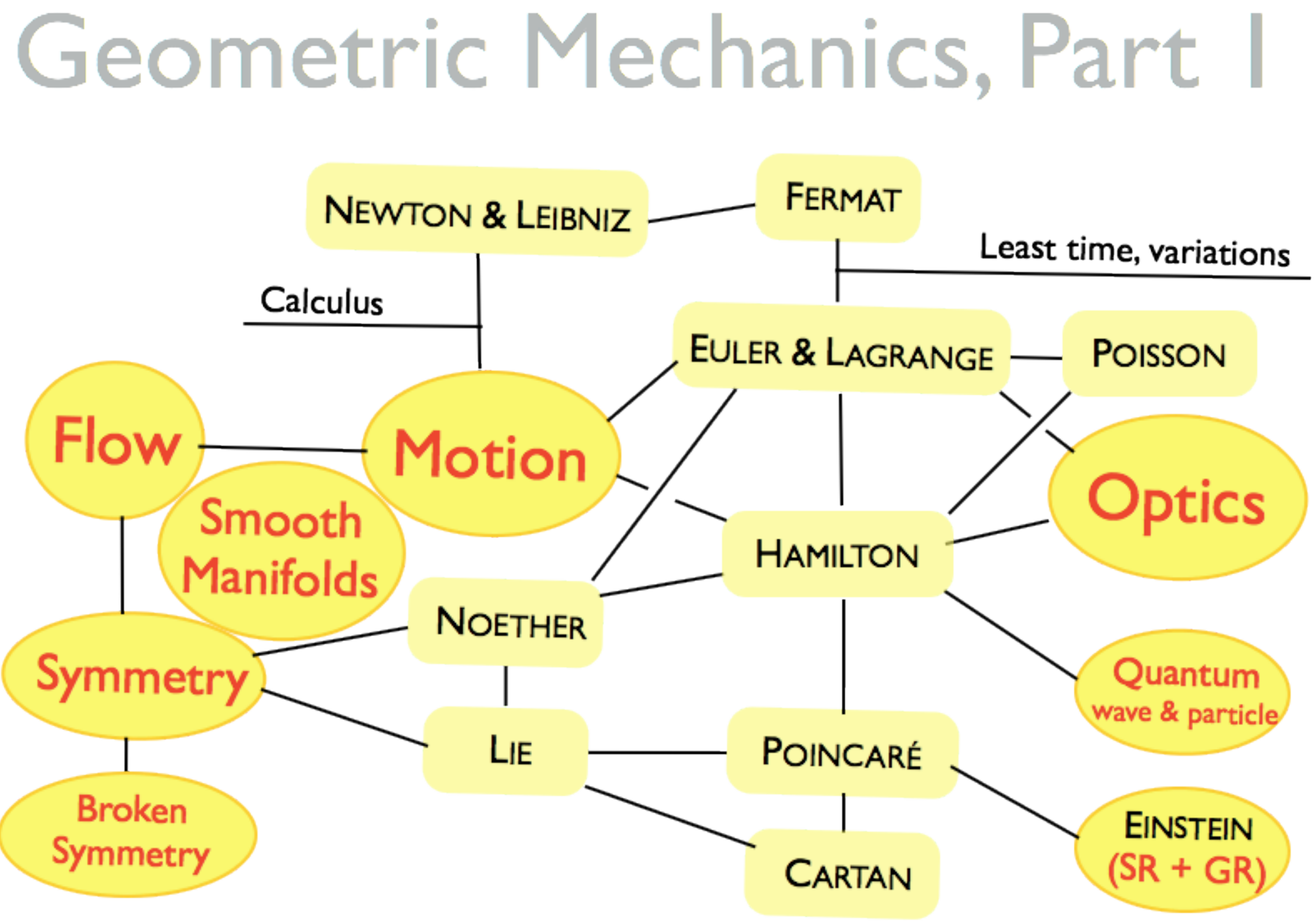
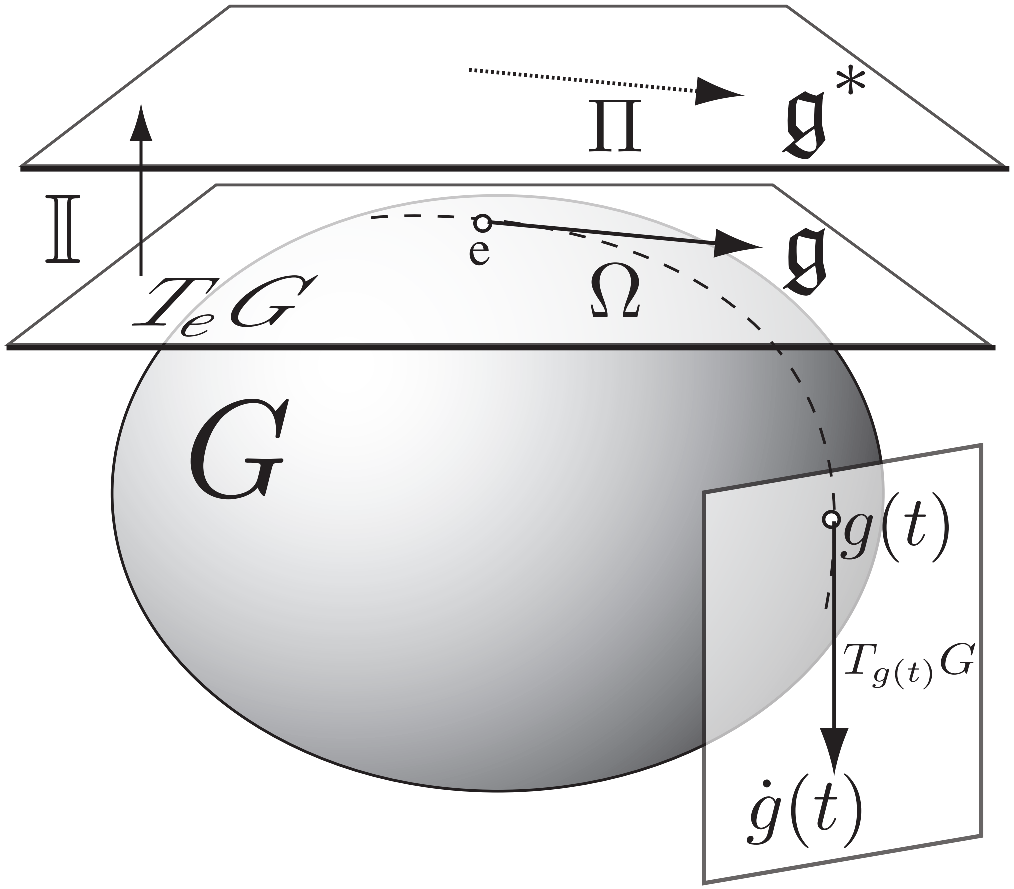


Figure 1: The fabric of geometric mechanics is woven by a network of fundamental contributions by at least a dozen people to the dual fields of optics and motion.



A Lie group G is a manifold. Its tangent space at the identity T_eG is its Lie algebra $\mathfrak{g}$.

2 Fermat's principle

Pierre de Fermat

According to Fermat's principle (1662):
The path between two points taken by a ray of light leaves the optical length stationary under variations in a family of nearby paths.

This principle accurately describes the properties of light that is reflected by mirrors, refracted at a boundary between different media or transmitted through a medium with a continuously varying index of refraction. Fermat's principle *defines* a light ray and provides an example that will guide us in recognising the principles of geometric mechanics.

Definition

2.1 (Optical length)**.** *The **optical length** of a path $\mathbf{r}(s)$ taken by a ray of light in passing from point A to point B in three-dimensional space is defined by*

$$\mathsf{A} := \int_A^B n(\mathbf{r}(s))\, ds\,, \tag{2.1}$$

where $n(\mathbf{r})$ is the index of refraction at the spatial point $\mathbf{r} \in \mathbb{R}^3$ and

$$ds^2 = d\mathbf{r}(s) \cdot d\mathbf{r}(s) \tag{2.2}$$

is the element of arc length ds along the ray path $\mathbf{r}(s)$ through that point.

Definition

2.2 (Fermat's principle for ray paths)**.** *The path* $\mathbf{r}(s)$ *taken by a ray of light passing from* A *to* B *in three-dimensional space leaves the optical length* ***stationary*** *under variations in a family of nearby paths* $\mathbf{r}(s,\varepsilon)$ *depending smoothly on a parameter* ε*. That is, the path* $\mathbf{r}(s)$ *satisfies*

$$\delta \mathsf{A} = 0 \quad \textit{with} \quad \delta \mathsf{A} := \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \int_A^B n(\mathbf{r}(s,\varepsilon))\, ds\,, \tag{2.3}$$

where the deviations of the ray path $\mathbf{r}(s,\varepsilon)$ *from* $\mathbf{r}(s)$ *are assumed to vanish when* $\varepsilon = 0$*, and to leave its endpoints fixed.*

Fermat's principle of stationary ray paths is dual to Huygens' principle (1678) of constructive interference of waves. According to Huygens, among all possible paths from an object to an image, the waves corresponding to the stationary path contribute most to the image because of their constructive interference. Both principles are models that approximate more fundamental physical results derived from Maxwell's equations.

The Fermat and Huygens principles for geometric optics are also foundational ideas in mechanics. Indeed, the founders of mechanics Newton, Lagrange and Hamilton all worked seriously in optics as well. We start with Fermat and Huygens, whose works preceded Newton's *Principia Mathematica* by 25 years, Lagrange's *Mécanique Analytique* by more than a century and Hamilton's *On a General Method in Dynamics* by more than 150 years.

After briefly discussing the geometric ideas underlying the principles of Fermat and Huygens and using them to derive and interpret the eikonal equation for ray optics, these notes show that Fermat's principle naturally introduces Hamilton's principle for the Euler–Lagrange equations, as well as the concepts of phase space, Hamiltonian formulation, Poisson brackets, Hamiltonian vector fields, symplectic transformations and momentum maps arising from reduction by symmetry.

In his time, Fermat discovered the geometric foundations of ray optics. This is our focus in these lectures.

2.1 Three-dimensional eikonal equation

Stationary paths Consider the possible ray paths in Fermat's principle leading from point A to point B in three-dimensional space as belonging to a family of C^2 curves $\mathbf{r}(s,\varepsilon) \in \mathbb{R}^3$ depending smoothly on a real parameter ε in an interval that includes $\varepsilon = 0$. This ε-family of paths $\mathbf{r}(s,\varepsilon)$ defines a set of smooth transformations of the ray path $\mathbf{r}(s)$. These transformations are taken to satisfy

$$\mathbf{r}(s,0) = \mathbf{r}(s)\,, \quad \mathbf{r}(s_A,\varepsilon) = \mathbf{r}(s_A)\,, \quad \mathbf{r}(s_B,\varepsilon) = \mathbf{r}(s_B)\,. \tag{2.4}$$

That is, $\varepsilon = 0$ is the identity transformation of the ray path and its endpoints are left fixed. An infinitesimal variation of the path $\mathbf{r}(s)$ is denoted by $\delta\mathbf{r}(s)$, defined by the ***variational derivative***,

$$\delta\mathbf{r}(s) := \frac{d}{d\varepsilon}\bigg|_{\varepsilon=0} \mathbf{r}(s,\varepsilon)\,, \tag{2.5}$$

where the fixed endpoint conditions in (2.4) imply $\delta\mathbf{r}(s_A) = 0 = \delta\mathbf{r}(s_B)$.

With these definitions, Fermat's principle in Definition 2.2 implies the fundamental equation for the ray paths, as follows.

Theorem

2.3 (Fermat's principle implies the eikonal equation)**.** *Stationarity of the optical length, or* ***action*** A*, under variations of the ray paths*

$$0 = \delta \mathsf{A} = \delta \int_A^B n(\mathbf{r}(s)) \sqrt{\frac{d\mathbf{r}}{ds} \cdot \frac{d\mathbf{r}}{ds}}\, ds\,, \tag{2.6}$$

defined using arc-length parameter s*, satisfying* $ds^2 = d\mathbf{r}(s) \cdot d\mathbf{r}(s)$ *and* $|\dot{\mathbf{r}}| = 1$*, implies the equation for the ray path* $\mathbf{r} \in \mathbb{R}^3$ *is*

$$\frac{d}{ds}\left(n(\mathbf{r}) \frac{d\mathbf{r}}{ds} \right) = \frac{\partial n}{\partial \mathbf{r}}\,. \tag{2.7}$$

In ray optics, this is called the ***eikonal equation****.*[1]*.*

Remark

2.4 (Invariance under reparameterisation)**.** *The integral in (2.6) is invariant under reparameterisation of the ray path. In particular, it is invariant under transforming* $\mathbf{r}(s) \to \mathbf{r}(\tau)$ *from arc length*

[1]The term ***eikonal*** (from the Greek $\epsilon\iota\kappa o\nu\alpha$ meaning image) was introduced into optics in Bruns [1895]

ds to optical length $d\tau = n(\mathbf{r}(s))ds$. That is,

$$\int_A^B n(\mathbf{r}(s)) \sqrt{\frac{d\mathbf{r}}{ds} \cdot \frac{d\mathbf{r}}{ds}}\, ds = \int_A^B n(\mathbf{r}(\tau)) \sqrt{\frac{d\mathbf{r}}{d\tau} \cdot \frac{d\mathbf{r}}{d\tau}}\, d\tau\,, \tag{2.8}$$

in which $|d\mathbf{r}/d\tau|^2 = n^{-2}(\mathbf{r}(\tau))$.

We now prove Theorem 2.3.

Proof. The equation for determining the ray path that satisfies the stationary condition (2.6) may be computed by introducing the ε-family of paths into the action A, then differentiating it with respect to ε under the integral sign, setting $\varepsilon = 0$ and integrating by parts with respect to s, as follows:

$$\begin{aligned} 0 &= \delta \int_A^B n(\mathbf{r}(s)) \sqrt{\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}}\, ds \\ &= \int_A^B \left[|\dot{\mathbf{r}}| \frac{\partial n}{\partial \mathbf{r}} \cdot \delta \mathbf{r} + \left(n(\mathbf{r}(s)) \frac{\dot{\mathbf{r}}}{|\dot{\mathbf{r}}|} \right) \cdot \delta \dot{\mathbf{r}} \right] ds \\ &= \int_A^B \left[|\dot{\mathbf{r}}| \frac{\partial n}{\partial \mathbf{r}} - \frac{d}{ds} \left(n(\mathbf{r}(s)) \frac{\dot{\mathbf{r}}}{|\dot{\mathbf{r}}|} \right) \right] \cdot \delta \mathbf{r}\, ds\,, \end{aligned}$$

where we have exchanged the order of the derivatives in s and ε, and used the homogeneous endpoint conditions in (2.4). We choose the arc-length variable $ds^2 = d\mathbf{r} \cdot d\mathbf{r}$, so that $|\dot{\mathbf{r}}| = 1$ for $\dot{\mathbf{r}} := d\mathbf{r}/ds$. (This means that $d|\dot{\mathbf{r}}|/ds = 0$.) Consequently, the three-dimension eikonal equation (2.7) emerges for the ray path $\mathbf{r} \in \mathbb{R}^3$. $\square$

Remark

2.5. *From the viewpoint of historical contributions in classical mechanics, the eikonal equation (2.7) is the three-dimensional*

$$\frac{d}{ds}\left(\frac{\partial L}{\partial \dot{\mathbf{r}}}\right) = \frac{\partial L}{\partial \mathbf{r}} \tag{2.9}$$

arising from ***Hamilton's principle***

$$\delta \mathsf{A} = 0 \quad \textit{with} \quad \mathsf{A} = \int_A^B L((\mathbf{r}(s), \dot{\mathbf{r}}(s))\, ds$$

for the ***Lagrangian*** *function*

$$L((\mathbf{r}(s), \dot{\mathbf{r}}(s)) = n(\mathbf{r}(s))\sqrt{\dot{\mathbf{r}} \cdot \dot{\mathbf{r}}}\,, \tag{2.10}$$

in Euclidean space.

Exercise. Verify that the same three-dimensional eikonal equation (2.7) also follows from Fermat's principle in the form

$$0 = \delta \mathsf{A} = \delta \int_A^B \frac{1}{2} n^2(\mathbf{r}(\tau))\, \frac{d\mathbf{r}}{d\tau} \cdot \frac{d\mathbf{r}}{d\tau}\, d\tau\,, \tag{2.11}$$

after transforming to a new arc-length parameter τ given by $d\tau = nds$. ★

Answer. Denoting $\mathbf{r}'(\tau) = d\mathbf{r}/d\tau$, one computes

$$0 = \delta \mathsf{A} = \int_A^B \frac{ds}{d\tau}\left[\frac{nds}{d\tau}\frac{\partial n}{\partial \mathbf{r}} - \frac{d}{ds}\left(\frac{nds}{d\tau}n\frac{d\mathbf{r}}{ds}\right)\right]\cdot \delta\mathbf{r}\, d\tau\,, \tag{2.12}$$

which agrees with the previous calculation upon reparameterising $d\tau = nds$. ▲

Remark

2.6 (Finsler geometry and singular Lagrangians).
The Lagrangian function in (2.10) for the three-dimensional eikonal equation (2.7)

$$L(\mathbf{r}, \dot{\mathbf{r}}) = n(\mathbf{r})\,\sqrt{\delta_{ij}\dot{r}^i\dot{r}^j} \tag{2.13}$$

is homogeneous of degree 1 in $\dot{\mathbf{r}}$*. That is,* $L(\mathbf{r}, \lambda\dot{\mathbf{r}}) = \lambda L(\mathbf{r}, \dot{\mathbf{r}})$ *for any* $\lambda > 0$*. Homogeneous functions of degree 1 satisfy* ***Euler's relation****,*

$$\dot{\mathbf{r}}\cdot\frac{\partial L}{\partial \dot{\mathbf{r}}} - L = 0\,. \tag{2.14}$$

Hence, Fermat's principle may be written as stationarity of the integral

$$\mathsf{A} = \int_A^B n(\mathbf{r}(s))\,ds = \int_A^B \mathbf{p}\cdot d\mathbf{r} \quad \textit{for the quantity} \quad \mathbf{p} := \frac{\partial L}{\partial \dot{\mathbf{r}}}\,.$$

The quantity $\mathbf{p}$ *defined here will be interpreted later as the* ***canonical momentum*** *for ray optics.*

Taking another derivative of Euler's relation (2.14) yields

$$\frac{\partial^2 L}{\partial \dot{r}^i \partial \dot{r}^j}\dot{r}^j = 0$$

so the Hessian of the Lagrangian L with respect to the tangent vectors is singular (has a zero determinant). A singular Lagrangian might become problematic in some situations. However, there is a simple way of obtaining a regular Lagrangian whose ray paths are the same as those for the singular Lagrangian arising in Fermat's principle.

The Lagrangian function in the integrand of action (2.11) is related to the Fermat Lagrangian in (2.13) by

$$\tfrac{1}{2}n^2(\mathbf{r})\,\delta_{ij}\dot{r}^i\dot{r}^j = \tfrac{1}{2}L^2 \ .$$

Computing the Hessian with respect to the tangent vector yields the Riemannian metric for the regular Lagrangian in (2.11),

$$\frac{1}{2}\frac{\partial^2 L^2}{\partial \dot{r}^i \partial \dot{r}^j} = n^2(\mathbf{r})\,\delta_{ij} \ .$$

The emergence of a Riemannian metric from the Hessian of the square of a homogeneous function of degree 1 is the hallmark of ***Finsler geometry****, of which Riemannian geometry is a special case. Finsler geometry, however, is beyond our present scope. For more discussions of the ideas underlying Finsler geometry, see, e.g., Rund [1959], Chern et al. [1999].*

In the present case, the variational principle for the regular Lagrangian in (2.11) leads to the same three-dimensional eikonal equation as that arising from Fermat's principle in (2.6), but parameterised by optical length, rather than arc length. This will be sufficient for the purposes of *studying ray trajectories because in geometric optics one is only concerned with the trajectories of the ray paths in space, not their parameterisation.*

Remark

2.7 (Newton's law form of the eikonal equation)**.** *Reparameterising the ray path in terms of a variable* $d\sigma = n^{-1}ds$ *transforms the eikonal equation (2.7) into the form of* ***Newton's law****,*

$$\frac{d^2\mathbf{r}}{d\sigma^2} = \frac{1}{2}\frac{\partial n^2}{\partial \mathbf{r}}\,. \tag{2.15}$$

Thus, in terms of the parameter σ *ray trajectories are governed by Newtonian dynamics. Interestingly, this equation has a conserved* ***energy integral****,*

$$E = \frac{1}{2}\left|\frac{d\mathbf{r}}{d\sigma}\right|^2 - \frac{1}{2}n^2(\mathbf{r})\,. \tag{2.16}$$

Exercise. Propose various forms of the squared index of refraction, e.g., cylindrically or spherically symmetric, then solve the eikonal equation in Newtonian form (2.15) for the ray paths and interpret the results as optical devices, e.g., lenses.

What choices of the index of refraction lead to closed ray paths? ★

2.2 Three-dimensional Huygens wave fronts

Ray vector Fermat's ray optics is complementary to Huygens' wavelets. According to the Huygens wavelet assumption, a level set of the wave front, $S(\mathbf{r})$, moves along the ***ray vector***, $\mathbf{n}(\mathbf{r})$, so that its incremental change over a distance $d\mathbf{r}$ along the ray is given by

$$\nabla S(\mathbf{r}) \cdot d\mathbf{r} = \mathbf{n}(\mathbf{r}) \cdot d\mathbf{r} = n(\mathbf{r})\, ds\,. \tag{2.17}$$

The geometric relationship between wave fronts and ray paths is illustrated in Figure 2.

Theorem

2.8 (Huygens–Fermat complementarity)**.** *Fermat's eikonal equation (2.7) follows from the Huygens wavelet equation (2.17)*

$$\nabla S(\mathbf{r}) = n(\mathbf{r})\frac{d\mathbf{r}}{ds} \qquad \textbf{\textit{(Huygens' equation)}} \tag{2.18}$$

by differentiating along the ray path.

Corollary

2.9. *The wave front level sets $S(\mathbf{r}) = constant$ and the ray paths $\mathbf{r}(s)$ are mutually orthogonal.*

Proof. The corollary follows once Equation (2.18) is proved, because $\nabla S(\mathbf{r})$ is along the ray vector and is perpendicular to the level set of $S(\mathbf{r})$. □

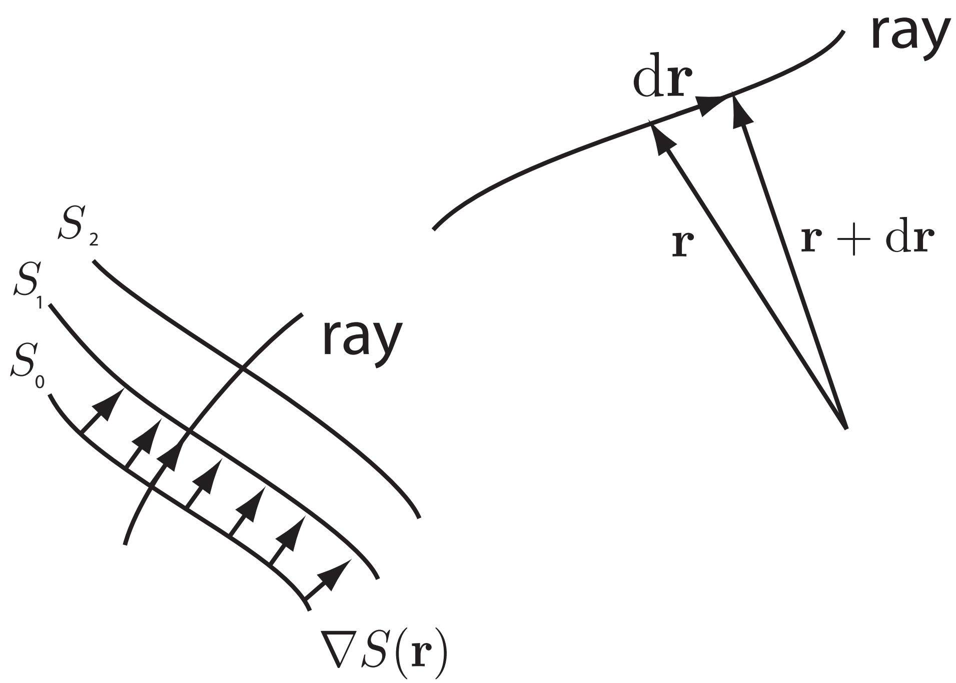


Figure 2: Huygens wave front and one of its corresponding ray paths. The wave front and ray path form mutually orthogonal families of curves. The gradient $\nabla S(\mathbf{r})$ is normal to the wave front and tangent to the ray through it at the point $\mathbf{r}$.

Proof. Theorem 2.8 may be proved by a direct calculation applying the operation

$$\frac{d}{ds} = \frac{d\mathbf{r}}{ds} \cdot \nabla = \frac{1}{n} \nabla S \cdot \nabla$$

to Huygens' equation (2.18). This yields the eikonal equation (2.7), by the following reasoning:

$$\frac{d}{ds}\left(n \frac{d\mathbf{r}}{ds} \right) = \frac{1}{n} \nabla S \cdot \nabla (\nabla S) = \frac{1}{2n} \nabla |\nabla S|^2 = \frac{1}{2n} \nabla n^2 = \nabla n \,.$$

In this chain of equations, the first step substitutes

$$d/ds = n^{-1}\nabla S \cdot \nabla\,.$$

The second step exchanges the order of derivatives. The third step uses the modulus of Huygens' equation (2.18) and invokes the property that $|d\mathbf{r}/ds|^2 = 1$. □

Corollary

2.10. *The modulus of Huygens' equation (2.18) yields*

$$|\nabla S|^2(\mathbf{r}) = n^2(\mathbf{r}) \qquad \textbf{\textit{(scalar eikonal equation)}} \tag{2.19}$$

which follows because $d\mathbf{r}/ds = \hat{\mathbf{s}}$ *in Equation (2.7) is a unit vector.*

Remark

2.11 (Hamilton–Jacobi equation)**.** *Corollary 2.10 arises as an algebraic result in the present considerations. However, it also follows at a more fundamental level from Maxwell's equations for electrodynamics in the slowly varying amplitude approximation of geometric optics, cf. Born and Wolf [1965, Chapter 3. See Keller [1962] for the modern extension of geometric optics to include diffraction. The scalar eikonal equation (2.19) is also known as the steady* ***Hamilton–Jacobi equation****.*

Theorem

2.12 (Ibn Sahl–Snell law of refraction)**.** *The gradient in Huygens' equation (2.18) defines the ray vector*

$$\mathbf{n} = \nabla S = n(\mathbf{r})\hat{\mathbf{s}} \tag{2.20}$$

of magnitude $|\mathbf{n}| = n$*. Integration of this gradient around a closed path vanishes, thereby yielding*

$$\oint_P \nabla S(\mathbf{r}) \cdot d\mathbf{r} = \oint_P \mathbf{n}(\mathbf{r}) \cdot d\mathbf{r} = 0\,. \tag{2.21}$$

Let's consider the case in which the closed path P *surrounds a boundary separating two different media. If we let the sides of the loop perpendicular to the interface shrink to zero, then only the parts of the line integral tangential to the interface path will contribute. Since these contributions must sum to zero, the tangential components of the ray vectors must be preserved. That is,*

$$(\mathbf{n} - \mathbf{n}') \times \hat{\mathbf{z}} = 0\,, \tag{2.22}$$

where the primes refer to the side of the boundary into which the ray is transmitted, whose normal vector is $\hat{\mathbf{z}}$*. Now imagine a ray piercing the boundary and passing into the region enclosed by the integration loop. If* θ *and* θ' *are the angles of incidence and transmission, measured from the direction* $\hat{\mathbf{z}}$ *normal to the boundary, then preservation of the tangential components of the ray vector means that, as in Fig. 1.1,*

$$n \sin\theta = n' \sin\theta'\,. \tag{2.23}$$

*This is the **Ibn Sahl–Snell law of refraction**, credited to Ibn Sahl (984) and Willebrord Snellius (1621). A similar analysis may be applied in the case of a reflected ray to show that the angle of incidence must equal the angle of reflection.*

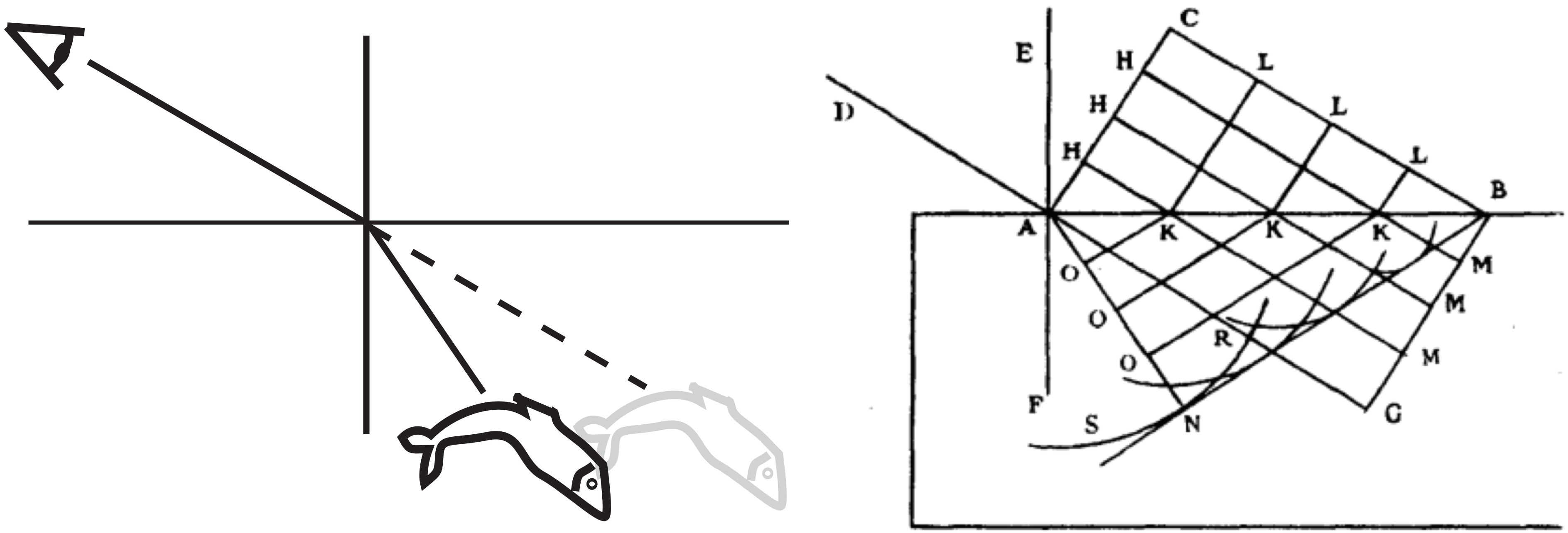


Figure 3: Ray tracing version (left) and Huygens' version (right) of the Ibn Sahl–Snell law of refraction that $n \sin\theta = n' \sin\theta'$. This law is implied in ray optics by preservation of the components of the ray vector tangential to the interface. According to Huygens' principle the law of refraction is implied by slower wave propagation in media of higher refractive index below the horizontal interface.

Remark

2.13 (Momentum form of Ibn Sahl–Snell law). *The phenomenon of refraction may be seen as a break in the direction* $\widehat{\mathbf{s}}$ *of the ray vector* $\mathbf{n}(\mathbf{r}(s)) = n\widehat{\mathbf{s}}$ *at a finite discontinuity in the refractive index* $n = |\mathbf{n}|$ *along the ray path* $\mathbf{r}(s)$*. According to the eikonal equation (2.7) the jump (denoted by* Δ*) in three-dimensional canonical momentum across the discontinuity must satisfy*

$$\Delta\left(\frac{\partial L}{\partial(d\mathbf{r}/ds)}\right) \times \frac{\partial n}{\partial \mathbf{r}} = 0\,.$$

This means the projections $\mathbf{p}$ *and* $\mathbf{p}'$ *of the ray vectors* $\mathbf{n}(\mathbf{q},z)$ *and* $\mathbf{n}'(\mathbf{q},z)$ *which lie tangent to the plane of the discontinuity in refractive index will be invariant. In particular, the lengths of these projections will be preserved. Consequently,*

$$|\mathbf{p}| = |\mathbf{p}'| \quad \textit{implies} \quad n\sin\theta = n'\sin\theta' \quad \textit{at } z = 0.$$

This is again the Ibn Sahl–Snell law, now written in terms of canonical momentum.

Exercise. How do the canonical momenta differ in the two versions of Fermat's principle in (2.6) and (2.11)? Do their Ibn Sahl–Snell laws differ? Do their Hamiltonian formulations differ? ★

Answer. The first stationary principle (2.6) gives $n(\mathbf{r}(s))d\mathbf{r}/ds$ for the optical momentum, while the second one (2.11) gives its reparameterised version $n^2(\mathbf{r}(\tau))d\mathbf{r}/d\tau$. Because $d/ds = nd/d\tau$, the values of the two versions of optical momentum agree in either parameterisation. Consequently, their Ibn Sahl–Snell laws agree. ▲

Remark

2.14. *As Newton discovered in his famous prism experiment, the propagation of a Huygens wave front depends on the light frequency, ω, through the frequency dependence $n(\mathbf{r},\omega)$ of the index of refraction of the medium. Having noted this possibility now, in what follows, we shall treat monochromatic light of fixed frequency ω and ignore the effects of frequency dispersion. We will also ignore finite-wavelength effects such as interference and diffraction of light. These effects were discovered in a sequence of pioneering scientific investigations during the 350 years after Fermat.*

2.3 Eikonal equation for axial ray optics

Most optical instruments are designed to possess a line of sight (or primary direction of propagation of light) called the ***optical axis***. Choosing coordinates so that the z-axis coincides with the optical axis expresses the arc-length element ds in terms of the increment along the optical axis, dz, as

$$\begin{aligned} ds &= [(dx)^2 + (dy)^2 + (dz)^2]^{1/2} \\ &= [1 + \dot{x}^2 + \dot{y}^2]^{1/2}\, dz = \frac{1}{\gamma}\, dz\,, \end{aligned} \tag{2.24}$$

in which the added notation defines $\dot{x} := dx/dz$, $\dot{y} := dy/dz$ and $\gamma := dz/ds$.

For such optical instruments, formula (2.24) for the element of arc length may be used to express the optical length in Fermat's principle as an integral called the ***optical action***,

$$\mathsf{A} := \int_{z_A}^{z_B} L(x, y, \dot{x}, \dot{y}, z)\, dz\,. \tag{2.25}$$

Definition

2.15. (Tangent vectors: Tangent bundle) *The coordinates* $(x, y, \dot{x}, \dot{y}) \in \mathbb{R}^2 \times \mathbb{R}^2$ *designate points along the ray path through the* ***configuration space*** *and the* ***tangent space*** *of possible vectors along the ray trajectory. The position on a ray passing through an image plane at a fixed value of* z *is denoted* $(x, y) \in \mathbb{R}^2$*. The notation* $(x, y, \dot{x}, \dot{y}) \in T\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2$ *for the combined space of positions and tangent vectors designates the* ***tangent bundle*** *of* $\mathbb{R}^2$*. The space* $T\mathbb{R}^2$ *consists of the union of all the position vectors* $(x, y) \in \mathbb{R}^2$ *and all the possible tangent vectors* $(\dot{x}, \dot{y}) \in \mathbb{R}^2$ *at each position* (x, y)*.*

The integrand in the optical action A is expressed as $L : T\mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$, or explicitly as

$$L(x, y, \dot{x}, \dot{y}, z) = n(x, y, z)[1 + \dot{x}^2 + \dot{y}^2]^{1/2} = \frac{n(x, y, z)}{\gamma}\,.$$

This is the ***optical Lagrangian***, in which

$$\gamma := \frac{1}{\sqrt{1 + \dot{x}^2 + \dot{y}^2}} \le 1\,.$$

We may write the coordinates equivalently as $(x, y, z) = (\mathbf{q}, z)$ where $\mathbf{q} = (x, y)$ is a vector with components in the plane perpendicular to the optical axis at displacement z.

The possible ray paths from point A to point B in space may be parameterised for axial ray optics as a family of C^2 curves $\mathbf{q}(z, \varepsilon) \in \mathbb{R}^2$ depending smoothly on a real parameter ε in an interval that includes $\varepsilon = 0$. The ε-family of paths $\mathbf{q}(z, \varepsilon)$ defines a set of smooth transformations of the ray path $\mathbf{q}(z)$. These transformations are taken to satisfy

$$\mathbf{q}(z, 0) = \mathbf{q}(z)\,, \quad \mathbf{q}(z_A, \varepsilon) = \mathbf{q}(z_A)\,, \quad \mathbf{q}(z_B, \varepsilon) = \mathbf{q}(z_B)\,, \tag{2.26}$$

so $\varepsilon = 0$ is the identity transformation of the ray path and its endpoints are left fixed. Define the variation of the optical action (2.25) using this parameter as

$$\begin{aligned} \delta \mathsf{A} &= \delta \int_{z_A}^{z_B} L(\mathbf{q}(z), \dot{\mathbf{q}}(z)) dz \\ &:= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int_{z_A}^{z_B} L(\mathbf{q}(z, \varepsilon), \dot{\mathbf{q}}(z, \varepsilon)) dz\,. \end{aligned} \tag{2.27}$$

In this formulation, Fermat's principle is expressed as the ***stationary condition*** $\delta \mathsf{A} = 0$ under infinitesimal variations of the path. This condition implies the axial eikonal equation, as follows.

Theorem

2.16 (Fermat's principle for axial eikonal equation)**.** *Stationarity under variations of the ac-*

tion A,

$$0 = \delta \mathsf{A} = \delta \int_{z_A}^{z_B} L(\mathbf{q}(z), \dot{\mathbf{q}}(z)) dz \,, \tag{2.28}$$

for the optical Lagrangian

$$L(\mathbf{q}, \dot{\mathbf{q}}, z) = n(\mathbf{q}, z)[1 + |\dot{\mathbf{q}}|^2]^{1/2} =: \frac{n}{\gamma} \,, \tag{2.29}$$

with

$$\gamma := \frac{dz}{ds} = \frac{1}{\sqrt{1 + |\dot{\mathbf{q}}|^2}} \leq 1 \,, \tag{2.30}$$

implies the ***axial eikonal equation***

$$\gamma \frac{d}{dz}\Big(n(\mathbf{q}, z)\, \gamma \frac{d\mathbf{q}}{dz} \Big) = \frac{\partial n}{\partial \mathbf{q}} \,, \quad \textit{with} \quad \frac{d}{ds} = \gamma \frac{d}{dz} \,. \tag{2.31}$$

Proof. As in the derivation of the eikonal equation (2.7), differentiating with respect to ε under the integral sign, denoting the variational derivative as

$$\delta \mathbf{q}(z) := \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathbf{q}(z, \varepsilon) \tag{2.32}$$

and integrating by parts produces the following variation of the optical action:

$$
\begin{aligned}
0 = \delta \mathsf{A} &= \delta \int L(\mathbf{q}, \dot{\mathbf{q}}, z) dz \\
&= \int \left(\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dz} \frac{\partial L}{\partial \dot{\mathbf{q}}} \right) \cdot \delta \mathbf{q} \, dz + \left[\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta \mathbf{q} \right]_{z_A}^{z_B} .
\end{aligned}
\tag{2.33}
$$

In the second line, one assumes equality of cross derivatives, $\mathbf{q}_{z\varepsilon} = \mathbf{q}_{\varepsilon z}$ evaluated at $\varepsilon = 0$, and thereby exchanges the order of derivatives; so that

$$
\delta \dot{\mathbf{q}} = \frac{d}{dz} \delta \mathbf{q} \, .
$$

The endpoint terms vanish in the ensuing integration by parts, because $\delta \mathbf{q}(z_A) = 0 = \delta \mathbf{q}(z_B)$. That is, the variation in the ray path must vanish at the prescribed spatial points A and B at z_A and z_B along the optical axis. Since $\delta \mathbf{q}$ is otherwise arbitrary, the principle of stationary action expressed in Equation (2.28) is equivalent to the following equation, written in a standard form later made famous by Euler and Lagrange:

$$
\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dz} \frac{\partial L}{\partial \dot{\mathbf{q}}} = 0 \quad \text{(Euler–Lagrange equation)} \, . \tag{2.34}
$$

After a short algebraic manipulation using the explicit form of the optical Lagrangian in (2.29), the Euler–Lagrange equation (2.34) for the light rays yields the eikonal equation (2.31), in which $\gamma \, d/dz = d/ds$ relates derivatives along the optical axis to derivatives in the arc-length parameter. □

Exercise. Check that the eikonal equation (2.31) follows from the Euler–Lagrange equation (2.34) with the Lagrangian (2.29). ★

Corollary

2.17 (Noether's theorem)**.** *Each smooth symmetry of the Lagrangian in an action principle implies a conservation law for its Euler–Lagrange equation Noether [1918].*

Proof. In a stationary action principle $\delta\mathsf{A} = 0$ for $\mathsf{A} = \int L\,dz$ as in (2.28), the Lagrangian L has a symmetry if it is *invariant* under the transformations $\mathbf{q}(z,0) \to \mathbf{q}(z,\varepsilon)$. In this case, stationarity $\delta\mathsf{A} = 0$ under the infinitesimal variations defined in (2.32) follows because of this invariance of the Lagrangian, even if these variations did *not* vanish at the endpoints in time. The variational calculation (2.33)

$$0 = \delta\mathsf{A} = \int \underbrace{\left(\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dz}\frac{\partial L}{\partial \dot{\mathbf{q}}} \right)}_{\text{Euler–Lagrange}} \cdot \delta\mathbf{q}\,dz + \underbrace{\left[\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q} \right]_{z_A}^{z_B}}_{\text{Noether}} \tag{2.35}$$

then shows that along the solution paths of the Euler–Lagrange equation (2.34) any smooth symmetry of the Lagrangian L implies

$$\left[\frac{\partial L}{\partial \dot{\mathbf{q}}} \cdot \delta\mathbf{q} \right]_{z_A}^{z_B} = 0.$$

Thus, the quantity $\delta\mathbf{q}\cdot(\partial L/\partial\dot{\mathbf{q}})$ is a constant of the motion (i.e., it is constant along the solution paths of the Euler–Lagrange equation) whenever $\delta\mathsf{A}=0$, because of the symmetry of the Lagrangian L in the action $\mathsf{A}=\int L\,dt$. □

Exercise. What does Noether's theorem imply for symmetries of the action principle given by $\delta\mathsf{A}=0$ for the following action?

$$\mathsf{A}=\int_{z_A}^{z_B}L(\dot{\mathbf{q}}(z))dz\,.$$

★

Answer. In this case, $\partial L/\partial\mathbf{q}=0$. This means the action A is invariant under translations $\mathbf{q}(z)\to\mathbf{q}(z)+\boldsymbol{\varepsilon}$ for any constant vector $\boldsymbol{\varepsilon}$. Setting $\delta\mathbf{q}=\boldsymbol{\varepsilon}$ in the Noether theorem associates conservation of any component of $\mathbf{p}:=(\partial L/\partial\dot{\mathbf{q}})$ with invariance of the action under spatial translations in that direction. For this case, the conservation law also follows immediately from the Euler–Lagrange equation (2.34). ▲

2.4 The eikonal equation for mirages

Air adjacent to a hot surface rises in temperature and becomes less dense (Figure 4). Thus over a flat hot surface, such as a desert expanse or a sun-baked roadway, air density locally increases with height and the

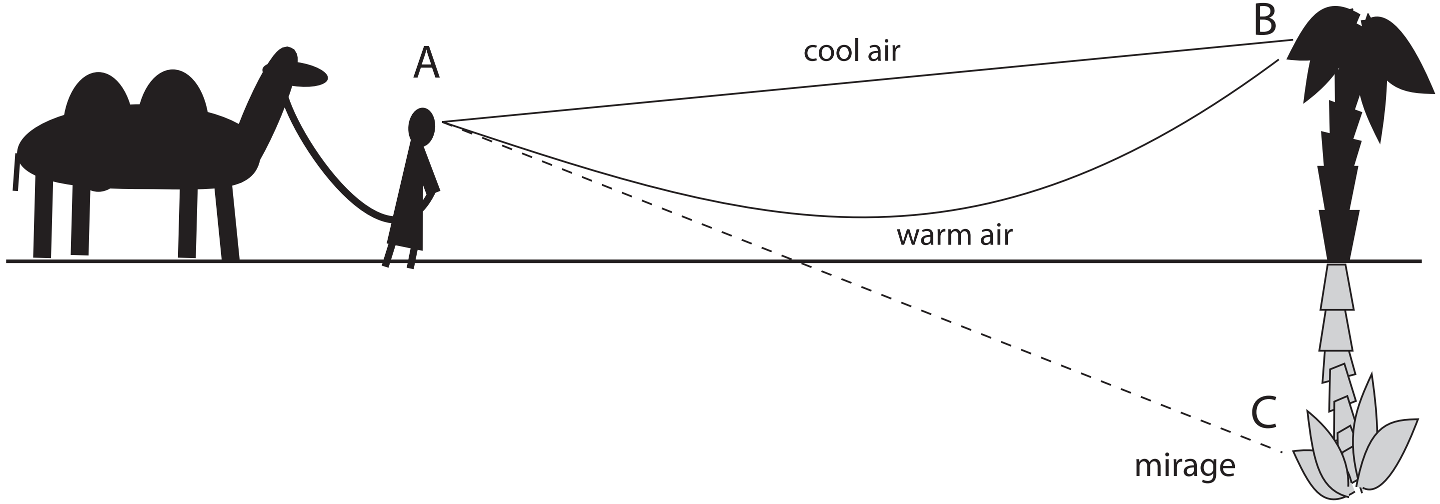


Figure 4: Fermat's principle states that the ray path from an observer at A to point B in space is a stationary path of optical length. For example, along a sun-baked road, the temperature of the air is warmest near the road and decreases with height, so that the index of refraction, n, increases in the vertical direction. For an observer at A, the curved path has the *same* optical path length as the straight line. Therefore, he sees not only the direct line-of-sight image of the tree top at B, but it also appears to him that the tree top has a mirror image at C. If there is no tree, the observer sees a direct image of the sky and also its mirror image of the same optical length, thereby giving the impression, perhaps sadly, that he is looking at water, when there is none.

average refractive index may be approximated by a linear variation of the form

$$n(x) = n_0(1 + \kappa x)\,,$$

where x is the vertical height above the planar surface, n_0 is the refractive index at ground level and κ is a positive constant. We may use the eikonal equation (2.31) to find an equation for the approximate ray trajectory. This will be an equation for the ray height x as a function of ground distance z of a light ray launched from a height x_0 at an angle θ_0 with respect to the horizontal surface of the earth.

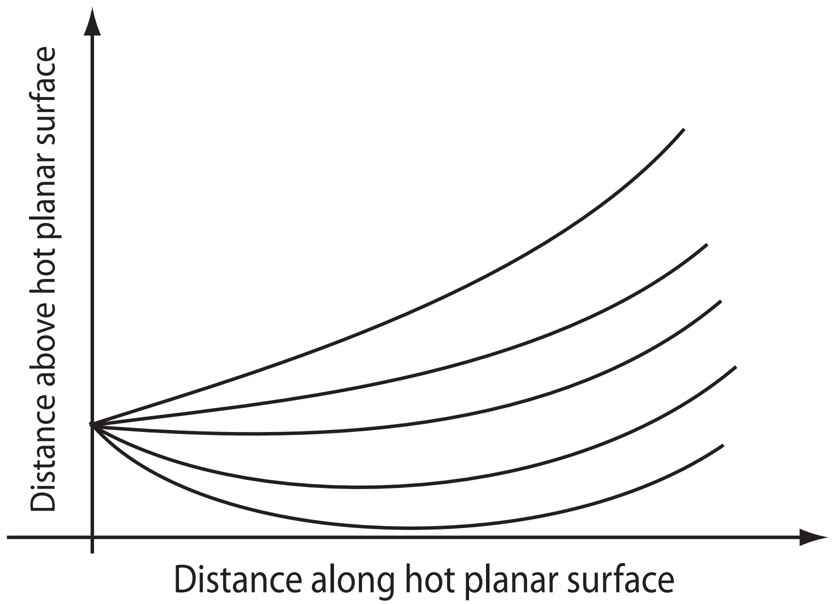


Figure 5: Ray trajectories are diverted in a spatially varying medium whose refractive index increases with height above a hot planar surface.

In this geometry, the eikonal equation (2.31) implies

$$\frac{1}{\sqrt{1+\dot{x}^2}}\frac{d}{dz}\left(\frac{(1+\kappa x)}{\sqrt{1+\dot{x}^2}}\,\dot{x}\right) = \kappa\,.$$

For nearly horizontal rays, $\dot{x}^2 \ll 1$, and if the variation in refractive index is also small then $\kappa x \ll 1$. In this case, the eikonal equation simplifies considerably to

$$\frac{d^2x}{dz^2} \approx \kappa \quad \text{for} \quad \kappa x \ll 1 \quad \text{and} \quad \dot{x}^2 \ll 1\,. \tag{2.36}$$

Thus, the ray trajectory is given approximately by

$$\begin{aligned}\mathbf{r}(z) &= x(z)\,\hat{\mathbf{x}} + z\,\hat{\mathbf{z}}\\ &= \Big(\frac{\kappa}{2}z^2 + \tan\theta_0 z + x_0\Big)\hat{\mathbf{x}} + z\,\hat{\mathbf{z}}\,.\end{aligned}$$

The resulting parabolic divergence of rays above the hot surface is shown in Figure 5.

Exercise. Explain how the ray pattern would differ from the rays shown in Figure 5 if the refractive index were *decreasing* with height x above the surface, rather than increasing. ★

2.5 Paraxial optics and classical mechanics

Rays whose direction is nearly along the optical axis are called ***paraxial***. In a medium whose refractive index is nearly homogeneous, paraxial rays remain paraxial and geometric optics closely resembles classical mechanics. Consider the trajectories of paraxial rays through a medium whose refractive index may be approximated by

$$n(\mathbf{q},z) = n_0 - \nu(\mathbf{q},z)\,,\quad \text{with}\quad \nu(\mathbf{0},z)=0 \quad\text{and}\quad \nu(\mathbf{q},z)/n_0 \ll 1\,.$$

Being nearly parallel to the optical axis, paraxial rays satisfy $\theta \ll 1$ and $|\mathbf{p}|/n \ll 1$; so the optical Hamiltonian (3.9) may then be approximated by

$$H = -\,n\left[1-\frac{|\mathbf{p}|^2}{n^2}\right]^{1/2} \simeq\; -\,n_0 + \frac{|\mathbf{p}|^2}{2n_0} + \nu(\mathbf{q},z)\,.$$

The constant n_0 here is immaterial to the dynamics. This calculation shows the following.

Lemma

2.18. *Geometric ray optics in the paraxial regime corresponds to classical mechanics with a time-dependent potential* $\nu(\mathbf{q}, z)$, *upon identifying* $z \leftrightarrow t$.

Exercise. Show that the canonical equations for paraxial rays recover the mirage equation (2.36) when $n = n_s(1 + \kappa x)$ for $\kappa > 0$.

Explain what happens to the ray pattern when $\kappa < 0$. ★

3 Lecture 2: Hamiltonian formulation of axial ray optics

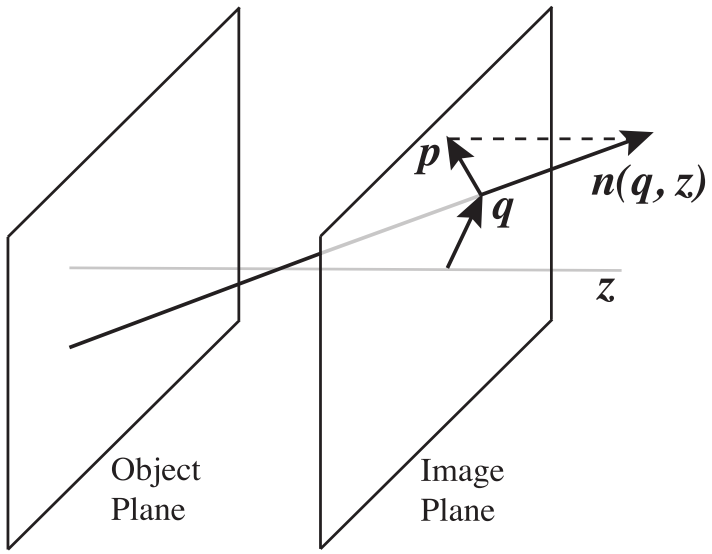


Figure 6: Geometrically, the momentum $\mathbf{p}$ associated with the coordinate $\mathbf{q}$ by Equation (3.1) on the image plane at z turns out to be the projection onto the plane of the ray vector $\mathbf{n}(\mathbf{q}, z) = \nabla S = n(\mathbf{q}, z)d\mathbf{r}/ds$ passing through the point $\mathbf{q}(z)$. That is, $|\mathbf{p}| = n(\mathbf{q}, z)\sin\theta$, where $\cos\theta = dz/ds$ is the direction cosine of the ray with respect to the optical z-axis.

Definition

3.1 (Canonical momentum)**.** *The canonical momentum (denoted as* $\mathbf{p}$*) associated with the ray path position* $\mathbf{q}$ *in an* ***image plane****, or* ***image screen****, at a fixed value of* z *along the optical axis is defined as*

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}\,, \quad \textit{with} \quad \dot{\mathbf{q}} := \frac{d\mathbf{q}}{dz}\,. \tag{3.1}$$

Remark

3.2. *For the optical Lagrangian (2.29), the corresponding canonical momentum for axial ray optics is found to be*

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} = n\gamma\,\dot{\mathbf{q}}\,, \quad \textit{which satisfies} \quad |\mathbf{p}|^2 = n^2(1-\gamma^2)\,. \tag{3.2}$$

Figure 6 illustrates the geometrical interpretation of this momentum for optical rays as the projection along the optical axis of the ray onto an image plane.

Remark

3.3. *From the definition of optical momentum (3.2), the corresponding* ***velocity*** $\dot{\mathbf{q}} = d\mathbf{q}/dz$ *is found as a function of position and momentum* $(\mathbf{q}, \mathbf{p})$ *as*

$$\dot{\mathbf{q}} = \frac{\mathbf{p}}{\sqrt{n^2(\mathbf{q}, z) - |\mathbf{p}|^2}}\,. \tag{3.3}$$

A Lagrangian admitting such an invertible relation between $\dot{\mathbf{q}}$ *and* $\mathbf{p}$ *is said to be* ***nondegenerate*** *(or* ***hyperregular*** *[MaRa1994]). Moreover, the velocity is real-valued, provided*

$$n^2 - |\mathbf{p}|^2 > 0\,. \tag{3.4}$$

The latter condition is explained geometrically, as follows.

3.1 Geometry, phase space and the ray path

Huygens' equation (2.18) summons the following geometric picture of the ray path, as shown in Figure 6. Along the optical axis (the z-axis) each image plane normal to the axis is pierced at a point $\mathbf{q} = (x, y)$ by the ***ray vector***, defined as

$$\mathbf{n}(\mathbf{q}, z) = \nabla S = n(\mathbf{q}, z)\frac{d\mathbf{r}}{ds}\,.$$

The ray vector is tangent to the ray path and has magnitude $n(\mathbf{q}, z)$. This vector makes an angle $\theta(z)$ with respect to the z-axis at point $\mathbf{q}$. Its direction cosine with respect to the z-axis is given by

$$\cos\theta := \hat{\mathbf{z}} \cdot \frac{d\mathbf{r}}{ds} = \frac{dz}{ds} = \gamma\,. \tag{3.5}$$

This definition of $\cos\theta$ leads by (3.2) to

$$|\mathbf{p}| = n\sin\theta \quad \text{and} \quad \sqrt{n^2 - |\mathbf{p}|^2} = n\cos\theta\,. \tag{3.6}$$

Thus, the projection of the ray vector $\mathbf{n}(\mathbf{q}, z)$ onto the image plane is the momentum $\mathbf{p}(z)$ of the ray (Figure 1.6). In three-dimensional vector notation, this is expressed as

$$\mathbf{p}(z) = \mathbf{n}(\mathbf{q}, z) - \hat{\mathbf{z}}\Big(\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z)\Big)\,. \tag{3.7}$$

The coordinates $(\mathbf{q}(z), \mathbf{p}(z))$ determine each ray's position and orientation completely as a function of propagation distance z along the optical axis.

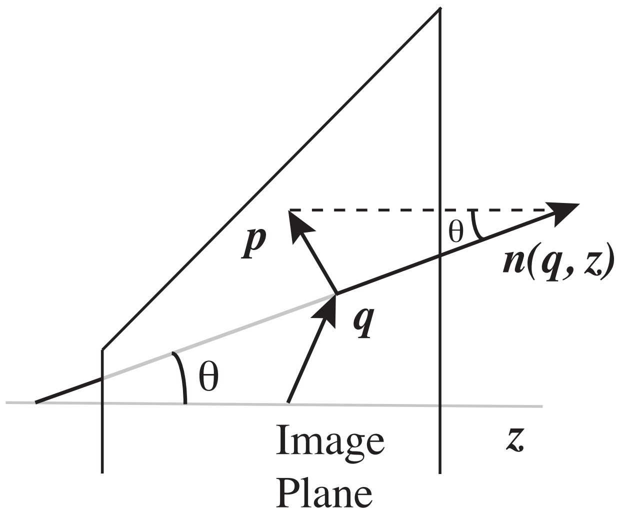


Figure 7: The canonical momentum $\mathbf{p}$ associated with the coordinate $\mathbf{q}$ by Equation (3.1) on the image plane at z has magnitude $|\mathbf{p}| = n(\mathbf{q}, z)\sin\theta$, where $\cos\theta = dz/ds$ is the direction cosine of the ray with respect to the optical z-axis.

Definition

3.4 (Optical phase space, or cotangent bundle)**.** *The coordinates* $(\mathbf{q}, \mathbf{p}) \in \mathbb{R}^2 \times \mathbb{R}^2$ *comprise the* ***phase-space*** *of the ray trajectory. The position on an image plane is denoted* $\mathbf{q} \in \mathbb{R}^2$*. Phase-space coordinates are denoted* $(\mathbf{q}, \mathbf{p}) \in T^*\mathbb{R}^2$*. The notation* $T^*\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2$ *for phase space designates the* ***cotangent bundle*** *of* $\mathbb{R}^2$*. The space* $T^*\mathbb{R}^2$ *consists of the union of all the position vectors* $\mathbf{q} \in \mathbb{R}^2$ *and all the possible canonical momentum vectors* $\mathbf{p} \in \mathbb{R}^2$ *at each position* $\mathbf{q}$*.*

Remark

3.5. *The phase space* $T^*\mathbb{R}^2$ *for ray optics is restricted to the disc*

$$|\mathbf{p}| < n(\mathbf{q}, z)\,,$$

so that $\cos\theta$ *in (3.6) remains real. When* $n^2 = |\mathbf{p}|^2$*, the ray trajectory is tangent to the image screen and is said to have* ***grazing incidence*** *to the screen at a certain value of* z*. Rays of grazing incidence are eliminated by restricting the momentum in the phase space for ray optics to lie in a disc* $|\mathbf{p}|^2 < n^2(\mathbf{q}, z)$*. This restriction implies that the velocity will remain real, finite and of a single sign, which we may choose to be positive* $(\dot{\mathbf{q}} > 0)$ *in the direction of propagation.*

3.2 Legendre transformation

The passage from the description of the eikonal equation for ray optics in variables $(\mathbf{q}, \dot{\mathbf{q}}, z)$ to its phase-space description in variables $(\mathbf{q}, \mathbf{p}, z)$ is accomplished by applying the ***Legendre transformation*** from the Lagrangian L to the Hamiltonian H, defined as

$$H(\mathbf{q}, \mathbf{p}) = \mathbf{p} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}, z)\,. \tag{3.8}$$

For the Lagrangian (2.29) the Legendre transformation (3.8) leads to the following ***optical Hamiltonian***,

$$H(\mathbf{q}, \mathbf{p}) = n\gamma\, |\dot{\mathbf{q}}|^2 - n/\gamma = -n\gamma = -\left[n(\mathbf{q}, z)^2 - |\mathbf{p}|^2 \right]^{1/2}, \tag{3.9}$$

upon using formula (3.3) for the velocity $\dot{\mathbf{q}}(z)$ in terms of the position $\mathbf{q}(z)$ at which the ray punctures the screen at z and its canonical momentum there is $\mathbf{p}(z)$. Thus, in the geometric picture of canonical screen optics in Figure 6, the component of the ray vector along the optical axis is (minus) the Hamiltonian. That is,

$$\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z) = n(\mathbf{q}, z) \cos\theta = -H\,. \tag{3.10}$$

Remark

3.6. *The optical Hamiltonian in (3.9) takes real values, so long as the phase space for ray optics is restricted to the disc $|\mathbf{p}| \le n(\mathbf{q}, z)$. The boundary of this disc is the zero level set of the Hamiltonian, $H = 0$. Thus, flows that preserve the value of the optical Hamiltonian will remain inside its restricted phase space.*

Theorem

3.7. *The phase-space description of the ray path follows from* ***Hamilton's canonical equations****, which are defined as*

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}\,, \qquad \dot{\mathbf{p}} = -\,\frac{\partial H}{\partial \mathbf{q}}\,. \tag{3.11}$$

With the optical Hamiltonian $H(\mathbf{q},\mathbf{p}) = -\left[n(\mathbf{q},z)^2 - |\mathbf{p}|^2\right]^{1/2}$ *in (3.9), these are*

$$\dot{\mathbf{q}} = \frac{-1}{H}\mathbf{p}\,, \qquad \dot{\mathbf{p}} = \frac{-1}{2H}\frac{\partial n^2}{\partial \mathbf{q}}\,. \tag{3.12}$$

Proof. Hamilton's canonical equations are obtained by differentiating both sides of the Legendre transformation formula (3.8) to find

$$\begin{aligned} dH(\mathbf{q},\mathbf{p},z) &= \mathbf{0}\cdot d\dot{\mathbf{q}} + \frac{\partial H}{\partial \mathbf{p}}\cdot d\mathbf{p} + \frac{\partial H}{\partial \mathbf{q}}\cdot d\mathbf{q} + \frac{\partial H}{\partial z}\cdot dz \\ &= \left(\mathbf{p} - \frac{\partial L}{\partial \dot{\mathbf{q}}}\right)\cdot d\dot{\mathbf{q}} + \dot{\mathbf{q}}\cdot d\mathbf{p} - \frac{\partial L}{\partial \mathbf{q}}\cdot d\mathbf{q} - \frac{\partial L}{\partial z}\,dz\,. \end{aligned}$$

The coefficient of $(d\dot{\mathbf{q}})$ vanishes in this expression by virtue of the definition of canonical momentum. The vanishing of this coefficient is required for H to be independent of $\dot{\mathbf{q}}$. Identifying the other coefficients yields the relations

$$\frac{\partial H}{\partial \mathbf{p}} = \dot{\mathbf{q}}\,, \qquad \frac{\partial H}{\partial \mathbf{q}} = -\frac{\partial L}{\partial \mathbf{q}} = -\frac{d}{dz}\frac{\partial L}{\partial \dot{\mathbf{q}}} = -\,\dot{\mathbf{p}}\,, \tag{3.13}$$

and

$$\frac{\partial H}{\partial z} = -\,\frac{\partial L}{\partial z}\,, \tag{3.14}$$

in which one uses the Euler–Lagrange equation to derive the second relation. Hence, one finds the canonical Hamiltonian formulas (3.11) in Theorem 3.7. □

Definition

3.8 (Canonical momentum)**.** *The momentum* $\mathbf{p}$ *defined in (3.1) that appears with the position* $\mathbf{q}$ *in Hamilton's canonical equations (3.11) is called the* ***canonical momentum****.*

4 Hamiltonian form of optical transmission

Proposition

4.1 (Canonical bracket)**.** *Hamilton's canonical equations (3.11) arise from a bracket operation,*

$$\{F,\,H\} = \frac{\partial F}{\partial \mathbf{q}} \cdot \frac{\partial H}{\partial \mathbf{p}} - \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{\partial F}{\partial \mathbf{p}}\,, \tag{4.1}$$

expressed in terms of position $\mathbf{q}$ *and momentum* $\mathbf{p}$*.*

Proof. One directly verifies

$$\dot{\mathbf{q}} = \{\mathbf{q},\,H\} = \frac{\partial H}{\partial \mathbf{p}} \quad \text{and} \quad \dot{\mathbf{p}} = \{\mathbf{p},\,H\} = -\frac{\partial H}{\partial \mathbf{q}}\,.$$

□

Definition

4.2 (Canonically conjugate variables)**.** *The components* q_i *and* p_j *of position* $\mathbf{q}$ *and momentum* $\mathbf{p}$ *satisfy*

$$\{q_i,\,p_j\} = \delta_{ij}\,, \tag{4.2}$$

with respect to the canonical bracket operation (4.1). Variables that satisfy this relation are said to be ***canonically conjugate****.*

Definition

4.3 (Dynamical systems in Hamiltonian form)**.** *A dynamical system on the tangent space TM of a space M*

$$\dot{\mathbf{x}}(t) = \mathbf{F}(\mathbf{x})\,, \quad \mathbf{x} \in M\,,$$

is said to be in ***Hamiltonian form*** *if it can be expressed as*

$$\dot{\mathbf{x}}(t) = \{\mathbf{x}, H\}\,, \quad \textit{for} \quad H : M \to \mathbb{R}\,, \tag{4.3}$$

in terms of a Poisson bracket operation $\{\cdot\,,\,\cdot\}$, which is a map among smooth real functions $\mathcal{F}(M) : M \to \mathbb{R}$ on M,

$$\{\cdot\,,\,\cdot\} : \mathcal{F}(M) \times \mathcal{F}(M) \to \mathcal{F}(M)\,, \tag{4.4}$$

so that $\dot{F} = \{F\,,\,H\}$ for any $F \in \mathcal{F}(M)$.

Definition

4.4 (Poisson bracket)**.** *A* ***Poisson bracket operation*** $\{\cdot\,,\,\cdot\}$ *is defined as possessing the following properties:*

- *It is* ***bilinear****.*
- *It is* ***skew-symmetric****,* $\{F\,,\,H\} = -\,\{H\,,\,F\}$*.*
- *It satisfies the* ***Leibniz rule*** *(product rule),*

$$\{FG\,,\,H\} = \{F\,,\,H\}G + F\{G\,,\,H\}\,,$$

  for the product of any two functions F *and* G *on* M*.*

- *It satisfies the* ***Jacobi identity****,*

$$\{F\,,\,\{G\,,\,H\}\} + \{G\,,\,\{H\,,\,F\}\} + \{H\,,\,\{F\,,\,G\}\} = 0\,, \tag{4.5}$$

  for any three functions F*,* G *and* H *on* M*.*

Remark

4.5. *Definition 4.4 of the Poisson bracket certainly includes the* ***canonical Poisson bracket*** *in (4.1) that produces Hamilton's canonical equations (3.11), with position* $\mathbf{q}$ *and conjugate momentum* $\mathbf{p}$*. However, this definition does not require the Poisson bracket to be expressed in the canonical form (4.1).*

Exercise. Show that the defining properties of a Poisson bracket hold for the canonical bracket expression in (4.1). ★

Exercise. Compute the Jacobi identity (4.5) using the canonical Poisson bracket (4.1) in one dimension for $F = p$, $G = q$ and H arbitrary. ★

Exercise. What does the Jacobi identity (4.5) imply about $\{F\,,\,G\}$ when F and G are constants of motion, so that $\{F\,,\,H\} = 0$ and $\{G\,,\,H\} = 0$ for a Hamiltonian H? ★

Exercise. How do the Hamiltonian formulations differ in the two versions of Fermat's principle in (2.6) and (2.11)? ★

Answer. The two Hamiltonian formulations differ, because the Lagrangian in (2.6) is homogeneous of degree 1 in its velocity, while the Lagrangian in (2.11) is homogeneous of degree 2. Consequently, under the Legendre transformation, the Hamiltonian in the first formulation vanishes identically, while the other Hamiltonian is quadratic in its momentum, namely,

$$H = |\mathbf{p}|^2/(2n)^2 \,.$$

▲

Definition

4.6 (Hamiltonian vector fields and flows)**.** *A* ***Hamiltonian vector field*** X_F *is a map from a function* $F \in \mathcal{F}(M)$ *on space* M *with Poisson bracket* $\{\,\cdot\,,\,\cdot\,\}$ *to a tangent vector on its tangent space* TM *given by the Poisson bracket. When* M *is the optical phase space* $T^*\mathbb{R}^2$*, this map is given by the partial differential operator obtained by inserting the phase-space function* F *into the canonical Poisson bracket,*

$$X_F = \{\,\cdot\,,\,F\} = \frac{\partial F}{\partial \mathbf{p}} \cdot \frac{\partial}{\partial \mathbf{q}} - \frac{\partial F}{\partial \mathbf{q}} \cdot \frac{\partial}{\partial \mathbf{p}}\,.$$

The solution $\mathbf{x}(t) = \phi_t^F \mathbf{x}$ *of the resulting differential equation,*

$$\dot{\mathbf{x}}(t) = \{\,\mathbf{x}\,,\,F\} \quad \textit{with} \quad \mathbf{x} \in M\,,$$

yields the ***flow*** $\phi_t^F : \mathbb{R} \times M \to M$ *of the Hamiltonian vector field* X_F *on* M*. Assuming that it exists and is unique, the solution* $\mathbf{x}(t)$ *of the differential equation is a* ***curve*** *on* M *parameterised by* $t \in \mathbb{R}$*. The tangent vector* $\dot{\mathbf{x}}(t)$ *to the flow represented by the curve* $\mathbf{x}(t)$ *at time* t *satisfies*

$$\dot{\mathbf{x}}(t) = X_F \mathbf{x}(t)\,,$$

which is the ***characteristic equation*** *of the Hamiltonian vector field on manifold* M*.*

Remark

4.7 (Caution about caustics)**.** *Caustics were discussed definitively in a famous unpublished paper written by Hamilton in 1823 at the age of 18. Hamilton later published what he called "supplements" to this paper in 1830 and 1837. The optical singularities discussed by Hamilton form in bright* ***caustic*** *surfaces when light reflects off a curved mirror. In the present context, we shall avoid caustics. Indeed, we shall avoid reflection altogether and deal only with smooth Hamiltonian flows in media whose spatial variation in refractive index is smooth. For a modern discussion of caustics, see Arnold [1994].*

4.1 Translation-invariant media

- If $n = n(\mathbf{q})$, so that the medium is ***invariant under translations*** along the optical axis with coordinate z, then

$$\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z) = n(\mathbf{q}, z) \cos\theta = -\,H$$

  in (3.10) is ***conserved***. That is, the projection $\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z)$ of the ray vector along the optical axis is constant in translation-invariant media.

- For translation-invariant media, the eikonal equation (2.31) simplifies via the canonical equations (3.12) to ***Newtonian dynamics***,

$$\ddot{\mathbf{q}} = -\,\frac{1}{2H^2}\frac{\partial n^2}{\partial \mathbf{q}} \quad \text{for} \quad \mathbf{q} \in \mathbb{R}^2\,. \tag{4.6}$$

- Thus, in translation-invariant media, geometric ray tracing formally reduces to Newtonian dynamics in z, with a potential $-\,n^2(\mathbf{q})$ and with time z rescaled along each path by the constant value of $\sqrt{2}\,H$ determined from the initial conditions for each ray at the ***object screen*** at $z = 0$.

Remark

4.8. *In media for which the index of refraction is not translation-invariant, the optical Hamiltonian $n(\mathbf{q}, z)\cos\theta = -H$ is not generally conserved.*

4.2 Axisymmetric, translation-invariant materials

In axisymmetric, translation-invariant media, the index of refraction may depend on the distance from the optical axis, $r = |\mathbf{q}|$, but does not depend on the azimuthal angle. As we have seen, translation invariance implies conservation of the optical Hamiltonian. Axisymmetry implies yet another constant of motion. This additional constant of motion allows the Hamiltonian system for the light rays to be reduced to phase-plane analysis. For such media, the index of refraction satisfies

$$n(\mathbf{q}, z) = n(r)\,, \quad \text{where} \quad r = |\mathbf{q}|\,. \tag{4.7}$$

Passing to polar coordinates (r, ϕ) yields

$$\begin{aligned}
\mathbf{q} &= (x, y) = r(\cos\phi,\, \sin\phi)\,, \\
\mathbf{p} &= (p_x, p_y) \\
&= (p_r \cos\phi - p_\phi \sin\phi/r,\; p_r \sin\phi + p_\phi \cos\phi/r)\,,
\end{aligned}$$

so that

$$|\mathbf{p}|^2 = p_r^2 + p_\phi^2/r^2\,. \tag{4.8}$$

Consequently, the optical Hamiltonian,

$$H = -\left[n(r)^2 - p_r^2 - p_\phi^2/r^2\right]^{1/2}, \tag{4.9}$$

is *independent* of the azimuthal angle ϕ. This independence of angle ϕ leads to conservation of its canonically conjugate momentum p_ϕ, whose interpretation will be discussed in a moment.

Exercise. Verify formula (4.9) for the optical Hamiltonian governing ray optics in axisymmetric, translation-invariant media by computing the Legendre transformation. ★

Answer. Fermat's principle $\delta S = 0$ for $S = \int L\,dz$ in axisymmetric, translation-invariant material may be written in polar coordinates using the Lagrangian

$$L = n(r)\sqrt{1 + \dot{r}^2 + r^2\dot{\phi}^2}\,, \tag{4.10}$$

from which one finds

$$p_r = \frac{\partial L}{\partial \dot{r}} = \frac{n(r)\dot{r}}{\sqrt{1 + \dot{r}^2 + r^2\dot{\phi}^2}}\,,$$

and

$$\frac{p_\phi}{r} = \frac{1}{r}\frac{\partial L}{\partial \dot{\phi}} = \frac{n(r)r\dot{\phi}}{\sqrt{1 + \dot{r}^2 + r^2\dot{\phi}^2}}\,.$$

Consequently, the velocities and momenta are related by

$$\frac{1}{\sqrt{1 + \dot{r}^2 + r^2\dot{\phi}^2}} = \sqrt{1 - \frac{p_r^2 + p_\phi^2/r^2}{n^2(r)}} = \sqrt{1 - |\mathbf{p}|^2/n^2(r)}\,,$$

which allows the velocities to be obtained from the momenta and positions. The Legendre transformation (3.9),

$$H(r, p_r, p_\phi) = \dot{r}p_r + \dot{\phi}p_\phi - L(r, \dot{r}, \dot{\phi})\,,$$

then yields formula (4.9) for the optical Hamiltonian. ▲

Exercise. Interpret the quantity p_ϕ in terms of the vector phase-space variables $\mathbf{p}$ and $\mathbf{q}$. ★

Answer. The vector $\mathbf{q}$ points from the optical axis and lies in the optical (x, y) or (r, ϕ) plane. Hence, the quantity p_ϕ may be expressed in terms of the vector image-screen phase-space variables $\mathbf{p}$ and $\mathbf{q}$ as

$$|\mathbf{p}\times\mathbf{q}|^2 = |\mathbf{p}|^2|\mathbf{q}|^2 - (\mathbf{p}\cdot\mathbf{q})^2 = p_\phi^2\,. \tag{4.11}$$

This may be obtained by using the relations

$$|\mathbf{p}|^2 = p_r^2 + \frac{p_\phi^2}{r^2}\,, \quad |\mathbf{q}|^2 = r^2 \quad \text{and} \quad \mathbf{q}\cdot\mathbf{p} = rp_r\,.$$

One interprets $p_\phi = \mathbf{p}\times\mathbf{q}$ as the ***oriented area*** spanned on the optical screen by the vectors $\mathbf{q}$ and $\mathbf{p}$. ▲

Exercise. Show that Corollary 2.17 (Noether's theorem) implies conservation of the quantity p_ϕ for the axisymmetric Lagrangian (4.10) in polar coordinates. ★

Answer. The Lagrangian (4.10) is invariant under $\phi \to \phi + \varepsilon$ for constant ε. Noether's theorem then implies conservation of $p_\phi = \partial L/\partial\dot\phi$. ▲

Exercise. What conservation law does Noether's theorem imply for the invariance of the Lagrangian (4.10) under translations in time $t \to t + \epsilon$ for a real constant $\epsilon \in \mathbb{R}$? ★

4.3 Hamiltonian optics in polar coordinates

Hamilton's equations in polar coordinates are defined for axisymmetric, translation-invariant media by the canonical Poisson brackets with the optical Hamiltonian (4.9),

$$
\begin{aligned}
\dot{r} &= \{r,\, H\} = \frac{\partial H}{\partial p_r} = -\,\frac{p_r}{H}\,,\\
\dot{p}_r &= \{p_r,\, H\} = -\,\frac{\partial H}{\partial r} = -\,\frac{1}{2H}\frac{\partial}{\partial r}\left(n^2(r) - \frac{p_\phi^2}{r^2}\right)\,,\\
\dot{\phi} &= \{\phi,\, H\} = \frac{\partial H}{\partial p_\phi} = -\,\frac{p_\phi}{Hr^2}\,,\\
\dot{p}_\phi &= \{p_\phi,\, H\} = -\,\frac{\partial H}{\partial \phi} = 0\,.
\end{aligned}
\tag{4.12}
$$

In the Hamiltonian for axisymmetric ray optics (4.9), the constant of the motion p_ϕ may be regarded as a parameter that is set by the initial conditions. Consequently, the motion governed by (4.12) restricts to canonical Hamiltonian dynamics for $r(z)$, $p_r(z)$ in a reduced phase space.

Remark

4.9 (An equivalent reduction). *Alternatively, the level sets of the Hamiltonian* $H = const$ *and angular momentum* $p_\phi = const$ *may be regarded as two surfaces in* $\mathbb{R}^3$ *with coordinates* $\boldsymbol{\chi} = (r, p_r, p_\phi)$*. In these* $\mathbb{R}^3$ *coordinates, Hamilton's equations (4.12) may be written as*

$$\dot{\boldsymbol{\chi}}(t) = -\frac{\partial p_\phi}{\partial \boldsymbol{\chi}} \times \frac{\partial H}{\partial \boldsymbol{\chi}} \quad \textit{with} \quad \boldsymbol{\chi} \in \mathbb{R}^3 \,. \tag{4.13}$$

This means that the evolution in $\mathbb{R}^3$ *with coordinates* $\boldsymbol{\chi} = (r, p_r, p_\phi)$ *takes place along the intersections of the level sets of the constants of motion* p_ϕ *and* H*. On a level set of* p_ϕ *the* $\mathbb{R}^3$ *equations (4.13) restrict to the first two equations in the canonical system (4.12).*

Remark

4.10 (Evolution of azimuthal angle)**.** *The evolution of the azimuthal angle, or phase, $\phi(z)$ in polar coordinates for a given value of p_ϕ decouples from the rest of the equations in (4.12) and may be found separately, after solving the canonical equations for $r(z)$ and $p_r(z)$.*

The polar canonical equation for $\phi(z)$ in (4.12) implies, for a given orbit $r(z)$, that the phase may be obtained as a ***quadrature***,

$$\Delta\phi(z) = \int^z \frac{\partial H}{\partial p_\phi}\, dz = -\frac{p_\phi}{H}\int^z \frac{1}{r^2(z)}\, dz\,, \tag{4.14}$$

where p_ϕ and H are constants of the motion. Because in this case the integrand is a square, the polar azimuthal angle, or phase, $\Delta\phi(z)$ must either increase or decrease monotonically in axisymmetric ray optics, depending on whether the sign of the conserved ratio p_ϕ/H is negative, or positive, respectively. Moreover, for a fixed value of the ratio p_ϕ/H, rays that are closer to the optical axis circulate around it faster.

The reconstruction of the phase for solutions of Hamilton's optical equations (4.12) for ray paths in an axisymmetric, translation-invariant medium has some interesting geometric features for periodic orbits in the radial (r, p_r) phase plane.

4.4 Geometric phase for Fermat's principle

One may decompose the total phase change around a closed periodic orbit of period Z in the phase space of radial variables (r, p_r) into the sum of the following two parts:

$$\oint p_\phi \, d\phi = p_\phi \Delta\phi = \underbrace{-\oint p_r \, dr}_{\text{geometric}} + \underbrace{\oint \mathbf{p} \cdot d\mathbf{q}}_{\text{dynamic}} \, . \tag{4.15}$$

On writing this decomposition of the phase as

$$\Delta\phi = \Delta\phi_{geom} + \Delta\phi_{dyn} \, ,$$

one sees that

$$p_\phi \Delta\phi_{geom} = \frac{1}{H} \oint p_r^2 dz = -\iint dp_r \wedge dr \tag{4.16}$$

is the area enclosed by the periodic orbit in the radial phase plane. Thus the name ***geometric phase*** for $\Delta\phi_{geom}$, because this part of the phase only depends on the geometric area of the periodic orbit. The rest of the phase is given by

$$\begin{aligned} p_\phi \Delta\phi_{dyn} &= \oint \mathbf{p}\cdot d\mathbf{q} \\ &= \oint \left(p_r \frac{\partial H}{\partial p_r} + p_\phi \frac{\partial H}{\partial p_\phi} \right) dz \\ &= \frac{-1}{H} \oint \left(p_r^2 + \frac{p_\phi^2}{r^2} \right) dz \\ &= \frac{1}{H} \oint \big(H^2 - n^2(|\mathbf{q}(z)|) \big) dz \\ &= ZH - \frac{Z}{H}\langle n^2 \rangle \,, \end{aligned} \tag{4.17}$$

where the loop integral $\oint n^2(|\mathbf{q}(z)|)dz = Z\langle n^2\rangle$ defines the average $\langle n^2\rangle$ over the orbit of period Z of the squared index of refraction. This part of the phase depends on the Hamiltonian, orbital period and average of the squared index of refraction over the orbit. Thus the name ***dynamic phase*** for $\Delta\phi_{dyn}$, because this part of the phase depends on the dynamics of the orbit, not just its area.

4.5 Skewness

Definition

4.11. *The quantity*

$$p_\phi = \mathbf{p} \times \mathbf{q} = y p_x - x p_y \,, \tag{4.18}$$

*is called the **skewness function**.*[2]

Remark

4.12. *By (4.12) the skewness is conserved for rays in axisymmetric media.*

Remark

4.13. *Geometrically, the skewness given by the cross product $S = \mathbf{p} \times \mathbf{q}$ is the area spanned on an image screen by the vectors $\mathbf{p}$ and $\mathbf{q}$ (Figure 1.7). This geometric conservation law for screen optics was first noticed by Lagrange in paraxial lens optics and it is still called **Lagrange's invariant** in that field. On each screen, the angle, length and point of intersection of the ray vector with the screen may vary. However, the oriented area $S = \mathbf{p} \times \mathbf{q}$ will be the same on each screen, for rays propagating in an axisymmetric medium. This is the geometric meaning of Lagrange's invariant.*

[2] This is short notation for $p_\phi = \hat{\mathbf{z}} \cdot \mathbf{p} \times \mathbf{q}$. Scalar notation is standard for a vector normal to a plane that arises as a cross product of vectors in the plane. Of course, the notation for skewness S cannot be confused with the action S.

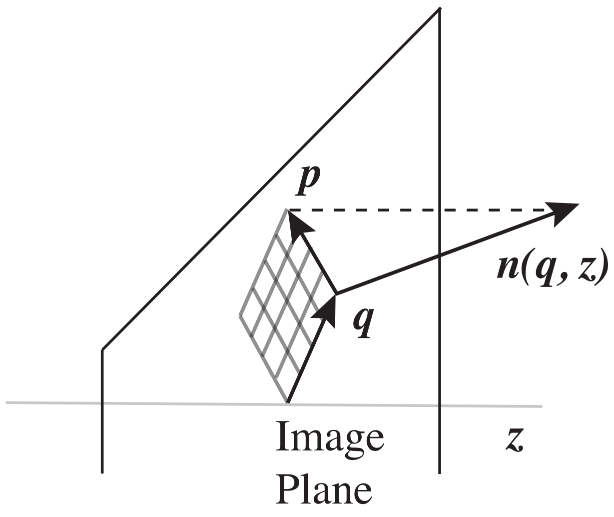


Figure 8: The skewness $S = \mathbf{p} \times \mathbf{q}$ of a ray $\mathbf{n}(\mathbf{q}, z)$ is an oriented area in an image plane. For axisymmetric media, skewness is preserved as a function of distance z along the optical axis. The projection $\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z)$ is also conserved, provided the medium is invariant under translations along the optical axis.

Conservation of the skewness function $p_\phi = \mathbf{p} \times \mathbf{q}$ follows in the reduced system (4.12) by computing

$$\frac{dp_\phi}{dz} = \{p_\phi,\, H\} = -\frac{\partial H}{\partial \phi} = 0\,,$$

which vanishes because the optical Hamiltonian for an axisymmetric medium is independent of the azimuthal angle ϕ about the optical axis.

Exercise. Check that Hamilton's canonical equations for ray optics (3.12) with the optical Hamiltonian (3.9) conserve the skewness function $p_\phi = \mathbf{p} \times \mathbf{q}$ when the refractive index satisfies (4.7). ★

Remark

4.14. *The values of the skewness function characterise the various types of rays [Wo2004].*

- *Vanishing of* $\mathbf{p} \times \mathbf{q}$ *occurs for* ***meridional rays****, for which* $\mathbf{p} \times \mathbf{q} = 0$ *implies that* $\mathbf{p}$ *and* $\mathbf{q}$ *are collinear in the image plane* ($\mathbf{p} \parallel \mathbf{q}$)*.*
- *On the other hand,* p_ϕ *takes its maximum value for* ***sagittal rays****, for which* $\mathbf{p} \cdot \mathbf{q} = 0$*, so that* $\mathbf{p}$ *and* $\mathbf{q}$ *are orthogonal in the image plane* ($\mathbf{p} \perp \mathbf{q}$)*.*
- *Rays that are neither collinear nor orthogonal are said to be* ***skew rays****.*

Remark

4.15 (Sign conventions in optics and mechanics)**.** *Unfortunately, the sign conventions differ between two fundamental ideas that are mathematically the same. Namely, in optics the skewness that characterises the rays is written in the two-dimensional image plane as* $S = \mathbf{p} \times \mathbf{q}$*, while in mechanics the angular momentum of a particle with momentum* $\mathbf{p}$ *at position* $\mathbf{q}$ *in three-dimensional is written as* $\mathbf{L} = \mathbf{q} \times \mathbf{p}$*. When working in either of these fields, it's probably best to adopt the customs of the natives. However, one must keep a sharp eye out for the difference in signs of rotations when moving between these fields.*

Exercise. ***(Phase-plane reduction)***

- Solve Hamilton's canonical equations for axisymmetric, translation-invariant media in the case of an optical fibre with a radially varying index of refraction in the following form,

$$n^2(r) = \lambda^2 + (\mu - \nu r^2)^2\,, \quad \lambda,\, \mu,\, \nu = constants,$$

  by reducing the problem to phase-plane analysis. How does the phase-space portrait differ between $p_\phi = 0$ and $p_\phi \neq 0$? What happens when ν changes sign?

- What regions of the phase plane admit real solutions? Is it possible for a phase point to pass from a region with real solutions to a region with complex solutions during its evolution? Prove it.

- Compute the dynamic and geometric phases for a periodic orbit of period Z in the (r, p_r) phase plane.

Hint: For $p_\phi \neq 0$ the problem reduces to a ***Duffing oscillator*** (Newtonian motion in a quartic potential) in a rotating frame, up to a rescaling of time by the value of the Hamiltonian on each ray "orbit".

See [HoKo1991] for a discussion of optical ray chaos under periodic perturbations of this solution.

★

4.6 Lagrange invariant: Poisson bracket relations

Under the canonical Poisson bracket (4.1), the skewness function, or Lagrange invariant,

$$S = p_\phi = \hat{\mathbf{z}} \cdot \mathbf{p} \times \mathbf{q} = y p_x - x p_y \,, \tag{4.19}$$

generates rotations of $\mathbf{q}$ and $\mathbf{p}$ jointly in the image plane. Both $\mathbf{q}$ and $\mathbf{p}$ are rotated by the same angle ϕ around the optical axis $\hat{\mathbf{z}}$. In other words, the equation $(d\mathbf{q}/d\phi, d\mathbf{p}/d\phi) = \{(\mathbf{q},\mathbf{p})\,,\,S\}$ defined by the Poisson bracket,

$$\frac{d}{d\phi} = X_S = \Big\{ \cdot \,,\, S \Big\} = \mathbf{q} \times \hat{\mathbf{z}} \cdot \frac{\partial}{\partial \mathbf{q}} + \mathbf{p} \times \hat{\mathbf{z}} \cdot \frac{\partial}{\partial \mathbf{p}} \,, \tag{4.20}$$

has the solution

$$\begin{pmatrix} \mathbf{q}(\phi) \\ \mathbf{p}(\phi) \end{pmatrix} = \begin{pmatrix} R_z(\phi) & 0 \\ 0 & R_z(\phi) \end{pmatrix} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} . \tag{4.21}$$

Here the matrix

$$R_z(\phi) = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \tag{4.22}$$

represents rotation of both $\mathbf{q}$ and $\mathbf{p}$ by an angle ϕ about the optical axis.

Remark

4.16. *The application of the Hamiltonian vector field X_S for skewness in (4.20) to the position vector $\mathbf{q}$ yields*

$$X_S \mathbf{q} = -\,\hat{\mathbf{z}} \times \mathbf{q} \,. \tag{4.23}$$

Likewise, the application of the Hamiltonian vector field X_S for skewness in (4.20) to the momentum vector yields

$$X_S\mathbf{p} = -\,\hat{\mathbf{z}} \times \mathbf{p}\,. \tag{4.24}$$

Thus, the application of X_S to the vectors $\mathbf{q} \in \mathbb{R}^3$ and $\mathbf{p} \in \mathbb{R}^3$ rotates them both about $\hat{\mathbf{z}}$ by the same angle.

Definition

4.17 (Diagonal action and cotangent lift)**.** *Together, formulas (4.23) and (4.24) comprise the **diagonal action** on $(\mathbf{q}, \mathbf{p})$ of axial rotations about $\hat{\mathbf{z}}$. The rotation of the momentum vector $\mathbf{p}$ that is induced by the rotation of the position vector $\mathbf{q}$ is called the **cotangent lift** of the action of the Hamiltonian vector field X_S. Namely, (4.24) is the lift of the action of rotation (4.23) from position vectors to momentum vectors.*

Remark

4.18 (Moment of momentum)**.** *Applying the Hamiltonian vector field X_S for skewness in (4.20) to screen coordinates $\mathbf{q} = \mathbb{R}^2$ produces the infinitesimal action of rotations about $\hat{\mathbf{z}}$, as*

$$X_S\mathbf{q} = \{\mathbf{q},\, S\} = -\,\hat{\mathbf{z}} \times \mathbf{q} = \frac{d\mathbf{q}}{d\phi}\bigg|_{\phi=0}\,.$$

thee skewness function S in (4.19) may be expressed in terms of two different pairings,

$$\begin{aligned} S &= \left\langle\!\!\left\langle \mathbf{p},\, X_S\mathbf{q} \right\rangle\!\!\right\rangle = \mathbf{p}\cdot(-\,\hat{\mathbf{z}} \times \mathbf{q}) \quad \textit{and} \\ S &= (\mathbf{p}\times\mathbf{q})\cdot\hat{\mathbf{z}} = \left\langle \mathbf{J}(\mathbf{p},\,\mathbf{q}),\,\hat{\mathbf{z}} \right\rangle = J^z(\mathbf{p},\,\mathbf{q})\,. \end{aligned} \tag{4.25}$$

Although these pairings are both written as dot products of vectors, strictly speaking they act on different spaces. Namely,

$$\begin{aligned} \left\langle\!\!\left\langle \,\cdot\,,\,\cdot\, \right\rangle\!\!\right\rangle &: \ (\textit{momentum}) \times (\textit{velocity}) \to \mathbb{R}\,, \\ \left\langle \,\cdot\,,\,\cdot\, \right\rangle &: \ (\textit{moment of momentum}) \times (\textit{rotation rate}) \to \mathbb{R}\,. \end{aligned} \tag{4.26}$$

The first pairing $\langle\!\langle\,\cdot\,,\,\cdot\,\rangle\!\rangle$ is between two vectors that are tangent to an optical screen. These vectors represent the projection of the ray vector on the screen $\mathbf{p}$ and the rate of change of the position $\mathbf{q}$ with azimuthal angle, $d\mathbf{q}/d\phi$, in (4.23). This is also the pairing $\langle\!\langle\,\cdot\,,\,\cdot\,\rangle\!\rangle$ between velocity and momentum that appears in the Legendre transformation. The second pairing $\langle\,\cdot\,,\,\cdot\,\rangle$ is between the oriented area

$\mathbf{p} \times \mathbf{q}$ *and the normal to the screen* $\hat{\mathbf{z}}$*. Thus, as we knew,* $J^z(\mathbf{p}, \mathbf{q}) = S(\mathbf{p}, \mathbf{q})$ *is the Hamiltonian for an infinitesimal rotation about the* $\hat{\mathbf{z}}$ *axis in* $\mathbb{R}^3$*.*

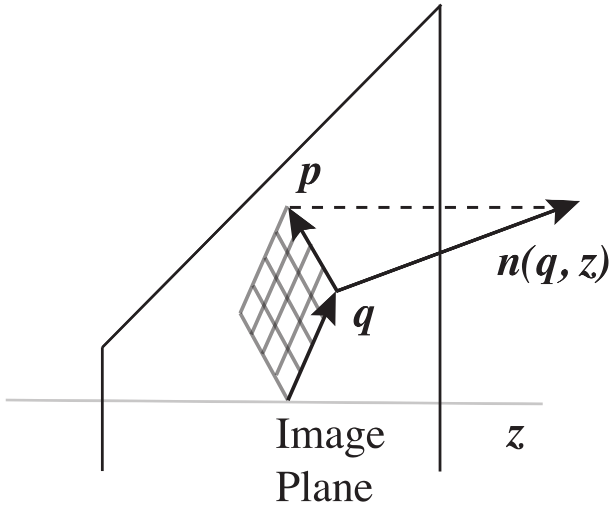


Figure 9: The skewness $S = \mathbf{p} \times \mathbf{q}$ of a ray $\mathbf{n}(\mathbf{q}, z)$ is an oriented area in an image plane. For axisymmetric media, skewness is preserved as a function of distance z along the optical axis. The projection $\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z)$ is also conserved, provided the medium is invariant under translations along the optical axis.

Definition

4.19. *Distinguishing between the pairings in (4.25) interprets the Lagrange invariant* $S = J^z(\mathbf{p},\,\mathbf{q}) = \mathbf{p}\times\mathbf{q}\cdot\hat{\mathbf{z}}$ *as the* $\hat{\mathbf{z}}$*-component of a map from phase space with coordinates* $(\mathbf{p},\,\mathbf{q})$ *to the oriented area* $\mathbf{J}(\mathbf{p},\,\mathbf{q}) = \mathbf{p}\times\mathbf{q}$, *or* ***moment of momentum***.

Definition

4.20 (Momentum map for cotangent lift)**.** *Formula (4.25) defines the momentum map for the cotangent lift of the action of rotations about* $\hat{\mathbf{z}}$ *from position vectors to their canonically conjugate momentum vectors in phase space. In general, a momentum map applies from phase space to the dual space of the Lie algebra of the Lie group whose action is involved. In this case, it is the map from phase space to the moment-of-momentum space,* $\mathcal{M}$,

$$\mathbf{J} : T^*\mathbb{R}^2 \to \mathcal{M}\,, \quad \textit{namely,} \quad \mathbf{J}(\mathbf{p},\,\mathbf{q}) = \mathbf{p}\times\mathbf{q}\,, \tag{4.27}$$

and $\mathbf{p}\times\mathbf{q}$ *is dual to the rotation rate about the axial direction* $\hat{\mathbf{z}}$ *under the pairing given by the three-dimensional scalar (dot) product. The corresponding Hamiltonian is the skewness*

$$S = J^z(\mathbf{p},\,\mathbf{q}) = \mathbf{J}\cdot\hat{\mathbf{z}} = \mathbf{p}\times\mathbf{q}\cdot\hat{\mathbf{z}}$$

in (4.25). This is the real-valued phase-space function whose Hamiltonian vector field X_S *rotates a point* $P = (\mathbf{q},\mathbf{p})$ *in phase space about the optical axis* $\hat{\mathbf{z}}$ *at its centre, according to*

$$-\hat{\mathbf{z}}\times P = X_{\mathbf{J}\cdot\hat{\mathbf{z}}}P = \{\,P\,,\,\mathbf{J}\cdot\hat{\mathbf{z}}\}\,. \tag{4.28}$$

Remark

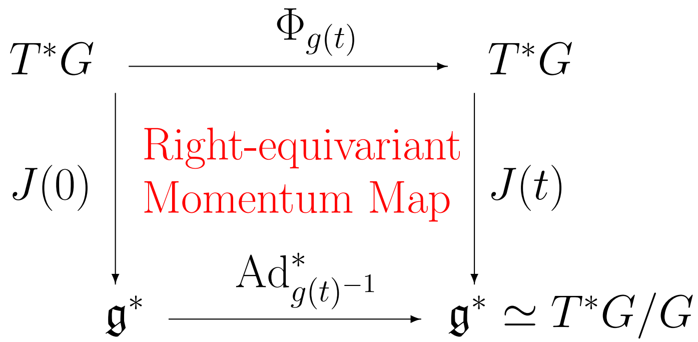
4.21. *The skewness function S and its square S^2 (called the* ***Petzval invariant*** *[Wo2004]) are conserved for ray optics in axisymmetric media. That is, the canonical Poisson bracket vanishes,*

$$\{S^2, H\} = 0\,, \tag{4.29}$$

for optical Hamiltonians of the form

$$H = -\Big[n(|\mathbf{q}|^2)^2 - |\mathbf{p}|^2\Big]^{1/2}. \tag{4.30}$$

The Poisson bracket (4.29) vanishes because $|\mathbf{q}|^2$ and $|\mathbf{p}|^2$ in H both remain invariant under the simultaneous rotations of $\mathbf{q}$ and $\mathbf{p}$ about $\hat{\mathbf{z}}$ generated by S in (4.20).



5 Axisymmetric invariant coordinates

Transforming to axisymmetric coordinates and the azimuthal angle in the optical phase space is similar to passing to polar coordinates (radius and angle) in the plane. Passing to polar coordinates by $(x, y) \to (r, \phi)$ decomposes the plane $\mathbb{R}^2$ into the product of the real line $r \in \mathbb{R}^+$ and the angle $\phi \in S^1$. Quotienting the plane by the angle leaves just the real line. The ***quotient map*** for the plane is

$$\pi : \mathbb{R}^2\{\mathbf{0}\} \to \mathbb{R}\backslash\{\mathbf{0}\} : \, (x, y) \to r \,. \tag{5.1}$$

The S^1 angle in optical phase space $T^*\mathbb{R}^2$ is the azimuthal angle. But how does one quotient the four-dimensional $T^*\mathbb{R}^2$ by the azimuthal angle?

As discussed in Section 4.3, azimuthal symmetry of the Hamiltonian summons the transformation to polar coordinates in phase space, as

$$(\mathbf{q}, \mathbf{p}) \to (r, p_r; p_\phi, \phi) \,.$$

This transformation reduces the motion to phase planes of radial (r, p_r) position and momentum, defined on level surfaces of the skewness p_ϕ. The trajectories evolve along intersections of the level sets of skewness (the planes $p_\phi = const$) with the level sets of the Hamiltonian $H(r, p_r, p_\phi) = const$. The motion along these intersections is independent of the ignorable phase variable $\phi \in S^1$, whose evolution thus decouples from that of the other variables. Consequently, the phase evolution may be reconstructed later by a quadrature, i.e., an integral that involves the parameters of the reduced phase space. Thus, in this case, azimuthal symmetry decomposes the phase space *exactly* into

$$T^*\mathbb{R}^2\backslash\{\mathbf{0}\} \simeq \big(T^*(\mathbb{R}\backslash\{0\}) \times \mathbb{R}\big) \times S^1 \,. \tag{5.2}$$

The corresponding quotient map for azimuthal symmetry is

$$\pi : T^*\mathbb{R}^2\backslash\{\mathbf{0}\} \to T^*(\mathbb{R}\backslash\{0\}) \times \mathbb{R} : (\mathbf{q},\mathbf{p}) \to (r, p_r; p_\phi)\,. \tag{5.3}$$

An alternative procedure exists for quotienting out the angular dependence of an azimuthally symmetric Hamiltonian system, which is independent of the details of the Hamiltonian function. This alternative procedure involves transforming to quadratic azimuthally invariant functions.

Definition

5.1 (Quotient map to quadratic S^1 invariants)**.** *The quadratic azimuthally invariant coordinates in $\mathbb{R}^3\backslash\{\mathbf{0}\}$ are defined by the quotient map* [3]

$$\pi : T^*\mathbb{R}^2\backslash\{0\} \to \mathbb{R}^3\backslash\{\mathbf{0}\} : (\mathbf{q},\mathbf{p}) \to \mathbf{X} = (X_1, X_2, X_3)\,, \tag{5.4}$$

given explicitly by the quadratic monomials

$$X_1 = |\mathbf{q}|^2 \geq 0\,, \quad X_2 = |\mathbf{p}|^2 \geq 0\,, \quad X_3 = \mathbf{p}\cdot\mathbf{q}\,. \tag{5.5}$$

The quotient map (5.4) can be written more succinctly as

$$\pi\,(\mathbf{p},\,\mathbf{q}) = \mathbf{X}\,. \tag{5.6}$$

[3] The transformation $T^*\mathbb{R}^2 \to \mathbb{R}^3$ in (5.5) will be recognised later as another example of a ***momentum map***.

Theorem

5.2. *The vector* (X_1, X_2, X_3) *of quadratic monomials in phase space all Poisson-commute with skewness* S*:*

$$\{S, X_1\} = 0\,, \quad \{S, X_2\} = 0\,, \quad \{S, X_3\} = 0\,. \tag{5.7}$$

Proof. These three Poisson brackets with skewness S all vanish because dot products of vectors are preserved by the joint rotations of $\mathbf{q}$ and $\mathbf{p}$ that are generated by S. □

Remark

5.3. *The orbits of* S *in (4.21) are rotations of both* $\mathbf{q}$ *and* $\mathbf{p}$ *by an angle* ϕ *about the optical axis at a fixed position* z*. According to the relation* $\{S,\, \mathbf{X}\} = 0$*, the quotient map* $\mathbf{X} = \pi\,(\mathbf{p},\, \mathbf{q})$ *in (5.4) collapses each circular orbit of* S *on a given image screen in phase space* $T^*\mathbb{R}^2\backslash\{0\}$ *to a point in* $\mathbb{R}^3\backslash\{\mathbf{0}\}$*. The converse also holds. Namely, the inverse of the quotient map* $\pi^{-1}\mathbf{X}$ *for* $\mathbf{X} \in \mathrm{Image}\,\pi$ *consists of the circle* (S^1) *generated by the rotation of phase space about its centre by the flow of* S*.*

Definition

5.4 (Orbit manifold)**.** *The image in* $\mathbb{R}^3$ *of the quotient map* $\pi : T^*\mathbb{R}^2\backslash\{0\} \to \mathbb{R}^3\backslash\{\mathbf{0}\}$ *in (5.4) is the* ***orbit manifold*** *for axisymmetric ray optics.*

Remark

5.5 (Orbit manifold for axisymmetric ray optics). *The image of the quotient map π in (5.4) may be conveniently displayed as the zero level set of the relation*

$$C(X_1,\, X_2,\, X_3,\, S) = S^2 - (X_1X_2 - X_3^2) = 0\,, \tag{5.8}$$

among the axisymmetric variables in Equation (5.5). Consequently, a level set of S in the quotient map $T^*\mathbb{R}^2\backslash\{0\} \to \mathbb{R}^3\backslash\{\mathbf{0}\}$ obtained by transforming to S^1 phase-space invariants yields an orbit manifold defined by $C(X_1,\, X_2,\, X_3,\, S) = 0$ in $\mathbb{R}^3\backslash\{\mathbf{0}\}$.

For axisymmetric ray optics, the image of the quotient map π in $\mathbb{R}^3$ turns out to be a family of hyperboloids of revolution.

6 Geometry of invariant coordinates

In terms of the axially invariant coordinates (5.5), the Petzval invariant and the square of the optical Hamiltonian satisfy

$$|\mathbf{p}\times\mathbf{q}|^2 = |\mathbf{p}|^2|\mathbf{q}|^2 - (\mathbf{p}\cdot\mathbf{q})^2 \quad \text{and} \quad H^2 = n^2(|\mathbf{q}|^2) - |\mathbf{p}|^2 \ge 0\,. \tag{6.1}$$

That is,

$$S^2 = X_1X_2 - X_3^2 \ge 0 \quad \text{and} \quad H^2 = n^2(X_1) - X_2 \ge 0\,. \tag{6.2}$$

The geometry of the solution is determined by the intersections of the level sets of the conserved quantities S^2 and H^2. The level sets of $S^2 \in \mathbb{R}^3$ are hyperboloids of revolution around the $X_1 = X_2$ axis in the horizontal plane defined by $X_3 = 0$ (Figure 1.8). The level-set hyperboloids lie in the interior of the $S = 0$ cone with $X_1 > 0$ and $X_2 > 0$. The level sets of H^2 depend on the functional form of the index of refraction, but they are X_3-independent. The ray path in the S^1-invariant variables $\mathbf{X} = (X_1, X_2, X_3) \in \mathbb{R}^3$ must occur along intersections of S^2 and H^2, since both of these quantities are conserved along the ray path in axisymmetric, translation-invariant media.

One would naturally ask how the quadratic phase-space quantities (X_1, X_2, X_3) Poisson-commute among themselves. However, before addressing that question, let us ask the following.

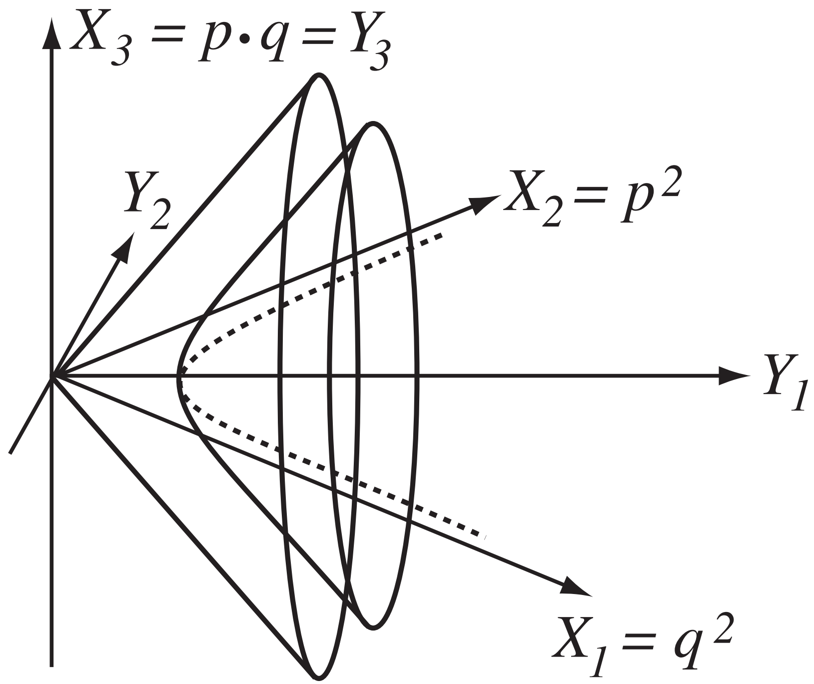


Figure 10: Level sets of the Petzval invariant $S^2 = X_1X_2 - X_3^2$ are hyperboloids of revolution around the $X_1 = X_2$ axis (along Y_1) in the horizontal plane, $X_3 = 0$. Level sets of the Hamiltonian H in (6.2) are independent of the vertical coordinate. The axisymmetric invariants $\mathbf{X} \in \mathbb{R}^3$ evolve along the intersections of these level sets by $\dot{\mathbf{X}} = \nabla S^2 \times \nabla H$, as the vertical Hamiltonian knife $H = constant$ slices through the hyperbolic onion of level sets of S^2. In the coordinates $Y_1 = (X_1 + X_2)/2\,, Y_2 = (X_2 - X_1)/2\,, Y_3 = X_3$, one has $S^2 = Y_1^2 - Y_2^2 - Y_3^2$. Being invariant under the flow of the Hamiltonian vector field $X_S = \{\,\cdot\,,\,S\}$, each point on any layer H^2 of the hyperbolic onion H^3 consists of an S^1 orbit in phase space under the diagonal rotation (4.21). This orbit is a circular rotation of both $\mathbf{q}$ and $\mathbf{p}$ on an image screen at position z by an angle ϕ about the optical axis.

Exercise. How does the Poisson bracket with each of the axisymmetric quantities (X_1, X_2, X_3) act as a derivative operation on functions of the phase-space variables $\mathbf{q}$ and $\mathbf{p}$? ★

Remark

6.1. *Answering this question introduces the concept of* ***flows of Hamiltonian vector fields****.*

6.1 Flows of Hamiltonian vector fields

Theorem

6.2 (Flows of Hamiltonian vector fields)**.** *Poisson brackets with the S^1-invariant phase-space functions X_1, X_2 and X_3 generate linear homogeneous transformations of $(\mathbf{q},\, \mathbf{p}) \in T^*\mathbb{R}^2$, obtained by regarding the Hamiltonian vector fields obtained as in Definition 4.6 from the Poisson brackets as derivatives,*

$$\frac{d}{d\tau_1} := \{\,\cdot\,,\, X_1\}\,, \quad \frac{d}{d\tau_2} := \{\,\cdot\,,\, X_2\} \quad \textit{and} \quad \frac{d}{d\tau_3} := \{\,\cdot\,,\, X_3\}\,, \tag{6.3}$$

in their ***flow parameters*** *τ_1, τ_2 and τ_3, respectively.*

The flows themselves may be determined by integrating the characteristic equations of these Hamiltonian vector fields.

Proof. • The sum $\frac{1}{2}(X_1 + X_2)$ is the harmonic-oscillator Hamiltonian. This Hamiltonian generates rotation of the $(\mathbf{q},\, \mathbf{p})$ phase space around its centre, by integrating the characteristic equations of its Hamiltonian vector field,

$$\frac{d}{d\omega} = \left\{\,\cdot\,,\, \frac{1}{2}(X_1 + X_2)\right\} = \mathbf{p}\cdot\frac{\partial}{\partial\mathbf{q}} - \mathbf{q}\cdot\frac{\partial}{\partial\mathbf{p}}\,. \tag{6.4}$$

To see this, write the simultaneous equations

$$\frac{d}{d\omega}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \left\{\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix}\,,\, \frac{1}{2}(X_1 + X_2)\right\}\,,$$

or in matrix form,[4]

$$\frac{d}{d\omega}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} =: \frac{1}{2}(m_1+m_2)\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix},$$

for the 2×2 traceless matrices m_1 and m_2 defined by

$$m_1 = \begin{pmatrix}0 & 0\\ -2 & 0\end{pmatrix} \quad\text{and}\quad m_2 = \begin{pmatrix}0 & 2\\ 0 & 0\end{pmatrix}.$$

These ω-dynamics may be rewritten as a complex equation,

$$\frac{d}{d\omega}(\mathbf{q}+i\mathbf{p}) = -i(\mathbf{q}+i\mathbf{p}), \tag{6.5}$$

whose immediate solution is

$$\mathbf{q}(\omega)+i\mathbf{p}(\omega) = e^{-i\omega}\Big(\mathbf{q}(0)+i\mathbf{p}(0)\Big).$$

This solution may also be written in matrix form as

$$\begin{pmatrix}\mathbf{q}(\omega)\\ \mathbf{p}(\omega)\end{pmatrix} = \begin{pmatrix}\cos\omega & \sin\omega\\ -\sin\omega & \cos\omega\end{pmatrix}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix}, \tag{6.6}$$

which is a diagonal clockwise rotation of $(\mathbf{q},\mathbf{p})$. This solution sums the following exponential series:

$$e^{\omega(m_1+m_2)/2} = \sum_{n=0}^{\infty}\frac{\omega^n}{n!}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}^n = \begin{pmatrix}\cos\omega & \sin\omega\\ -\sin\omega & \cos\omega\end{pmatrix}. \tag{6.7}$$

This may also be verified by summing its even and odd powers separately.

[4]For rotational symmetry, it is sufficient to restrict attention to rays lying in a fixed azimuthal plane and thus we may write these actions using 2×2 matrices, rather than 4×4 matrices.

Likewise, a nearly identical calculation yields

$$\frac{d}{d\gamma}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \frac{1}{2}(m_2 - m_1)\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix}$$

for the dynamics of the Hamiltonian $H = (|\mathbf{p}|^2 - |\mathbf{q}|^2)/2$. This time, the solution is the hyperbolic rotation

$$\begin{pmatrix}\mathbf{q}(\gamma)\\ \mathbf{p}(\gamma)\end{pmatrix} = \begin{pmatrix}\cosh\gamma & \sinh\gamma\\ \sinh\gamma & \cosh\gamma\end{pmatrix}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix} \tag{6.8}$$

which, in turn, sums the exponential series

$$\begin{aligned} e^{\gamma(m_2-m_1)/2} &= \sum_{n=0}^{\infty}\frac{\gamma^n}{n!}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}^n \\ &= \begin{pmatrix}\cosh\gamma & \sinh\gamma\\ \sinh\gamma & \cosh\gamma\end{pmatrix}. \end{aligned} \tag{6.9}$$

- In ray optics, the canonical Poisson bracket with the quadratic phase-space function $X_1 = |\mathbf{q}|^2$ defines the action of the following *linear* Hamiltonian vector field:

$$\frac{d}{d\tau_1} = \Big\{\,\cdot\,,\, X_1\Big\} = -2\mathbf{q}\cdot\frac{\partial}{\partial\mathbf{p}}\,. \tag{6.10}$$

This action may be written equivalently in matrix form as

$$\frac{d}{d\tau_1}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix}0 & 0\\ -2 & 0\end{pmatrix}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = m_1\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix}.$$

Integration of this system of equations yields the finite transformation

$$\begin{aligned} \begin{pmatrix} \mathbf{q}(\tau_1) \\ \mathbf{p}(\tau_1) \end{pmatrix} &= e^{\tau_1 m_1} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ -2\tau_1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &=: M_1(\tau_1) \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} . \end{aligned} \tag{6.11}$$

This is an easy result, because the matrix m_1 is nilpotent. That is, $m_1^2 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, so the formal series representing the exponential of the matrix

$$e^{\tau_1 m_1} = \sum_{n=0}^{\infty} \frac{1}{n!} (\tau_1 m_1)^n \tag{6.12}$$

truncates at its second term. This solution may be interpreted as the ***action of a thin lens*** [Wo2004].

- Likewise, the canonical Poisson bracket with $X_2 = |\mathbf{p}|^2$ defines the linear Hamiltonian vector field,

$$\frac{d}{d\tau_2} = \Big\{ \cdot\,,\, X_2 \Big\} = 2\mathbf{p} \cdot \frac{\partial}{\partial \mathbf{q}} \,. \tag{6.13}$$

In matrix form, this is

$$\frac{d}{d\tau_2} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} = m_2 \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} ,$$

in which the matrix m_2 is also nilpotent. Its integration generates the finite transformation

$$\begin{aligned} \begin{pmatrix} \mathbf{q}(\tau_2) \\ \mathbf{p}(\tau_2) \end{pmatrix} &= e^{\tau_2 m_2} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2\tau_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &=: M_2(\tau_2) \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} , \end{aligned} \tag{6.14}$$

corresponding to ***free propagation*** of light rays in a homogeneous medium.

- The transformation generated by $X_3 = \mathbf{q}\cdot\mathbf{p}$ compresses phase space along one coordinate and expands it along the other, while preserving skewness. Its Hamiltonian vector field is

$$\frac{d}{d\tau_3} = \Big\{ \cdot\,,\, X_3 \Big\} = \mathbf{q}\cdot\frac{\partial}{\partial \mathbf{q}} - \mathbf{p}\cdot\frac{\partial}{\partial \mathbf{p}}\,.$$

Being linear, this may be written in matrix form as

$$\frac{d}{d\tau_3}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix} =: m_3\begin{pmatrix}\mathbf{q}\\ \mathbf{p}\end{pmatrix}.$$

The integration of this linear system generates the flow, or finite transformation,

$$\begin{aligned}\begin{pmatrix}\mathbf{q}(\tau_3)\\ \mathbf{p}(\tau_3)\end{pmatrix} &= e^{\tau_3 m_3}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix}\\ &= \begin{pmatrix}e^{\tau_3} & 0\\ 0 & e^{-\tau_3}\end{pmatrix}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix}\\ &=: M_3(\tau_3)\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix}, \end{aligned} \tag{6.15}$$

whose exponential series is easily summed, because m_3 is diagonal and constant. Thus, the quadratic quantity X_3 generates a transformation that turns one harmonic-oscillator Hamiltonian into another one corresponding to a different natural frequency. This transformation is called ***squeezing*** of light.

The proof of Theorem 6.2 is now finished. □

++

7 Lecture 4: Symplectic matrices

Remark

7.1 (Symplectic matrices)**.** *Poisson brackets with the quadratic monomials on phase space* X_1, X_2, X_3 *correspond respectively to multiplication by the traceless constant matrices* m_1, m_2, m_3*. For example,*

$$\frac{d}{d\tau_3}\begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} =: m_3 \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} .$$

In turn, exponentiation of these traceless constant matrices leads to the corresponding matrices $M_1(\tau_1), M_2(\tau_2), M_3(\tau_3)$ *in Equations (6.11), (6.14) and (6.15). The latter are* 2×2 ***symplectic matrices****. That is, these three matrices each satisfy*

$$M_i(\tau_i) J M_i(\tau_i)^T = J \quad (\textit{no sum on } i = 1, 2, 3), \tag{7.1}$$

where

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} . \tag{7.2}$$

By their construction from the axisymmetric invariants X_1, X_2, X_3*, each of the symplectic matrices* $M_1(\tau_1), M_2(\tau_2), M_3(\tau_3)$ *preserves the cross product* $S = \mathbf{p} \times \mathbf{q}$*.*

Definition

7.2 (Lie transformation groups)**.**

- *A* ***transformation*** *is a one-to-one mapping of a set onto itself.*
- *A collection of transformations is called a* ***group****, provided*
  - *it includes the identity transformation and the inverse of each transformation;*
  - *it contains the result of the consecutive application of any two transformations; and*
  - *composition of that result with a third transformation is associative.*
- *A group is a* ***Lie group****, provided its transformations depend smoothly on a set of parameters.*

Theorem

7.3 (Symplectic group $Sp(2,\mathbb{R})$)**.** *Under matrix multiplication, the set of 2×2 symplectic matrices forms a group.*

Exercise. Prove that the matrices $M_1(\tau_1), M_2(\tau_2), M_3(\tau_3)$ defined above all satisfy the defining relation (7.1) required to be symplectic. Prove that these matrices form a group under matrix multiplication. Conclude that they form a three-parameter Lie group. ★

Theorem

7.4 (Fundamental theorem of planar optics)**.** *Any planar paraxial optical system, represented by a 2×2 symplectic matrix $M \in Sp(2, \mathbb{R})$, may be factored into subsystems consisting of products of three subgroups of the symplectic group, as*

$$M = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} e^{\tau_3} & 0 \\ 0 & e^{-\tau_3} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2\tau_1 & 1 \end{pmatrix} . \tag{7.3}$$

This is a general result, called the ***Iwasawa decomposition*** *of the symplectic matrix group, usually written as [Ge1961]*

$$Sp(2, \mathbb{R}) = \mathsf{KAN} \, . \tag{7.4}$$

The rightmost matrix factor (nilpotent subgroup N*) corresponds to a* ***thin lens****, whose parameter $2\tau_1$ is called its* ***Gaussian power*** *[Wo2004]. This factor, generated by the Hamiltonian $H_N = -|\mathbf{q}|^2$ leaves $\mathbf{q}$ invariant, but changes $\mathbf{p}$ and thus alters the direction of the rays that fall on each point of the screen. The middle factor (abelian subgroup* A*) magnifies the image by the factor e^{τ_3}, while squeezing the light so that its Hamiltonian $H_A = \mathbf{q} \cdot \mathbf{p}$ remains invariant. The leftmost factor (the maximal compact subgroup* K*) is a rotation by angle $\omega \in S^1$ generated by $H_K = \frac{1}{2}|\mathbf{q}|^2 + \frac{1}{2}|\mathbf{p}|^2$.*

For insightful discussions and references to the literature on the design and analysis of optical systems using the symplectic matrix approach, see, e.g., [Wo2004]. For many extensions of these ideas with applications to charged-particle beams, see [Dr2007].

Definition

7.5 (Hamiltonian matrices)**.** *The traceless constant matrices*

$$m_1 = \begin{pmatrix} 0 & 0 \\ -2 & 0 \end{pmatrix}, \quad m_2 = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \quad m_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{7.5}$$

whose exponentiation defines the $Sp(2,\mathbb{R})$ *symplectic matrices*

$$e^{\tau_1 m_1} = M_1(\tau_1)\,, \quad e^{\tau_2 m_2} = M_2(\tau_2)\,, \quad e^{\tau_3 m_3} = M_3(\tau_3)\,, \tag{7.6}$$

and which are the tangent vectors at their respective identity transformations,

$$\begin{aligned} m_1 &= \Big[M\,'_1(\tau_1) M_1^{-1}(\tau_1) \Big]_{\tau_1=0}\,, \\ m_2 &= \Big[M\,'_2(\tau_2) M_2^{-1}(\tau_2) \Big]_{\tau_2=0}\,, \\ m_3 &= \Big[M\,'_3(\tau_3) M_3^{-1}(\tau_3) \Big]_{\tau_3=0}\,, \end{aligned} \tag{7.7}$$

are called ***Hamiltonian matrices****.*

Remark

7.6 (Hamiltonian matrices and their symplectic orbits)**.**

- *From the definition, the Hamiltonian matrices* m_i *with* $i = 1, 2, 3$ *each satisfy*

$$Jm_i + m_i^T J = 0\,, \quad \textit{where} \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{7.8}$$

That is, $Jm_i = (Jm_i)^T$ *is a symmetric matrix.*

> ***Exercise.*** *Take the derivative of the definition of symplectic matrices (7.1) to prove statement (7.8) about Hamiltonian matrices. Verify that the Hamiltonian matrices in (7.5) satisfy (7.8).* ★

- *The respective actions of the symplectic matrices* $M_1(\tau_1)$, $M_2(\tau_2)$, $M_3(\tau_3)$ *in (7.6) on the phase-space vector* $(\mathbf{q},\, \mathbf{p})^T$ *are the flows of the Hamiltonian vector fields* $\{\,\cdot\,,\, X_1\}$, $\{\,\cdot\,,\, X_2\}$, *and* $\{\,\cdot\,,\, X_3\}$ *corresponding to the axisymmetric invariants* X_1, X_2 *and* X_3 *in (5.5).*

- *The quadratic Hamiltonian,*

$$
\begin{aligned}
H &= \frac{\omega}{2}(X_1 + X_2) + \frac{\gamma}{2}(X_2 - X_1) + \tau X_3 \\
&= \frac{\omega}{2}(|\mathbf{p}|^2 + |\mathbf{q}|^2) + \frac{\gamma}{2}(|\mathbf{p}|^2 - |\mathbf{q}|^2) + \tau \mathbf{q}\cdot\mathbf{p}\,,
\end{aligned}
\tag{7.9}
$$

is associated with the ***Hamiltonian matrix****,*

$$
\begin{aligned}
m(\omega,\gamma,\tau) &= \frac{\omega}{2}(m_1 + m_2) + \frac{\gamma}{2}(m_2 - m_1) + \tau m_3 \\
&= \begin{pmatrix} \tau & \gamma+\omega \\ \gamma-\omega & -\tau \end{pmatrix}.
\end{aligned}
\tag{7.10}
$$

The eigenvalues of the Hamiltonian matrix (7.10) are determined from

$$
\lambda^2 + \Delta = 0\,, \quad \textit{with} \quad \Delta = \det m = \omega^2 - \gamma^2 - \tau^2\,.
\tag{7.11}
$$

Consequently, the eigenvalues come in pairs, given by

$$
\lambda^{\pm} = \pm\sqrt{-\Delta} = \pm\sqrt{\tau^2 + \gamma^2 - \omega^2}\,.
\tag{7.12}
$$

Orbits of Hamiltonian flows *in the space* $(\gamma+\omega, \gamma-\omega, \tau) \in \mathbb{R}^3$ *are obtained from the action of a symplectic matrix* $M(\tau_i)$ *on a Hamiltonian matrix* $m(\omega, \gamma, \tau)$ *by matrix conjugation* *(This is secretly Ad action, but we will not have time to discuss it more.)*

$$m \to m' = M(\tau_i) m M^{-1}(\tau_i) \quad \textit{(no sum on } i = 1, 2, 3\textit{)}$$

may alter the values of (ω, γ, τ) *in (7.10). However, this action preserves eigenvalues, so it preserves the value of the determinant* Δ*. This means the orbits of the Hamiltonian flows lie on the level sets of the determinant* Δ*.*

The Hamiltonian flows corresponding to these eigenvalues change type, depending on whether $\Delta < 0$ *(hyperbolic),* $\Delta = 0$ *(parabolic), or* $\Delta > 0$ *(elliptic), as illustrated in Figure 11 and summarised in the table below, cf. [Wo2004].*

Harmonic (elliptic) orbit	*Trajectories: Ellipses*
$\Delta = 1\,, \quad \lambda^{\pm} = \pm\, i$	$m_H = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$
Free (parabolic) orbit	*Trajectories: Straight lines*
$\Delta = 0\,, \quad \lambda^{\pm} = 0$	$m_H = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$
Repulsive (hyperbolic) orbit	*Trajectories: Hyperbolas*
$\Delta = -1\,, \quad \lambda^{\pm} = \pm\, 1$	$m_H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$

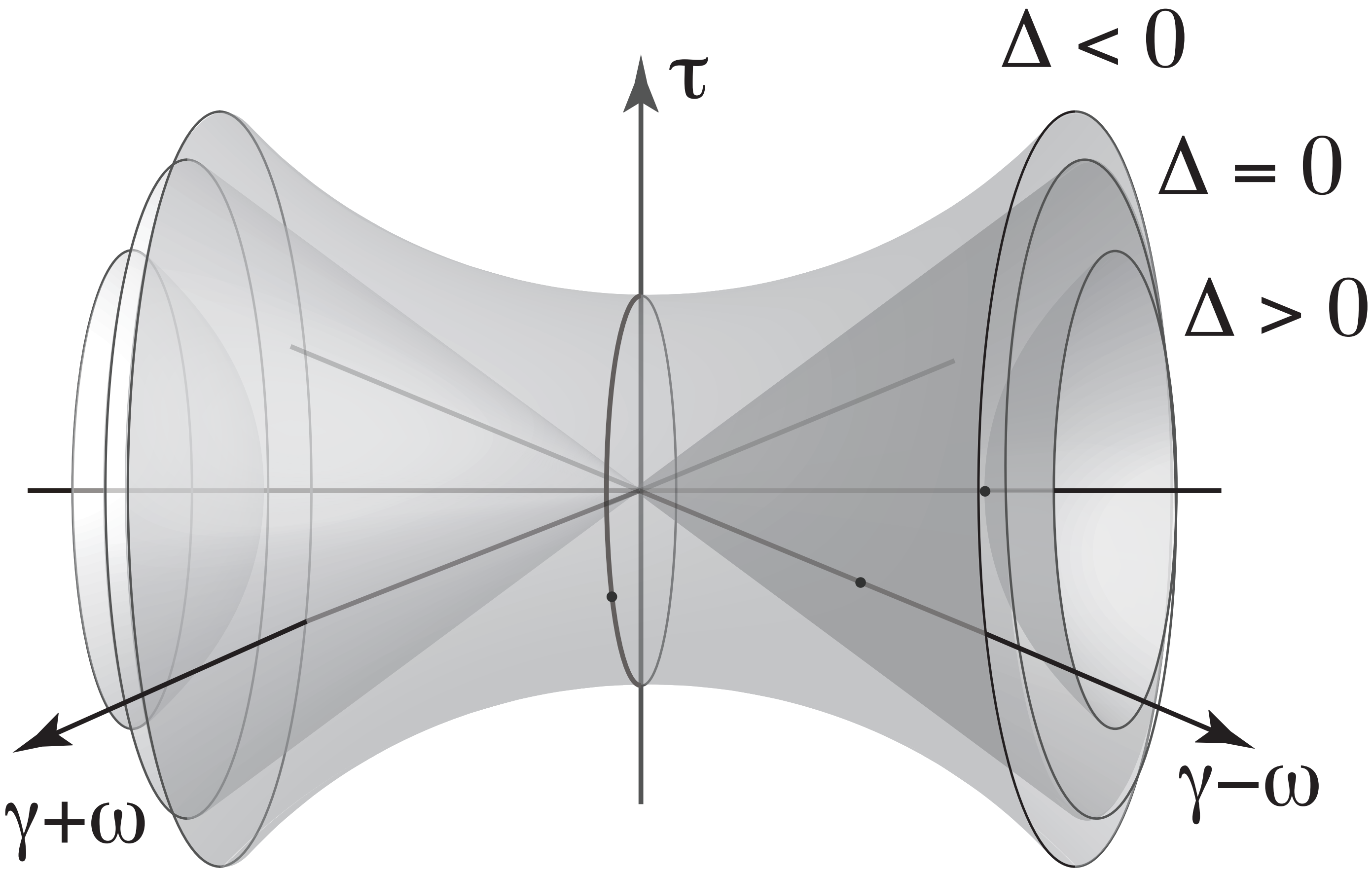


Figure 11: The Ad action by matrix conjugation of a Hamiltonian matrix by a symplectic matrix changes its parameters $(\omega, \gamma, \tau) \in \mathbb{R}^3$, while preserving the value of the discriminant $\Delta = \omega^2 - \gamma^2 - \tau^2$. The resulting flows corresponding to exponentiation of the Hamiltonian matrices with parameters $(\omega, \gamma, \tau) \in \mathbb{R}^3$ are divided into three families of orbits defined by the sign of Δ. These three families of orbits are hyperbolic $(\Delta < 0)$, parabolic $(\Delta = 0)$ and elliptic $(\Delta > 0)$.

Remark

7.7 (Prelude to Lie algebras).

- *In terms of the Hamiltonian matrices the KAN decomposition (7.3) may be written as*

$$M = e^{\omega(m_1+m_2)/2} e^{\tau_3 m_3} e^{\tau_1 m_1} \,. \tag{7.13}$$

- *Under the* ***matrix commutator*** $[m_i, m_j] := m_i m_j - m_j m_i$, *the Hamiltonian matrices* m_i *with* $i = 1, 2, 3$ *close among themselves, as*

$$[m_1, m_2] = 4m_3 \,, \quad [m_2, m_3] = -\,2m_2 \,, \quad [m_3, m_1] = -\,2m_1 \,.$$

The last observation (closure of the commutators) summons the definition of a Lie algebra. For this, we follow [Olv00].

8 Lie algebras

8.1 Definition

Definition

8.1. *A **Lie algebra** is a vector space $\mathfrak{g}$ together with a bilinear operation*

$$[\,\cdot\,,\,\cdot\,] : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}\,,$$

*called the **Lie bracket** for $\mathfrak{g}$, that satisfies the defining properties:*

- *bilinearity, e.g.,*

  $$[a\mathbf{u} + b\mathbf{v}\,,\,\mathbf{w}] = a[\mathbf{u}\,,\,\mathbf{w}] + b[\mathbf{v}\,,\,\mathbf{w}]\,,$$

  for constants $(a, b) \in \mathbb{R}$ and any vectors $(\mathbf{u}, \mathbf{v}, \mathbf{w}) \in \mathfrak{g}$;

- *skew-symmetry,*

  $$[\mathbf{u}\,,\,\mathbf{w}] = -[\mathbf{w}\,,\,\mathbf{u}]\,;$$

- *Jacobi identity,*

  $$[\mathbf{u}\,,\,[\mathbf{v}\,,\,\mathbf{w}]\,] + [\mathbf{v}\,,\,[\mathbf{w}\,,\,\mathbf{u}]\,] + [\mathbf{w}\,,\,[\mathbf{u}\,,\,\mathbf{v}]\,] = 0\,,$$

  for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in $\mathfrak{g}$.

8.2 Structure constants

Suppose $\mathfrak{g}$ is any finite-dimensional Lie algebra. The Lie bracket for any choice of basis vectors $\{\mathbf{e}_1, \ldots, \mathbf{e}_r\}$ of $\mathfrak{g}$ must again lie in $\mathfrak{g}$. Thus, constants c_{ij}^k exist, where $i, j, k = 1, 2, \ldots, r$, called the ***structure constants*** of the Lie algebra $\mathfrak{g}$, such that

$$[\,\mathbf{e}_i\,,\,\mathbf{e}_j\,] = c_{ij}^k \mathbf{e}_k\,. \tag{8.1}$$

Since $\{\mathbf{e}_1, \ldots, \mathbf{e}_r\}$ form a vector basis, the structure constants in (8.1) determine the Lie algebra $\mathfrak{g}$ from the bilinearity of the Lie bracket. The conditions of skew-symmetry and the Jacobi identity place further constraints on the structure constants. These constraints are

- skew-symmetry

$$c_{ji}^k = -\,c_{ij}^k\,, \tag{8.2}$$

- Jacobi identity

$$c_{ij}^k c_{lk}^m + c_{li}^k c_{jk}^m + c_{jl}^k c_{ik}^m = 0\,. \tag{8.3}$$

Conversely, any set of constants c_{ij}^k that satisfy relations (8.2) and (8.3) defines a Lie algebra $\mathfrak{g}$.

Exercise. Prove that the Jacobi identity requires the relation (8.3). ★

Answer. The Jacobi identity involves cyclically summing three terms of the form

$$[\,\mathbf{e}_l\,,\,[\,\mathbf{e}_i\,,\,\mathbf{e}_j\,]\,] = c_{ij}^k[\,\mathbf{e}_l\,,\,\mathbf{e}_k] = c_{ij}^k c_{lk}^m \mathbf{e}_m\,.$$

▲

8.3 Commutator tables

A convenient way to display the structure of a finite-dimensional Lie algebra is to write its commutation relations in tabular form. If $\mathfrak{g}$ is an r-dimensional Lie algebra and $\{\mathbf{e}_1, \ldots, \mathbf{e}_r\}$ forms a basis of $\mathfrak{g}$, then its ***commutator table*** will be the $r \times r$ array whose (i, j)th entry expresses the Lie bracket $[\, \mathbf{e}_i \,,\, \mathbf{e}_j \,]$. Commutator tables are always antisymmetric since $[\, \mathbf{e}_j \,,\, \mathbf{e}_i \,] = -[\, \mathbf{e}_i \,,\, \mathbf{e}_j \,]$. Hence, the diagonal entries all vanish. The structure constants may be easily read off the commutator table, since c_{ij}^k is the coefficient of $\mathbf{e}_k$ in the (i, j)th entry of the table.

For example, the commutator table of the Hamiltonian matrices in Equation (7.7) is given by

$$
[m_i,\, m_j] = c_{ij}^k m_k = \begin{array}{|c|ccc|}
\hline
[\,\cdot\,,\,\cdot\,] & m_1 & m_2 & m_3 \\
\hline
m_1 & 0 & 4m_3 & 2m_1 \\
m_2 & -4m_3 & 0 & -2m_2 \\
m_3 & -2m_1 & 2m_2 & 0 \\
\hline
\end{array} \,. \tag{8.4}
$$

The structure constants are immediately read off the table as

$$
c_{12}^3 = 4 = -c_{21}^3 \,, \quad c_{13}^1 = c_{32}^2 = 2 = -c_{23}^2 = -c_{31}^1 \,,
$$

and all the other c_{ij}^k's vanish.

These commutation relations for the 2×2 Hamiltonian matrices define the structure constants for the ***symplectic Lie algebra*** $sp(2, \mathbb{R})$.

8.4 Poisson brackets among axisymmetric variables

Theorem

8.2. *The canonical Poisson brackets among the axisymmetric variables X_1, X_2 and X_3 in (5.5) close among themselves:*

$$\{X_1, X_2\} = 4X_3\,, \quad \{X_2, X_3\} = -\,2X_2\,, \quad \{X_3, X_1\} = -\,2X_1\,.$$

In tabular form, this is

$$\{X_i,\, X_j\} = \begin{array}{|c|ccc|} \hline \{\,\cdot\,,\,\cdot\,\} & X_1 & X_2 & X_3 \\ \hline X_1 & 0 & 4X_3 & 2X_1 \\ X_2 & -4X_3 & 0 & -2X_2 \\ X_3 & -2X_1 & 2X_2 & 0 \\ \hline \end{array}\,. \tag{8.5}$$

Proof. The proof is a direct verification using the chain rule for Poisson brackets,

$$\{X_i,\, X_j\} = \frac{\partial X_i}{\partial z_A}\{z_A,\, z_B\}\frac{\partial X_j}{\partial z_A}\,, \tag{8.6}$$

for the invariant quadratic monomials $X_i(z_A)$ in (5.5). Here one denotes $z_A = (q_A, p_A)$, with $A = 1, 2, 3$.

The table in (8.5) implies the suggestive form for the Poisson bracket of functions, with a pairing $\langle\,\cdot\,,\,\cdot\,\rangle$,

$$\{F, H\} = -\,X_k\, c^k_{ij}\frac{\partial F}{\partial X_i}\frac{\partial H}{\partial X_j} =: -\left\langle \mathbf{X}\,,\, \left[\frac{\partial F}{\partial \mathbf{X}}\,,\, \frac{\partial H}{\partial \mathbf{X}}\right]\right\rangle\,, \tag{8.7}$$

that uses the same structure constants as in (8.4). □

Definition

8.3 (Lie Poisson bracket)**.** *A **Lie–Poisson bracket** is a bracket operation defined as a linear functional of a Lie algebra bracket by a real-valued pairing between a Lie algebra and its dual space.*

Remark

8.4. *By direct computation, we have found such a bracket for the quadratic axisymmetric phase space invariants for ray optics,*

$$\{F, H\} = -\, X_k\, c^k_{ij} \frac{\partial F}{\partial X_i} \frac{\partial H}{\partial X_j} =: -\left\langle \mathbf{X}\,,\, \left[\frac{\partial F}{\partial \mathbf{X}}\,,\, \frac{\partial H}{\partial \mathbf{X}}\right]\right\rangle, \tag{8.8}$$

Remark

8.5. *Equation (8.8) defines a Lie–Poisson bracket. Being a linear functional of an operation (the Lie bracket* $[\,\cdot\,,\,\cdot\,]$*) which satisfies the Jacobi identity, any Lie–Poisson bracket also satisfies the Jacobi identity.*

Exercise. Why did this work? Why did passage to the quadratic axisymmetric phase space invariants for ray optics produce a Lie-Poisson bracket? ★

9 Back to momentum maps

9.1 The action of $Sp(2,\mathbb{R})$ on $T^*\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2$

The Lie group $Sp(2,\mathbb{R})$ of symplectic real matrices $M(s)$ acts diagonally on $\mathbf{z} = (\mathbf{q},\, \mathbf{p})^T \in T^*\mathbb{R}^2$ by matrix multiplication as

$$\mathbf{z}(s) = M(s)\mathbf{z}(0) = \exp(s\xi)\mathbf{z}(0)\,,$$

in which $M(s)JM^T(s) = J$ is a symplectic 2×2 matrix. The 2×2 matrix tangent to the symplectic matrix $M(s)$ at the identity $s = 0$ is given by

$$\xi = \Big[M\,'(s)M^{-1}(s)\Big]_{s=0}\,.$$

This is a 2×2 Hamiltonian matrix in $sp(2,\mathbb{R})$, satisfying (7.8) as

$$J\xi + \xi^T J = 0 \quad \text{so that} \quad J\xi = (J\xi)^T\,. \tag{9.1}$$

That is, for $\xi \in sp(2,\mathbb{R})$, the matrix $J\xi$ is symmetric.
The vector field $\xi_M(\mathbf{z}) \in T\mathbb{R}^2$ may be expressed as a derivative,

$$\xi_M(\mathbf{z}) = \frac{d}{ds}\big[\exp(s\xi)\mathbf{z}\big]\Big|_{s=0} = \xi\mathbf{z}\,,$$

in which the diagonal action $(\xi\mathbf{z})$ of the Hamiltonian matrix (ξ) and the two-component real multi-vector $\mathbf{z} = (\mathbf{q},\, \mathbf{p})^T$ has components given by $(\xi_{kl}q_l, \xi_{kl}p_l)^T$, with $k,l = 1,2$. The matrix ξ is any linear combination of the traceless constant Hamiltonian matrices (7.5).

Definition

9.1 (Momentum map $\mathcal{J} : T^*\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2 \to sp(2,\mathbb{R})^*$)**.**
The map $\mathcal{J} : T^*\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2 \to sp(2,\mathbb{R})^*$ *is defined as follows, where* $\mathbf{z} = (\mathbf{q},\, \mathbf{p})^T \in \mathbb{R}^2 \times \mathbb{R}^2$

$$\begin{aligned} \mathcal{J}^\xi(\mathbf{z}) &:= \Big\langle \mathcal{J}(\mathbf{z}),\, \xi \Big\rangle_{sp(2,\mathbb{R})^* \times sp(2,\mathbb{R})} \\ &= \Big(\mathbf{z},\, J\xi\mathbf{z} \Big)_{\mathbb{R}^2\times\mathbb{R}^2} := z_k (J\xi)_{kl} z_l = \mathbf{z}^T \cdot J\xi\mathbf{z} \\ &= \operatorname{tr}\Big((\mathbf{z} \otimes \mathbf{z}^T J)\xi \Big) \,. \end{aligned} \tag{9.2}$$

Remark

9.2. *The map* $\mathcal{J}(\mathbf{z})$ *given in (9.2) by*

$$\mathcal{J}(\mathbf{z}) = (\mathbf{z} \otimes \mathbf{z}^T J) \in sp(2,\mathbb{R})^* \tag{9.3}$$

sends $\mathbf{z} = (\mathbf{q},\, \mathbf{p})^T \in \mathbb{R}^2 \times \mathbb{R}^2$ *to* $\mathcal{J}(\mathbf{z}) = (\mathbf{z} \otimes \mathbf{z}^T J)$*, which is an element of* $sp(2,\mathbb{R})^*$*, the dual space to* $sp(2,\mathbb{R})$*. Under the pairing* $\langle \cdot\,,\, \cdot \rangle : sp(2,\mathbb{R})^* \times sp(2,\mathbb{R}) \to \mathbb{R}$ *given by the trace of the matrix product, one finds the Hamiltonian, or phase-space function,*

$$\Big\langle \mathcal{J}(\mathbf{z}),\, \xi \Big\rangle = \operatorname{tr}\Big(\mathcal{J}(\mathbf{z})\, \xi \Big) \,, \tag{9.4}$$

for $\mathcal{J}(\mathbf{z}) = (\mathbf{z} \otimes \mathbf{z}^T J) \in sp(2,\mathbb{R})^*$ *and* $\xi \in sp(2,\mathbb{R})$*.*

Remark

9.3 (Map to axisymmetric invariant variables)**.** *The map,* $\mathcal{J}: T^*\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2 \to sp(2,\mathbb{R})^*$ *in (9.2) for* $Sp(2,\mathbb{R})$ *acting diagonally on* $\mathbb{R}^2 \times \mathbb{R}^2$ *in Equation (9.3) may be expressed in matrix form as*

$$
\begin{aligned}
\mathcal{J} &= (\mathbf{z} \otimes \mathbf{z}^T J) \\
&= 2 \begin{pmatrix} \mathbf{p}\cdot\mathbf{q} & -|\mathbf{q}|^2 \\ |\mathbf{p}|^2 & -\mathbf{p}\cdot\mathbf{q} \end{pmatrix} \\
&= 2 \begin{pmatrix} X_3 & -X_1 \\ X_2 & -X_3 \end{pmatrix} .
\end{aligned}
\tag{9.5}
$$

This is none other than the matrix form of the map to axisymmetric invariant variables,

$$
T^*\mathbb{R}^2 \to \mathbb{R}^3 : \ (\mathbf{q},\mathbf{p})^T \to \mathbf{X} = (X_1, X_2, X_3) \,,
$$

defined as

$$
X_1 = |\mathbf{q}|^2 \ge 0 \,, \quad X_2 = |\mathbf{p}|^2 \ge 0 \,, \quad X_3 = \mathbf{p}\cdot\mathbf{q} \,. \tag{9.6}
$$

Applying the momentum map $\mathcal{J}$ *to the vector of Hamiltonian matrices* $\mathbf{m} = (m_1,\, m_2,\, m_3)$ *in Equation (7.5) yields the individual components,*

$$
\mathcal{J}\cdot\mathbf{m} = 2\mathbf{X} \quad \Longleftrightarrow \quad \mathbf{X} = \frac{1}{2}\, z_k (J\mathbf{m})_{kl} z_l \,. \tag{9.7}
$$

Thus, the map $\mathcal{J}: T^*\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2 \to sp(2,\mathbb{R})^*$ *recovers the components of the vector* $\mathbf{X} = (X_1,\, X_2,\, X_3)$ *that are related to the components of the Petzval invariant by* $S^2 = X_1X_2 - X_3^2$.

Exercise. Verify Equation (9.7) explicitly by computing, for example,

$$
\begin{aligned}
X_1 &= \frac{1}{2}(\mathbf{q},\, \mathbf{p}) \cdot (Jm_1) \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} \\
&= \frac{1}{2}(\mathbf{q},\, \mathbf{p}) \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ -2 & 0 \end{pmatrix} \begin{pmatrix} \mathbf{q} \\ \mathbf{p} \end{pmatrix} \\
&= |\mathbf{q}|^2 \,.
\end{aligned}
$$

★

Remark

9.4. *[Summary]*

- The momentum map $\mathcal{J} : T^*\mathbb{R}^2 \simeq \mathbb{R}^2 \times \mathbb{R}^2 \to sp(2,\mathbb{R})^*$ in (9.5) is a Poisson map. That is, it maps Poisson brackets on phase space into Poisson brackets on the target space.

- The corresponding Lie algebra product in $sp(2,\mathbb{R})$ may also be identified with the Lie algebra of divergence-free vector fields in the space $\mathbf{X} \in \mathbb{R}^3$ by using the Nambu bracket in $\mathbb{R}^3$,

$$\frac{d}{dt}F(\mathbf{X}) = \{F,\, H\}_{S^2} = \underbrace{(\nabla S^2 \times \nabla H) \cdot \nabla F}_{\text{Div-free VF in } \mathbb{R}^3} \quad \text{for} \quad S^2 = X_1 X_2 - X_3^2 \tag{9.8}$$

  This gives another proof that the $\mathbb{R}^3$ brackets among (X_1, X_2, X_3) close among themselves.

- When the corresponding Nambu bracket relations are all linear, they will be ***Lie–Poisson brackets***. This has been the case for Lie-Poisson reduction of the Hamiltonian formulation of Fermat's principle for ray optics, which yielded the momentum map for $M \in Sp(2,\mathbb{R})$

$$\begin{array}{ccc}
T^*\mathbb{R}^2 & \xrightarrow{\quad \Phi_{M(t)} \quad} & T^*\mathbb{R}^2 \\
\mathcal{J}(0) \Big\downarrow & \text{Right-equivariant Momentum Map} & \Big\downarrow \mathcal{J}(t) = \mathbf{z} \otimes \mathbf{z}^T J \\
\mathfrak{sp}(2,\mathbb{R})^* & \xrightarrow{\quad \mathrm{Ad}^*_{M(t)^{-1}} \quad} & \mathfrak{sp}(2,\mathbb{R})^* \simeq T^*\mathbb{R}^2/Sp(2,\mathbb{R})
\end{array}$$

Definition

9.5 (Dual pair). *Let (M, ω) be a symplectic manifold and P_1, P_2 be two Poisson manifolds. A pair of Poisson mappings*

$$P_1 \xleftarrow{J_1} (M, \omega) \xrightarrow{J_2} P_2$$

is called a dual pair [We1983b] if $\ker TJ_1$ *and* $\ker TJ_2$ *are symplectic orthogonal complements of one another. That is*

$$(\ker TJ_1)^\omega = \ker TJ_2. \tag{9.9}$$

A systematic treatment of dual pairs can be found in Chapter 11 of [OrRa2004].

Exercise. Show that the pair of momentum maps

$$\mathbb{R} \xleftarrow{S} (T^*\mathbb{R}^2, \omega) \xrightarrow{\mathcal{J}} \mathfrak{sp}(2, \mathbb{R})^* \simeq \mathbb{R}^3 \tag{9.10}$$

is a dual pair.

Hint: The momentum maps for the commuting Hamiltonian actions of $Sp(2, \mathbb{R})$ and S^1 on $(T^*\mathbb{R}^2, \omega)$ are equivariant, hence they form a pair of Poisson maps. S is obviously $Sp(2, \mathbb{R})$ invariant, so the dual pair property $(\ker TS)^\omega = \ker T\mathcal{J}$ follows if one shows that $Sp(2, \mathbb{R})$ acts transitively on fibers of S. ★

Thanks for listening!

++

9.2 Lie–Poisson brackets with quadratic Casimirs

An interesting class of Lie–Poisson brackets emerges from the $\mathbb{R}^3$ Poisson bracket,

$$\{F, H\}_C := -\nabla C \cdot \nabla F \times \nabla H \,, \tag{9.11}$$

when its Casimir gradient is the linear form on $\mathbb{R}^3$ given by $\nabla C = \mathsf{K}\mathbf{X}$ associated with the 3×3 symmetric matrix $\mathsf{K}^T = \mathsf{K}$. This bracket may be written equivalently in various notations, including index form, $\mathbb{R}^3$ vector form, and Lie–Poisson form, as

$$\begin{aligned} \{F,\, H\}_{\mathsf{K}} &= -\nabla C \cdot \nabla F \times \nabla H \\ &= -X_l \mathsf{K}^{li} \epsilon_{ijk} \frac{\partial F}{\partial X_j} \frac{\partial H}{\partial X_k} \\ &= -\mathbf{X} \cdot \mathsf{K} \left(\frac{\partial F}{\partial \mathbf{X}} \times \frac{\partial H}{\partial \mathbf{X}} \right) \\ &=: -\left\langle \mathbf{X} \,,\, \left[\frac{\partial F}{\partial \mathbf{X}} \,,\, \frac{\partial H}{\partial \mathbf{X}} \right]_{\mathsf{K}} \right\rangle . \end{aligned} \tag{9.12}$$

Remark

9.6. *The triple scalar product of gradients in the $\mathbb{R}^3$ bracket (9.11) is the determinant of the Jacobian matrix for the transformation $(X_1, X_2, X_3) \to (C, F, H)$, which is known to satisfy the Jacobi identity. Being a special case, the Poisson bracket $\{F,\, H\}_{\mathsf{K}}$ also satisfies the Jacobi identity.*

In terms of the $\mathbb{R}^3$ components, the Poisson bracket (9.12) yields

$$\{X_j,\, X_k\}_{\mathsf{K}} = -X_l \mathsf{K}^{li} \epsilon_{ijk}\,. \tag{9.13}$$

The Lie–Poisson form in (9.12) associates the $\mathbb{R}^3$ bracket with a Lie algebra with structure constants given in the dual vector basis by

$$[\mathbf{e}_j,\, \mathbf{e}_k]_{\mathsf{K}} = \mathbf{e}_l \mathsf{K}^{li} \epsilon_{ijk} =: \mathbf{e}_l c^{\,l}_{\ jk}\,. \tag{9.14}$$

The Lie group belonging to this Lie algebra is the invariance group of the quadratic Casimir. Namely, it is the transformation group G_{K} with elements $O(s) \in G_{\mathsf{K}}$, with $O(t)|_{t=0} = Id$, whose action from the left on $\mathbb{R}^3$ is given by $\mathbf{X} \to O\mathbf{X}$, such that

$$O^T(t)\mathsf{K}O(t) = \mathsf{K} \tag{9.15}$$

or, equivalently,

$$\mathsf{K}^{-1}O^T(t)\mathsf{K} = O^{-1}(t)\,, \tag{9.16}$$

for the 3×3 symmetric matrix $\mathsf{K}^T = \mathsf{K}$.

Definition

9.7. *An* $n \times n$ ***orthogonal*** *matrix* $O(t)$ *satisfies*

$$O^T(t)\mathrm{Id}\,O(t) = \mathrm{Id} \tag{9.17}$$

in which Id *is the* $n \times n$ *identity matrix. These matrices represent the orthogonal Lie group in* n *dimensions, denoted* $O(n)$*.*

Definition

9.8 (Quasi-orthogonal matrices)**.** *A matrix* $O(t)$ *satisfying (9.15) is called* ***quasi-orthogonal*** *with respect to the symmetric matrix* K*.*

The quasi-orthogonal transformations $\mathbf{X} \to O\mathbf{X}$ are *not* orthogonal, unless $\mathsf{K} = \mathrm{Id}$. However, they do form a Lie group under matrix multiplication, since for any two of them O_1 and O_2, we have

$$(O_1O_2)^T\mathsf{K}(O_1O_2) = O_2^T(O_1^T\mathsf{K}O_1)O_2 = O_2^T\mathsf{K}O_2 = \mathsf{K}\,. \tag{9.18}$$

The corresponding Lie algebra $\mathfrak{g}_\mathsf{K}$ is the derivative of the defining condition of the Lie group (9.15), evaluated at the identity. This yields

$$0 = \big[\dot{O}^TO^{-T}\big]_{t=0}\mathsf{K} + \mathsf{K}\big[O^{-1}\dot{O}\big]_{t=0}\,.$$

Consequently, if $\widehat{X} = [O^{-1}\dot{O}]_{t=0} \in \mathfrak{g}_\mathsf{K}$, the quantity $\mathsf{K}\widehat{X}$ is skew. That is,

$$(\mathsf{K}\widehat{X})^T = -\,\mathsf{K}\widehat{X}\,.$$

A vector representation of this skew matrix is provided by the following ***hat map*** from the Lie algebra $\mathfrak{g}_{\mathsf{K}}$ to vectors in $\mathbb{R}^3$,

$$\widehat{\ }: \mathfrak{g}_{\mathsf{K}} \to \mathbb{R}^3 \quad \text{defined by} \quad (\mathsf{K}\widehat{X})_{jk} = -\, X_l \mathsf{K}^{li} \epsilon_{ijk} \,. \tag{9.19}$$

When K is invertible, the hat map $(\,\widehat{\cdot}\,)$ in (9.19) is a linear isomorphism. For any vectors $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^3$ with components u^j, v^k where $j, k = 1, 2, 3$, one computes

$$\begin{aligned} u^j(\mathsf{K}\widehat{X})_{jk}v^k &= -\,\mathbf{X}\cdot\mathsf{K}(\mathbf{u}\times\mathbf{v}) \\ &=: -\,\mathbf{X}\cdot[\mathbf{u}\,,\,\mathbf{v}]_{\mathsf{K}}\,. \end{aligned}$$

This is the Lie–Poisson bracket for the Lie algebra structure represented on $\mathbb{R}^3$ by the vector product

$$[\mathbf{u}, \mathbf{v}]_{\mathsf{K}} = \mathsf{K}(\mathbf{u}\times\mathbf{v})\,. \tag{9.20}$$

Thus, the Lie algebra of the Lie group of transformations of $\mathbb{R}^3$ leaving invariant the quadratic form $\frac{1}{2}\mathbf{X}^T\cdot\mathsf{K}\mathbf{X}$ may be identified with the cross product of vectors in $\mathbb{R}^3$ by using the K-pairing instead of the usual dot product. For example, in the case of the Petzval invariant we have

$$\nabla S^2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix}\mathbf{X}\,, \quad \text{so that,} \quad \mathsf{K} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -2 \end{pmatrix},$$

for ray optics, with $\mathbf{X} = (X_1,\, X_2,\, X_3)^T$.

Exercise. Verify that inserting this formula for K into formula (9.13) recovers the Lie–Poisson bracket relations (8.5) for ray optics (up to an inessential constant). ★

Hence, we have proved the following theorem.

Theorem

9.9. *Consider the* $\mathbb{R}^3$ *bracket in Equation (9.12)*

$$\{F, H\}_{\mathsf{K}} := -\nabla C_{\mathsf{K}} \cdot \nabla F \times \nabla H \quad \textit{with} \quad C_{\mathsf{K}} = \frac{1}{2}\mathbf{X} \cdot \mathsf{K}\mathbf{X}\,, \tag{9.21}$$

in which $\mathsf{K}^T = \mathsf{K}$ *is a* 3×3 *real symmetric matrix and* $\mathbf{X} \in \mathbb{R}^3$. *The quadratic form* C_{K} *is the Casimir function for the Lie–Poisson bracket given by*

$$\begin{aligned}\{F,\, H\}_{\mathsf{K}} &= -\mathbf{X} \cdot \mathsf{K}\left(\frac{\partial F}{\partial \mathbf{X}} \times \frac{\partial H}{\partial \mathbf{X}}\right), &(9.22)\\ &=: -\left\langle \mathbf{X}\,,\, \left[\frac{\partial F}{\partial \mathbf{X}}, \frac{\partial H}{\partial \mathbf{X}}\right]_{\mathsf{K}}\right\rangle, &(9.23)\end{aligned}$$

defined on the dual of the three-dimensional Lie algebra $\mathfrak{g}_{\mathsf{K}}$, *whose bracket has the following vector product representation for* $\mathbf{u}$, $\mathbf{v} \in \mathbb{R}^3$,

$$[\,\mathbf{u}, \mathbf{v}\,]_{\mathsf{K}} = \mathsf{K}(\mathbf{u} \times \mathbf{v})\,. \tag{9.24}$$

This is the Lie algebra bracket for the Lie group G_{K} *of transformations of* $\mathbb{R}^3$ *given by action from the left,* $\mathbf{X} \to O\mathbf{X}$, *such that* $O^T\mathsf{K}O = \mathsf{K}$, *thereby leaving the quadratic form* C_{K} *invariant.*

Definition

9.10 (The ad and ad* operations). *The adjoint* (ad) *and coadjoint* (ad*) *operations are defined for the Lie–Poisson bracket (9.23) with the quadratic Casimir,* $C_{\mathsf{K}} = \frac{1}{2}\mathbf{X}\cdot\mathsf{K}\mathbf{X}$, *as follows:*

$$
\begin{aligned}
\langle \mathbf{X}\,,\,[\,\mathbf{u},\mathbf{v}\,]_{\mathsf{K}}\rangle &= \langle \mathbf{X}\,,\,\mathrm{ad}_{\mathbf{u}}\mathbf{v}\rangle = \langle \mathrm{ad}^*_{\mathbf{u}}\mathbf{X}\,,\,\mathbf{v}\rangle \\
&= \mathsf{K}\mathbf{X}\cdot(\mathbf{u}\times\mathbf{v}) = (\mathsf{K}\mathbf{X}\times\mathbf{u})\cdot\mathbf{v}\,.
\end{aligned}
\tag{9.25}
$$

Thus, we have explicitly

$$
\mathrm{ad}_{\mathbf{u}}\mathbf{v} = \mathsf{K}(\mathbf{u}\times\mathbf{v}) \quad \textit{and} \quad \mathrm{ad}^*_{\mathbf{u}}\mathbf{X} = -\,\mathbf{u}\times\mathsf{K}\mathbf{X}\,.
\tag{9.26}
$$

These definitions of the ad and ad* operations yield the following theorem for the dynamics.

Theorem

9.11 (Lie Poisson dynamics). *The Lie–Poisson dynamics (9.22) and (9.23) expressed in terms of the* ad *and* ad* *operations by*

$$
\begin{aligned}
\frac{dF}{dt} = \{F,\,H\}_{\mathsf{K}} &= \left\langle \mathbf{X}\,,\,\mathrm{ad}_{\partial H/\partial \mathbf{X}}\,\frac{\partial F}{\partial \mathbf{X}}\right\rangle \\
&= \left\langle \mathrm{ad}^*_{\partial H/\partial \mathbf{X}}\mathbf{X}\,,\,\frac{\partial F}{\partial \mathbf{X}}\right\rangle ,
\end{aligned}
\tag{9.27}
$$

so that the Lie–Poisson dynamics expresses itself as coadjoint motion,

$$\frac{d\mathbf{X}}{dt} = \{\mathbf{X},\, H\}_{\mathsf{K}} = \mathrm{ad}^*_{\partial H/\partial \mathbf{X}}\mathbf{X} = -\,\frac{\partial H}{\partial \mathbf{X}} \times \mathsf{K}\mathbf{X}\,. \tag{9.28}$$

By construction, this equation conserves the quadratic Casimir, $C_{\mathsf{K}} = \frac{1}{2}\,\mathbf{X}\cdot\mathsf{K}\mathbf{X}$.

Exercise. Write the equations of coadjoint motion (9.28) for $\mathsf{K} = \mathrm{diag}(1,1,1)$ and $H = X_1^2 - X_3^2/2$. ★

10 Divergenceless vector fields

10.1 Jacobi identity

One may verify directly that the $\mathbb{R}^3$ bracket in (9.8) and in the class of brackets (9.21) does indeed satisfy the defining properties of a Poisson bracket. Clearly, it is a bilinear, skew-symmetric form. To show that it is also a Leibniz operator that satisfies the Jacobi identity, we identify the bracket in (9.8) with the following ***divergenceless vector field*** on $\mathbb{R}^3$ defined by

$$X_H = \{\,\cdot\,, H\} = \nabla S^2 \times \nabla H \cdot \nabla \in \mathfrak{X}\,. \tag{10.1}$$

This isomorphism identifies the bracket in (10.1) acting on functions on $\mathbb{R}^3$ with the action of the divergenceless vector fields $\mathfrak{X}$. It remains to verify the Jacobi identity explicitly, by using the properties of the commutator of divergenceless vector fields.

Definition

10.1 (Jacobi–Lie bracket)**.** *The* ***commutator*** *of two divergenceless vector fields $u, v \in \mathfrak{X}$ is defined to be*

$$\big[v,\, w\big] = \big[\mathbf{v}\cdot\nabla,\, \mathbf{w}\cdot\nabla\big] = \Big((\mathbf{v}\cdot\nabla)\mathbf{w} - (\mathbf{w}\cdot\nabla)\mathbf{v}\Big)\cdot\nabla\,. \tag{10.2}$$

The coefficient of the commutator of vector fields is called the ***Jacobi–Lie bracket****. It may be written without risk of confusion in the same notation as*

$$\big[v,\, w\big] = (\mathbf{v}\cdot\nabla)\mathbf{w} - (\mathbf{w}\cdot\nabla)\mathbf{v}\,. \tag{10.3}$$

In Euclidean vector components, the Jacobi–Lie bracket (10.3) is expressed as

$$\big[v,\, w\big]_i = w_{i,j}v_j - v_{i,j}w_j \,. \tag{10.4}$$

Here, a subscript comma denotes partial derivative, e.g., $v_{i,j} = \partial v_i/\partial x_j$*, and one sums repeated indices over their range, for example,* $i, j = 1, 2, 3$*, in three dimensions.*

Exercise. Show that $[v, w]_{i,i} = 0$ for the expression in (10.4), so the commutator of two divergenceless vector fields yields another one. ★

Remark

10.2 (Interpreting commutators of vector fields)**.** *We may interpret a smooth vector field in* $\mathbb{R}^3$ *as the tangent at the identity* $(\epsilon = 0)$ *of a one-parameter flow* ϕ_ϵ *in* $\mathbb{R}^3$ *parameterised by* $\epsilon \in \mathbb{R}$ *and given by integrating*

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} = v_i(\mathbf{x})\frac{\partial}{\partial x_i} \,. \tag{10.5}$$

The characteristic equations of this flow are

$$\frac{dx_i}{d\epsilon} = v_i(\mathbf{x}(\epsilon))\,, \quad \textit{so that} \quad \frac{dx_i}{d\epsilon}\bigg|_{\epsilon=0} = v_i(\mathbf{x})\,, \quad i = 1, 2, 3\,. \tag{10.6}$$

Integration of the characteristic equations (10.6) yields the solution for the flow $\mathbf{x}(\epsilon) = \phi_\epsilon \mathbf{x}$ *of the vector field defined by (10.5), whose initial condition starts from* $\mathbf{x} = \mathbf{x}(0)$*. Suppose the characteristic* *equations for two such flows parameterised by* $(\epsilon, \sigma) \in \mathbb{R}$ *are given respectively by*

$$\frac{dx_i}{d\epsilon} = v_i(\mathbf{x}(\epsilon)) \quad \textit{and} \quad \frac{dx_i}{d\sigma} = w_i(\mathbf{x}(\sigma))\,.$$

The difference of their cross derivatives evaluated at the identity yields the Jacobi–Lie bracket,

$$\begin{aligned}
\frac{d}{d\epsilon}\Big|_{\epsilon=0}\frac{dx_i}{d\sigma}\Big|_{\sigma=0} - \frac{d}{d\sigma}\Big|_{\sigma=0}\frac{dx_i}{d\epsilon}\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}w_i(\mathbf{x}(\epsilon))\Big|_{\epsilon=0} - \frac{d}{d\sigma}v_i(\mathbf{x}(\sigma))\Big|_{\sigma=0} \\
&= \frac{\partial w_i}{\partial x_j}\frac{dx_i}{d\epsilon}\Big|_{\epsilon=0} - \frac{\partial v_i}{\partial x_j}\frac{dx_i}{d\sigma}\Big|_{\sigma=0} \\
&= w_{i,j}v_j - v_{i,j}w_j \\
&= [v, w]_i\,.
\end{aligned}$$

Thus, the Jacobi–Lie bracket of vector fields is the difference between the cross derivatives with respect to their corresponding characteristic equations, evaluated at the identity. Of course, this difference of cross derivatives would vanish if each derivative were not evaluated before taking the next one.

The composition of Jacobi–Lie brackets for three divergenceless vector fields $u, v, w \in \mathfrak{X}$ has components given by

$$\begin{aligned}
\big[u, [v, w]\big]_i &= u_k v_j w_{i,kj} + u_k v_{j,k} w_{i,j} - u_k w_{j,k} v_{i,j} \\
&\quad - u_k w_j v_{i,jk} - v_j w_{k,j} u_{i,k} + w_j v_{k,j} u_{i,k}\,.
\end{aligned} \tag{10.7}$$

Equivalently, in vector form,

$$\begin{aligned}
\big[u, [v, w]\big] &= \mathbf{uv} : \nabla\nabla\mathbf{w} + \mathbf{u}\cdot\nabla\mathbf{v}^T\cdot\nabla\mathbf{w}^T - \mathbf{u}\cdot\nabla\mathbf{w}^T\cdot\nabla\mathbf{v}^T \\
&\quad - \mathbf{uw} : \nabla\nabla\mathbf{v} - \mathbf{v}\cdot\nabla\mathbf{w}^T\cdot\nabla\mathbf{u}^T + \mathbf{w}\cdot\nabla\mathbf{v}^T\cdot\nabla\mathbf{u}^T\,.
\end{aligned}$$

Theorem

10.3. *The Jacobi–Lie bracket of divergenceless vector fields satisfies the Jacobi identity,*

$$[u,[v,w]\,]+[v,[w,u]\,]+[w,[u,v]\,]=0\,. \tag{10.8}$$

Proof. Direct verification using (10.7) and summing over cyclic permutations. □

Exercise. Prove Theorem 10.3 in streamlined notation obtained by writing

$$[v,\ w]=v(w)-w(v)\,,$$

and using bilinearity of the Jacobi–Lie bracket. ★

Lemma

10.4. *The* $\mathbb{R}^3$ *bracket (9.8) may be identified with the divergenceless vector fields in (10.1) by*

$$[X_G,X_H]=-X_{\{G,H\}}\,, \tag{10.9}$$

where $[X_G,X_H]$ *is the Jacobi–Lie bracket of vector fields* X_G *and* X_H*.*

Proof. Equation (10.9) may be verified by a direct calculation,

$$\begin{aligned}
[X_G, X_H] &= X_G X_H - X_H X_G \\
&= \{G,\,\cdot\,\}\{H,\,\cdot\,\} - \{H,\,\cdot\,\}\{G,\,\cdot\,\} \\
&= \{G,\,\{H,\,\cdot\,\}\} - \{H,\,\{G,\,\cdot\,\}\} \\
&= \{\{G,\,H\},\,\cdot\,\} = -X_{\{G,H\}}\,.
\end{aligned}$$

□

Remark

10.5. *The last step in the proof of Lemma 10.4 uses the Jacobi identity for the* $\mathbb{R}^3$ *bracket, which follows from the Jacobi identity for divergenceless vector fields, since*

$$X_F X_G X_H = -\{F,\,\{G,\,\{H,\,\cdot\,\}\}\}\,.$$

10.2 Geometric forms of Poisson brackets

10.2.1 Determinant and wedge-product forms of the canonical bracket

For one degree of freedom, the canonical Poisson bracket $\{F, H\}$ is the same as the determinant for a change of variables $(q, p) \to (F(q, p), H(q, p))$,

$$\{F, H\} = \frac{\partial F}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial H}{\partial q}\frac{\partial F}{\partial p} = \det\frac{\partial(F, H)}{\partial(q, p)}\,. \tag{10.10}$$

This may be written in terms of the differentials of the functions $(F(q,p), H(q,p))$ defined as

$$dF = \frac{\partial F}{\partial q}\,dq + \frac{\partial F}{\partial p}\,dp \quad \text{and} \quad dH = \frac{\partial H}{\partial q}\,dq + \frac{\partial H}{\partial p}\,dp\,, \tag{10.11}$$

by writing the canonical Poisson bracket $\{F, H\}$ as a phase-space density

$$dF \wedge dH = \det \frac{\partial(F,H)}{\partial(q,p)} dq \wedge dp = \{F,H\} dq \wedge dp\,. \tag{10.12}$$

Here the wedge product $\wedge$ in $dF \wedge dH = -\, dH \wedge dF$ is introduced to impose the antisymmetry of the Jacobian determinant under interchange of its columns.

Definition

10.6 (Wedge product of differentials)**.** *The wedge product of differentials (dF, dG, dH) of any smooth functions (F, G, H) is defined by the following three properties:*

- $\wedge$ *is anticommutative:* $dF \wedge dG = -dG \wedge dF$*;*
- $\wedge$ *is bilinear:* $(adF + bdG) \wedge dH = a(dF \wedge dH) + b(dG \wedge dH)$*;*
- $\wedge$ *is associative:* $dF \wedge (dG \wedge dH) = (dF \wedge dG) \wedge dH$*.*

Remark

10.7. *These are the usual properties of area elements and volume elements in integral calculus. These properties also apply in computing changes of variables.*

Exercise. Verify formula (10.12) from Equation (10.11) and the linearity and antisymmetry of the wedge product, so that, e.g., $dq \wedge dp = -dp \wedge dq$ and $dq \wedge dq = 0$. ★

10.2.2 Determinant and wedge-product forms of the $\mathbb{R}^3$ bracket

The $\mathbb{R}^3$ bracket in Equation (9.8) may also be rewritten equivalently as a Jacobian determinant, namely,

$$\{F,\, H\} = -\,\nabla S^2 \cdot \nabla F \times \nabla H = -\,\frac{\partial(S^2, F, H)}{\partial(X_1, X_2, X_3)}\,, \tag{10.13}$$

where

$$\frac{\partial(F_1, F_2, F_3)}{\partial(X_1, X_2, X_3)} = \det\left(\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right)\,. \tag{10.14}$$

The determinant in three dimensions may be defined using the antisymmetric tensor symbol ϵ_{ijk} as

$$\epsilon_{ijk}\det\left(\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right) = \epsilon_{abc}\,\frac{\partial F_a}{\partial X_i}\frac{\partial F_b}{\partial X_j}\frac{\partial F_c}{\partial X_k}\,, \tag{10.15}$$

where, as mentioned earlier, we sum repeated indices over their range. We shall keep track of the antisymmetry of the determinant in three dimensions by using the wedge product ($\wedge$):

$$\det\left(\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right) dX_1 \wedge dX_2 \wedge dX_3 = dF_1 \wedge dF_2 \wedge dF_3\,. \tag{10.16}$$

Thus, the $\mathbb{R}^3$ bracket in Equation (9.8) may be rewritten equivalently in wedge-product form as

$$
\begin{aligned}
\{F,\,H\}\,dX_1\wedge dX_2\wedge dX_3 &= -(\nabla S^2\cdot\nabla F\times\nabla H)\,dX_1\wedge dX_2\wedge dX_3\\
&= -\,dS^2\wedge dF\wedge dH\,.
\end{aligned}
$$

Poisson brackets of this type are called Nambu brackets, since [Na1973] introduced them in three dimensions. They can be generalised to any dimension, but this requires additional compatibility conditions [Ta1994].

10.3 Nambu brackets

Theorem

10.8 (Nambu brackets [?]). *For any smooth functions* $F\,H\in\mathcal{F}(\mathbb{R}^3)$ *of coordinates* $\mathbf{X}\in\mathbb{R}^3$ *with volume element* $d^{\,3}X$*, the* ***Nambu bracket***

$$\{F,\,H\}:\mathcal{F}(\mathbb{R}^3)\times\mathcal{F}(\mathbb{R}^3)\to\mathcal{F}(\mathbb{R}^3)$$

defined by

$$
\begin{aligned}
\{F,\,H\}\,d^{\,3}X &= -\nabla C\cdot\nabla F\times\nabla H d^{\,3}X\\
&= -\,dC\wedge dF\wedge dH
\end{aligned}
\tag{10.17}
$$

possesses the properties (4.4) required of a Poisson bracket for any choice of distinguished smooth function $C:\mathbb{R}^3\to\mathbb{R}$.

Proof. The bilinear skew-symmetric Nambu $\mathbb{R}^3$ bracket yields the divergenceless vector field

$$X_H = \{\,\cdot\,,\, h\} = \nabla C \times \nabla H \cdot \nabla\,,$$

in which

$$\operatorname{div}(\nabla C \times \nabla H) = 0\,.$$

Divergenceless vector fields are derivative operators that satisfy the Leibniz product rule and the Jacobi identity. These properties hold in this case for any choice of smooth functions $C,\, H \in \mathcal{F}(\mathbb{R}^3)$. The other two properties – bilinearity and skew-symmetry – hold as properties of the wedge product. Hence, the Nambu $\mathbb{R}^3$ bracket in (10.17) satisfies all the properties required of a Poisson bracket specified in Definition 4.4. □

11 Geometry of solution behaviour

11.1 Restricting axisymmetric ray optics to level sets

Having realised that the $\mathbb{R}^3$ bracket in Equation (9.8) is associated with Jacobian determinants for changes of variables, it is natural to transform the dynamics of the axisymmetric optical variables (5.5) from three dimensions $(X_1, X_2, X_3) \in \mathbb{R}^3$ to one of its level sets $S^2 > 0$. For convenience, we first make a linear change of Cartesian coordinates in $\mathbb{R}^3$ that explicitly displays the axisymmetry of the level sets of S^2 under rotations, namely,

$$S^2 = X_1X_2 - X_3^2 = Y_1^2 - Y_2^2 - Y_3^2 \,, \tag{11.1}$$

with

$$Y_1 = \frac{1}{2}(X_1 + X_2)\,, \quad Y_2 = \frac{1}{2}(X_2 - X_1)\,, \quad Y_3 = X_3\,.$$

In these new Cartesian coordinates $(Y_1, Y_2, Y_3) \in \mathbb{R}^3$, the level sets of S^2 are manifestly invariant under rotations about the Y_1-axis.

Exercise. Show that this linear change of Cartesian coordinates preserves the orientation of volume elements, but scales them by a constant factor of one-half. That is, show

$$\{F,\, H\}\, dY_1 \wedge dY_2 \wedge dY_3 = \frac{1}{2}\{F,\, H\}\, dX_1 \wedge dX_2 \wedge dX_3\,.$$

The overall constant factor of one-half here is unimportant, because it may be simply absorbed into the units of axial distance in the dynamics induced by the $\mathbb{R}^3$ bracket for axisymmetric ray optics in the Y-variables. ★

Each of the family of hyperboloids of revolution in (11.1) comprises a layer in the hyperbolic onion preserved by axisymmetric ray optics. We use hyperbolic polar coordinates on these layers in analogy to spherical coordinates,

$$Y_1 = S\cosh u\,, \quad Y_2 = S\sinh u\cos\psi\,, \quad Y_3 = S\sinh u\sin\psi\,. \tag{11.2}$$

The $\mathbb{R}^3$ bracket (9.8) thereby transforms into hyperbolic coordinates (11.2) as

$$\{F,\, H\}\, dY_1\wedge dY_2\wedge dY_3 = -\,\{F,\, H\}_{hyperb}\, S^2\, dS\wedge d\psi\wedge d\cosh u\,. \tag{11.3}$$

Note that the oriented quantity

$$S^2 d\cosh u\wedge d\psi = -S^2 d\psi\wedge d\cosh u$$

is the area element on the hyperboloid corresponding to the constant S^2.

On a constant level surface of S^2 the function $\{F, H\}_{hyperb}$ only depends on $(\cosh u, \psi)$ so the Poisson bracket for optical motion on any *particular* hyperboloid is then

$$\begin{aligned}\{F,\, H\}\, d^{\,3}Y &= -\,S^2 dS\wedge dF\wedge dH \\ &= -\,S^2 dS\wedge\{F,\, H\}_{hyperb}\, d\psi\wedge d\cosh u\,,\end{aligned} \tag{11.4}$$

with

$$\{F,\, H\}_{hyperb} = \left(\frac{\partial F}{\partial\psi}\frac{\partial H}{\partial\cosh u} - \frac{\partial H}{\partial\cosh u}\frac{\partial F}{\partial\psi}\right)\,.$$

Being a constant of the motion, the value of S^2 may be absorbed by a choice of units for any given initial condition and the Poisson bracket for the optical motion thereby becomes canonical on each hyperboloid,

$$\frac{d\psi}{dz} = \{\psi,\, H\}_{hyperb} = \frac{\partial H}{\partial \cosh u}\,, \tag{11.5}$$

$$\frac{d\cosh u}{dz} = \{\cosh u,\, H\}_{hyperb} = -\frac{\partial H}{\partial \psi}\,. \tag{11.6}$$

In the Cartesian variables $(Y_1, Y_2, Y_3) \in \mathbb{R}^3$, one has $\cosh u = Y_1/S$ and $\psi = \tan^{-1}(Y_3/Y_2)$. In the original variables,

$$\cosh u = \frac{X_1 + X_2}{2S} \quad \text{and} \quad \psi = \tan^{-1}\frac{2X_3}{X_2 - X_1}\,.$$

Example

11.1. *For a paraxial harmonic guide, whose Hamiltonian is*

$$H = \frac{1}{2}(|\mathbf{p}|^2 + |\mathbf{q}|^2) = \frac{1}{2}(X_1 + X_2) = Y_1\,, \tag{11.7}$$

the ray paths consist of circles cut by the intersections of level sets of the planes $Y_1 = const$ with the hyperboloids of revolution about the Y_1-axis, given by $S^2 = const$.

The dynamics for $\mathbf{Y} \in \mathbb{R}^3$ is given by

$$\dot{\mathbf{Y}} = \{\mathbf{Y},\, H\} = \nabla_{\mathbf{Y}} S^2 \times \widehat{\mathbf{Y}}_\mathbf{1} = 2\widehat{\mathbf{Y}}_\mathbf{1} \times \mathbf{Y}\,, \tag{11.8}$$

on using (11.1) to transform the $\mathbb{R}^3$ bracket in (9.8). Thus, for the paraxial harmonic guide, the rays spiral down the optical axis following circular helices whose radius is determined by their initial conditions.

Exercise. Verify that Equation (11.3) transforms the $\mathbb{R}^3$ bracket from Cartesian to hyperboloidal coordinates, by using the definitions in (11.2). ★

Exercise. Reduce $\{F, H\}_{hyperb}$ to the conical level set $S = 0$. ★

Exercise. Reduce the $\mathbb{R}^3$ dynamics of (9.8) to level sets of the Hamiltonian

$$H = aX_1 + bX_2 + cX_3\,,$$

for constants (a, b, c). Explain how this reduction simplifies the equations of motion. ★

11.2 Geometric phase on level sets of $S^2 = p_\phi^2$

In polar coordinates, the axisymmetric invariants are

$$
\begin{aligned}
Y_1 &= \frac{1}{2}\Big(p_r^2 + p_\phi^2/r^2 + r^2\Big)\,,\\
Y_2 &= \frac{1}{2}\Big(p_r^2 + p_\phi^2/r^2 - r^2\Big)\,,\\
Y_3 &= rp_r\,.
\end{aligned}
$$

Hence, the corresponding volume elements are found to be

$$
\begin{aligned}
d^{\,3}Y &=: dY_1 \wedge dY_2 \wedge dY_3\\
&= d\frac{S^3}{3} \wedge d\cosh u \wedge d\psi\\
&= dp_\phi^2 \wedge dp_r \wedge dr\,.
\end{aligned}
\tag{11.9}
$$

On a level set of $S = p_\phi$ this implies

$$
S\, d\cosh u \wedge d\psi = 2\, dp_r \wedge dr\,, \tag{11.10}
$$

so the transformation of variables $(\cosh u,\, \psi) \to (p_r,\, r)$ is canonical on level sets of $S = p_\phi$.

Stokes theorem on phase space One recalls Stokes theorem on phase space,

$$
\iint_A dp_j \wedge dq_j = \oint_{\partial A} p_j dq_j\,, \tag{11.11}
$$

where the boundary of the phase-space area ∂A is taken around a loop on a closed orbit. Either in polar coordinates or on an invariant hyperboloid $S = p_\phi$ this loop integral becomes

$$\oint \mathbf{p}\cdot d\mathbf{q} := \oint p_j dq_j \;=\; \oint \Big(p_\phi d\phi + p_r dr \Big) \;=\; \oint \Big(\frac{S^3}{3} d\phi + \cosh u \, d\psi \Big) .$$

Thus we may compute the total phase change around a closed periodic orbit on the level set of hyperboloid S from

$$\oint \frac{S^3}{3} d\phi \;=\; \frac{S^3}{3}\Delta\phi \;=\; \underbrace{-\oint \cosh u \, d\psi}_{\text{geometric } \Delta\phi} \;+\; \underbrace{\oint \mathbf{p}\cdot d\mathbf{q}}_{\text{dynamic } \Delta\phi} . \qquad (11.12)$$

Evidently, one may denote the total change in phase as the sum

$$\Delta\phi = \Delta\phi_{geom} + \Delta\phi_{dyn} \,,$$

by identifying the corresponding terms in the previous formula. By the Stokes theorem (11.11), one sees that the geometric phase associated with a periodic motion on a particular hyperboloid is given by the hyperbolic solid angle enclosed by the orbit, times a constant factor depending on the level set value $S = p_\phi$. Thus the name: ***geometric phase***.

12 Geometric ray optics in anisotropic media

> Every ray of light has therefore two opposite sides.... And since the crystal by this disposition or virtue does not act upon the rays except when one of their sides of unusual refraction looks toward that coast, this argues a virtue or disposition in those sides of the rays which answers to and sympathises with that virtue or disposition of the crystal, as the poles of two magnets answer to one another....
>
> – Newton, *Opticks* (1704)

Some media have directional properties that are exhibited by differences in the transmission of light in different directions. This effect is seen, for example, in certain crystals. Fermat's principle for such media still conceives light rays as lines in space (i.e., no polarisation vectors yet), but the refractive index along the paths of the rays in the medium is allowed to depend on both position and *direction*. In this case, Theorem 2.16 adapts easily to yield the expected three-dimensional eikonal Equation (2.7).

12.1 Fermat's and Huygens' principles for anisotropic media

Let us investigate how Fermat's principle (2.6) changes, when Huygens' equation (2.18) allows the velocity of the wave fronts to depend on direction because of material anisotropy.

We consider a medium in which the tangent vector of a light ray does not point along the normal to its wave front, $\nabla S(\mathbf{r})$, but instead satisfies a matrix relation depending on the spatial location along the ray path, $\mathbf{r}(s)$, as

$$\frac{d\mathbf{r}}{ds} = \mathcal{D}^{-1}(\mathbf{r})\nabla S\,, \tag{12.1}$$

with $\dot{\mathbf{r}}(s) := d\mathbf{r}/ds$ and $|\dot{\mathbf{r}}| = 1$ for an invertible matrix function $\mathcal{D}$ that characterises the medium. In components, this ***anisotropic Huygens equation*** is written

$$\frac{\partial S}{\partial r^i} = \mathcal{D}_{ij}(\mathbf{r})\frac{dr^j}{ds}\,. \tag{12.2}$$

We shall write the Euler–Lagrange equation for $\mathbf{r}(s)$ that arises from Fermat's principle for this anisotropic version of Huygens' equation and derive its corresponding Snell's law.

We begin by taking the square on both sides of the vector Equation (12.2), which produces the anisotropic version of the scalar eikonal Equation (2.19), now written in the form

$$|\nabla S|^2 = \frac{d\mathbf{r}}{ds}^T(\mathcal{D}^T\mathcal{D})\frac{d\mathbf{r}}{ds} = \frac{dr^i}{ds}(\mathcal{D}_{ik}\mathcal{D}_{kj})\frac{dr^j}{ds} = n^2(\mathbf{r},\dot{\mathbf{r}}). \tag{12.3}$$

Substituting this expression into Fermat's principle yields

$$\begin{aligned} 0 = \delta \mathsf{A} \;&=\; \delta\int_A^B n(\mathbf{r}(s),\dot{\mathbf{r}}(s))\,ds \\ &=\; \delta\int_A^B \sqrt{\dot{\mathbf{r}}(s)\cdot(\mathcal{D}^T\mathcal{D})(\mathbf{r})\cdot\dot{\mathbf{r}}(s)}\,ds. \end{aligned} \tag{12.4}$$

This is the variational principle for the curve that leaves the distance between two points A and B stationary under variations in a space with ***Riemannian metric*** given by $\mathcal{G} := \mathcal{D}^T\mathcal{D}$, whose element of length is defined by

$$n^2(\mathbf{r},\dot{\mathbf{r}})ds^2 = d\mathbf{r}\cdot\mathcal{G}(\mathbf{r})\cdot d\mathbf{r}. \tag{12.5}$$

When $\mathcal{G}_{ij} = n^2(\mathbf{r})\,\delta_{ij}$, one recovers the isotropic case discussed in Section 2.1.

Remark

12.1. *By construction, the quantity under the integral in (12.4), now rewritten as*

$$\mathsf{A} = \int_A^B \sqrt{\dot{\mathbf{r}} \cdot \mathcal{G}(\mathbf{r}) \cdot \dot{\mathbf{r}}}\, ds = \int_A^B n(\mathbf{r}(s), \dot{\mathbf{r}}(s))\, ds\,, \tag{12.6}$$

is homogeneous of degree 1 and thus is invariant under reparameterising the ray path $\mathbf{r}(s) \to \mathbf{r}(\sigma)$*.*

Continuing the variational computation leads to

$$\begin{aligned} 0 = \delta\mathsf{A} \;&=\; \delta \int_A^B \sqrt{\dot{\mathbf{r}}(s) \cdot \mathcal{G}(\mathbf{r}) \cdot \dot{\mathbf{r}}(s)}\, ds \\ &= \int_A^B \left[\frac{\partial n(\mathbf{r}, \dot{\mathbf{r}})}{\partial \mathbf{r}} - \frac{d}{ds} \left(\frac{\mathcal{G}(\mathbf{r}) \cdot \dot{\mathbf{r}}}{n(\mathbf{r}, \dot{\mathbf{r}})} \right) \right] \cdot \delta\mathbf{r}\, ds \end{aligned}$$

upon substituting $\sqrt{\dot{\mathbf{r}}(s) \cdot \mathcal{G}(\mathbf{r}) \cdot \dot{\mathbf{r}}(s)} = n(\mathbf{r}, \dot{\mathbf{r}})$.

Stationarity $\delta\mathsf{A} = 0$ now yields the Euler–Lagrange equation

$$\dot{\mathbf{p}} = \frac{\partial n(\mathbf{r}, \dot{\mathbf{r}})}{\partial \mathbf{r}}\,, \quad \text{with} \quad \mathbf{p} := \frac{\mathcal{G}(\mathbf{r}) \cdot \dot{\mathbf{r}}}{n(\mathbf{r}, \dot{\mathbf{r}})} = \frac{\partial n(\mathbf{r}, \dot{\mathbf{r}})}{\partial \dot{\mathbf{r}}}\,, \tag{12.7}$$

where the vector $\mathbf{p}$ is the ***optical momentum***. This is the eikonal equation for an anisotropic medium whose Huygens' equation is (12.1).

Note the elegant formula

$$\mathbf{p} \cdot d\mathbf{r} = n(\mathbf{r}, \dot{\mathbf{r}})\, ds\,, \quad \text{or equivalently} \quad \mathbf{p} \cdot \dot{\mathbf{r}} = n(\mathbf{r}, \dot{\mathbf{r}})\,,$$

which follows, as in the isotropic case, from Euler's relation (2.14) for homogeneous functions of degree 1, and in particular for the Lagrangian in (12.6).

Exercise. Transform the arc-length variable in Fermat's principle (12.4) to $d\sigma = n(\mathbf{r}, \dot{\mathbf{r}})ds$ so that

$$d\sigma^2 = d\mathbf{r} \cdot \mathcal{G}(\mathbf{r}) \cdot d\mathbf{r} \tag{12.8}$$

and recalculate its Euler–Lagrange equation. ★

Answer. Denoting $\mathbf{r}'(\sigma) = d\mathbf{r}/d\sigma$ in Fermat's principle (12.4) and recalculating the stationary condition yields

$$0 = \delta \mathsf{A} \;=\; \int_A^B \left[\frac{d\sigma}{2\,nds} r'^k \frac{\partial \mathcal{G}_{kl}}{\partial \mathbf{r}} r'^l - \frac{d}{d\sigma}\left(\frac{d\sigma}{nds} \mathcal{G} \cdot \mathbf{r}' \right) \right] \cdot \delta\mathbf{r}\, d\sigma.$$

Applying $n(\mathbf{r}, \dot{\mathbf{r}})ds = d\sigma$ leads to the Euler–Lagrange equation

$$\frac{1}{2} r'^k \frac{\partial \mathcal{G}_{kl}}{\partial \mathbf{r}} r'^l - \frac{d}{d\sigma}\left(\mathcal{G}(\mathbf{r}) \cdot \mathbf{r}'(\sigma) \right) = 0\,. \tag{12.9}$$

A curve $\mathbf{r}(\sigma)$ satisfying this equation leaves the length between points A and B stationary under variations and is called a ***geodesic*** with respect to the metric $\mathcal{G}(\mathbf{r})$, whose length element is defined in Equation (12.8). ▲

Exercise. Does the variational principle

$$0 = \delta \mathsf{A} = \delta \int_A^B \tfrac{1}{2}\mathbf{r}'(\sigma)\cdot \mathcal{G}(\mathbf{r})\cdot \mathbf{r}'(\sigma)\,d\sigma$$

also imply the Euler–Lagrange Equation (12.9)? Prove it.

Hint: Take a look at Equation (2.12) and the remark after it about Finsler geometry and singular Lagrangians. Does this calculation reveal a general principle about the stationary conditions of Fermat's principle for singular Lagrangians of degree 1 and their associated induced Lagrangians of degree 2? ★

Exercise. Show that the Euler–Lagrange Equation (12.9) may be written as

$$\mathbf{r}''(\sigma) = -\mathbf{G}(\mathbf{r},\mathbf{r}')\,, \quad \text{with} \quad G^i := \Gamma^i_{jk}(\mathbf{r}){r'}^j{r'}^k \tag{12.10}$$

and

$$\Gamma^i_{jk}(\mathbf{r}) = \frac{1}{2}\mathcal{G}^{il}\left[\frac{\partial \mathcal{G}_{lk}(\mathbf{r})}{\partial r^j} + \frac{\partial \mathcal{G}_{lj}(\mathbf{r})}{\partial r^k} - \frac{\partial \mathcal{G}_{jk}(\mathbf{r})}{\partial r^l}\right], \tag{12.11}$$

where $\mathcal{G}_{ki}\mathcal{G}^{il} = \delta^l_k$. Equation (12.11) identifies $\Gamma^i_{jk}(\mathbf{r})$ as the ***Christoffel coefficients*** of the Riemannian metric $\mathcal{G}_{ij}(\mathbf{r})$ in Equation (12.8). Note that the Christoffel coefficients are symmetric under exchange of their lower indices, $\Gamma^i_{jk}(\mathbf{r}) = \Gamma^i_{kj}(\mathbf{r})$.

The vector $\mathbf{G}$ in Equation (12.10) is called the ***geodesic spray*** of the Riemannian metric $\mathcal{G}_{ij}(\mathbf{r})$. Its analytical properties (e.g., smoothness) govern the behaviour of the solutions $\mathbf{r}(\sigma)$ for the geodesic paths. ★

12.2 Ibn Sahl–Snell law for anisotropic media

The statement of the Ibn Sahl–Snell law relation at discontinuities of the refractive index for isotropic media is a bit more involved for an anisotropic medium with Huygens' equation (12.1).

A break in the direction $\widehat{\mathbf{s}}$ of the ray vector is still expected for anisotropic media at any interface of finite discontinuity in the material properties (refractive index and metric) that may be encountered along the ray path $\mathbf{r}(s)$. According to the eikonal equation for anisotropic media (12.7) the jump in three-dimensional optical momentum across a material interface, $\Delta\mathbf{p} := \mathbf{p} - \bar{\mathbf{p}}$ must satisfy the relation

$$\Delta\mathbf{p} \times \frac{\partial n}{\partial \mathbf{r}} = \Delta\left(\frac{\mathcal{G}(\mathbf{r}) \cdot \dot{\mathbf{r}}}{n(\mathbf{r}, \dot{\mathbf{r}})}\right) \times \frac{\partial n}{\partial \mathbf{r}} = 0\,. \tag{12.12}$$

This means the jump in optical momentum $\Delta\mathbf{p}$ at a material interface can only occur in the direction normal to the interface. Hence, the transverse projection of the three-dimensional optical momenta onto the tangent plane will be invariant across the interface.

Let ψ and ψ' denote the angles of incident and transmitted ***momentum directions***, measured from the normal $\hat{\mathbf{z}}$ through the interface. Preservation of the transverse components of the momentum vector

$$\big(\mathbf{p} - \bar{\mathbf{p}}\big) \times \hat{\mathbf{z}} = 0 \tag{12.13}$$

means that these momentum vectors must lie in the same plane and the angles $\psi = \cos^{-1}(\mathbf{p}\cdot\hat{\mathbf{z}}/|\mathbf{p}|)$ and $\bar{\psi} = \cos^{-1}(\bar{\mathbf{p}}\cdot\hat{\mathbf{z}}/|\bar{\mathbf{p}}|)$ must satisfy

$$|\mathbf{p}|\sin\psi = |\bar{\mathbf{p}}|\sin\bar{\psi} \tag{12.14}$$

Relation (12.14) determines the angles of incidence and transmission of the optical momentum directions. However, the transverse components of the optical momentum alone are not enough to determine the ray directions, in general, because the inversion using Equation (12.7) involves all three of the momentum and velocity components.

13 Ten geometrical features of ray optics

1. The design of axisymmetric planar optical systems reduces to multiplication of symplectic matrices corresponding to each element of the system, as in Theorem 7.4.

2. Hamiltonian evolution occurs by canonical transformations. Such transformations may be obtained by integrating the characteristic equations of Hamiltonian vector fields, which are defined by Poisson bracket operations with smooth functions on phase space, as in the proof of Theorem 6.2.

3. The Poisson bracket is associated geometrically with the Jacobian for canonical transformations in Section 10.2. Canonical transformations are generated by Poisson-bracket operations and these transformations preserve the Jacobian.

4. A one-parameter symmetry, that is, an invariance under canonical transformations that are generated by a Hamiltonian vector field $X_{p_\phi} = \{ \cdot \,,\, p_\phi \}$, separates out an angle, ϕ, whose canonically conjugate momentum p_ϕ is conserved. As discussed in Section 4.2, the conserved quantity p_ϕ may be an important bifurcation parameter for the remaining reduced system. The dynamics of the angle ϕ decouples from the reduced system and can be determined as a quadrature after solving the reduced system.

5. Given a symmetry of the Hamiltonian, it may be wise to transform from phase-space coordinates to invariant variables as in (5.5). This transformation defines the quotient map, which quotients out the angle(s) conjugate to the symmetry generator. The image of the quotient map produces the orbit manifold, a reduced manifold whose points are orbits under the symmetry transformation. The corresponding transformation of the Poisson bracket is done using the chain rule as in (8.6). Closure of the Poisson brackets of the invariant variables amongst themselves is a necessary condition for the

quotient map to be a momentum map, as discussed in Section **??**.

6. Closure of the Poisson brackets among an odd number of invariant variables is no cause for regret. It only means that this Poisson bracket among the invariant variables is not canonical (symplectic). For example, the Nambu $\mathbb{R}^3$ bracket (10.17) arises this way.

7. The bracket resulting from transforming to invariant variables could also be Lie–Poisson. This will happen when the new invariant variables are quadratic in the phase-space variables, as occurs for the Poisson brackets among the axisymmetric variables X_1, X_2 and X_3 in (5.5). Then the Poisson brackets among them are *linear* in the new variables with constant coefficients. Those constant coefficients are dual to the structure constants of a Lie algebra. In that case, the brackets will take the Lie–Poisson form (8.7) and the transformation to invariant variables will be the momentum map associated with the action of the symmetry group on the phase space.

8. The orbits of the solutions in the space of axisymmetric invariant variables in ray optics lie on the intersections of level sets of the Hamiltonian and the Casimir for the noncanonical bracket. The Petzval invariant $S^2 = |\mathbf{p} \times \mathbf{q}|^2$ is the Casimir for the Nambu bracket in $\mathbb{R}^3$, which for axisymmetric, translation-invariant ray optics is also a Lie–Poisson bracket. In this case, the ray paths are revealed when the Hamiltonian knife slices through the level sets of the Petzval invariant. These level sets are the layers of the hyperbolic onion shown in Figure 10. When restricted to a level set of the Petzval invariant, the dynamics becomes symplectic.

9. The phases associated with reconstructing the solution from the reduced space of invariant variables by going back to the original space of canonical coordinates and momenta naturally divide into their geometric and dynamic parts as in Equation (11.12). In ray optics as governed by Fermat's principle,

the geometric phase is related to the area enclosed by a periodic solution on a symplectic level set of the Petzval invariant S^2. This is no surprise, because the Poisson bracket on the level set is determined from the Jacobian using that area element.

10. The Lagrangian in Fermat's principle is typically homogeneous of degree 1 and thus is singular; that is, its Hessian determinant vanishes, as occurs in the example of Fermat's principle in anisotropic media discussed in Section 12. In this case, the ray directions cannot be determined from the optical momentum and the coordinate along the ray. In particular, Snell's law of refraction at an interface determines momentum directions, but not ray directions. Even so, the Euler–Lagrange equations resulting from Fermat's principle may be regularised by using an induced Lagrangian that is homogeneous of degree 2, in the sense that reparameterised solutions for the ray paths may be obtained from the resulting ordinary differential equations.

    The Euler–Lagrange Equation (12.9) for geometric optics may be written equivalently in the form (12.10), which shows that ray paths follow geodesic motion through a Riemannian space with a metric that is determined from optical material parameters whose physical meaning is derived from Huygens' equation relating the ray path to the motion of the wave front.

References

[AS06] Ablowitz, M. J. and Segur, H. [2006] *Solitons and the Inverse Scattering Transform*. Cambridge University Press, New York.

[AbMa1978] Abraham, R. and Marsden, J. E. [1978] *Foundations of Mechanics*, 2nd ed. Reading, MA: Addison-Wesley.

[BlIs2006] Bloch, A. M. and Iserles, A. [2006] On an isospectral Lie-Poisson system and its Lie algebra. *Found. Comput. Math.* **6**, 121–144.

[Bo1985] Bogoyavlensky, O. I. [1985] Integrable Euler equations on Lie algebras arising in problems of mathematical physics. *Math. USSR Izvestiya* **25**, 207–257.

[CH93] Camassa,R. and Holm, D. D. [1993] An integrable shallow water equation with peaked solitons. *Phys. Rev. Lett.*, 71:1661–64.

[LDP2000] *Memoirs of Lorenzo Da Ponte*, [2000] Translated by E. Abbott; edited and annotated by A. Livingston; introduction by C. Rosen. The New York Review of Books, New York.

[Da1994] Darling, R. W. R. [1994] *Differential Forms and Connections*. Cambridge: Cambridge University Press.

[DHH03] Degasperis, A., Holm, D. D. and Hone,A. N. W. [2003] Integrable and nonintegrable equations with peakons. In M. J. Ablowitz, M. Boiti, F. Pempinelli, and B. Prinari, editors, *Nonlinear Physics: Theory and Experiment (Gallipoli 2002) Vol II*, pages 37–43. World Scientific, Singapore.

[doCa1976] do Carmo, M. P. [1976] *Differential Geometry of Curves and Surfaces*. Englewood Cliffs, NJ: Prentice-Hall.

[DGH04] Dullin, H., Gottwald, G. and Holm, D. D. [2004] On asymptotically equivalent shallow water wave equations. *Physica D*, 190:1–14.

[EM70] Ebin, D. G. and Marsden, J. E. [1970] Groups of diffeomorphisms and the motion of an incompressible fluid. *Ann. of Math.*, 92:102–163.

[FaTa1987] Faddeev, L. D. and Takhtajan, L. A. [1987] *Hamiltonian Methods in the Theory of Solitons*. New York: Springer.

[Fl1963] Flanders, H. [1963] *Differential Forms with Applications to the Physical Sciences*. New York: Academic Press.

[FH01] Fringer, O. B. and Holm, D. D. [2001] Integrable vs nonintegrable geodesic soliton behavior. *Physica D*, 150:237–263.

[Fuc96] Fuchssteiner, B. [1996] Some tricks from the symmetry-toolbox for nonlinear equations: Generalizations of the Camassa-Holm equation. *Physica D*, 95:229–243.

[GBRa2008] Gay-Balmaz, F. and Ratiu, T.S. [2008] The geometric structure of complex fluids. *Adv. in Appl. Math.* **42** (2), 176–275.

[GaVi2010] Gay-Balmaz F. and Vizman, C. [2011] Dual pairs in fluid dynamics. *Ann. Glob. Anal. Geom.* **41**, 1–24.

[GD79] Gelfand, I. M. and Dorfman, I. Ya. R. [1979] Hamiltonian operators and algebraic structures associated with them. *Funct. Anal. Appl.*, 13:248–254.

[GiHoKu1982] Gibbons, J., Holm, D. D. and Kupershmidt, B. A. [1982] Gauge-invariant Poisson brackets for chromohydrodynamics. *Phys. Lett. A* **90**, 281–283.

[GrKrMa1988] Grossman, R., Krishnaprasad, P. S. and Marsden, J. E. [1988] The dynamics of two coupled rigid bodies. In *Dynamical Systems Approaches to Nonlinear Problems in Systems and Circuits*, edited by F. M. Abdel Salam and M. Levi, Engineering Foundation, United States Army Research Office, and U.S. National Science Foundation. Philadelphia, PA: SIAM, pp. 373–378.

[Ho2002] Holm, D. D. [2002] Euler–Poincaré dynamics of perfect complex fluids. In *Geometry, Mechanics, and Dynamics: In Honor of the 60th Birthday of Jerrold E. Marsden*, edited by P. Newton, P. Holmes and A. Weinstein. New York: Springer, pp. 113–167.

[Ho2005] Holm, D. D. [2005] The Euler–Poincaré variational framework for modeling fluid dynamics. In *Geometric Mechanics and Symmetry: The Peyresq Lectures*, edited by J. Montaldi and T. Ratiu. London Mathematical Society Lecture Notes Series 306. Cambridge: Cambridge University Press.

[Ho2011GM] Holm, D. D. [2011]
Geometric Mechanics I: Dynamics and Symmetry,
Second edition, World Scientific: Imperial College Press, Singapore, .

[Ho2011] Holm, D. D. [2011]
Applications of Poisson geometry to physical problems,
Geometry & Topology Monographs **17**, 221–384.

[HoIvPe2012] Holm, D. D., Ivanov, R. I. and Percival, J. R. [2012] G-Strands, *J. Nonlin. Sci.*, to appear.

[HoKu1988] Holm, D. D. and Kupershmidt, B. A. [1988] The analogy between spin glasses and Yang–Mills fluids. *J. Math Phys.* **29**, 21–30.

[HoMa2004] Holm, D. D. and Marsden, J. E. [2004] Momentum maps and measure valued solutions (peakons, filaments, and sheets) for the EPDiff equation. In *The Breadth of Symplectic and Poisson Geometry*, edited by J. Marsden, and T. Ratiu. Boston: Birkhäuser. arXiv.nlin.CD/0312048.

[HoMaRa1998] Holm, D. D., Marsden, J. E. and Ratiu, T. S. [1998] The Euler–Poincaré equations and semidirect products with applications to continuum theories. *Adv. Math.* **137**, 1–81.

[HoSmSt2009] Holm, D. D., Schmah, T. and Stoica, C. [2009]
Geometric Mechanics and Symmetry: From Finite to Infinite Dimensions,
Oxford University Press.

[HS03] Holm, D. D. and Staley, M. F. [2003] Wave structures and nonlinear balances in a family of 1+1 evolutionary pdes. *SIAM J. Appl. Dyn. Syst.*, 2:323–380.

[HoVi2010] Holm, D. D. and Vizman, C. [2010] *The n:m resonance dual pair*. Preprint at `http://arxiv.org/abs/1102.4377`.

[HZ94] Hunter, J. K. and Zheng, Y. [1994] On a completely integrable nonlinear hyperbolic variational equation. *Physica D*, 79:361–386.

[Iv2006] Ivanov, R. [2006] Hamiltonian formulation and integrability of a complex symmetric nonlinear system. *Phys. Lett. A* **350**, 232–235.

[JKLOR2011] Jia, S., Kirsten, J., Lai, E., Ortmann, K. and Raithatha, A. [2011] An Application (and Translation) of Henri Poincaré's Paper, "*Sur une forme nouvelle des équations de la mécanique*" [Po1901].

[Ku1978] Kummer, M. [1978] On resonant classical Hamiltonians with two equal frequencies. *Commun. Math. Phys.* **58**, 85–112.

[Mag78] Magri, F. [1978] A simple model of the integrable Hamiltonian equation. *J. Math. Phys.*, 19:1156–1162.

[Man1976] Manakov, S. V. [1976] Note on the integration of Euler's equations of the dynamics of and n-dimensional rigid body. *Funct. Anal. Appl.* **10**, 328–329.

[MaRa1994] Marsden, J. E. and Ratiu, T. S. [1994]
Introduction to Mechanics and Symmetry.
Texts in Applied Mathematics, Vol. 75. New York: Springer-Verlag.

[MaRaWe1984a] Marsden, J. E., Ratiu, T. S. and Weinstein, A. [1984a] Semidirect products and reduction in mechanics. *Trans. Amer. Math. Soc.* **281**(1), 147–177.

[MaRaWe1984b] Marsden, J. E., Ratiu, T. S. and Weinstein, A. [1984b] Reduction and Hamiltonian structures on duals of semidirect product Lie algebras. *Contemporary Math.* **28**, 55–100.

[MN02] Mikhailov, A. V. and Novikov, V. S. [2002] Perturbative symmetry approach. *J. Phys. A*, 35:4775–4790.

[Olv00] Olver, P. J. [2000] *Applications of Lie Groups to Differential Equations*. Springer, New York.

[OrRa2004] Ortega, J.-P. and Ratiu, T. S. [2004] *Momentum Maps and Hamiltonian Reduction*. Progress in Mathematics, Vol. 222. Boston: Birkhäuser.

[Po1901] Poincaré, H. [1901] Sur une forme nouvelle des équations de la méchanique, *C.R. Acad. Sci.* **132** 369-371.

[Ra1982] Ratiu, T. [1982] Euler–Poisson equations on Lie algebras and the N-dimensional heavy rigid body. *Am. J. Math.* **104**, 409–448.

[RaTuSbSoTe2005] Ratiu, T. S., Tudoran, R., Sbano, L., Sousa Dias, E. and Terra, G. [2005] A crash course in geometric mechanics. In *Geometric Mechanics and Symmetry: The Peyresq Lectures*, edited by J. Montaldi and T. Ratiu. London Mathematical Society Lecture Notes Series 306. Cambridge: Cambridge University Press.

[RaVM1982] Ratiu, T. and van Moerbeke, P. [1982] The Lagrange rigid body motion. *Annales de l´institut Fourier* **32**(1), 211–234.

[Sk1961] Skyrme, T. H. R. [1961] A nonlinear field theory. *Proc. R. Soc. A* **260**, 127–138.

[Tay96] Taylor, M. E. [1996] *Partial Differential Equations I.* Springer-Verlag, New York.

[We1983a] Weinstein, A. [1983a] Sophus Lie and symplectic geometry. *Expo. Math.* **1**, 95–96.

[We1983b] Weinstein, A. [1983b] The local structure of Poisson manifolds. *J. Differ. Geom.* **18**, 523–557.

[We2002] Weinstein, A. [2002] Geometry of momentum. ArXiv:math/SG0208108 v1.

[Wo2004] Wolf, K. B. [2004] *Geometric Optics on Phase Space*. Berlin: Springer.