Chapter 1

Fermat's ray optics

1.1 Fermat's principle

Fermat

Fermat's principle states that *the path between two points taken by a ray of light is the one traversed in the extremal time.*

This principle accurately describes the properties of light reflected off mirrors, refracted through different media, or transmitted through a medium with a continuously varying index of refraction. Fermat's principle of optimal paths is complementary to Huygens' principle of constructive interference of waves.

According to Huygens, among all possible paths from an object to an image, the waves corresponding to the extremal (stationary) paths contribute most to the image because of their constructive interference. Both principles are approximations to more fundamental physical results derived from Maxwell's equations.

For us, Fermat's principle provides an example that will guide us in recognising the principles of geometric mechanics.

Fermat's principle *defines* a ray of light. Announced in 1662, Fermat's principle for geometric optics preceded Lagrange by a century and Hamilton by more than 150 years. This chapter shows that Fermat's principle naturally introduces Hamilton's principle, phase space, symplectic transformations and momentum maps arising from reduction by symmetry.

Mathematically, Fermat's principle states that the ray path $\mathbf{r}(s)$ from a point A to a point B in space is an *extremal* of ***optical length***, defined by

$$\delta \int_A^B n(\mathbf{r}(s))\, ds = 0\,.$$

Here $n(\mathbf{r})$ is the index of refraction at the spatial point $(\mathbf{r})$ and

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

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

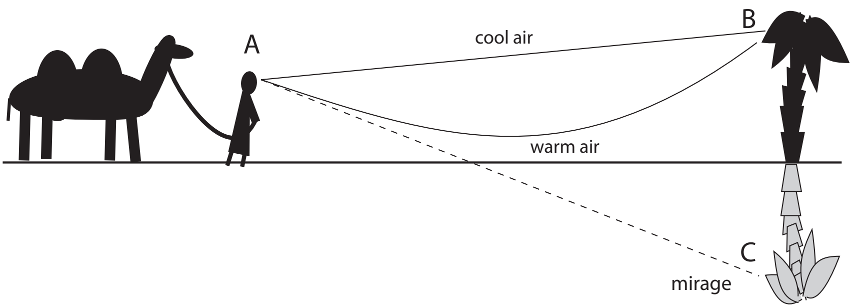

Figure 1.1: Fermat's principle states that the ray path from an observer at A to a point B in space is an *extremal of optical length*. For example, along a sun-baked road, the temperature of the air is warmest near the road and decreases with height, so that the index of refraction, n, increases in the vertical direction. For an observer at A, the curved path has a smaller optical path than 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, thereby giving the impression, perhaps sadly, that he is looking at water.

1.1.1 Eikonal equation

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

$$\begin{aligned} ds &= [(dx)^2 + (dy)^2 + (dz)^2]^{1/2} \\ &= [1 + \dot{x}^2 + \dot{y}^2]^{1/2}\, dz = \frac{1}{\gamma}\, dz \,, \end{aligned} \qquad (1.1.1)$$

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

Fermat's principle can be interpreted as a ***principle of stationary action***, with z playing the role of time,

$$0 = \delta S := \delta \int_{z_A}^{z_B} L(x, y, \dot{x}, \dot{y}, z)\, dz \,. \qquad (1.1.2)$$

Here, the ***optical Lagrangian*** is

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

with

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

We may think of $(x, y, z) = (\mathbf{q}, z)$ where $\mathbf{q} = (x, y)$ is a vector with components in the plane perpendicular to the optical axis at displacement z. In this formulation, Fermat's principle implies the ***eikonal equation***, as follows:[1]

Theorem 1.1.1 *Fermat's principle*

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

for the optical Lagrangian

$$L(\mathbf{q}, \dot{\mathbf{q}}, z) = n(\mathbf{q}, z)[1 + |\dot{\mathbf{q}}|^2]^{1/2} =: \frac{n}{\gamma} \,, \qquad (1.1.3)$$

[1] The term ***eikonal*** (from the Greek $\epsilon\iota\kappa o\nu\alpha$ meaning image) was introduced into optics in [Br1895].

with

$$\gamma := \frac{dz}{ds} = \frac{1}{\sqrt{1+|\dot{\mathbf{q}}|^2}} \le 1\,, \tag{1.1.4}$$

implies the ***eikonal equation***

$$\gamma \frac{d}{dz}\Big(n(\mathbf{q},z)\,\gamma \frac{d\mathbf{q}}{dz}\Big) = \frac{\partial n}{\partial \mathbf{q}}\,, \quad \textit{with} \quad \frac{d}{ds} = \gamma \frac{d}{dz}\,. \tag{1.1.5}$$

Proof. Consider a family of C^2 curves $\mathbf{q}(z,\varepsilon) \in \mathbb{R}^2$ representing the possible ray paths from a point A to a point B in space. These paths satisfy

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

for a parameter ε in some bounded interval. Define the ***variation*** of the optical action (1.1.2) using this parameter as

$$\begin{aligned} \delta S &= \delta \int_{z_A}^{z_B} L(\mathbf{q}(z),\dot{\mathbf{q}}(z))dz \\ &:= \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \int_{z_A}^{z_B} L(\mathbf{q}(z,\varepsilon),\dot{\mathbf{q}}(z,\varepsilon))dz\,. \end{aligned} \tag{1.1.6}$$

Differentiating with respect to ε under the integral sign, denoting the ***variational derivative*** as

$$\delta \mathbf{q}(z) := \frac{d}{d\varepsilon}\Big|_{\varepsilon=0} \mathbf{q}(z,\varepsilon)\,, \tag{1.1.7}$$

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

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

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 is equivalent to the following equation, written in a standard form later made famous by Euler and Lagrange,

$$\frac{\partial L}{\partial \mathbf{q}} - \frac{d}{dz}\frac{\partial L}{\partial \dot{\mathbf{q}}} = 0 \qquad \textbf{\textit{(Euler-Lagrange equations).}} \qquad (1.1.8)$$

After a short algebraic manipulation using the explicit form of the optical Lagrangian in (1.1.3), the Euler-Lagrange equation (1.1.8) for the light rays yields the eikonal equation (1.1.5), 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 (1.1.5) follows from the Euler-Lagrange equations (1.1.8) with Lagrangian (1.1.3). ★

1.1.2 Huygens wave fronts

We may write the three-dimensional version of the Euler-Lagrange eikonal equation (1.1.5) for the ray path $\mathbf{r} \in \mathbb{R}^3$ as

$$\frac{d}{ds}\left(\frac{\partial L}{\partial(d\mathbf{r}/ds)}\right) = \frac{d}{ds}\left(n(\mathbf{r})\frac{d\mathbf{r}}{ds}\right) = \frac{\partial n}{\partial \mathbf{r}}\,. \qquad (1.1.9)$$

This is the ***3D eikonal equation*** for the vector function $\mathbf{r}(s)$ defining the ray path.

> **Exercise.** Verify that the three-dimensional Eikonal equation (1.1.9) follows from Fermat's principle in the form
>
> $$\begin{aligned} 0 &= \delta \int_A^B n(\mathbf{r}(s))\, ds \\ &= \delta \int_A^B n(\mathbf{r}(s)) \sqrt{\frac{d\mathbf{r}}{ds} \cdot \frac{d\mathbf{r}}{ds}}\, ds \,, \end{aligned} \qquad (1.1.10)$$
>
> with $ds^2 = d\mathbf{r}(s) \cdot d\mathbf{r}(s)$ for the arclength parameter s.
>
> ★

Answer. As in the calculation leading to the Euler-Lagrange equation in (1.1.8), one finds

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

where one denotes $\dot{\mathbf{r}} := d\mathbf{r}/ds$. The 3D eikonal equation (1.1.9) emerges, upon choosing the arclength variable $ds^2 = d\mathbf{r} \cdot d\mathbf{r}$, so that $|\dot{\mathbf{r}}| = 1$. (This means that $d|\dot{\mathbf{r}}|/ds = 0$.) ▲

> **Exercise.** Verify that the *same* three-dimensional Eikonal equation (1.1.9) *also* follows from Fermat's principle in the form
>
> $$0 = \delta S = \delta \int_A^B \frac{1}{2} n^2(\mathbf{r}(\tau)) \frac{d\mathbf{r}}{d\tau} \cdot \frac{d\mathbf{r}}{d\tau}\, d\tau \,, \qquad (1.1.11)$$
>
> with $d\tau = n ds$ for the arclength parameter s. ★

Answer. Denoting $\mathbf{r}'(\tau) = d\mathbf{r}/d\tau$ one computes as in deriving (1.1.8),

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

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

Remark 1.1.2 *The ray path* $\mathbf{r}(s)$ *in 3D is the orthogonal trajectory to the* ***Huygens wave front*** *defined by a level set of* ***Hamilton's characteristic function*** $S(\mathbf{r}) = constant$. *The geometric relationship between wave fronts and ray paths is illustrated in Figure 1.2.*

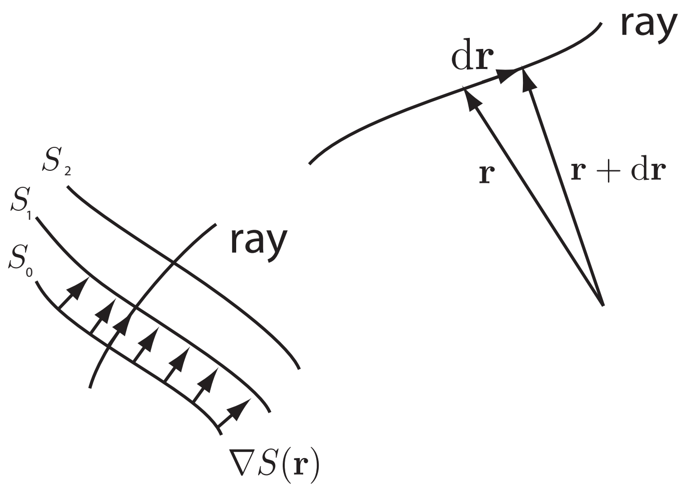

Figure 1.2: Huygens wave front and one of its corresponding ray paths. The wave fronts and ray paths form mutually orthogonal families of curves. The gradient $\nabla S(\mathbf{r})$ is normal to the wave front and tangent to the ray through it at the point $\mathbf{r}$.

Remark 1.1.3 *As Newton discovered in his famous prism experiment, the propagation of a Huygens wave front* $S(\mathbf{r}, \omega) =$ *constant 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, in what follows, we shall treat monochromatic light of fixed frequency* ω *and ignore 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. In his time, Fermat discovered the geometric foundations of ray optics. This is our focus.*

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

$$dS(\mathbf{r}) = \mathbf{n}(\mathbf{r}) \cdot d\mathbf{r} = n(\mathbf{r})\, ds\,. \tag{1.1.12}$$

Theorem 1.1.4 (Huygens-Fermat complementarity) *Fermat's eikonal equation (1.1.9) follows from the Huygens wavelet equation (1.1.12)*

$$\nabla S(\mathbf{r}) = n(\mathbf{r}) \frac{d\mathbf{r}}{ds} \qquad \textit{(Huygens equation)} \tag{1.1.13}$$

by differentiating along the ray path.

Corollary 1.1.5 *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 (1.1.13) is proved, because $\nabla S(\mathbf{r})$ is along the ray vector and is perpendicular to the level set of $S(\mathbf{r})$. ■

Proof. Theorem 1.1.4 is proved by applying the operation

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

to Huygens equation (1.1.13). This yields the eikonal equation (1.1.9), 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 order of derivatives. The third step uses the modulus of the Huygens equation (1.1.13) and invokes the property $|d\mathbf{r}/ds|^2 = 1$. ■

Corollary 1.1.6 *The modulus of the Huygens equation (1.1.13) yields*

$$|\nabla S|^2(\mathbf{r}) = n^2(\mathbf{r}) \qquad \textit{\textbf{(scalar eikonal equation)}} \qquad (1.1.14)$$

which follows because $d\mathbf{r}/ds = \hat{\mathbf{s}}$ in equation (1.1.9) is a unit vector.

This corollary 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. [BoWo1965], Chapter 3. See Keller [Ke1962] for the modern extension of geometric optics to include diffraction. The scalar eikonal equation (1.1.14) is also known as the ***Hamilton-Jacobi equation***.

Theorem 1.1.7 (Ibn Sahl-Snell law of refraction)
The gradient in Huygens equation (1.1.13) defines the ***ray vector****,*

$$\mathbf{n} = \nabla S = n(\mathbf{r})\hat{\mathbf{s}} \qquad (1.1.15)$$

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

$$\oint_P \nabla S(\mathbf{r}) \cdot d\mathbf{r} = \oint_P \mathbf{n}(\mathbf{r}) \cdot d\mathbf{r} = 0\,. \qquad (1.1.16)$$

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,*

$$(\mathbf{n} - \mathbf{n}') \times \hat{\mathbf{z}} = 0\,, \tag{1.1.17}$$

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 through the region enclosed by the integration loop. If θ and θ' are the angles of incidence and transmission, measured from the normal $\hat{\mathbf{z}}$ through the boundary, then preservation of the tangential components of the ray vector means that

$$n \sin\theta = n' \sin\theta'\,. \tag{1.1.18}$$

*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.*

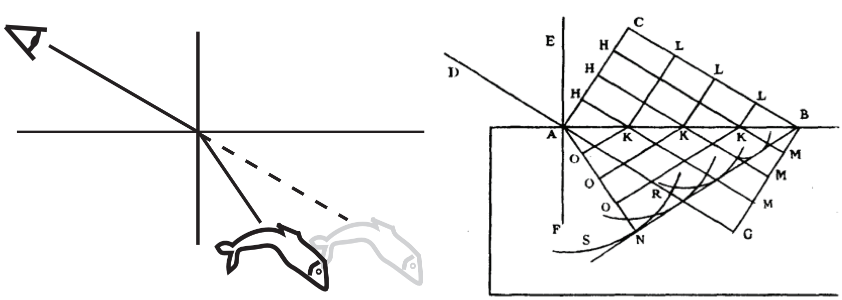

Figure 1.3: Ray tracing version (left) and Huygens version (right) of the Ibn Sahl-Snell law of refraction that $n \sin\theta = n' \sin\theta'$. This law is implied in ray optics by preservation of the components of the ray vector tangential to the interface. According to Huygens principle the law of refraction is implied by slower wave propagation in media of higher refractive index below the horizontal interface.

1.1.3 The eikonal equation for mirages

Air adjacent to a hot surface rises in temperature and becomes less dense. Thus over a flat hot surface, such as a desert expanse or a sun-baked roadway, air density locally increases with height and the 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 (1.1.5) 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.

In this geometry, the eikonal equation (1.1.5) 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

$$\frac{d^2x}{dz^2} \approx \kappa \quad \text{for} \quad \kappa x \ll 1 \quad \text{and} \quad \dot{x}^2 \ll 1\,. \tag{1.1.19}$$

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 1.4.

> **Exercise.** Explain how the ray pattern would differ from the rays shown in Figure 1.4 if the refractive index were *decreasing* with height x above the surface, rather than increasing. ★

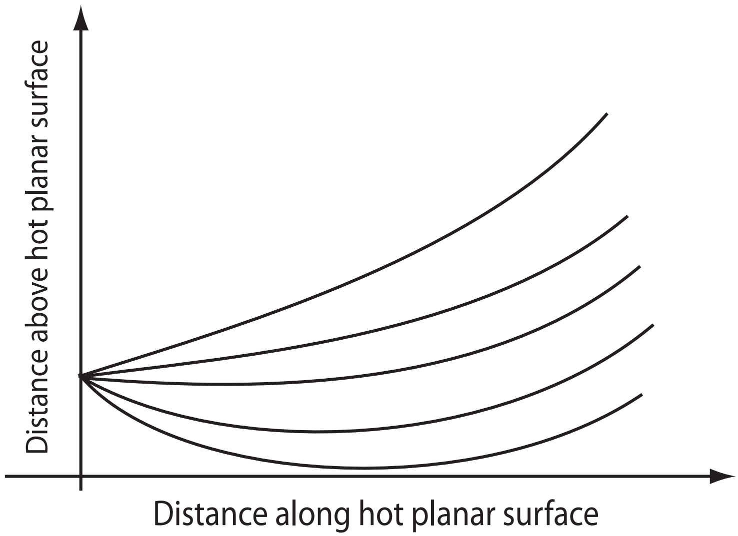

Figure 1.4: Ray trajectories are diverted in a spatially varying medium whose refractive index increases with height above a hot planar surface.

1.2 Hamiltonian formulation of ray optics

Definition 1.2.1 (Optical momentum)
The ***optical momentum*** *(denoted as* $\mathbf{p}$*) associated to 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 to be*

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}}\,, \quad \textit{with} \quad \dot{\mathbf{q}} := \frac{d\mathbf{q}}{dz}\,. \tag{1.2.1}$$

Remark 1.2.2 *For the optical Lagrangian (1.1.3), this momentum is found to be*

$$\mathbf{p} = \frac{\partial L}{\partial \dot{\mathbf{q}}} = n\gamma\, \dot{\mathbf{q}}\,, \quad \textit{which satisfies} \quad |\mathbf{p}|^2 = n^2(1-\gamma^2)\,. \tag{1.2.2}$$

Figure 1.5 illustrates the geometrical interpretation of this momentum for optical rays as the projection along the optical axis of the ray onto an image plane.

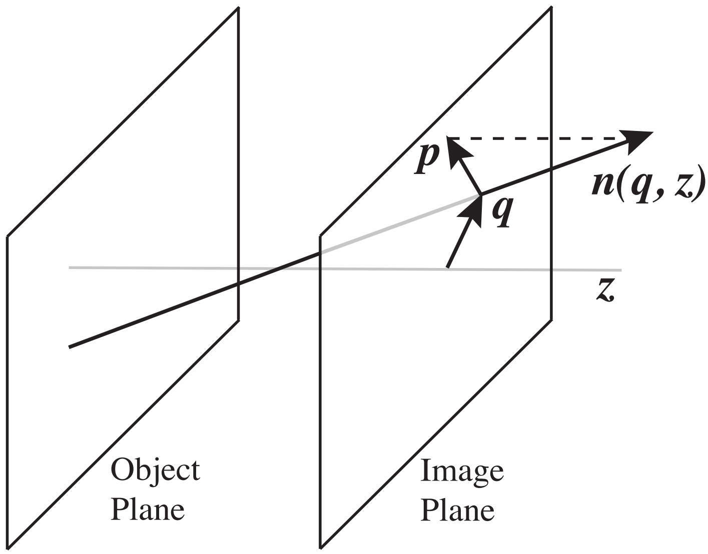

Figure 1.5: Geometrically, the momentum $\mathbf{p}$ associated to the coordinate $\mathbf{q}$ by equation (1.2.1) on the image plane at z turns out to be the projection onto the plane of the ray vector $\mathbf{n}(\mathbf{q}, z) = \nabla S = n(\mathbf{q}, z)d\mathbf{r}/ds$ passing through the point $\mathbf{q}(z)$. 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.

From the definition of optical momentum (1.2.2), the corresponding ***velocity*** $\dot{\mathbf{q}} = d\mathbf{q}/dz$ is found as a function of position and momentum $(\mathbf{q}, \mathbf{p})$ as

$$\dot{\mathbf{q}} = \frac{\mathbf{p}}{\sqrt{n^2(\mathbf{q}, z) - |\mathbf{p}|^2}}\,. \tag{1.2.3}$$

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

$$n^2 - |\mathbf{p}|^2 > 0\,. \tag{1.2.4}$$

The latter condition is explained geometrically, as follows.

1.2.1 Geometry, phase space and the ray path

Huygens' equation (1.1.13) summons the following geometric picture of the ray path, as shown in Figure 1.5. 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 the point $\mathbf{q}$. Its direction cosine with respect to the z-axis is given by

$$\cos\theta := \hat{\mathbf{z}} \cdot \frac{d\mathbf{r}}{ds} = \frac{dz}{ds} = \gamma\,. \tag{1.2.5}$$

This definition of $\cos\theta$ leads by (1.2.2) to

$$|\mathbf{p}| = n\sin\theta \quad \text{and} \quad \sqrt{n^2 - |\mathbf{p}|^2} = n\cos\theta\,. \tag{1.2.6}$$

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. In three-dimensional vector notation, this is expressed as

$$\mathbf{p}(z) = \mathbf{n}(\mathbf{q}, z) - \hat{\mathbf{z}}\Big(\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z)\Big)\,. \tag{1.2.7}$$

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.

Definition 1.2.3 (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. 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 1.2.4 *The phase space* $T^*\mathbb{R}^2$ *for ray optics is restricted to the disc,*

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

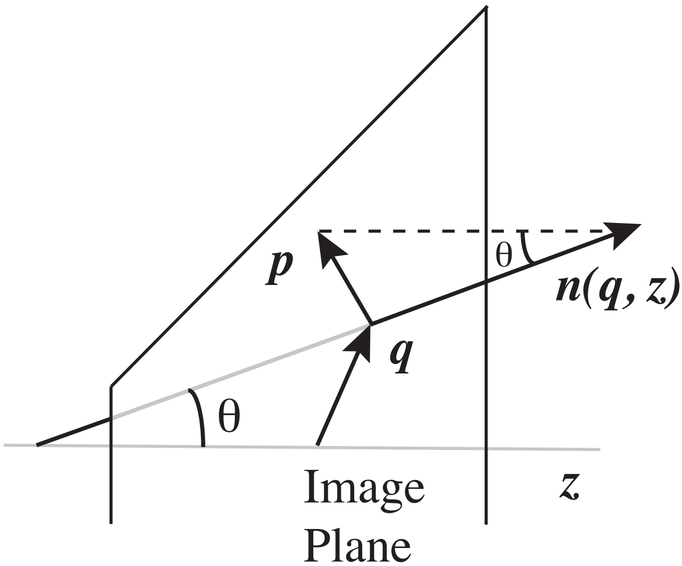

Figure 1.6: The canonical momentum $\mathbf{p}$ associated to the coordinate $\mathbf{q}$ by equation (1.2.1) on the image plane at z has magnitude $|\mathbf{p}| = n(\mathbf{q}, z)\sin\theta$, where $\cos\theta = dz/ds$ is the direction cosine of the ray with respect to the optical z-axis.

so that $\cos\theta$ *in (1.2.6) 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.*

1.2.2 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,

$$H(\mathbf{q}, \mathbf{p}) = \mathbf{p} \cdot \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}, z)\,. \tag{1.2.8}$$

For the Lagrangian (1.1.3) the Legendre transformation (1.2.8) leads to the following ***optical Hamiltonian***,

$$H(\mathbf{q},\mathbf{p}) = n\gamma\,|\dot{\mathbf{q}}|^2 - n/\gamma = -n\gamma = -\left[n(\mathbf{q},z)^2 - |\mathbf{p}|^2\right]^{1/2}, \quad (1.2.9)$$

upon using formula (1.2.3) 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 $\mathbf{p}(z)$. Thus, in the geometric picture of canonical screen optics in Figure 1.5, the component of the ray vector along the optical axis is (minus) the Hamiltonian. That is,

$$\hat{\mathbf{z}}\cdot\mathbf{n}(\mathbf{q},z) = n(\mathbf{q},z)\cos\theta = -H\,. \qquad (1.2.10)$$

Remark 1.2.5 *The optical Hamiltonian in (1.2.9) 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 1.2.6 *The phase space description of the ray path follows from* ***Hamilton's canonical equations****, which are defined as*

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}\,, \qquad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}\,. \qquad (1.2.11)$$

With the optical Hamiltonian $H(\mathbf{q},\mathbf{p}) = -\left[n(\mathbf{q},z)^2 - |\mathbf{p}|^2\right]^{1/2}$ in (1.2.9), these are

$$\dot{\mathbf{q}} = \frac{1}{H}\mathbf{p}\,, \qquad \dot{\mathbf{p}} = \frac{-1}{2H}\frac{\partial n^2}{\partial \mathbf{q}}\,. \qquad (1.2.12)$$

Proof. Hamilton's canonical equations are obtained by differentiating both sides of the Legendre transformation formula (1.2.8) 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. Vanishing of this coefficient is required for H to be independent of $\dot{\mathbf{q}}$. Identifying the other coefficients yields the relations

$$\frac{\partial H}{\partial \mathbf{p}} = \dot{\mathbf{q}}\,, \qquad \frac{\partial H}{\partial \mathbf{q}} = -\frac{\partial L}{\partial \mathbf{q}} = -\frac{d}{dz}\frac{\partial L}{\partial \dot{\mathbf{q}}} = -\,\dot{\mathbf{p}}\,, \tag{1.2.13}$$

and

$$\frac{\partial H}{\partial z} = -\,\frac{\partial L}{\partial z}\,, \tag{1.2.14}$$

in which ones uses the Euler-Lagrange equation to derive the second relation. Hence, one finds the canonical Hamiltonian formulas (1.2.11) in Theorem 1.2.6. ■

Definition 1.2.7 (Canonical momentum)
The momentum $\mathbf{p}$ *defined in (1.2.1) that appears with the position* $\mathbf{q}$ *in Hamilton's canonical equations (1.2.11) is called the* ***canonical momentum.***

Remark 1.2.8 (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 (1.1.9) 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 (1.1.10) and (1.1.11)? Do their Ibn Sahl-Snell laws differ? Do their Hamiltonian formulations differ? ★

Answer. The first stationary principle (1.1.10) gives $n(\mathbf{r}(s))d\mathbf{r}/ds$ for the optical momentum, while the second one (1.1.11) 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.

The two Hamiltonian formulations differ, because the Lagrangian in (1.1.10) is homogeneous of degree one in its velocity, while the Lagrangian in (1.1.11) is homogeneous of degree two. Consequently, under the Legendre transformation, the Hamiltonian in the first formulation vanishes identically, while the other Hamiltonian is quadratic in its momentum, namely, $H = |\mathbf{p}|^2/(2n)^2$. ▲

1.2.3 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 (1.2.9) 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 1.2.9 *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 (1.1.19) when $n = n_s(1 + \kappa x)$ for $\kappa > 0$.
>
> Explain what happens to the ray pattern when $\kappa < 0$.
>
> ★

1.3 Hamiltonian form of optical transmission

Proposition 1.3.1 (Canonical bracket)
Hamilton's canonical equations (1.2.11) arise from a bracket operation,

$$\big\{F,\,H\big\} = \frac{\partial F}{\partial \mathbf{q}} \cdot \frac{\partial H}{\partial \mathbf{p}} - \frac{\partial H}{\partial \mathbf{q}} \cdot \frac{\partial F}{\partial \mathbf{p}}\,, \tag{1.3.1}$$

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

Proof. One directly verifies,

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

■

Definition 1.3.2 (Canonically conjugate variables)
The components q_i *and* p_j *of position* $\mathbf{q}$ *and momentum* $\mathbf{p}$ *satisfy*

$$\big\{q_i,\,p_j\big\} = \delta_{ij}\,, \tag{1.3.2}$$

with respect to the canonical bracket operation (1.3.1). Variables that satisfy this relation are said to be ***canonically conjugate****.*

Definition 1.3.3 (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*

$$\dot{\mathbf{x}}(t) = \{\mathbf{x}, H\} \,, \quad for \quad H : M \to \mathbb{R} \,, \tag{1.3.3}$$

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,*

$$\{\cdot\,,\,\cdot\} : \mathcal{F}(M) \times \mathcal{F}(M) \to \mathcal{F}(M) \,, \tag{1.3.4}$$

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

Definition 1.3.4 (Poisson bracket)
A ***Poisson bracket operation*** *$\{\cdot\,,\,\cdot\}$ is defined as possessing the following properties:*

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

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

   for the product of any two functions F and G on M, and
4. *satisfies the* ***Jacobi identity***

   $$\{F\,,\,\{G\,,\,H\}\} + \{G\,,\,\{H\,,\,F\}\} + \{H\,,\,\{F\,,\,G\}\} = 0 \,, \tag{1.3.5}$$

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

Remark 1.3.5 *Definition 1.3.4 of Poisson bracket certainly includes the* ***canonical Poisson bracket*** *in (1.3.1) that produces Hamilton's* *canonical equations (1.2.11), in position* $\mathbf{q}$ *and conjugate momentum* $\mathbf{p}$*. However, this definition does not require the Poisson bracket to be expressed in the canonical form (1.3.1).*

> **Exercise.** Show that the defining properties of a Poisson bracket hold for the canonical bracket expression in (1.3.1). ★

> **Exercise.** Compute the Jacobi identity (1.3.5) using the canonical Poisson bracket (1.3.1) in one dimension for $F = p$, $G = q$ and H arbitrary. ★

> **Exercise.** What does the Jacobi identity (1.3.5) imply about $\{F\,,\,G\}$ when F and G are constants of motion, so that $\{F\,,\,H\} = 0$ and $\{G\,,\,H\}$ for a Hamiltonian H?
>
> ★

Definition 1.3.6 (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 are the **characteristic equations** of the Hamiltonian vector field on manifold M.*

Remark 1.3.7 (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 [Ha1830, Ha1837]. 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 [Ar1994].*

1.3.1 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 (1.2.10) 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 (1.1.5) simplifies via the canonical equations (1.2.12) to ***Newtonian dynamics***,

$$\ddot{\mathbf{q}} = -\,\frac{1}{2H^2}\,\frac{\partial n^2}{\partial \mathbf{q}} \quad \text{for} \quad \mathbf{q} \in \mathbb{R}^2\,. \qquad (1.3.6)$$

- 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 1.3.8 *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.*

1.3.2 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

$$n(\mathbf{q}, z) = n(r)\,, \quad \text{where} \quad r = |\mathbf{q}|\,. \tag{1.3.7}$$

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

$$|\mathbf{p}|^2 = p_r^2 + p_\phi^2/r^2\,. \tag{1.3.8}$$

Consequently, the optical Hamiltonian,

$$H = -\left[n(r)^2 - p_r^2 - p_\phi^2/r^2\right]^{1/2}, \tag{1.3.9}$$

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 (1.3.9) 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 $S = \int L\, dz$ an axisymmetric, translation-invariant material may be written in polar coordinates using the Lagrangian

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

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 (1.2.9),

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

then yields formula (1.3.9) for the optical Hamiltonian. ▲

Exercise. Interpret the quantity p_ϕ in terms of the vector image-screen 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

$$|\mathbf{p}\times\mathbf{q}|^2 = |\mathbf{p}|^2|\mathbf{q}|^2 - (\mathbf{p}\cdot\mathbf{q})^2 = p_\phi^2 \,. \tag{1.3.10}$$

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 ***area spanned on the optical screen*** by the vectors $\mathbf{q}$ and $\mathbf{p}$. ▲

1.3.3 Hamiltonian optics equations, polar coordinates

Hamilton's equations in polar coordinates are defined for axisymmetric, translation-invariant by the canonical Poisson brackets with the optical Hamiltonian (1.3.9),

$$\begin{aligned}
\dot{r} &= \{r,\, H\} = \frac{\partial H}{\partial p_r} = -\,\frac{p_r}{H}\,,\\
\dot{p}_r &= \{p_r,\, H\} = -\,\frac{\partial H}{\partial r} = -\,\frac{1}{2H}\frac{\partial}{\partial r}\left(n^2(r) - \frac{p_\phi^2}{r^2}\right),\\
\dot{\phi} &= \{\phi,\, H\} = \frac{\partial H}{\partial p_\phi} = -\,\frac{p_\phi}{Hr^2}\,,\\
\dot{p}_\phi &= \{p_\phi,\, H\} = -\,\frac{\partial H}{\partial \phi} = 0\,.
\end{aligned} \tag{1.3.11}$$

In the reduced phase space dynamics for $r(z)$, $p_r(z)$, the constants of the motion p_ϕ, H, may be regarded as parameters that are set by the initial conditions. The dynamics of the azimuthal angle, or phase, $\phi(z)$ in polar coordinates decouples from the rest and may be found separately, after solving for $r(z)$ and $p_r(z)$.

Remark 1.3.9 (Evolution of azimuthal angle)
*The polar canonical equation for $\phi(z)$ in (1.3.11) implies, for a given orbit $r(z)$, that the phase may be obtained as a **quadrature**,*

$$\Delta\phi(z) = \int^z \frac{\partial H}{\partial p_\phi}\,dz = -\,\frac{p_\phi}{H}\int^z \frac{1}{r^2(z)}\,dz\,, \tag{1.3.12}$$

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 (1.3.11) 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.

1.3.4 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:

$$\oint p_\phi \, d\phi = p_\phi \Delta\phi = \underbrace{-\oint p_r \, dr}_{\text{Geometric}} + \underbrace{\oint \mathbf{p} \cdot d\mathbf{q}}_{\text{Dynamic}} . \tag{1.3.13}$$

On writing this decomposition of the phase as

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

one sees that

$$p_\phi \Delta\phi_{geom} = \frac{1}{H} \oint p_r^2 dz = - \iint dp_r \wedge dr \tag{1.3.14}$$

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

$$\begin{aligned} p_\phi \Delta\phi_{dyn} &= \oint \mathbf{p} \cdot d\mathbf{q} \\ &= \oint \left(p_r \frac{\partial H}{\partial p_r} + p_\phi \frac{\partial H}{\partial p_\phi} \right) dz \\ &= \frac{-1}{H} \oint \left(p_r^2 + \frac{p_\phi^2}{r^2} \right) dz \\ &= \frac{1}{H} \oint \left(H^2 - n^2(|\mathbf{q}(z)|) \right) dz \\ &= ZH - \frac{Z}{H} \langle n^2 \rangle \,, \end{aligned} \tag{1.3.15}$$

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.

1.3.5 Skewness

Definition 1.3.10 *The quantity*

$$p_\phi = \mathbf{p} \times \mathbf{q} = y p_x - x p_y \,, \tag{1.3.16}$$

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

Remark 1.3.11 *By (1.3.11) the skewness is conserved for rays in axisymmetric media.*

Remark 1.3.12 *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* $\mathbf{q}$*. 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*

[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.

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.*

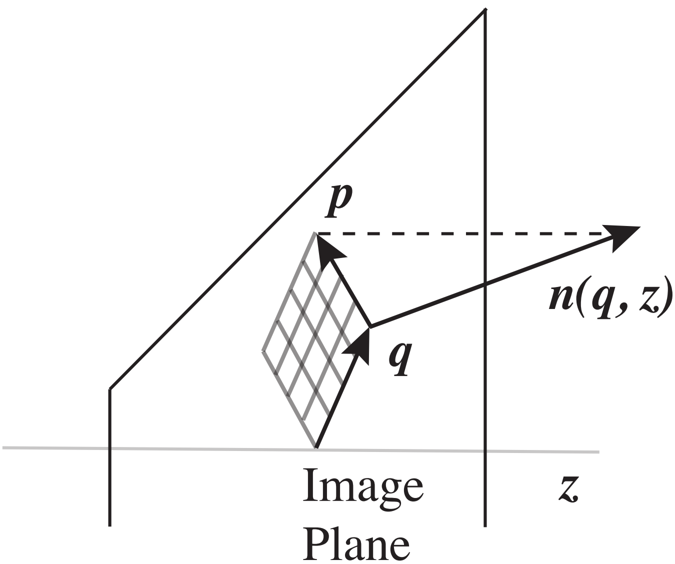

Figure 1.7: The skewness $S = \mathbf{p} \times \mathbf{q}$ of a ray $\mathbf{n}(\mathbf{q}, z)$ is an oriented area in an image plane. For axisymmetric media, skewness is preserved as a function of distance z along the optical axis. The projection $\hat{\mathbf{z}} \cdot \mathbf{n}(\mathbf{q}, z)$ is 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 (1.3.11) 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 (1.2.12) with the optical Hamiltonian (1.2.9)

conserve the skewness function $p_\phi = \mathbf{p} \times \mathbf{q}$ when the refractive index satisfies (1.3.7). ★

Remark 1.3.13 *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****.*

Exercise.
(Phase plane reduction)
(1) Solve Hamilton's canonical equations for axisymmetric, translation invariant media in the case of an optical fibre with radially varying index of refraction in the following form:

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

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?

(2) 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.

(3) 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. ★

1.3.6 Lagrange invariant: Poisson bracket relations

Under the canonical Poisson bracket (1.3.1), the skewness function, or Lagrange invariant,

$$S = p_\phi = \hat{\mathbf{z}} \cdot \mathbf{p} \times \mathbf{q} = yp_x - xp_y \,, \tag{1.3.17}$$

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,

$$\frac{d}{d\phi} = X_S = \Big\{ \,\cdot\,,\, S \Big\} = \mathbf{q} \times \hat{\mathbf{z}} \cdot \frac{\partial}{\partial \mathbf{q}} + \mathbf{p} \times \hat{\mathbf{z}} \cdot \frac{\partial}{\partial \mathbf{p}} \,, \tag{1.3.18}$$

has the solution,

$$\begin{pmatrix} \mathbf{q}(\phi) \\ \mathbf{p}(\phi) \end{pmatrix} = \begin{pmatrix} R_z(\phi) & 0 \\ 0 & R_z(\phi) \end{pmatrix} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} . \tag{1.3.19}$$

Here the matrix

$$R_z(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{1.3.20}$$

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

Remark 1.3.14 *The application of the Hamiltonian vector field X_S for skewness in (1.3.18) to the position vector $\mathbf{q}$ yields*

$$X_S \mathbf{q} = -\,\hat{\mathbf{z}} \times \mathbf{q} \,. \tag{1.3.21}$$

Likewise, the application of the Hamiltonian vector field X_S for skewness in (1.3.18) to the momentum vector yields

$$X_S \mathbf{p} = -\,\hat{\mathbf{z}} \times \mathbf{p} \,. \tag{1.3.22}$$

Thus, the application of X_S to the vectors $\mathbf{p}$ and $\mathbf{q}$ rotates them both by the same angle.

Definition 1.3.15 (Diagonal action & cotangent lift)
Together, formulas (1.3.21) - (1.3.22) 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, (1.3.22) is the lift of the action of rotation (1.3.21) from position vectors to momentum vectors.*

Remark 1.3.16 (Moment of momentum)
Applying the Hamiltonian vector field X_S *for skewness in (1.3.18) to screen coordinates* $\mathbf{q} = \mathbb{R}^2$ *produces the infinitesimal action of rotations about* $\hat{\mathbf{z}}$*, as*

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

The skewness function S *in (1.3.17) may be expressed in terms of* ***two different pairings****,*

$$\begin{aligned} S &= \left\langle\!\!\left\langle \mathbf{p},\, X_S\mathbf{q} \right\rangle\!\!\right\rangle = \mathbf{p}\cdot(-\,\hat{\mathbf{z}}\times\mathbf{q}) \quad \textit{and} \\ S &= (\mathbf{p}\times\mathbf{q})\cdot\hat{\mathbf{z}} = \left\langle \mathbf{J}(\mathbf{p},\,\mathbf{q}),\, \hat{\mathbf{z}} \right\rangle = J^z(\mathbf{p},\,\mathbf{q}) \,. \end{aligned} \tag{1.3.23}$$

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

$$\begin{aligned} &\left\langle\!\!\left\langle \,\cdot\,,\,\cdot\, \right\rangle\!\!\right\rangle :\ (\textit{momentum}) \times (\textit{velocity}) \to \mathbb{R}\,, \qquad (1.3.24) \\ &\left\langle \,\cdot\,,\,\cdot\, \right\rangle :\ (\textit{moment of momentum}) \times (\textit{rotation rate}) \to \mathbb{R}\,. \end{aligned}$$

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 (1.3.21). 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* $\mathbb{R}^3$*.*

Definition 1.3.17 *Distinguishing between the pairings in (1.3.23) 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 1.3.18 (Momentum map for cotangent lift) *Formula (1.3.23) defines the momentum map for the* ***cotangent lift*** *from position vectors to their canonically conjugate momentum vectors in phase space of the action of rotations about* $\hat{\mathbf{z}}$*. 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,* $\mathcal{M}$*,*

$$\mathbf{J} : T^*\mathbb{R}^2 \to \mathcal{M}\,, \quad \textit{namely,} \quad \mathbf{J}(\mathbf{p},\,\mathbf{q}) = \mathbf{p}\times\mathbf{q}\,, \tag{1.3.25}$$

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 (1.3.23). 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*

$$-\hat{\mathbf{z}}\times P = X_{\mathbf{J}\cdot\hat{\mathbf{z}}}P = \{\,P\,,\,\mathbf{J}\cdot\hat{\mathbf{z}}\}\,. \tag{1.3.26}$$

Remark 1.3.19 *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*

$$\{S^2, H\} = 0\,, \tag{1.3.27}$$

for optical Hamiltonians of the form,

$$H = -\left[n(|\mathbf{q}|^2)^2 - |\mathbf{p}|^2\right]^{1/2}. \tag{1.3.28}$$

The Poisson bracket (1.3.27) 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 (1.3.18).

1.4 Axisymmetric invariant coordinates

Transforming to axisymmetric coordinates and 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

$$\pi : \mathbb{R}^2\{\mathbf{0}\} \to \mathbb{R}\backslash\{\mathbf{0}\} : (x, y) \to r\,. \tag{1.4.1}$$

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 1.3.3, 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 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* as

$$T^*\mathbb{R}^2\backslash\{\mathbf{0}\} \simeq \big(T^*(\mathbb{R}\backslash\{0\}) \times \mathbb{R}\big) \times S^1 \,. \tag{1.4.2}$$

The corresponding ***quotient map for azimuthal symmetry*** is

$$\pi : T^*\mathbb{R}^2\backslash\{\mathbf{0}\} \rightarrow T^*(\mathbb{R}\backslash\{0\}) \times \mathbb{R} : (\mathbf{q}, \mathbf{p}) \rightarrow (r, p_r; p_\phi) \,. \tag{1.4.3}$$

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

$$\pi : T^*\mathbb{R}^2\backslash\{0\} \rightarrow \mathbb{R}^3\backslash\{\mathbf{0}\} : (\mathbf{q}, \mathbf{p}) \rightarrow \mathbf{X} = (X_1, X_2, X_3) \,, \tag{1.4.4}$$

given explicitly by the quadratic monomials,

$$X_1 = |\mathbf{q}|^2 \geq 0 \,, \quad X_2 = |\mathbf{p}|^2 \geq 0 \,, \quad X_3 = \mathbf{p} \cdot \mathbf{q} \,. \tag{1.4.5}$$

The quotient map (1.4.4) is written a bit more succinctly as

$$\pi\,(\mathbf{p}, \mathbf{q}) = \mathbf{X} \,. \tag{1.4.6}$$

Theorem 1.4.2 *The vector (X_1, X_2, X_3) of quadratic monomials in phase space all Poisson-commute with skewness S,*

$$\{S, X_1\} = 0 \,, \quad \{S, X_2\} = 0 \,, \quad \{S, X_3\} = 0 \,. \tag{1.4.7}$$

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. ■

[3]The transformation $T^*\mathbb{R}^2 \rightarrow \mathbb{R}^3$ in (1.4.5) will be recognised later as another example of a ***momentum map***.

Remark 1.4.3 *The orbits of S in (1.3.19) 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 (1.4.4) 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 1.4.4 (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 (1.4.4) is the ***orbit manifold*** *for axisymmetric ray optics.*

Remark 1.4.5 (Orbit manifold for axisymmetric ray optics)
The image of the quotient map π in (1.4.4) may be conveniently displayed as the zero-level set of the the relation

$$C(X_1,\,X_2,\,X_3,\,S)=S^2-(X_1X_2-X_3^2)=0\,, \qquad (1.4.8)$$

among the axisymmetric variables in equation (1.4.5). 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.

1.5 Geometry of invariant coordinates

In terms of the axially invariant coordinates (1.4.5), the Petzval invariant and the square of the optical Hamiltonian satisfy

$$|\mathbf{p}\times\mathbf{q}|^2=|\mathbf{p}|^2|\mathbf{q}|^2-(\mathbf{p}\cdot\mathbf{q})^2 \quad\text{and}\quad H^2=n^2(|\mathbf{q}|^2)-|\mathbf{p}|^2\geq 0\,. \quad (1.5.1)$$

That is,

$$S^2=X_1X_2-X_3^2\geq 0 \quad\text{and}\quad H^2=n^2(X_1)-X_2\geq 0\,. \qquad (1.5.2)$$

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 $X_3 = 0$. 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.

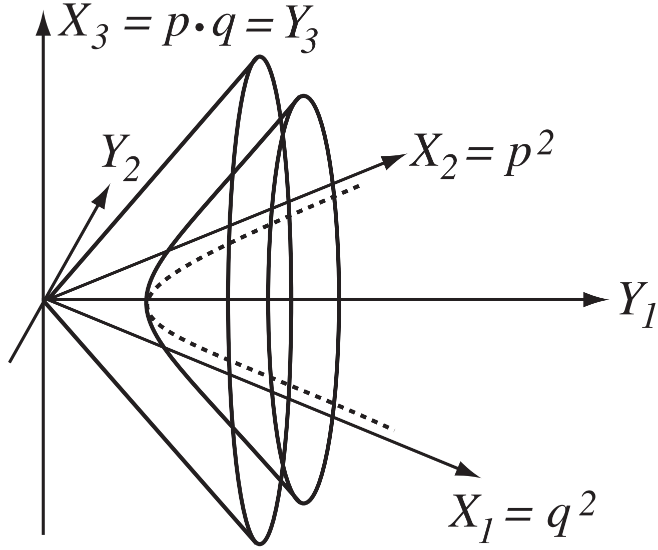

Figure 1.8: Level sets of the Petzval invariant $S^2 = X_1X_2 - X_3^2$ are hyperboloids of revolution around the $X_1 = X_2$ axis (along Y_1) in the horizontal plane, $X_3 = 0$. Level sets of the Hamiltonian H in (1.5.2) are independent of the vertical coordinate. The axisymmetric invariants $\mathbf{X} \in \mathbb{R}^3$ 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 (1.3.19). 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.

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.

Question 1.5.1
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 1.5.2
Answering this question introduces the concept of ***flows of Hamiltonian vector fields****.*

1.5.1 Flows of Hamiltonian vector fields

Theorem 1.5.3 (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 1.3.6 from the Poisson brackets as derivatives,*

$$\frac{d}{d\tau_1} := \{\,\cdot\,,\, X_1\}\,, \quad \frac{d}{d\tau_2} := \{\,\cdot\,,\, X_2\} \quad \textit{and} \quad \frac{d}{d\tau_3} := \{\,\cdot\,,\, X_3\}\,, \tag{1.5.3}$$

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***,

$$\frac{d}{d\omega} = \Big\{\,\cdot\,,\, \frac{1}{2}(X_1 + X_2)\Big\} = \mathbf{p}\cdot\frac{\partial}{\partial\mathbf{q}} - \mathbf{q}\cdot\frac{\partial}{\partial\mathbf{p}}\,. \tag{1.5.4}$$

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,

$$\frac{d}{d\omega}(\mathbf{q}+i\mathbf{p}) = -i(\mathbf{q}+i\mathbf{p}), \tag{1.5.5}$$

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

$$\begin{pmatrix}\mathbf{q}(\omega)\\ \mathbf{p}(\omega)\end{pmatrix} = \begin{pmatrix}\cos\omega & \sin\omega\\ -\sin\omega & \cos\omega\end{pmatrix}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix}, \tag{1.5.6}$$

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

$$\begin{aligned} e^{\omega(m_1+m_2)/2} &= \sum_{n=0}^{\infty}\frac{\omega^n}{n!}\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}^n \\ &= \begin{pmatrix}\cos\omega & \sin\omega\\ -\sin\omega & \cos\omega\end{pmatrix}. \end{aligned} \tag{1.5.7}$$

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

$$\begin{pmatrix}\mathbf{q}(\gamma)\\ \mathbf{p}(\gamma)\end{pmatrix}=\begin{pmatrix}\cosh\gamma&\sinh\gamma\\ \sinh\gamma&\cosh\gamma\end{pmatrix}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix}, \qquad (1.5.8)$$

which, in turn, sums the exponential series

$$\begin{aligned} e^{\gamma(m_2-m_1)/2} &= \sum_{n=0}^{\infty}\frac{\gamma^n}{n!}\begin{pmatrix}0&1\\ 1&0\end{pmatrix}^n \\ &= \begin{pmatrix}\cosh\gamma&\sinh\gamma\\ \sinh\gamma&\cosh\gamma\end{pmatrix}. \end{aligned} \qquad (1.5.9)$$

- 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:

$$\frac{d}{d\tau_1}=\Big\{\,\cdot\,,\,X_1\Big\}=-2\mathbf{q}\cdot\frac{\partial}{\partial\mathbf{p}}\,. \qquad (1.5.10)$$

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

$$\begin{aligned} \begin{pmatrix}\mathbf{q}(\tau_1)\\ \mathbf{p}(\tau_1)\end{pmatrix} &= e^{\tau_1 m_1}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix} \\ &= \begin{pmatrix}1&0\\ -2\tau_1&1\end{pmatrix}\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix} \\ &=: M_1(\tau_1)\begin{pmatrix}\mathbf{q}(0)\\ \mathbf{p}(0)\end{pmatrix}. \end{aligned} \qquad (1.5.11)$$

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

$$e^{\tau_1 m_1} = \sum_{n=0}^{\infty} \frac{1}{n!} (\tau_1 m_1)^n \tag{1.5.12}$$

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,

$$\frac{d}{d\tau_2} = \Big\{ \cdot \,, \, X_2 \Big\} = 2\mathbf{p} \cdot \frac{\partial}{\partial \mathbf{q}} \,. \tag{1.5.13}$$

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

$$\begin{aligned} \begin{pmatrix} \mathbf{q}(\tau_2) \\ \mathbf{p}(\tau_2) \end{pmatrix} &= e^{\tau_2 M_2} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2\tau_2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &=: M_2(\tau_2) \begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} , \end{aligned} \tag{1.5.14}$$

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,

$$\begin{aligned} \begin{pmatrix} \mathbf{q}(\tau_3) \\ \mathbf{p}(\tau_3) \end{pmatrix} &= e^{\tau_3 m_3}\begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &= \begin{pmatrix} e^{\tau_3} & 0 \\ 0 & e^{-\tau_3} \end{pmatrix}\begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} \\ &=: M_3(\tau_3)\begin{pmatrix} \mathbf{q}(0) \\ \mathbf{p}(0) \end{pmatrix} , \end{aligned} \tag{1.5.15}$$

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

The proof of Theorem 1.5.3 is now finished. ■

1.6 Symplectic matrices

Remark 1.6.1 (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. 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 (1.5.11), (1.5.14) and (1.5.15). The latter are 2×2 ***symplectic matrices****. That is, these three matrices each satisfy*

$$M_i(\tau_i) J M_i(\tau_i)^T = J \quad \textit{(no sum on } i = 1,2,3\textit{)}, \tag{1.6.1}$$

where

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} . \tag{1.6.2}$$

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 1.6.2 (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 1.6.3 (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 (1.6.1) required to be symplectic. Prove that these matrices form a group under matrix multiplication. Conclude that they form a three-parameter Lie group. ★

Theorem 1.6.4 (Fundamental theorem of planar optics)
Any plane 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

$$M = \begin{pmatrix} \cos\omega & \sin\omega \\ -\sin\omega & \cos\omega \end{pmatrix} \begin{pmatrix} e^{\tau_3} & 0 \\ 0 & e^{-\tau_3} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2\tau_1 & 1 \end{pmatrix}. \tag{1.6.3}$$

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

$$Sp(2,\mathbb{R}) = \mathsf{KAN}\,. \tag{1.6.4}$$

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 does not affect the image at all, since* $\mathbf{q}(\tau_1) = \mathbf{q}(0)$ *from equation (1.5.11). However, the rightmost factor does change 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 the product* $\mathbf{q}\cdot\mathbf{p}$ *remains the invariant as in equation (1.5.15). The leftmost factor (the maximal compact subgroup* K*) is a type of Fourier transform in angle* $\omega \in S^1$ *on a circle as in equation (1.5.6).*

For insightful discussions and references to the literature in 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 1.6.5 (Hamiltonian matrices)
The traceless constant matrices

$$m_1 = \begin{pmatrix} 0 & 0 \\ -2 & 0 \end{pmatrix}, \; m_2 = \begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix}, \; m_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{1.6.5}$$

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

$$e^{\tau_1 m_1} = M_1(\tau_1)\,, \quad e^{\tau_2 m_2} = M_2(\tau_2)\,, \quad e^{\tau_3 m_3} = M_3(\tau_3)\,, \tag{1.6.6}$$

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

$$\begin{aligned} m_1 &= \Big[M\,'_1(\tau_1) M_1^{-1}(\tau_1) \Big]_{\tau_1=0}\,, \\ m_2 &= \Big[M\,'_2(\tau_2) M_2^{-1}(\tau_2) \Big]_{\tau_2=0}\,, \\ m_3 &= \Big[M\,'_3(\tau_3) M_3^{-1}(\tau_3) \Big]_{\tau_3=0}\,, \end{aligned} \tag{1.6.7}$$

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

Remark 1.6.6

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

$$m_i J + J m_i = 0\,, \quad \textit{where} \quad J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}. \tag{1.6.8}$$

> ***Exercise.*** *Take the derivative of the definition of symplectic matrices (1.6.1) to prove statement (1.6.8) about Hamiltonian matrices.* ★

- *The respective actions of the symplectic matrices* $M_1(\tau_1)$, $M_2(\tau_2)$, $M_3(\tau_3)$ *in (1.6.6) 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 (1.4.5).*

- *The quadratic Hamiltonian,*

$$\begin{aligned} H &= \frac{\omega}{2}(X_1 + X_2) + \frac{\gamma}{2}(X_2 - X_1) + \tau X_3 \qquad (1.6.9)\\ &= \frac{\omega}{2}(|\mathbf{p}|^2 + |\mathbf{q}|^2) + \frac{\gamma}{2}(|\mathbf{p}|^2 - |\mathbf{q}|^2) + \tau \mathbf{q}\cdot\mathbf{p}\,, \end{aligned}$$

is associated to the Hamiltonian matrix,

$$\begin{aligned} m(\omega,\gamma,\tau) &= \frac{\omega}{2}(m_1 + m_2) + \frac{\gamma}{2}(m_2 - m_1) + \tau m_3 \\ &= \begin{pmatrix} \tau & \gamma+\omega \\ \gamma-\omega & -\tau \end{pmatrix}. \qquad (1.6.10) \end{aligned}$$

The eigenvalues of the Hamiltonian matrix (1.6.10) are determined from

$$\lambda^2 + \Delta = 0\,, \quad \textit{with} \quad \Delta = \det m = \omega^2 - \gamma^2 - \tau^2\,. \tag{1.6.11}$$

Consequently, the eigenvalues come in pairs, given by

$$\lambda^{\pm} = \pm\sqrt{-\Delta} = \pm\sqrt{\tau^2 + \gamma^2 - \omega^2}\,. \tag{1.6.12}$$

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 1.9 and summarised in the table below, cf. [Wo2004]. The action of a symplectic matrix* $M(\tau_i)$ *on a Hamiltonian matrix* $m(\omega, \gamma, \tau)$ *by matrix conjugation* $m \to m' = M(\tau_i) m M^{-1}(\tau_i)$ *(no sum on* $i = 1, 2, 3$*) may alter the values of* (ω, γ, τ)*. However, this action preserves eigenvalues, so it preserves the value of the determinant* Δ*.*

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
$\Delta = -1\,, \quad \lambda^{\pm} = \pm\, 1$	$m_H = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$

Remark 1.6.7 (Prelude to Lie algebras)

- *In terms of the Hamiltonian matrices the KAN decomposition (1.6.3) may be written as*

$$M = e^{\omega(m_1+m_2)/2} e^{\tau_3 m_3} e^{\tau_1 m_1}\,. \tag{1.6.13}$$

- *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 [Ol2000].

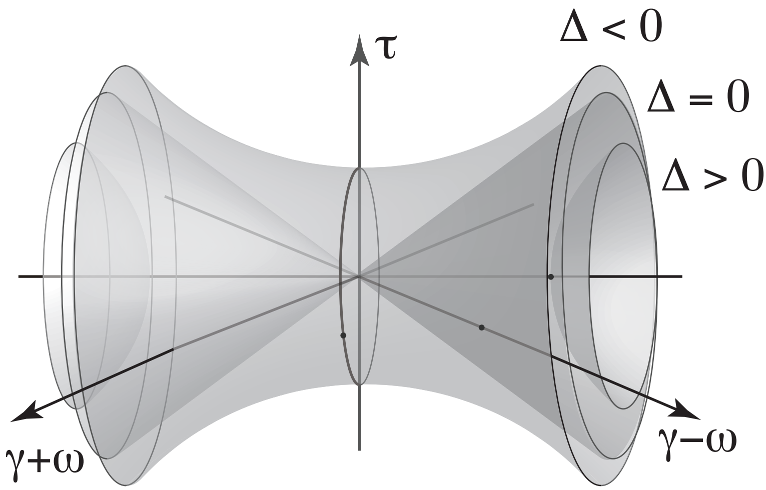

Figure 1.9: The flows corresponding to exponentiation of the Hamiltonian matrices with parameters $(\omega, \gamma, \tau) \in \mathbb{R}^3$ are divided into three families of orbits defined by the sign of the discriminant $\Delta = \omega^2 - \gamma^2 - \tau^2$. These three families of orbits are hyperbolic ($\Delta < 0$), parabolic ($\Delta = 0$) and elliptic ($\Delta > 0$). The action by matrix conjugation of a symplectic matrix on a Hamiltonian matrix changes the parameters $(\omega, \gamma, \tau) \in \mathbb{R}^3$, while preserving the value of the determinant Δ.

1.7 Lie algebras

1.7.1 Definitions

Definition 1.7.1 *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:*

(a) 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* $(\mathbf{u}, \mathbf{v}, \mathbf{w}) \in \mathfrak{g}$*;*

(b) Skew symmetry

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

(c) 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* $\mathfrak{g}$.

1.7.2 Structure constants

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

$$[\,\mathbf{e}_i\,,\,\mathbf{e}_j\,]=c_{ij}^k\mathbf{e}_k\,. \tag{1.7.1}$$

Since the $\{\mathbf{e}_1,\,\dots,\,\mathbf{e}_r\}$ form a vector basis, the structure constants in (1.7.1) 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:

(i) Skew symmetry

$$c_{ji}^k=-\,c_{ij}^k\,, \tag{1.7.2}$$

and

(ii) Jacobi identity

$$c_{ij}^kc_{lk}^m+c_{li}^kc_{jk}^m+c_{jl}^kc_{ik}^m=0\,. \tag{1.7.3}$$

Conversely, any set of constants c_{ij}^k that satisfy relations (1.7.2) and (1.7.3) defines a Lie algebra $\mathfrak{g}$.

> **Exercise.** Prove that the Jacobi identity requires the relation (1.7.3). ★

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

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

Summing over the three cyclic permutations of (l,i,j) of this expression yields the required relation (1.7.3) among the structure constants for the Jacobi identity to hold.

▲

1.7.3 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^k_{ij} 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 (1.6.7) is given by

$$[m_i,\, m_j] = c^k_{ij} m_k = \begin{array}{|c|ccc|} \hline [\,\cdot\,,\,\cdot\,] & m_1 & m_2 & m_3 \\ \hline m_1 & 0 & 4m_3 & 2m_1 \\ m_2 & -4m_3 & 0 & -2m_2 \\ m_3 & -2m_1 & 2m_2 & 0 \\ \hline \end{array} \tag{1.7.4}$$

The structure constants are immediately read off the table as

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

and all the other c^k_{ij}'s vanish.

Proposition 1.7.2 (Structure constants for $sp(2, \mathbb{R})$)
The commutation relations in (1.6.7) for the 2×2 Hamiltonian matrices define the structure constants for the ***symplectic Lie algebra*** *$sp(2, \mathbb{R})$.*

Proof. The exponentiation of the Hamiltonian matrices was shown in Theorem 1.5.3 of Section 1.5.1 to produce the symplectic Lie group, $Sp(2, \mathbb{R})$. Likewise, the tangent space at the identity of the symplectic Lie group $Sp(2, \mathbb{R})$ is the symplectic Lie algebra, $sp(2, \mathbb{R})$, a vector space whose basis may be chosen as the 2×2 Hamiltonian matrices. Thus, the commutation relations among these matrices yield the structure constants for $sp(2, \mathbb{R})$ in this basis. ∎

1.7.4 Poisson brackets among axisymmetric variables

Theorem 1.7.3 *The canonical Poisson brackets among the axisymmetric variables X_1, X_2 and X_3 in (1.4.5) 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

$$\{X_i,\,X_j\} = \begin{array}{|c|ccc|}\hline \{\,\cdot\,,\,\cdot\,\} & X_1 & X_2 & X_3 \\ \hline X_1 & 0 & 4X_3 & 2X_1 \\ X_2 & -4X_3 & 0 & -2X_2 \\ X_3 & -2X_1 & 2X_2 & 0 \\ \hline \end{array} \tag{1.7.5}$$

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

$$\{X_i,\,X_j\} = \frac{\partial X_i}{\partial z_A}\{z_A,\,z_B\}\frac{\partial X_j}{\partial z_A}\,, \tag{1.7.6}$$

for the invariant quadratic monomials $X_i(z_A)$ in (1.4.5). Here one denotes $z_A = (q_A, p_A)$, with $A = 1, 2, 3$. ■

Remark 1.7.4 *The closure in the Poisson commutator table (1.7.5) among the set of axisymmetric phase space functions $(X_1,\,X_2,\,X_3)$ is possible, because these functions are all quadratic monomials in the canonical variables. That is, the canonical Poisson bracket preserves the class of quadratic monomials in phase space.*

Summary 1.7.5 *The 2×2 traceless matrices m_1, m_2 and m_3 in equation (1.6.7) provide a matrix commutator representation of the Poisson bracket relations in equation (1.7.5) for the quadratic monomials in phase space, X_1, X_2 and X_3, from which the matrices m_1, m_2 and m_3 were derived. Likewise, the 2×2 symplectic matrices $M_1(\tau_1)$, $M_2(\tau_2)$ and $M_3(\tau_3)$ provide a matrix representation of the transformations of the phase space vector $(\mathbf{q},\mathbf{p})^T$. These transformations are generated by integrating the characteristic equations of the Hamiltonian vector fields $\{\,\cdot\,,\,X_1\}$, $\{\,\cdot\,,\,X_2\}$ and $\{\,\cdot\,,\,X_3\}$.*

1.7.5 Non-canonical $\mathbb{R}^3$ Poisson bracket for ray optics

The canonical Poisson bracket relations in (1.7.5) may be used to transform to another Poisson bracket expressed solely in terms of the variables $\mathbf{X} = (X_1, X_2, X_3) \in \mathbb{R}^3$ by using the chain rule again,

$$\frac{dF}{dt} = \{F, H\} = \frac{\partial F}{\partial X_i}\{X_i,\, X_j\}\frac{\partial H}{\partial X_j}\,. \tag{1.7.7}$$

Here, the quantities $\{X_i,\, X_j\}$, with $i, j = 1, 2, 3$, are obtained from Poisson commutator table in (1.7.5).

This chain rule calculation reveals that the Poisson bracket in the $\mathbb{R}^3$ variables (X_1, X_2, X_3) repeats the commutator table $[m_i,\, m_j] = c_{ij}^k m_k$ for the Lie algebra $sp(2, \mathbb{R})$ of Hamiltonian matrices in (1.7.4). Consequently, we may write this Poisson bracket equivalently as

$$\{F, H\} = X_k\, c_{ij}^k\, \frac{\partial F}{\partial X_i}\frac{\partial H}{\partial X_j}\,. \tag{1.7.8}$$

In particular, the Poisson bracket between two of these quadratic-monomial invariants is a linear function of them

$$\{X_i,\, X_j\} = c_{ij}^k X_k\,, \tag{1.7.9}$$

and we also have

$$\{X_l,\, \{X_i,\, X_j\}\} = c_{ij}^k\{X_l,\, X_k\} = c_{ij}^k c_{lk}^m X_m\,. \tag{1.7.10}$$

Hence, the Jacobi identity is satisfied for the Poisson bracket (1.7.7) as a consequence of

$$\begin{aligned}
&\{X_l,\, \{X_i,\, X_j\}\} + \{X_i,\, \{X_j,\, X_l\}\} + \{X_j,\, \{X_l,\, X_i\}\} \\
&= c_{ij}^k\{X_l,\, X_k\} + c_{jl}^k\{X_i,\, X_k\} + c_{li}^k\{X_j,\, X_k\} \\
&= \Big(c_{ij}^k c_{lk}^m + c_{jl}^k c_{ik}^m + c_{li}^k c_{jk}^m\Big) X_m = 0\,,
\end{aligned}$$

followed by comparison with equation (1.7.3) for the Jacobi identity in terms of the structure constants.

Remark 1.7.6 *This calculation for the Poisson bracket (1.7.8) provides an independent proof that it satisfies the Jacobi identity.*

The chain rule calculation (1.7.7) also reveals the following.

Theorem 1.7.7 *Under the map*

$$T^*\mathbb{R}^2 \to \mathbb{R}^3 : (\mathbf{q},\mathbf{p}) \to \mathbf{X} = (X_1, X_2, X_3)\,, \tag{1.7.11}$$

the Poisson bracket among the axisymmetric optical variables (1.4.5)

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

may be expressed for $S^2 = X_1X_2 - X_3^2$ as

$$\begin{aligned} \frac{dF}{dt} = \{F, H\} &= \nabla F \cdot \nabla S^2 \times \nabla H \\ &= -\frac{\partial S^2}{\partial X_l}\,\epsilon_{ljk}\frac{\partial F}{\partial X_j}\,\frac{\partial H}{\partial X_k}\,. \end{aligned} \tag{1.7.12}$$

Proof. This is a direct verification using formula (1.7.7). For example,

$$2\epsilon_{123}\frac{\partial S^2}{\partial X_3} = -4X_3, \quad 2\epsilon_{132}\frac{\partial S^2}{\partial X_2} = 2X_1, \quad 2\epsilon_{231}\frac{\partial S^2}{\partial X_1} = -2X_2\,.$$

(The inessential factors of 2 may be absorbed into the definition of the independent variable, which here is the time, t.) ■

The standard symbol ϵ_{klm} used in the last relation in (1.7.12) to write the triple scalar product of vectors in index form is defined as follows.

Definition 1.7.8 (Antisymmetric symbol ϵ_{klm})
The symbol ϵ_{klm} with $\epsilon_{123} = 1$ is the totally antisymmetric tensor in three dimensions: it vanishes if any of its indices are repeated and it equals the parity of the permutations of the set $\{1,2,3\}$ when $\{k,l,m\}$ are all different. That is,

$$\epsilon_{kkm} = 0 \text{ (no sum)}$$

and

$$\epsilon_{klm} = +1 \text{ (resp. } -1\text{) for even (resp. odd) permutations of } \{1,2,3\}\,.$$

Remark 1.7.9 *For three-dimensional vectors* **A**, **B**, **C**, *one has*

$$(\mathbf{B}\times\mathbf{C})_l = \epsilon_{klm}B_mC_n \quad \textit{and} \quad \big(\mathbf{A}\times(\mathbf{B}\times\mathbf{C})\big)_i = \epsilon_{ijk}A_j\epsilon_{klm}B_lC_m$$

Hence, the relation

$$\epsilon_{ijk}\epsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}$$

verifies the familiar BAC minus CAB rule for the triple vector product. That is,

$$\mathbf{A}\times(\mathbf{B}\times\mathbf{C}) = \mathbf{B}(\mathbf{A}\cdot\mathbf{C}) - \mathbf{C}(\mathbf{A}\cdot\mathbf{B})\,.$$

Corollary 1.7.10 *The equations of Hamiltonian ray optics in axisymmetric translation-invariant media may be expressed with* $H = H(X_1, X_2)$ *as*

$$\dot{\mathbf{X}} = \nabla S^2 \times \nabla H\,, \quad \textit{with} \quad S^2 = X_1X_2 - X_3^2 \geq 0\,. \qquad (1.7.13)$$

Thus, the flow preserves volume (that is, it satisfies $div\,\dot{\mathbf{X}} = 0$*) and the evolution along the curve* $\mathbf{X}(z) \in \mathbb{R}^3$ *takes place on intersections of level surfaces of the axisymmetric media invariants* S^2 *and* $H(X_1, X_2)$ *in* $\mathbb{R}^3$.

Remark 1.7.11 *The Petzval invariant* S^2 *satisfies* $\{S^2, H\} = 0$ *with the bracket (1.7.12) for every Hamiltonian* $H(X_1, X_2, X_3)$ *expressed in these variables.*

Definition 1.7.12 (Casimir, or distinguished function) *A function that Poisson-commutes with all other functions on a certain space is the Poisson bracket's* ***Casimir****, or* ***distinguished function****.*

1.8 Equilibrium solutions

1.8.1 Energy-Casimir stability

Remark 1.8.1 (Critical energy plus Casimir equilibria) *A point of tangency of the level sets of Hamiltonian H and Casimir*

S^2 is an equilibrium solution of equation (1.7.13). This is because, at such a point, the gradients of the Hamiltonian H and Casimir S^2 are collinear; so the right-hand side of (1.7.13) vanishes. At such points of tangency, the variation of the sum $H_\Phi = H + \Phi(S^2)$ vanishes, for some smooth function Φ. That is,

$$\begin{aligned} \delta H_\Phi(\mathbf{X}_e) &= DH_\Phi(\mathbf{X}_e) \cdot \delta\mathbf{X} \\ &= \Big[\nabla H + \Phi'(S^2)\nabla S^2\Big]_{\mathbf{X}_e} \cdot \delta\mathbf{X} = 0 \,, \end{aligned}$$

when evaluated at equilibrium points $\mathbf{X}_e$ where the level sets of H and S^2 are tangent.

> **Exercise.** Show that a point $\mathbf{X}_e$ at which H_Φ has a critical point (i.e., $\delta H_\Phi = 0$) must be an equilibrium solution of equation (1.7.13). ★

Energy-Casimir stability of equilibria

The second variation of the sum $H_\Phi = H + \Phi(S^2)$ is a quadratic form in $\mathbb{R}^3$ given by

$$\delta^2 H_\Phi(\mathbf{X}_e) = \delta\mathbf{X} \cdot D^2 H_\Phi(\mathbf{X}_e) \cdot \delta\mathbf{X} \,.$$

Thus we have, by Taylor's theorem,

$$H_\Phi(\mathbf{X}_e + \delta\mathbf{X}) - H_\Phi(\mathbf{X}_e) = \frac{1}{2}\delta^2 H_\Phi(\mathbf{X}_e) + o(|\delta\mathbf{X}|^2) \,,$$

when evaluated at the critical point $\mathbf{X}_e$. Remarkably, the quadratic form $\delta^2 H_\Phi(\mathbf{X}_e)$ is the Hamiltonian for the dynamics linearised around the critical point. Consequently, the second variation $\delta^2 H_\Phi$ is preserved by the linearised dynamics in a neighbourhood of the equilibrium point.

> **Exercise.** Linearise the dynamical equation (1.7.13) about an equilibrium $\mathbf{X}_e$ for which the quantity H_Φ has a critical point and show that the linearised dynamics conserves the quadratic form arising from the second variation. Show that the quadratic form is the Hamiltonian for the

linearised dynamics.
What is the corresponding Poisson bracket?
Does this process provides a proper bracket for the linearised dynamics? Prove that it does. ★

The ***signature*** of the second variation provides a method for determining the stability of the critical point. This is the ***energy-Casimir stability method***. This method is based on the following.

Theorem 1.8.2 *A critical point $\mathbf{X}_e$ of $H_\Phi = H + \Phi(S^2)$ whose second variation is definite in sign is a stable equilibrium solution of equation (1.7.13).*

Proof. A critical point $\mathbf{X}_e$ of $H_\Phi = H + \Phi(S^2)$ is an equilibrium solution of equation (1.7.13). Sign definiteness of the second variation provides a norm $\|\delta\mathbf{X}\|^2 = |\delta^2 H_\Phi(\mathbf{X}_e)|$ for the perturbations around the equilibrium $\mathbf{X}_e$ that is conserved by the linearised dynamics. Being conserved by the dynamics linearised around the equilibrium, this sign-definite distance from $\mathbf{X}_e$ must remain constant. Therefore, in this case, the absolute value of sign-definite second variation $|\delta^2 H_\Phi(\mathbf{X}_e)|$ provides a distance from the equilibrium $\|\delta\mathbf{X}\|^2$ which is bounded in time under the linearised dynamics. Hence, the equilibrium solution is stable.

■

Remark 1.8.3 *Even when the second variation is indefinite, it is still linearly conserved. However, an indefinite second variation does not provide a norm for the perturbations. Consequently, an indefinite second variation does not limit the growth of a perturbation away from its equilibrium.*

Definition 1.8.4 (Geometrical nature of equilibria)
*An equilibrium whose second variation is sign-definite are called **elliptic**, because the level sets of the second variation in this case*

make closed, nearly elliptical contours in its Euclidean neighbourhood. Hence, the orbits on these closed level sets remain near the equilibria in the sense of the Euclidean norm on $\mathbb{R}^3$. (In $\mathbb{R}^3$ all norms are equivalent to the Euclidean norm.)

An equilibrium with sign-indefinite second variation is called ***hyperbolic****, because the level sets of the second variation do not close locally in its Euclidean neighbourhood. Hence, in this case, an initial perturbation following a hyperbolic level set of the second variation may move out of the Euclidean neighbourhood of the equilibrium.*

1.9 Momentum maps

1.9.1 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

$$\xi J + J\xi = 0 \quad \text{so that} \quad J\xi J = \xi\,. \tag{1.9.1}$$

Exercise. Verify (1.9.1), cf. (1.6.8). ★

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 2-component real multi-vector $\mathbf{z} = (\mathbf{q},\, \mathbf{p})^T$ has components $(\xi_{kl}q_l, \xi_{kl}p_l)^T$, with $k, l = 1, 2$. The matrix ξ is any linear combination of the traceless constant Hamiltonian matrices (1.6.5).

Definition 1.9.1 (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 by

$$
\begin{aligned}
\mathcal{J}^\xi(\mathbf{z}) &:= \Big\langle \mathcal{J}(\mathbf{z}),\, \xi \Big\rangle_{sp(2,\mathbb{R})^* \times sp(2,\mathbb{R})} \\
&= \Big(\mathbf{z},\, J\xi\mathbf{z} \Big)_{\mathbb{R}^2\times\mathbb{R}^2} \\
&:= z_k (J\xi)_{kl} z_l \\
&= \mathbf{z}^T \cdot J\xi\mathbf{z} \\
&= \mathrm{tr}\Big((\mathbf{z} \otimes \mathbf{z}^T J)\xi \Big) \,,
\end{aligned}
\tag{1.9.2}
$$

where $\mathbf{z} = (\mathbf{q},\, \mathbf{p})^T \in \mathbb{R}^2 \times \mathbb{R}^2$.

Remark 1.9.2 *The map $\mathcal{J}(\mathbf{z})$ given in (1.9.2) by*

$$
\mathcal{J}(\mathbf{z}) = (\mathbf{z} \otimes \mathbf{z}^T J) \in sp(2,\mathbb{R})^* \,, \tag{1.9.3}
$$

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,

$$
\Big\langle \mathcal{J}(\mathbf{z}),\, \xi \Big\rangle = \mathrm{tr}\,\big(\mathcal{J}(\mathbf{z})\, \xi \big) \,, \tag{1.9.4}
$$

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 1.9.3 (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 (1.9.2) for $Sp(2,\mathbb{R})$ acting diagonally on $\mathbb{R}^2 \times \mathbb{R}^2$ in equation (1.9.3) may be expressed in matrix form as

$$
\begin{aligned}
\mathcal{J} &= (\mathbf{z} \otimes \mathbf{z}^T J) \\
&= 2 \begin{pmatrix} \mathbf{p}\cdot\mathbf{q} & -|\mathbf{q}|^2 \\ |\mathbf{p}|^2 & -\mathbf{p}\cdot\mathbf{q} \end{pmatrix} \\
&= 2 \begin{pmatrix} X_3 & -X_1 \\ X_2 & -X_3 \end{pmatrix} .
\end{aligned}
\tag{1.9.5}
$$

This is none other than matrix form of the map (1.7.11) 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

$$X_1 = |\mathbf{q}|^2 \geq 0\,, \quad X_2 = |\mathbf{p}|^2 \geq 0\,, \quad X_3 = \mathbf{p}\cdot\mathbf{q}\,. \qquad (1.9.6)$$

Applying the momentum map $\mathcal{J}$ *to the vector of Hamiltonian matrices* $\mathbf{m} = (m_1,\, m_2,\, m_3)$ *in equation (1.6.5) yields the individual components,*

$$\mathcal{J}\cdot\mathbf{m} = 2\mathbf{X} \quad\Longleftrightarrow\quad \mathbf{X} = \frac{1}{2}\, z_k (J\mathbf{m})_{kl} z_l\,. \qquad (1.9.7)$$

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 (1.9.7) 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 1.9.4 (Momentum maps for ray optics)
Our previous discussions have revealed that the axisymmetric variables (X_1, X_2, X_3) *in (1.9.6) generate the Lie group of symplectic transformations (1.6.3) as flows of Hamiltonian vector fields. It turns out that this result is connected to the theory of* ***momentum maps****. Momentum maps take phase space coordinates* $(\mathbf{q},\,\mathbf{p})$ *to the space of Hamiltonians whose flows are canonical transformations of phase space. An example of a momentum map already appeared in Definition 1.3.18.*

The Hamiltonian functions for the one-parameter subgroups of the symplectic group $Sp(2,\mathbb{R})$ in the KAN decomposition (1.6.13) are given by

$$H_K = \frac{1}{2}(X_1 + X_2)\,, \quad H_A = X_3 \quad and \quad H_N = -\,X_1\,. \qquad (1.9.8)$$

The three phase space functions,

$$H_K = \frac{1}{2}(|\mathbf{q}|^2 + |\mathbf{p}|^2)\,, \quad H_A = \mathbf{q}\cdot\mathbf{p}\,, \quad H_N = -\,|\mathbf{q}|^2\,, \qquad (1.9.9)$$

map the phase space $(\mathbf{q},\, \mathbf{p})$ to these Hamiltonians whose corresponding Poisson brackets are the Hamiltonian vector fields for the corresponding one-parameter subgroups. These three Hamiltonians and, equally well, any other linear combinations of (X_1, X_2, X_3), arise from a single ***momentum map****, as we shall explain in Section 1.9.2.*

Remark 1.9.5 *Momentum maps are* ***Poisson maps****. That is, they map Poisson brackets on phase space into Poisson brackets on the target space.*

The corresponding Lie algebra product in $sp(2,\mathbb{R})$ was identified using Theorem 1.7.7 with the vector cross product in the space $\mathbb{R}^3$ by using the $\mathbb{R}^3$-bracket. The $\mathbb{R}^3$-brackets among the (X_1, X_2, X_3) closed among themselves. Therefore, as expected, the momentum map was found to be Poisson. In general, when the Poisson bracket relations are all linear, they will be Lie-Poisson brackets, defined below in Section 1.10.1.

1.9.2 Summary: Properties of momentum maps

A momentum map takes phase space coordinates $(\mathbf{q},\, \mathbf{p})$ to the space of Hamiltonians, whose flows are canonical transformations of phase space. The ingredients of the momentum map are: (i) a representation of the infinitesimal action of the Lie algebra of the transformation group on the coordinate space; and (ii) an appropriate pairing with the conjugate momentum space. For example, one may construct a momentum map by using the familiar pairing $\langle\!\langle \cdot\,, \cdot \rangle\!\rangle$ between momentum in phase space and the velocity in the tangent space of the configuration manifold that also appears in the Legendre transformation. For this pairing, the momentum map is derived from the cotangent lift of the infinitesimal action $\xi_M(\mathbf{q})$ of the Lie algebra of the transformation group on the configuration manifold to its action on the canonical momentum. In this case, the formula for the momentum map $\mathcal{J}(\mathbf{q}, \mathbf{p})$ is

$$\mathcal{J}^{\xi}(\mathbf{q}, \mathbf{p}) = \Big\langle \mathcal{J}(\mathbf{q}, \mathbf{p}),\, \xi \Big\rangle = \Big\langle\!\!\Big\langle \mathbf{p},\, \xi_M(\mathbf{q}) \Big\rangle\!\!\Big\rangle\,, \tag{1.9.10}$$

in which the other pairing $\langle \cdot\,, \cdot \rangle$ is between the Lie algebra and its dual. This means the momentum map $\mathcal{J}$ for the Hamiltonian $\mathcal{J}^{\xi}$ lives in the dual space of the Lie algebra belonging to the Lie symmetry. The flow of its vector field $X_{\mathcal{J}^{\xi}} = \{\cdot\,, \mathcal{J}^{\xi}\}$ is the transformation of phase space by the cotangent lift of a Lie group symmetry infinitesimally generated for configuration space by $\xi_M(\mathbf{q})$. The computation of the Lagrange invariant S in (1.3.23) is an example of this type of momentum map.

Not all momentum maps arise as cotangent lifts. Momentum maps may also arise from the infinitesimal action of the Lie algebra on the phase space manifold $\xi_{T^*M}(\mathbf{z})$ with $\mathbf{z} = (\mathbf{q}, \mathbf{p})$ by using the pairing with the symplectic form. The formula for the momentum map is then

$$\mathcal{J}^{\xi}(\mathbf{z}) = \Big\langle \mathcal{J}(\mathbf{z}),\, \xi \Big\rangle = \Big(\mathbf{z},\, J\xi_{T^*M}(\mathbf{z})\Big)\,, \tag{1.9.11}$$

where J is the symplectic form and $(\cdot\,, \cdot)$ is the inner product on phase space $T^*\mathbb{R} \simeq \mathbb{R} \times \mathbb{R}$ for n degrees of freedom. The transformation to axisymmetric variables in (1.9.5) is an example of a momentum map obtained from the symplectic pairing. Both of these approaches are useful and we have seen that both types of momentum maps are summoned when reduction by S^1 axisymmetry is applied in ray optics. The present chapter explores the consequences of S^1 symmetry and the reductions of phase space associated with the momentum maps for this symmetry.

The level sets of the momentum maps provide the geometrical setting for dynamics with symmetry. The components of the momentum map live on the dual of the Lie symmetry algebra, which is a linear space. The level sets of the components of the momentum map provide the natural coordinates for the reduced dynamics. Thus, the motion takes place in a ***reduced space*** whose coordinates are invariant under the original S^1 symmetry. The motion in the reduced space lies on a level set of the momentum map for the S^1 symmetry. It also lies on a level set of the Hamiltonian. Hence, the dynamics in the reduced space of coordinates that are invariant under the S^1 symmetry is confined to the intersections in the reduced space of the level sets of the Hamiltonian and the momentum map associated with that symmetry. Moreover, in most cases, restriction to either level set results in symplectic (canonical) dynamics.

After the solution for this S^1-reduced motion is determined, one must reconstruct the phase associated with the S^1 symmetry, which decouples from the dynamics of the rest of system through the process of reduction. Thus, each point on the manifolds defined by the level sets of the Hamiltonian and the momentum map in the reduced space is associated with an orbit of the phase on S^1. This S^1 phase must be reconstructed from the solution on the reduced space of S^1-invariant functions. The reconstruction of the phase is of interest in its own right, because it contains both geometric and dynamic components, as discussed in Section 1.12.2.

One advantage of this geometric setting is that it readily reveals how ***bifurcations*** arise under changes of parameters in the system, for example, under changes in parameters in the Hamiltonian. In this setting, bifurcations are topological transitions in the intersections of level surfaces of ***orbit manifolds*** of the Hamiltonian and momentum map. The motion proceeds along these intersections in the reduced space whose points are defined by S^1-invariant coordinates. These topological changes in the intersections of the orbit manifolds accompany qualitative changes in the solution behaviour, such as the change of stability of an equilibrium, or the creation or destruction of equilibria. The display of these changes of topology in the reduced space of S^1-invariant functions also allows a visual classification of potential bifurcations. That is, it affords an opportunity to organise the ***choreography of bifurcations*** that are available to the system as its parameters are varied. For an example of this type of geometric bifurcation analysis, see Section 4.5.5.

Remark 1.9.6 *The two results:*
(1) that the action of a Lie group G with Lie algebra $\mathfrak{g}$ on a symplectic manifold P should be accompanied by a momentum map $J : P \to \mathfrak{g}^*$; and
(2) that the orbits of this action are themselves symplectic manifolds, both occur already in [Lie1890]. See [We1983] for an interesting discussion of Lie's contributions to the theory of momentum maps.

The reader should consult [MaRa1994, OrRa2004] for more discussions and additional examples of momentum maps.

1.10 Lie-Poisson brackets

1.10.1 The $\mathbb{R}^3$-bracket for ray optics is Lie-Poisson

The Casimir invariant $S^2 = X_1X_2 - X_3^2$ for the $\mathbb{R}^3$-bracket (1.7.12) is quadratic. In such cases, one may write the Poisson bracket on $\mathbb{R}^3$ in the suggestive form with a pairing $\langle\,\cdot\,,\,\cdot\,\rangle$,

$$\{F,H\} = -\,X_k\, c^k_{ij}\frac{\partial F}{\partial X_i}\frac{\partial H}{\partial X_j} =: -\left\langle \mathbf{X}\,,\,\left[\frac{\partial F}{\partial \mathbf{X}}\,,\,\frac{\partial H}{\partial \mathbf{X}}\right]\right\rangle, \qquad (1.10.1)$$

where c^k_{ij} with $i,j,k = 1,2,3$ are the structure constants of a three-dimensional Lie algebra operation denoted as $[\,\cdot\,,\,\cdot\,]$. In the particular case of ray optics, $c^3_{12} = 4$, $c^2_{23} = 2$, $c^1_{31} = 2$ and the rest of the structure constants either vanish, or are obtained from antisymmetry of c^k_{ij} under exchange of any pair of its indices. These values are the structure constants of the 2×2 Hamiltonian matrices (1.6.5), which represent any of the Lie algebras $sp(2,\mathbb{R})$, $so(2,1)$, $su(1,1)$, or $sl(2,\mathbb{R})$. Thus, the reduced description of Hamiltonian ray optics in terms of axisymmetric $\mathbb{R}^3$ variables may be defined on the dual space of any of these Lie algebras, say, $sp(2,\mathbb{R})^*$ for definiteness, where duality is defined by pairing $\langle\,\cdot\,,\,\cdot\,\rangle$ in $\mathbb{R}^3$ (contraction of indices). Since $\mathbb{R}^3$ is dual to itself under this pairing, upper and lower indices are equivalent.

Definition 1.10.1 (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 1.10.2 *Equation (1.10.1) 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.*

1.10.2 Lie-Poisson brackets with quadratic Casimirs

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

$$\{F,H\}_C := -\,\nabla C\cdot\nabla F\times\nabla H\,, \tag{1.10.2}$$

when its Casimir is the quadratic form on $\mathbb{R}^3$ given by $C=\frac{1}{2}\mathbf{X}^T\cdot\mathsf{K}\mathbf{X}$ 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

$$\begin{aligned}
\{F,\,H\}_{\mathsf{K}} &= -\,\nabla C\cdot\nabla F\times\nabla H\\
&= -X_l\mathsf{K}^{li}\epsilon_{ijk}\frac{\partial F}{\partial X_j}\frac{\partial H}{\partial X_k}\\
&= -\,\mathbf{X}\cdot\mathsf{K}\left(\frac{\partial F}{\partial\mathbf{X}}\times\frac{\partial H}{\partial\mathbf{X}}\right)\\
&=: -\,\left\langle\mathbf{X}\,,\,\left[\frac{\partial F}{\partial\mathbf{X}}\,,\,\frac{\partial H}{\partial\mathbf{X}}\right]_{\mathsf{K}}\right\rangle.
\end{aligned} \tag{1.10.3}$$

Remark 1.10.3 *The triple scalar product of gradients in the* $\mathbb{R}^3$*-bracket (1.10.2) is the determinant of the Jacobian matrix for the transformation* $(X_1,X_2,X_3)\rightarrow(C,F,H)$*, 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 (1.10.3) yields

$$\{X_j,\,X_k\}_{\mathsf{K}} = -X_l\mathsf{K}^{li}\epsilon_{ijk}\,. \tag{1.10.4}$$

The Lie-Poisson form in (1.10.3) associates the $\mathbb{R}^3$ bracket to a Lie algebra with structure constants given in the dual vector basis by

$$[\mathbf{e}_j,\,\mathbf{e}_k]_{\mathsf{K}} = \mathbf{e}_l \mathsf{K}^{li}\epsilon_{ijk} =: \mathbf{e}_l c^{\,l}_{jk}\,. \tag{1.10.5}$$

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

$$O^T(t)\mathsf{K}O(t) = \mathsf{K} \tag{1.10.6}$$

or, equivalently,

$$\mathsf{K}^{-1}O^T(t)\mathsf{K} = O^{-1}(t)\,, \tag{1.10.7}$$

for the 3×3 symmetric matrix $\mathsf{K}^T = \mathsf{K}$. A matrix $O(t)$ satisfying (1.10.6) is called a ***quasi-orthogonal matrix*** with respect to K. That is, $O(t)$ is the similarity transformation of an orthogonal matrix by the symmetric matrix K.

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

$$(O_1O_2)^T\mathsf{K}(O_1O_2) = O_2^T(O_1^T\mathsf{K}O_1)O_2 = O_2^T\mathsf{K}O_2 = \mathsf{K}\,. \tag{1.10.8}$$

The corresponding Lie algebra $\mathfrak{g}_{\mathsf{K}}$ is the derivative of the defining condition of the Lie group (1.10.6), evaluated at the identity. This yields,

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

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 $\mathbb{R}^3$,

$$\widehat{\ }: \mathfrak{g}_{\mathsf{K}} \to \mathbb{R}^3 \quad \text{defined by} \quad (\mathsf{K}\widehat{X})_{jk} = -\,X_l\mathsf{K}^{li}\epsilon_{ijk}\,. \tag{1.10.9}$$

When K is invertible, the hat map $(\,\widehat{\cdot}\,)$ in (1.10.9) 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,

$$[\mathbf{u},\mathbf{v}]_{\mathsf{K}} = \mathsf{K}(\mathbf{u}\times\mathbf{v})\,. \tag{1.10.10}$$

Thus, the Lie algebra of the Lie group of transformations of $\mathbb{R}^3$ leaving invariant the quadratic form $C = \frac{1}{2}\mathbf{X}^T\cdot\mathsf{K}\mathbf{X}$ 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

$$S^2 = X_1X_2 - X_3^2 = \mathbf{X}\cdot\begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix}\mathbf{X}\,.$$

Consequently,

$$\mathsf{K} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{pmatrix},$$

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

> **Exercise.** Verify that inserting this formula for K into formula (1.10.4) recovers the Lie-Poisson bracket relations (1.7.5) for ray optics (up to an inessential constant).
> ★

Hence, we have proved the following theorem.

Theorem 1.10.4 *Consider the $\mathbb{R}^3$ bracket in equation (1.10.3)*

$$\{F,H\}_{\mathsf{K}} := -\nabla C_{\mathsf{K}}\cdot\nabla F\times\nabla H \quad \textit{with} \quad C_{\mathsf{K}} = \frac{1}{2}\mathbf{X}\cdot\mathsf{K}\mathbf{X}\,, \tag{1.10.11}$$

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*

$$\begin{aligned} \{F, H\}_\mathsf{K} &= -\,\mathbf{X} \cdot \mathsf{K} \left(\frac{\partial F}{\partial \mathbf{X}} \times \frac{\partial H}{\partial \mathbf{X}} \right), \qquad (1.10.12) \\ &=: -\left\langle \mathbf{X}\,,\, \left[\frac{\partial F}{\partial \mathbf{X}}\,,\, \frac{\partial H}{\partial \mathbf{X}} \right]_\mathsf{K} \right\rangle, \qquad (1.10.13) \end{aligned}$$

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

$$[\,\mathbf{u}, \mathbf{v}\,]_\mathsf{K} = \mathsf{K}(\mathbf{u} \times \mathbf{v})\,. \qquad (1.10.14)$$

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 1.10.5 (The ad **and** ad* **operations)**
The adjoint (ad) *and coadjoint* (ad*) *operations are defined for the Lie-Poisson bracket (1.10.13) with quadratic Casimir,* $C_\mathsf{K} = \frac{1}{2}\,\mathbf{X} \cdot \mathsf{K}\mathbf{X}$, *as follows.*

$$\begin{aligned} \langle \mathbf{X}\,,\, [\,\mathbf{u}, \mathbf{v}\,]_\mathsf{K} \rangle &= \langle \mathbf{X}\,,\, \mathrm{ad}_\mathbf{u}\mathbf{v} \rangle = \langle \mathrm{ad}^*_\mathbf{u}\mathbf{X}\,,\, \mathbf{v} \rangle \qquad (1.10.15) \\ &= \mathsf{K}\mathbf{X} \cdot (\mathbf{u} \times \mathbf{v}) = (\mathsf{K}\mathbf{X} \times \mathbf{u}) \cdot \mathbf{v}\,. \end{aligned}$$

Thus, we have explicitly,

$$\mathrm{ad}_\mathbf{u}\mathbf{v} = \mathsf{K}(\mathbf{u} \times \mathbf{v}) \quad \textit{and} \quad \mathrm{ad}^*_\mathbf{u}\mathbf{X} = -\,\mathbf{u} \times \mathsf{K}\mathbf{X}\,. \qquad (1.10.16)$$

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

Theorem 1.10.6 (Lie-Poisson dynamics)
The Lie-Poisson dynamics (1.10.12) - (1.10.13) is expressed in terms of the ad *and* ad* *operations by*

$$\begin{aligned} \frac{dF}{dt} = \{F, H\}_\mathsf{K} &= \left\langle \mathbf{X}\,,\, \mathrm{ad}_{\partial H/\partial \mathbf{X}}\, \frac{\partial F}{\partial \mathbf{X}} \right\rangle \\ &= \left\langle \mathrm{ad}^*_{\partial H/\partial \mathbf{X}}\mathbf{X}\,,\, \frac{\partial F}{\partial \mathbf{X}} \right\rangle, \qquad (1.10.17) \end{aligned}$$

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

$$\frac{d\mathbf{X}}{dt} = \{\mathbf{X},\, H\}_{\mathsf{K}} = \mathrm{ad}^*_{\partial H/\partial\mathbf{X}}\mathbf{X} = -\,\frac{\partial H}{\partial \mathbf{X}} \times \mathsf{K}\mathbf{X}\,. \quad (1.10.18)$$

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 (1.10.18) for $\mathsf{K} = \mathrm{diag}(1,1,1)$ and $H = X_1^2 - X_3^2/2$. ★

1.11 Divergenceless vector fields

1.11.1 Jacobi identity

One may verify directly that the $\mathbb{R}^3$ bracket in (1.7.12) and in the class of brackets (1.10.11) 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 Leibnitz operator that satisfies the Jacobi identity, we identify the bracket in (1.7.12) with the following ***divergenceless vector field*** on $\mathbb{R}^3$ defined by

$$X_H = \{\,\cdot\,, H\} = \nabla S^2 \times \nabla H \cdot \nabla \in \mathfrak{X}\,. \quad (1.11.1)$$

This isomorphism identifies the bracket in (1.11.1) 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 1.11.1 (Jacobi-Lie bracket)
*The **commutator** of two divergenceless vector fields $u, v \in \mathfrak{X}$ is defined to be*

$$\big[v,\, w\big] = \big[\mathbf{v}\cdot\nabla,\, \mathbf{w}\cdot\nabla\big] = \Big((\mathbf{v}\cdot\nabla)\mathbf{w} - (\mathbf{w}\cdot\nabla)\mathbf{v}\Big)\cdot\nabla\,. \quad (1.11.2)$$

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*

$$\big[v,\, w\big] = (\mathbf{v}\cdot\nabla)\mathbf{w} - (\mathbf{w}\cdot\nabla)\mathbf{v}\,. \tag{1.11.3}$$

In Euclidean vector components, the Jacobi-Lie bracket (1.11.3) is expressed as

$$\big[v,\, w\big]_i = w_{i,j}v_j - v_{i,j}w_j\,. \tag{1.11.4}$$

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 (1.11.4); so the commutator of two divergenceless vector fields yields another one. ★

Remark 1.11.2 (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*

$$\frac{d}{d\epsilon}\bigg|_{\epsilon=0} = v_i(\mathbf{x})\frac{\partial}{\partial x_i}\,. \tag{1.11.5}$$

The characteristic equations of this flow are

$$\frac{dx_i}{d\epsilon} = v_i(\mathbf{x}(\epsilon))\,, \quad \textit{so that} \quad \frac{dx_i}{d\epsilon}\bigg|_{\epsilon=0} = v_i(\mathbf{x})\,, \quad i = 1, 2, 3\,. \tag{1.11.6}$$

Integration of the characteristic equations (1.11.6) yields the solution for the ***flow*** $\mathbf{x}(\epsilon) = \phi_\epsilon\mathbf{x}$ *of the vector field defined by (1.11.5), 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

$$\begin{aligned}
\big[u,[v,w]\big]_i &= u_k v_j w_{i,kj} + u_k v_{j,k} w_{i,j} - u_k w_{j,k} v_{i,j} \\
&\quad - u_k w_j v_{i,jk} - v_j w_{k,j} u_{i,k} + w_j v_{k,j} u_{i,k}\,. \qquad (1.11.7)
\end{aligned}$$

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 1.11.3 *The Jacobi-Lie bracket of divergenceless vector fields satisfies the* ***Jacobi identity,***

$$[u,[v,w]\,] + [v,[w,u]\,] + [w,[u,v]\,] = 0\,. \qquad (1.11.8)$$

Proof. Direct verification using (1.11.7) and summing over cyclic permutations. ■

Exercise. Prove Theorem 1.11.3 in streamlined notation obtained by writing

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

and using bilinearity of the Jacobi-Lie bracket. ★

Lemma 1.11.4 *The $\mathbb{R}^3$-bracket (1.7.12) may be identified with the divergenceless vector fields in (1.11.1) by*

$$[X_G, X_H] = -X_{\{G,H\}}\,, \tag{1.11.9}$$

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

Proof. Equation (1.11.9) 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 1.11.5 *The last step in the proof of Lemma 1.11.4 uses the Jacobi identity for the class of $\mathbb{R}^3$-brackets in equation (1.10.2).*

1.11.2 Geometric forms of Poisson brackets

Determinant & 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 $(q, p) \to (F(q, p), H(q, p))$,

$$\{F, H\} = \frac{\partial F}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial H}{\partial q}\frac{\partial F}{\partial p} = \det\frac{\partial(F, H)}{\partial(q, p)}\,. \tag{1.11.10}$$

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

$$dF = \frac{\partial F}{\partial q}\,dq + \frac{\partial F}{\partial p}\,dp \quad \text{and} \quad dH = \frac{\partial H}{\partial q}\,dq + \frac{\partial H}{\partial p}\,dp\,, \tag{1.11.11}$$

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

$$dF \wedge dH = \det\frac{\partial(F, H)}{\partial(q, p)}dq \wedge dp = \{F, H\}dq \wedge dp\,. \tag{1.11.12}$$

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

(i) $\wedge$ is anticommutative: $dF \wedge dG = -dG \wedge dF$;

(ii) $\wedge$ is bilinear: $(adF + bdG) \wedge dH = a(dF \wedge dH) + b(dG \wedge dH)$;

(iii) $\wedge$ is associative: $dF \wedge (dG \wedge dH) = (dF \wedge dG) \wedge dH$.

Remark 1.11.7 *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 (1.11.12) from equation (1.11.11) and the linearity and antisymmetry of the wedge product, so that, e.g., $dq \wedge dp = -dp \wedge dq$ and $dq \wedge dq = 0$.
>
> ★

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

The $\mathbb{R}^3$-bracket in equation (1.7.12) may also be rewritten equivalently as a Jacobian determinant, namely,

$$\{F,\, H\} = -\,\nabla S^2 \cdot \nabla F \times \nabla H = -\,\frac{\partial(S^2, F, H)}{\partial(X_1, X_2, X_3)}\,, \qquad (1.11.13)$$

where

$$\frac{\partial(F_1, F_2, F_3)}{\partial(X_1, X_2, X_3)} = \det\left(\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right). \qquad (1.11.14)$$

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

$$\epsilon_{ijk} \det\left(\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right) = \epsilon_{abc}\, \frac{\partial F_a}{\partial X_i} \frac{\partial F_b}{\partial X_j} \frac{\partial F_c}{\partial X_k}\,, \qquad (1.11.15)$$

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

$$\det\left(\frac{\partial \mathbf{F}}{\partial \mathbf{X}}\right) dX_1 \wedge dX_2 \wedge dX_3 = dF_1 \wedge dF_2 \wedge dF_3 \,. \qquad (1.11.16)$$

Thus, the $\mathbb{R}^3$-bracket in equation (1.7.12) 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].

1.11.3 Nambu brackets

Theorem 1.11.8 (Nambu brackets [Na1973])
*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

$$\begin{aligned} \{F,\, H\}\, d^{\,3}X &= -\nabla C \cdot \nabla F \times \nabla H d^{\,3}X \\ &= -\, dC \wedge dF \wedge dH \,, \end{aligned} \qquad (1.11.17)$$

possesses the properties (1.3.4) 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 Leibnitz 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 (1.11.17) satisfies all the properties required of a Poisson bracket specified in Definition 1.3.4. ■

1.12 Geometry of solution behaviour

1.12.1 Restricting axisymmetric ray optics to level sets

Having realised that the $\mathbb{R}^3$-bracket in equation (1.7.12) is associated to Jacobian determinants for changes of variables, it is natural to transform the dynamics of the axisymmetric optical variables (1.4.5) 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,

$$S^2 = X_1X_2 - X_3^2 = Y_1^2 - Y_2^2 - Y_3^2\,, \tag{1.12.1}$$

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 (1.12.1) 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,

$$Y_1 = S\cosh u\,, \quad Y_2 = S\sinh u\cos\psi\,, \quad Y_3 = S\sinh u\sin\psi\,. \tag{1.12.2}$$

The $\mathbb{R}^3$-bracket (1.7.12) thereby transforms into hyperbolic coordinates (1.12.2) as

$$\{F,\,H\}\,dY_1\wedge dY_2\wedge dY_3 = -\,\{F,\,H\}_{hyperb}\,S^2\,dS\wedge d\psi\wedge d\cosh u\,. \tag{1.12.3}$$

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

$$\begin{aligned}\{F,\,H\}\,d^{\,3}Y &= -\,S^2 dS\wedge dF\wedge dH \qquad (1.12.4)\\ &= -\,S^2 dS\wedge\{F,\,H\}_{hyperb}\,d\psi\wedge d\cosh u\,,\end{aligned}$$

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***,

$$\frac{d\psi}{dz} = \{\psi,\,H\}_{hyperb} = \frac{\partial H}{\partial\cosh u}\,, \tag{1.12.5}$$

$$\frac{d\cosh u}{dz} = \{\cosh u,\,H\}_{hyperb} = -\,\frac{\partial H}{\partial\psi}\,. \tag{1.12.6}$$

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 1.12.1 *For a paraxial harmonic guide, whose Hamiltonian is,*

$$H = \frac{1}{2}(|\mathbf{p}|^2 + |\mathbf{q}|^2) = \frac{1}{2}(X_1 + X_2) = Y_1\,, \tag{1.12.7}$$

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

$$\dot{\mathbf{Y}} = \big\{\mathbf{Y},\, H\big\} = \nabla_{\mathbf{Y}} S^2 \times \widehat{\mathbf{Y}}_{\mathbf{1}} = 2\widehat{\mathbf{Y}}_{\mathbf{1}} \times \mathbf{Y}\,, \tag{1.12.8}$$

on using the (1.12.1) to transform the $\mathbb{R}^3$ bracket in (1.7.12). 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 (1.12.3) transforms the $\mathbb{R}^3$-bracket from Cartesian to hyperboloidal coordinates, by using the definitions in equations (1.12.2). ★

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

Exercise. Reduce the $\mathbb{R}^3$ dynamics of (1.7.12) 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. ★

1.12.2 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

$$
\begin{aligned}
d^{\,3}Y &=: dY_1 \wedge dY_2 \wedge dY_3\\
&= d\frac{S^3}{3} \wedge d\cosh u \wedge d\psi\\
&= dp_\phi^2 \wedge dp_r \wedge dr\,.
\end{aligned}
\tag{1.12.9}
$$

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

$$
S\, d\cosh u \wedge d\psi = 2\, dp_r \wedge dr\,, \tag{1.12.10}
$$

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

One recalls Stokes Theorem on phase space

$$
\iint_A dp_j \wedge dq_j = \oint_{\partial A} p_j dq_j\,, \tag{1.12.11}
$$

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

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

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

$$\begin{aligned} \oint \frac{S^3}{3} d\phi &= \frac{S^3}{3} \Delta\phi \\ &= \underbrace{-\oint \cosh u \, d\psi}_{\text{Geometric } \Delta\phi} + \underbrace{\oint \mathbf{p} \cdot d\mathbf{q}}_{\text{Dynamic } \Delta\phi} . \end{aligned} \tag{1.12.12}$$

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 (1.12.11), 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***.

1.13 Singular 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, *Optiks* 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 1.1.1 adapts easily to yield the expected 3D eikonal equation (1.1.9). However, in general, the Lagrangian in such a description is singular, as we shall explain. The Euler-Lagrange equation,

$$\frac{d}{ds}\left(\frac{\partial L}{\partial \dot{\mathbf{r}}(s)}\right) = \frac{\partial L}{\partial \mathbf{r}(s)}\,, \tag{1.13.1}$$

follows from the variational principle,

$$0 = \delta S = \delta \int_A^B L(\mathbf{r}(s), \dot{\mathbf{r}}(s))ds\,. \tag{1.13.2}$$

The Lagrangian in the case of an anisotropic optical medium is given by

$$L(\mathbf{r}(s), \dot{\mathbf{r}}(s)) = n(\mathbf{r}(s), \dot{\mathbf{r}}(s))\,|\dot{\mathbf{r}}(s)|\,. \tag{1.13.3}$$

Here the refractive index is modelled as a function both of position along the ray $\mathbf{r}(s)$ and the ray direction $\dot{\mathbf{r}}(s)$, which is a unit vector. The latter is defined when s is the arclength as

$$\widehat{\mathbf{s}} = \dot{\mathbf{r}}(s)/|\dot{\mathbf{r}}(s)| \quad \text{with} \quad |\dot{\mathbf{r}}| = 1\,. \tag{1.13.4}$$

> **Exercise.** Show that the variation of the ray direction in (1.13.4) is related to the variation of the path $\delta\mathbf{r}(s)$ by
>
> $$\delta\,\widehat{\mathbf{s}} = \frac{-1}{|\dot{\mathbf{r}}|}\,\widehat{\mathbf{s}} \times \Big(\widehat{\mathbf{s}} \times \delta\dot{\mathbf{r}}(s)\Big)\,.$$
>
> ★

The variational principle (1.13.2) with optical Lagrangian (1.13.3) implies the following ***3D eikonal equation*** for the vector $\mathbf{r}(s)$ defining the ray path,

$$\frac{d}{ds}\Big(n(\mathbf{r},\widehat{\mathbf{s}}\,)\widehat{\mathbf{s}} + \mathbf{A}(\mathbf{r},\widehat{\mathbf{s}}\,)\Big) = \frac{\partial n(\mathbf{r},\,\widehat{\mathbf{s}}\,)}{\partial \mathbf{r}}\,. \tag{1.13.5}$$

Here, the ***anisotropy vector*** $\mathbf{A}(\mathbf{r},\widehat{\mathbf{s}}\,)$ is defined as

$$\mathbf{A} := \frac{\partial n}{\partial \dot{\mathbf{r}}}\bigg|_{|\dot{\mathbf{r}}|=1} = -\,\widehat{\mathbf{s}} \times \left(\widehat{\mathbf{s}} \times \frac{\partial n}{\partial \widehat{\mathbf{s}}}\right). \tag{1.13.6}$$

The anisotropy vector $\mathbf{A}$ is the projection of the vector $\partial n/\partial\dot{\mathbf{r}}$ onto the plane that is normal to $\dot{\mathbf{r}}$ and tangent to the direction sphere $|\dot{\mathbf{r}}| = 1$.

In the $\dot{\mathbf{r}}$ notation, the ***3D optical momentum*** is defined as

$$\mathbf{p}_3 := \frac{\partial L}{\partial \dot{\mathbf{r}}(s)} = n(\mathbf{r}, \dot{\mathbf{r}}\,)\dot{\mathbf{r}} + \mathbf{A}(\mathbf{r}, \dot{\mathbf{r}})\,. \tag{1.13.7}$$

Thus, the 3D optical momentum $\mathbf{p}_3$ lies in the plane spanned by the vectors $\dot{\mathbf{r}}$ and $\partial n/\partial\dot{\mathbf{r}}$, and these two vectors are orthogonal because of the constraint $|\dot{\mathbf{r}}| = 1$. The optical momentum is related to the tangent vector $\dot{\mathbf{r}}(s)$ along the ray path $\mathbf{r}(s)$ by

$$n(\mathbf{r}, \dot{\mathbf{r}}\,)\dot{\mathbf{r}} = \mathbf{p}_3 - \mathbf{A}(\mathbf{r}, \dot{\mathbf{r}})\,, \tag{1.13.8}$$

whose norm is

$$|\mathbf{p}_3 - \mathbf{A}(\mathbf{r}, \dot{\mathbf{r}})| = n(\mathbf{r}, \dot{\mathbf{r}}\,) \quad \text{since} \quad \dot{\mathbf{r}} \cdot \mathbf{A} = 0\,. \tag{1.13.9}$$

Remark 1.13.1 *The anisotropy vector is orthogonal to the desired ray direction and is a prescribed function of it and the position along the ray path.*

Unfortunately, it is not possible to solve for the ray direction $\dot{\mathbf{r}}$, given the 3D optical momentum $\mathbf{p}_3$ and position $\mathbf{r}$. The 3D optical momentum decomposes conveniently into components which are parallel and perpendicular to $\dot{\mathbf{r}}$, as

$$\mathbf{p}_3 = n(\mathbf{r}, \dot{\mathbf{r}}\,)\dot{\mathbf{r}} + \mathbf{A}(\mathbf{r}, \dot{\mathbf{r}}) =: \mathbf{p}_{3\|} + \mathbf{p}_{3\perp}\,.$$

However, media for which these functional relations are nontrivial do not in general admit solutions for the tangent vector $\dot{\mathbf{r}}(s)$ as a function of $(\mathbf{r}(s), \mathbf{p}_3(s))$. Thus, the ray direction is not solvable in general from the optical momentum and ray path.[5] However, the 3D eikonal equation (1.13.5) still holds and so does its associated ***anisotropic Huygens wave front description***,

$$\frac{\partial S(\mathbf{r}, \dot{\mathbf{r}})}{\partial \mathbf{r}} = n(\mathbf{r}, \dot{\mathbf{r}}\,)\dot{\mathbf{r}} + \mathbf{A}(\mathbf{r}, \dot{\mathbf{r}})\,,$$

[5] This is an example of a singular, or degenerate, Lagrangian.

whose norm yields the ***scalar Huygens equation for anisotropic media***,

$$\left|\frac{\partial S}{\partial \mathbf{r}}\right|^2 = n^2(\mathbf{r}, \dot{\mathbf{r}}\,) + \left|\mathbf{A}(\mathbf{r}, \dot{\mathbf{r}})\right|^2 .$$

Remark 1.13.2 (Ibn Sahl-Snell law for anisotropic media) *The statement of the Ibn Sahl-Snell law relation at discontinuities of the refractive index in anisotropic media is rather more involved than for isotropic media. A break in the direction* $\widehat{\mathbf{s}}$ *of the ray vector is still expected at any finite discontinuity in the refractive index* $n = |\mathbf{n}|$ *encountered along the ray path* $\mathbf{r}(s)$. *According to the eikonal equation for anisotropic media (1.13.5) the jump (denoted by* Δ*) in 3D optical momentum across the discontinuity must satisfy the relation*

$$\Delta\mathbf{p}_3 \times \frac{\partial n}{\partial \mathbf{r}} = \Delta\bigg(n(\mathbf{r}, \widehat{\mathbf{s}}\,)\widehat{\mathbf{s}} + \mathbf{A}(\mathbf{r}, \widehat{\mathbf{s}})\bigg) \times \frac{\partial n}{\partial \mathbf{r}} = 0\,. \qquad (1.13.10)$$

This means the 2D projections of the 3D optical momenta $\mathbf{p}_3$ *and* $\mathbf{p}_3'$ *onto the plane of the discontinuity in refractive index will be invariant across the interface.*

Thus, preservation of the components of 3D optical momentum tangential to the discontinuity still holds, but a difficulty occurs because the ray direction and optical momentum are no longer colinear. Instead, they differ by the anisotropy vector, which is orthogonal to the desired ray direction and also depends as a prescribed function of ray direction on either side of the discontinuity.

The geometry for determining the refracted ray direction in an anisotropic medium thus becomes considerably more involved than the simple Ibn Sahl-Snell law of ray projection for isotropic media. There does exist a graphical construction (see, e.g., [Wo2004]), but its application in the Ibn Sahl-Snell law for construction of the break in ray direction at a discontinuity in refractive index in an anisotropic medium is problematic, unless the prescribed dependence of the anisotropy vector on the ray direction is rather simple.

An alternative argument

The loop integral argument in equations (1.1.16) - (1.1.18) reaches the same conclusion about the difficulty in determining the ray direc-

tions in general at an interface where the refractive index is discontinuous in an anisotopic medium. This argument proceeds by evaluating the loop integral of the Huygens phase,

$$\oint_P \nabla S(\mathbf{r}) \cdot d\mathbf{r} = \oint_P \Big(\mathbf{n}(\mathbf{r}) + \mathbf{A}(\mathbf{r}, \dot{\mathbf{r}}) \Big) \cdot d\mathbf{r} = 0 \,, \qquad (1.13.11)$$

taken around a closed path P that surrounds a boundary separating two different media. Letting the sides of the loop perpendicular to the interface shrink to zero implies that the tangential components of the momentum vectors must be preserved, in agreement with the previous argument. Consequently,

$$\Big(\big(\mathbf{n}(\mathbf{r}) + \mathbf{A}(\mathbf{r}, \dot{\mathbf{r}}) \big) - \big(\mathbf{n}'(\mathbf{r}) + \mathbf{A}'(\mathbf{r}, \dot{\mathbf{r}}) \big) \Big) \times \hat{\mathbf{z}} = 0 \,, \qquad (1.13.12)$$

in agreement with equation (1.13.10). If ψ and ψ' are the angles of incident and transmitted ***momentum directions****, measured from the normal $\hat{\mathbf{z}}$ through the boundary, then preservation of the tangential components of the momentum vector means that the momentum vectors must lie in the same plane and the angles ψ and ψ' must satisfy*

$$\sqrt{n^2 + A^2} \sin \psi = \sqrt{(n')^2 + (A')^2} \sin \psi' \,. \qquad (1.13.13)$$

Relation (1.13.13) determines the angles of incidence and transmission of the momentum directions. However, in general, it does not determine the ray directions. The ray directions are not invertible from the co-planar momentum directions, because the anisotropy vector in equation (1.13.7) shifts the ray vectors into different planes by an amount depending on the ray direction itself, not the momentum direction.

1.14 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, see Theorem 1.6.4.

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 1.5.3.

3. The Poisson bracket is associated geometrically with the Jacobian for canonical transformations in Section 1.11.2. 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 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 1.3.2, 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 (1.4.5). 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 (1.7.6). 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 1.9.2.

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 (1.11.17) 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 (1.4.5). 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 (1.10.1) and the transformation to invariant variables will be the momentum map associated as in Remark 1.9.4 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 in (1.7.13) is the Casimir for the Nambu bracket in $\mathbf{R}^3$, 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 1.8. 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 (1.12.12). 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. A Lagrangian may be singular; that is, it may be degenerate, as occurs in the example of Fermat's principle in anisotropic media discussed in Section 1.13. This means the velocity cannot be

solved from the momentum and its conjugate coordinate. Even so, the dynamics resulting from the Lagrangian formulation of the problem may still be well-defined, in the sense that the solutions may still exist for the resulting ordinary differential equations.