UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

MORSE FAMILIES IN ANISOTROPIC OPTICS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

MARTIN CARLSON
Norman, Oklahoma
2025

MORSE FAMILIES IN ANISOTROPIC OPTICS

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MATHEMATICS

BY THE COMMITTEE CONSISTING OF

Dr. Nikola P. Petrov, Chair

Dr. Pengfei Zhang, Co-Chair

Dr. Stacy Reeder

Dr. John Albert

© Copyright by MARTIN CARLSON 2025
All Rights Reserved.

Dedicated

to

My wife

and

My kids

Acknowledgments

When I began graduate school, I arrogantly predicted that I would finish my dissertation in five years. That did not happen, although if one considers modulo five years, then my predection was correct! Graduate school has been one of the most challenging experiences of my life, but it has also been one of the most rewarding. Enduring these many years helped me see what I'm capable of, and there are many people to thank.

First, I wouldn't have made it without the support of my advisor, Dr. Nikola Petrov. During my most difficult times, he could have washed his hands of me. However, he continously encouraged me when I needed it, and his high standards helped me finally get over the hump. I think I can say with confidence that if anyone other than Dr. Petrov were my advisor, I wouldn't have made it. So, thank you, Dr. Petrov!

I would also like to thank my co-advisor, Dr. Pengfei Zhang. As with Dr. Petrov, he would have been completely justified in walking away from our project when I was lagging far behind. However, his incredible knowledge of Finsler geometry and mathematics in general was instrumental in helping me break through and achieve a result. He has been so helpful from beginning to end in getting the dissertation to where it is, with much of his help coming remotely as well. Dr. Zhang's help was invaluable, and I cannot thank him enough.

Next, I want to thank Dr. Michael Jablonski. He was the first professor I had in graduate school who truly humbled me academically. While I hated it in the moment, the experience made me a much stronger mathematician. But on top of that, he has been a very trusted colleague in whom I have been able to confide many things over the years, and he is someone who has always had my back. Dr. Jablonski, thank you for everything you did for me, including the things you may not even realize you did!

I have also been blessed to work alongside so many incredible people in my role as an instructor. Ashley Berger, Ashley Taylor, Candace Andrews, Dustin Gaskins, Jonathan Lee, Dr. Deborah Moore-Russo, Haleigh Watson-Wilkes, Gary Barksdale, Christine Tinsley, Roshini Gallage, Carye Chapman, Malva Asaad, Cory Wilson, Beth Rawlins, Kayla Pierson, Casey Haskins, Hanah Goetz, Ling Jin, Wenwen Li, Erin Hausmann, Thomas Morgan, Mahesh Sunkula, Dania Sheaib, Jose Lozano, Nirjal Sapkota, and so many more: each of you have had a positive impact on me at some point over the years. All of you have helped me out at some point, and together, you helped me get to where I am now. From the bottom of my heart, thank you all for everything!

One person I would like to single out is an old friend, Ryan Reynolds. We began graduate school at the same time, and being the two oldest students in our cohort, we quickly became great friends. Ryan was — and still is — someone I could always lean on in tough times, and getting to know him and his family was a privilege I don't deserve. Breaking bread with him and his family was an incredible experience, and I am proud to call Ryan my friend. Thank you, Ryan, for being such a great friend!

I have also been blessed to see many of my former students from Carl Albert Middle School come through the University of Oklahoma over the years. Watching them grow and become professionals in their fields was an honor. So, to Adam Ward, Ashley Esser, Peter Vo, Ivan Ma, Garrett Davis, Austin Bullock, Connor Burns, Theo Stephens, Maddie and Ethan Sparks, Gavin Hamilton, Riley Pascual, and Paige Scott: thank you for letting me a small part of your lives!

Naturally, I would be remiss if I didn't say thank you to my beautiful wife, Sam, and to our three perfect kids: Jessica, Gordon, and Aaron. Simply attending graduate school would have been impossible without Sam's support; surviving graduate school would have been impossible without her love and encouragement. To my children: while you undoubtedly slowed down my progress, I wouldn't have it any other way. Every day, you gave me the inspriation I desperately needed to make it through, because I want to be the best dad I can be for you, and I want to show you that you can truly do anything you set your mind to. I love you! Thank you for loving me too!

Contents

List of Figures

Abstract

The main objects of study in this dissertation are Morse families, which are helpful for gaining an understanding of the propagation of light rays. Previous literature on Morse families limited discussion to isotropic media, where the speed of ray propagation is independent of the direction of propagation. We extend the discussion to anisotropic media using tools of Finsler geometry. We prove that in homogeneous media, where the speed of propagation is independent of the point in the medium, any Finsler metric F generates a Morse family. We also prove that in inhomogeneous, anisotropic media, any Finsler metric F which is a perturbation of the flat Euclidean metric is a Morse family. Additionally, Morse families can be used to determine the envelopes of a family of light rays in anisotropic media.

Chapter 1

Introduction and Literature Review

1.1 Physical Framework

1.1.1 Wave Phenomena in Nature

Wave phenomena in nature are described by the so-called *wave equation*. This is a second-order partial differential equation, with time t and the spatial coordinates $(x^1, ..., x^n)$ as independent variables, and the propagating quantity as the unknown function. Depending on the context, the propagating quantity may be light (in optics), seismic waves (in seismology), forest fires, etc. In this work, we will generally mean optical waves, i.e., electromagnetic fields propagating through space or through some medium.

In general, PDE's are difficult to explicitly solve, except in cases when the geometry of the problem is very simple (e.g., the source of the wave is a sphere, a straight line, or an infinitely long cylinder, and there is nothing to disturb the propagation of the wave). Additionally, the solutions of PDE's generally blow up in finite time. However, it is possible to construct geometrical methods as an alternate way of studying wave phenomena.

1.1.2 From Waves to Rays: the Hamilton-Jacobi Equation

In our work, we discuss the most popular approximate method in optics, called *geometric optics* or *ray optics*, as opposed to the full solution (i.e., *physical optics* or *wave optics*). In geometric optics, one assumes that light propagates along curves called *rays*. This approximation is quite good if the length scale of the optical system is considerably larger than the wavelength of the propagating wave. The following is a brief description of the derivation of the *eikonal equation*, a major focus of our work (see [22], Section 9.1 and [23], Section 12.7 for a detailed derivation).

Let Q be a manifold. Let c denote the speed of light, $n : Q \to \mathbb{R}$ be the refractive index of a medium, and $u : Q \times \mathbb{R} \to \mathbb{R}$ be an unknown function. The scalar wave equation is

$$\frac{n^2(x)}{c^2}\frac{\partial^2 u}{\partial t^2}(x,t) - \Delta u(x,t) = 0. \tag{1.1}$$

If we apply the ansatz $u(x,t) = U(x)e^{-i\omega t}$ in the case of monochromatic light, where ω is the frequency of light, then (1.1) becomes the so-called *Helmholtz equation*:

$$n^2(x)U(x) + \frac{1}{k^2}\Delta U(x) = 0, \tag{1.2}$$

where $k = \dfrac{\omega}{c}$ is the *wave number*. Now, if the length scale r is sufficiently larger than the wavelength of the propagating wave, we can study the Helmholtz equation in the framework of short wave analysis. We apply the ansatz

$$U(x,k) = a(x,k)e^{ikS(x)}, \tag{1.3}$$

where $a(x,k)$ describes the amplitude of the wave, and $S(x)$ is the phase. The function $S : Q \to \mathbb{R}$ is known as the *eikonal function*. By plugging in (1.3) into (1.2), we arrive at the following:

$$\left(\frac{1}{k^2}\Delta a + \frac{2i}{k}\nabla a \cdot \nabla S + \frac{ia}{k}\Delta S - a(\nabla S)^2 + n^2 a\right)e^{ikS} = 0. \tag{1.4}$$

The amplitude a varies comparitively slowly, on the length of the optical system. On the other hand, the phase S varies on the length scale of the wavelength $\lambda = \frac{2\pi}{k}$. So, we can expand $a(x,k)$ in powers of $\frac{1}{k}$:

$$a(x,k) = a_0(x) + \frac{1}{k}a_1(x) + \frac{1}{k^2}a_2(x) + \ldots \ .$$

Plugging this into (1.4) and comparing coefficients of $\frac{1}{k}$ yields, among infinitely many equations, the following two on which we focus:

- Order 0 in $\frac{1}{k}$: $(\nabla S)^2 = n^2(x)$; (Eikonal Equation)
- Order 1 in $\frac{1}{k}$: $\nabla S \cdot \nabla \ln a_0^2 + \Delta S = 0$. (Transport Equation)

The roughest approximation of the wave equation is the so-called *eikonal equation*, a first-order, non-linear PDE. It is a special case of the *Hamilton-Jacobi equation*, whose unknown function is called the *eikonal function*. In [25], Rund discusses the origin, motivation, and properties of the Hamilton-Jacobi equation in depth.

The eikonal equation in $\mathbb{R}^n$ is solved by using the method of characteristics. As we will show in Chapter 2, the characteristics are a $(2n-1)$-parameter family of curves which satisfy certain conditions, and they help determine the general solution to the eikonal equation. Taken together with some initial condition, the characteristics help to find the particular solution. In the geometric construction, the characteristics will weave together a solution surface with the help of the initial condition. This surface will satisfy certain conditions; such surfaces are *Lagrangian submanifolds* (see the definition in Section 2.2).

1.1.3 Fermat's Principle and the Optical Path

In 1662, Pierre de Fermat introduced a principle which bears his name: when traveling from one point to another, a light ray follows the path which minimizes time. Let c denote the speed of light in vacuum. In some medium (e.g., water, air, etc.), the speed v of propagation of light is smaller than c, and the ratio $\frac{c}{v}$, denoted by n, is called the *index of refraction*. Thus, the speed of propagation of light in a medium is $v = \frac{c}{n}$.

Let Q be an n-dimensional space, and let $x \in Q$ be a given point in our space. For $k < n$, consider a k-dimensional source of light $\mathscr{S}$ parameterized by $X(y) = \Big(X^1(y^1, ..., y^k), ..., X^n(y^1, ...y^k)\Big)$. Assume that the light ray emitted from a point $X(y)$ on the light source $\mathscr{S}$ goes to $x \in Q$, and let $M(x, y)$ stand for the *optical path* of this ray, i.e., the time it takes for light to get from $X(y)$ to x. Let us divide the ray into k subintervals of length Δs_i, and let the speed of light on the i-th subinverval be given by $v_i = \frac{c}{n_i}$. The, the total optical path is the sum of all the short time intervals of length $\frac{\Delta s_i}{v_i}$ needed for the light to traverse them, i.e.,

$$M(x, y) \approx \sum_{i=1}^{k} \frac{\Delta s_i}{v_i} = \frac{1}{c} \sum_{i=1}^{k} n_i \Delta s_i \,.$$

Taking $k \to \infty$, we get

$$M(x, y) = \frac{1}{c} \int n \, \mathrm{d}s \,. \tag{1.5}$$

This function plays an important role in our work. For now, note that for each

$x \in Q$, there may be multiple points $X(y)$ on the source $\mathscr{S}$ from which light can come to x.

1.1.4 Non-Homogeneous and Anisotropic Media

During the course of our work, we must consider the case where a medium may be non-homogeneous (speed of propagation depends on position) and/or anisotropic (speed of propagation depends on the direction of propagation).

First, consider the case of a non-homogeneous medium. While light rays are straight lines in empty space (or, more generally, if n is constant), light rays are curves in a non-homogeneous medium. This is the physical reason for seeing mirages while driving on a hot road — what we perceive as water on the road is actually the sky, because light rays are bent due to differing air temperatures above the road. Another example of the bending of light rays is far more interesting: the so-called *gravitational lensing* which occurs when light coming from a distant star is bent when there are massive objects (e.g., black holes, quasars, neutron stars, etc.) on the line connecting Earth and the distant star that bend the light (in accordance with the general theory of relativity). If the medium is non-homogeneous, the simple formula for the optical path $M(x, y)$ given by (1.5) becomes

$$M(x, y) = \int \sqrt{g_{ij}(\gamma(s))\dot{\gamma}^i(s)\dot{\gamma}^j(s)}\,\mathrm{d}s\,, \tag{1.6}$$

where γ is the path that light traverses from the source point $X(y) \in \mathscr{S}$ to the point $x \in Q$, the vector $\dot{\gamma} = \dfrac{d\gamma}{ds}$, and g_{ij} are the components of a Riemannian metric tensor. As before, light rays are the curves that minimize the optical path $M(x, y)$; such curves are *geodesics*.

In the case of an anisotropic medium, things become even more complicated.

Anisotropy isn't an uncommon property. For example, most crystals are anisotropic. Seismologists mostly work with anisotropic media; in fact, seismic waves are far more complicated to describe than light! Other simpler examples include a moving medium, in which light travels faster in the direction of motion than in the opposite direction, and even the propagation of a forest fire in the presence of wind. For anisotropic media, the equation for $M(x, y)$ from (1.6) becomes

$$M(x, y) = \int \sqrt{g_{ij}(\gamma(s), \dot{\gamma}(s))\dot{\gamma}^i(s)\dot{\gamma}^j(s)}\, \mathrm{d}s\,. \tag{1.7}$$

In this case, the metric tensor depends not only on the point $\gamma(s)$, but also on the tangent vector $\dot{\gamma}(s)$ of the ray passing through this point. Thus, the problem of finding the wave at a point x becomes much more challenging. The geometry underlying this concept of distance goes beyond Riemannian geometry. One requires *Finsler geometry*, named after Paul Finsler, who studied this geometry in 1918. In essence, Finsler geometry is a generalization of Riemannian geometry; said another way, Riemannian geometry is just a special case of Finsler geometry. In [12], Chern discusses precisely this topic. As stated above, metric tensors in Finsler geometry depend on both the point $x \in Q$ on a manifold as well as the velocity v at x, whereas in Riemannian geometry, the metric tensor doesn't depend on v. While Finsler geometry does have properties which are not particularly convenient — for example, there is no connection which is both metric-compatible and torsion-free like the Levi-Civita connection in Riemannian geometry — we can nonetheless define geodesics in a very similar manner to Riemannian geometry. More importantly, the unit sphere in the slit tangent bundle $TQ\backslash\{0\}$ (called the *indicatrix*) becomes a valuable tool for allowing us to write (1.7) in a more convenient way in certain situations.

Rund wrote one of the first big works on Finsler geometry in [24]. Rund discusses a select number of concepts of Finsler geometry in great depth, defining the indicatrix of a Finsler metric and its dual, the figuratrix. He also discusses geodesic equations and certain connections which arise in Finsler geometry. Several decades later, in [3], Bao et al gave a comprehensive view into all aspects of Finsler geometry. They define a Minkowski norm, then use this to define Finsler metrics. They give many examples, and then discuss many topics with analogues in Riemannian geometry, including connections, curvature, etc. We discuss many of these notions in Chapter 4.

1.1.5 Morse Families

As one can already see, propagation of waves in anisotropic, non-homogeneous media is not an easy problem. While in homogeneous, isotropic media (media in which the speed of propagation is independent of direction), the light rays are straight lines, computing the geodesics which minimize $M(x, y)$ in the general Finsler case is very non-trivial.

However, things are even more complicated than one might expect. When we wrote the expression for (1.7), we assumed that the light ray γ began at the source point $X(y) \in \mathscr{S}$ and ended at a given point $x \in Q$. However, if we are given an arbitrary $x \in Q$, finding a y - and thus a source point $X(y) \in \mathscr{S}$ - from which γ travels to reach x is already difficult on its own. Additionally, there may be more than one source point that emits a ray going to x. However, there is an elegant geometric way to deal with these issues: Morse families.

The notion of a Morse family has its mathematical roots with Hörmander [17] and Weinstein [27]. In [5], Benenti and Tulczyew use the notion of a Morse family for the first time in a physical setting. The construction is as follows:

Instead of working in strictly the n-dimensional space Q, we examine the larger $(n+k)$-dimensional space B, in which points $b = (x, y)$ are defined from a point $x \in Q$ and a point $X(y)$ on the source determined by $y \in \mathbb{R}^k$. Now, consider the optical path function $M(x, y)$ on this extended configuration space, where x is a given observation point. We would like to determine the curves which define light rays. First, we impose a condition that determines the points y from the source from which a light ray travels to x. According to variational principles, for a fixed $x \in Q$, this condition is

$$\frac{\partial M}{\partial y^\alpha}(x, y) = 0, \text{ for every } \alpha = 1, ..., k. \tag{1.8}$$

Next, we use these (x, y)-pairs to evaluate $p_{x,i} = \dfrac{\partial M}{\partial x^i}(x, y)$. This determines the (initial) direction of a light ray. Combining the $p_{x,i}$ with the fact that

$$p_{y,\alpha} := \frac{\partial M}{\partial y^\alpha}(x, y) = 0, \text{ for every } \alpha = 1, ..., k,$$

we can construct a surface in T^*B which we call L_M:

$$L_M = \left\{ \left(x, y; \frac{\partial M}{\partial x}(x, y), 0 \right) \in T^*B \, : \frac{\partial M}{\partial y^\alpha}(x, y) = 0, \; \forall \alpha = 1, ..., k \right\}.$$

The surface L_M encodes both the points (x, y) which determine light rays, as well as the direction of the light ray coming from the source $\mathscr{S}$ to $x \in Q$.

The final step is to "remove" the y- and p_y-coordinates from L_M, as these were artificially introduced in the first place. However, we cannot simply remove these coordinates. This is where the notion of *symplectic reduction* comes in; it allows us to project L_M to T^*Q in a well-defined manner. (In [4], Benenti discusses symplectic reductions in great depth, including as a category.) Having said that, it is essential that the projection satisfies certain conditions. For these conditions to be satisfied, the submanifolds $\mathrm{Im}\ \mathrm{d}M \subset T^*B$ and $\{p_y = 0\} \subset T^*B$ (to be named the *conormal bundle*) must intersect transversally. When this is satisfied, the optical math M is said to be a *Morse family.*

1.2 Potential Applications

To this point, Morse families have only been used in the isotropic case. However, as anisotropy is very common in natural settings, applying the concepts of Morse families to anisotropic settings is worthwhile. Even though our work focuses on optics, the methods we use can be applied to problems in many different areas, as ray and wavefront propagation occur in a variety of contexts.

One such area is seismology. In [26], Slawinski gives an in-depth discussion of ray and wave propagation in elastic continua. He also derives the eikonal equation from a different starting point than the scalar wave equation. Bona and Slawinski also describe these phenomena from a more strictly PDE point of view in [7]. In [9], Bucataru and Slawinski describe a generalized orthogonality of wavefronts and rays in anisotropic, inhomogeneous media. In [1], Antonelli, Bona, and Slawinski further demonstrate that seismic rays are, in fact, Finsler geodesics. Yajima and Nagahama build upon these results in [28] and [29], studying wavefront propagation in rock and fluid flow. Collaborating with Yamasaki in [30], they discuss, among other things, the relationship between Finsler geometry and wavefronts.

Another area is wildfire modeling. In [20], Markvorsen shows how the so-called *Richards' equations* for large-scale elliptic wildfire spreads have a simple Finsler geometric formulation. On the other hand, in [13], Dekhordi takes a different approach in modeling wildfire propagation. If one were to take a spark and drop it into dry grass with no wind, the grass would begin to burn in a slowly expanding circle. However, if there is wind, then the shape of the fire's front isn't a circle, but rather an elllipse. As it turns out, this is how one can describe the Huygens's Principle (i.e., the wavefront propagation) in a Randers space (i.e., an environment with wind). Dehkordi describes how to model wildfires using Huygens's Principle.

Gravitational theory and field theory are other areas of potential application. In [14] and [15], Frolov describes the propagation of high-frquency electromagnetic waves in curved spacetime, which may or may not be anisotropic. He also formulates a generalized Fermat's principle for light propagation in curved spacetime. In [21], Pfeifer defines a Finsler spacetime, as Finsler geometry occurs naturally in physics. In [16], Hohmann, Pfeifer, and Voicu build upon Pfeifer's definition of a Finsler spacetime to describe many aspects of field theory in a Finsler geometric context.

1.3 Plan of the Dissertation

In this dissertation, we establish the motivation for Morse families and their symplectic structure. We also discuss the tools of Finsler geometry necessary to expand the concept of Morse families to anisotropic media. Then, we show that in the case of media which is homogeneous but anisotropic, the optical path M is a Morse family. Furthermore, we show that even in the case of non-homogeneous media, M may still be a Morse family as long as the Finsler structure is a small perturbation of the Euclidean metric.

In Chapter 2, we begin by describing how one generally solves the Hamilton-Jacobi equation as a partial differential equation. Then, we give some symplectic preliminaries and discuss the geometric method of characteristics.

In Chapter 3, we discuss the construction of Morse families, discuss symplectic reduction and the transversality condition in detail, and show that the distance function in Euclidean space is a Morse family. We also show how one uses Morse families to find caustics.

In Chapter 4, we define Minkowski norms and Finsler structures on manifolds. Then, we discuss the utility of the indicatrix of a Finsler structure F, namely, its role in the speed of propagation in anisotropic media. We discuss the relationship Finsler geometry and the Legendre transform have with Huygens's principle. We show that light rays are F-orthogonal to wavefronts, and describe conditions light rays must satisfy. Furthermore, we discuss a particular type of metric — Randers metrics — and show that in homogeneous media, the indicatrix is always an ellipse with a focus at the origin.

In Chapter 5, we show that in spatially homogeneous media, any Finsler metric defines a Morse family. Furthermore, we demonstrate that in the general case (i.e., a medium which may be inhomogeneous and/or anisotropic), the optical path is still a Morse family if the Finsler structure is a small perturbation of the Euclidean metric. We also describe what types of optical paths this includes.

In Chapter 6, we give a few examples of Morse families in various contexts. We also show that in a Randers space, we can find the caustic for a light source by either (a) strictly using tools of Finsler geometry, or (b) using Morse families. We finish by giving a few examples.

Chapter 2

Symplectic Geometry and the Geometric Method of Characteristics

The eikonal equation can be solved as a first-order, nonlinear PDE. More interesting is obtaining the solution as a Lagrangian submanifold in the phase space. In order to do this, we will consider the geometric method of characteristics.

2.1 Motivation

The topics covered in this work revolve around solving the so-called eikonal equation

$$|\nabla S|^2 = n^2(x), \tag{2.1}$$

where $n : Q \to \mathbb{R}$ is the refractive index of the medium, and Q is a manifold. The eikonal equation is a first-order, non-linear PDE. To motivate the geometric method of characteristics, we must first understand how such a PDE is solved in general.

Consider the following Hamilton-Jacobi equation-type initial value problem for an unknown function $S : \mathbb{R}^n \to \mathbb{R}$:

$$H\left(x, \frac{\partial S}{\partial x}\right) = 0 \tag{2.2}$$

$$S|_D = S_0,$$

where $H : \mathbb{R}^{2n} \to \mathbb{R}$, $D \subset \mathbb{R}^{n-1}$ a hypersurface, and $S_0 : D \to \mathbb{R}$ is a given function. Let $\iota_D : D \to \mathbb{R}^n$ be an embedding (i.e., a homeomorphism onto its image). Let $\sigma = (\sigma^1, ..., \sigma^{n-1})$ be local coordinates in D, and let the function $x_0 : D \to \mathbb{R}^n$ be defined as

$$x_0(\sigma) := \iota_D(\sigma),$$

so that the parametric equation of D is

$$x = x_0(\sigma), \ \ \sigma \in D.$$

The IVP (2.2) becomes

$$H\left(x, \frac{\partial S}{\partial x}\right) = 0 \tag{2.3}$$

$$S(x_0(\sigma)) = S_0(\sigma).$$

To solve (2.3), we also need the values of the derivative $p := \dfrac{\partial S}{\partial x}$ on D, i.e., functions $p_0 : D \to \mathbb{R}^n$ such that

$$\frac{\partial S}{\partial x}(x_0(\sigma)) =: p_0(\sigma).$$

These p_i are determined by using the first equation of (2.3) on D, as well as taking the σ^μ derivatives of the second equation to obtain the $(n-1)$ identities

$$\frac{\partial S}{\partial x^i}(x_0(\sigma))\frac{\partial x_0^i}{\partial \sigma^\mu}(\sigma) = \frac{\partial S_0}{\partial \sigma^\mu}(\sigma), \quad \mu = 1, ..., n-1,$$

which gives us the n equations

$$H\left(x_0(\sigma), \frac{\partial S}{\partial x}(x_0(\sigma))\right) = 0 \tag{2.4}$$
$$p_{0,i}(\sigma)\frac{\partial x_0^i}{\partial \sigma^\mu}(\sigma) = \frac{\partial S_0}{\partial \sigma^\mu}(\sigma), \quad \mu = 1, ..., n-1.$$

We can now solve for the p_0 from these equations.

Now that we know the values of x, p, and S on D, we can solve the *characteristic system*, i.e., the system of $(2n+1)$ first-order ODE's given by

$$\begin{aligned} \frac{dx}{dt} &= H_p(x,p) \\ \frac{dS}{dt} &= p \cdot H_p(x,p) \\ \frac{dp}{dt} &= -H_x(x,p). \end{aligned} \tag{2.5}$$

Solve the system of $2n$ ODE's known as *Hamilton's Equations*,

$$\begin{aligned} \frac{dx}{dt} &= H_p(x,p) \\ \frac{dp}{dt} &= -H_x(x,p), \end{aligned} \tag{2.6}$$

with initial conditions

$$x(0) = x_0(\sigma) \tag{2.7}$$

$$p(0) = p_0(\sigma).$$

The geometric interpretation of the characteristic system is as follows:

- The initial conditions (2.7) can be considered as an $(n-1)$-dimensional surface

$$\tilde{D} := \left\{ (x,p) \in \mathbb{R}^{2n} : \, x = x_0(\sigma), p = p_0(\sigma), \sigma \in D \right\} \subset \mathbb{R}^{2n}.$$

- The general solution of (2.6) is a $(2n-1)$-parameter family of curves in $\mathbb{R}^{2n}$, known as the *characteristics of the Hamilton-Jacobi Equation*.

- The particular solution of the IVP (2.6), (2.7) consists of those characteristics which intersect $\tilde{D}$. Since dim $\tilde{D} = n-1$, this amounts to imposing $(n-1)$ conditions on the family of characteristics. Therefore, the general solution of the IVP is an n-dimensional surface $\mathscr{S} \in \mathbb{R}^{2n}$ "woven" by the characteristics. The characteristic

$$\big\{(x,p) = \big(x(t,\sigma), p(t,\sigma)\big) : t \in I \subset \mathbb{R}\big\} \tag{2.8}$$

  satisfies (2.6) and passes through the point $\big(x_0(\sigma), p_0(\sigma)\big) \in \tilde{D}$.

To find the value of S which solves the IVP (2.2), we let $\sigma \in D$ be fixed, and consider the characteristic which passes through the point $\big(x_0(\sigma), p_0(\sigma)\big) \in \tilde{D}$ at $t = 0$. Recall from (2.5) that S satisfies the ODE

$$\frac{dS}{dt} = p \cdot H_p(x,p) = p \cdot \frac{dx}{dt}$$

with initial condition

$$S\big(x_0(\sigma)\big) = S_0(\sigma).$$

Using (2.8), we can write the IVP for S as

$$\frac{dS}{dt} = p(t,\sigma) \cdot \frac{dx}{dt}(t,\sigma) \tag{2.9}$$

$$S|_{t=0} = S_0(\sigma), \tag{2.10}$$

whose solution can be written explicitly as

$$S(t,\sigma) = S_0(\sigma) + \int_0^t p(s,\sigma) \cdot \frac{dx}{dt}(s,\sigma)\, ds. \tag{2.11}$$

Finally, we express t and σ as functions of x from the equations $x = x(t,\sigma)$, and we plug these into $S(t,\sigma)$ from (2.11) to get $S(x) = S\big(t(x), \sigma(x)\big)$.

There are two main issues regarding solvability. First, for the particular solution of (2.6),(2.7) to exist, characteristics must intersect $\tilde{D}$ transversally. Geometrically, this means that the tangent vector to the characteristic at the point $\big(x_0(\sigma), p_0(\sigma)\big) \in \tilde{D}$, namely

$$\left(\frac{dx}{dt}(0), \frac{dp}{dt}(0)\right) = \Big(H_p\big(x_0(\sigma), p_0(\sigma)\big), -H_0\big(x_0(\sigma), p_0(\sigma)\big)\Big), \tag{2.12}$$

must not be in the tangent space to $\tilde{D}$ at this point, which is spanned by the $2n$ vectors

$$\frac{\partial x_0}{\partial \sigma^1}(\sigma), ..., \frac{\partial x_0}{\partial \sigma^n}(\sigma), \frac{\partial p_0}{\partial \sigma^1}(\sigma), ..., \frac{\partial p_0}{\partial \sigma^n}(\sigma). \tag{2.13}$$

This condition can be written as the maximality of the rank of the $2n \times n$ matrix whose entries are the components of the n vectors (2.12) and the $2n$ vectors (2.13) tangent to $\tilde{D}$. In physical problems, this condition is naturally satisfied.

The second, more serious issue is in regards to inverting the functions $x(t,\sigma)$ to obtain functions $t = t(x)$ and $\sigma = \sigma(x)$. Geometrically, the surface $\mathscr{S}$ must be projectible to the n-dimensional x-space. In general, this condition is often violated. The set points of $\mathscr{S}$ at which this condition is violated are called *caustics*.

The geometric treatment of the problem allows us to continue the solution of the Hamilton-Jacobi equation beyond the caustics.

Before continuing to the geometric construction, let us consider a specific type of Hamilton-Jacobi equation, namely the eikonal equation

$$H\left(x, \frac{\partial S}{\partial x}\right) = \left(\sum_{i=1}^{n} \left(\frac{\partial S}{\partial x^i}\right)^2\right) - n^2(x) \tag{2.14}$$

$$S|_D = S_0,$$

where D represents an $(n-1)$-dimensional source of light. The eikonal equation is naturally encountered as a 0-th order approximation to the wave equation. The solution to the eikonal equation can be interpreted as the minimal amount of time required to travel from a point $x(\sigma) \in D$ to a point x not on the source. Furthermore, the eikonal equation has the following physical interpretations:

- Owing to the *Fermat principle*, i.e., that light chooses the path that minimizes time, we can interpret the characteristic that passes through the points $x(\sigma) \in D$ and x as a *light ray*. In other words, a light ray is a solution to Hamilton's Equations (2.6).

- The level sets $S(x) = c$, $c \in \mathbb{R}$, describe surfaces of constant phase, called *wavefronts*.

2.2 Symplectic Preliminaries

We would like to generalize the method of characteristics in $\mathbb{R}^n$ described in the previous section to a general manifold Q. To do so, we need some preliminary facts and definitions.

Let (Q, ω) be a symplectic manifold (i.e., the two-form ω on Q is closed and non-degenerate). Let TQ denote the tangent bungle of Q, and T^*Q the cotangent bundle of Q, which we will sometimes refer to as the *phase space*. Let $E \subset TQ$ be a vector subbundle of TQ.

Definition 2.2.1. The *ω-orthogonal vector subbundle* E^ω is defined fiberwise as, for every $q \in Q$,

$$E_q^\omega := \Big\{ v \in T_q Q \,|\, \omega(q)(u, v) = 0, \forall u \in E_q \}. \tag{2.15}$$

▲

Definition 2.2.2. The vector subbundle $E \subset TQ$ is called *isotropic* if $E \subseteq E^\omega$, *coisotropic* if $E^\omega \subseteq E$, and *Lagrangian* if $E = E^\omega$. Alternatively, E is Lagrangian if and only if E is isotropic and its rank is half the dimension of Q. ▲

These notions of (co-)isotropy can be extended to submanifolds of a manifold as well.

Definition 2.2.3. A submanifold $P \subset Q$ is called isotropic (coisotropic, Lagrangian) if its tangent bundle is isotropic (coisotropic, Lagrangian) in the restriction $TQ|_P$ of the tangent bundle TQ to P. If we let $\iota : P \hookrightarrow Q$ denote the natural inclusion, a submanifold P of Q is isotropic if and only if $\iota^*\omega = 0$, and is Lagrangian if we additionally have dim $P = \frac{1}{2}$ dim Q. ▲

We can also extend these definitions to maps.

Definition 2.2.4. A *Lagrangian immersion* is a map $\phi : P \to Q$ satisfying the following: for all $p \in P$,

(1) the tangent map $T_p\phi : T_pP \to T_{\phi(p)}Q$ is injective, and

(2) $T\phi(T_pP)$ is a Lagrangian subspace of $(T_{\phi(p)}Q, \omega_{\phi(p)})$ in the sense of Definition 2.2.2.

▲

Remark 2.2.5. The word "Lagrangian" in the previous defintion can be replaced with "isotropic" and "coisotropic", but we will be solely interested in Lagrangian immersions. ▲

As we proceed, we will study the tangent bundle TP of a submanifold P. However, we will only be interested in a specific type of subset of TP which pertains to the method of characteristics.

Definition 2.2.6. A *distribution* on P is a subset $G \subset TP$ such that for every $p \in P$, the following hold:

(1) $G_p := G \cap T_pP$ is a linear subspace of T_pP,

(2) For every vector $v \in G_p$, there exists a vector field $X \in \mathfrak{X}^G(P)$ (vector fields in P taking values in G) such that $X_p = v$.

▲

Since we will shortly define the method of characteristics, we restrict our focus to a certain type of distribution, namely the characteristic distribution.

Definition 2.2.7. Let $\alpha \in \Omega^k(Q)$ be a k-form on Q.

(1) The subspace $F_q^\alpha := \ker \alpha_q \cap \ker (\mathrm{d}\alpha)_q$ of T_qQ is called the *characteristic subspace* of α at q. This subspace is of interest, as both α_q and $(\mathrm{d}\alpha)_q$ are invariant under F_q^α.

(2) A local vector field X on Q with domain U is called *characteristic* for α if $X_q \in F_q^\alpha$ for every $q \in U$.

(3) The distribution spanned by the characteristic local vector fields for α is called the *characteristic distribution* of α, denoted by D^α.

▲

As we will shortly show, characteristic distributions define foliations, i.e., they partition Q into submanifolds of lower dimension.

Remark 2.2.8. Note that $F^\alpha = \bigcup_{q \in Q} F_q^\alpha$ need not define a distribution, as its dimension need not remain constant. Rather, we should think of D^α as the maximal distribution on Q which is contained in F^α. However, if dim F_q^α is constant on Q, then F^α and D^α coincide. ▲

Definition 2.2.9. Let D^α be a distribution on a manifold Q. A connected submanifold (N, ι) of Q (where $\iota : N \hookrightarrow Q$ is the natural inclusion) is called an *integral manifold* of D^α through $q \in Q$ if $q \in \iota(N)$ and

$$T_p\iota(T_pN) = D^\alpha_{\iota(p)},$$

for all $p \in N$. Furthermore, D^α is said to be *integrable* if, for every $q \in Q$, there exists an integral manifold of D^α through q. ▲

In the language of foliations, the integral manifolds of D^α are referred to as the *leaves* of the foliation.

Now, let (Q, ω) be a symplectic manifold, $P \subset Q$ a coisotropic submanifold, and $\iota : P \hookrightarrow Q$ the natural inclusion. Since P is coisotropic, we have that $(TP)^\omega \subset TP$.

Proposition 2.2.10. *The space $(T_pP)^\omega$ coincides with the characteristic subspace $F_p^{\omega_P}$ of the two-form $\omega_P = \iota^*\omega$, where ι^* denotes the pull-back of ω to P.*

Proof. By the closedness of ω, we know that $\mathrm{d}\omega_P = 0$. By Definition 2.2.7,

$$\begin{aligned} F_p^{\omega_P} &= \ker{(\omega_P)_p} \cap \ker{(\mathrm{d}\omega_P)_p} \\ &= \ker{(\omega_P)_p} \\ &= T_pP \,\cap\, (T_pP)^\omega, \end{aligned}$$

by Definition 2.2.1. Since P is coisotropic, we get

$$F_p^{\omega_P} = (T_pP)^\omega.$$

□

Using the fact that dim (T_pP) + dim $(T_pP)^\omega$ = dim $T_{\iota(p)}Q$, we have that dim $(T_pP)^\omega$ = codim P. Hence, the dimension is independent of $p \in P$. Therefore, the characteristic distribution D^{ω_P} of ω_P is regular (i.e., its rank is constant), and $(TP)^\omega$ coincides with this distribution. We will call D^{ω_P} the characteristic distribution of P. Furthermore, by the Frobenius Theorem, we know that D^{ω_P} is, in fact, integrable. (A full proof can be found in [23], Proposition 4.2.20.) We call the integral submanifolds of D^{ω_P} *characteristics* of P.

2.3 Relating Hamiltonian Vector Fields and Distributions

2.3.1 Hamiltonian Vector Fields

When we consider the geometric method of characteristics, we will be working with the Hamilton-Jacobi equation in the phase space T^*Q of Q. Thus, we need to understand its associated vector field and how it relates to characteristics and characteristic distributions.

Definition 2.3.1. Let (Q, ω) be a symplectic manifold, $f \in C^\infty(Q)$. The vector field X_f satisfying $\omega(X_f, \cdot) = -df$ is called the *Hamiltonian vector field* generated by f. The flow of X_f is called the *Hamiltonian flow* generated by f. ▲

We can express X_f in local coordinates (q^i, p_i). Write $X_f = A_i \frac{\partial}{\partial p_i} + B^i \frac{\partial}{\partial q^i}$. Plugging this into $\omega = \mathrm{d}q^i \wedge \mathrm{d}p_i$ yields

$$\begin{aligned}\omega(X_f, \cdot) &= (\mathrm{d}q^i \wedge \mathrm{d}p_i)\left(A_i \frac{\partial}{\partial p_i} + B^i \frac{\partial}{\partial q^i}, \cdot\right) \\ &= A_i \mathrm{d}q^i - B^i \mathrm{d}p_i.\end{aligned}$$

On the other hand,

$$-df = -\frac{\partial f}{\partial q^i}\mathrm{d}q^i - \frac{\partial f}{\partial p_i}\mathrm{d}p_i.$$

By comparing coefficients, we obtain

$$X_f = \left(\frac{\partial f}{\partial p_i}\right)\frac{\partial}{\partial q^i} - \left(\frac{\partial f}{\partial q^i}\right)\frac{\partial}{\partial p_i}. \tag{2.16}$$

Furthermore, suppose $\gamma(t) = (q(t), p(t))$ is an integral curve of X_f. Then, it must satisfy the following equations, known as *Hamilton's equations*:

$$\dot{q}(t) = \frac{\partial f}{\partial p_i} \tag{2.17}$$

$$\dot{p}(t) = -\frac{\partial f}{\partial q^i}. \tag{2.18}$$

2.3.2 From Distributions to Hamiltonian Vector Fields

We now have all the tools necessary to connect the concept of a characteristic distribution with Hamiltonian vector fields.

Lemma 2.3.2. *Let (Q, ω) be a symplectic manifold, $P \subset Q$ be a coisotropic submanifold, and $f \in C^\infty(Q)$. The vector field X_f on Q restricts to a characteristic vector field on P only if f is locally constant on P.*

Proof. See [23], Lemma 8.5.4. □

Let (Q, ω) be a symplectic manifold of dimension $2n$. Consider a smooth function $H : T^*Q \to \mathbb{R}$ for which 0 is a regular value (i.e., for every $q \in H^{-1}(0)$, rank $(\mathrm{d}H)_q = 1$). Consider the level set $\mathscr{C} := H^{-1}(0)$. Notice that $\mathscr{C}$ is a coisotropic submanifold of T^*Q since it is of codimension 1. Characteristics of $\mathscr{C}$ are the integral manifolds of $D^{\omega_{\mathscr{C}}}$, which has rank 1. Notice that the Hamiltonian vector field X_H has no zeros on the smooth submanifold $\mathscr{C}$. By Lemma 2.3.2, we have the following:

Corollary 2.3.3. *The characteristic distribution $D^{\omega_{\mathscr{C}}}$ is spanned by X_H.*

Thus, the integral curves of X_H coincide with the characteristics of $\mathscr{C}$.

2.4 Geometric Method of Characteristics

Let Q be an n-dimensional manifold, D an embedded submaifold of dimension $k < n$, and $S_0 : D \to \mathbb{R}$ a smooth function. Given a smooth function $H : T^*Q \to \mathbb{R}$ for which 0 is a regular value, solving the Hamilton-Jacobi equation amounts to finding a smooth function $S : Q \to \mathbb{R}$ satisfying the following initial value problem:

$$\begin{aligned} H\left(x, \frac{\partial S}{\partial x}(x)\right) &= 0; \\ S|_D &= S_0. \end{aligned} \tag{2.19}$$

Solving this initial value problem via the method of characteristics involves the following steps:

(1) "Lift" the initial condition (D, S_0) into T^*Q, and take its intersection with $\mathscr{C} = H^{-1}(0)$. Call this submanifold $\mathscr{S}_0$.

(2) From $\mathscr{S}_0$, "weave" the solution surface by means of the flow of the Hamiltonian vector field X_H. This generates a Lagrangian immersion, and locally, this immersion induces Lagrangian submanifolds $\mathscr{S}$.

(3) If $\mathscr{S}$ intersects the fibers of T^*Q transversally (i.e., the intersection of their tangent spaces is empty) and at most once, then we obtain a local function S on Q which satisfies $\mathrm{d}S = \mathscr{S}$, and thus (2.19).

(4) It is possible that the surface could have folds, cusps, or other types of singularities. To determine the caustic of the IVP, we determine the critical points of the projection $\pi_Q : T^*Q \to Q$. Project this set of critical points to Q to obtain the caustic.

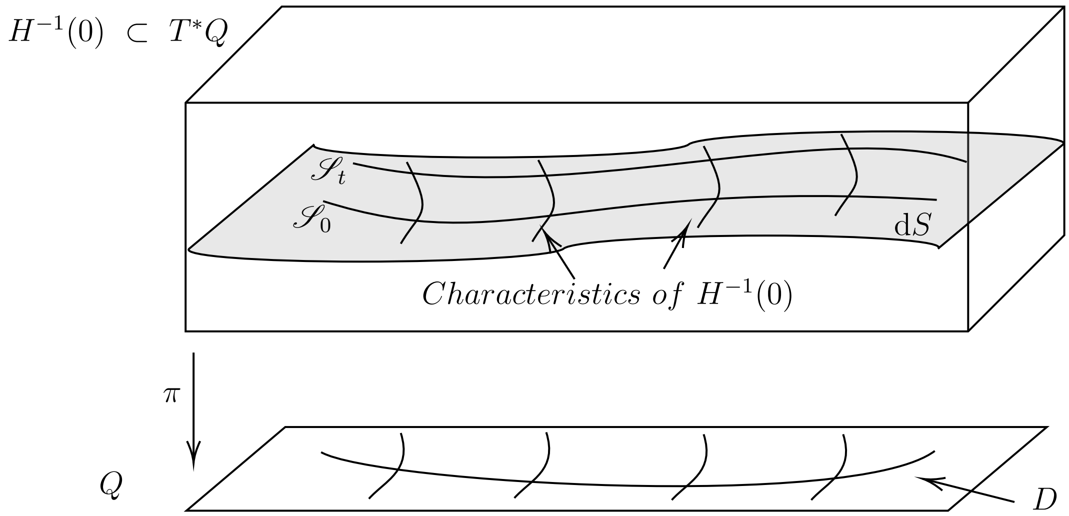

Figure 2.1: Continuation of $\mathscr{S}_0$ by the flow of X_H

2.4.1 Lift of the Initial Condition

Let (Q, ω) be a manifold with Liouville two-form ω. Take the pair (D, S_0). We wish to lift this canonically to a submanifold in T^*Q; denote this lift $\widehat{(D, S_0)}$, where

$$\widehat{(D, S_0)} := \Big\{ \xi \in (T^*Q)|_D : \langle \xi, X \rangle = \langle \mathrm{d}S_0, X \rangle, \, \forall X \in T_{\pi_Q(\xi)} D \Big\}. \tag{2.20}$$

Proposition 2.4.1. *The lift* $\widehat{(D, S_0)}$ *is a Lagrangian submanifold of* T^*Q.

Proof. Clearly, $\widehat{(D, S_0)}$ has half the dimension of T^*Q. So, we need to show that for $j : \widehat{(D, S_0)} \hookrightarrow T^*Q$ the natural inclusion, we have $j^*\omega = 0$.

Let $\pi_D : \widehat{(D, S_0)} \to D$ be the restriction of the projection. Let Θ be the Liouville 1-form on T^*Q. We first show that $j^*\Theta = \pi_D^*(\mathrm{d}S_0)$. Indeed, for any $X \in T_\xi \widehat{(D, S_0)}$,

$$\begin{aligned} (j^*\Theta)_\xi &= \Theta_{j(\xi)}(j_* X) \\ &= \Theta_\xi(X), \end{aligned}$$

since j is the inclusion (j_* denotes the push-forward of j). But by the definition of Θ at the point ξ,

$$\Theta_\xi(X) = \pi_D^*(\mathrm{d}S_0)_\xi(X), \tag{2.21}$$

see ([18], p.222). Thus, we have our desired equality. To show that $\widehat{(D, S_0)}$ is isotropic, we show that $j^*\omega = 0$, where ω is the canonical 2-form on T^*Q. By the smoothness of j and π_D, and by (2.21),

$$\begin{aligned}(j^*\omega) &= j^*(\mathrm{d}\Theta)\\ &= \mathrm{d}(j^*\Theta)\\ &= \mathrm{d}(\pi_D^*(\mathrm{d}S_0))\\ &= \pi_D^*(\mathrm{d}(\mathrm{d}S_0))\\ &= 0.\end{aligned}$$

Thus, $\widehat{(D, S_0)}$ is a Lagrangian submanifold of T^*Q. □

Now, consider $\mathscr{C} = H^{-1}(0)$. Recall that codim $\mathscr{C} = 1$, so $\mathscr{C}$ is a coisotropic submanifold of T^*Q. If the intersection $\mathscr{S}_0 := \mathscr{C} \cap \widehat{(D, S_0)}$ is transverse, we have that

$$\dim \mathscr{S}_0 = \dim \mathscr{C} + \dim \widehat{(D, S_0)} - \dim T^*Q = n - 1.$$

Note that $\mathscr{S}_0$ is an isotropic submanifold of T^*Q, since subspaces of isotropic subspaces are isotropic.

In partial differential equations, it is required that the characteristic curves be transverse to the initial condition. Therefore, since the characteristic curves in this context are defined by the integral curves of X_H, we require that $\mathscr{S}_0$ be transverse in $\mathscr{C} = H^{-1}(0)$ to the integral curves of X_H. In this case, we call $\mathscr{S}_0$ an *admissable* *initial condition.* Since X_H has no zeros on $\mathscr{C}$, we say that $\mathscr{S}_0$ is transverse to the integral curves of X_H if and only if, for all $p \in \mathscr{S}_0$,

$$T_p\mathscr{S}_0 \cap D_p^{\omega_{\mathscr{C}}} = \{0\}.$$

2.4.2 Weaving the Solution $\mathscr{S}$

Now that we have "lifted" the initial condition into $\mathscr{C} \subset T^*Q$, we use the flow of X_H to construct the solution surface $\mathscr{S}$.

Let Φ be the flow of X_H, let $\mathscr{D} \subset \mathbb{R} \times T^*Q$ denote the domain of Φ, and let $\Lambda = (\mathbb{R} \times \mathscr{S}_0) \cap \mathscr{D}$.

Theorem 2.4.2. *The restriction Ψ of Φ to Λ is a Lagrangian immersion.*

Proof. Clearly, Λ is an open subset of $\mathbb{R} \times \mathscr{S}_0$, and $\Psi : \Lambda \to T^*Q$ is smooth. Additionally, since $\Psi(0,p) = p$ for every $p \in \mathscr{S}_0$, $\Psi(\Lambda)$ contains $\mathscr{S}_0$. Tangent vectors at $(t,p) \in \Lambda$ are of the form $\left(\lambda \frac{d}{dt}, Y\right)$, where $\lambda \in \mathbb{R}$, $\frac{d}{dt}$ is the standard vector field on $\mathbb{R}$, and $Y \in T_p\mathscr{S}_0$. The tangent mapping is given by

$$T_{(t,p)}\Psi\left(\lambda\frac{d}{dt}, Y\right) = T_p(\Phi_t)(\lambda X_H + Y).$$

Since $T_p(\Phi_t)$ is a bijection, and since X_H has no zeros on $\mathscr{C}$, we have that

$$\ker T_{(t,p)}\Psi = \{0\}.$$

Hence, $T\Psi$ is injective, and hence Ψ is an immersion.

To show we have a Lagrangian immersion, we show isotropy. Let λ,$\mu \in \mathbb{R}$, X,

$Y \in T_p\mathscr{S}_0$.

$$(\Psi^*\omega)_{(t,p)}\left(\left(\lambda\frac{d}{dt}, X\right), \left(\mu\frac{d}{dt}, Y\right)\right)$$
$$= \omega_{\Phi_t(p)}\Big(T_p(\Phi_t)(\lambda X_H + X), T_p(\Phi_t)(\mu X_H + Y)\Big).$$

Since Φ_t is a symplectic map, this means

$$= \omega_p(\lambda X_H + X, \mu X_H + Y)$$
$$= \lambda\,\mu\,\omega_p(X_H, X_H) + \omega_p(X, Y),$$

by the bilinearity of ω. Since H is constant along the characteristics of X_H, we get that

$$\omega_p(X_H, X_H) = 0.$$

Finally, since $\mathscr{S}_0$ is isotropic,

$$\omega_p(X, Y) = 0.$$

Thus, we have our desired result. □

Let $\mathscr{S}_t := \psi_t(\mathscr{S}_0)$ denote the transformation of the initial condition $\mathscr{S}_0$ under the flow of X_H. Then, we define

$$\mathscr{S} = \bigcup_{t\in\mathbb{R}} \mathscr{S}_t.$$

2.4.3 Obtaining the Function S

The final step is to find the local function S satisfying (2.19). Let $\pi : T^*Q \to Q$ be the canonical projection.

Proposition 2.4.3. *Let $L \subset T^*Q$ be a Lagrangian submanifold which is transversal to the fibers and which intersects each fiber at most once. Then, $U := \pi(L)$ is open in Q, and there exists a closed 1-form α on U such that $L = \alpha(U)$. If L is also contractible, then there exists a smooth function $S : U \to \mathbb{R}$ such that $L = \mathrm{d}S(U)$.*

Proof. By the assumptions of tranversality and L intersecting the fibers at most once, the restriction $\pi|_L : L \to Q$ is injective. Hence, $U = \pi(L)$ is a submanifold of Q. Furthermore, since U has the same dimension as Q, U is open in Q, and $\pi|_L$ is a diffeomorphism onto U. If $\iota_L : L \hookrightarrow T^*Q$ is the natural inclusion mapping of L, then $\alpha = \iota_L \circ (\pi|_L)^{-1}$ is a local 1-form over U satisfying $L = \alpha(U)$. Since L is a Lagrangian submanifold, α must be closed. If L is contractible, then so is U. By the Poincaré Lemma (every closed form on a contractible manifold is exact), there exists a smooth function $S : U \to \mathbb{R}$ such that $\alpha = \mathrm{d}S$, and hence, $L = \mathrm{d}S(U)$. $\square$

The preceding proposition is critical to proving the key result of the method of characteristics:

Proposition 2.4.4. *Let $\mathscr{S}_0$ be an admissible initial condition for the initial value problem* (2.19)*, defined by the pair (D, S_0). Assume that D is contractible and that $\mathscr{S}_0$ is transversal to the fibers of T^*Q and intersects every fiber at most once. Let $\mathscr{S}$ be defined as above. Then, there exists a neighborhood U of D in Q and a smooth function $S : U \to \mathbb{R}$ such that $\mathrm{d}S \subset \mathscr{S}$.*

Proof. Since we assume $\mathscr{S}_0$ is transversal to the fibers of T^*Q and intersects each fiber at most once, then there exists an open neighborhood V of $\mathscr{S}_0$ in $\mathscr{S}$ for which this is still true. Since V is open in $\mathscr{S}$, it is a Lagrangian submanifold of T^*Q. Since $\mathscr{S}_0$ is contractible (since D is contractible), we can choose V to be contractible as well. Therefore, Proposition 2.4.3 yields the existence of U and S.

Since $\mathscr{S} \subset \mathscr{C} = H^{-1}(0)$, we have that $H \circ \mathrm{d}S = 0$, i.e., S is a solution to the Hamilton-Jacobi equation.

Next, we check the initial condition. Note that $\mathscr{S}_0 = \mathrm{d}S(D)$. On the other hand, by the definition of the canonical lift (2.20), we have that for every $\xi \in \mathscr{S}_0$ and for every $X \in T_{\pi_Q(\xi)}D$, we have

$$\langle \xi, X \rangle = \langle \mathrm{d}S_0, X \rangle.$$

Thus, for all $q \in D$ and for all $X \in T_qD$, $\langle \mathrm{d}S(x), X \rangle = \langle \mathrm{d}S_0(x), X \rangle$. Hence, $\mathrm{d}(S|_D) = \mathrm{d}S_0$. By adding an appropriate constant, we get $S|_D = S_0$. □

Remark 2.4.5 (Terminology)**.** We refer to the pair (Λ, Ψ) as a *generalized solution*, the Lagrangian submanifold $\mathscr{S}$ as a *geometric solution*, and the function S as an *analytic solution* to the initial value problem given in (2.19). ▲

Remark 2.4.6. In general, one does not always explicitly find the function S. Rather, we are interested in the existence of such a function. ▲

2.4.4 Example

Let $Q = \mathbb{R}^2$ endowed with the standard Euclidean metric. Suppose we wish to solve the following initial value problem for the vacuum eikonal equation,

$$\left(\frac{\partial S}{\partial x^1}\right)^2 + \left(\frac{\partial S}{\partial x^2}\right)^2 = 1 \tag{2.22}$$

$$S|_{x=y^2} = 0.$$

The corresponding Hamilon-Jacobi equation is given by $H : T^*Q \to \mathbb{R}$,

$$H(x,p) = |p|^2 - 1 = 0. \tag{2.23}$$

The Hamiltonian vector field X_H is given by

$$X_H(x,p) = 2p_i \frac{\partial}{\partial x^i}, \tag{2.24}$$

and Hamilton's equations read

$$\begin{aligned} \dot{x}^i &= 2p_i \\ \dot{p}_i &= 0. \end{aligned} \tag{2.25}$$

Given a point $(x_0, p_0) \in \mathscr{S}_0$, the integral curves of X_H are given by

$$x^i(t) = (x^i)_0 + 2(p_i)_0 t, \;\; p_i(t) = (p_i)_0. \tag{2.26}$$

Thus, the generalized solution $(\mathscr{S}, \Psi)$ is given by

$$\mathscr{S} = \mathscr{S}_0 \times \mathbb{R}, \quad \Psi(x,p,t) = (\mathbf{x}_0 + t\mathbf{p}_0, \mathbf{p}_0). \tag{2.27}$$

Now, we discuss the particular case where the light source (i.e., the initial condition) is given by the submanifold $D = \{(y^2, y) : y \in \mathbb{R}\} \subset \mathbb{R}^2$. Its canonical lift is given by

$$\widehat{(D, S_0)} = \big\{(y^2, y; p, -2py) : y, p \in \mathbb{R}\big\}, \tag{2.28}$$

and the initial condition $\mathscr{S}_0 \subset T^*Q$ is given by

$$\mathscr{S}_0 = \left\{ \left(y^2, y; \pm\frac{1}{\sqrt{1+4y^2}}, \mp\frac{2y}{\sqrt{1+4y^2}} \right) : y \in \mathbb{R} \right\}, \tag{2.29}$$

where the p_i-components of (2.28) and (2.29) are derived from the IVP (2.22). Thus, we can write the solution as $\mathscr{S} = \mathscr{S}_0^{\pm} \times \mathbb{R}$, and the image $\operatorname{Im}\Psi \subset T^*Q$ of the flow of X_H is given by

$$\Psi(y,t) = \left\{ \left(y^2 \pm \frac{t}{\sqrt{1+4y^2}}, y \mp \frac{2yt}{\sqrt{1+4y^2}}; \pm\frac{1}{\sqrt{1+4y^2}}, \mp\frac{2y}{\sqrt{1+4y^2}} \right) \right\}. \tag{2.30}$$

Furthermore, since Ψ is injective, we have that $\mathscr{S}$ is a geometric solution to the IVP.

To find the caustic, we find the critical points of the projection

$$\pi : \mathscr{S} \subset T^*Q \to Q.$$

We find the values for which

$$\det(D\pi) = \sqrt{1+4y^2} - \frac{2t}{1+4y^2} = 0.$$

Solving this for t yields

$$t = \frac{(1+4y^2)^{\frac{3}{2}}}{2}. \tag{2.31}$$

Define $\Sigma \subset T^*Q$ to be the set of critical points. Plugging (2.31) into (2.30) gives us

$$\Sigma = \left\{ \left(\frac{1}{2} + 3y^2, 4y^3; \pm\frac{1}{\sqrt{1+4y^2}}, \mp\frac{2y}{\sqrt{1+4y^2}} \right) : y \in \mathbb{R} \right\}, \tag{2.32}$$

and the caustic $\Gamma \subset Q$ is given by projecting Σ to Q:

$$\Gamma = \left\{ \left(\frac{1}{2} + 3y^2, 4y^3 \right) : y \in \mathbb{R} \right\}. \tag{2.33}$$

Chapter 3

Morse Families

In the method of characteristics, the flow of the Hamiltonian vector field generates a Lagrangian immersion in the phase space which satisfies the Hamilton-Jacobi equation. Locally, this immersion induces Lagrangian submanifolds in the phase space. These submanifolds are, locally, referred to as geometric solutions to the initial-value problem (see Chapter 2), and the functions $S : Q \to \mathbb{R}$ which generate these submanifolds are analytic solutions.

Morse families allow us to view these solutions in a different light. By adding extra dimensions to the configuration space Q, we arrive at a submanifold of an "extended" phase space (see Section 3.3) which contains all of the information about the solution (achieved by projecting this submanifold to the "true" phase space).

3.1 Motivation

The notion of a Morse family generalizes the concepts of the previous chapter. The ideas were first introduced by Hörmander in 1971 (see [17]). In physics, Morse families were used for the first time by Benenti and Tulczyew (see [5]).

The physical motivation for Morse families is as follows. Let Q be a manifold of dimension n, which we call the *configuration space*, and let $\mathscr{S}$ be a k-dimensional source of light which is emitting rays. Let $x \in Q$ be a point in the configuration space, and suppose $\mathscr{S}$ can be parameterized by $X(y) = X(y^1, ..., y^k)$, where $(y^1, ..., y^k) \in \mathbb{R}^k_{\text{loc}}$. Let $M(x, y)$ be the optical path from $X(y) \in \mathscr{S}$ to x, and let the constant c denote the speed of light in vacuum. Let n be the coefficient of refraction at x.

Suppose we want to find the time it takes to travel from a point $A = X(y)$ in the light source to $B = x$ in the configuration space. These points lie along some curve γ parameterized by s. We have:

$$
\begin{aligned}
\text{Time} &= M(x, y) \\
&= \lim_{i \to \infty} \sum_i \frac{\Delta s_i}{\frac{c}{n}} \\
&= \frac{1}{c} \lim_{i \to \infty} \sum_i n\, \Delta s_i \\
&= \frac{1}{c} \int_A^B n\, ds.
\end{aligned}
$$

A special case is that of vacuum, where $n \equiv 1$. Letting $c = 1$ for simplicity, $M(x, y)$ is just the distance from A to B.

According to variational principles, light chooses the path that takes the shortest time. In other words, light travels along the geodesics of Q. This leads us to the defining condition of minimizing the time it takes for a light ray emitted at $X(y) \in \mathscr{S}$ to reach some fixed $x^* \in Q$; namely, for a fixed x^*, the light rays coming to x^* satisfy

$$\frac{\partial M}{\partial y^\alpha}(x^*, y) = 0, \text{ for every } \alpha = 1, ..., k. \tag{3.1}$$

Remark 3.1.1. If Q is a Riemannian manifold, this condition is equivalent to the light rays being orthogonal to the source at the point of emission. ▲

Equations (3.1) allow us to do the following:

- Given some $x^* \in Q$, solve (3.1) for the value(s) of $y \in \mathbb{R}^k_{\text{loc}}$ which corresponds to a light ray from $X(y) \in \mathscr{S}$ to $x^* \in Q$ (possibly more than one).
- For these x^* and y, compute $\dfrac{\partial M}{\partial x^i}(x^*, y)$ for $i = 1, ..., n$. This will give us the direction of the light ray.

Consider T^*Q with coordinates (x, p_x). Consider the subset $\Lambda_M \subset T^*Q$ defined as

$$\Lambda_M := \left\{ (x, p_x) \in T^*Q : \exists y \in \mathbb{R}^k_{\text{loc}} \text{ such that } \frac{\partial M}{\partial y^\alpha}(x, y) = 0,\, p_{x,i} = \frac{\partial M}{\partial x^i}(x, y) \right\}.$$

This set gives us a nice physical representation of light rays: for each point $x \in Q$, the p_x gives the direction of the light ray(s) emanating from the source which pass through that point.

Unfortunately, T^*Q is too small of a space for this set to live in naturally. The reason for this is that the $p_{x,i}$ are multi-valued: for each $x \in Q$, there may be multiple $y \in \mathbb{R}^k_{\text{loc}}$ which correspond to it. Thus, for each $x \in Q$, there may be multiple $p_{x,i}$ associated with it. Thus, we require an "extended" space which takes the y into account, and thus eliminates the multi-valued issues.

Consider the $(n + k)$-dimensional product space $B := Q \times \mathbb{R}^k_{\text{loc}}$ with local coordinates (x, y), where y represents the parameters on the source $\mathscr{S}$. Consider

$\mathrm{d}M \in \Omega^1(Q)$. We can consider it as a map $\mathrm{d}M : B \to T^*B$. The submanifold

$$\mathrm{Im}\ \mathrm{d}M = \left\{ \left(x, y; p_{x,i} = \frac{\partial M}{\partial x^i}(x,y), p_{y,\alpha} = \frac{\partial M}{\partial y^\alpha}(x,y) \right) \right\} \subset T^*B$$

generated by the function $M : B \to \mathbb{R}$ is a graph over T^*B, as the $p_{x,i}$ and $p_{y,\alpha}$ are single-valued. It is also a Lagrangian submanifold of T^*B, as we will show in Section 3.3. Note that

$$\dim\left(\mathrm{Im}\,\mathrm{d}M\right) = n + k.$$

Now, intersect Im $\mathrm{d}M$ with the set $\{p_y = 0\}$ (which will be defined later as the *conormal bundle*, which is a canonically well-defined submanifold of T^*B). Call this intersection L_M, where in local coordinates on T^*B,

$$L_M = \left\{ \left(x, y; \frac{\partial M}{\partial x}(x,y), 0 \right) \in T^*B \; : \frac{\partial M}{\partial y^\alpha}(x,y) = 0,\ \forall \alpha = 1, ..., k \right\}.$$

Note that since there are k conditions on the x and y in L_M, we have

$$\dim L_M = (n + k) - k = n.$$

Now, we project L_M to T^*Q, giving us a set which gives us a more well-defined depiction of the pair (point, direction of ray); namely, it is the projection of a well-defined set in T^*B.

We also consider the set $B_M \subset B$ defined by

$$B_M = \left\{ (x,y) \in B : \frac{\partial M}{\partial y^\alpha}(x,y) = 0,\ \forall \alpha = 1, ..., k \right\}.$$

This set has a nice physical meaning in homogeneous media. When the light rays are straight lines, B_M describes the pairs (configuration point, source point) which define a light ray. Thus, in homogeneous media, one can think of B_M as the collection of light rays.

Our goal is to derive a coordinate-free method for the above construction. To do this, the main requirement is that the intersection $\mathrm{Im}\ \mathrm{d}M \cap \{p_y = 0\}$ is transverse. When this is the case, we refer to M as a *Morse family*.

3.2 Deriving a Coordinate-Free Construction

In this section and the next, we develop the coordinate-free foundation required for our construction.

3.2.1 Some Concepts of Linear Algebra

Let B be a linear space and $A \subseteq B$ be a subspace of B. We have the natural short exact sequence

$$0 \longrightarrow A \longrightarrow B \longrightarrow NA \longrightarrow 0\ , \tag{3.2}$$

where $NA := B/A$. The dual sequence to (3.2) is

$$0 \longleftarrow A^* \longleftarrow B^* \longleftarrow N^*A \longleftarrow 0\ , \tag{3.3}$$

where $N^*A := (NA)^* = (B/A)^*$ is the *conormal subspace*.

What is the general form of a $\xi \in N^*A$? Let $b \in B$, $[b] := b + A \in B/A$. For the natural paring $\langle \xi, [b] \rangle = \langle \xi, b + A \rangle$ to be well-defined, we have to have

$$\langle \xi, a \rangle = 0 \quad \forall a \in A\ .$$

Define the *annihilator* of A in B:

$$A^\circ := \operatorname{Ann} A := \{\eta \in B^* \,:\, \langle \eta, a \rangle = 0 \ \forall a \in A\} \subseteq B^* \ . \tag{3.4}$$

Hence, there is a natural isomorphism

$$N^* A \simeq \operatorname{Ann} A \ . \tag{3.5}$$

3.2.2 Conormal bundle

Let B and Q be manifolds of dimensions $\dim B = n + k$, $\dim Q = n$. Let

$$\pi : B \to Q \tag{3.6}$$

be a submersion (i.e., $T_b\pi : T_bB \to T_{\pi(b)}Q$ be surjective $\forall b \in B$). The submersion π defines a foliation Π of the k-dimensional fibers (level sets) of π. For $b \in B$, let $\Pi_b := \pi^{-1}(\pi(b)) \subseteq B$ be the fiber through b. Let

$$T_b\Pi := T_b\Pi_b := \ker T_b\pi \subseteq T_bB \ , \qquad N_b\Pi := N(T_b\Pi_b) \simeq T_bB/T_b\Pi_b$$

be respectively the tangent space at the point b to the fiber Π_b through b, and its quotient space in T_bB. We have short exact sequences analogous to (3.2) and (3.3)

$$0 \longrightarrow T_b\Pi \longrightarrow T_bB \longrightarrow N_b\Pi \longrightarrow 0 \ ,$$

$$0 \longleftarrow T_b^*\Pi \longleftarrow T_b^*B \longleftarrow N_b^*\Pi \longleftarrow 0 \ .$$

From (3.4) and (3.5), we have

$$N_b^*\Pi = \{\beta_b \in T_b^*B \,:\, \langle \beta_b, T_b\Pi\rangle = 0\}\ . \tag{3.7}$$

Since $\langle \beta_b, T_b\Pi\rangle = 0$ is equivalent to $k = \dim T_b\Pi$ independent conditions, (3.7) implies that

$$\dim N_b^*\Pi = (n+k) - k = n\ .$$

Additionally, we have the following dimensions of linear spaces:

- $\dim T_b\Pi = \dim T_b^*\Pi = k$,
- $\dim T_bB = \dim T_b^*B = n+k$,
- $\dim N_b\Pi = \dim N_b^*\Pi = n$.

Let us introduce local coordinates in B that are adapted to the submersion π (3.6). Let $x = (x^i) = (x^1, \ldots, x^n)$ be local coordinates in Q, and $(x, y) = (x^i, y^\alpha) = (x^1, \ldots, x^n, y^1, \ldots, y^k)$ be coordinates in B. In these coordinates, the submersion π (3.6) is given by $\pi(x, y) = (x)$, and, if $\tilde{b} = (\tilde{x}, \tilde{y}) \in B$, then the fiber through $\tilde{b}$ is

$$\Pi_{\tilde{b}} = \{(\tilde{x}, y) \,:\, \tilde{x} \in Q\}\ .$$

Having identified the local neighborhoods with their coordinates, let us write $X = (X^i) \in \mathbb{R}^n$, $Y = (Y^\alpha) \in \mathbb{R}^k$, and $b = (x, y) \in B$. Then, the general form of a vector from T_bB is

$$(x, y; X, Y) = X\left(\frac{\partial}{\partial x}\right)_b + Y\left(\frac{\partial}{\partial y}\right)_b := X^i\left(\frac{\partial}{\partial x^i}\right)_{(x,y)} + Y^\alpha\left(\frac{\partial}{\partial y^\alpha}\right)_{(x,y)} \in T_{(x,y)}B\ .$$

The tangent lift $T_b\pi$ of the submersion π is given by

$$T_b\pi : T_bB \to T_{\pi(b)}Q : X\left(\frac{\partial}{\partial x}\right)_b + Y\left(\frac{\partial}{\partial y}\right)_b \mapsto X\left(\frac{\partial}{\partial x}\right)_{\pi(b)} .$$

Therefore, the general form of a vector from $T_b\Pi = \ker T_b\pi$ is

$$(x, y; 0, Y) = Y\left(\frac{\partial}{\partial y}\right)_b \in T_b\Pi ,$$

and the general form of a vector from $N_b\Pi = T_bB/T_b\Pi$ is

$$(x, y; X, *^k) = X\left(\frac{\partial}{\partial x}\right)_b + \operatorname{span}\left\{\left(\frac{\partial}{\partial y}\right)_b\right\} \in N_b\Pi ,$$

where $*^k$ represents all vectors from $\mathbb{R}^k$.

Let $p_x = (p_{x,i}) \in \mathbb{R}^n$ and $p_y = (p_{y,\alpha}) \in \mathbb{R}^k$. Then, the general form of a covector from T_b^*B is

$$(x, y; p_x, p_y) = p_x(\mathrm{d}x)_b + p_y(\mathrm{d}y)_b \in T_bB .$$

The general form of a covector from $T_b^*\Pi = (T_b\Pi)^*$ is

$$(x, y; *^n, p_y) = p_y(\mathrm{d}y)_b + \operatorname{span}\{(\mathrm{d}x)_b\} \in T_b^*\Pi ,$$

(where $*^n$ stands for all vectors from $\mathbb{R}^n$), and the general form of a covector from $N_b^*\Pi = (N_b\Pi)^* = (T_bB/T_b\Pi)^*$ is

$$(x, y; p_x, 0) = p_x(\mathrm{d}x)_b \in N_b^*\Pi .$$

Let $N^*\Pi$ stand for the *conormal bundle* over B, i.e., the disjoint union of the linear spaces $N_b^*\Pi$ for all $b \in B$ (with transition functions induced by the transition functions of T^*B). The conormal bundle $N^*\Pi$ is a vector bundle over B; it is a subbundle of T^*B (see [23], Remark 2.7.18). The dimension of the fibers $N_b^*\Pi$ of $N^*\Pi$ is n, and the dimension of $N^*\Pi$ as a submanifold of T^*B is

$$\dim N^*\Pi = (n+k) + n = 2n + k \ . \tag{3.8}$$

It is important to note that $N^*\Pi$ is a coisotropic submanifold of T^*B. The proof is simple. Let

$$\omega_B = \mathrm{d}x \wedge \mathrm{d}p_x + \mathrm{d}y \wedge \mathrm{d}p_y \in \Omega^2(T^*B)$$

be the canonical symplectic form on T^*B. Take any point $b \in B$, and take $\alpha_b \in N_b^*\Pi$. If $(T_{\alpha_b} N^*\Pi)^{\omega_B}$ stands for the ω_B-orthogonal complement of $T_{\alpha_b} N^*\Pi$ in $T_{\alpha_b} T^*B$, then $(T_{\alpha_b} N^*\Pi)^{\omega_B} \subseteq T_{\alpha_b} N^*\Pi$. In coordinates: let $b = (x, y) \in B$, $\alpha_b = (x, y; p_x, 0) \in N_b^*\Pi$,

$$T_{\alpha_b} N^*\Pi = \operatorname{span}\left\{ \left(\frac{\partial}{\partial x}\right)_{\alpha_b}, \left(\frac{\partial}{\partial y}\right)_{\alpha_b}, \left(\frac{\partial}{\partial p_x}\right)_{\alpha_b} \right\} \ . \tag{3.9}$$

Hence,

$$(T_{\alpha_b} N^*\Pi)^{\omega_B} = \operatorname{span}\left\{ \left(\frac{\partial}{\partial y}\right)_{\alpha_b} \right\} \subseteq T_{\alpha_b} N^*\Pi \ .$$

3.3 Morse families

Let $M : B \to \mathbb{R}$ be a smooth function. Its exterior derivative, in local adapted coordinates, is

$$\mathrm{d}M = \frac{\partial M}{\partial x}\,\mathrm{d}x + \frac{\partial M}{\partial y}\,\mathrm{d}y \in \Omega^1(B) \ .$$

We can think of $\mathrm{d}M$ as a map $\mathrm{d}M : B \to T^*B$. In the local adapted coordinates, the image $\mathrm{Im}\,\mathrm{d}M \subseteq T^*B$, has the form

$$\left(x, y; \frac{\partial M}{\partial x}(x,y), \frac{\partial M}{\partial y}(x,y)\right) = \frac{\partial M}{\partial x}(x,y)\,(\mathrm{d}x)_b + \frac{\partial M}{\partial y}(x,y)\,(\mathrm{d}y)_b \ ,$$

where $b = (x, y) \in B$. Clearly, the dimension of $\mathrm{Im}\,\mathrm{d}M$ is the same as the dimension of B, namely, $\dim(\mathrm{Im}\,\mathrm{d}M) = n + k$.

An important fact to remember is that Im $\mathrm{d}M$ is a Lagrangian submanifold of T^*B, as it is the image of an exact 1-form $\alpha = \mathrm{d}M$. In more detail, we know that for the canonical 2-form ω_B on T^*B, $(\mathrm{d}M)^*\omega_B = \mathrm{d}(\mathrm{d}M) = 0$. Thus, Im $\mathrm{d}M$ is isotropic, and since its dimension is $n + k$, it is Lagrangian.

A recurring condition for our construction is that the submanifolds $N^*\Pi$ and $\mathrm{Im}\mathrm{d}M$ intersect transversally in T^*B. From this point forward, our definition of transversality will differ from the one we used in Chapter 2. Specifically, going forward, $N^*\Pi \pitchfork \mathrm{Im}\,\mathrm{d}M$ means that, at every point $\eta \in \mathrm{Im}\,\mathrm{d}M \cap N^*\Pi \subseteq T^*B$,

$$T_\eta N^*\Pi \oplus T_\eta \mathrm{Im}\,\mathrm{d}M = T_\eta T^*B \ .$$

Both the conormal bundle $N^*\Pi$ and $\mathrm{Im}\,\mathrm{d}M$ are submanifolds of the $2(n + k)$-dimensional manifold T^*B; recall that $\dim N^*\Pi = 2n + k$ (3.8). By counting dimensions, we see that $\dim(N^*\Pi \cap \mathrm{Im}\,\mathrm{d}M)$ can vary

- from n, in the case $N^*\Pi \pitchfork \mathrm{Im}\,\mathrm{d}M$, i.e., when $N^*\Pi$ is transversal to $\mathrm{Im}\,\mathrm{d}M$ in T^*B,
- to $n + k$, in the case $\mathrm{Im}\,\mathrm{d}M \subseteq N^*\Pi$.

Definition 3.3.1. Let $\pi : B \to Q$ be a submersion, and $M : B \to \mathbb{R}$ be a smooth function. The triple (B, π, M) is called a *Morse family* along π if $N^*\Pi \pitchfork \operatorname{Im} \mathrm{d}M$ (in T^*B). ▲

Hereafter assume that (B, π, M) is a Morse family.

Definition 3.3.2. Define the submanifold $L_M \in T^*B$ by

$$L_M := N^*\Pi \cap \operatorname{Im} \mathrm{d}M \subseteq T^*B. \tag{3.10}$$

Note that L_M is an isotropic submanifold of T^*B, since it a submanifold of the isotropic (in fact, Lagrangian) submanifold $\operatorname{Im} \mathrm{d}M$. Note also that $\dim L_M = n$. ▲

Definition 3.3.3. The *fiber-critical submanifold* B_M of B is defined as follows: if $\pi_B : T^*B \to B$, then

$$B_M := \pi_B(L_M) \subseteq B.$$

▲

Note that $\dim B_M = n$. In local adapted coordinates, L_M (3.10) is defined by the conditions

$$p_x = \frac{\partial M}{\partial x}(x, y) \ , \quad p_y = \frac{\partial M}{\partial y}(x, y) = 0 \ .$$

Therefore,

$$L_M = \left\{ \left(x, y; \frac{\partial M}{\partial x}(x, y), 0 \right) \in T^*B \ : \ (x, y) \in B \text{ with } \frac{\partial M}{\partial y}(x, y) = 0 \right\} \ , \tag{3.11}$$

and the fiber-critical submanifold B_M is given by

$$B_M = \left\{ (x, y) \in B \ : \ \frac{\partial M}{\partial y}(x, y) = 0 \right\} \ . \tag{3.12}$$

We see that

$$\dim B_M = \dim L_M = \dim Q = n \ .$$

3.4 Generating Lagrangian Submanifolds

Now that we have defined the various submanifolds essential to our construction, we demonstrate the reason for our construction. Namely, we will construct a map which — taking advantage of a process called *symplectic reduction* — generates the same submanifolds in T^*Q generated by the geometric method of characteristics in Chapter 2. Thus, our construction gives us an alternate method of solving the Hamilton-Jacobi equation.

3.4.1 Constructing the Map Λ_M

We begin by constructing the map which will ultimately give us the Lagrangian submanifold in T^*Q which will be a generalized solution to the Hamilton-Jacobi equation.

First, we need to address a few technical details. Let B and Q be arbitrary manifolds (later, they will be the same B and Q we have already discussed), and let $\pi : B \to Q$ be a submersion. Let $\pi_B : T^*B \to B$ and $\pi_Q : T^*Q \to Q$ be canonical projections, and let Θ_B and Θ_Q be the Liouville 1-forms in T^*B and T^*Q, respectively. Define the pull-back bundle

$$\pi^*(T^*Q) := \Big\{(b, \xi) \in B \times T^*Q : \pi(b) = \pi_Q(\xi)\Big\}. \tag{3.13}$$

The following can be shown about $\pi^*(T^*Q)$ (see [11], Proposition 1 and [23], section 12.4):

- The bundle $\pi^*(T^*Q)$ is a vector bundle over B with projection

$$\rho : \pi^*(T^*Q) \to B, \quad \rho(b, \xi) = b,$$

  and fibers $\rho^{-1}(b) = T^*_{\pi(b)}Q$.

- The bundle $\pi^*(T^*Q)$ is an embedded submanifold of $B \times T^*Q$.

- The natural projection $B \times T^*Q \to T^*Q$ restricts to a vector bundle morphism

$$\phi : \pi^*(T^*Q) \to T^*Q, \quad \phi(b, \xi) = \xi$$

  which covers π,

$$\pi_Q \circ \phi = \pi \circ \rho.$$

Since π is a submersion, we have that ker $T\pi$ is a vertical subbundle of TB (i.e., it is a regular distribution on B), and its annihilator $(\ker T\pi)^0 = N^*\Pi$ is a vertical subbundle of T^*B.

The following lemma demonstrates the relationship between the pull-back bundle and the conormal bundle (proof can be found in [23], Lemma 12.4.1):

Lemma 3.4.1. *The mapping $\psi : \pi^*(T^*Q) \to T^*B$, $\psi(b, \xi) := (T_b\pi)^T(\xi)$ is an injective vertical (i.e., fiber-preserving, fiberwise linear, and satisfying $\pi_B \circ \psi = \rho$) vector bundle morphism, with image $N^*\Pi$, and satisfying $\psi^*\Theta_B = \phi^*\Theta_Q$. Furthermore, ψ is an embedding, and thus, $\pi^*(T^*Q) \simeq N^*\Pi$ as vector bundles.*

Now, since ψ restricts to an isomorphism $\tilde{\psi} : \pi^*(T^*Q) \to N^*\Pi$, there exists a unique vector bundle morphism $\lambda : N^*\Pi \to T^*Q$ such that

$$\phi = \lambda \circ \tilde{\psi}. \tag{3.16}$$

$$
\begin{tikzcd}
T^*B \arrow[ddr, "\pi_B"'] & & & T^*Q \arrow[dd, "\pi_Q"] \\
& N^*\Pi \arrow[ul, "\iota"] \arrow[urr, "\lambda"] \arrow[d, "\pi_B|_{N^*\Pi}"] & \pi^*(T^*Q) \arrow[l, "\widetilde{\psi}"] \arrow[ull, "\psi"'] \arrow[ur, "\phi"] \arrow[dl, "\rho"] & \\
& B \arrow[rr, "\pi"] & & Q
\end{tikzcd}
\tag{3.14}
$$

$$
\begin{tikzcd}
\{(x,y;p_x,p_y)\} \arrow[ddr, "\pi_B"'] & & & \{(x;p_x)\} \arrow[dd, "\pi_Q"] \\
& \{(x,y;p_x,0)\} \arrow[ul, "\iota"] \arrow[urr, "\lambda"] \arrow[d, "\pi_B|_{N^*\Pi}"] & \{(x,y;p_x)\} \arrow[l, "\widetilde{\psi}"] \arrow[ull, "\psi"'] \arrow[ur, "\phi"] \arrow[dl, "\rho"] & \\
& \{(x,y)\} \arrow[rr, "\pi"] & & \{(x)\}
\end{tikzcd}
\tag{3.15}
$$

Figure 3.1: The important maps in the construction of a Morse family

Locally, the morphism λ has the form

$$\lambda(x, y, p_x, 0) = (x, p_x).$$

Recall that

$$L_M = \left\{ \left(x, y; \frac{\partial M}{\partial x}(x, y), 0 \right) : (x, y) \in B_M \right\} \subset T^*B.$$

Let $\lambda_M : L_M \to T^*Q$ be the restriction of λ to L_M, and define

$$\Lambda_M := \lambda_M \circ (dM)|_{B_M} : B_M \to T^*Q. \tag{3.17}$$

In local coordinates, we have

$$\Lambda_M(x, y) = \left\{ \left(x; \frac{\partial M}{\partial x}(x, y) \right) : (x, y) \in B_M \right\} \subset T^*Q. \tag{3.18}$$

The image of Λ_M is what will ultimately become the solution surface we seek.

3.4.2 Symplectic Reduction

Symplectic reduction is a key component of the construction of Morse families. An in-depth investigation into symplectic reduction can be found in [4], [5], [11], [19], and [27].

In our work, when we projected from $L_M \in T^*B$ to $\operatorname{Im} \Lambda_M \in T^*Q$ via λ_M, we essentially "forgot" the y and p_y components. However, it turns out that the morphism λ from (3.16) is not simply a projection. In fact, it has certain properties which result in its image being precisely the solution surface found by using the method of characteristics!

Definition 3.4.2. Let (T^*B, ω_B) be a symplectic manifold, N a submanifold of T^*B where $\iota : N \hookrightarrow T^*B$ is the natural inclusion, and (T^*Q, ω_Q) another symplectic manifold. A mapping $\kappa : N \to T^*Q$ is called a *symplectic reduction* if it is a surjective submersion satisfying

$$\kappa^* \omega_Q = \iota^* \omega_B. \tag{3.19}$$

Furthermore, the symplectic reduction is called *strict* if N is coisotropic. The manifold (T^*Q, ω_Q) is called a *reduced* symplectic manifold for N. ▲

Note that the above definition applies to general symplectic manifolds.

Proposition 3.4.3. *The morphism $\lambda : N^*\Pi \to T^*Q$ is a strict symplectic reduction.*

Proof. By Lemma 3.4.1, we have that $\psi^*\Theta_B = \phi^*\Theta_Q$. From the decomposition of ϕ in (3.16), we get

$$\begin{aligned} \psi^*\Theta_B &= (\lambda \circ \tilde{\psi})^*\Theta_Q \\ &= \tilde{\psi}^* \circ \lambda^*(\Theta_Q). \end{aligned}$$

But this means that $\tilde{\psi}^* \circ j^*(\Theta_B) = \tilde{\psi}^* \circ \lambda^*(\Theta_Q)$. Since ψ is a diffeomorphism, we conclude that

$$\lambda^*(\Theta_Q) = j^*(\Theta_B). \tag{3.20}$$

Taking the exterior derivative of both sides yields the definition of a symplectic reduction λ. The strictness is immediate since $N^*\Pi$ is coisotropic. □

Symplectic reduction allows us to take what we are studying in a larger space, to a reduced space where the physical system's ”true” dynamics lie. In our specific case, our configuration space is some manifold Q. The system's “true” dynamics are studied in the associated phase space (T^*Q, ω_Q). In the construction of Morse families, we used the "extended" configuration space B and the "extended" phase space (T^*B, ω_B).

As previously stated, in our construction, the symplectic reduction λ allows us to "forget" the y and p_y components from T^*B. But, it does more than that, as we now demonstrate.

Proposition 3.4.4. *Let $\lambda : N^*\Pi \to T^*Q$ be a symplectic reduction onto the symplectic manifold (T^*Q, ω_Q). Then,* $\operatorname{rank} \omega_{N^*\Pi} = \dim T^*Q$*, and the characteristic distribution* $D^{\omega_{N^*\Pi}} = \ker \omega_{N^*\Pi} = \ker T\lambda$.

Remark 3.4.5. If V is any finite-dimensional vector space over $\mathbb{R}^n$, then for a two-form $\omega \in \bigwedge^2 V^*$, $\operatorname{rank} \omega$ is defined to be the rank of the linear mapping $\omega^\flat : V \to V^*$, where for $X \in V$, $\omega^\flat(X) := \omega(X, \cdot)$. ▲

Proof. Since $d\omega_{N^*\Pi} = 0$, we have that $\ker \omega_{N^*\Pi} = F^{\omega_{N^*\Pi}}$, i.e., the characteristic subspace of $\omega_{N^*\Pi}$ (see the proof of Proposition 2.2.10). This gives us

$$\begin{aligned}
\operatorname{rank} \omega_{N^*\Pi} &= \dim N^*\Pi - \dim \ker \omega_{N^*\Pi} \\
&= \dim \ker \mathrm{T}\lambda + \dim T^*Q - \dim \ker \omega_{N^*\Pi} \\
&= \dim T^*Q.
\end{aligned}$$

Therefore, $\ker \omega_{N^*\Pi} = F^{\omega_{N^*\Pi}}$ has constant rank. By Remark 2.2.8, it coincides with the distribution $D^{\omega_{N^*\Pi}}$. □

What does this mean? The distribution $D^{\omega_{N^*\Pi}} = \ker T\lambda$ foliates T^*B, i.e., it partitions T^*B into submanifolds of lower dimension. The image of these submanifolds under λ correspond to the set of characteristics in T^*Q. See [23], Section 8.7 for a more detailed discussion.

3.4.3 Showing that λ_M and Λ_M are Lagrangian Immersions

The final step is showing that the map $\Lambda_M : B_M \to T^*Q$ generates Lagrangian submanifolds in T^*Q. This will demonstrate that, just like the geometric method of characteristics, Morse families generate solutions to the Hamilton-Jacobi equation via Lagrangian immersions. Before proceeding, we need a few facts.

Proposition 3.4.6. *Let (M, ω) be a symplectic manifold, and let (N, ι_N) be a coisotropic submanifold of M, where ι_N is the natural inclusion mapping. Define $\omega_N := \iota_N^* \omega \in \Omega^2(N)$. Then, for any $n \in N$,*

$$\ker(\omega_N)_n = T_n N \cap (T_n N)^\omega = (T_n N)^\omega.$$

Proof. This follows immediately from the definition of coisotropy. See [23], Formula 8.1.5. □

Proposition 3.4.7. *Let (M, ω) be a symplectic manifold, and let (N, ι_N) be a submanifold (not necessarily coisotropic). If $\zeta : N \to P$ is a symplectic reduction onto a symplectic manifold (P, ω_P), then*

$$\ker \omega_N = \ker T\zeta.$$

Proof. See [23], Proposition 8.7.2. Since ζ is a symplectic reduction, we have that $\zeta^* \omega_P = \iota_N^* \omega$. But, since ω_P is non-degenerate by definition, and since $\omega_N = \iota_N^* \omega$ by definition, the result follows from (3.4.4). □

Proposition 3.4.8. *Let (V, ω) be a symplectic vector space, and let W_1 and W_2 be subspaces of V. Then,*

$$W_1^\omega \cap W_2^\omega = (W_1 + W_2)^\omega.$$

Proof. The proof is clear. See [23], Proposition 7.2.1/5 for details. □

Now, we are ready to show how symplectic reduction plays its role in our construction, as well as showing that Morse families generate Lagrangian submanifolds. Let $i : L_M \to T^*B$ and $\iota : N^*\Pi \to T^*B$ be natural inclusions, and let (B, π, M) be a Morse family over Q.

Proposition 3.4.9. *The maps λ_M and Λ_M are Lagrangian immersions.*

Proof. For brevity, we will show that λ_M is a Lagrangian immersion, since its proof uses facts about transversality. First, we show that λ_M is an immersion. For $\eta \in L_M$, we have that

$$\ker\, T_\eta \lambda_M = T_\eta(L_M) \cap \ker\, T_\eta \lambda.$$

By definition of L_M and Proposition 3.4.7, we get

$$\ker\, T_\eta \lambda_M = T_\eta(\mathrm{Im}\ \mathrm{d}M) \cap T_\eta(N^*\Pi) \cap (T_\eta N^*\Pi)^{\omega_B}.$$

Since $N^*\Pi$ is a coisotropic submanifold of T^*B, this reduces by Proposition 3.4.6 to

$$\ker\, T_\eta \lambda_M = T_\eta(\mathrm{Im}\ \mathrm{d}M) \cap (T_\eta N^*\Pi)^{\omega_B}.$$

Since $\mathrm{Im}\ \mathrm{d}M$ is a Lagrangian submanifold of T^*B, we have that

$$T_\eta(\mathrm{Im}\ \mathrm{d}M) = (T_\eta(\mathrm{Im}\ \mathrm{d}M))^{\omega_B}.$$

Thus,

$$\ker\, T_\eta \lambda_M = (T_\eta(\mathrm{Im}\ \mathrm{d}M))^{\omega_B} \cap (T_\eta N^*\Pi)^{\omega_B}.$$

By Proposition 3.4.8, this becomes

$$\ker T_\eta \lambda_M = (T_\eta(\mathrm{Im}\ \mathrm{d}M) + (T_\eta N^*\Pi))^{\omega_B}.$$

Since M is a Morse family, Im dM and $N^*\Pi$ intersect transversally. Therefore,

$$\ker T_\eta \lambda_M = (T_\eta(T^*B))^{\omega_B} = \{0\}.$$

To show that λ_M is Lagrangian, note that L_M has dimension n; therefore, it suffices to show that λ_M is isotropic. Note that $\lambda_M = \lambda \circ i$. Using this fact and (3.20),

$$\begin{aligned}\lambda_M^*(\Theta_Q) &= i^* \circ \lambda^*(\Theta_Q)\\ &= i^* \circ \iota^*(\Theta_B).\end{aligned}$$

Therefore, $\lambda_M^*(d\Theta_Q) = i^* \circ \iota^*(d\Theta_B) = 0$ since L_M is an isotropic submanifold of T^*B. Thus, λ_M is a Lagrangian immersion. The proof for Λ_M can be found in [23], Lemma 12.4.4. □

To summarize: our construction allows us to work in the "extended" phase space T^*B to avoid the multi-value issues arising when we compute $\dfrac{\partial M}{\partial x}$ in T^*Q. The conormal bundle $N^*\Pi$ grants us a foliation of T^*B, and via symplectic reduction, we can project from T^*B to T^*Q in such a manner that allows us to generate Lagrangian submanifolds in T^*Q which coincide with the general solution given by the geometric method of characteristics.

3.5 The Transversality Condition in Detail

We now examine the transversality condition from Definition 3.3.1 in greater detail. Recall that in order for the map $\lambda_M : L_M \to T^*Q$ to be an immersion, Im $\mathrm{d}M$ and $N^*\Pi$ must intersect transversally. The following discussion will allow us to derive an alternate definition of a Morse family.

3.5.1 Tangent space to the fiber-critical manifold B_M

Let $b = (x, y) \in B$, $M : B \to \mathbb{R}$ a smooth function,

$$\begin{pmatrix} X \\ \\ Y \end{pmatrix}_b := X \left(\frac{\partial}{\partial x}\right)_{(x,y)} + Y \left(\frac{\partial}{\partial y}\right)_{(x,y)} \in T_{(x,y)}B \; ; \tag{3.21}$$

$$\mathrm{d}M : B \to T^*B : (x, y) \mapsto (\mathrm{d}M)_{(x,y)} = \left(x, y; \frac{\partial M}{\partial x}(x, y), \frac{\partial M}{\partial y}(x, y)\right) \; . \tag{3.22}$$

The tangent lift of $\mathrm{d}M : B \to T^*B$ at $(x, y) \in B$ is a mapping

$$T_{(x,y)}\mathrm{d}M : T_{(x,y)}B \to T_{\mathrm{d}M_{(x,y)}}T^*B \; . \tag{3.23}$$

Let $\gamma(t) = (x(t), y(t))$ be a curve in the fiber-critical submanifold $B_M \subseteq B$ (where t belongs to some open interval in $\mathbb{R}$ containing 0). From the expression (3.12) for B_M in adapted coordinates, we obtain that

$$\frac{\partial M}{\partial y}(\gamma(t)) = \frac{\partial M}{\partial y}(x(t), y(t)) = 0 \quad \forall t \; .$$

Let $\begin{pmatrix} X_{\gamma(t)} \\ \\ Y_{\gamma(t)} \end{pmatrix} = \begin{pmatrix} \dot{x}(t) \\ \\ \dot{y}(t) \end{pmatrix} = \dot{\gamma}(t) \in T_{\gamma(t)}B_M$ be a vector tangent to γ — thus tangent to B_M — at the point $\gamma(t) \in B_M$. Differentiate the above condition with respect to t to obtain

$$\frac{\partial^2 M}{\partial y\,\partial x}(\gamma(t)) \cdot \dot{x}(t) + \frac{\partial^2 M}{\partial y^2}(\gamma(t)) \cdot \dot{y}(t) = 0 \ ,$$

which can be written as

$$M_{yx}(\gamma(t)) \cdot X_{\gamma(t)} + M_{yy}(\gamma(t)) \cdot Y_{\gamma(t)} = 0 \ ,$$

or, in matrix notations, as

$$\begin{pmatrix} M_{yx}(\gamma(t)) & M_{yy}(\gamma(t)) \end{pmatrix} \begin{pmatrix} X_{\gamma(t)} \\ \\ Y_{\gamma(t)} \end{pmatrix} = 0 \ ,$$

where $M_{yx}(\gamma(t)) := \dfrac{\partial^2 M}{\partial y\,\partial x}(\gamma(t))$ is a $(k \times n)$ matrix of the mixed second derivatives of M; similarly for $M_{yy}(\gamma(t))$.

Conclusion: For a point $b = (x, y) \in B_M$ in the fiber-critical manifold B_M, i.e., satisfying $\dfrac{\partial M}{\partial y}(b) = 0$ (3.12),

$$\begin{pmatrix} X \\ \\ Y \end{pmatrix}_b \in T_b B_M \quad \Longleftrightarrow \quad M_{yx}(b) \cdot X_b + M_{yy}(b) \cdot Y_b = 0 \ .$$

3.5.2 Computing the map $T_{(x,y)}\mathrm{d}M$ and the tangent space $T_{(\mathrm{d}M)_b}\mathrm{Im}\,\mathrm{d}M$

Let temporarily $\gamma(t) = (x(t), y(t))$ be a curve in B (later we will require it to be in B_M) with

$$\gamma(0) = b \in B \ , \quad \dot{\gamma}(0) = \begin{pmatrix} X_b \\ \\ Y_b \end{pmatrix} \in T^*B.$$

The image, $(\mathrm{d}M) \circ \gamma(t) \in T^*_{\gamma(t)}B$, of γ under $\mathrm{d}M$ is a curve in T^*B. To find how the tangent lift

$$T_b\mathrm{d}M : T_bB \to T_{(\mathrm{d}M)_b}T^*B$$

acts on the vector $\dot{\gamma}(0) \in T_bB$, we will use that

$$(T_b\mathrm{d}M)\,\dot{\gamma}(0) = \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{d}M) \circ \gamma(t)\bigg|_{t=0} \in T_{(\mathrm{d}M)_b}T^*B \ .$$

Recalling (3.22), we have

$$(\mathrm{d}M) \circ \gamma(t) = \left(x(t), y(t); \frac{\partial M}{\partial x}(x(t), y(t)), \frac{\partial M}{\partial x}(x(t), y(t)) \right) \ .$$

Differentiate this with respect to t and set $t = 0$ to obtain

$$
\begin{aligned}
(T_b \mathrm{d}M) \begin{pmatrix} X_b \\ \\ Y_b \end{pmatrix} &= (T_b \mathrm{d}M)\, \dot{\gamma}(0) = \frac{\mathrm{d}}{\mathrm{d}t} (\mathrm{d}M) \circ \gamma(t) \Big|_{t=0} \\
&= \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} x(t) \\ \\ y(t) \\ \\ \frac{\partial M}{\partial x}(x(t), y(t)) \\ \\ \frac{\partial M}{\partial x}(x(t), y(t)) \end{pmatrix} \Bigg|_{t=0} \\
&= \begin{pmatrix} \dot{x}(0) \\ \\ \dot{y}(0) \\ \\ M_{xx}(\gamma(0))\, \dot{x}(0) + M_{xy}(\gamma(0))\, \dot{y}(0) \\ \\ M_{yx}(\gamma(0))\, \dot{x}(0) + M_{yy}(\gamma(0))\, \dot{y}(0) \end{pmatrix} \\
&= \begin{pmatrix} I_n & 0 \\ \\ 0 & I_k \\ \\ M_{xx}(b) & M_{xy}(b) \\ \\ M_{yx}(b) & M_{yy}(b) \end{pmatrix} \begin{pmatrix} X_b \\ \\ Y_b \end{pmatrix} ,
\end{aligned}
$$

so that $T_b(\mathrm{d}M)$ is the $2(n+k)\times(n+k)$ matrix

$$T_b(\mathrm{d}M) = \begin{pmatrix} I_n & 0 \\ 0 & I_k \\ M_{xx}(b) & M_{xy}(b) \\ M_{yx}(b) & M_{yy}(b) \end{pmatrix} .$$

Therefore the tangent space $T_{(\mathrm{d}M)_b}\mathrm{Im}\,\mathrm{d}M$ to $\mathrm{Im}\,\mathrm{d}M$ at $(\mathrm{d}M)_b \in \mathrm{Im}\,\mathrm{d}M$ consists of vectors with coordinates

$$T_{(\mathrm{d}M)_b}\mathrm{Im}\,\mathrm{d}M = \Big\{(X_b, Y_b; M_{xx}(b)X_b + M_{xy}(b)Y_b, M_{yx}(b)X_b + M_{yy}(b)Y_b)\Big\} \tag{3.24}$$

in the basis

$$\left\{\left(\frac{\partial}{\partial x}\right)_{(\mathrm{d}M)_b}, \left(\frac{\partial}{\partial y}\right)_{(\mathrm{d}M)_b}, \left(\frac{\partial}{\partial p_x}\right)_{(\mathrm{d}M)_b}, \left(\frac{\partial}{\partial p_y}\right)_{(\mathrm{d}M)_b}\right\} . \tag{3.25}$$

3.5.3 Deriving the transversality condition

Recall that $N^*\Pi \pitchfork \mathrm{Im}\,\mathrm{d}M$ means that, at every point $\eta \in L_M \subseteq T^*B$,

$$T_\eta N^*\Pi \oplus T_\eta \mathrm{Im}\,\mathrm{d}M = T_\eta T^*B .$$

From (3.9), it is clear that, in the basis (3.25), $T_\eta N^*\Pi$ consists of vectors with coordinates

$$T_\eta N^*\Pi = \{(X, Y; P_x, 0)\} .$$

Therefore, when we add $T_\eta \mathrm{Im}\,\mathrm{d}M$ (3.24) to $T_\eta N^*\Pi$, the only components that are missing are the the $\dfrac{\partial}{\partial p_{y,\alpha}}$ components (for $\alpha = 1, \ldots, k$), whose coefficients are $M_{yx}(b)X_b + M_{yy}(b)Y_b$ (cf. (3.24)). For the linear combination

$$\sum_{\alpha=1}^{k}\left[\sum_{i=1}^{n}\frac{\partial^2 M}{\partial y^\alpha\,\partial x^i}(b)\,X_b^i + \sum_{\beta=1}^{k}\frac{\partial^2 M}{\partial y^\alpha\,\partial y^\beta}(b)\,Y_b^\beta\right]\left(\frac{\partial}{\partial p_{y,\alpha}}\right)_b$$

to have span of dimension k, the $k \times (n+k)$ matrix

$$\begin{pmatrix} M_{yx}(b) & M_{yy}(b) \end{pmatrix} \tag{3.26}$$

must have full rank, namely, k. The tangent space to $L_M = N^*\Pi \cap \operatorname{Im} \mathrm{d}M$ at η can be written as

$$T_\eta L_M = \Big\{(X_b, Y_b; M_{xx}(b)X_b{+}M_{xy}(b)Y_b, 0) : \\ M_{yx}(b)X_b + M_{yy}(b)Y_b = 0\Big\}\ , \quad b \in B_M\ .$$

Because of condition of full rank of the matrix (3.26), the equation

$$M_{yx}(b)X_b + M_{yy}(b)Y_b = 0,$$

or, in matrix form,

$$\begin{pmatrix} M_{yx}(b) & M_{yy}(b) \end{pmatrix}\begin{pmatrix} X_b \\ Y_b \end{pmatrix} = 0_{k\times 1}$$

(where $0_{k\times 1}$ is a column vector of length k), can be solved to obtain k of the $(n+k)$ components (X_b^i, Y_b^α), so that among them exactly n are independent. Therefore, for a given $b \in B_M$, the tangent space $T_\eta L_M$ is n-dimensional, i.e., it has the same dimension as B_M. This leads to an alternate definition of a Morse family, to which we will frequently refer:

Definition 3.5.1. The triple (B, π, M) is a Morse family if and only if the matrix $\begin{pmatrix} \dfrac{\partial^2 M}{\partial y^\alpha \partial x^i} & \dfrac{\partial^2 M}{\partial y^\alpha \partial y^\beta} \end{pmatrix}$ has maximum rank k on the fiber-critical manifold B_M. ▲

3.6 Example of a Morse Family

The following is a standard example of a Morse family which solves the eikonal equation in Euclidean space, assuming the speed of light is constant (which we can set indentically equal to 1).

Let $\mathscr{S} \subset Q = \mathbb{R}^n$ be a k-dimensional embedded submanifold of Q. This will play the role of a light source in the configuration space. Let

$$X : \mathbb{R}^k_{\text{loc}} \to \mathscr{S}$$

$$(y^1, ...y^k) \mapsto \Big(X^1(y^1, ..., y^k), ..., X^n(y^1, ..., y^k)\Big),$$

so that $\mathscr{S} = \Big(X^1(y^1, ..., y^k), ..., X^n(y^1, ..., y^k)\Big)$. Local tangent vector fields in $\mathscr{S}$ are

$$\mathbf{e}_\alpha(X(y)) = \frac{\partial X}{\partial y^\alpha}(y) \in T_{X(y)}\mathscr{S}.$$

Proposition 3.6.1. *Let $\mathscr{S}$ be an embedded submanifold of $Q = \mathbb{R}^n$ of dimension $k < n$. Let $B = \Big\{(x, y) \in Q \times \mathbb{R}^k_{loc} : x \neq X(y)\Big\}$. The Euclidean distance function $M : B \to \mathbb{R}$ defined by*

$$M(x, y) = \sqrt{\sum_{i=1}^{n} (x - X(y))^2} =: |x - X(y)|_E \tag{3.27}$$

is a Morse family along the submersion $\pi : B \to Q$ induced by the natural projection onto the first argument. It generates a generalized solution to the eikonal equation for the initial condition $M\Big(X(y), y\Big) = 0\ \forall y \in \mathbb{R}^k_{loc}$ on $\mathscr{S}$, given by

$$B_M = \left\{ (x, y) \in B : (x - X(y)) \perp T_{X(y)}\mathscr{S} \right\}; \tag{3.28}$$

$$\Lambda_M(x, y) = \left\{ \left(x, \frac{x - X(y)}{|x - X(y)|_E} \right) \right\}. \tag{3.29}$$

Proof. We need to show that the matrix $\begin{pmatrix} \dfrac{\partial^2 M}{\partial y^\alpha \partial x^i} & \dfrac{\partial^2 M}{\partial y^\alpha \partial y^\beta} \end{pmatrix}$ has rank k on B_M. Straightforward calculations yield the following:

$$\frac{\partial M}{\partial x^i} = \frac{x^i - X^i(y)}{|x - X(y)|_E} =: p_i \tag{3.30}$$

$$\frac{\partial M}{\partial y^\alpha} = -\mathbf{p} \cdot \mathbf{e}_\alpha \tag{3.31}$$

$$\frac{\partial^2 M}{\partial x^i \partial y^\alpha} = \frac{(\mathbf{p} \cdot \mathbf{e}_\alpha) p_i - \mathbf{e}_{\alpha,i}}{|x - X(y)|_E} \tag{3.32}$$

$$\frac{\partial^2 M}{\partial y^\alpha \partial y^\beta} = \frac{\mathbf{e}_\alpha \cdot \mathbf{e}_\beta - (\mathbf{p} \cdot \mathbf{e}_\alpha)(\mathbf{p} \cdot \mathbf{e}_\beta)}{|x - X(y)|_E} - \mathbf{p} \cdot \frac{\partial \mathbf{e}_\beta}{\partial y^\alpha}. \tag{3.33}$$

On B_M, we have $\mathbf{p} \cdot \mathbf{e}_\alpha = 0$. Writing this out in more detail gives us

$$\frac{1}{|x - X(y)|_E} \left[(x - X(y)) \cdot \frac{\partial X}{\partial y^\alpha}(y) \right] = 0. \tag{3.34}$$

As both $(x - X(y))$ and $\dfrac{\partial X}{\partial y^\alpha}(y)$ can be considered as vectors, (3.34) yields the orthogonality condition given in (3.28), so B_M can be described as it is in (3.28).

Next, to show (3.29), recall from (3.17) that

$$\Lambda_M := \lambda_M \circ (dM)|_{B_M} : B_M \to T^*Q.$$

The submanifold L_M (3.11) obtained as the image of $(\mathrm{d}M)|_{B_M}$ is

$$L_M = \left\{ \left(x, y; \frac{x - X(y)}{|x - X(y)|_E}, 0 \right) : (x - X(y)) \perp T_{X(y)}\mathscr{S} \right\}.$$

Via λ_M, we "forget" the y and p_y coordinates. Therefore,

$$\begin{aligned}\Lambda_M &= \lambda_M(L_M) \\ &= \left\{ \left(x; \frac{x - X(y)}{|x - X(y)|_E} \right) : (x - X(y)) \perp T_{X(y)}\mathscr{S} \right\}.\end{aligned}$$

Now, we show that M is a Morse family. On B_M, we have

$$\left.\frac{\partial^2 M}{\partial x^i \partial y^\alpha}\right|_{B_M} = -\frac{\mathbf{e}_{\alpha i}}{|x - X(y)|_E}.$$

By the linear independence of the $\mathbf{e}_\alpha$, the submatrix $\left(\left.\frac{\partial^2 M}{\partial x^i \partial y^\alpha}\right|_{B_M} \right)$ of (3.5.1) has rank k. Therefore, M is a Morse family.

Finally, we show that M generates a solution to the eikonal equation (2.14). The associated Hamilton-Jacobi equation is given by

$$H(x, p) = |p|^2 - 1.$$

So, we need to show that Λ_M takes values in $H^{-1}(0)$. Using the p_i given in (3.30), this is clearly true. Thus, the pair (B_M, Λ_M) is a generalized solution to the eikonal equation. $\square$

3.7 Caustics

Recall that if a 1-parameter family of curves is given implicitly by some C^2 function $U : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $U(x, y) = 0$, where y is a parameter, then their envelope (if it exists) is determined by excluding y from the following system of equations:

$$\begin{aligned} U(x,y) &= 0 \\ \frac{\partial U}{\partial y}(x,y) &= 0. \end{aligned} \tag{3.35}$$

In higher dimensions, we use the general Implicit Function Theorem. Suppose we have a k-parameter family of surfaces in $\mathbb{R}^n$ which is given by some C^2 function $U : \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}$, and the level surfaces

$$\Sigma_y := \Big\{ x : U(x,y) = 0 \Big\}$$

satisfy the non-degeneracy condition

$$\det \left(\frac{\partial U}{\partial y}\bigg|_{\Sigma} \right) \neq 0.$$

Then, we find the envelope by excluding the y from the following system of equations:

$$\begin{aligned} U(x,y) &= 0 \\ \frac{\partial U}{\partial y^1}(x,y) &= 0 \\ &\dots \\ \frac{\partial U}{\partial y^k}(x,y) &= 0. \end{aligned} \tag{3.36}$$

In the context of Morse families, we find the caustic of a light source by finding the envelope of the set of light rays, namely B_M. However, B_M is determined by the k conditions (3.12), not a single condition. So, it's not immediately clear how to find the envelope using these conditions.

However, it turns out that we already have a method of determing a caustic from the method of characteristics. Recall that in that context, we determined the caustic by finding the critical points (and values) of the projection $\pi_Q : T^*Q \to Q$. In other words, we find where $\det(D\pi_Q) = 0$. In the context of Morse families, we only need to look at the subset of B_M where the matrix $\left(M_{yy}\right)$ of second derivatives of the y has determinant zero, as the following lemma demonstrates.

Lemma 3.7.1. *Let L be a Lagrangian submanifold of T^*Q, and let $\iota : L \to T^*Q$ be a Lagrangian immersion. Let $\xi \in L$. Let (B, π, M) be a Morse family of fiber dimension k generating (L, ι) at ξ. Assume $(x, y) \in B_M$ such that $\xi := \Lambda_M(x, y)$. Let $\pi_Q : T^*Q \to Q$ be the projection map, $\Pi_Q := \pi_Q \circ \iota : L \to Q$. Then,*

$$k - \mathit{rank}\ M_{yy}(x, y) = n - \mathit{rank}\ (T\Pi_Q)_\xi.$$

Idea of Proof. Since

$$k - \operatorname{rank} M_{yy}(x, y) = \dim \ker M_{yy}(x, y)$$

and

$$n - \operatorname{rank} (T\Pi_Q)_\xi = \dim \ker (T\Pi_Q)_\xi$$

by the rank-nullity theorem, one can prove the lemma by looking at the dimension of the kernels of M_{yy} and $(T\Pi_Q)_\xi$. Indeed, if dim ker $M_{yy}(x, y) \neq 0$, then since M is a Morse family, the submatrix M_{xy} has maximum rank. Thus, one can obtain an isomorphism between $\ker M_{yy}(x, y)$ and $\ker (T\Pi_Q)_\xi$. $\square$

This lemma states that if we want to find the caustic of our light source, we find the points $(x, y) \in B_M$ where the matrix $\left(M_{yy}\right)$ has determinant equal to zero.

Definition 3.7.2 (Singular Set and Caustic)**.** We define the *singular set of* B_M as

$$\Sigma(B_M) := \left\{ (x, y) \in B_M : \det\left(\frac{\partial^2 M}{\partial y^\alpha \partial y^\beta}(x, y) \right) = 0 \right\} \subset B. \tag{3.37}$$

The *caustic* $\Gamma(B_M) \subset Q$ is the image of $\Sigma(B_M)$ under the projection $\pi : B \to Q$. ▲

Using (3.37), we can determine the caustic of a light source $\mathscr{S}$ by solving for $x^1, ..., x^n$ in the system of equations

$$\frac{\partial M}{\partial y^\alpha}(x, y) = 0 \tag{3.38}$$

$$\det\left(\frac{\partial^2 M}{\partial y^\alpha \partial y^\beta} \right) = 0. \tag{3.39}$$

Remark 3.7.3. The Morse family (3.27) can be used to determine the caustic of a plane curve $\mathscr{S} \in (\mathbb{R}^2, \delta_{ij})$. If $\mathscr{S}$ is given by the parameterization $X(y)$, then we can explicitly write conditions (3.38) and (3.39):

$$\frac{(x^1 - X^1(y))\dot{X}^1 + (x^2 - X^2(y))\dot{X}^2}{|x - X(y)|_E} = 0$$

$$\frac{(\dot{X}^1)^2 + (\dot{X}^2)^2 - (x^1 - X^1(y))\ddot{X}^1 - (x^2 - X^2(y))\ddot{X}^2}{|x - X(y)|_E} = 0,$$

where $\dot{X} := \dfrac{\partial M}{\partial y}(y)$, and $\ddot{X} := \dfrac{\partial^2 M}{\partial y^2}(y)$. Straightforward calculations yield the following parameterization for the caustic $\Gamma(B_M) \in \mathbb{R}^2$:

$$
\begin{aligned}
x^1(y) &= X^1(y) + \frac{(\dot{X}^1)^2 + (\dot{X}^2)^2}{\dot{X}^1\ddot{X}^2 - \ddot{X}^1\dot{X}^2}\dot{X}^2 \\
x^2(y) &= X^2(y) - \frac{(\dot{X}^1)^2 + (\dot{X}^2)^2}{\dot{X}^1\ddot{X}^2 - \ddot{X}^1\dot{X}^2}\dot{X}^1.
\end{aligned}
\tag{3.40}
$$

▲

This result can be found using different methods as well. See, for example, [6], Theorem 7.3 and [10], Theorem 4.3.

Chapter 4

Finsler Geometry

When discussing propagation of rays in anisotropic media, Riemannian geometry is insufficient to describe what is happening. This is due to the fact that, when rays propagate in anisotropic media, the direction of propagation matters, which is not the case in Riemannian geometry. So, we need to develop new tools to understand what happens in anisotropic media. Hence, we utilize Finsler geometry. In a nutshell, Finsler geometry is a generalization of Riemannian geometry which accounts for both the position on Q *and* the direction of a vector in TQ. In this section, we will define the indicatrix of a Finsler metric and show how it can be used to describe the speed of propagation in anisotropic media. We will also define conditions that rays must satisfy. Finally, we will give an example of a Finsler metric, namely a Randers metric.

4.1 Minkowski Norms

Definition 4.1.1. A *Minkowski norm* on $\mathbb{R}^n$ is a function $F : \mathbb{R}^n \to [0, \infty)$ satisfying the following conditions:

(a) F is C^∞ on $\mathbb{R}^n \backslash \{0\}$;

(b) $F(\lambda y) = \lambda F(y)$ for every $\lambda > 0$, i.e., F is positively homogeneous of order 1;

(c) the $n \times n$ matrix $\big(g_{ij}(y)\big) := \Big(\big[\frac{1}{2}F^2\big]_{y^i y^j}(y)\Big)$ is positive-definite for every $y \neq 0$.

▲

The following theorem summarizes some important properties of Minkowski norms (see [3], Theorem 1.2.2):

Theorem 4.1.2. *Let $F : \mathbb{R}^n \to [0, \infty)$ satisfy the conditions of Definition 4.1.1. Then, we have the following:*

(a) ***positivity:*** *$F(y) > 0$ for every $y \neq 0$;*

(b) ***triangle inequality:*** *$F(y_1 + y_2) \leq F(y_1) + F(y_2)$, and equality holds if and only if $y_1 = \alpha y_2$, for some $\alpha \geq 0$;*

(c) ***fundamental inequality:*** *for any $w \in \mathbb{R}^n$, $w^j F_{y^j}(y) \leq F(w)$ for every $y \neq 0$, and equality holds if and only if $w = \alpha y$, for some $\alpha \geq 0$.*

4.2 Finsler Structures

Let Q be an n-dimensional C^∞ manifold, and let (x, v) denote elements of the tangent bundle TQ, where $x \in Q$, $v \in T_x Q$.

Definition 4.2.1. The *slit tangent bundle* is defined as

$$TQ \backslash \{0\} := \bigcup_{x \in Q} \Big\{ T_x Q \backslash \{0\} \Big\}.$$

▲

Definition 4.2.2. A (globally defined) *Finsler structure on* Q is a function F : $\mathrm{T}Q \to [0, \infty)$ with the properties

(1) **regularity:** F is C^∞ on $TQ \backslash \{0\}$;

(2) **positive homogeneity:** $F(x, \lambda v) = \lambda F(x, v)$ for every $\lambda > 0$;

(3) **strong convexity:** $g_{ij}(x, v) = \left[\frac{1}{2}F^2\right]_{v^i v^j}(x, v)$ is positive-definite on all of $\mathrm{T}Q \backslash \{0\}$.

▲

Note that the restriction of F to any specific tangent space T_xQ is a Minkowski norm.

Definition 4.2.3. The pair (Q, F), where F is a Finsler structure on Q, is called a *Finsler manifold.* ▲

Before continuing, we list a few immediate consequences of the positive homogeneity which will be referred to later.

$$F(v)F_{v^i}(v)v^i = F(v)^2, \;\; F_{v^i v^j}(v)v^i v^j = 0. \tag{4.1}$$

4.3 Indicatrix of a Finsler Structure

Definition 4.3.1. Let $x \in Q$ be given. The *indicatrix* I_x of F at x is defined as the set

$$I_x := \Big\{ v \in T_xQ : F(v) = 1 \Big\}.$$

▲

By the strong convexity assumption on F, I_x is a closed, smooth, strictly convex surface in T_xQ which contains the origin in its interior. In other words, given any two points $v, w \in I_x$, the line segment L containing v and w is strictly in the interior of I_x.

The indicatrix turns out to be useful in tying light ray propagation into the metric of the space. We will use it to write a general formula for the Morse family from the previous chapter, as well as to find certain conditions that light rays must satisfy.

Note that F can be completely determined by its indicatrix. To show this, let $v \in T_xQ$ for some $x \in Q$, and let $\tilde{v} \in I_x$ be in the same direction as v. Note that we can write

$$v = \frac{|v|}{|\tilde{v}|}\tilde{v},$$

where $|v|$ denotes some metric. Then, using the homogeneity of F (4.1) and the fact that $F(\tilde{v}) = 1$, we obtain

$$F(v) = F\left(\frac{|v|}{|\tilde{v}|}\tilde{v}\right) = \frac{|v|}{|\tilde{v}|}F(\tilde{v}) = \frac{|v|}{|\tilde{v}|}.$$

4.4 Perturbing the Euclidean Metric

It should be noted that for $n = 2$, it is quite simple to show that, in fact, *any* Minkowski metric is a perturbation of the standard Euclidean metric. To show this, we need to be able to do the following:

(1) Define a map whose graph is a simple, closed, strongly convex curve Γ, as these are conditions for the indicatrix of a Minkowski metric.

(2) Translate Γ such that its "geometric center" (to be defined) is translated to

the origin.

(3) Find a parameterization such that as the curve Γ is deformed into some other curve $\tilde{\Gamma}$, it remains a simple, closed, strongly convex curve.

4.4.1 Preliminaries

Let $\mathbb{T} := \mathbb{R}/2\pi$, and let $\rho : \mathbb{T} \to (0, \infty)$ be a 2π-periodic smooth function with

$$\int_0^{2\pi} \rho(t)e^{it}\,dt = 0. \tag{4.2}$$

Given a point γ_0, consider the induced map

$$\gamma : \mathbb{T} \to \mathbb{C}, \quad \phi \mapsto \gamma_0 + \int_0^{\phi} \rho(t)e^{it}\,dt.$$

Condition (4.2) ensures that $\Gamma := \gamma(\mathbb{T})$ is a closed curve. Note that

$$\gamma'(\phi) = \rho(\phi)e^{i\phi}.$$

So, the unit tangent vector $T(\phi) = e^{i\phi}$ induces a map $T : \mathbb{T} \to S^1$ which is one-to-one. Note also that

$$\gamma''(\phi) = \rho'(\phi)e^{i\phi} + \rho(\phi)ie^{i\phi}; \tag{4.3}$$

$$\kappa(\phi) = \frac{|\gamma'(\phi) \times \gamma''(\phi)|}{|\gamma'(\phi)|^3} = \frac{1}{\rho(\phi)}. \tag{4.4}$$

Since $\rho > 0$ by definition, this means that $\kappa > 0$, and thus γ defines a simple, closed, strongly convex curve.

Next, we seek a parameterization for Γ. Given a strongly convex curve Γ, let

$G : \Gamma \to S^1$ be the Gauss map. In other words, $G(p)$ is the outer unit normal vector of Γ at p. Then, the map $iG : \Gamma \to S^1$ gives the unit tangent vector of Γ at p counterclockwise. Since iG is one-to-one, its inverse gives a parameterization $\gamma : \mathbb{T} \to \Gamma$ such that $\gamma'(t)$ is positive parallel to e^{it}. In other words, $\gamma'(t) = \rho(t)e^{it}$ for some positive function $\rho(t)$. From (4.4), it follows that $\rho(t)$ is the radius of curvature of Γ at $\gamma(t)$.

We must also discuss the geometric center of Γ, as we will soon be translating this center to the origin. Let $K \subset \mathbb{R}^2$ be a compact, convex set with nonempty interior. Let $\Gamma = \partial K$, which is a simple, closed, convex curve. Then, the geometric center of Γ is defined as

$$c(\Gamma) = \frac{\int_\Gamma x\, ds}{\int_\Gamma 1\, ds} \in \mathbb{R}^2. \tag{4.5}$$

Note that this is analogous to finding the center of mass of an object with uniform density. Let $\gamma : \mathbb{T} \to \Gamma$ be the parameterization given above with $\gamma'(t) = \rho(t)e^{it}$. Then,

$$c(\Gamma) = \frac{1}{L} \int_0^{2\pi} \gamma(t)\rho(t)\, dt, \tag{4.6}$$

where L denotes the length of Γ. Note that $c(\Gamma)$ is contained in the interior of K.

Finally, we discuss translations. For any point $p \in \mathbb{C}$, let $T_p(x) = x - p$ be the translation that translates p to the origin. Given a simple, closed, convex curve Γ, the translation $T_{c(\Gamma)}$ translates the geometric center of Γ to the origin.

4.4.2 Deformations of Minkowski Metrics

Now, we are ready to discuss how any Minkowski metric can be perturbed into the Euclidean metric. Let $F : \mathbb{R}^2 \to \mathbb{R}$ be a Minkowski metric, $I = F^{-1}(1)$ be its indicatrix, which is a simple, closed, strongly convex curve. Let $c(I)$ be the geometric center of I.

First, we translate. Suppose $c(I) \neq 0$. Consider the translation $I_s = T_{sc(I)}(I)$, $0 \leq s \leq 1$. Note that $I_0 = I$ and $c(I_1) = c(I) - c(I) = 0$. In this process, the shape of I_s stays the same (in particular, strongly convex). Therefore, it is the indicatrix of some Minkowski metric, say F_s. This gives the first deformation of Minkowski metrics that translates the geometric center to the origin.

Now, we deform in such a way that the strong convexity is preserved. Suppose $c(I) = 0$. Let $\gamma : \mathbb{T} \to I$ be the parameterization of I as given in 4.4.1. Then, $\gamma'(t) = \rho(t)e^{it}$, where $\rho(t)$ is the radius of curvature of I. Let

$$\bar{\rho} = \frac{1}{2\pi}\int_0^{2\pi} \rho(t)\,dt = \frac{L}{2\pi}, \quad \hat{\rho}(t) = \rho(t) - \bar{\rho}. \tag{4.7}$$

Clearly, $\bar{\rho} > 0$ since L is the arc length of the curve.

Now, let $s \in [0,1]$, and define

$$\rho_s(t) = \bar{\rho} + (1-s)\hat{\rho}(t) = s\bar{\rho} + (1-s)\rho(t), \;\; t \in \mathbb{T}. \tag{4.8}$$

Note that $\rho_0 = \rho$, $\rho_1 \equiv \bar{\rho}$, and ρ_s satisfies (4.2) for all $s \in [0,1]$. Furthermore, $\rho_s(t) > 0$ for any t, since it interpolates two positive functions ρ and $\bar{\rho}$. Let I_s be the curve constructed using $\rho_s(t)$ in Section 4.4.1, with the translation so that $c(I_s) = 0$. Then, $I_0 = I$, I_1 is a circle of radius $\bar{\rho}$ centered at the origin, and for each $s \in [0,1]$, I_s is a simple, closed, strongly convex curve centered at the origin. In particular, each I_s is the indicatrix of some Minkowski metric, say F_s. Moreover, F_1 is Euclidean F_e.

Combining these two steps of deformations, we get a path of Minkowski metrics connecting F to the Euclidean metric F_e.

4.5 Speed of Propagation in Homogeneous Anisotropic Media

We now wish to find more general version of the optical path M (1.5) which represents the shortest time to get from a point $X(y)$ on the source $\mathscr{S}$ to a point $x \in Q$ in the configuration space. In the case of isotropic, homogeneous media in $\mathbb{R}^n$, M was simply the Euclidean distance function in $\mathbb{R}^n$, divided by the speed of propagation. However, in the case of homogeneous but anisotropic media, this does not suffice because the speed of propagation can differ depending on direction. Having said that, we are able to write a more general version of M with the assistance of the indicatrix.

Let (Q, ν) be the configuration space with metric tensor ν, which is some Riemannian metric. Let $x \in Q$ with coordinates (x^i). The infinitesimal line element is given by

$$ds = \sqrt{\nu_{ij}\, \mathrm{d}x^i \mathrm{d}x^j}.$$

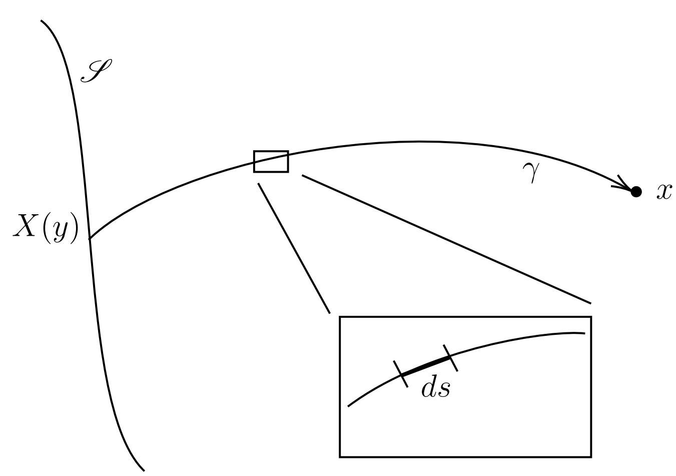

Figure 4.1: A curve γ from a source point $X(y) \in \mathscr{S}$ to a point x, and the line element ds.

Let $v \in T_xQ$. The time of propagation from $x \in Q$ to some infinitesimally close point (call it $x + \mathrm{d}x$) is given by

$$\frac{\mathrm{d}s}{\text{velocity}} = \frac{\sqrt{\nu_{ij}\,\mathrm{d}x^i \mathrm{d}x^j}}{V(x, v)},$$

where $V : TQ \backslash \{0\} \to \mathbb{R}$ is the velocity at $x \in Q$ in the direction of v.

For $v \in T_xQ$, define

$$|v|_\nu := \sqrt{\nu_{ij}(x) v^i v^j}.$$

So, we can write the optical path $M : B \to \mathbb{R}$ as

$$\begin{aligned} M(x, y) &= \int_0^1 \frac{|v|_\nu}{V(x, v)}\, dt \\ &=: \int_0^1 F(x, v)\, dt, \end{aligned} \tag{4.9}$$

where $F(x, v)$ is positively homogeneous of degree one with respect to v in order to have reparameterization invariance. For this to hold, V must not depend on v, but only on the *direction* of v, i.e., on $\dfrac{v(t)}{|v(t)|_\nu}$. Let $\Theta : TQ \backslash \{0\} \to S^{n-1}$ be a mapping which takes a vector v and gives the direction of v. Then, the speed of propagation can be given as a function of x and Θ, namely $\tilde{V}(x, \Theta(v))$. Then,

$$F(x, v) = \frac{\sqrt{\nu_{ij}(x) v^i v^j}}{\tilde{V}(x, \Theta(v))}.$$

Particular case: let $x \in \mathbb{R}^2$, and let $\nu_{ij} = \delta_{ij}$ be the usual flat metric in Cartesian coordinates. If the medium is homogeneous, we can write $\tilde{V} = \tilde{V}(\Theta)$ (i.e., $\tilde{V}$ has no dependence on x), and thus we can write $F(x, v) = F(v)$. We obtain:

$$F(v) = \frac{\sqrt{v^i v^j}}{\tilde{V}(\Theta(v))} = \frac{|v|_E}{\tilde{V}(\Theta(v))}, \tag{4.10}$$

where Θ is the direction of v, and the indicatrix I_x can be written as a function $r = \tilde{V}(\Theta)$. Later, we will do an example where we use the indicatrix I_x of a Finsler metric F to recover the expression for F itself.

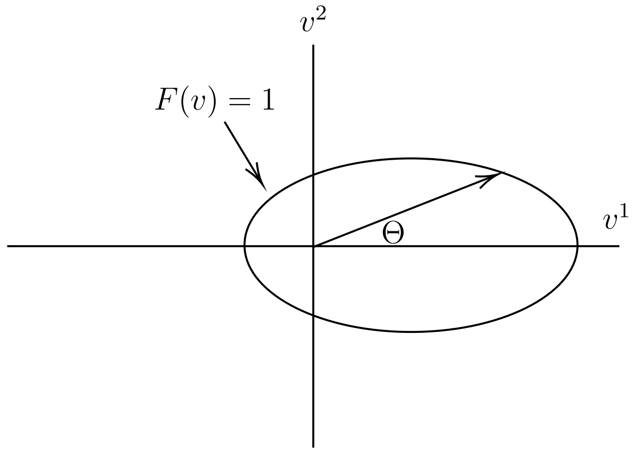

Figure 4.2: A typical indicatrix for $x \in \mathbb{R}^2$

4.6 Legendre Transform and Huygens's Principle

In the case of isotropic $\mathbb{R}^n$ with the Euclidean metric, light rays are orthogonal (in the Euclidean sense) to wavefronts (which we will soon redefine in the language of Finsler geometry). This is not the case in anisotropic media in general. In this section, we develop a notion of orthogonality with respect to a Finsler metric. Then, we will demonstrate that in general, light rays are always orthogonal to wavefronts in this sense.

4.6.1 Legendre Transform

Let (Q, F) be a Finsler manifold, let $x \in Q$, and let $I_x \subset T_xQ$ be the indicatrix of F at x.

Definition 4.6.1. The *Legendre transform* on T_xQ is defined by

$$\mathscr{L} : T_xQ \to T_x^*Q$$

$$v \mapsto \mathscr{L}(v) = \mathrm{d}\left[\frac{1}{2}F(v)^2\right] \in T^*Q.$$

▲

In components,

$$\mathscr{L}(v)_i = \frac{\partial}{\partial v^i}\left[\frac{1}{2}F(v)^2\right] = F(v)F_{v^i}(v).$$

Therefore,

$$\mathscr{L}(v) = \mathscr{L}(v)_i \mathrm{d}v^i = F(v)F_{v^i}(v)\mathrm{d}v^i \in T_x^*Q.$$

Pair with $v \in T_xQ$:

$$\langle \mathscr{L}(v), v\rangle = F(v)F_{v^i}(v)v^i = F(v)^2. \tag{4.11}$$

by the homogeneity of F (4.1). Now, consider the Legendre transform

$$p := \mathscr{L}(v) \in T_x^*Q$$

of the vector $v \in I_x \subset T_xQ$ in the indicatrix. Since $F(v) = 1$, (4.11) yields

$$\langle p, v\rangle = F(v)^2 = 1.$$

The components of p are $p_i = \dfrac{\partial F}{\partial v^i}(v)$ since $F(v) = 1$.

A general benefit of the Legendre transform is that it allows us to work in T^*Q, which has a canonical symplectic form. Additionally, T^*Q is where Hamiltonian mechanics, Hamilton's equations, the Hamilton-Jacobi equation, etc., happen. Thus, it is beneficial to be able to transition from TQ to T^*Q in a natural way.

Now, we show the criteria for a vector w to belong to T_vI_x: Let $v \in T^*Q$, $p := \mathscr{L}(v) \in T_x^*Q$, or equivalently, $p : T_xQ \to \mathbb{R}$.

Proposition 4.6.2. *The tangent space to I_x is the kernel of p.*

Proof. Let $\sigma : (-\epsilon, \epsilon) \to I_x$ be a (local) curve in I_x such that $\sigma(0) = v$ and such that $\dot{\sigma}(0) = w \in T_vI_x$.

$$\begin{aligned} 0 &= \frac{d}{dt}\bigg|_{t=0} \left[\frac{1}{2}F(\sigma(t))^2\right] \\ &= F(\sigma(0))F_{v^i}(\sigma(0))\dot{\sigma}^i(0) \\ &= F(v)F_{v^i}(v)w^i \\ &= \langle p, w\rangle. \end{aligned}$$

because F is constant along γ. □

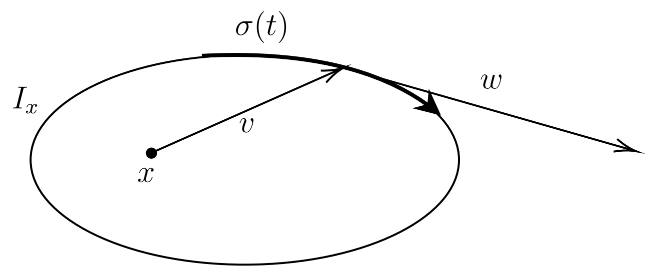

Figure 4.3: The relationship between v and w.

The Legendre transform helps determine a notion of orthogonality with respect to the Finsler metric F.

Definition 4.6.3. Let (Q, F) be a Finsler manifold, and $A \subset Q$ be a submanifold. Let $x \in A \subseteq Q$. A vector $v \in T_xQ$ is *F-orthogonal* to A if for every vector $w \in T_vA$,

$$F_{v^i}(v)w^i = 0. \tag{4.12}$$

See [24], Chapter 1, Section 6. We denote this as $v \vdash_F w$. Note that multiplying by $F(v)$ allows to write (4.12) as

$$F(v)F_{v^i}(v)w^i = 0, \tag{4.13}$$

▲

which is precisely the dual pairing $\langle \mathscr{L}(v), w\rangle$ which arises from the Lengendre transform.

Remark 4.6.4. This is an asymmetric relationship, i.e., $v \vdash_F w$ does not imply $w \vdash_F v$. ▲

Now, we observe an important relationship.

Proposition 4.6.5. *Given $x \in Q$, let $v \in I_x$, and let $w \in T_v I_x$. We have that $v \vdash_F w$.*

Proof. Follows the same idea as Proposition 4.6.2. □

4.6.2 Huygens's Principle

In Chapter 2, we defined a wavefront as a surface of constant phase. We now define them in a different manner.

To understand how waves propagate in simple terms, imagine dropping a tiny stone into still water. One will observe a circular ripple with center at the point where the stone entered the water, which expands as time passes. Now, if the medium were some anisotropic fluid rather than water, then the ripple would not be a circle, but rather some other convex curve. As it turns out, the shape of that ripple is determined completely by the Finsler metric of the medium. In fact, it is a scaled version of the indicatrix whose base point is where the stone entered the medium. This leads us to a second definition of wavefronts.

Definition 4.6.6. A *wavefront* $\psi_x(t)$ starting at a given point $x_0 \in Q$ at time t is the set

$$\psi_{x_0}(t) := \Big\{ v \in T_{x_0}Q : F(v) = t \Big\}.$$

It is a scaled version of the indicatrix. ▲

Now, take some wavefront $\psi_{x_0}(t)$. What is the wavefront at time $t + \delta t$, for δ very small? Take two points $x, \tilde{x} \in \psi_{x_0}(t)$ very close to each other. Consider the indicatrices I_x and $I_{\tilde{x}}$; these describe propagation at x and $\tilde{x}$, respectively, after time δt. One can consider the envelope $\mathscr{E}$ of I_x and $I_{\tilde{x}}$, i.e., the curve whose tangent space coincides with both $T_v I_x$ (for some $v \in I_x$) and $T_{\tilde{v}} I_{\tilde{x}}$ (for some $\tilde{v} \in I_{\tilde{x}}$). Said another way, the tangent space of $\mathscr{E}$ should belong to the kernel of both $\mathscr{L}(v)$ and $\mathscr{L}(\tilde{v})$. As it turns out, the envelope $\mathscr{E}$ is exactly the wavefront at time $t + \delta t$. This was first discovered by the great Dutch astronomer, Christiaan Huygens, whose name emblazens the principle:

Theorem 4.6.7 (Huygens's Principle)**.** *Let $\psi_{x_0}(t)$ denote the wavefront at time t starting at x_0 at time 0. For every $x \in \psi_{x_0}(t)$, consider the wavefront after time s, $\psi_x(s)$. Then, the wavefront of the point x_0 after time $s + t$, $\psi_{x_0}(s + t)$, will be the envelope of the wavefronts $\psi_x(s)$.*

Proof. See [2], section 46, for a proof which involves the Fermat principle. □

Now, we are ready to show the vital relationship between rays and wavefronts.

Proposition 4.6.8. *Light rays and wavefronts are F-orthogonal.*

Proof. Let $x_0 \in Q$ be given. Let $\psi_{x_0}(t)$ denote the wavefront at time t starting at x_0. Let $\gamma : [0,1] \to Q$ be a given light ray such that $\gamma(0) = x_0$. To show F-orthogonality, we show that for a given wavefront $\psi_{x_0}(t_0)$, the tangent spaces $T_{\gamma(t_0)}\psi_{x_0}(t_0)$ and $T_{\dot{\gamma}(t_0)}I_{\gamma(t_0)}$ are the same.

Let $\psi_{\gamma(t_0)}(\delta t)$ be the wavefront at the point $\gamma(t_0) \in \psi_{x_0}(t_0)$ after time δt, and let $\psi_{x_0}(t_0 + \delta t)$ be the wavefront at x_0 after time $t + \delta t$. By Huygens's Principle, $\psi_{x_0}(t_0 + \delta t)$ is the envelope of all $\psi_x(\delta t)$ for $x \in \psi_{x_0}(t_0)$. Thus, any vector v which is F-orthogonal to a vector $w \in T_v\psi_x(\delta t)$ is also F-orthogonal to a vector which is in $T_v\psi_{x_0}(t_0 + \delta t)$.

Note that $\psi_{\gamma(t_0)}(\delta t)$ is simply a scaled version of the indicatrix $I_{\gamma(t_0)}$ at the point $\gamma(t_0)$. By Proposition 4.6.5, given $V := \dfrac{\dot{\gamma}(t_0)}{F(\dot{\gamma}(t_0))} \in I_{\gamma(t_0)}$, $W \in T_V I_{\gamma(t_0)}$, we have that $V \vdash_F W$. As $\delta \to 0$, the tangent spaces $T_V I_{\gamma(t_0)}$ and $T_{\gamma(t_0)}\psi_{x_0}(t_0)$ are identified. So, W can be thought of as belonging to $T_{\gamma(t_0)}\psi_{x_0}(t_0)$. Therefore, for $\tilde{W} \in T_{\gamma(t_0)}\psi_{x_0}(t_0)$, $V \vdash_F \tilde{W}$. In other words, the light ray γ is F-orthogonal to the wavefront $\psi_{x_0}(t_0)$. $\square$

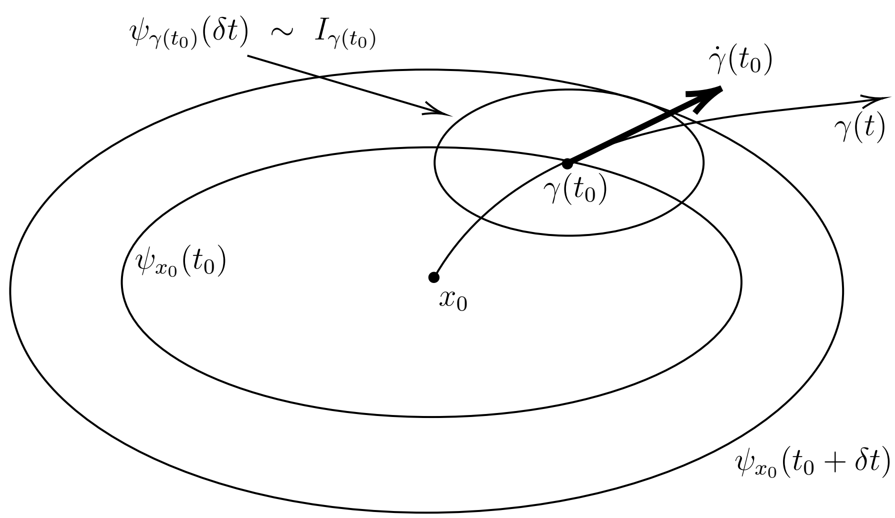

Figure 4.4: A ray $r(t)$ intersecting the wavefronts

Corollary 4.6.9. *Light rays are F-orthogonal to the light source.*

Proof. This follows because a light source can be thought of as a wavefront at time $t = 0$. $\square$

4.7 Conditions Rays Must Satisfy

As a consequence of the previous section, we can now explicitly state conditions a ray must satisfy. Let $\psi_{x_0}(t)$ denote the wavefront at $x_0 \in Q$ at time t. Let $\gamma : [0,1] \to Q$ be a curve such that $\gamma(0) = x_0$, $\gamma(t) \in \psi_{x_0}(t)$. Let us write $v := \dfrac{\dot{\gamma}(t)}{F(\dot{\gamma}(t))} \in I_{\gamma(t)}$, and let $w \in T_v I_{\gamma(t)}$. Then, a ray must satisfy the following conditions: for all $t \in (-\epsilon, \epsilon)$,

(1) $F(v) = 1$ (or equivalently, $F^2(v) = 1$)

(2) $F_{v^i}(v) w^i = 0$.

Remark 4.7.1. When we discuss caustics, we wish to express the caustic in terms of the light source $\mathscr{S}$. Let $X(y)$ be a parameterization for $\mathscr{S}$, and let $\dot{X}(y)$ be a tangent vector to $\mathscr{S}$. The specific conditions we will use to determine the caustic are:

(1) $F(v) = 1$ (or equivalently, $F^2(v) = 1$)

(2) $F_{v^i}(v) \dot{X}^i(y) = 0$.

▲

4.8 Example: Randers Metric

A useful class of Finsler metrics are the so-called Randers metrics.

Definition 4.8.1. A *Randers metric* is a Finsler structure F on TQ of the form

$$F(x, v) = \alpha(x, v) + \beta(x, v)\,, \tag{4.14}$$

where

- $\alpha(x, v) = \sqrt{\alpha_{ij}(x)v^i v^j}$, where the α_{ij} are the components of a Riemannian metric
- $\beta(x, v) = \beta_i(x)v^i$, where the β_i are the components of a one-form $\beta \in \Omega^1(Q)$.

▲

In order for the positivity criterion of F to be satisfied, a sufficient (but not necessary) condition is

$$\sqrt{\beta_i \alpha^{ij} \beta_j} < 1. \tag{4.15}$$

As it turns out, this detail also ensures the strong convexity condition; see [3], section 1.3C.

Randers metrics are useful for describing propagation in a variety of asymmetric contexts. Such examples include:

- Light propagation in a moving homogeneous medium
- Wildfire propagation in the presence of wind.

To show an interesting property regarding the indicatrix of a Randers metric, we first need the following lemma.

Lemma 4.8.2. *A Finsler metric F in a 2-dimensional space is Riemannian iff its indicatrix is an ellipse centered at the origin.*

Proof. Suppose that for each $x \in Q$, the indicatrix of a Finsler metric F can be written as

$$\frac{(v^1)^2}{a^2(x)} + \frac{(v^2)^2}{b^2(x)} = 1.$$

This is the indicatrix of a metric whose metric tensor is given by the matrix

$$\alpha = \begin{pmatrix} \frac{1}{a^2(x)} & 0 \\ 0 & \frac{1}{b^2(x)} \end{pmatrix}.$$

Such a metric is Riemannian.

Now, suppose F is a Riemannian metric with metric tensor α (see (4.14)). Since F is Riemannian, it is known that α is a quadratic form and is symmetric. Therefore, the solution set to

$$\alpha_{ij}(x)v^i v^j = 1$$

in TQ is an ellipse. Thus, we have our desired result. □

Theorem 4.8.3. *A Finsler metric F is Randers iff its indicatrix is an ellipse with one focus at the origin and major axis in coincidence with the coordinate axis.*

Proof. For simplicity, we will prove this for the case $n = 2$.

"$\Longleftarrow$": First, assume that, without loss of generality, for a fixed x, I_x is an ellipse with the major axis on the v^1-axis. The equation for such an ellipse is

$$\frac{(v^1 - c)^2}{a^2} + \frac{v^2}{b^2} = 1,$$

where

- a is the length of the semi-major axis,

- b is the length of the semi-minor axis, and

- c is the focal length, i.e.,

$$c = \sqrt{a^2 - b^2}. \tag{4.16}$$

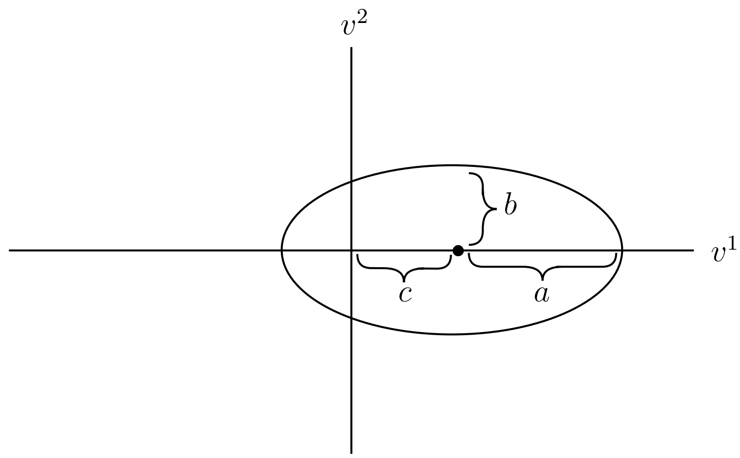

Figure 4.5: An elliptical indicatrix with a focus at the origin

The polar equation for such an ellipse is

$$v(\theta) = \frac{p}{1 - \epsilon \cos \theta},$$

where $\epsilon \in (0, 1)$, $p = a(1 - \epsilon^2)$. Using (4.10), we can write

$$F(v) = \frac{1}{p}\Big[|v| - |v|\epsilon \cos \theta\Big].$$

Since $\cos\,\theta = \dfrac{v^1}{|v|}$, we can write

$$F(v) = \frac{1}{p}\Big[|v| - \epsilon\, v^1\Big], \tag{4.17}$$

which is Randers since $|v|$ is a Riemannian metric.

"$\Longrightarrow$": Now, suppose F is Randers. Then, its indicatrix can be given by an ellipse (due to the Riemannian component) which is shifted, but still contains the origin in its interior due to the condition (4.15) on the one-form. To see this, let $v \in I_x$, so it satisfies

$$1 = \sqrt{A_{ij}v^i v^j} + B_i v^i.$$

Squaring both sides and rearranging terms gives us

$$(A_{ij} - B_i B_j)v^i v^j + 2B_i v^i - 1 = 0,$$

which is the equation of an ellipse in TQ due to the constraint (4.15) on the B_i.

Now, we discuss the specific case where $n = 2$. Given a general ellipse

$$A_{11}(v^1)^2 + 2A_{12}v^1 v^2 + A_{22}(v^2)^2 + B_1 v^1 + B_2 v^2 + C = 0, \qquad (4.18)$$

our goal will be:

(1) Rotate the ellipse via an orthogonal transformation, so that the major axis is parallel to a coordinate axis,

(2) Perform a twist/shear so the major axis of the ellipse coincides with a coordinate axis, and

(3) Rescale so one focus of the ellipse is at the origin.

We now perform these steps in turn.

(1) & (2) Conveniently, we can do the first two steps at once. We want to get rid of the mixed term $v^1 v^2$ in (4.18) and either v^1 or v^2. This will both rotate the ellipse appropriately, as well as make an axis of the ellipse coincide with a

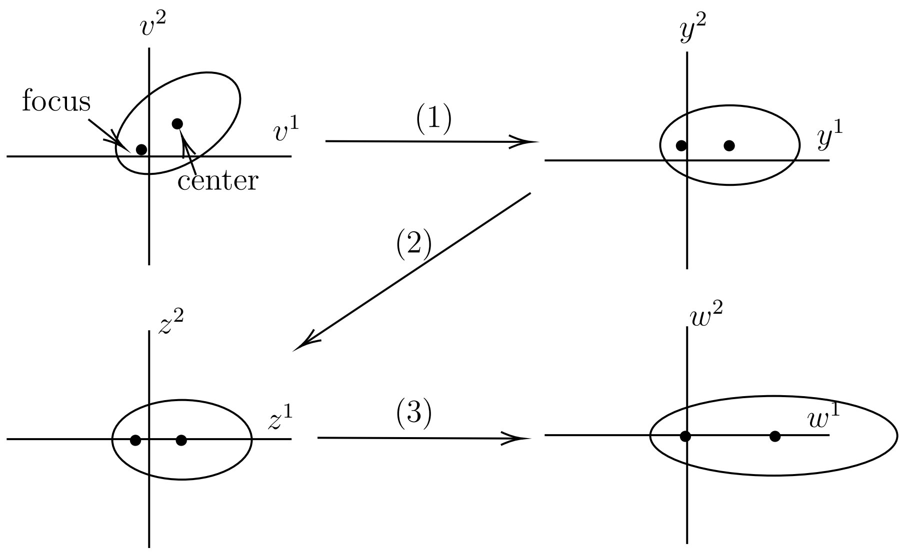

Figure 4.6: Illustration of the transformation of coordinates

coordinate axis. Let

$$v^1 = ay^1 + by^2 \tag{4.19}$$

$$v^2 = cy^1 + dy^2. \tag{4.20}$$

(Note: a, b, and c are arbitrary constants here, not to be confused with the dimensions of the ellipse.) This is an underdetermined system, so we can find nice solutions for the coefficients. Substituting v^1 and v^2 into (4.18) yields

$$\begin{aligned} &A_{11}\Big[a^2(y^1)^2 + 2aby^1y^2 + b^2(y^2)^2\Big] \\ &+ 2A_{12}\Big[ac(y^1)^2 + (ad+bc)y^1y^2 + bd(y^2)^2\Big] \\ &+ A_{22}\Big[c^2(y^1)^2 + 2cdy^1y^2 + d^2(y^2)^2\Big] \\ &+ B_1(ay^1 + by^2) + B_2(cy^1 + dy^2) + C = 0. \end{aligned} \tag{4.21}$$

Combining like y-terms gives us

$$\begin{aligned}&\Big[A_{11}a^2 + 2A_{12}ac + A_{22}c^2\Big](y^1)^2 + \Big[A_{11}b^2 + 2A_{12}bd + A_{12}d^2\Big](y^2)^2\\&+ \Big[2A_{11}ab + 2A_{12}ad + 2A_{12}bc + A_{22}cd\Big]y^1y^2 + \Big[B_1a + B_2c\Big]y^1\\&+ \Big[B_1b + B_2d\Big]y^2 + C = 0.\end{aligned} \tag{4.22}$$

Since we want the major axis to coincide with a coordinate axis, we need to get rid of the mixed terms and either the first-order y^1 or y^2 term. Thus, we need the following to be satisfied:

$$A_{11}ab + A_{12}ad + A_{12}bc + A_{22}cd = 0, \tag{4.23}$$

and at least one of the following to be satisfied:

$$B_1a + B_2c = 0 \tag{4.24}$$

$$B_1b + B_2d = 0. \tag{4.25}$$

Without loss of generality, assume the first equation can be satisfied. Solving for c gives

$$c = -\frac{B_1}{B_2}a. \tag{4.26}$$

Plugging this into (4.23) gives

$$A_{11}ab + A_{12}ad + A_{12}b\left[-\frac{B_1}{B_2}a\right] + A_{22}d\left[-\frac{B_1}{B_2}a\right] = 0. \tag{4.27}$$

We can factor out a:

$$a\left[A_{11}b + A_{12}d - \frac{A_{12}B_1}{B_2}b - \frac{A_{22}B_1}{B_2}d\right] = 0.$$

If $a \neq 0$, we have

$$A_{11}b + A_{12}d - \frac{A_{12}B_1}{B_2}b - \frac{A_{22}B_1}{B_2}d = 0.$$

Solve for d in terms of b:

$$d = b\frac{A_{12}B_1 - A_{11}B_2}{A_{12}B_2 - A_{22}B_1}. \tag{4.28}$$

Applying c and d to the change of coordinates given by (4.19) and (4.20), and applying this change of coordinates to (4.18) yields

$$\begin{aligned} A_{11}(ay^1 + by^2)^2 + 2A_{12}(ay^1 + by^2)\bigg(-\frac{B_1}{B_2}a + b\frac{A_{12}B_1 - A_{11}B_2}{A_{12}B_2 - A_{22}B_1}\bigg) \\ + A_{22}\bigg(-\frac{B_1}{B_2}a + b\frac{A_{12}B_1 - A_{11}B_2}{A_{12}B_2 - A_{22}B_1}\bigg)^2 + B_1(ay^1 + by^2) \\ + B_2\bigg(-\frac{B_1}{B_2}a + b\frac{A_{12}B_1 - A_{11}B_2}{A_{12}B_2 - A_{22}B_1}\bigg) + C = 0. \end{aligned} \tag{4.29}$$

Reshuffling terms, completing the square, etc., gives an ellipse of the form

$$\frac{(y^1)^2}{m^2} + \frac{(y^2 - \alpha)^2}{n^2} = 1. \tag{4.30}$$

Note: if $\alpha = 0$, we are in the case of a Riemannian metric.

Remark 4.8.4. In the special case where, say, $B_2 = 0$, then either $a = 0$ or $b = 0$ by equations (4.24) and (4.25). Without loss of generality, if $b = 0$ and

$a \neq 0$, then equation (4.23) becomes

$$A_{12}ad + A_{22}cd = 0.$$

Since $d \neq 0$, we have

$$c = -\frac{A_{12}}{A_{22}}a.$$

The numbers a and d are free. ▲

Remark 4.8.5. In the special case that the denominator in (4.28) equals zero, we arrive at a similar conclusion. We have

$$0 = b[A_{12}B_1 - A_{11}B_2].$$

We can take $b = 0$ and arrive at the same conclusion as in the previous remark. ▲

(3) Finally, we rescale so that the focus of the ellipse is at the origin. Let

$$y^1 = w^1 \tag{4.31}$$

$$y^2 = \frac{w^2}{\eta}. \tag{4.32}$$

Substituting y^1 and y^2 into (4.30) yields

$$\frac{(w^1)^2}{m^2} + \frac{(w^2 - \alpha\eta)^2}{n^2\eta^2} = 1. \tag{4.33}$$

For this to represent the equation of an ellipse with a focus at the origin, (4.16) must be satisfied. Therefore,

$$\eta = \sqrt{\frac{m^2}{n^2 - \alpha^2}}.$$

Plugging this value of η into (4.33) gives us our desired result.

$\square$

Chapter 5

Morse Families in Anisotropic Media

Now armed with tools from Finsler geometry, we are equipped to explore whether Morse families can be applied to anisotropic media.

5.1 Geodesics in the Finsler Setting

According to variational principles (see Section 1.1.3), light travels along the paths of a manifold Q which minimize time. In other words, light travels along the geodesics of the Q.

However, if Q is equipped with a Finsler structure, how does one determine geodesics? There isn't quite an analogue to the Levi-Civita connection in the Finsler setting. In fact, in a general Finsler setting, there is no connection which is both metric-compatible and torsion-free; see [3] for many examples of such connections. Nevertheless, one must still be able to write the equations geodesics must satisfy in a given Finsler space.

Let $\gamma : [0,1] \to Q$ be a parameterized curve with endpoints denoted by $\gamma(0)$ and $\gamma(1)$. The length of γ is given by

$$L(\gamma) = \int_0^1 F\Big(\gamma(t), \dot{\gamma}(t)\Big)\, dt. \tag{5.1}$$

One can show that L doesn't depend on a change of coordinates on TQ, nor does it depend on the paramaterization of γ due to the homogeneity of F. Thus, L only depends on the curve γ itself.

Now, consider the variational problem for (5.1) with fixed endpoints. Consider the set of curves

$$\gamma_\epsilon : t \in [0,1] \mapsto \Big(\gamma^i(t) + \epsilon V^i(t)\Big) \in Q$$

which have the same endpoints $\gamma^i(0)$ and $\gamma^i(1)$ as γ, $V^i(t) = V^i(\gamma(t))$ a regular vector field on γ with the property $V^i(0) = V^i(1) = 0$, and $\epsilon \in \mathbb{R}$ sufficiently small. The length of γ_ϵ is

$$L(\gamma_\epsilon) = \int_0^1 F\Big(\gamma + \epsilon V, \dot{\gamma} + \epsilon \frac{\mathrm{d}V^i}{\mathrm{d}t}\Big)\, dt. \tag{5.2}$$

The necessary condition for γ to be an extremal of Ł(γ_ϵ) is

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}(L(\gamma_\epsilon))\bigg|_{\epsilon=0} = 0. \tag{5.3}$$

Taking $\dfrac{\mathrm{d}}{\mathrm{d}\epsilon}$ of (5.2) and setting it equal to zero, we arrive at the so-called Euler-Lagrange equations:

Theorem 5.1.1. *The functional $L(\gamma)$ is an extremal of $L(\gamma_\epsilon)$ if γ satisfies the Euler-Lagrange equations for F:*

$$\frac{\partial F}{\partial \gamma^i} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial F}{\partial \dot{\gamma}^i} = 0.$$

From here, one can then show (see [8], Section 5.3) that the geodesics of Q are solutions of the equations

$$\ddot{\gamma} + 2G^i(\dot{\gamma}) = 0, \tag{5.4}$$

where for $v \in TQ\backslash\{0\}$,

$$G^i(x,v) = \frac{1}{4} g^{ik}(x,v)\left(2\frac{\partial g_{jk}}{\partial x^l}(x,v) - \frac{\partial g_{jl}}{\partial x^k}(x,v)\right)v^j v^l, \tag{5.5}$$

where $\big(g\big) = \big(g_{ik}(x,v)\big)$ is a matrix whose i,k-th component is defined as in Definition 4.2.2, and where $g^{ik}(x,v)$ is the i,k-th component of $\big(g^{-1}\big)$.

5.2 Light Rays in Anisotropic, Homogeneous Media

First, we will consider the case in which the media is anisotropic but homogeneous, as it is quite instructive. In homogeneous media, the composition of the media is completely uniform throughout. In the case of *isotropic* media, propagation is independent of the direction of travel. Recalling Huygens's principle, such propagation can be modeled by an indicatrix which is a circle centered at the origin of the tangent space of some point x. On the other hand, in *anisotropic* media, propagation depends on the direction of travel at a given point. Thus, the indicatrices here will be given by a general convex curve.

Consider the case of homogeneous media. Here, the indicatrix will be the same at every point of the manifold. Furthermore, since the metric F is completely determined by its indicatrix, then we can say that F has no dependence on the point x, i.e., we can write $F(x,v) = F(v)$.

What do light rays look like in anisotropic, homogeneous media? If F has no dependence on x, then the G^i from (5.5) vanish, and thus (5.4) reduces to

$$\ddot{\gamma} = 0.$$

Thus, geodesics are straight lines.

5.3 Optical Path in Homogeneous Medium is a Morse Family

We begin with the most basic case of a manifold Q which is a vector space. Recall that the time of propagation along a curve $\gamma : [0, 1] \to Q$ is given by

$$M(x, y) = \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt, \tag{5.6}$$

where $\gamma(0) = X(y) \in \mathscr{S}$ is a point in the light source $\mathscr{S}$, $\gamma(1) = x \in Q$. We wish to write the right-hand side of (5.6) in terms of x and y. Since the choice of γ depends on $x \in Q$ and the point $X(y) \in \mathscr{S}$, we can now write $\gamma(t) =: \tilde{x}(x, y, t)$ and $\dot{\gamma}(t) =: \tilde{v}(x, y, t)$. Suppose we are in a homogeneous medium, i.e., F does not depend on the point on the manifold. We write $F(\tilde{x}, \tilde{v}) = F(\tilde{v})$, and thus

$$M(x, y) = \int_0^1 F(\tilde{v}(x, y, t))\, dt.$$

Recall that in this case, light rays are straight lines. Furthermore, by homogeneity, F is independent of parameterization, so we can explicitly define $\tilde{x}$ and $\tilde{v}$ in a convenient manner:

$$\begin{aligned}\tilde{x}: \ \gamma(t) &= X(y)(1-t)+xt \\ \tilde{v}: \ \dot{\gamma}(t) &= x - X(y).\end{aligned} \tag{5.7}$$

Note: $\tilde{v}$ is independent of t, so we can drop the integral and write

$$\begin{aligned}M(x,y) &= F(\tilde{v}(x,y,t)) \\ &= F(x-X(y)).\end{aligned} \tag{5.8}$$

Theorem 5.3.1. *The function M is a Morse family.*

Proof. We need to show that the matrix

$$\begin{pmatrix} \dfrac{\partial^2 M}{\partial x^i \partial y^\alpha} & \dfrac{\partial^2 M}{\partial y^\beta \partial y^\alpha} \end{pmatrix} \tag{5.9}$$

has maximal rank on the fiber-critical manifold B_M. We have:

$$\frac{\partial M}{\partial x^i}(x,y) = \frac{\partial F}{\partial \tilde{v}^i}(\tilde{v}); \tag{5.10}$$

$$\begin{aligned}\frac{\partial M}{\partial y^\alpha}(x,y) &= \frac{\partial F}{\partial \tilde{v}^l}(\tilde{v}) \cdot \frac{\partial \tilde{v}^l}{\partial y^\alpha}(y) \\ &= -\frac{\partial F}{\partial \tilde{v}^l}(\tilde{v}) \cdot \frac{\partial X^l}{\partial y^\alpha}(y);\end{aligned} \tag{5.11}$$

$$\frac{\partial^2 M}{\partial x^i \partial y^\alpha}(x,y) = -\frac{\partial^2 F}{\partial \tilde{v}^i \partial \tilde{v}^l}(\tilde{v}) \cdot \frac{\partial X^l}{\partial y^\alpha}(y); \tag{5.12}$$

$$\frac{\partial^2 M}{\partial y^\beta \partial y^\alpha}(x,y) = \left(\frac{\partial^2 F}{\partial \tilde{v}^j \partial \tilde{v}^l}(\tilde{v}) \cdot \frac{\partial X^l}{\partial y^\beta}(y)\right)\frac{\partial X^j}{\partial y^\alpha}(y) - \frac{\partial F}{\partial \tilde{v}^j}(\tilde{v}) \cdot \frac{\partial^2 X^j}{\partial y^\beta \partial y^\alpha}(y). \tag{5.13}$$

Now, we show that (5.9) has full rank on B_M. Recall that by definition,

$$
\begin{aligned}
g_{il}(\tilde{v}) &= \left[\frac{1}{2}F^2(\tilde{v})\right]_{\tilde{v}^i\tilde{v}^l} \\
&= F_{\tilde{v}^i}(\tilde{v})F_{\tilde{v}^l}(\tilde{v}) + F(\tilde{v})F_{\tilde{v}^i\tilde{v}^l}(\tilde{v}).
\end{aligned} \tag{5.14}
$$

Therefore,

$$F_{\tilde{v}^i\tilde{v}^l} = \frac{g_{il} - F_{\tilde{v}^i}F_{\tilde{v}^l}}{F}.$$

So, we plug this into (5.12):

$$\frac{\partial^2 M}{\partial x^i \partial y^\alpha} = -\left(\frac{g_{il} - F_{\tilde{v}^i}F_{\tilde{v}^l}}{F}\right) \cdot \frac{\partial X^l}{\partial y^\alpha}. \tag{5.15}$$

Recall the condition for B_M in (3.12). Thus, $\dfrac{\partial M}{\partial y^\alpha} = -\dfrac{\partial F}{\partial \tilde{v}^l}(\tilde{v}) \cdot \dfrac{\partial X^l}{\partial y^\alpha}(y) = 0$, so we get

$$\left.\frac{\partial^2 M}{\partial x^i \partial y^\alpha}\right|_{B_M} = -\frac{1}{F} g_{il}(\tilde{v}) \cdot \frac{\partial X^l}{\partial y^\alpha}. \tag{5.16}$$

Note that this is the α, i-th entry of the product of two matrices: the $n \times n$ matrix $\Big(g_{il}(\tilde{v})\Big)$ and the $n \times k$ matrix $\left(\dfrac{\partial X^l}{\partial y^\alpha}\right)$. We have that $\left(\dfrac{\partial X^l}{\partial y^\alpha}\right)$ has full rank k, since the $\dfrac{\partial X}{\partial y^\alpha}$ are assumed to be linearly independent. Furthermore, the matrix $\Big(g_{il}(\tilde{v})\Big)$ is full-rank since it is positive-definite. Thus, their product is rank k. Therefore, the submatrix $\left.\left(\dfrac{\partial^2 M}{\partial x^i \partial y^\alpha}\right)\right|_{B_M}$ is always rank k, and thus we have achieved our desired result.

$\square$

Corollary 5.3.2. *If F defines a Randers metric in a homogeneous medium, then M is a Morse family.*

Remark 5.3.3. Corollary 5.3.2 can also be proven directly, similar to Proposition

3.6.1. ▲

Of course, the above proof relies heavily on the manifold Q having a vector space structure. However, if Q is a general manifold whose metric tensor does not depend on $x \in Q$ a local neighborhood of x, but whose tangent bundle is trivial, then a point $(x.v) \in TQ$ can be described as a point in $Q \times \mathbb{R}^n$. Thus, M as defined in (5.8) can be used locally on Q, provided the choice of neighborhood of $x \in Q$ doesn't cross the cut locus of Q, as the function M would no longer be smooth in this neighborhood of x. (The *cut locus* of a point $x \in Q$ is defined as the closure of the set of points $\tilde{x} \in Q$ such that x and $\tilde{x}$ are connected by two or more distinct "shortes" geodesics.)

5.4 Optical Path in the General Case

Now, we consider the case where the media may be inhomogeneous and anisotropic. Here, we aren't as fortunate as in the case of homogeneous media, as light rays are no longer straight lines. In fact, the equations of the light rays can be incredibly complicated in general. So, there isn't a convenient way to rewrite the optical path M as there was in the case of homogeneous media. Furthermore, if the manifold lacks a natural vector space structure and doesn't have a trivial tangent bundle, we need more than the previous theorem. But, as we will show, if the Finsler metric is "close" to being Euclidean, then we have a Morse family.

Theorem 5.4.1. *Let M be as described in* (5.6)*. Suppose F is a small perturbation of the Euclidean metric, and suppose the light source paramaterized by $X(y)$ has ϵ-small curvature. Then, M is a Morse family on any compact subset of the fiber-critical manifold B_M.*

Proof. We need to show that $\mathrm{Im}\,\mathrm{d}M$ and $N^*\Pi$ have a transverse intersection. Recall that this means that for every $\nu \in L_M$,

$$T_\nu N^*\Pi \oplus T_\nu \mathrm{Im}\,\mathrm{d}M = T_\nu T^*B \ .$$

Notice that this condition is equivalent to

$$(T_\nu N^*\Pi)^{\omega_B} \cap T_\nu \mathrm{Im}\,\mathrm{d}M = \{0\} \ ,$$

(see [6], Remark 3.6). Furthermore, in the proof of Proposition 3.4.9, we showed that this is precisely the condition

$$\ker\, T_\nu \lambda_M = \{0\},$$

where $\lambda_M : L_M \to T^*Q$, $(x, y; p_x, 0) \mapsto (x, p_x)$.

Let $\eta : [0,1] \to L_M$ be a smooth curve in L_M. As the x- and y-components of η come from the fiber-critical manifold B_M, they must satisfy

$$\frac{\partial M}{\partial y^\alpha}\big(x(t), y(t)\big) = 0, \tag{5.17}$$

for every $\alpha = 1, ..., k$. The tangent vector $\dot\eta \in T_\eta L_M$ is

$$\dot\eta(t) = \dot x^i \frac{\partial}{\partial x^i} + \dot y^\alpha \frac{\partial}{\partial y^\alpha} + \left(\frac{\partial^2 M}{\partial x^i \partial x^j} \dot x^j + \frac{\partial^2 M}{\partial x^i \partial y^\alpha} \dot y^\alpha \right) \frac{\partial}{\partial (p_x)_i} \ , \tag{5.18}$$

with conditions on $\dot x^i$ and $\dot y^\alpha$ derived by taking $\dfrac{\mathrm{d}}{\mathrm{dt}}$ of (5.17)

$$\frac{\partial^2 M}{\partial y^\alpha \partial x^i} \dot x^i + \frac{\partial^2 M}{\partial y^\alpha \partial y^\beta} \dot y^\beta = 0, \quad \alpha = 1, ..., k\,. \tag{5.19}$$

We observe that

$$T_\eta \lambda_M(\dot{\eta}) = \dot{x}^i \frac{\partial}{\partial x^i} + \left(\frac{\partial^2 M}{\partial x^i \partial x^j} \dot{x}^j + \frac{\partial^2 M}{\partial x^i \partial y^\alpha} \dot{y}^\alpha \right) \frac{\partial}{\partial (p_x)_i},$$

with conditions (5.19) on the x^i and y^α.

Now, suppose $\dot{\eta} \in \ker T_\eta \lambda_M$. Then, the following conditions are satisfied for all i:

$$\begin{aligned} \dot{x}^i &= 0 \\ \frac{\partial^2 M}{\partial x^i \partial x^j} \dot{x}^j + \frac{\partial^2 M}{\partial x^i \partial y^\alpha} \dot{y}^\alpha &= 0 \\ \frac{\partial^2 M}{\partial y^\alpha \partial x^i} \dot{x}^i + \frac{\partial^2 M}{\partial y^\alpha \partial y^\beta} \dot{y}^\beta &= 0. \end{aligned} \tag{5.20}$$

This simplifies to

$$\begin{aligned} \dot{x}^i &= 0 \\ \frac{\partial^2 M}{\partial x^i \partial y^\alpha} \dot{y}^\alpha &= 0 \\ \frac{\partial^2 M}{\partial y^\alpha \partial y^\beta} \dot{y}^\beta &= 0. \end{aligned} \tag{5.21}$$

So, since $\dot{\eta} \in \ker T_\eta \lambda_M$, its projection to the tangent vector $\dot{G} \in T_{\pi_B(\eta)} B_M$ is

$$\dot{G} = \dot{y}^\alpha \frac{\partial}{\partial y^\alpha},$$

with conditions

$$\begin{aligned}\frac{\partial^2 M}{\partial x^i \partial y^\alpha}\dot{y}^\alpha &= 0 \\ \frac{\partial^2 M}{\partial y^\alpha \partial y^\beta}\dot{y}^\beta &= 0.\end{aligned} \tag{5.22}$$

For M to be a Morse family, we require $\ker T_\eta \lambda_M = \{0\}$. However, in general,

$$\ker T_\eta \lambda_M \subset \operatorname{span}\left\{ \left(\frac{\partial}{\partial y^\alpha}\right)_\eta \right\}.$$

Thus, we need to show that for the curve G satisfying (5.22), we always have $\dot{y}^\alpha = 0$ for all $\alpha = 1, ..., k$.

It may happen that there are points in G where the matrix $\left(\frac{\partial^2 M}{\partial y^\alpha \partial y^\beta}\right)$ fails to have full rank. Recall from (3.37) that the singular set $\Sigma(B_M)$ is defined as

$$\Sigma(B_M) := \left\{ (x, y) \in B_M : \det\left(\frac{\partial^2 M}{\partial y^\alpha \partial y^\beta}(x, y)\right) = 0 \right\}.$$

First, suppose G doesn't intersect $\Sigma(B_M)$. In this case, $\left(\frac{\partial^2 M}{\partial y^\alpha \partial y^\beta}\right)$ has full rank on the entire domain of G. Thus, it follows from (5.21) that $\dot{y}^\alpha = 0$.

Next, suppose G intersects $\Sigma(B_M)$ at some point b. As G is a smooth curve, the set R of points on G for which $\dot{y}^\alpha = 0$ is a closed set. Since b can be approximated by a sequence of points $\{b_r\}_{r=1}^\infty \in R$, b is a limit point of R. As R is closed, $b \in R$. Thus, at b, we also have that $\dot{y}^\alpha = 0$.

Before considering the case where $G \subset \Sigma(B_M)$ locally, we need to show that for the function M (5.6) which describes light ray propagation, the condition

$$\dim\, \Sigma(B_M) < \dim\, B_M$$

is satisfied. (Up to this point, we haven't used the fact that M describes propagation.)

Recall that for a light ray $\gamma : [0,1] \to Q$, the time of propagation from a point $\gamma(0) = X(y) \in \mathscr{S}$ in the light source $\mathscr{S}$ to a point $\gamma(1) = x \in Q$ is

$$M(x,y) = \int_0^1 F(\gamma(t), \dot{\gamma}(t))\, dt,$$

where F is the Finsler structure (4.9) which describes propagation. Consider another Finsler structure $\tilde{F}$ defined as

$$\tilde{F}(x,v) := F(x,-v).$$

We now have

$$M(x,y) = \int_0^1 \tilde{F}(\gamma(t), -\dot{\gamma}(t))\, dt.$$

Now, reparameterize γ: define $\mu : [0,1] \to Q$, and let $\mu(t) := \gamma(1-t)$. This allows to write M as

$$M(x,y) = \int_0^1 \tilde{F}(\mu(t), \dot{\mu}(t))\, dt.$$

We will consider this interpretation of M. When we compute $\dfrac{\partial M}{\partial y^\alpha}$ and $\dfrac{\partial^2 M}{\partial y^\alpha \partial y^\beta}$, we consider x to be fixed. Previously, this meant we considered $\gamma(1) = x$ fixed. Now, we consider $\mu(0) = x$ to be fixed. So, under this new interpretation, $M(x,y)$ is the travel time *from* the fixed point x *to* the point $X(y) \in \mathscr{S}$.

Pick $x^* \in Q$ fixed. Consider the travel time from x^* to some other $x \in Q$ (so in this case, we are only considering M as a function of x). As we will soon use a perturbative argument, let us first consider F to be the standard flat Euclidean metric in a homogeneous medium, i.e., where F doesn't depend on $x \in Q$. One can easily show that rank $\Big(M_{xx}\Big) = n - 1$. Indeed, if we let $x^* = 0$ for simplicity, then

$$M(x) = \sqrt{\sum_{i=1}^{n} (x^i)^2}.$$

Then,

$$M_{x^i} = \frac{x^i}{M} \tag{5.23}$$

$$M_{x^i x^j} = -\frac{x^i x^j}{M^3} \quad (i \neq j) \tag{5.24}$$

$$M_{x^i x^j} = \frac{M^2 - (x^i)^2}{M^3} \quad (i = j). \tag{5.25}$$

Therefore, we write

$$\Big(M_{xx}\Big) = \frac{1}{M^3} \begin{pmatrix} M^2 - (x^1)^2 & -x^1 x^2 & ... & -x^1 x^n \\ -x^2 x^1 & M^2 - (x^2)^2 & ... & -x^2 x^n \\ ... & ... & ... & ... \\ -x^n x^1 & -x^n x^2 & ... & M^2 - (x^n)^2 \end{pmatrix}. \tag{5.26}$$

By solving the system of equations

$$x^1\Big(M^2-(x^1)^2-(x^2)^2-...-(x^n)^2\Big)=0$$

$$x^2\Big(-(x^1)^2+M^2-(x^2)^2-...-(x^n)^2\Big)=0$$

$$...$$

$$x^n\Big(-(x^1)^2-(x^2)^2-...+M^2-(x^n)^2\Big)=0,$$

one can easily show that the vector $(x^1,...,x^n)\in\ker M_{xx}$, as the parenthetical terms are all zero. Thus, the rank of $\Big(M_{xx}\Big)$ can be no greater than $n-1$.

Now, observe that for any point $x\neq 0$ of the form $(x^1,0,...,0)$, the matrix $\Big(M_{xx}\Big)$ will have rank $n-1$. By rotational symmetry of $\mathbb{R}^n$, one can conclude that the rank will be $n-1$ for *any* $x\in\mathbb{R}^n$.

We now observe a key fact: on the space of $n\times n$ matrices of rank $n-1$ (or less), the maximality of rank is an open condition. For example, given some $n\times n$ matrix A of rank $(n-1)$, one can find an $(n-1)\times(n-1)$ submatrix $\tilde{A}$ of A whose rank is maximal, and slightly perturbing $\tilde{A}$ will not cause it to lose maximal rank. What does mean in our context? If the Finsler metric F is a small perturbation of the Euclidean metric, the rank of $\Big(M_{xx}\Big)$ will not change.

Now, suppose F is such a perturbation of the Euclidean metric. When we deal with light propagation, we always assume that the dimension of the source is strictly less than that of the ambient space. Said another way, when we deal with the optical path M, we always assume that $k\leq n-1$. So, we claim that for a generic light source $\mathscr{S}$ with parameterization $X(y)$, when we consider $M(x^*,y)=N\big(x^*,X(y)\big)$, the rank of $\left(\dfrac{\partial^2 M}{\partial y^\alpha\partial y^\beta}\right)$ will be k almost everywhere. Indeed, as in the proof of

Theorem 5.3.1, we can write

$$M_{y^\alpha} = \frac{\partial N}{\partial x^i}\big(x^*, X(y)\big)\frac{\partial X^i}{\partial y^\alpha}(y).$$

In vector form, this is

$$M_y = N_x \cdot X_y,$$

where X_y is an $n \times k$ matrix. Now, in an analogous manner to (5.13), we get

$$M_{yy} = X_y^T\, N_{xx}\, X_y + N_x \cdot X_{yy}.$$

Note that since $k \leq n-1$, we want

$$\operatorname{rank}\Big(X_y^T\, N_{xx}\, X_y\Big) = k \tag{5.27}$$

for almost any light source (one needs to take care that a ray from $x \in Q$ to the source $\mathscr{S}$ doesn't intersect $\mathscr{S}$ tangentially).

Call $X(y) \in \mathscr{S}$ *non-tangential relative to* $x \in Q$ if (5.27) is satisfied for this $X(y)$ and x. Since this rank is maximal, there exists an open neighborhood U of y and an open neighborhood V of x such that $X(y)$ is non-tangential relative to x for any $y \in U$ and for any $x \in V$. But, we can say more: there exists a neighborhood $\mathscr{W}$ of the Euclidean metric (in the space of Finsler metrics) and an open neighborhood $\mathscr{U}$ of $X(y)$ (in the space of parameterized light sources) such that for any Finsler metric $F \in \mathscr{W}$ and any light source $\tilde{X}(y) \in \mathscr{U}$, $\tilde{X}(y)$ is non-tangential relative to x under F for any $y \in U$, $x \in V$.

So, given a metric F and a light source parameterized by $X(y)$, we determine B_M. If a point $(x, y) \in B_M$ satisfies (5.27), then (5.27) is also satisfied in a neighborhood

U of (x, y). These neighborhoods form an open cover $\{U_r\}_{r=1}^{\infty}$ of B_M. Thus, we claim that (5.27) is satisfied on any compact subset $K \subset B_M$, as K can be covered by a finite subcover of the open cover.

We now make another pertubative argument: for $\epsilon > 0$ sufficiently small, if X_{yy} is ϵ-small (i.e., if the source is "straight enough"), the $N_x \cdot X_{yy}$ term can be viewed as a perturbation. In this case,

$$\operatorname{rank}\Big(M_{yy}\Big) = \operatorname{rank}\Big(X_y^T\, N_{xx}\, X_y\Big) = k \text{ a.e. on } K \subset B_M.$$

Thus, when M describes light propagation, we have dim $\Sigma(B_M) <$ dim B_M on the compact subset K of B_M.

At last, we are ready for the final case. Suppose $G \subset \Sigma(B_M)$ locally. Take another smooth curve $H \in K \subset B_M$ such that $H(0) \in \Sigma(B_M)$, $H(0) = G(0)$, $H \notin \Sigma(B_M)$ for any other point of H, and such that $\dot{H}(0) = \dot{G}(0)$. (Note that this is only possible because dim $\Sigma(B_M) <$ dim B_M on K.) Since H is a curve of the previous case, and since the two tangent vectors are equal, then $\dot{y}^\alpha = 0$ for G.

In summary: if F is a Finsler metric which is a perturbation of the Euclidean metric, if $G \in K \subset B_M$ is a curve on a compact subset K of B_M such that its "lift" $\eta \in L_M$ has tangent vector $\dot{\eta} \in \ker T_\eta \lambda_M$, and if the curvature of the light source is ϵ-small, then ker $T_\eta \lambda_M = \{0\}$. Thus, M is a Morse family on K. □

Remark 5.4.2. The proof merely demonstrates the *existence* of such an ϵ which guarantees that M is a Morse family. The upper bound of ϵ depends not only on the light source, but also on the Finsler structure itself. Finding the upper bound is a highly non-trivial task. ▲

Our proof demonstrates that as long as a Finsler structure is "close" to being Euclidean, and if the X_{yy}-term is sufficiently small, then the function M which describes light propagation is a Morse family.

What might such an F look like? Define a Minkowski metric by

$$|v|_r := \left(\sum_{i=1}^{n} |v^i|^r \right)^{\frac{1}{r}}.$$

Note that $r = 2$ is the standard Euclidean metric. Note also that there may be dependence on x. Then, a perturbation of the Euclidean metric would include (but not necessarily be limited to) metrics of the form

$$F(x, v) = |v|_2 + a(x)v + b(x)|v|_3 + c(x)|v|_4 + ..., \tag{5.28}$$

for sufficiently small a, b, c, etc.

Remark 5.4.3. Notice that Randers metrics are included in the set of metrics (5.28). ▲

Chapter 6

Applications of Morse Families

Now, we look at some concrete examples of Morse families, and we examine the effect changing the metric of the space has on the system. We also discuss what can happen if we don't require the light source to be an embedded submanifold. Finally, we give an explicit formulation for the caustic of a plane curve in a homogeneous Randers space using Morse families.

6.1 Parabolic Light Source in $\mathbb{R}^2$ with the Euclidean Metric

Let (x^1, x^2) denote coordinates in $Q = \mathbb{R}^2$ endowed with the standard Euclidean metric, and let y be the coordinate on the parabolic light source $\mathscr{S}$ given by

$$X(y) = (y^2, y) \in \mathbb{R}^2.$$

Recall from Section 5.2 that light rays are straight lines in this context. Using the Morse family from Proposition 3.27, we have

$$M(x, y) = \sqrt{(x^1 - y^2)^2 + (x^2 - y)^2}\,.$$

Simple calculations yield

$$\frac{\partial M}{\partial x^1} = \frac{x^1 - y^2}{M} \tag{6.1}$$

$$\frac{\partial M}{\partial x^2} = \frac{x^2 - y}{M} \tag{6.2}$$

$$\frac{\partial M}{\partial y} = \frac{2y^3 - 2x^1 y - x^2 + y}{M}. \tag{6.3}$$

The condition (3.12) for the fiber-critical manifold B_M is

$$\frac{\partial M}{\partial y} = 2y^3 - 2x^1 y - x^2 + y = 0. \tag{6.4}$$

The fiber-critical manifold $B_M \subset B$ is given by

$$B_M = \left\{ (x^1, x^2, y) \in B : 2y^3 - 2x^1 y - x^2 + y = 0 \right\}.$$

The straight lines defined by points $(x^1, x^2, y) \in B_M$ are light rays, and by Corollary 4.6.9, they are orthogonal to the source of light $\mathscr{S}$. The orthogonality is also easy to show explicitly. Let $v = x - X(y) = \langle x^1 - y^2, x^2 - y \rangle$ be a vector corresponding to the direction of a light ray, and let $w = \langle 2y, 1 \rangle \in T_{X(y)}\mathscr{S}$ be a vector tangent to the light source $\mathscr{S}$ at the point $X(y)$. In this context, orthogonality simply means that $v \cdot w = 0$. We get

$$v \cdot w = (x^1 - y^2)(2y) + (x^2 - y)(1) = 2x^1y - 2y^3 + x^2 - y = 0,$$

which is equivalent to the condition 6.4.

Now, we can find the defining matrix of a Morse family:

$$\left.\begin{pmatrix} \dfrac{\partial^2 M}{\partial x^1 \partial y} & \dfrac{\partial^2 M}{\partial x^2 \partial y} & \dfrac{\partial^2 M}{\partial y^2} \end{pmatrix}\right|_{B_M} = \begin{pmatrix} \dfrac{-2y}{M} & \dfrac{1}{M} & \dfrac{6y^2 - 2x^1 + 1}{M} \end{pmatrix}.$$

Clearly, this matrix has maximum rank 1, so M is a Morse family. The submanifold $L_M \subset T^*B$ 3.11 is given by

$$L_M = \left\{ \left(x^1, x^2, y; \frac{x^1 - y^2}{M}, \frac{x^2 - y}{M}, 0 \right) : (x^1, x^2, y) \in B_M \right\}.$$

Now, we project L_M to T^*Q via $\lambda_M : L_M \to T^*Q$. Recall that λ_M maps a point $\left(x, y; \dfrac{\partial M}{\partial x}(x, y), 0 \right) \in L_M$ to $\left(x; \dfrac{\partial M}{\partial x}(x, y) \right) \in T^*Q$. We have

$$\Lambda(B_M) = \lambda_M(L_M) \tag{6.5}$$

$$= \left\{ \left(x^1, x^2; \frac{x^1 - y^2}{M}, \frac{x^2 - y}{M} \right) : (x^1, x^2, y) \in B_M \right\} \subset T^*Q\,. \tag{6.6}$$

The caustic can be determined by solving the following system of equations for x^1 and x^2:

$$\frac{\partial M}{\partial y} = \frac{2y^3 - 2x^1y - x^2 + y}{M} = 0 \tag{6.7}$$

$$\frac{\partial^2 M}{\partial y^2} = \frac{6y^2 - 2x^1 + 1}{M} = 0. \tag{6.8}$$

This yields the singular set

$$\Sigma(B_M) = \left\{ \left(\frac{1}{2} + 3y^2, -4y^3, y \right) : y \in \mathbb{R} \right\} \subset B, \tag{6.9}$$

and the caustic

$$\Gamma(B_M) = \left\{ \left(\frac{1}{2} + 3y^2, -4y^3 \right) : y \in \mathbb{R} \right\} \subset Q. \tag{6.10}$$

What is the physical meaning of B_M in this context? Given some fixed $x \in Q$, we can solve (6.4) for the value(s) of y such that $X(y)$ is the point on the source $\mathscr{S}$ from which a light ray comes to x. However, much more can be said.

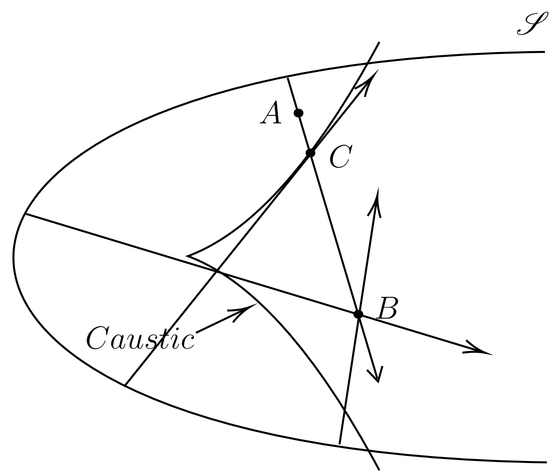

Figure 6.1: Points A, B, and C along light rays eminating from $\mathscr{S}$.

Take points $A, B, C \in Q$ along a particular light ray; see Figure 6.1.

- At A, there is only one light ray which passes through it. Thus, (6.4) has one unique solution. Note that A is a point along the ray before the ray has passed through the caustic, and the ray intersects the caustic transversally.

- At B, which is beyond the caustic, there are three rays which pass through

B. Thus, (6.4) has three solutions. Each ray through B passes through the caustic transversally.

- At C, which is on the caustic, (6.4) has two solutions, one of which is a multiple root. Notice also that the ray which corresponds to the multiple root intersects the caustic *tangentially*, while the other one intersects the caustic transversally.

From a purely mathematical standpoint, this is no accident. One can rewrite (6.4) as a monic polynomial equation in y:

$$y^3 - \left(x^1 - \frac{1}{2}\right)y - \frac{1}{2}x^2 = 0. \tag{6.11}$$

The discrimant of the polynomial in (6.11) is

$$\Delta(x) = 4\left(x^1 - \frac{1}{2}\right)^3 - \frac{27}{4}(x^2)^2. \tag{6.12}$$

One can easily show that (6.12) coincides with $\Gamma(B_M)$. When $\Delta(x) < 0$ (to the "left" of the caustic), (6.11) has one real solution with multiplicity one. When $\Delta(x) > 0$ (to the"right" of the caustic), (6.11) has three real solutions. When $\Delta(x) = 0$ (i.e., on the caustic), (6.11) has two real solutions, one of which is multiplicity two. Furthermore, if one were to solve (6.11) at the cusp point, one would obtain a single real solution of multiplicity three. Thus, one can view the caustic as a sort of bifurcation with regards to the number of solutions for (6.11).

Now, we show how $\Lambda_M(B_M) = \lambda \circ \mathrm{d}M(B_M)$ generates a solution to the eikonal equation (2.14). The associated Hamilton-Jacobi equation is given by

$$H(x,p) = (p_{x,1})^2 + (p_{x,2})^2 - 1 = 0. \tag{6.13}$$

For $\Lambda_M(B_M)$, we have $p_{x,1} = \dfrac{x^1 - y^2}{M}$, $p_{x,2} = \dfrac{x^2 - y}{M}$. Clearly, these p solve (6.13). In fact, if one takes

$$x^1 = y^2 \pm \frac{t}{\sqrt{1+4y^2}} \tag{6.14}$$

$$x^2 = y \mp \frac{2yt}{\sqrt{1+4y^2}}, \tag{6.15}$$

this choice of x^1, x^2 satisfy the condition (6.4) of the fiber-critical manifold B_M. One can check that substituting (6.14) and (6.15) into the generalized solution $\Lambda_M(B_M)$ (6.5), one obtains precisely the solution (2.30) generated by the method of characteristics.

6.2 Parabolic Light Source in $\mathbb{R}^2$ with Randers Metric

6.2.1 Perturbation is Parallel to Axis of Symmetry

Now, let (x^1, x^2) be coordinates in $Q = \mathbb{R}^2$ endowed with the Randers metric F given by (4.17) with $p = 1$, and let y be the coordinate on a parabolic light source $\mathscr{S}$ given by $X(y) = (y^2, y)$. Since this metric doesn't depend on $x \in Q$, it defines a homogeneous media, so light rays are once again straight lines. We now write the function for the optical path as

$$M(x,y) = \sqrt{(x^1-y^2)^2 + (x^2-y)^2} - \epsilon(x^1 - y^2)\,.$$

(**Notation:** Let $M_E := \sqrt{(x^1-y^2)^2 + (x^2-y)^2}$.)

Simple calculations yield

$$\frac{\partial M}{\partial x^1} = \frac{x^1 - y^2}{M_E} - \epsilon \tag{6.16}$$

$$\frac{\partial M}{\partial x^2} = \frac{x^2 - y}{M_E} \tag{6.17}$$

$$\frac{\partial M}{\partial y} = \frac{2y^3 - 2x^1 y - x^2 + y}{M_E} + 2\epsilon y. \tag{6.18}$$

The condition for the fiber-critical manifold B_M is $\dfrac{2y^3 - 2x^1 y - x^2 + y}{M_E} + 2\epsilon y = 0$. The fiber-critical manifold $B_M \subset B$ is given by

$$B_M = \left\{ (x^1, x^2, y) \in B : \frac{2y^3 - 2x^1 y - x^2 + y}{M_E} + 2\epsilon y = 0 \right\}.$$

Before continuing, we must ask a natural question: what does the condition for the fiber-critical manifold B_M mean in this context? If one examines the conditions for F-orthogonality from Remark 4.7.1, one will obtain exactly the condition for B_M. In more detail, let $v = x - X(y) = \langle x^1 - y^2, x^2 - y \rangle$ be a vector corresponding to the direction of a light ray, and let $w = \langle 2y, 1 \rangle \in T_{X(y)}\mathscr{S}$ be a vector tangent to the light source $\mathscr{S}$ at the point $X(y)$. From Remark 4.7.1, the vectors v and w must satisfy $F_{v^i}(v)w^i = 0$. We have that

$$\begin{aligned} F(v) &= |v|_E - \epsilon v^1 \\ F_{v^1}(v) &= \frac{v^1}{|v|_E} - \epsilon \\ F_{v^2}(v) &= \frac{v^2}{|v|_E}. \end{aligned}$$

Using v and w from above gives us

$$\begin{aligned} F_{v^i}(v)w^i &= \left(\frac{x^1 - y^2}{M_E} - \epsilon\right)(2y) + \left(\frac{x^2 - y}{M_E}\right)(1) \\ &= \frac{2x^1y - 2y^3 + x^2 - y}{M_E} - 2\epsilon y = 0, \end{aligned}$$

which is equivalent to the condition on B_M. Thus, in this context, the points $(x^1, x^2, y) \in B_M$ define straight lines — i.e., light rays — which are F-orthogonal to the light source.

By Corollary 5.3.2, we know that M is a Morse family, and the submanifold $L_M \subset T^*B$ is given by

$$L_M = \left\{\left(x^1, x^2, y; \frac{x^1 - y^2}{M_E} - \epsilon, \frac{x^2 - y}{M_E}, 0\right) : (x^1, x^2, y) \in B_M\right\}.$$

The projection $\lambda_M(L_M)$ (which has the same meaning as in the case of the Euclidean metric) is

$$\lambda_M(L_M) = \left\{\left(x^1, x^2; \frac{x^1 - y^2}{M_E} - \epsilon, \frac{x^2 - y}{M_E}\right) : (x^1, x^2, y) \in B_M\right\} \subset T^*Q\,.$$

The caustic can be determined by solving the system

$$\frac{\partial M}{\partial y} = 0 \tag{6.19}$$

$$\frac{\partial^2 M}{\partial y^2} = 0 \tag{6.20}$$

for x^1 and x^2. This system is given by

$$\frac{2y^3 - 2x^1y - x^2 + y}{\sqrt{(x^1 - y^2)^2 + (x^2 - y)^2}} + 2\epsilon y = 0 \tag{6.21}$$

$$\frac{-2x^1 + 6y^2 + 1 - 4\epsilon^2y^2}{\sqrt{(x^1 - y^2)^2 + (x^2 - y)^2}} + 2\epsilon = 0. \tag{6.22}$$

Unlike in the case of the standard Euclidean metric, here we actually get two solutions:

$$x^1 = \frac{-2\left(\epsilon^2 + 3\right)y^2 \pm \sqrt{-\left(4\epsilon^3y^2 + 4\epsilon y^2 + \epsilon\right)^2\left(4\left(\epsilon^2 - 1\right)y^2 - 1\right)} - 1}{2\left(\epsilon^2 - 1\right)} \tag{6.23}$$

$$x^2 = -4(1 + \epsilon^2)y^3. \tag{6.24}$$

However, as is often the case with systems of equations like (6.21) and (6.22), we have to check for false solutions. Indeed, only the x^1 with the negative root is an actual solution (one can verify this by choosing explicit values for ϵ and y, plugging these values into x^1 and x^2, then plugging all of these into (6.21) and (6.22)). This yields the singular set

$$\Sigma(B_M) = \left\{\left(\frac{-2\left(\epsilon^2 + 3\right)y^2 - \sqrt{E} - 1}{2\left(\epsilon^2 - 1\right)}, -4(1 + \epsilon^2)y^3, y\right) : y \in \mathbb{R}\right\} \subset B, \tag{6.25}$$

and the caustic

$$\Gamma(B_M) = \left\{\left(\frac{-2\left(\epsilon^2 + 3\right)y^2 - \sqrt{E} - 1}{2\left(\epsilon^2 - 1\right)}, -4(1 + \epsilon^2)y^3\right) : y \in \mathbb{R}\right\} \subset Q, \tag{6.26}$$

where $E = -\left(4\epsilon^3y^2 + 4\epsilon y^2 + \epsilon\right)^2\left(4\left(\epsilon^2 - 1\right)y^2 - 1\right)$.

Remark 6.2.1. Note that throughout this example, if we allow $\epsilon \to 0$, we obtain the results of the first example. ▲

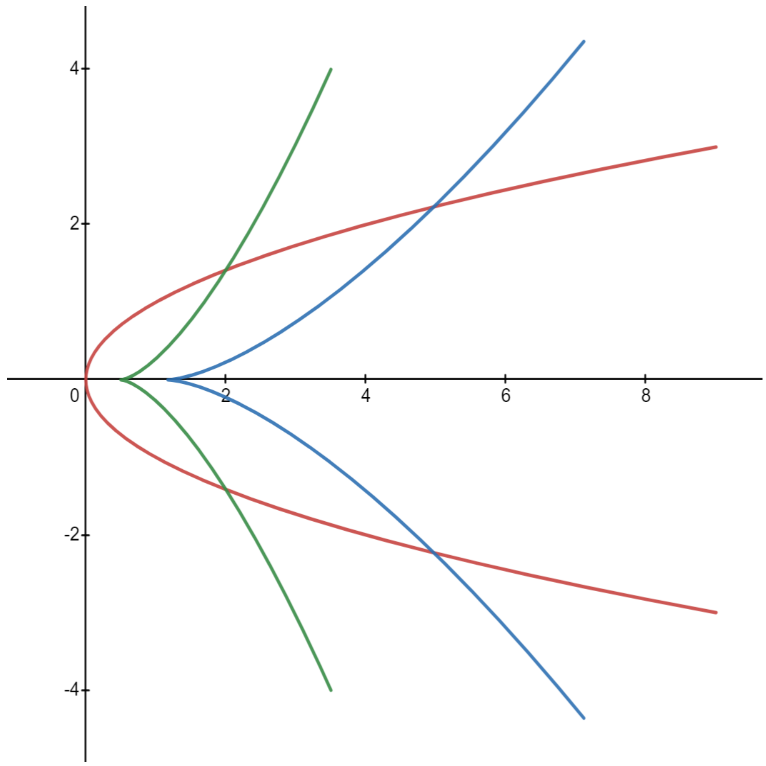

Figure 6.2: Parabolic source with the caustic from a Euclidean metric (left cusp) and a Randers metric (right cusp) when $\epsilon = 0.3$

6.2.2 Perturbation is Perpendicular to Axis of Symmetry

In the previous example, there was a subtle element of symmetry. Namely, the horizontal coordinate axis in Q was an axis of symmetry of the parabolic light source, while the horizontal axis in TQ coincided with the major axis of the ellipse which determined the indicatrix of F. What happens if we alter either F or the source in some way?

Let us alter F so that the major axis of the indicatrix is now on the vertical axis of TQ, while maintaining the same light source. The optical path is now given by

$$M(x,y) = \sqrt{(x^1 - y^2)^2 + (x^2 - y)^2} - \epsilon(x^2 - y)\,.$$

Focusing strictly on what happens with the caustic, we solve the system of equations

$$\frac{\partial M}{\partial y} = \frac{2y^3 - 2x^1 y - x^2 + y}{M_E} + \epsilon = 0 \tag{6.27}$$

$$\frac{\partial^2 M}{\partial y^2} = \frac{6y^2 - 2x^1 + 1 - \epsilon^2}{M_E} = 0 \tag{6.28}$$

for x^1 and x^2. We get

$$x^1 = \frac{1}{2} + 3y^2 - \frac{1}{2}\epsilon^2 \tag{6.29}$$

$$x^2 = \pm\frac{-8y^3 + \sqrt{-\epsilon^2(-1+\epsilon^2-4y^2)^3}}{2(-1+\epsilon^2}. \tag{6.30}$$

One can easily verify that, like in the previous example, only the x^2 where the $\pm$ is negative is an actual solution. So, we obtain the caustic

$$\Gamma(B_M) = \left\{ \left(\frac{1}{2} + 3y^2 - \frac{1}{2}\epsilon^2, -\frac{-8y^3 + \sqrt{-\epsilon^2(-1+\epsilon^2-4y^2)^3}}{2(-1+\epsilon^2)} \right) : y \in \mathbb{R} \right\} \subset Q. \tag{6.31}$$

6.3 Caustic of a Plane Curve in a 2-Dimensional Homogeneous Randers Space

Let $(\mathbb{R}^2, F)$ be such that the Finsler metric F has no dependence on $x \in \mathbb{R}^2$. Suppose we have a family of light rays in $(\mathbb{R}^2, F)$ — i.e., a family of straight lines — and a

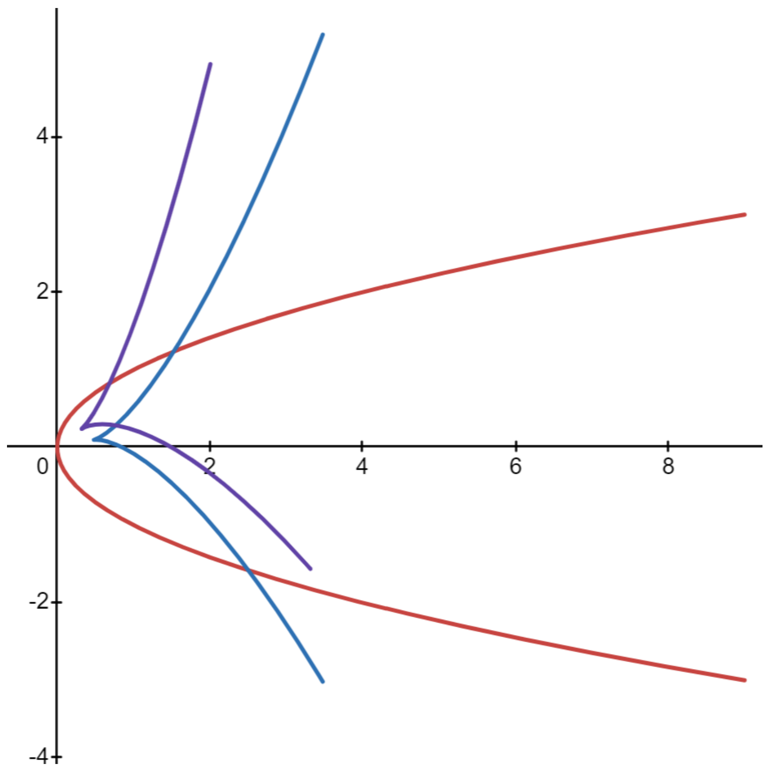

Figure 6.3: Parabolic source with the caustic from a Randers metric whose indicatrix is aligned with the vertical axis, when $\epsilon = 0.2$ and $\epsilon = 0.6$, moving counterclockwise.

curve $\mathscr{S}$ which intersects each light ray transversally. It is possible to write equations for the caustic of this family of light rays in terms of the curve $\mathscr{S}$. Let $X : \mathbb{R}_{\text{loc}} \to \mathbb{R}^2$ be a parameterization for $\mathscr{S}$. A light ray can be given by the expression

$$X(y) + t\mathbf{v}(y),$$

for some vector $\mathbf{v}(y)$ and for $t \in \mathbb{R}$. Recall from Section 3.7 that the caustic of a family of rays is their envelope. In other words, it is a curve which is tangent to each light ray. Since the light ray will tangentially intersect the caustic at one point, the caustic can be given by

$$\Gamma(y) = X(y) + f(y)\mathbf{v}(y) \tag{6.32}$$

for some unknown function $f : \mathbb{R}_{\text{loc}} \to \mathbb{R}$.

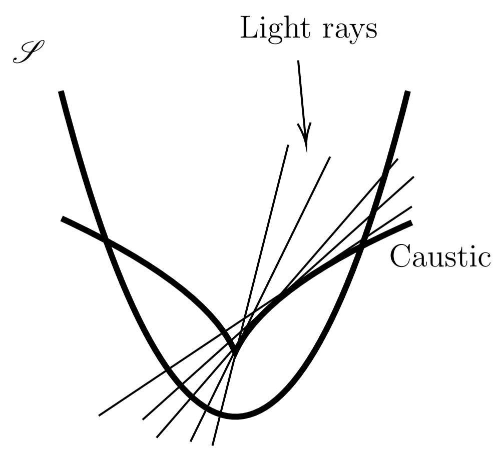

Figure 6.4: A curve $\mathscr{S}$, and the envelope of a set of light rays

A vector tangent to the caustic is parallel to $\mathbf{v}(y)$ if and only if the following determinant equation is satisfied:

$$\det \begin{pmatrix} (X(y) + f(y)\mathbf{v}(y))' \\ \mathbf{v}(y) \end{pmatrix} = 0 \tag{6.33}$$

Solving this equation for $f(y)$ yields

$$f(y) = \frac{\dot{X}^2 v^1 - \dot{X}^1 v^2}{\dot{v}^1 v^2 - \dot{v}^2 v^1}, \tag{6.34}$$

where $\dot{X}^i = \dfrac{\partial X^i}{\partial y}(y)$. Thus, we obtain a description of the caustic:

$$\Gamma(y) = X(y) + \frac{\dot{X}^2 v^1 - \dot{X}^1 v^2}{\dot{v}^1 v^2 - \dot{v}^2 v^1}\mathbf{v}(y). \tag{6.35}$$

Now, suppose we wish to find explicit expressions for v^1 and v^2 in terms of $X(y)$. Then, the caustic will be given by (6.35), where v^1 and v^2 satisfy the conditions of light rays given in Remark 4.7.1, namely

(1) $F(v) = 1$ (or equivalently, $F^2(v) = 1$), i.e., v is in the indicatrix of F

(2) $F_{v^i}(v)\dot{X}^i(y) = 0$, i.e., v is F-orthogonal to $\mathscr{S}$.

6.3.1 Example: Caustic in Homogeneous Randers Space

Let $Q = (\mathbb{R}^2, F)$, where F is the Randers metric

$$F(v) = |v|_E - \epsilon v^1, \tag{6.36}$$

where $|v|_E = \sqrt{(v^1)^2 + (v^2)^2}$. We wish to find explicit expressions for v^1 and v^2. We can find an explicit expression for the caustic by either using only the tools of

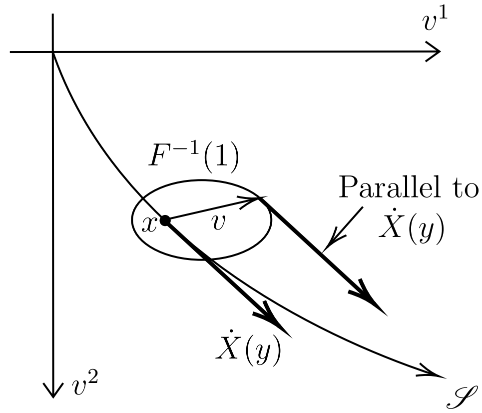

Figure 6.5: Illustration of the conditions light rays must satisfy

Finsler geometry, or by using Morse families.

Using Finsler Geometry

For $F(v) = |v|_E - \epsilon v^1$, we can write the system from Remark 4.7.1 as follows:

$$(1+\epsilon^2)(v^1)^2 + (v^2)^2 - 2\epsilon v^1 |v|_E = 1 \tag{6.37}$$

$$\left(\frac{v^1}{|v|_E} - \epsilon\right)\dot{X}^1 + \left(\frac{v^2}{|v|_E}\right)\dot{X}^2 = 0, \tag{6.38}$$

or equivalently,

$$(1+\epsilon^2)(v^1)^2 + (v^2)^2 - 2\epsilon v^1 |v|_E = 1 \tag{6.39}$$

$$\frac{v^1\dot{X}^1 + v^2\dot{X}^2}{|v|_E} = \epsilon \dot{X}^1. \tag{6.40}$$

Solve (6.39) for $|v|_E$ to get

$$|v|_E = \frac{(1+\epsilon^2)(v^1)^2 + (v^2)^2 - 1}{2\epsilon v^1}.$$

Plug this into (6.40) to solve for v^1 in terms of v^2:

$$v^1 = \frac{-v^2\dot{X}^2 + \sqrt{(v^2)^2(\dot{X}^2)^2 - (1-(v^2)^2)(1-\epsilon^2)(\dot{X}^1)^2}}{(1-\epsilon^2)\dot{X}^1}. \tag{6.41}$$

For ease of notation, let

$$R = (1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2. \tag{6.42}$$

Plugging (6.41) into (6.39) allows us to explicitly solve for v^1 and v^2, and thus also $\dot{v}^1$ and $\dot{v}^2$:

$$v^2 = \pm\frac{\dot{X}^1}{\sqrt{R}} \tag{6.43}$$

$$v^1 = \frac{1}{\epsilon^2 - 1}\left(\pm\,\epsilon + \frac{\dot{X}^2}{\sqrt{R}}\right) \text{ (for } v^2 \text{ +)} \tag{6.44}$$

$$v^1 = \frac{1}{\epsilon^2 - 1}\left(\pm\,\epsilon - \frac{\dot{X}^2}{\sqrt{R}}\right) \text{ (for } v^2 \text{ -)} \tag{6.45}$$

$$\dot{v}^1 = \pm\frac{\dot{X}^1[\dot{X}^2\ddot{X}^1 - \dot{X}^1\ddot{X}^2]}{R^{3/2}} \tag{6.46}$$

$$\dot{v}^2 = \pm\frac{\dot{X}^2[\dot{X}^2\ddot{X}^1 - \dot{X}^1\ddot{X}^2]}{R^{3/2}}. \tag{6.47}$$

To obtain the function $f(y)$ from (6.32), we must now find $\dfrac{\dot{X}^2v^1 - \dot{X}^1v^2}{\dot{v}^1v^2 - \dot{v}^2v^1}$. First, we find its numerator. We will use the + version of v^2 in the calculations.

$$
\begin{aligned}
\dot{X}^2 v^1 - \dot{X}^1 v^2 &= \frac{1}{\epsilon^2 - 1}\left(\pm\, \epsilon \dot{X}^2 + \frac{(\dot{X}^2)^2}{\sqrt{R}} \right) - \dot{X}^1 \left(\frac{\dot{X}^1}{\sqrt{R}} \right) \\
&= \pm \frac{\epsilon \dot{X}^2}{\epsilon^2 - 1} + \frac{\left(\dot{X}^2\right)^2}{(\epsilon^2 - 1)\sqrt{R}} - \frac{\left(\dot{X}^1\right)^2}{\sqrt{R}} \\
&= \frac{R \pm \sqrt{R}\epsilon \dot{X}^2}{(\epsilon^2 - 1)\sqrt{R}}.
\end{aligned}
$$

Now, we find the denominator of $\dfrac{\dot{X}^2 v^1 - \dot{X}^1 v^2}{\dot{v}^1 v^2 - \dot{v}^2 v^1}$.

$$
\begin{aligned}
\dot{v}^1 v^2 - \dot{v}^2 v^1 &= \frac{\dot{X}^1\big(\dot{X}^2 \ddot{X}^1 - \dot{X}^1 \ddot{X}^2\big)}{R^{3/2}} \cdot \frac{\dot{X}^1}{\sqrt{R}} - \\
&\quad \frac{\dot{X}^2\big(\dot{X}^2 \ddot{X}^1 - \dot{X}^1 \ddot{X}^2\big)}{R^{3/2}} \left(\pm \frac{\epsilon}{\epsilon^2 - 1} + \frac{\dot{X}^2}{\sqrt{R}(\epsilon^2 - 1)} \right) \\
&= \frac{\dot{X}^2 \ddot{X}^1 - \dot{X}^1 \ddot{X}^2}{R^{3/2}} \left(\frac{\left(\dot{X}^1\right)^2}{\sqrt{R}} - \frac{(\pm\epsilon)\dot{X}^2}{\epsilon^2 - 1} - \frac{\left(\dot{X}^2\right)^2}{\sqrt{R}(\epsilon^2 - 1)} \right) \\
&= -\frac{\big(\dot{X}^2 \ddot{X}^1 - \dot{X}^1 \ddot{X}^2\big)\big[(1 - \epsilon^2)\big(\dot{X}^1\big)^2 + \big(\dot{X}^2\big)^2 \pm \sqrt{R}\epsilon \dot{X}^2\big]}{R^2(\epsilon^2 - 1)} \\
&= -\frac{\big(\dot{X}^2 \ddot{X}^1 - \dot{X}^1 \ddot{X}^2\big)\big[R \pm \sqrt{R}\epsilon \dot{X}^2\big]}{R^2(\epsilon^2 - 1)}.
\end{aligned}
$$

Therefore,

$$\begin{aligned}
f(y) &= \frac{\dot{X}^1 v^2 - \dot{X}^2 v^1}{\dot{v}^1 v^2 - \dot{v}^2 v^1} \\
&= \frac{R \pm \sqrt{R}\epsilon \dot{X}^2}{(\epsilon^2 - 1)\sqrt{R}} \cdot - \frac{R^2(\epsilon^2 - 1)}{\left(\dot{X}^2 \ddot{X}^1 - \dot{X}^1 \ddot{X}^2\right)[R \pm \sqrt{R}\epsilon \dot{X}^2]} \\
&= \frac{R^{3/2}}{\dot{X}^1 \ddot{X}^2 - \dot{X}^2 \ddot{X}^1}.
\end{aligned}$$

Finally, we can explicitly write the parametric equations for the caustic:

$$\begin{aligned}
\Gamma^1(y) &= X^1(y) + f(y) v^1(y) \\
&= X^1(y) + \frac{R^{3/2}}{\dot{X}^1 \ddot{X}^2 - \dot{X}^2 \ddot{X}^1} \left(\frac{\pm \epsilon}{\epsilon^2 - 1} + \frac{\dot{X}^2}{\sqrt{R}(\epsilon^2 - 1)} \right) \\
&= X^1(y) + \frac{R}{\dot{X}^1 \ddot{X}^2 - \dot{X}^2 \ddot{X}^1} \left[\frac{-\dot{X}^2 \mp \epsilon \sqrt{R}}{1 - \epsilon^2} \right];
\end{aligned}$$

$$\begin{aligned}
\Gamma^2(y) &= X^2(y) + f(y) v^2(y) \\
&= X^2(y) + \frac{R^{3/2}}{\dot{X}^1 \ddot{X}^2 - \dot{X}^2 \ddot{X}^1} \cdot \frac{\dot{X}^1}{\sqrt{R}} \\
&= X^2(y) + \frac{R}{\dot{X}^1 \ddot{X}^2 - \dot{X}^2 \ddot{X}^1} \dot{X}^1.
\end{aligned}$$

Replacing R with (6.42), we get

$$\Gamma^1(y) = X^1(y) + \frac{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}{\dot{X}^1\ddot{X}^2 - \dot{X}^2\ddot{X}^1}\left[\frac{-\dot{X}^2 \mp \epsilon\sqrt{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}}{1-\epsilon^2}\right];$$

$$\Gamma^2(y) = X^2(y) + \frac{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}{\dot{X}^1\ddot{X}^2 - \dot{X}^2\ddot{X}^1}\dot{X}^1. \tag{6.48}$$

Using Morse Families

Now, we obtain the same result using a Morse family. Let

$$M(x,y) = \sqrt{(x^1 - X^1(y))^2 + (x^2 - X^2(y))^2} - \epsilon(x^1 - X^1(y))$$

be the Morse family corresponding to the Randers metric (6.36) in homogeneous media (i.e., F has no dependence on $x \in \mathbb{R}^2$). Recall from (3.38) and (3.39) that the conditions for the caustic are

$$\frac{\partial M}{\partial y^\alpha}(x,y) = 0\,,\ \text{for all } \alpha = 1, ..., k; \tag{6.49}$$

$$\det\left(\frac{\partial^2 M}{\partial y^\alpha \partial y^\beta}\right) = 0. \tag{6.50}$$

For the specific case of $\mathbb{R}^2$, this gives us a system of two equations:

$$\frac{\partial M}{\partial y}(x,y) = 0 \tag{6.51}$$

$$\frac{\partial^2 M}{\partial y^2}(x,y) = 0. \tag{6.52}$$

Equation (6.51) becomes

$$\frac{\partial M}{\partial y} = -\frac{(x^1 - X^1(y))\dot{X}^1 + (x^2 - X^2(y))\dot{X}^2}{|x - X(y)|_E} + \epsilon \dot{X}^1. \tag{6.53}$$

Thus, we can write the condition for the fiber-critical manifold B_M as

$$\frac{(x^1 - X^1(y))\dot{X}^1 + (x^2 - X^2(y))\dot{X}^2}{|x - X(y)|_E} = \epsilon \dot{X}^1. \tag{6.54}$$

On B_M, we have for Equation (6.52)

$$\begin{aligned}\frac{\partial^2 M}{\partial y^2} &= \frac{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2 - (x^1 - X^1(y))\ddot{X}^1 + (x^2 - X^2(y))\ddot{X}^2}{|x - X(y)|_E} + \epsilon \ddot{X}^1 \\ &= 0.\end{aligned} \tag{6.55}$$

We wish to solve for x^1 and x^2 in terms of $X(y)$, $\dot{X}$, and $\ddot{X}$. First, we solve (6.54) for $|x - X(y)|_E$:

$$|x - X(y)|_E = \frac{(x^1 - X^1(y))\dot{X}^1 + (x^2 - X^2(y))\dot{X}^2}{\epsilon \dot{X}^1}.$$

Plug this into (6.55) and solve for x^2:

$$x^2 = X^2(y) + \frac{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}{\dot{X}^1\ddot{X}^2 - \ddot{X}^1\dot{X}^2}\dot{X}^1. \tag{6.56}$$

Note that the expression for x^2 is precisely $\Gamma^2(y)$ from (6.48). For ease of notation, let

$$A = \frac{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}{\dot{X}^1\ddot{X}^2 - \ddot{X}^1\dot{X}^2}. \tag{6.57}$$

Plug (6.56) into (6.54):

$$\frac{(x^1 - X^1(y)) + A\dot{X}^2}{\sqrt{(x^1 - X^1(y))^2 + A^2(\dot{X}^1)^2}} = \epsilon. \tag{6.58}$$

Square both sides and get rid of the fraction to get

$$\xi^2 + 2A\dot{X}^2\xi + A^2(\dot{X}^2)^2 = \epsilon^2\xi^2 + \epsilon^2 A^2(\dot{X}^1)^2, \tag{6.59}$$

where $\xi = x^1 - X^1(y)$. Rearrange to form a quadratic equation in ξ:

$$(1-\epsilon^2)\xi^2 + 2A\dot{X}^2\xi + A^2((\dot{X}^2)^2 - \epsilon^2(\dot{X}^1)^2) = 0. \tag{6.60}$$

By the quadratic formula,

$$\begin{aligned}
\xi &= \frac{-2A\dot{X}^2 \pm \sqrt{4A^2(\dot{X}^2)^2 - 4(1-\epsilon^2)A^2((\dot{X}^2)^2 - \epsilon^2(\dot{X}^1)^2)}}{2(1-\epsilon^2)} \\
&= \frac{-2A\dot{X}^2 \pm \sqrt{4A^2[(\dot{X}^2)^2 - (1-\epsilon^2)((\dot{X}^2)^2 - \epsilon^2(\dot{X}^1)^2)]}}{2(1-\epsilon^2)} \\
&= \frac{-2A\dot{X}^2 \pm 2A\sqrt{\epsilon^2(\dot{X}^1)^2 + \epsilon^2(\dot{X}^2)^2 - \epsilon^4(\dot{X}^1)^2}}{2(1-\epsilon^2)} \\
&= A\left(\frac{-\dot{X}^2 \pm \epsilon\sqrt{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}}{1-\epsilon^2}\right).
\end{aligned}$$

But since $\xi = x^1 - X^1(y)$, we now have

$$x^1 = X^1(y) + A\left(\frac{-\dot{X}^2 \pm \epsilon\sqrt{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}}{1-\epsilon^2}\right). \tag{6.61}$$

Replace A with (6.57). Comparing (6.61) to (6.48) from Finsler geometry, note that the expression for x^1 is the same as $\Gamma^1(y)$.

6.4 Caustic Examples

For the next few examples, we need the following parameterizations for light sources:

(1) $\mathscr{S}_E$: $X^1(y) = \cos\, y, X^2(y) = \sin\, y$ (the unit circle in $\mathbb{R}^2$);

(2) $\mathscr{S}_R$: $X^1(y) = \dfrac{\cos\, y}{1-\epsilon\cos\, y}, X^2(y) = \dfrac{\sin\, y}{1-\epsilon\cos\, y}$ (an ellipse with one focus centered at the origin).

Example 6.4.1. Let $Q = (\mathbb{R}^2, \delta_{ij})$, so the indicatrix is the unit circle in $T\mathbb{R}^2$. Consider the light source $\mathscr{S}_E$, so the light source and indicatrix are represented by the same curve in $\mathbb{R}^2$ and $T\mathbb{R}^2$, respectively. The caustic $\Gamma(y)$ of $\mathscr{S}$ is given by (3.40):

$$\Gamma^1(y) = \cos y - \frac{(-\sin y)^2 + \cos^2 y}{(-\sin y)(-\sin y) - (\cos y)(-\cos y)}(\cos y) = 0;$$

$$\Gamma^2(y) = \sin y + \frac{(-\sin y)^2 + \cos^2 y}{(-\sin y)(-\sin y) - (\cos y)(-\cos y)}(-\sin y) = 0,$$

so the caustic is the origin in $\mathbb{R}^2$. ▲

Example 6.4.2. Now, let the light source be $\mathscr{S}_E$, but consider the case where the underlying metric is the Randers metric (6.36). Here, the caustic is given by (6.48):

$$\Gamma^1(y) = \cos y + [1-\epsilon^2\sin^2 y]\left(\frac{-\cos y \pm \epsilon\sqrt{1-\epsilon^2\sin^2 y}}{1-\epsilon^2}\right);$$

$$\Gamma^2(y) = \epsilon^2\sin^3 y.$$

Clearly, this isn't the origin! However, note that as $\epsilon \to 0$, the caustic tends to the origin. ▲

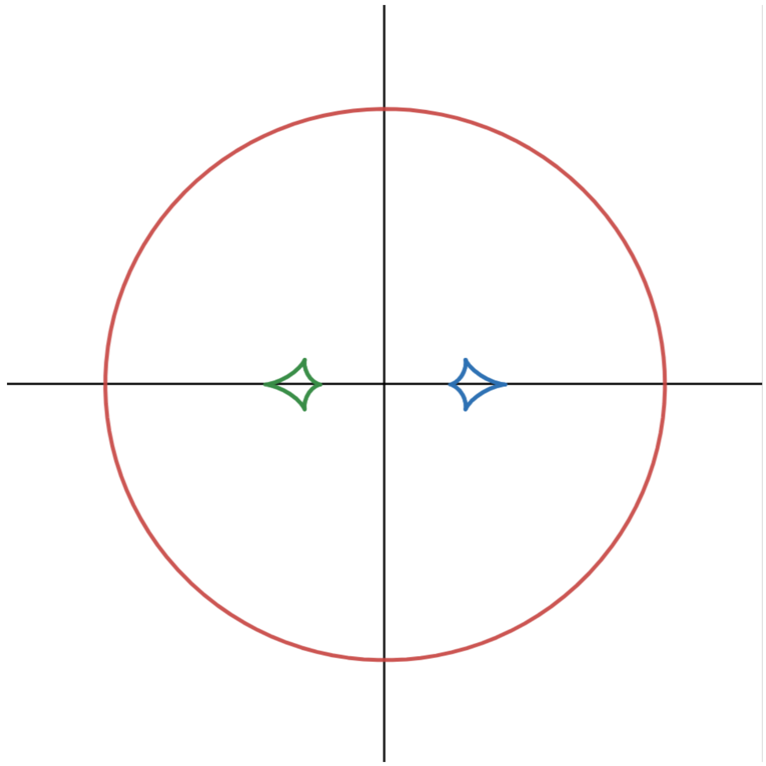

Figure 6.6: Caustic when the light source is the unit circle and the metric is the Randers metric (6.36) ($\epsilon = 0.3$)

Example 6.4.3. Now, the light source be given by $\mathscr{S}_R$, and consider the Randers metric (6.36) once again. We have:

$$\begin{aligned}
X^1(y) &= \frac{\cos y}{1-\epsilon\cos y} \\
\dot{X}^1(y) &= \frac{-\sin y}{(1-\epsilon\cos y)^2} \\
\ddot{X}^1(y) &= \frac{\epsilon\sin^2 y - \cos y + \epsilon}{(1-\epsilon\cos y)^2} \\
X^2(y) &= \frac{\sin y}{1-\epsilon\cos y} \\
\dot{X}^2(y) &= \frac{\cos y - \epsilon}{(1-\epsilon\cos y)^2} \\
\ddot{X}^(y) &= \frac{-(\epsilon\cos y - 2\epsilon^2 + 1)\sin y}{(1-\epsilon\cos y)^3}.
\end{aligned}$$

Some useful expressions are:

$$\begin{aligned}
(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2 &= (1-\epsilon\cos y)^{-2}; \\
\dot{X}^1\ddot{X}^2 - \ddot{X}^1\dot{X}^2 &= (1-\epsilon\cos y)^{-3}.
\end{aligned}$$

Therefore,

$$\frac{(1-\epsilon^2)(\dot{X}^1)^2 + (\dot{X}^2)^2}{\dot{X}^1\ddot{X}^2 - \ddot{X}^1\dot{X}^2} = 1 - \epsilon\cos y.$$

Using (6.48), the equations for the caustic are

$$\begin{aligned}
\Gamma^1(y) &= \frac{\cos y}{1-\epsilon\cos y} + \frac{1-\epsilon\cos y}{1-\epsilon^2}\left(\frac{\epsilon-\cos y}{(1-\epsilon\cos y)^2} \pm \epsilon\frac{1}{1-\epsilon\cos y}\right) \\
&= \frac{\cos y}{1-\epsilon\cos y} + \frac{\epsilon - \cos y \pm \epsilon(1-\epsilon\cos y)}{(1-\epsilon^2)(1-\epsilon\cos y)}; \\
\Gamma^2(y) &= \frac{\sin y}{1-\epsilon\cos y} + (1-\epsilon\cos y)\frac{-\sin y}{(1-\epsilon\cos y)^2} \\
&= 0.
\end{aligned}$$

- If the $\pm$ is $-$, we get

$$\Gamma^1_-(y) = 0.$$

- If the $\pm$ is $+$, we get

$$\Gamma^1_+(y) = \frac{2\epsilon}{1-\epsilon^2}.$$

Recalling that $\epsilon = \dfrac{c}{a}$, and $c^2 = a^2 - b^2$, we can write

$$\Gamma^1_+ = 2c.$$

Therefore, the caustic point is $(2c, 0)$, the other focus of the ellipse. ▲

Bibliography

[1] Peter L. Antonelli, Andrej Bóna, and Michael A. Slawiński. Seismic rays as Finsler geodesics. *Nonlinear Anal. Real World Appl.*, 4(5):711–722, 2003.

[2] V. I. Arnol'd. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1989.

[3] D. Bao, S.-S. Chern, and Z. Shen. *An Introduction to Riemann-Finsler Geometry*, volume 200 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2000.

[4] S. Benenti. The category of symplectic reductions. In *Proceedings of the international meeting on geometry and physics (Florence, 1982)*, pages 11–41. Pitagora, Bologna, 1983.

[5] S. Benenti and W. M. Tulczyjew. The geometrical meaning and globalization of the Hamilton-Jacobi method. In *Differential geometrical methods in mathematical physics (Proc. Conf., Aix-en-Provence/Salamanca, 1979)*, volume 836 of *Lecture Notes in Math.*, pages 9–21. Springer, Berlin, 1980.

[6] Sergio Benenti. *Hamiltonian structures and generating families.* Universitext. Springer, New York, 2011.

[7] Andrej Bóna and Michael A. Slawinski. *Wavefronts and Rays as Characteristics and Asymptotics.* World Scientific, Hackensack, NJ, second edition, 2015.

[8] Ioan Bucataru and Radu Miron. *Finsler-Lagrange Geometry.* Editura Academiei Române, Bucharest, 2007.

[9] Ioan Bucataru and Michael A. Slawinski. Generalized orthogonality between rays and wavefronts in anisotropic inhomogeneous media. *Nonlinear Anal. Real World Appl.*, 6(1):111–121, 2005.

[10] Franco Cardin. *Elementary symplectic topology and mechanics*, volume 16 of *Lecture Notes of the Unione Matematica Italiana.* Springer, Cham, 2015.

[11] Marco Castrillón López and Tudor S. Ratiu. Morse families and Lagrangian submanifolds. In *Geometry, Algebra and Applications: From Mechanics to Cryptography*, pages 65–78. Springer, [Cham], 2016.

[12] Shiing-shen Chern. Riemannian geometry as a special case of Finsler geometry. In *Finsler geometry (Seattle, WA, 1995)*, volume 196 of *Contemp. Math.*, pages 51–58. Amer. Math. Soc., Providence, RI, 1996.

[13] Hengameh R. Dehkordi. Applications of Randers geodesics for wildfire spread modelling. *Appl. Math. Model.*, 106:45–59, 2022.

[14] Valeri P. Frolov. Generalized fermat's principle and action for light rays in a curved spacetime. *Phys. Rev. D*, 88(6):064039, 7, 2013.

[15] Valeri P. Frolov. Maxwell equations in a curved spacetime: spin optics approximation. *Phys. Rev. D*, 102(8):084013, 14, 2020.

[16] Manuel Hohmann, Christian Pfeifer, and Nicoleta Voicu. Mathematical foundations for field theories on Finsler spacetimes. *J. Math. Phys.*, 63(3):Paper No. 032503, 33, 2022.

[17] Lars Hörmander. Fourier integral operators. I. *Acta Math.*, 127(1-2):79–183, 1971.

[18] John M. Lee. *Introduction to smooth manifolds*, volume 218 of *Graduate Texts in Mathematics*. Springer, New York, second edition, 2013.

[19] Paulette Libermann and Charles-Michel Marle. *Symplectic geometry and analytical mechanics*, volume 35 of *Mathematics and its Applications*. D. Reidel Publishing Co., Dordrecht, 1987. Translated from the French by Bertram Eugene Schwarzbach.

[20] Steen Markvorsen. A Finsler geodesic spray paradigm for wildfire spread modelling. *Nonlinear Anal. Real World Appl.*, 28:208–228, 2016.

[21] Christian Pfeifer. Finsler spacetime geometry in physics. *Int. J. Geom. Methods Mod. Phys.*, 16(suppl. 2):1941004, 18, 2019.

[22] Hartmann Römer. *Theoretical Optics: An Introduction*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005.

[23] Gerd Rudolph and Matthias Schmidt. *Differential Geometry and Mathematical Physics. Part I: Manifolds, Lie Groups and Hamiltonian Systems*. Springer, Dordrecht, 2013.

[24] Hanno Rund. *The Differential Geometry of Finsler Spaces*. Die Grundlehren der mathematischen Wissenschaften, Band 101. Springer-Verlag, Berlin-Göttingen-Heidelberg, 1959.

[25] Hanno Rund. *The Hamilton-Jacobi Theory in the Calculus of Variations: Its Role in Mathematics and Physics*. D. Van Nostrand Co., London-Toronto, Ont.-New York, 1966.

[26] Michael A. Slawinski. *Waves and Rays in Elastic Continua*. World Scientific, Hackensack, NJ, 2015.

[27] Alan Weinstein. *Lectures on symplectic manifolds*. Regional Conference Series in Mathematics, No. 29. American Mathematical Society, Providence, RI, 1977. Expository lectures from the CBMS Regional Conference held at the University of North Carolina, March 8–12, 1976.

[28] Takahiro Yajima and Hiroyuki Nagahama. Finsler geometry of seismic ray path in anisotropic media. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 465(2106):1763–1777, 2009.

[29] Takahiro Yajima and Hiroyuki Nagahama. Finsler geometry for nonlinear path of fluids flow through inhomogeneous media. *Nonlinear Anal. Real World Appl.*, 25:1–8, 2015.

[30] Takahiro Yajima, Kazuhito Yamasaki, and Hiroyuki Nagahama. Finsler metric and elastic constants for weak anisotropic media. *Nonlinear Anal. Real World Appl.*, 12(6):3177–3184, 2011.