



Snell's law revisited and generalized via Finsler geometry

Markvorsen, Steen; Pendás-Recondo, Enrique

Published in:
International Journal of Geometric Methods in Modern Physics

Link to article, DOI:
[10.1142/S0219887823501384](https://doi.org/10.1142/S0219887823501384)

Publication date:
2023

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Markvorsen, S., & Pendás-Recondo, E. (2023). Snell's law revisited and generalized via Finsler geometry. *International Journal of Geometric Methods in Modern Physics*, 20(8), Article 2350138. <https://doi.org/10.1142/S0219887823501384>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

SNELL'S LAW REVISITED AND GENERALIZED VIA FINSLER GEOMETRY

STEEN MARKVORSEN AND ENRIQUE PENDÁS-RECONDO

ABSTRACT. We study the variational problem of finding the fastest path between two points that belong to different anisotropic media, each with a prescribed speed profile and a common interface. The optimal curves are Finsler geodesics that are refracted – broken – as they pass through the interface, due to the discontinuity of their velocities. This “breaking” must satisfy a specific condition in terms of the Finsler metrics defined by the speed profiles, thus establishing the generalized Snell’s law. In the same way, optimal paths bouncing off the interface – without crossing into the second domain – provide the generalized law of reflection. The classical Snell’s and reflection laws are recovered in this setting when the velocities are isotropic. If one considers a wave that propagates in all directions from a given ignition point, the trajectories that globally minimize the traveltime generate the wavefront at each instant of time. We study in detail the global properties of such wavefronts in the Euclidean plane with anisotropic speed profiles. Like the individual rays, they break when they encounter the discontinuity interface. But they are also broken due to the formation of cut loci – stemming from the self-intersection of the wavefronts – which typically appear when they approach a high-speed profile domain from a low-speed profile.

Keywords: Snell’s law, Fermat’s principle, Huygens’ principle, Zermelo navigation, Finsler metrics, anisotropic discontinuous media, rays with least traveltime, wave propagation.

CONTENTS

1. Introduction	2
Outline of the paper	3
2. Initial concepts	3
2.1. Traveltime and Finsler metrics	3
2.2. Snell's law; the general setting	5
3. Time-minimizing trajectories	6
3.1. Variational definitions and conventions	6
3.2. Critical points of the traveltime functional	8
3.3. Generalized Snell's law	9
3.4. The two-dimensional case	10
4. Reflected trajectories	11
4.1. The law of reflection	11
4.2. The total reflection phenomenon	13
5. Global considerations in \mathbb{R}^2	13
5.1. Existence and uniqueness of solution	14
5.2. New reflected trajectories	15
5.3. Globally time-minimizing trajectories	17
5.4. Anisotropic example	18
5.5. Construction of the wavefront	20
5.6. The cut locus	22
Acknowledgments	24
References	25

1. INTRODUCTION

Snell's law is – in physical terms – concerned with the refraction of light which in space passes from a point q_1 to another point q_2 through an interface between two (usually isotropic) media Q_1 and Q_2 with different refraction indices n_1 and n_2 , say. In its simplest form the law expresses the break of the light path at the interface in terms of the incoming angle, the outgoing angle, and the ratio between the two refraction indices. The light is supposed to have a constant speed c/n_1 in all directions in Q_1 and similarly c/n_2 in Q_2 (where c denotes the speed of light in vacuum). For convenience of the reader, this classical formula is presented and discussed in Examples 3.7 and 4.5 below – in the more general framework of the present paper.

Snell's law is classically derived from Fermat's principle, which states that the path of the light is the one which takes the least time to go from q_1 to q_2 . This principle will also be our starting point. It motivates directly the variational approach to the task of finding optimal traveltimes – also through the interface of *anisotropic* inhomogeneous domains which is the main concern of the present paper. The anisotropy is here given by having different prescribed speed profiles V_1 and V_2 in the two domains. These speed profiles, that we now allow to depend on both the position and the *direction* of the travel path, are considered in a given Riemannian manifold (Q, g) which thus defines the background for the domains $Q_1 \subset Q$, $Q_2 \subset Q$, their interface $\eta \subset Q$ as well as the space of paths under consideration.

We then show how the traveltime through the interface η in this general situation is minimized by broken geodesics of two Finsler metrics – one on each domain – which are determined by the ambient metric g and the prescribed speed profiles V_i . And we show and illustrate how the “breaking” at the interface must satisfy a specific condition on these Finsler metrics, i.e. the generalized Snell’s law. We note that the classical Snell’s law (with isotropic constant speed profiles c/n_i in (Q, g)) of course also follows from the general version – but then with simpler Finsler metrics which are just the conformally modified background metrics $\rho_i^2 \cdot g$, where the conformal factors are $\rho_i = n_i/c$.

Once the full set of relevant (possibly broken) time-parametrized Finsler geodesics issuing from one point q_1 has been found, it gives rise to a foliation of the domain Q by the wavefronts, each consisting of those points which have the same traveltime from q_1 – assuming that the geodesic equations have solutions for sufficiently large time. Some geodesics are not globally time-minimizing, so they do not contribute to the wavefront. This means that they have encountered a cut locus on their way from q_1 . We discuss in detail, and illustrate via a simple example (with constant speed profiles V_i in \mathbb{R}^2), how such cut loci typically appear in situations where V_2 is “larger” than V_1 . In this way – already in this example – the creation of cut loci follows from the discontinuous speed profile in Q notwithstanding the fact that there is neither curvature nor structural topology present in neither Q_1 nor Q_2 .

Outline of the paper. We begin in § 2 with the equivalence between the time spent when following a curve with a prescribed anisotropic velocity and the length of the curve computed with a certain Finsler metric. As alluded to above, when the velocity changes discontinuously from one medium to another, we have a situation that generalizes the classical Snell’s law. In § 3 we apply calculus of variations to find the critical points of the traveltime functional, i.e., the refracted trajectories, thus arriving at the generalized Snell’s law (9). The precise results are summarized in Theorem 3.4. Following an analogous procedure, we derive in § 4 a similar result, Theorem 4.1, for the reflected trajectories, including the generalized law of reflection (11). Finally, in § 5 we reduce the setting to \mathbb{R}^2 with constant Finsler metrics (i.e., Minkowski norms). This allows us to address some global issues, such as the existence and uniqueness of solutions to Snell’s law and the law of reflection, which are formulated in Theorems 5.2 and 5.3, as well as the study of the global minima of the traveltime functional. When considering rays that are shot in all directions from a given point of ignition, the trajectories that globally minimize the traveltime are the ones that generate the wavefront at each given time. This phenomenon is described and illustrated in Theorem 5.8. When a wavefront overlaps itself (at the intersection of two rays) it forms the cut locus, which is studied and characterized for a specific example in § 5.6.

2. INITIAL CONCEPTS

2.1. Traveltime and Finsler metrics. Let (Q, g) be a smooth Riemannian manifold of dimension $n \geq 2$, and consider also its tangent bundle

TQ . The metric g will measure actual distances in the space Q , i.e., if $\gamma : [0, t_0] \rightarrow Q$ is a (piecewise smooth) curve with tangent vector $\gamma'(t)$, then the distance traveled – when following this trajectory – is

$$L_g[\gamma] := \int_0^{t_0} \sqrt{g(\gamma'(t), \gamma'(t))} dt = \int_0^{t_0} \|\gamma'(t)\|_g dt \quad (1)$$

(splitting the integral accordingly at any break). This distance, this length, is independent of the parametrization of γ .

Assume now that we travel along γ with a prescribed velocity. We will allow this velocity to be anisotropic, i.e., it may depend on the point of the space and also on the direction. This means that there is a function

$$V : TQ \setminus \mathbf{0} \rightarrow (0, \infty)$$

such that $V(p, v)$ represents a fixed speed at $p \in Q$ in the direction $v \in T_pQ \setminus \{0\}$. Since $V(p, v)$ depends on the (oriented) direction of v but not on its length, this function must be positively homogeneous of degree zero, i.e., $V(p, v) = V(p, \lambda v)$ for all $\lambda > 0$. Then the time spent traveling along γ with this prescribed velocity is

$$T[\gamma] := \int_0^{t_0} \frac{\|\gamma'(t)\|_g}{V(\gamma(t), \gamma'(t))} dt, \quad (2)$$

which is again independent of reparametrizations (as long as they preserve the orientation of γ) thanks to the zero-homogeneity of V . The preservation of orientation is necessary, since in general we will not assume that $V(p, v) = V(p, -v)$. Nevertheless, observe that if γ is parametrized in such a way that its tangent vector always coincides with the velocity given by V , i.e. $\|\gamma'(t)\|_g = V(\gamma(t), \gamma'(t))$, then the parameter t directly measures the time, as $T[\gamma] = t_0$. We say in this case that γ is *time-parametrized*.

Let us denote by F the function given by the integrand in (2), i.e.,

$$\begin{aligned} F: TQ &\longrightarrow [0, \infty) \\ (p, v) &\longmapsto F(p, v) := \frac{\|v\|_g}{V(p, v)}, \end{aligned}$$

with $F(p, 0) := 0$ for all $p \in Q$. Then F is a Finsler metric if and only if V is smooth and the following oval (the indicatrix of F), that it defines at each $p \in Q$ is a strongly convex hypersurface of T_pQ :¹

$$\Sigma_p := \{v \in T_pQ : \|v\|_g = V(p, v)\} = \{v \in T_pQ : F(p, v) = 1\}.$$

Note that the length of γ computed with F coincides with the traveltime along γ with the prescribed velocity V :

$$L_F[\gamma] := \int_0^{t_0} F(\gamma(t), \gamma'(t)) dt = T[\gamma]. \quad (3)$$

¹Using any Euclidean metric in T_pQ , this means that the second fundamental form of Σ_p with respect to the inner normal vector is positive definite or, equivalently, that Σ_p has positive sectional curvature everywhere (see [17, Prop. 2.3]).

2.2. Snell's law; the general setting. In the classical Snell's law setting, a light ray travels between two different regions in \mathbb{R}^2 separated by a straight line. Let us generalize this framework by having two open subsets $Q_1, Q_2 \subset Q$ such that their closures $\overline{Q}_1, \overline{Q}_2$ are smooth manifolds with the same smooth boundary $\eta := \partial Q_1 = \partial Q_2$ and $\overline{Q}_1 \cup \overline{Q}_2 = Q$. Assume also that we have two different velocity functions – speed profiles – V_1, V_2 defined on TQ_1 and TQ_2 , respectively, and let F_1, F_2 be their corresponding Finsler metrics, so that (Q_1, F_1) and (Q_2, F_2) are smooth Finsler manifolds (and (Q, g) is still a Riemannian one).

Then, if $\gamma : [0, t_0] \rightarrow Q$ is a curve going from $\gamma(0) = q_1 \in Q_1$ to $\gamma(t_0) = q_2 \in Q_2$ and crossing η once at $\gamma(\tau)$, for some $\tau \in (0, t_0)$, we can still compute the traveltime by

$$T[\gamma] = \int_0^\tau F_1(\gamma(t), \gamma'(t))dt + \int_\tau^{t_0} F_2(\gamma(t), \gamma'(t))dt, \quad (4)$$

where, formally, $\int_0^\tau := \lim_{h \rightarrow \tau^-} \int_0^h$ and $\int_\tau^{t_0} := \lim_{h \rightarrow \tau^+} \int_h^{t_0}$. To avoid integrability and differentiability issues, we will further assume that F_1 and F_2 can be smoothly extended to η , so that the integrals always exist and we can also compute $F_1(\gamma(\tau), \gamma'(\tau^-))$, $F_2(\gamma(\tau), \gamma'(\tau^+))$ and their derivatives.

Definition 2.1. Fixing an interval $[0, t_0]$, some $\tau \in (0, t_0)$ and two points $q_1 \in Q_1, q_2 \in Q_2$, let \mathcal{N} be the set of all (regular) piecewise smooth curves $\gamma : [0, t_0] \rightarrow Q$ from $\gamma(0) = q_1$ to $\gamma(t_0) = q_2$ and crossing η once at $\gamma(\tau)$. We will call $T_{\mathcal{N}} : \mathcal{N} \rightarrow \mathbb{R}$ defined by (4) the *traveltime functional* on \mathcal{N} .

Note that any curve from q_1 to q_2 (piecewise smooth and crossing η once) can be reparametrized so that it belongs to \mathcal{N} , without affecting its traveltime. Therefore, the choice of t_0 and τ is arbitrary and non-restrictive (as well as the choice of 0 as the initial time).

Among all the possible paths in \mathcal{N} , light (or any wave, in general) will choose the one that minimizes the traveltime functional, according to the classical Fermat's principle. In a smooth Finsler manifold (Q, F) , it is well known that this trajectory (if it exists) must be a geodesic of F (see e.g. [2, § 5]). In our setting, we will have an F_1 -geodesic from q_1 to η (the incident trajectory) and an F_2 -geodesic from η to q_2 (the refracted trajectory), with a break at the point where the curve meets η . The goal is to find the change in the direction at the break point, i.e., the generalized Snell's law.

Remark 2.2. The functions V_1 and V_2 can represent not only the propagation velocities of a wave, but also the maximum possible velocities of an object or a person in two different media. In the former case, the problem is related to Fermat's principle (see e.g. [3, 4, 5, 12]) and Huygens' principle (see e.g. [8, 14, 15, 20]), with real-world applications in seismic theory [1, 21], sound waves [13] or wildfire modeling [7, 11, 16, 19], among others. In the latter, it becomes a general version of Zermelo's navigation problem, which seeks the fastest trajectory between two points for a moving object in a medium which, in turn, may also be moving with respect to the observer (see e.g. [6, 18]). Thus, real-world applications of this generalized Snell's law are multiple and widely varied. Some examples are naturally encountered when studying for example:

- The trajectories of light rays between two anisotropic media.
- The firefront of a wildfire that spreads over two different crop or fuel fields.
- The propagation of seismic waves when passing through layers of different anisotropic inhomogeneous materials.
- The fastest trajectory between a point on the ground and a point in the sea for a person who first has to run and then swim, taking the water current into account.
- The fastest trajectory for a travel first by boat and then by zeppeling, taking into account both the water current and the wind.

3. TIME-MINIMIZING TRAJECTORIES

3.1. Variational definitions and conventions. In order to find time-minimizing trajectories, one looks for the critical points of $T_{\mathcal{N}}$. Let $\gamma \in \mathcal{N}$ be our candidate curve, with $\gamma(0) = q_1 \in Q_1$, $\gamma(t_0) = q_2 \in Q_2$ and $\gamma(\tau) \in \eta$. Recall from Definition 2.1 that the choice of t_0 and τ is arbitrary, so we can further assume, without loss of generality, that γ is time-parametrized, i.e.,

$$\begin{aligned} F_1(\gamma(t), \gamma'(t)) &= 1, & \forall t \in [0, \tau), \\ F_2(\gamma(t), \gamma'(t)) &= 1, & \forall t \in (\tau, t_0], \end{aligned} \quad (5)$$

(so $F_1(\gamma(\tau), \gamma'(\tau^-)) = 1$ and $F_2(\gamma(\tau), \gamma'(\tau^+)) = 1$). The curve γ will be the base curve in a family of curves defined by a variation, in the following sense.

Definition 3.1. An *admissible variation* H of γ is a continuous map

$$\begin{aligned} H: (-\varepsilon, \varepsilon) \times [0, t_0] &\longrightarrow Q \\ (\omega, t) &\longmapsto H(\omega, t) := \gamma_\omega(t) \end{aligned}$$

such that:

- H is smooth on each $(-\varepsilon, \varepsilon) \times [s_i, s_{i+1}]$, where $0 = s_1 < \dots < s_k = \tau < \dots < s_{k+l} = t_0$ is a partition of $[0, t_0]$.²
- $H(\omega, \cdot) = \gamma_\omega \in \mathcal{N}$ for all $\omega \in (-\varepsilon, \varepsilon)$, with $\gamma_0 = \gamma$. In particular, $\gamma_\omega(0) = q_1$, $\gamma_\omega(t_0) = q_2$ and $\gamma_\omega(\tau) \in \eta$.

The first condition ensures that the tangent vector of the curve $\omega \mapsto \gamma_\omega(t)$ at $\omega = 0$ is always well-defined. Let us denote by $W(t) \in T_{\gamma(t)}Q$ this tangent vector. Then the curve in TQ given by

$$\begin{aligned} [0, t_0] &\longrightarrow TQ \\ t &\longmapsto (\gamma(t), W(t)) \end{aligned}$$

is called the *variational field* of γ associated with H .

Finally, we say that γ is a *critical point* of $T_{\mathcal{N}}$ if, for every admissible variation H , we have that

$$\left. \frac{d}{d\omega} T_{\mathcal{N}}[\gamma_\omega] \right|_{\omega=0} = 0.$$

²For simplicity, we will only focus on the break $s_k = \tau$ and omit writing the others explicitly, as they are not relevant for our study here.

Remark 3.2. Note that the only time-parametrized curve in the variation is the base curve $\gamma = \gamma_0$. In particular, the traveltime for γ is t_0 and τ is the time at which γ meets η . However, the parametrization of any other curve γ_ω is not arbitrary: it must satisfy the conditions $\gamma_\omega(0) = q_1$, $\gamma_\omega(t_0) = q_2$ and $\gamma_\omega(\tau) \in \eta$. So, we cannot assume that γ_ω is time-parametrized for $\omega \neq 0$, which means that the traveltime must be computed through (4) and τ does not have to coincide with the actual time at which γ_ω intersects η for $\omega \neq 0$.

Notation 3.3. From now on we will work in coordinates with the following usual conventions. Let $q = (q^1, \dots, q^n)$ be a coordinate system on Q and consider also the coordinate system (q, \dot{q}) on TQ , being $\dot{q} = (\dot{q}^1, \dots, \dot{q}^n)$ the natural coordinates on the fiber induced by q . This means that every $(p, v) \in TQ$ will be written in these coordinates as $(p^1, \dots, p^n, v^1, \dots, v^n) \in \mathbb{R}^{2n}$, where $p^i = q^i(p)$ and $v^i = \dot{q}^i(v) = v(q^i)$, i.e., $v = v^i \frac{\partial}{\partial q^i} \Big|_p$ (using Einstein's summation convention). Then:

- Any function on TQ such as the Finsler metrics F_j can be written in coordinates as $F_j \equiv F_j(q, \dot{q})$ and it makes sense to compute the following derivatives for $j = 1, 2$:

$$\begin{aligned} \frac{\partial F_j}{\partial q}(p, v) &:= \left(\frac{\partial F_j}{\partial q^1}(p, v), \dots, \frac{\partial F_j}{\partial q^n}(p, v) \right) \in \mathbb{R}^n, \\ \frac{\partial F_j}{\partial \dot{q}}(p, v) &:= \left(\frac{\partial F_j}{\partial \dot{q}^1}(p, v), \dots, \frac{\partial F_j}{\partial \dot{q}^n}(p, v) \right) \in \mathbb{R}^n, \end{aligned}$$

- A curve $\gamma : [0, t_0] \rightarrow Q$ will be written in these coordinates as $\gamma(t) \equiv q(t) = (q^1(t), \dots, q^n(t))$, where $q^i(t) := q^i(\gamma(t))$, and its tangent vector as $\gamma'(t) \equiv \dot{q}(t) = (\dot{q}^1(t), \dots, \dot{q}^n(t))$, where

$$\dot{q}^i(t) := \dot{q}^i(\gamma'(t)) = \frac{dq^i}{dt}(t), \quad i = 1, \dots, n.$$

This way, we can also write for $j = 1, 2$:

$$\begin{aligned} F_j(t) &:= F_j(q(t), \dot{q}(t)), \\ \frac{\partial F_j}{\partial q}(t) &:= \frac{\partial F_j}{\partial q}(q(t), \dot{q}(t)), \\ \frac{\partial F_j}{\partial \dot{q}}(t) &:= \frac{\partial F_j}{\partial \dot{q}}(q(t), \dot{q}(t)), \end{aligned}$$

- If we consider an admissible variation H of γ , then we write in coordinates $H(\omega, t) \equiv q_\omega(t)$, being $\dot{q}_\omega(t)$ the tangent vector of the curve $t \mapsto q_\omega(t)$. Also, the variational field is $t \mapsto (q(t), W(t))$, with $W(t) = W^i(t) \frac{\partial}{\partial q^i} \Big|_{q(t)} \equiv (W^1(t), \dots, W^n(t))$ and

$$W^i(t) = \frac{\partial q_\omega^i(t)}{\partial \omega} \Big|_{\omega=0}, \quad i = 1, \dots, n.$$

Finally, observe that even if q is not a global coordinate system on Q , the compactness of $[0, t_0]$ allows one to find a finite number of coordinate charts $\{(U_i, q_i)\}_{i=1}^{\tilde{k}+\tilde{l}}$ and a partition of the interval $0 = r_1 < \dots < r_{\tilde{k}} = \tau < \dots <$

$r_{\tilde{k}+\tilde{l}} = t_0$ such that $\gamma([r_i, r_{i+1}]) \subset U_i$. So, we can compute

$$\begin{aligned} T[\gamma] &= \int_0^\tau F_1(\gamma(t), \gamma'(t)) dt + \int_\tau^{t_0} F_2(\gamma(t), \gamma'(t)) dt = \\ &= \sum_{i=1}^{\tilde{k}-1} \int_{r_i}^{r_{i+1}} F_1(q_i(t), \dot{q}_i(t)) dt + \sum_{i=\tilde{k}}^{\tilde{k}+\tilde{l}-1} \int_{r_i}^{r_{i+1}} F_2(q_i(t), \dot{q}_i(t)) dt \end{aligned}$$

and therefore, without loss of generality, we can assume that q is a global coordinate system on Q (and so is (q, \dot{q}) on TQ).

3.2. Critical points of the traveltime functional. Let us now derive the conditions for the time-parametrized $\gamma \in \mathcal{N}$ to be a critical point of $T_{\mathcal{N}}$. If such a curve exists, then

$$\begin{aligned} 0 &= \left. \frac{d}{d\omega} T_{\mathcal{N}}[\gamma_\omega] \right|_{\omega=0} = \\ &= \left. \frac{d}{d\omega} \int_0^\tau F_1(q_\omega(t), \dot{q}_\omega(t)) dt \right|_{\omega=0} + \left. \frac{d}{d\omega} \int_\tau^{t_0} F_2(q_\omega(t), \dot{q}_\omega(t)) dt \right|_{\omega=0} \end{aligned} \quad (6)$$

for every admissible variation of γ . The first integral is

$$\begin{aligned} I_1 &= \int_0^\tau \left. \frac{\partial}{\partial \omega} F_1(q_\omega(t), \dot{q}_\omega(t)) \right|_{\omega=0} dt = \\ &= \int_0^\tau \left(\frac{\partial F_1}{\partial q^i}(t) W^i(t) + \frac{\partial F_1}{\partial \dot{q}^i}(t) \frac{d}{dt} W^i(t) \right) dt = \\ &= \int_0^\tau \left[\left(\frac{\partial F_1}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_1}{\partial \dot{q}^i}(t) \right) W^i(t) + \frac{d}{dt} \left(\frac{\partial F_1}{\partial \dot{q}^i}(t) W^i(t) \right) \right] dt = \\ &= \frac{\partial F_1}{\partial \dot{q}^i}(\tau^-) W^i(\tau) + \int_0^\tau \left(\frac{\partial F_1}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_1}{\partial \dot{q}^i}(t) \right) W^i(t) dt. \end{aligned}$$

Analogously, the second integral is

$$I_2 = - \frac{\partial F_2}{\partial \dot{q}^i}(\tau^+) W^i(\tau) + \int_\tau^{t_0} \left(\frac{\partial F_2}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_2}{\partial \dot{q}^i}(t) \right) W^i(t) dt.$$

Note that in both cases we have used that $W(0) = W(t_0) = 0$, since the variation has fixed endpoints. Then (6) reduces to

$$\begin{aligned} 0 &= \left[\frac{\partial F_1}{\partial \dot{q}^i}(\tau^-) - \frac{\partial F_2}{\partial \dot{q}^i}(\tau^+) \right] W^i(\tau) + \int_0^\tau \left(\frac{\partial F_1}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_1}{\partial \dot{q}^i}(t) \right) W^i(t) dt \\ &\quad + \int_\tau^{t_0} \left(\frac{\partial F_2}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_2}{\partial \dot{q}^i}(t) \right) W^i(t) dt \end{aligned}$$

and this must hold for every admissible variation $H(\omega, t) \equiv q_\omega(t)$, which implies that γ must satisfy:

$$\begin{aligned} \frac{\partial F_1}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_1}{\partial \dot{q}^i}(t) &= 0, & \forall t \in [0, \tau), \\ \frac{\partial F_2}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_2}{\partial \dot{q}^i}(t) &= 0, & \forall t \in (\tau, t_0], \end{aligned} \quad (7)$$

and the following additional condition at the break point $\gamma(\tau)$:

$$\left[\frac{\partial F_1}{\partial \dot{q}^i}(\tau^-) - \frac{\partial F_2}{\partial \dot{q}^i}(\tau^+) \right] W^i(\tau) = 0, \quad \forall H. \quad (8)$$

On the one hand, the two equations in (7) are the well-known Euler-Lagrange equations, equivalent to the (pre-)geodesic equations for the metrics F_1, F_2 . Along with the time-parametrization conditions in (5), this tells us that γ is an F_1 -unit speed geodesic from $\gamma(0) = q_1 \in Q_1$ to $\gamma(\tau) \in \eta$ and an F_2 -unit speed geodesic from there to $\gamma(t_0) = q_2 \in Q_2$. On the other hand, (8) is the condition at the break point $\gamma(\tau)$ that provides the desired Snell's law, as we will see next.

3.3. Generalized Snell's law. Let us now examine the equation (8). First, note that for every admissible variation $H(\omega, t) \equiv q_\omega(t)$, we have that $q_\omega(t) \in \eta$ for all $\omega \in (-\varepsilon, \varepsilon)$, where τ is fixed (recall Definition 3.1). Therefore:

$$\left. \frac{\partial q_\omega(\tau)}{\partial \omega} \right|_{\omega=0} = W(\tau) \in T_{q(\tau)}\eta$$

and this restriction must hold for all H , so (8) can be rewritten as

$$\left[\frac{\partial F_1}{\partial \dot{q}^i}(\tau^-) - \frac{\partial F_2}{\partial \dot{q}^i}(\tau^+) \right] u^i = 0, \quad \forall u \in T_{q(\tau)}\eta, \quad (9)$$

where $u^i = \dot{q}^i(u)$. This is the condition at the break point that governs the change in the trajectory's direction, and thus we arrive at the general version of Snell's law:

Theorem 3.4. *Let $\gamma : [0, t_0] \rightarrow Q$ be a time-parametrized curve in \mathcal{N} . Then γ is a critical point of $T_{\mathcal{N}}$ if and only if it is an F_1 -unit speed geodesic from $\gamma(0) \in Q_1$ to $\gamma(\tau) \in \eta$, an F_2 -unit speed geodesic from $\gamma(\tau) \in \eta$ to $\gamma(t_0) \in Q_2$, and it satisfies the generalized Snell's law given by (9) at the break point $\gamma(\tau) \in \eta$. In this case we say that γ is a refracted trajectory.*

Remark 3.5. The formulation we have carried out here is typical in classical mechanics, where instead of a Finsler metric F one has a Lagrangian defined as $L = K - P$, being K and P the total kinetic and potential energy of a system of particles, respectively, and instead of minimizing the traveltime functional $T = \int F$ one looks for the critical points of the action functional $S = \int L$. In this setting, (9) states that the total linear momentum of the system must be conserved in the tangent directions to η (see e.g. [9]).

Snell's law is essentially an equation system involving $q(\tau)$, $\dot{q}(\tau^-)$ and $\dot{q}(\tau^+)$. Typically, $q(\tau)$, $\dot{q}(\tau^-)$ are known from the initial conditions and the geodesic equations in the first region Q_1 , so one looks for $\dot{q}(\tau^+)$. Since $\dim(\eta) = \dim(T_{q(\tau)}\eta) = n - 1$, (9) provides exactly $n - 1$ independent equations. With the additional condition $F_2(q(\tau), \dot{q}(\tau^+)) = 1$, we have n equations to find the n -component vector $\dot{q}(\tau^+)$. Alternatively, one can just find the *direction* of $\dot{q}(\tau^+)$ from (9), which is usually more convenient in dimension 2, as we will see next.

3.4. The two-dimensional case. If $\dim(Q) = n = 2$, then the direction of the incident and refracted trajectory can be written in terms of the following angles.

Definition 3.6. Let $n = 2$ and $\gamma \in \mathcal{N}$. We define the *angles of incidence* and *refraction* $\theta_1, \theta_2 \in (-\pi, \pi]$ as the (oriented) directions of $\dot{q}(\tau^-), \dot{q}(\tau^+)$, respectively, measured with respect to the \dot{q}^1 -axis. Namely,

$$\tan \theta_1 = \frac{\dot{q}^2(\tau^-)}{\dot{q}^1(\tau^-)}, \quad \tan \theta_2 = \frac{\dot{q}^2(\tau^+)}{\dot{q}^1(\tau^+)},$$

with the restriction that $v_{\theta_1} := (\cos \theta_1, \sin \theta_1)$ and $v_{\theta_2} := (\cos \theta_2, \sin \theta_2)$ must point to the Q_2 -side of $T_{q(\tau)}\eta$.

This way, for a fixed break point $q(\tau) \in \eta$, Snell's law is transformed into an equation between the angles of incidence and refraction. Specifically, it can always be written explicitly in terms of $\tan \theta_1$ and $\tan \theta_2$. Indeed, since F_1 and F_2 are positively homogeneous of degree one, $F_j(\lambda v) = |\lambda|F_j(\operatorname{sgn}(\lambda)v)$ for all $v \in T_{q(\tau)}\eta$, $\lambda \in \mathbb{R}$ and, as a consequence,

$$\frac{\partial F_j}{\partial \dot{q}}(\lambda v) = \frac{\partial F_j}{\partial \dot{q}}(\operatorname{sgn}(\lambda)v), \quad j = 1, 2,$$

so, in particular,

$$\frac{\partial F_j}{\partial \dot{q}^i}(\dot{q}^1, \dot{q}^2) = \frac{\partial F_j}{\partial \dot{q}^i} \left(\operatorname{sgn}(\dot{q}^1) \left(1, \frac{\dot{q}^2}{\dot{q}^1} \right) \right) = \frac{\partial F_j}{\partial \dot{q}^i}(\operatorname{sgn}(\dot{q}^1)(1, \tan \theta_j)),$$

for $i, j = 1, 2$, where the references to $q(\tau)$ and τ^\pm have been omitted.

Therefore, in general one can solve (9) for $\tan \theta_2$ and then obtain $\theta_2 \in (-\pi, \pi]$ by imposing the restriction that v_{θ_2} must point to the Q_2 -side of $T_{q(\tau)}\eta$.

Example 3.7. Let us recover the classical Snell's law in the Euclidean plane. In this usual setting, $Q = \mathbb{R}^2$, $(q, \dot{q}) = (x, y, \dot{x}, \dot{y})$ and the background Riemannian metric $g = dx^2 + dy^2$ is the natural one in \mathbb{R}^2 . Let $Q_1 = \{(x, y) \in \mathbb{R}^2 : x < 0\}$ and $Q_2 = \{(x, y) \in \mathbb{R}^2 : x > 0\}$, so η is the vertical line $x = 0$. In this situation, assume that the propagation of light is isotropic and its velocity depends on the refractive index of the medium. Assume also that Q_1 and Q_2 are homogeneous, i.e., their respective refractive indices n_1, n_2 are constant. Then:

$$\begin{aligned} V_1(x, y, \dot{x}, \dot{y}) = \frac{c}{n_1} &\Rightarrow F_1(x, y, \dot{x}, \dot{y}) = \frac{n_1}{c} \sqrt{\dot{x}^2 + \dot{y}^2}, \\ V_2(x, y, \dot{x}, \dot{y}) = \frac{c}{n_2} &\Rightarrow F_2(x, y, \dot{x}, \dot{y}) = \frac{n_2}{c} \sqrt{\dot{x}^2 + \dot{y}^2}, \end{aligned}$$

where c is the speed of light in vacuum. The angles of incidence and refraction are given by

$$\tan \theta_1 = \frac{\dot{y}(\tau^-)}{\dot{x}(\tau^-)}, \quad \tan \theta_2 = \frac{\dot{y}(\tau^+)}{\dot{x}(\tau^+)},$$

so, omitting the reference to τ^\pm we get:

$$\begin{aligned} \frac{\partial F_j}{\partial \dot{y}} &= \frac{n_j}{c} \frac{\dot{y}}{\sqrt{\dot{x}^2 + \dot{y}^2}} = \frac{n_j}{c} \frac{\dot{y}}{|\dot{x}| \sqrt{1 + \dot{y}^2/\dot{x}^2}} = \pm \frac{n_j}{c} \frac{\tan \theta_j}{\sqrt{1 + \tan^2 \theta_j}} = \\ &= \pm \frac{n_j}{c} \tan \theta_j |\cos \theta_j| = \frac{n_j}{c} \sin \theta_j, \quad j = 1, 2, \end{aligned}$$

where the \pm vanishes because $\text{sgn}(\dot{x}(\tau^-)) = \text{sgn}(\cos \theta_1)$ and $\text{sgn}(\dot{x}(\tau^+)) = \text{sgn}(\cos \theta_2)$. This way, since in this particular example $\{(0, 1)\}$ is a basis of $T_{(0, y(\tau))}\eta$ at any break point $(0, y(\tau)) \in \eta$, (9) yields the classical Snell's law:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2, \quad (10)$$

with the restriction that $\text{sgn}(\cos \theta_1) = \text{sgn}(\cos \theta_2) > 0$, so that both v_{θ_1} and v_{θ_2} point to Q_2 .

4. REFLECTED TRAJECTORIES

4.1. The law of reflection. So far, given a curve γ departing from Q_1 and intersecting η , we have been looking for the refracted trajectory, i.e., the one that crosses η and continues into Q_2 . It is interesting also to seek the reflected trajectory, which will be the one that minimizes the traveltime in a slightly different set of curves than \mathcal{N} in Definition 2.1.

Specifically, fixing $[0, t_0]$, $\tau \in (0, t_0)$ and two points $q_1, q_2 \in Q_1$, we define \mathcal{N}^* as the set of all (regular) piecewise smooth curves $\gamma : [0, t_0] \rightarrow Q$ from $\gamma(0) = q_1$ to $\gamma(t_0) = q_2$ such that γ remains entirely in $\overline{Q_1}$, touching η once at $\gamma(\tau)$.

If one now looks for the critical points of $T_{\mathcal{N}^*} : \mathcal{N}^* \rightarrow \mathbb{R}$ in a similar way as in the previous section, one finds, on the one hand, the Euler-Lagrange equations (compare with (7)):

$$\frac{\partial F_1}{\partial q^i}(t) - \frac{d}{dt} \frac{\partial F_1}{\partial \dot{q}^i}(t) = 0, \quad \forall t \in [0, \tau] \cup (\tau, t_0],$$

which now state that the candidate curve $\gamma(t) \equiv q(t)$ must be an F_1 -geodesic from q_1 to η and also from η to q_2 , and on the other hand, instead of Snell's law (9) (i.e., the refraction law), we get the following reflection law:

$$\left[\frac{\partial F_1}{\partial \dot{q}^i}(\tau^-) - \frac{\partial F_1}{\partial \dot{q}^i}(\tau^+) \right] u^i = 0, \quad \forall u \in T_{q(\tau)}\eta. \quad (11)$$

Therefore, analogously to Theorem 3.4, we have the following result.

Theorem 4.1. *Let $\gamma : [0, t_0] \rightarrow Q$ be a time-parametrized curve in \mathcal{N}^* . Then γ is a critical point of $T_{\mathcal{N}^*}$ if and only if it is an F_1 -unit speed geodesic from $\gamma(0) \in Q_1$ to $\gamma(\tau) \in \eta$ and from $\gamma(\tau) \in \eta$ to $\gamma(t_0) \in Q_1$, and it satisfies the law of reflection given by (11) at the break point $\gamma(\tau) \in \eta$. In this case we say that γ is a reflected trajectory.*

Moreover, the discussion developed in § 3.4 becomes valid here as well, i.e., in the two-dimensional case the law of reflection can be written in terms of the following angles.

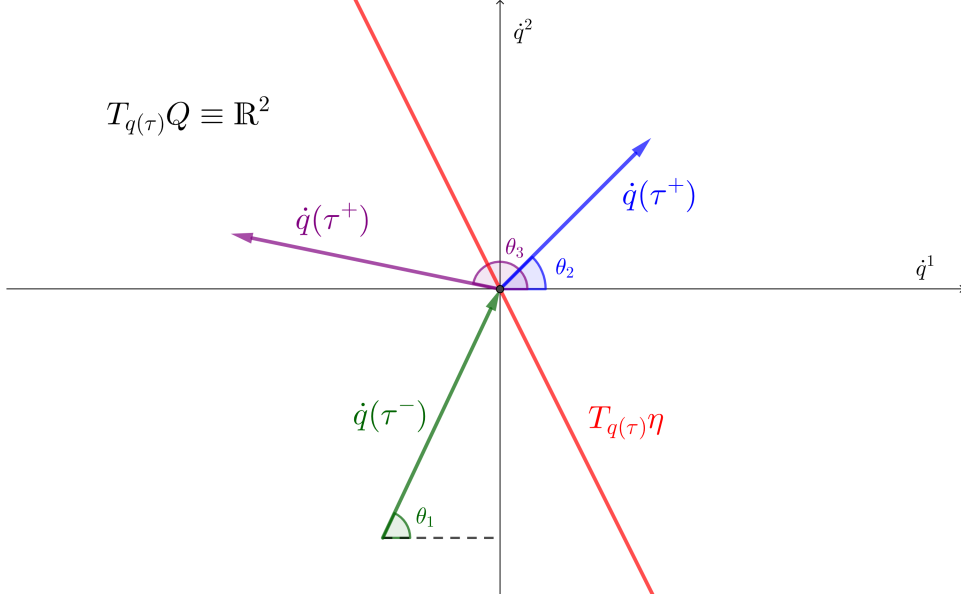


FIGURE 1. In the two-dimensional case, Snell's law or the law of reflection can be written in terms of the angle of incidence θ_1 and the angle of refraction θ_2 or reflection θ_3 , respectively. When looking for the refracted trajectory $\gamma \in \mathcal{N}$ (resp. the reflected trajectory $\gamma \in \mathcal{N}^*$), $\dot{q}(\tau^+)$ must point to the Q_2 -side (resp. the Q_1 -side) of $T_{q(\tau)}\eta$.

Definition 4.2. Let $n = 2$ and $\gamma \in \mathcal{N}^*$. We define the *angles of incidence* and *reflection* $\theta_1, \theta_3 \in (-\pi, \pi]$ as the (oriented) directions of $\dot{q}(\tau^-), \dot{q}(\tau^+)$, respectively, measured with respect to the \dot{q}^1 -axis. Namely,

$$\tan \theta_1 = \frac{\dot{q}^2(\tau^-)}{\dot{q}^1(\tau^-)}, \quad \tan \theta_3 = \frac{\dot{q}^2(\tau^+)}{\dot{q}^1(\tau^+)},$$

with the restriction that $v_{\theta_1} := (\cos \theta_1, \sin \theta_1)$ must point to the Q_2 -side and $v_{\theta_3} := (\cos \theta_3, \sin \theta_3)$ to the Q_1 -side of $T_{q(\tau)}\eta$.

Note the difference between the angles of refraction and reflection (compare with Definition 4.2): now $\dot{q}(\tau^+)$ must point to the Q_1 -side of $T_{q(\tau)}\eta$ because otherwise, γ would cross to Q_2 and thus, $\gamma \notin \mathcal{N}^*$ (see Figure 1).

Example 4.3. In the same setting as in Example 3.7, let us recover the classical reflection law. The angles of incidence and reflection are

$$\tan \theta_1 = \frac{\dot{y}(\tau^-)}{\dot{x}(\tau^-)}, \quad \tan \theta_3 = \frac{\dot{y}(\tau^+)}{\dot{x}(\tau^+)},$$

with $\text{sgn}(\dot{x}(\tau^-)) = \text{sgn}(\cos \theta_1)$ and $\text{sgn}(\dot{x}(\tau^+)) = \text{sgn}(\cos \theta_2)$, so

$$\frac{\partial F_1}{\partial \dot{y}}(\tau^-) = \frac{n_1}{c} \sin \theta_1, \quad \frac{\partial F_1}{\partial \dot{y}}(\tau^+) = \frac{n_1}{c} \sin \theta_3.$$

Thus, (11) directly yields

$$\sin \theta_1 = \sin \theta_3, \tag{12}$$

with the restriction that $\operatorname{sgn}(\cos \theta_1) > 0$ and $\operatorname{sgn}(\cos \theta_3) < 0$, so that v_{θ_1} points to Q_2 and v_{θ_3} to Q_1 . For example, if $\theta_1 \in [0, \frac{\pi}{2})$, then $\theta_3 = \pi - \theta_1$. If $\theta_1 \in (-\frac{\pi}{2}, 0)$, then $\theta_3 = -\pi - \theta_1$.³

4.2. The total reflection phenomenon. In general, the existence of solutions to Snell's law (9) is not guaranteed. Namely, given an incident velocity $\dot{q}(\tau^-)$, there might not be a solution $\dot{q}(\tau^+)$ pointing to the Q_2 -side of $T_{q(\tau)}\eta$. In this case, there is no refraction and the incoming trajectory is totally reflected without being able to cross into Q_2 . This is known as the *total reflection phenomenon*. The limit between the existence and non-existence of solution is given by the situation where $\dot{q}(\tau^+)$ is tangent to $T_{q(\tau)}\eta$, which motivates the following definition.

Definition 4.4. We say that an incident velocity $\dot{q}(\tau^-)$ (or an angle of incidence θ_1 in the two-dimensional case) is *critical* if there exists a solution $\dot{q}(\tau^+)$ to Snell's law (9) which is tangent to $T_{q(\tau)}\eta$.

The trajectory γ associated with the critical velocity can still be considered as an actual refracted trajectory as long as immediately after touching η , it passes to Q_2 or remains on η with the F_2 metric, i.e. if there exists $\delta > 0$ such that $\gamma(\tau+t) \in (\overline{Q_2}, F_2)$ for all $t \in (0, \delta)$. We say in this case that γ is a *critical trajectory*. In fact, critical trajectories along η are relevant in situations such as the one we will describe in the next section.

Example 4.5. In the classical setting of Example 3.7, v_{θ_2} is tangent to η when $\theta_2 = \pm \frac{\pi}{2}$, so the critical angles are given by

$$\theta_c^\pm = \pm \arcsin\left(\frac{n_2}{n_1}\right).$$

Note that these angles only exist if $n_1 > n_2$, i.e. if the velocity V_2 on Q_2 is greater than V_1 on Q_1 . If, for example, $n_1 = 1$ and $n_2 = \frac{1}{2}$, then $\theta_c^\pm = \pm \frac{\pi}{6}$. In the cases $\theta_1 > \frac{\pi}{6}$ or $\theta_1 < -\frac{\pi}{6}$, there is no angle θ_2 satisfying the classical Snell's law (10) and pointing to Q_2 . In consequence, there does not exist a time-minimizing path from any $q_1 \in Q_1$ to any $q_2 \in Q_2$ which meets η with $\theta_1 > \frac{\pi}{6}$ or $\theta_1 < -\frac{\pi}{6}$, so there is no refraction and the incident trajectory is totally reflected with an angle θ_3 satisfying the classical law of reflection (12). In other terms, suppose q_1 has Euclidean distance d to η , then the Euclidean size of the "window" through which a time-minimizing trajectory can go through η is $\frac{2d \cdot n_2}{\sqrt{n_1^2 - n_2^2}}$.

5. GLOBAL CONSIDERATIONS IN \mathbb{R}^2

In order to reduce the casuistry we can find in the most general case and fix some ideas, we will reduce our study in this section to the classical setting (see Example 3.7), although maintaining the Finslerian nature of the metrics. Namely, $Q = \mathbb{R}^2$, $(q, \dot{q}) = (x, y, \dot{x}, \dot{y})$, $g = dx^2 + dy^2$, $Q_1 = \{(x, y) \in$

³Another usual way to define the angles of incidence, refraction and reflection is to restrict their values to $[0, \frac{\pi}{2})$ and measure them with respect to the normal direction to $T_{(x(\tau), y(\tau))}\eta$. In this case, the classical Snell's law would still be (10) but without any additional restriction, and the classical reflection law would read $\theta_1 = \theta_3$.

$\mathbb{R}^2 : x < 0\}$, $Q_2 = \{(x, y) \in \mathbb{R}^2 : x > 0\}$ and F_1, F_2 are Finsler metrics independent of the position (i.e., Minkowski norms):

$$F_j(x, y, \dot{x}, \dot{y}) = F_j(\dot{x}, \dot{y}) = \frac{\sqrt{\dot{x}^2 + \dot{y}^2}}{V_j(\dot{x}, \dot{y})}, \quad j = 1, 2.$$

In particular, this means that the geodesics of F_1, F_2 are straight lines. Note also that the angles of incidence, refraction and reflection are restricted to the values $\theta_1 \in (-\frac{\pi}{2}, \frac{\pi}{2})$, $\theta_2 \in [-\frac{\pi}{2}, \frac{\pi}{2}]$, $\theta_3 \in (-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$, and there are at most two critical angles θ_c^\pm , with $\theta_c^+ > \theta_c^-$: θ_c^+ corresponds to $\theta_2 = \frac{\pi}{2}$ and θ_c^- to $\theta_2 = -\frac{\pi}{2}$ (the uniqueness of θ_c^\pm when they exist is proven below in Lemma 5.1).

5.1. Existence and uniqueness of solution. In this setting, Snell's law (9) reduces to

$$\frac{\partial F_1}{\partial \dot{y}}(\tau^-) = \frac{\partial F_2}{\partial \dot{y}}(\tau^+).$$

Now, being F_1 and F_2 Minkowski norms, we have that $\frac{\partial F_j}{\partial y} = 0$ and thus, from the Euler-Lagrange equations (7):

$$\begin{aligned} \frac{d}{dt} \frac{\partial F_1}{\partial \dot{y}}(t) &= 0, & \forall t \in [0, \tau), \\ \frac{d}{dt} \frac{\partial F_2}{\partial \dot{y}}(t) &= 0, & \forall t \in (\tau, t_0]. \end{aligned}$$

which means, together with Snell's law, that there is a conserved quantity along the entire refracted curve $\gamma \in \mathcal{N}$.⁴ This constant is called the *raypath parameter* (see [3]). It only depends on the initial direction, due to the zero-homogeneity of $\frac{\partial F_j}{\partial \dot{y}}$. So, Snell's law becomes

$$P_1(\theta_1) = P_2(\theta_2), \quad (13)$$

with

$$P_j(\theta) := \frac{\partial F_j}{\partial \dot{y}}(\theta) = \frac{\sin \theta}{V_j(\theta)} + \cos \theta \frac{\partial}{\partial \theta} \left(\frac{1}{V_j(\theta)} \right), \quad j = 1, 2, \quad (14)$$

where we have used that $\frac{\partial}{\partial \dot{y}} = \frac{\cos^2 \theta}{x} \frac{\partial}{\partial \theta}$. Note that $P_j(\theta)$ is smooth, thanks to the smoothness and positiveness of $V_j(\theta)$.

Lemma 5.1. *The functions $P_j : [-\frac{\pi}{2}, \frac{\pi}{2}] \rightarrow \mathbb{R}$ are strictly increasing. Also, the critical angles θ_c^\pm exist (and are necessarily unique) if and only if $V_1(\pm\frac{\pi}{2}) < V_2(\pm\frac{\pi}{2})$, respectively.⁵*

Proof. One can easily verify that

$$\frac{dP_j}{d\theta} = \frac{\cos \theta}{V_j^3} \left[V_j^2 + 2 \left(\frac{dV_j}{d\theta} \right)^2 - V_j \frac{d^2 V_j}{d\theta^2} \right],$$

where the term in brackets is exactly the second fundamental form of the indicatrix Σ_j of F_j (given by the curve $(\pi, \pi] \ni \theta \mapsto V_j(\cos \theta, \sin \theta)$) with

⁴This is related to Noether's theorem: the translational symmetry of the setting along the y -axis implies the existence of a *constant of motion*.

⁵As in the classical case, for both critical angles θ_c^\pm to exist we need the velocity V_2 to be greater than V_1 on η in both directions $\pm\frac{\pi}{2}$.

respect to the natural Euclidean metric in \mathbb{R}^2 and the inner normal vector. The strong convexity of Σ_j and the positiveness of V_j ensure that $\frac{dP_j}{d\theta} > 0$ for all $\theta \in (-\frac{\pi}{2}, \frac{\pi}{2})$, so we conclude that $P_j(\theta)$ is strictly increasing in $\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$. In particular, this means that the maximum of $P_j(\theta)$ is $P_j(\frac{\pi}{2}) = \frac{1}{V_j(\frac{\pi}{2})}$ and the minimum $P_j(-\frac{\pi}{2}) = -\frac{1}{V_j(-\frac{\pi}{2})}$. So $\theta_c^\pm \in (-\frac{\pi}{2}, \frac{\pi}{2})$ exists if and only if $P_1(\theta_c^\pm) = P_2(\pm\frac{\pi}{2})$, i.e. if and only if $\frac{1}{V_1(\pm\frac{\pi}{2})} > \frac{1}{V_2(\pm\frac{\pi}{2})}$ (and in this case θ_c^\pm is unique due to the fact that $P_1(\theta)$ is strictly increasing). \square

Theorem 5.2. *Given any angle of incidence $\theta_1 \in (\theta_c^-, \theta_c^+)$ (replace θ_c^\pm with $\pm\frac{\pi}{2}$ if the critical angles do not exist), there exists a unique angle of refraction $\theta_2 \in (-\frac{\pi}{2}, \frac{\pi}{2})$ satisfying Snell's law. Equivalently, given any $q_1 \in Q_1$ and $q_2 \in Q_2$, there exists a unique refracted trajectory $\gamma \in \mathcal{N}$ joining both points.*

Proof. Given $\theta_1 \in (\theta_c^-, \theta_c^+)$ (or $\theta_1 \in (-\frac{\pi}{2}, \frac{\pi}{2})$ if the critical angles do not exist), from Lemma 5.1 we know that $P_2(-\frac{\pi}{2}) < P_1(\theta_1) < P_2(\frac{\pi}{2})$, so there exists $\theta_2 \in (-\frac{\pi}{2}, \frac{\pi}{2})$ satisfying (13) and since $P_2(\theta)$ is strictly increasing, this angle is necessarily unique. Note that the converse is also true, i.e., given any $\theta_2 \in (-\frac{\pi}{2}, \frac{\pi}{2})$ we can find $\theta_1 \in (\theta_c^-, \theta_c^+)$ (or $\theta_1 \in (-\frac{\pi}{2}, \frac{\pi}{2})$) satisfying Snell's law. In \mathbb{R}^2 with straight lines as geodesics, this ensures that all the refracted trajectories cover Q_2 entirely, independently of the initial point $q_1 \in Q_1$. \square

Of course, one can follow the same steps and derive an analogous result for the existence and uniqueness of solutions to the law of reflection, which reads $P_1(\theta_1) = P_1(\theta_3)$ in terms of the raypath parameter.

Theorem 5.3. *Given any angle of incidence $\theta_1 \in (-\frac{\pi}{2}, \frac{\pi}{2})$, there exists a unique angle of reflection $\theta_3 \in (-\pi, -\frac{\pi}{2}) \cup (\frac{\pi}{2}, \pi]$ satisfying the law of reflection. Equivalently, given any $q_1, q_2 \in \hat{Q}_1$, there exists a unique refracted trajectory $\gamma \in \mathcal{N}^*$ joining both points.*

5.2. New reflected trajectories. The existence and uniqueness of refracted and reflected trajectories ensure that, once the initial and final points are fixed, there is only one critical point of $T_{\mathcal{N}}$ and $T_{\mathcal{N}^*}$, which therefore must be a global minimum. However, recall that curves in \mathcal{N} and \mathcal{N}^* have the restriction that they must intersect η only once. One might wonder if more contact points with η could result in a curve that spends even less time than the usual refracted or reflected trajectory. Namely, we can define new sets $\hat{\mathcal{N}}$ and $\hat{\mathcal{N}}^*$ in the same way as \mathcal{N} and \mathcal{N}^* , respectively, but with the restriction that the (piecewise smooth) curves must meet η at least once (but possibly at more points), and look for the critical points of the traveltime functional on these sets.

Let γ be a curve from q_1 to q_2 that meets η for the first time at τ_1 and leaves it for the last time at τ_2 . In order to minimize the traveltime, the curve from $\gamma(\tau_1)$ to $\gamma(\tau_2)$ must be the segment of the straight line η joining these points (since it is a geodesic in the present \mathbb{R}^2 -setting under consideration). Now, if $\gamma \in \hat{\mathcal{N}}$, by the triangle inequality (see e.g. [2, Thm. 1.2.2]) any two-segment trajectory from $q_1 \in Q_1$ to $\gamma(\tau) \in \eta$ (for any $\tau \in [\tau_1, \tau_2]$) and from there to $q_2 \in Q_2$ spends less time than γ . So the critical points of

$T_{\mathcal{N}}$ must always belong to \mathcal{N} , i.e., they are the usual two-segment refracted trajectories crossing η once. However, if $\gamma \in \hat{\mathcal{N}}^*$, the travel along η with the metric F_2 can be an advantage. A real-world example is when moving between two points in the sea next to the shore: it can be faster to swim first to the shore, then run along the beach line edge and finally swim again to the final point. Obviously, the traveltime on η has to be computed through F_2 (otherwise, the usual reflected trajectory and hence also the direct straight line trajectory in Q_1 is always faster):

$$T_{\hat{\mathcal{N}}^*}[\gamma] = \int_0^{\tau_1} F_1(\gamma'(t))dt + \int_{\tau_1}^{\tau_2} F_2(\gamma'(t))dt + \int_{\tau_2}^{t_0} F_1(\gamma'(t))dt.$$

Taking this into account, we can now define a suitable variation and apply the calculus of variations in the same way as in § 3.2. This gives the following conditions for a time-parametrized curve γ to be a critical point of $T_{\hat{\mathcal{N}}^*}$:

- γ is composed of three straight lines: an F_1 -unit segment γ_1 from $q_1 \in Q_1$ to $\gamma(\tau_1) \in \eta$, an F_2 -unit segment $\gamma_2 \subset \eta$ between $\gamma(\tau_1)$ and $\gamma(\tau_2)$, and a final F_1 -unit segment γ_3 from $\gamma(\tau_2) \in \eta$ to $q_2 \in Q_1$ (see Figure 2).
- At the break points, $\gamma(t) \equiv (x(t), y(t))$ (for $t = \tau_1$ and $t = \tau_2$, respectively) must satisfy:

$$\begin{aligned} \frac{\partial F_1}{\partial \dot{y}}(\tau_1^-) &= \frac{\partial F_2}{\partial \dot{y}}(\tau_1^+), \\ \frac{\partial F_2}{\partial \dot{y}}(\tau_2^-) &= \frac{\partial F_1}{\partial \dot{y}}(\tau_2^+). \end{aligned}$$

Equivalently, in terms of the (constant) directions $\theta_1, \theta_2, \theta_3 \in (-\pi, \pi]$ of each segment and the raypath parameter (14):

$$P_1(\theta_1) = P_2(\theta_2) = P_1(\theta_3). \quad (15)$$

In particular, this means that θ_1 must be critical (so that $\theta_2 = \pm \frac{\pi}{2}$) and θ_3 must be exactly the angle of reflection we would obtain at τ_1 , being γ_1 the incident trajectory (see Figure 2).

If it exists, this three-segment trajectory is always faster than the usual reflection. However, its existence is not guaranteed: it depends on whether the critical angle exists (which in turn depends on whether the velocity V_2 on η is greater than V_1 on Q_1 ; recall Lemma 5.1) and, if it does, on the initial and final points. So, fixing two points $q_1, q_2 \in Q_1$, the global minimum of $T_{\hat{\mathcal{N}}^*}$ is the three-segment trajectory satisfying the above conditions, if it exists, or the usual reflected trajectory, if it does not. This global comparison should, of course, be benchmarked against the traveltime for the direct straight line between q_1 and q_2 . This will be done in § 5.3 below.

Remark 5.4. At first glance, these new reflected trajectories might only seem relevant in situations when you can willingly change the directions midway at points of η . Obviously, a light ray refracted at the critical angle will continue along η indefinitely, without returning to Q_1 . When computing the whole propagation of a wave though, it is crucial to take into account all the possible trajectories. Indeed, Huygens' principle states that each point of the wavefront can be regarded as an independent source that shoots wave

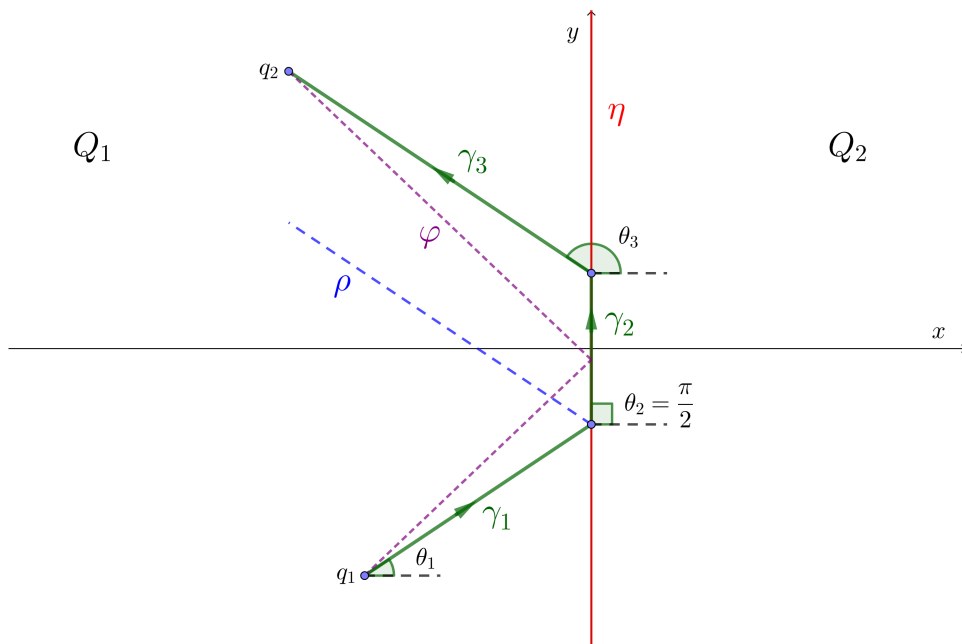


FIGURE 2. There are situations where moving along γ is faster than following the usual reflected trajectory φ or even the straight line between two fixed points $q_1, q_2 \in Q_1$. This can only happen if γ_2 is traveled with velocity $V_2(\frac{\pi}{2})$ along η and if θ_1 and θ_3 are, respectively, the critical and reflection angles of γ_1 . In particular, γ_3 must be parallel to the reflection ρ of γ_1 .

trajectories in all directions. Among all of them, those that globally minimize the traveltime are the ones that keep providing points in the wavefront at later instants of times. So, a curve such as γ in Figure 2 can indeed represent a realistic wave trajectory that contributes to the formation of the wavefront (see § 5.5).

5.3. Globally time-minimizing trajectories. Taking a step further, we can now remove the restriction that the curves must touch η .

Definition 5.5. Fixing any two points $q_1, q_2 \in Q$, let \mathcal{M} be the set of all (regular) piecewise smooth curves from q_1 to q_2 . We say that $\gamma \in \mathcal{M}$ globally minimizes the traveltime if $T[\gamma] \leq T[\varphi]$ for all $\varphi \in \mathcal{M}$.

If $q_1 \in Q_1, q_2 \in Q_2$, then it is clear that $\mathcal{M} = \hat{\mathcal{N}}$, so the refracted trajectories always globally minimize the traveltime. This includes the critical trajectories along η , as stated in the following result.

Proposition 5.6. If $\gamma : [0, t_0] \rightarrow Q$ is a critical trajectory (associated with one of the critical angles θ_c^\pm), then γ globally minimizes the traveltime between $\gamma(0) \in Q_1$ and $\gamma(t_0) \in \eta$.

Proof. Given $\gamma(t_0) = (0, y_0) \in \eta$, suppose there exists another curve $\xi_1 \in \mathcal{M}$ that arrives earlier at this point from $\gamma(0) \in Q_1$. Consider now $(\delta, y_0) \in Q_2$

with $\delta > 0$ and let φ_δ be the unique refracted trajectory from $\gamma(0)$ to (δ, y_0) , which globally minimizes the traveltime. The smoothness of the raypath parameter given by (14) ensures that

$$\lim_{\delta \rightarrow 0} T[\varphi_\delta] = T[\gamma] > T[\xi_1].$$

So, for a sufficiently small $\delta > 0$, the straight line ξ_2 from $(0, y_0)$ to (δ, y_0) satisfies that $T[\varphi_\delta] > T[\xi_1] + T[\xi_2]$, which means that we have built a path in \mathcal{M} faster than φ_δ , arriving at a contradiction. \square

If $q_1, q_2 \in Q_1$ though, the global minimum of $T_{\mathcal{N}^*}$ competes against the straight line in Q_1 between q_1 and q_2 . It is clear that the straight line is always faster than the usual reflection, due to the triangle inequality for the metric F_1 in Q_1 , but in some situations the three-segment reflected trajectory can globally minimize the traveltime between q_1 and q_2 , thus creating a point on the cut locus from q_1 – we elaborate further on this issue in § 5.6 below. We first illustrate the setting for these possibilities next with an example.

5.4. Anisotropic example. Let us consider the following velocity functions in \mathbb{R}^2 :

$$V_j(\theta) = \frac{a_j(1 - \varepsilon_j^2)}{1 - \varepsilon_j \cos(\theta - \phi_j)}, \quad j = 1, 2,$$

which define ellipses of eccentricity ε_j , respectively, centered at one of their foci, with semi-major axis a_j oriented in the direction ϕ_j (with respect to the \hat{x} -axis) and such that $\phi_j = 0$ points to their other respective focus point. Precisely, $V_j(\theta)$ is the (Euclidean) length of the vector from the centered focus to the ellipse in the direction θ , and this ellipse becomes the indicatrix of the Finsler metric⁶

$$F_j(\hat{x}, \hat{y}) = \frac{\sqrt{\hat{x}^2 + \hat{y}^2}}{a_j(1 - \varepsilon_j^2)} \left[1 - \varepsilon_j \cos \left(\arctan \left(\frac{\hat{y}}{\hat{x}} \right) - \phi_j \right) \right], \quad j = 1, 2.$$

One can easily verify that

$$\frac{\partial F_j}{\partial \hat{y}}(\theta) = \frac{\sin \theta - \varepsilon_j \sin \phi_j}{a_j(1 - \varepsilon_j^2)}, \quad j = 1, 2,$$

so (15) becomes

$$\frac{\partial F_1}{\partial \hat{y}}(\theta_k) = \frac{\partial F_2}{\partial \hat{y}}(\theta_2) \quad \Rightarrow \quad \frac{\sin \theta_k - \varepsilon_1 \sin \phi_1}{a_1(1 - \varepsilon_1^2)} = \frac{\sin \theta_2 - \varepsilon_2 \sin \phi_2}{a_2(1 - \varepsilon_2^2)},$$

for $k = 1, 3$. In particular, in order to find the three-segment reflected trajectory, θ_1 must be the (critical) angle such that $\theta_2 = \pm \frac{\pi}{2}$, i.e.,

$$\theta_1 = \arcsin \left(\frac{a_1(1 - \varepsilon_1^2)}{a_2(1 - \varepsilon_2^2)} (\pm 1 - \varepsilon_2 \sin \phi_2) + \varepsilon_1 \sin \phi_1 \right), \quad (16)$$

and θ_3 must satisfy the same equation but pointing to Q_1 .⁷

⁶This type of Finsler metrics are used to model the spread of wildfires (see e.g. [10, 16]).

⁷Even though the velocities are anisotropic, the law of reflection in this example is the same as in the case with Euclidean metrics (see Example 4.3). However, this is not true in general.

Now, suppose we want to go from $q_1 = (-x_0, -y_0)$ to $q_2 = (-x_0, y_0)$, for some $x_0, y_0 > 0$, and let us set $a_1 = a_2 = 1$, $\varepsilon_1 = \varepsilon_2 = \frac{1}{2}$, $\phi_1 = 0$ and $\phi_2 = \frac{\pi}{2}$. Then (16) becomes $\theta_1 = \arcsin(\pm 1 - \frac{1}{2})$, which does not have solution for the $-$ sign (corresponding to $\theta_2 = -\frac{\pi}{2}$), but it does for the $+$ sign (corresponding to $\theta_2 = \frac{\pi}{2}$). Namely, there is only one critical angle $\theta_1 = \theta_c^+ = \arcsin \frac{1}{2} = \frac{\pi}{6}$ and thus, $\theta_3 = \pi - \theta_1 = \frac{5\pi}{6}$. We can now compute the intersection points with η to see if the three-segment reflection exists. Since we can completely determine γ_1 and γ_3 (we know q_1, q_2 and θ_1, θ_3), it is easy to show that

$$\gamma_1 \cap \eta = \left(0, -y_0 + \frac{x_0}{\sqrt{3}}\right), \quad \gamma_3 \cap \eta = \left(0, y_0 - \frac{x_0}{\sqrt{3}}\right).$$

We conclude (see Figure 3):

- If $y_0 > \frac{x_0}{\sqrt{3}}$, then the three-segment curve exists and it is the global minimum of $T_{\mathcal{N}^*}$.
- If $y_0 = \frac{x_0}{\sqrt{3}}$, then γ_2 reduces to the point $(0, 0)$ and γ_3 becomes the reflection of γ_1 corresponding to the critical angle.
- If $y_0 < \frac{x_0}{\sqrt{3}}$, then you cannot reach q_2 through a three-segment curve with the given angles $\theta_1, \theta_2, \theta_3$. In this case, the global minimum of $T_{\mathcal{N}^*}$ is the usual reflected trajectory, with an angle of incidence smaller than the critical angle.

Finally, we can find the trajectory that globally minimizes the traveltime. This can be done simply by comparing the traveltimes of the straight line ξ and the three-segment reflection γ in the case $y_0 > \frac{x_0}{\sqrt{3}}$ (in any other case the straight line is always the fastest trajectory). Using the endpoints and direction of each segment, they can be parametrized as follows:

$$\begin{aligned} \xi(t) &= (-x_0, -y_0) + t(0, 2), & t \in [0, y_0], \\ \gamma_1(t) &= (-x_0, -y_0) + t(\sqrt{3}, 1), & t \in \left[0, \frac{x_0}{\sqrt{3}}\right], \\ \gamma_2(t) &= \left(0, -y_0 + \frac{x_0}{\sqrt{3}}\right) + t(0, 2), & t \in \left[0, y_0 - \frac{x_0}{\sqrt{3}}\right], \\ \gamma_3(t) &= \left(0, y_0 - \frac{x_0}{\sqrt{3}}\right) + t(-\sqrt{3}, 1), & t \in \left[0, \frac{x_0}{\sqrt{3}}\right] \end{aligned}$$

(note that this is not the time-parametrization). Therefore:

$$\begin{aligned} T[\xi] &= \int_0^{y_0} F_1(0, 2) dt = \frac{4}{3}y_0, \\ T[\gamma] &= \int_0^{\frac{x_0}{\sqrt{3}}} F_1(\sqrt{3}, 1) dt + \int_0^{y_0 - \frac{x_0}{\sqrt{3}}} F_2(0, 2) dt + \int_0^{\frac{x_0}{\sqrt{3}}} F_1(-\sqrt{3}, 1) dt = \\ &= \frac{4}{3} \left(\frac{x_0}{\sqrt{3}}(4 - \sqrt{3}) + 1 \right), \end{aligned}$$

and we conclude (see Figure 3):

- If $y_0 = \frac{x_0}{\sqrt{3}}(4 - \sqrt{3}) + 1$, then γ and ξ reach q_2 at the same time and both globally minimize the traveltime. In this case q_2 belongs to the cut locus of q_1 – see § 5.6 below.

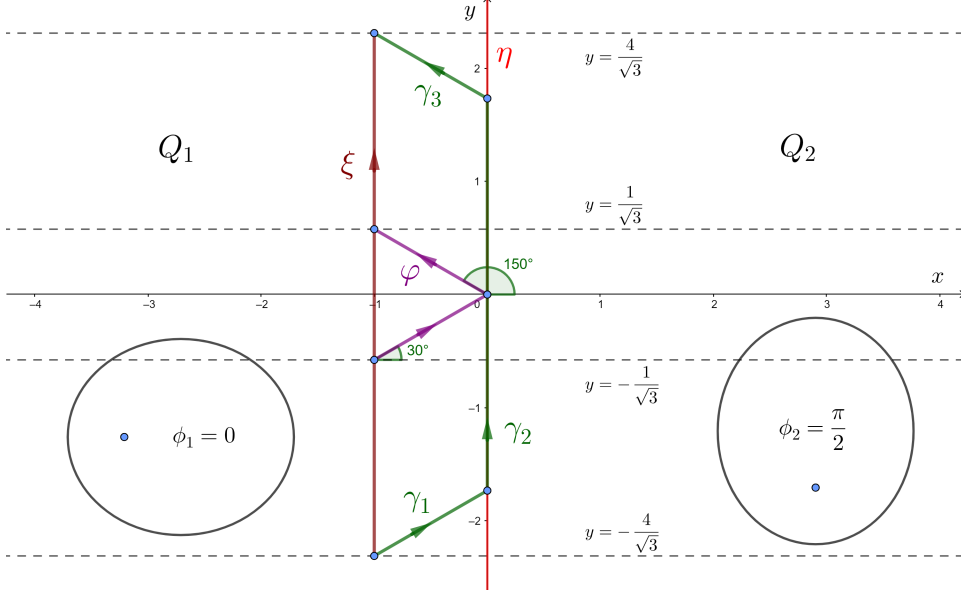


FIGURE 3. In this example, the velocity at each point is given by a focus-centered ellipse with orientation $\phi_1 = 0$ in Q_1 and $\phi_2 = \frac{\pi}{2}$ in Q_2 , which favors the travel along η . Fixing $x_0 = 1$ and always considering the travel from $q_1 = (-1, -y_0)$ to its symmetrical $q_2 = (-1, y_0)$, there are three different regions for y_0 : if $0 < y_0 < \frac{1}{\sqrt{3}}$, the fastest trajectory via η is the usual reflection passing through $(0, 0)$, with an angle of incidence smaller than the critical one; if $\frac{1}{\sqrt{3}} < y_0 < \frac{4}{\sqrt{3}}$, then the three-segment reflection becomes faster than the usual reflected trajectory, but still in these first two cases ξ is the fastest path; when $y_0 > \frac{4}{\sqrt{3}}$, the three-segment reflection becomes even faster than ξ . The limit cases are depicted explicitly: φ when $y_0 = \frac{1}{\sqrt{3}}$ and γ when $y_0 = \frac{4}{\sqrt{3}}$.

- If $y_0 > \frac{x_0}{\sqrt{3}}(4 - \sqrt{3}) + 1$, only γ globally minimizes the traveltime.
- If $y_0 < \frac{x_0}{\sqrt{3}}(4 - \sqrt{3}) + 1$, only ξ globally minimizes the traveltime.

5.5. Construction of the wavefront. Fixing a time $t_0 > 0$, consider all the time-parametrized trajectories γ that globally minimize the traveltime from $q_1 \in Q_1$ to $\gamma(t_0)$ (and thus, to any $\gamma(t)$ with $t \in (0, t_0]$). They provide the outermost points that can be reached by a wave spreading from q_1 during the time lapse t_0 , i.e., each one of these trajectories provide a point in the wavefront.

Throughout this section we have seen that only three types of curves can globally minimize the traveltime in the classical Finslerian setting: the straight lines in Q_1 , the refracted trajectories (crossing η once or remaining on it) and the three-segment reflected trajectories. Separately, they provide the *standard*, *refracted* and *reflected wavefronts*, respectively, which compete against each other to generate the actual wavefront.

Let τ_η and τ^\pm be the times when the standard wavefront and the critical trajectories reach η , respectively (τ_η coincides with the F_1 -distance from q_1 to η). The refracted wavefront only comes into existence from τ_η , and the reflected ones from τ^\pm .

Lemma 5.7. *Fixing any $q_1 \in Q_1$ and a time $t_0 > \tau^\pm$:*

- (i) *The standard wavefront $\tilde{\Gamma}_{t_0}$ at $t = t_0$ is given by the indicatrix of F_1 scaled up by t_0 , which is equivalent to*

$$\tilde{\Gamma}_{t_0} = \{p \in \mathbb{R}^2 : F_1(p - q_1) = t_0\}, \quad (17)$$

and $\tilde{\Gamma}_{t_0}$ intersects η at two points.

- (ii) *There exists one reflected wavefront $\hat{\Gamma}_{t_0}^\pm$ at $t = t_0$ for each critical angle θ_c^\pm . If φ^\pm is the usual reflection at θ_c^\pm and γ^\pm is the corresponding critical trajectory along η (both time-parametrized), then $\hat{\Gamma}_{t_0}^\pm$ is the straight line joining $\gamma^\pm(t_0)$ and $\varphi^\pm(t_0)$.*

- (iii) *The intersection between $\tilde{\Gamma}_{t_0}$ and each $\hat{\Gamma}_{t_0}^\pm$ is a unique point, provided that $\hat{\Gamma}_{t_0}^\pm$ exists.*

Proof. For the claim (i), if we consider F_1 -unit speed straight lines ξ , note that they travel a distance $t_0 \|\xi'(t)\|_g$ with $\|\xi'(t)\|_g$ constant and $F_1(\xi'(t)) = 1$, so $t_0 \Sigma_1$ provides the distance traveled by ξ in all directions, being Σ_1 the indicatrix of F_1 . Centered at q_1 , $t_0 \Sigma_1$ gives the points (17). Also, as $t_0 > \tau^\pm \geq \tau_\eta$, the strong convexity of Σ_1 ensures that $\tilde{\Gamma}_{t_0}$ intersects η at exactly two points.

We will prove (ii) and (iii) for $\hat{\Gamma}_{t_0}^+$, being the proof for $\hat{\Gamma}_{t_0}^-$ completely analogous. For (ii), let ρ be any (time-parametrized) three-segment reflected trajectory with angle of incidence $\theta_1 = \theta_c^+$. Assume ρ reaches η at time $\tau_1 > 0$ (the same as φ^+ and γ^+) and leaves it at τ_2 , with $\tau_1 < \tau_2 < t_0$. Splitting the curves in their different segments, note that $\rho_1 = \varphi_1^+ = \gamma_1^+$, ρ_2 is part of γ_2^+ and ρ_3 is parallel to φ_2^+ . Thus:

$$\frac{\|\rho_3'(t)\|_g(t_0 - \tau_2)}{\|(\gamma_2^+)'(t)\|_g(t_0 - \tau_2)} = \frac{\|\rho_3'(t)\|_g}{\|(\gamma_2^+)'(t)\|_g} = \frac{\|(\varphi_2^+)'(t)\|_g}{\|(\gamma_2^+)'(t)\|_g} = \frac{\|(\varphi_2^+)'(t)\|_g(t_0 - \tau_1)}{\|(\gamma_2^+)'(t)\|_g(t_0 - \tau_1)},$$

which means, by Thales' theorem, that $\rho(t_0)$ is in the segment $\hat{\Gamma}_{t_0}^+$ joining $\gamma^+(t_0)$ and $\varphi^+(t_0)$. This also ensures that given any point $p \in \hat{\Gamma}_{t_0}^+$, there exists a three-segment reflected trajectory from q_1 to p . So $\hat{\Gamma}_{t_0}^+$ is the reflected wavefront.

For the claim (iii), note that γ^+ always globally minimizes the traveltime by Proposition 5.6, but φ^+ never does. In particular, this means that

$$F_1(\varphi^+(t_0) - q_1) < t_0 < F_1(\gamma^+(t_0) - q_1),$$

so $\varphi^+(t_0)$ lies inside $\tilde{\Gamma}_{t_0}$, whereas $\gamma^+(t_0)$ is outside. Given the fact that $\hat{\Gamma}_{t_0}^+$ is the straight line joining both points and $\tilde{\Gamma}_{t_0}$ is strongly convex, there is a unique intersection point. \square

Note that $\tilde{\Gamma}_{t_0}$ and $\hat{\Gamma}_{t_0}^\pm$ can be explicitly parametrized as (smooth) curves in \mathbb{R}^2 as follows:

$$\begin{aligned}\tilde{\Gamma}_{t_0}(\theta) &= t_0 V_1(\theta)(\cos \theta, \sin \theta) + q_1, & \theta &\in (-\pi, \pi], \\ \hat{\Gamma}_{t_0}^\pm(s) &= \gamma^\pm(t_0) + s(\varphi^\pm(t_0) - \gamma^\pm(t_0)), & s &\in [0, 1],\end{aligned}\tag{18}$$

and Lemma 5.7 ensures there are exactly two angles θ_η^\pm such that $\tilde{\Gamma}_{t_0}(\theta_\eta^\pm) \in \eta$ and, if θ_c^\pm exists, we can find s_0^\pm and θ_0^\pm such that $\tilde{\Gamma}_{t_0}(\theta_0^\pm) = \hat{\Gamma}_{t_0}(s_0^\pm)$. Note that θ_η^\pm , s_0^\pm and θ_0^\pm depend on t_0 .

Theorem 5.8. *Assume for simplicity that θ_c^+ exists but θ_c^- does not (analogous results are obtained in any other case). Then the actual wavefront Γ_{t_0} from $q_1 \in Q_1$ at $t = t_0$ is a piecewise smooth closed curve composed of the following parts:*

- If $t_0 \in [0, \tau_\eta]$, then Γ_{t_0} is the whole standard wavefront at $t = t_0$ (see Figure 4(a)).
- If $t_0 \in (\tau_\eta, \tau^+]$, then $\Gamma_{t_0} = \Gamma_1 \cup \Gamma_2$, where

$$\Gamma_1 = \{\tilde{\Gamma}_{t_0}(\theta) \in \mathbb{R}^2 : \theta \in (-\pi, \theta_\eta^-] \cup [\theta_\eta^+, \pi]\}$$

and Γ_2 is the whole refracted wavefront at $t = t_0$ (see Figure 4(b)).

- If $t_0 > \tau^+$, then $\Gamma_{t_0} = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$, where

$$\Gamma_1 := \{\tilde{\Gamma}_{t_0}(\theta) \in \mathbb{R}^2 : \theta \in (-\pi, \theta_\eta^-] \cup [\theta_\eta^+, \pi]\},$$

$$\Gamma_3 := \{\hat{\Gamma}_{t_0}^+(s) \in \mathbb{R}^2 : s \in [0, s_0^+]\}$$

and Γ_2 is the whole refracted wavefront at $t = t_0$ (see Figures 4(c) and 4(d)).

Proof. Note first that Γ_1 and Γ_3 are always smooth curves by (18) and so is Γ_2 by Theorem 5.2 and the smoothness of the raypath parameter (14). Now, if $t_0 \in [0, \tau_\eta]$, only the standard wavefront exists. If $t_0 \in (\tau_\eta, \tau^+]$, the reflected wavefront still does not exist, so the part of the standard wavefront Γ_1 in Q_1 globally minimizes the traveltime, and so does the whole refracted wavefront Γ_2 (recall the discussion in § 5.3). Also, the endpoints of Γ_1 and Γ_2 coincide, which ensures that Γ_{t_0} is a piecewise smooth closed curve. Finally, if $t_0 > \tau^+$, it is clear from § 5.3 and Lemma 5.7 that each part $\Gamma_1, \Gamma_2, \Gamma_3$ globally minimizes the traveltime. In this case, the endpoints of Γ_1 are $\Gamma_2(\theta_\eta^-)$ and $\Gamma_3(0)$, and $\Gamma_2(\theta_0^+) = \Gamma_3(s_0^+)$. So Γ_{t_0} is still a piecewise smooth closed curve. \square

5.6. The cut locus. Fixing any $q \in Q$, let us define the following curves:

$$\begin{aligned}\sigma^\pm: (\tau^\pm, \infty) &\longrightarrow \mathbb{R}^2 \\ t &\longmapsto \sigma^\pm(t) = \tilde{\Gamma}_t \cap \hat{\Gamma}_t^\pm,\end{aligned}$$

where $\tilde{\Gamma}_t, \hat{\Gamma}_t^\pm$ are the standard and reflected wavefronts from q at time t and τ^\pm is the time at which the critical trajectory γ^\pm reaches η . Provided that the critical angles exist, note that each σ^\pm is well defined, as τ^\pm is the time from which the reflected wavefront begins to exist and $\tilde{\Gamma}_t \cap \hat{\Gamma}_t^\pm$ is just one point by Lemma 5.7 (see Figures 4(c) and 4(d)).

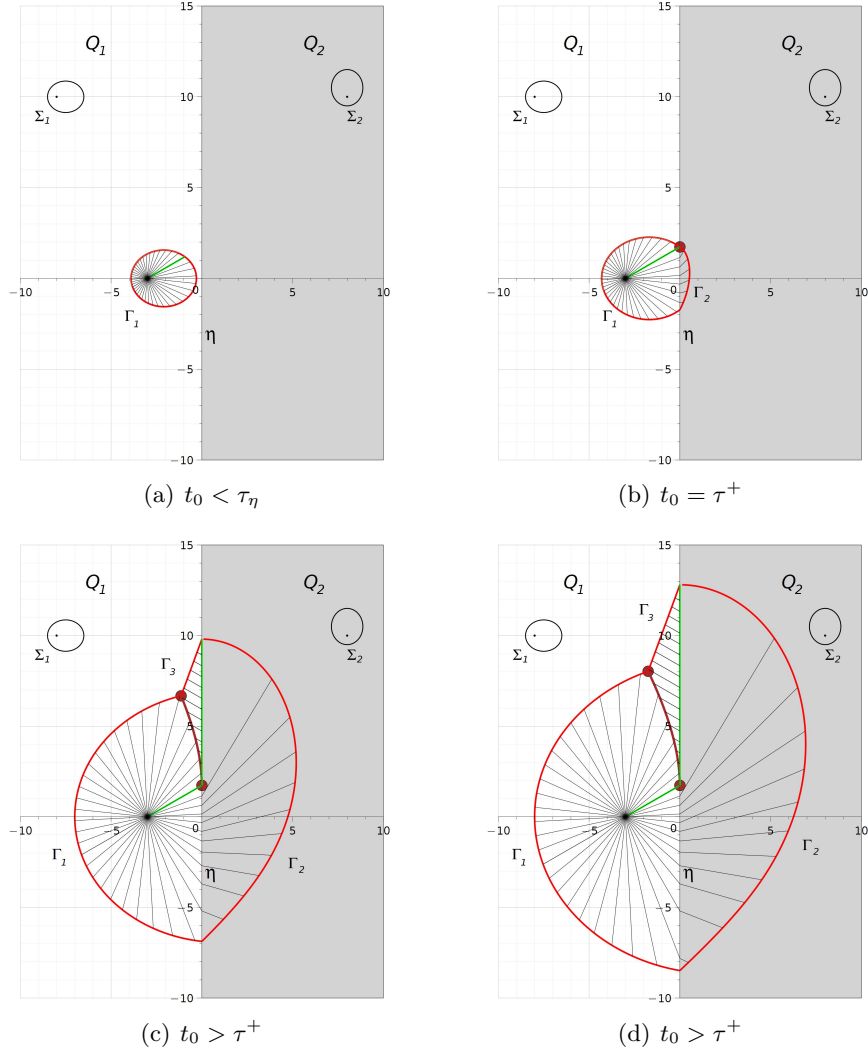


FIGURE 4. Evolution of the wavefront over time. The cut locus in (c) and (d) given by the intersection of Γ_1 and Γ_3 for $t \in (\tau^+, t_0]$ develops into Q_1 from the point of first encounter of the expanding ellipse with η in the critical direction, as shown in (b). The Finsler metrics are the same as in § 5.4.

By definition, each point of these curves is reached at the same time by two different globally minimizing trajectories from q , constituting what is known as the *cut locus* of q :

$$\mathcal{C}_q := \text{Im}(\sigma^+) \cup \text{Im}(\sigma^-).$$

Proposition 5.9. Fix $q \in Q$ and let \mathcal{M}_q be the set of all piecewise smooth curves $\gamma : [0, \infty) \rightarrow Q$ (and thus, future inextendible) with $\gamma(0) = q$. If we define the map $t_q : \mathcal{M}_q \rightarrow (0, \infty]$ by

$$t_q(\gamma) := \sup\{t \in \mathbb{R} : \gamma|_{[0,t]} \text{ globally minimizes the traveltime}\},$$

then the cut locus of q is

$$\mathcal{C}_q = \{\gamma(t_q(\gamma)) \in \mathbb{R}^2 : \gamma \in \mathcal{M}_q, 0 < t_q(\gamma) < \infty\}. \quad (19)$$

Proof. Let $p = \tilde{\Gamma}_{t_0} \cap \hat{\Gamma}_{t_0}^\pm \in \mathcal{C}_q$ for any $t_0 > \tau^\pm$. There exist two (time-parametrized) curves $\xi, \gamma \in \mathcal{M}_q$ such that $\xi(t_0) = \gamma(t_0) = p$ and $\xi|_{[0,t_0]}, \gamma|_{[0,t_0]}$ globally minimize the traveltime. Suppose $\gamma|_{[0,t_0+\delta]}$ is still globally time-minimizing for some $\delta > 0$. Then the concatenation of $\xi|_{[0,t_0]}$ with $\gamma|_{[t_0,t_0+\delta]}$ also globally minimizes the traveltime. But if we concatenate $\xi|_{[0,t_0-\delta]}$ with the straight line joining $\xi(t_0 - \delta)$ and $\gamma(t_0 + \delta)$ we reduce the traveltime by the triangle inequality (note that γ and ξ cannot be parallel), which is a contradiction. So $t_0 = t_q(\gamma)$ and p is in the set (19).

Conversely, let $p = \gamma(t_q(\gamma))$ for some $\gamma \in \mathcal{M}_q$ with $0 < t_q(\gamma) < \infty$. Since $\gamma|_{[0,t_q(\gamma)]}$ globally minimizes the traveltime, γ is either a refracted trajectory, a straight line or a three-segment reflected trajectory. Now, if it were a refracted trajectory, then $t_q(\gamma) = \infty$, which is a contradiction, so assume without loss of generality that $\gamma(t_q(\gamma)) \in \hat{\Gamma}_{t_q(\gamma)}^+$. For each $\delta > 0$, the global time-minimizing curve from q to $\gamma(t_q(\gamma) + \delta)$ must be the straight line joining both points (it cannot be a three-segment reflected trajectory because its last segment would not intersect γ) and in the limit $\delta \rightarrow 0$, we find the (time-parametrized) straight line ξ from q to $\gamma(t_q(\gamma))$ satisfying $\xi(t_q(\gamma)) \in \tilde{\Gamma}_{t_q(\gamma)}$. So $p = \gamma(t_q(\gamma)) \in \tilde{\Gamma}_{t_q(\gamma)} \cap \hat{\Gamma}_{t_q(\gamma)}^+ \in \mathcal{C}_q$. \square

Essentially, any wavefront trajectory crossing one of the curves σ^\pm loses the property of being globally time-minimizing, no longer providing points in the wavefront afterwards. In practice, σ^\pm represent the merging regions of two wavefronts coming from different directions, becoming extremely relevant in some situations. When dealing with wildfires, for instance, these curves are danger zones for the firefighters, as they can easily become trapped by the fire and the convergence of several fire trajectories leads to the corresponding increase in the heat intensity (see also the discussion in [16]).

ACKNOWLEDGMENTS

The authors warmly thank the anonymous referee, whose comments helped improve the paper.

This work is the result of a research stay of EPR at DTU Compute, partially supported by The International Doctoral School of the University of Murcia (EIDUM) and by Ayudas para la Formación de Profesorado Universitario (FPU) from the Spanish Government. EPR was supported by the project PID2021-124157NB-I00, funded by MCIN/AEI/10.13039/501100011033/ and by "ERDF A way of making Europe", and also by Ayudas a proyectos para el desarrollo de investigación científica y técnica por grupos competitivos (Comunidad Autónoma de la Región de Murcia), included in the Programa Regional de Fomento de la Investigación Científica y Técnica (Plan de Actuación 2022) of the Fundación Séneca-Agencia de Ciencia y Tecnología de la Región de Murcia, REF. 21899/PI/22.

REFERENCES

- [1] P.L. ANTONELLI, A. BÓNA AND M.A. SLAWIŃSKI. Seismic rays as Finsler geodesics. *Nonlinear Anal. RWA* 4 (5), 711–722 (2003).
- [2] D. BAO, S.-S. CHERN AND Z. SHEN. *An Introduction to Riemann-Finsler geometry*. In: Graduate Texts in Mathematics, vol. 200, Springer-Verlag, New York, 2000.
- [3] A. BÓNA AND M.A. SLAWIŃSKI. Raypaths as parametric curves in anisotropic, nonuniform media: differential-geometry approach *Nonlinear Analysis* 51, 983–994 (2002).
- [4] E. CAPONIO, M.A. JAVALOYES AND A. MASIELLO. On the energy functional on Finsler manifolds and applications to stationary spacetimes. *Mathematische Annalen* 351, 365–392 (2011).
- [5] E. CAPONIO, M.A. JAVALOYES AND M. SÁNCHEZ. On the interplay between Lorentzian causality and Finsler metrics of Randers type. *Rev. Mat. Iberoam.* 27 (3), 919–952 (2011).
- [6] E. CAPONIO, M.A. JAVALOYES AND M. SÁNCHEZ. Wind Finslerian structures: from Zermelo's navigation to the causality of spacetimes. *ArXiv e-prints*, arXiv:1407.5494 [math.DG] (2014). To appear in *Memoirs of AMS*.
- [7] H.R. DEHKORDI. Applications of Randers geodesics for wildfire spread modelling. *Applied Mathematical Modelling* 106, 45–59 (2022).
- [8] H.R. DEHKORDI AND A. SAA. Huygens' envelope principle in Finsler spaces and analogue gravity. *Classical Quantum Gravity* 36 (8), 085008 (2019).
- [9] R.C. FETEAU, J.E. MARSDEN, M. ORTIZ AND M. WEST. Nonsmooth Lagrangian Mechanics and Variational Collision Integrators. *SIAM J. Applied Dynamical Systems* 2 (3), 381–416 (2019).
- [10] M.A. FINNEY. *FARSITE: Fire Area Simulator-model development and evaluation*. USDA Forest Service, Res. Pap. RMRS-RP-4, Rocky Mountain Research Station, Ogden, UT, 1998 (revised 2004).
- [11] G. W. GIBBONS. A Spacetime Geometry picture of Forest Fire Spreading and of Quantum Navigation. *ArXiv e-prints*, arXiv:1708.02777 [gr-qc] (2017).
- [12] G.W. GIBBONS, C.A.R. HERDEIRO, C.M. WARNICK AND M.C. WERNER. Stationary metrics and optical Zermelo-Randers-Finsler geometry. *Phys. Rev. D* 79, 044022 (2009).
- [13] G.W. GIBBONS AND C.M. WARNICK. The geometry of sound rays in a wind. *Contemp. Phys.* 52 (3), 197–209 (2011).
- [14] M.A. JAVALOYES AND E. PENDÁS-RECONDO. Lightlike hypersurfaces and time-minimizing geodesics in cone structures. *ArXiv e-prints*, arXiv:2109.07969 [math.DG] (2021). To appear in: A.L. Albuje et al. (eds.), *Developments in Lorentzian Geometry*, Springer Proceedings in Mathematics & Statistics 389.
- [15] M.A. JAVALOYES, E. PENDÁS-RECONDO AND M. SÁNCHEZ. Applications of cone structures to the anisotropic rheonomic Huygens' principle. *Nonlinear Analysis* 209, 112337 (2021).
- [16] M.A. JAVALOYES, E. PENDÁS-RECONDO AND M. SÁNCHEZ. A general model for wildfire propagation with wind and slope. *ArXiv e-prints*, arXiv:2110.03364 [math.DG] (2021).
- [17] M.A. JAVALOYES AND M. SÁNCHEZ. On the definition and examples of Finsler metrics. *Ann. Sc. Norm. Super. Pisa Cl. Sci.* 13 (3), 813–858 (2014).
- [18] M.A. JAVALOYES AND M. SÁNCHEZ. On the definition and examples of cones and Finsler spacetimes. *RACSAM* 114, 30 (2020).
- [19] S. MARKVORSEN. A Finsler geodesic spray paradigm for wildfire spread modelling. *Nonlinear Anal. RWA* 28, 208–228 (2016).
- [20] S. MARKVORSEN. Geodesic sprays and frozen metrics in rheonomic Lagrange manifolds. *ArXiv e-prints*, arXiv:1708.07350 [math.DG] (2017).
- [21] T. YAJIMA AND H. NAGAHAMA. Finsler geometry of seismic ray path in anisotropic media. *Proc. R. Soc. A* 469, 1763–1777 (2009).

DTU COMPUTE, MATHEMATICS,
TECHNICAL UNIVERSITY OF DENMARK,
DK-2800 KGS. LYNGBY, DENMARK.
Email address: `stema@dtu.dk`

DEPARTAMENTO DE MATEMÁTICAS,
UNIVERSIDAD DE MURCIA,
CAMPUS DE ESPINARDO,
30100 ESPINARDO, MURCIA, SPAIN.
Email address: `e.pendasrecondo@um.es`