Nonlinear Analysis 157 (2017) 123–145

Contents lists available at ScienceDirect

Nonlinear Analysis

www.elsevier.com/locate/na

On the numerical solution of the far field refractor problem

Roberto De Leo[a], Cristian E. Gutiérrez[b,*], Henok Mawi[a]

[a] Department of Mathematics, Howard University, Washington, D.C. 20059, United States
[b] Department of Mathematics, Temple University, Philadelphia, PA 19122, United States

ARTICLE INFO

Article history:
Received 1 March 2017
Accepted 18 March 2017
Communicated by Enzo Mitidieri

Keywords:
Geometric pdes and optics
Lipschitz estimates
Numerical analysis

ABSTRACT

The far field refractor problem with a discrete target is solved with a numerical scheme that uses and simplify ideas from Caffarelli et al. (1999). A numerical implementation is carried out and examples are shown.

1. Introduction

The purpose of this paper is to present an algorithm to construct far field one source refractors with arbitrary precision. We use the ideas from the paper [4] by Caffarelli, Kochengin and Oliker, where they develop an algorithm to construct far field point source global reflectors, i.e., the source domain Ω is the whole sphere S^2, and the density is smooth. For our refraction problem, we are able to simplify and extend these ideas to deal with densities that are only bounded and work in general domains. In particular, we do not need to consider derivatives of the refractor measure, we only need to prove an appropriate Lipschitz bound for the refractor measure which considerably simplifies the approach proposed in [4]. In addition, our approach does not use the mass transport structure of the far field problem, and therefore it can be used in near field problems. Since we are working in general domains Ω and with a non smooth density, the differentiability of the refractor measure might not hold in general. This depends on the shape and regularity of the domain and the smoothness of the density. The nature of refraction problems demands for domains for which total internal reflection does not occur, see condition (2.5). Therefore the global problem does not make sense in this case.

To place our results in perspective we mention the following. Recently, Castro, Mérigot and Thibert [5] introduced numerical methods to solve the reflector problem. These are based on optimal transport ideas

* Corresponding author.
E-mail addresses: roberto.deleo@howard.edu (R. De Leo), gutierre@temple.edu (C.E. Gutiérrez), henok.mawi@howard.edu (H. Mawi).

http://dx.doi.org/10.1016/j.na.2017.03.009

introducing a concave function arising from the Kantorovich functional. This function is analyzed numerically and their results follow, combined with other numerical packages. An advantage of this approach is that the convergence of their algorithm is faster than the one proposed in [4]. For general cost functions satisfying the Ma, Trudinger and Wang condition arising in optimal transport [13], the algorithm in [4] is extended in [12] when the density is C^∞ and the domains are convex with respect to the cost function. We remark that this does not include our results when the density is smooth, since the refractor considered in the present paper is for $\kappa < 1$ and the condition of Ma, Trudinger and Wang does not hold in this case; see [7, Section 5]. We believe the case $\kappa < 1$ is more interesting for lens design since lenses are made of materials that are denser than the surrounding medium. In fact, if the material around the source is cut out with sphere centered at the source, then the lens sandwiched between that sphere and the constructed refractor surface will perform the desired refracting job.

The far field refractor problem has been considered and solved for the first time in [7] using optimal mass transport. Several models and variants have been introduced to reflect more accurately the physical features of the problem; see [8,10,11], and [9]. For numerical results to design reflectors solving Monge–Ampère type pdes we refer to [3,2] both containing many references.

The organization of the paper is as follows. In Section 2, we explain the set up and the problem solved. In Section 3.1 we prove lemmas concerning the tracing map and the refractor measure to be used in solving the problem. Section 3.2 contains a few results about geodesic disks that are needed in the proof of the Lipschitz estimates. The algorithm is explained in detail in Section 4, and the convergence in a finite number of steps in Section 4.3. Section 5 contains the Lipschitz estimate in Proposition 5.1 needed to show the convergence of the algorithm in a finite number of steps. Finally, in Section 6 we give a numerical implementation of our algorithm to construct various examples.

2. Set up, definitions, and statement of results

Suppose Γ is a surface in $\mathbb{R}^3$ that separates two homogeneous, isotropic and dielectric media I and II having refractive indices n_1 and n_2, respectively. If a ray of light having direction $x \in S^2$, the unit sphere in $\mathbb{R}^3$, and traveling through the medium I strikes Γ at the point P, then this ray is refracted in the direction $m \in S^2$ through the medium II according to the law of refraction (Snell's Law)

$$n_1(x \times \nu) = n_2(m \times \nu), \tag{2.1}$$

where ν is the unit normal to Γ at P pointing towards medium II. If we set $\kappa = n_2/n_1$, then we can also write (2.1) as

$$x - \kappa m = \lambda \nu \tag{2.2}$$

where $\lambda \in \mathbb{R}$ is given by $\lambda = x \cdot \nu - m \cdot \nu = x \cdot \nu - \kappa\sqrt{1 - \kappa^{-2}(1 - (x \cdot \nu)^2)}$. When medium I is optically denser than medium II, that is, $\kappa < 1$, the vector m bends away from the normal, and total internal reflection might occur. That is, the ray with direction m is transmitted to medium II if and only if $x \cdot m \geq \kappa$, or equivalently $x \cdot \nu \geq \sqrt{1 - \kappa^2}$; see [7, Section 2.1].

When $\kappa < 1$, the surfaces having the uniform refracting property, are ellipsoids of revolution having a focus at the origin, see [7, Section 2.2]. That is, the surface written in polar coordinates $\rho(x)x$ with $x \in S^2$ and with

$$\rho(x) = \frac{b}{1 - \kappa m \cdot x}, \tag{2.3}$$

$b > 0$, is an ellipsoid of revolution with axis m, eccentricity κ, foci 0 and $\frac{2\kappa b}{1-\kappa^2}m$, and refracts all rays emanating from 0 into the direction m for $x \cdot m \geq \kappa$. We then denote this semi-ellipsoid by

$$E(m,b) = \left\{ \rho(x)x : \rho(x) = \frac{b}{1-\kappa m \cdot x},\ x \in S^2,\ m \cdot x \geq \kappa \right\}. \tag{2.4}$$

We assume throughout the paper that medium I is denser than medium II and therefore $\kappa = n_2/n_1 < 1$. We also point out that similar analysis can be done for the case $\kappa = n_2/n_1 > 1$, changing ellipsoids for hyperboloids, see [7, Section 2.2].

Suppose that Ω and Ω^* are two domains of the unit sphere S^2 of $\mathbb{R}^3$[1] with the property, to avoid total reflection [1, Sect. 1.5.4], that

$$\inf_{m \in \bar{\Omega}^*, x \in \bar{\Omega}} m \cdot x \geq \kappa, \tag{2.5}$$

where $m \cdot x$ is the usual inner product of m and x in $\mathbb{R}^3$; and the boundary of Ω has surface measure zero.

Definition 2.1. A surface $\mathcal{R}$ in $\mathbb{R}^3$ parameterized by $\rho(x)x$ is a refractor from $\bar{\Omega}$ to $\bar{\Omega}^*$ if for any $x_o \in \bar{\Omega}$ there exists a semi-ellipsoid $E(m,b)$ with $m \in \bar{\Omega}^*$ such that $\rho(x_o) = \frac{b}{1-\kappa m \cdot x_o}$ and $\rho(x) \leq \frac{b}{1-\kappa m \cdot x}$ for all $x \in \bar{\Omega}$. We call $E(m,b)$ a supporting semi-ellipsoid to $\mathcal{R}$ at $\rho(x_o)x_o$ or simply at x_o.

From the definition, it is easy to see that refractors are Lipschitz continuous in $\bar{\Omega}$, i.e., $|\rho(x) - \rho(y)| \leq C_\kappa\ (\inf_\Omega \rho)\ |x-y|$ for $x, y, \in \bar{\Omega}$ with C_κ a constant depending only on κ.

Definition 2.2. Given a refractor $\mathcal{R} = \{\rho(x)x : x \in \bar{\Omega}\}$, the refractor mapping of $\mathcal{R}$ is the multi-valued map defined for $x_o \in \bar{\Omega}$ by

$$\mathcal{N}_\mathcal{R}(x_o) = \{m \in \bar{\Omega}^* : E(m,b) \text{ supports } \mathcal{R}\ at \rho(x_o)x_o \text{ for some } b > 0\}.$$

Given $m_o \in \bar{\Omega}^*$ the tracing mapping of $\mathcal{R}$ is defined by

$$\mathcal{T}_\mathcal{R}(m_o) = \{x \in \bar{\Omega} : m_o \in \mathcal{N}_\mathcal{R}(x)\}.$$

Suppose that we are given a nonnegative function $g \in L^1(\bar{\Omega})$. We then recall the notion of refractor measure, see [7, Section 3.1].

Definition 2.3. The refractor measure associated with the refractor $\mathcal{R}$ and the function g is the Borel measure given by

$$G_\mathcal{R}(\omega) = \int_{\mathcal{T}_\mathcal{R}(\omega)} g(x)\, dx$$

for every Borel subset ω of $\bar{\Omega}^*$.

Given a Borel measure μ in Ω^* satisfying the energy conservation condition $\int_\Omega g(x)\, dx = \mu(\Omega^*)$, the *far field refractor problem* consists in finding a refractor $\mathcal{R}$ from Ω to Ω^* such that $G_\mathcal{R} = \mu$ in Ω^*. Existence of refractors and uniqueness up to dilations is proved in [7] using mass transport techniques. This is also proved in [8] with a different method where a more general case that takes into account internal reflection is considered.

For the remaining part of the discussion fix $m_1, m_2, \ldots, m_N$, $N \geq 2$, distinct points in $\bar{\Omega}^* \subset S^2$. Given $\mathbf{b} = (b_1, \ldots, b_N) \in \mathbb{R}^N_+$, i.e., with each $b_i > 0$, we denote by $\mathcal{R}(\mathbf{b})$ the refractor defined by a finite number of semi-ellipsoids and given by

$$\mathcal{R}(\mathbf{b}) = \left\{ \rho(x)x : x \in \bar{\Omega}, \rho(x) = \min_{1 \leq i \leq N} \frac{b_i}{1-\kappa m_i \cdot x} \right\}. \tag{2.6}$$

[1] The physical problem considered is three dimensional; the mathematical extension to n dimensions is straightforward.

In this setting, we recall the following theorem from [8, Remark 6.10] for discrete targets.

Theorem 2.4. *Let $g \in L^1(\bar{\Omega})$ with $g > 0$ a.e., $f_1, \dots, f_N$ are positive numbers, $m_1, \dots, m_N \in S^2$ are distinct points with $x \cdot m_j \geq \kappa$ for all $x \in \Omega$ and $1 \leq j \leq N$. Assume the energy conservation condition*

$$\int_{\bar{\Omega}} g(x)\, dx = f_1 + \cdots + f_N. \tag{2.7}$$

Then there exists a refractor unique up to dilations,[2] *having the form (2.6), and solving $G_{\mathcal{R}(\mathbf{b})}(m_i) = f_i$ for all $i = 1, \dots, N$.*

The main result of this paper is to describe an iterative scheme to construct this refractor with arbitrary precision. That is, given $g \in L^\infty$ nonnegative, and $f_1, \dots, f_N; m_1, \dots, m_N$, as in Theorem 2.4, and $\epsilon > 0$ we find a vector $\mathbf{b} \in \mathbb{R}^N_+$, which depends on ϵ, such that the refractor $\mathcal{R}(\mathbf{b})$ of the form (2.6) satisfies

$$|G_{\mathcal{R}(\mathbf{b})}(m_i) - f_i| \leq \epsilon, \quad 1 \leq i \leq N. \tag{2.8}$$

3. Preliminary results

The purpose of this section is to present results on the refractor map and measure that will be used in the remaining part of the paper. Geodesic disks are also introduced and will be used in the proof of the Lipschitz estimate in Proposition 5.1.

3.1. Lemmas for the tracing map and refractor measures

Lemma 3.1. *Let $\mathbf{b} = (b_1, \dots, b_N) \in \mathbb{R}^N$ with each $b_i > 0$. Consider the family of refractors obtained from $\mathcal{R}(\mathbf{b}) = \{\rho(x)x : x \in \Omega\}$, by changing only b_i and fixing b_j for all $j \neq i$. Then:*

i. *$G_{\mathcal{R}(\mathbf{b})}(m_i) = 0$ for $b_i > (1+\kappa)\min_{j\neq i} b_j$.*
ii. *$G_{\mathcal{R}(\mathbf{b})}(m_i) = \int_\Omega g(x)\, dx$ for $0 < b_i < \frac{\min_{j\neq i} b_j}{1+\kappa}$.*

Proof. To prove (i) suppose $x \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)$, and x is not a singular point of $\mathcal{R}(\mathbf{b})$. Then $E(m_i, b_i)$ is a supporting semi-ellipsoid to $\mathcal{R}(\mathbf{b})$ at $\rho(x)x$. So we have

$$\frac{b_i}{1 - \kappa m_i \cdot x} \leq \frac{b_j}{1 - \kappa m_j \cdot x}$$

for all $j = 1, \dots, N$. Therefore

$$b_i \leq \frac{1 - \kappa m_i \cdot x}{1 - \kappa m_j \cdot x} b_j \leq \frac{1 - \kappa^2}{1 - \kappa} b_j = (1+\kappa) b_j, \quad j = 1, \dots, N.$$

Hence if $b_i > (1+\kappa)\min_{j\neq i} b_j$, then $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) \subset S$, where S is the singular set of $\mathcal{R}(\mathbf{b})$. The first part of the lemma is then proved.

Let us prove (ii). Let $b_0 = \min_{j\neq i} b_j$, and take $0 < b_i < b_0/(1+\kappa)$. Then for any $x \in \Omega$ and for any $j \neq i$ we have

$$\frac{b_i}{1 - \kappa m_i \cdot x} < \frac{b_0/(1+\kappa)}{1 - \kappa m_i \cdot x} \leq \frac{b_o}{1 - \kappa^2} \leq \frac{b_j}{1 - \kappa^2} \leq \frac{b_j}{1 - \kappa m_j \cdot x}.$$

[2] The assumption $g > 0$ a.e. is only used to prove uniqueness up to dilations.

So for $0 < b_i < b_0/(1+\kappa)$ we obtain

$$\min_{1\le l\le N} \frac{b_l}{1-\kappa m_l\cdot x} = \frac{b_i}{1-\kappa m_i\cdot x}$$

and consequently $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) = \bar{\Omega}$ completing the proof of part (ii) of the lemma. □

Remark 3.2. For each fixed $1 \le i \le N$, from Lemma 3.1, the function $G_{\mathcal{R}(\mathbf{b})}(m_i)$ is constant on the set defined by linear inequalities

$$F_i := \bigcup_{j\ne i} \{\mathbf{b} = (b_1,\dots,b_N) : b_i \ge (1+\kappa)\, b_j\} \bigcup \bigcap_{j\ne i} \left\{\mathbf{b} = (b_1,\dots,b_N) : b_i \le \frac{1}{1+\kappa}\, b_j\right\}.$$

If we set $G_i(\mathbf{b}) = G_{\mathcal{R}(\mathbf{b})}(m_i), 1 \le i \le N$, and consider the map $\mathbf{b} = (b_1,\dots,b_N) \mapsto (G_1(\mathbf{b}),\dots,G_N(\mathbf{b}))$, the Jacobian of this map is zero on the set $\cup_{i=1}^N F_i$.

Lemma 3.3. *Let $\mathbf{b} = (b_1,\dots,b_N)$ and $\mathbf{b}^* = (b_1^*,\dots,b_N^*)$ be in $\mathbb{R}_+^N$. Suppose that for some $l, b_l^* \le b_l$ and for all $i \ne l, b_i^* = b_i$, where $1 \le l, i \le N$. Then*

$$\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l) \subseteq \mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_l) \tag{3.1}$$

and

$$\mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_i) \subseteq \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) \quad \text{for } i \ne l, \tag{3.2}$$

where the inclusions are up to a set of measure zero. Consequently

$$G_{\mathcal{R}(\mathbf{b})}(m_l) \le G_{\mathcal{R}(\mathbf{b}^*)}(m_l) \quad \text{and} \quad G_{\mathcal{R}(\mathbf{b})}(m_i) \ge G_{\mathcal{R}(\mathbf{b}^*)}(m_i) \quad \text{for } i \ne l.$$

Proof. We use here that if $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)$ and x_0 is not a singular point, then the ellipsoid $E(m_l, b_l)$ supports $\mathcal{R}(\mathbf{b})$ at x_0, this holds for any refractor $\mathcal{R}(\mathbf{b})$ and any $1 \le l \le N$; see [8, Lemma 5.1].[3]

We first prove (3.2) when x_0 is not a singular point of $\mathcal{R}(\mathbf{b}^*)$. Since $b_l^* \le b_l$, we obviously have $\rho^*(x) \le \rho(x)$ for all $x \in \Omega$, where ρ^* is the parametrization of $\mathcal{R}(\mathbf{b}^*)$ and ρ is the parametrization of $\mathcal{R}(\mathbf{b})$. Suppose $i \ne l$ and let $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_i)$. Then, since x_0 is not a singular point of $\mathcal{R}(\mathbf{b}^*)$, the ellipsoid with polar radius $\frac{b_i}{1-\kappa x\cdot m_i}$ supports $\mathcal{R}(\mathbf{b}^*)$ at x_0. We have $\rho(x) \le \frac{b_i}{1-\kappa x\cdot m_i}$. Therefore

$$\frac{b_i}{1-\kappa x_0\cdot m_i} = \rho^*(x_0) \le \rho(x_0) \le \frac{b_i}{1-\kappa x_0\cdot m_i},$$

that is, $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)$.

We now prove (3.1). That is, if x_0 is neither a singular point of $\mathcal{R}(\mathbf{b})$ nor a singular point of $\mathcal{R}(\mathbf{b}^*)$, and $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)$, then $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_l)$. We may assume $b_l^* < b_l$. We have that $E(m_l, b_l)$ supports $\mathcal{R}(\mathbf{b})$ at x_0. We claim that the ellipsoid with polar radius $\frac{b_l^*}{1-\kappa\, x\cdot m_l}$ supports $\mathcal{R}(\mathbf{b}^*)$ at x_0. Suppose this is not true. Since by definition $\rho^*(x) \le \frac{b_l^*}{1-\kappa\, x\cdot m_l}$, we would have $\rho^*(x_0) < \frac{b_l^*}{1-\kappa\, x_0\cdot m_l}$. So $\rho^*(x_0) = \frac{b_j}{1-\kappa\, x_0\cdot m_j}$ for some $j \ne l$, and therefore $\frac{b_j}{1-\kappa\, x\cdot m_j}$ supports $\mathcal{R}(\mathbf{b}^*)$ at x_0. Since x_0 is not a singular point of $\mathcal{R}(\mathbf{b}^*)$, then by the inclusion previously proved we get that $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_j)$. Since $j \ne l$ we obtain that x_0 is a singular point of $\mathcal{R}(\mathbf{b})$, a contradiction. □

[3] The restriction that x_0 is not a singular point cannot be disposed of. For example, consider a refractor $\mathcal{R}$ that is the min of only two semi-ellipsoids $E(m_2, b_2)$ and $E(m_3, b_3)$. Take a singular point x_0 of this refractor and consider a supporting semi-ellipsoid $E(m_1, b)$ at x_0 having another direction m_1. Take now $E(m_1, b_1)$ with $b_1 > b$. The refractor can be defined with the three ellipsoids $E(m_i, b_i), 1 \le i \le 3$, because the definition of refractor does not see $E(m_1, b_1)$, but $E(m_1, b)$ is supporting at x_0 and $b < b_1$ and $E(m_1, b_1)$ does not support $\mathcal{R}$ at x_0.

Remark 3.4. We show that if Ω is connected, $0 < |\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)| < |\Omega|$, and $b_\ell^* < b_\ell$, then $|\mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_l) \setminus \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)| > 0$. Therefore, if $g > 0$ a.e., then this implies that if $0 < G_{\mathcal{R}(\mathbf{b})}(m_l) < \int_\Omega g(x)\,dx$ we obtain $G_{\mathcal{R}(\mathbf{b})}(m_l) < G_{\mathcal{R}(\mathbf{b}^*)}(m_l)$ when $b_\ell^* < b_\ell$.

In fact, the proof follows the argument in [6, Lemma 4.12]. Since $|\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)| > 0$, by [6, Lemma 4.11] if $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)$, then the semi-ellipsoid $E(m_\ell, b_\ell)$ supports $\mathcal{R}(\mathbf{b})$ at x_0. Hence $\frac{b_j}{1-\kappa\, m_j\cdot x_0} \geq \frac{b_\ell}{1-\kappa\, m_\ell\cdot x_0}$ for all j. Since $b_\ell^* < b_\ell$, we then get

$$\frac{b_j}{1-\kappa\, m_j\cdot x_0} > \frac{b_\ell^*}{1-\kappa\, m_\ell\cdot x_0} \quad \forall j.$$

By continuity there is a neighborhood V_{x_0} such that

$$\frac{b_j}{1-\kappa\, m_j\cdot x} > \frac{b_\ell^*}{1-\kappa\, m_\ell\cdot x} \quad \forall j,\ \forall x \in V_{x_0}.$$

Thus $V_{x_0} \subset \mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_l)$. Therefore, we have the inclusion

$$\mathcal{T} := \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l) \subset \textit{interior}\left(\mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_l)\right) := \mathcal{O}.$$

On the other hand, from the proof of [6, Lemma 3.12], the set $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)$ is compact. Therefore the set $\mathcal{O} \setminus \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)$ is an open set. Now since $G_{\mathcal{R}(\mathbf{b})}(m_l) < \int_\Omega g(x)\,dx$, by the continuity of the refractor measure as a function of b_ℓ, Lemma 3.6(ii), we have that $G_{\mathcal{R}(\mathbf{b}^*)}(m_l) < \int_\Omega g(x)\,dx$ for b_ℓ^* sufficiently close to b_ℓ. Since $G_{\mathcal{R}(\mathbf{b}^*)}(m_l)$ increases when b_ℓ^* decreases, it is enough to prove the desired inequality when b_ℓ^* is sufficiently close to b_ℓ. This implies that $\mathcal{O} \neq \bar{\Omega}$. So we have the configuration $\mathcal{T}$ closed, $\mathcal{T} \subset \mathcal{O} \subsetneqq \bar{\Omega}$. If the set $\mathcal{O} \setminus \mathcal{T} \neq \emptyset$, then since $\mathcal{O} \setminus \mathcal{T}$ is open, we have $|\mathcal{O} \setminus \mathcal{T}| > 0$. Since $\mathcal{O} \setminus \mathcal{T} \subset \mathcal{T}_{\mathcal{R}(\mathbf{b}^*)}(m_l) \setminus \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_l)$, we obtain the desired result. It then remains to show that $\mathcal{O} \setminus \mathcal{T} \neq \emptyset$. Suppose by contradiction that $\mathcal{O} \setminus \mathcal{T} = \emptyset$. That is, $\mathcal{O} \cap (\bar{\Omega} \setminus \mathcal{T}) = \emptyset$. We shall prove this implies that

$$\bar{\Omega} = \mathcal{O} \cup (\bar{\Omega} \cap \mathcal{T}^c). \tag{3.3}$$

Since both sets in this union are open ($\mathcal{T}$ is closed) relative to $\bar{\Omega}$, and $\mathcal{O} \neq \bar{\Omega}$, we obtain that $\bar{\Omega}$ is disconnected, contradicting the assumption that $\bar{\Omega}$ is connected. So let us prove (3.3). Write

$$\begin{aligned}\mathcal{O} \cup (\bar{\Omega} \cap \mathcal{T}^c) &= (\mathcal{O} \cup \bar{\Omega}) \cap (\mathcal{O} \cup \mathcal{T}^c) = \bar{\Omega} \cap (\mathcal{O} \cup \mathcal{T}^c)\\ &\supset \bar{\Omega} \cap (\mathcal{O} \cup \mathcal{O}^c) = \bar{\Omega} \quad \text{since } \mathcal{T} \subset \mathcal{O}.\end{aligned}$$

This completes the remark.

By [8, Lemma 3.6], we also have the following:

Lemma 3.5. *Let $\mathcal{R}_j = \{\rho_j(x)x : x \in \bar{\Omega}\}, j \geq 1$ be refractors from $\bar{\Omega}$ to $\bar{\Omega}^*$. Suppose that $0 < a_1 \leq \rho_j \leq a_2$ and $\rho_j \to \rho$ pointwise on $\bar{\Omega}$. Then:*

i. *$\mathcal{R} := \{\rho(x)x : x \in \bar{\Omega}\}$ is a refractor from $\bar{\Omega}$ to $\bar{\Omega}^*$.*
ii. *The measures $G_{\mathcal{R}_j}$ converge weakly to the measure $G_{\mathcal{R}}$.*

Lemma 3.6. *If in Lemma 3.5, $\mathcal{R}_j$ and $\mathcal{R}$ are defined by finite number of semi-ellipsoids as:*

$$\mathcal{R}_j = \mathcal{R}(\mathbf{b}_j) = \left\{\rho(x)x : x \in \bar{\Omega}, \rho(x) = \min_{1\leq i\leq N} \frac{b_i^j}{1-\kappa m_i\cdot x}\right\}$$

and

$$\mathcal{R} = \mathcal{R}(\mathbf{b}) = \left\{\rho(x)x : x \in \bar{\Omega}, \rho(x) = \min_{1\leq i\leq N} \frac{b_i}{1-\kappa m_i\cdot x}\right\}$$

then

i. $G_{\mathcal{R}_j} = \sum G_{\mathcal{R}_j}(m_i)\delta_{m_i}, G_{\mathcal{R}} = \sum G_{\mathcal{R}}(m_i)\delta_{m_i}$
ii. $G_{\mathcal{R}_j}(m_i) \to G_{\mathcal{R}}(m_i)$ *for all* $1 \le i \le N$, *when* $\mathbf{b}_j \to \mathbf{b}$.

For a proof of this lemma see [6, Lemma 4.7].

3.2. Geodesic disks

Recall that if $\alpha, \beta \in S^2$, the geodesic distance between them is given by $\cos^{-1}(\alpha \cdot \beta)$. We define a geodesic disk with center α and radius r to be the set of points x on S^2 for which $x \cdot \alpha \ge \cos r$.

Lemma 3.7. *Let* $b_1, b_2 > 0$ *and* $m_1, m_2 \in S^2$ *be such that* $m_1 \ne m_2$. *Consider the set*

$$V_{12} = \left\{ x \in S^2 : \frac{b_1}{1 - \kappa\, x \cdot m_1} \le \frac{b_2}{1 - \kappa\, x \cdot m_2} \right\}.$$

This set is non empty if and only if $\frac{b_1 - b_2}{\kappa\,|b_1 m_2 - b_2 m_1|} \le 1$, *and* V_{12} *is the geodesic disk with center at*

$$A_{12} = \frac{b_1 m_2 - b_2 m_1}{|b_1 m_2 - b_2 m_1|}$$

and radius

$$\tau_{12} = \cos^{-1} \frac{b_1 - b_2}{\kappa |b_1 m_2 - b_2 m_1|},$$

that is,

$$V_{12} = \left\{ x \in S^2 : x \cdot A_{12} \ge \cos \tau_{12} \right\}.$$

In addition, if $\frac{b_1 - b_2}{\kappa\,|b_1 m_2 - b_2 m_1|} \le -1$, *then* $V_{12} = S^2$. *If* $\mathcal{R}(\mathbf{b})$ *is the refractor in* Ω *with polar radius* $\rho(x) = \min\left\{ \frac{b_1}{1 - \kappa\, x \cdot m_1}, \frac{b_2}{1 - \kappa\, x \cdot m_2} \right\}$, *then* $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1) \subset V_{12}$.

Proof. If $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1) = \emptyset$, there is nothing to prove. Otherwise, let $x \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1)$. Then $\frac{b_1}{1 - \kappa m_1 \cdot x} \le \frac{b_2}{1 - \kappa m_2 \cdot x}$. So $b_1 - b_2 \le x \cdot \kappa(b_1 m_2 - b_2 m_1)$, and we obtain

$$x \cdot \frac{b_1 m_2 - b_2 m_1}{|b_1 m_2 - b_2 m_1|} \ge \frac{1}{\kappa} \frac{b_1 - b_2}{|b_1 m_2 - b_2 m_1|},$$

in particular, we must have $\frac{b_1 - b_2}{\kappa\,|b_1 m_2 - b_2 m_1|} \le 1$. Thus $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1)$ is contained in the geodesic disk with center at $\frac{b_1 m_2 - b_2 m_1}{|b_1 m_2 - b_2 m_1|}$ and geodesic radius $\cos^{-1}\left(\frac{1}{\kappa} \frac{b_1 - b_2}{|b_1 m_2 - b_2 m_1|} \right)$. $\square$

Remark 3.8. If $\mathcal{R}(\mathbf{b})$ is a refractor of the form (2.6), then

$$\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) = \Omega \cap \cap_{j=1}^N V_{ij}, \text{ except possibly on the singular set of } \mathcal{R}(\mathbf{b}), \tag{3.4}$$

where $V_{ij} = \left\{ x \in S^2 : \frac{b_i}{1 - \kappa\, x \cdot m_i} \le \frac{b_j}{1 - \kappa\, x \cdot m_j} \right\}$. In fact, if $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)$ and x_0 is not singular, then by [8, Lemma 5.1] the semi-ellipsoid $E(m_i, b_i)$ supports $\mathcal{R}(\mathbf{b})$ at x_0 implying $\frac{b_i}{1 - \kappa\, x_0 \cdot m_i} \le \frac{b_j}{1 - \kappa\, x_0 \cdot m_j}$ for all j. Vice versa, if $x_0 \in \Omega \cap \cap_{j=1}^N V_{ij}$, then the polar radius ρ satisfies $\rho(x_0) = \frac{b_i}{1 - \kappa\, x_0 \cdot m_i}$, and so $\frac{b_i}{1 - \kappa\, x \cdot m_i}$ supports ρ at x_0.

The following example shows that in (3.4) it is necessary to remove the singular points. In fact, take two ellipsoids E_1 and E_2 with polar radii $\frac{b_1}{1 - \kappa\, x \cdot m_1}$ and $\frac{b_2}{1 - \kappa\, x \cdot m_2}$ respectively, with $m_1 \ne m_2$ and take the corresponding refractor minimum of the two ellipsoids and let $\rho(x)$ be the polar radius. Suppose the refractor ρ has a singular point x_0. At x_0 take a supporting semi-ellipsoid to ρ having axis m_3, with m_3

different from m_1 and m_2. Let the polar radius of this ellipsoid be $\frac{b_3}{1-\kappa x\cdot m_3}$. Now take an ellipsoid E of the form $\frac{b*}{1-\kappa x\cdot m_3}$ with $b*$ much larger than b_3 so that the ellipsoids E_1 and E_2 are contained in the interior of the solid E, that is, $\frac{b_i}{1-\kappa\, x\cdot m_i}, i = 1, 2$, are both strictly smaller than $\frac{b*}{1-\kappa x\cdot m_3}$. Then refractor $\min\left\{\frac{b_1}{1-\kappa\, x\cdot m_1}, \frac{b_2}{1-\kappa\, x\cdot m_2}, \frac{b*}{1-\kappa x\cdot m_3}\right\}$ is the same as the refractor $\rho(x)$. We have that $x_0 \in \mathcal{T}_\rho(m_3)$. On the other hand, the sets $V_{31} = \left\{\frac{b*}{1-\kappa x\cdot m_3} \le \frac{b_1}{1-\kappa\, x\cdot m_1}\right\} = \emptyset$ and $V_{32} = \left\{\frac{b*}{1-\kappa x\cdot m_3} \le \frac{b_2}{1-\kappa\, x\cdot m_2}\right\} = \emptyset$.

4. The algorithm

We assume the energy conservation condition (2.7).

4.1. The set W of admissible vectors

Let $N \ge 2, f_o = \min\left\{\frac{f_1}{N-1}, f_2, \dots, f_N\right\}$, and $0 < \delta < f_o$. Consider the set of admissible vectors

$$W = \{\mathbf{b} = (1, b_2, \dots, b_N) : b_i > 0 \,\text{and}\, G_{\mathcal{R}(\mathbf{b})}(m_i) \le f_i + \delta \text{ for } i = 2, \dots, N\}.$$

This set is non empty and their coordinates are bounded away from zero. This is the contents of the following lemma.

Lemma 4.1. *We have the following*

(1) *if* $b_i > 1 + \kappa$ *for* $2 \le i \le N$, *then* $(1, b_2, b_3, \dots, b_N) \in W$;
(2) *if* $0 < \delta < f_1/(N-1)$ *and* $\mathbf{b} = (1, b_2, \dots, b_N) \in W$, *then*

$$b_i \ge \frac{1}{1+\kappa} \quad \textit{for } 2 \le i \le N. \tag{4.1}$$

Proof. We prove (1). Let $\mathbf{b} = (1, b_2, \dots, b_N)$ with $b_i > 0$. Fix $j \ge 2$ and let $x \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_j)$ be a non singular point. Then from [8, Lemma 5.1], the semi-ellipsoid $E(m_j, b_j)$ supports $\mathcal{R}(\mathbf{b})$ at x. Since $x \cdot m_j \ge \kappa$, we have

$$\frac{b_j}{1-\kappa^2} \le \frac{b_j}{1-\kappa\, x\cdot m_j} = \rho(x) \le \frac{1}{1-\kappa\, x\cdot m_1} \le \frac{1}{1-\kappa},$$

and so $b_j \le 1 + \kappa$. Therefore, if $b_j > 1 + \kappa$ with $j \ge 2$, and $x \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_j)$, then x is a singular point and therefore $\mathcal{T}_{\mathcal{R}}(m_j)$ has measure zero, and so $G_{\mathcal{R}(b)}(m_j) = 0 < f_j + \delta$.

To show (2), we first prove that $G_{\mathcal{R}(\mathbf{b})}(m_1) > 0$ for each $\mathbf{b} \in W$. In fact, from (2.7) and the definition of W we have

$$G_{\mathcal{R}(\mathbf{b})}(m_1) = f_1 + \sum_{i=2}^{N}(f_i - G_{\mathcal{R}(\mathbf{b})}(m_i)) > f_1 - (N-1)\delta > 0$$

from the choice of δ. Since $g \ge 0$, the set $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1)$ has positive measure. This implies that for each $\mathbf{b} \in W, \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1) \cap \left(\cup_{i=2}^N \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)\right)^c \ne \emptyset$. Otherwise, $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1) \subset \cup_{i=2}^N \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)$ which means that each point in $\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1)$ is singular, and therefore $|\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1)| = 0$; a contradiction. From this we conclude (4.1) because, if $\mathbf{b} \in W$, then we can pick $x_0 \in \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_1) \cap \left(\cup_{i=2}^N \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)\right)^c$ and we have

$$\rho(x_0) = \frac{1}{1-\kappa\, x_0\cdot m_1} \le \frac{b_i}{1-\kappa\, x_0\cdot m_i}, \quad i = 2, \dots, N$$

so

$$b_i \ge \frac{1-\kappa\, x_0\cdot m_i}{1-\kappa\, x_0\cdot m_1} \ge \frac{1-\kappa\, x_0\cdot m_i}{1-\kappa^2} \ge \frac{1-\kappa}{1-\kappa^2} = \frac{1}{1+\kappa}. \quad \square$$

4.2. Detailed description of the algorithm

From Lemma 4.1(2), we can pick $\mathbf{b}^1 = (1, b_2, \dots, b_N) \in W$. We will construct $N-1$ intermediate consecutive vectors $\mathbf{b}^2, \dots, \mathbf{b}^N$ associated with $\mathbf{b}^1$ in the following way.

Step 1. We first test if $\mathbf{b}^1$ satisfies the inequality:

$$f_2 - \delta \le G_{\mathcal{R}(\mathbf{b}^1)}(m_2) \le f_2 + \delta. \tag{4.2}$$

If $\mathbf{b}^1$ satisfies this inequality, then we set $\mathbf{b}^2 = \mathbf{b}^1$ and we proceed to Step 2 below. Notice that the inequality on the right hand side of (4.2) holds since $\mathbf{b}^1 \in W$. If $\mathbf{b}^1$ does not satisfy (4.2), then

$$G_{\mathcal{R}(\mathbf{b}^1)}(m_2) < f_2 - \delta. \tag{4.3}$$

We shall pick $b_2^* \in (0, b_2)$, and leave all other components fixed, so that the new vector $\mathbf{b}^2 = (1, b_2^*, b_3, \dots, b_N)$ belongs to W, and satisfies

$$f_2 \le G_{\mathcal{R}(\mathbf{b}^2)}(m_2) \le f_2 + \delta. \tag{4.4}$$

In fact, this is possible because applying Lemma 3.3 with $\ell = 2$ we get that $G_{\mathcal{R}(\mathbf{b}^2)}(m_j) \le G_{\mathcal{R}(\mathbf{b}^1)}(m_j)$ for $j \ne 2$ and $b_2^* \in (0, b_2]$ from (3.2); and applying Lemma 3.1(ii.) we get that $G_{\mathcal{R}(\mathbf{b}^2)}(m_2) \to \int_\Omega g(x)\,dx = f_1 + \dots + f_N$ as $b_2^* \to 0$, from the energy conservation assumption. Since the f_i's are all positive, $f_1 + \dots + f_N > f_2$, and from the choice of δ we have $f_1 + \dots + f_N > f_2 + \delta$. We obviously have $f_1 + \dots + f_N > f_2 - \delta$. From (4.3), the value of $G_{\mathcal{R}(1,t,b_3,\dots,b_N)}(m_2)$ when $t = b_2$ is strictly less than $f_2 - \delta$. Also from Lemma 3.1(ii.) we get that $G_{\mathcal{R}(1,t,b_3,\dots,b_N)}(m_2) \to \int_\Omega g(x)\,dx = f_1 + \dots + f_N$ as $t \to 0$. From Lemma 3.6ii, $G_{\mathcal{R}(1,t,b_3,\dots,b_N)}(m_2)$ is continuous for t in the interval $(0, b_2]$. Therefore, by the intermediate value theorem there is $b_2^* \in (0, b_2)$ such that $G_{\mathcal{R}(1,b_2^*,b_3,\dots,b_N)}(m_2) = f_2$.[4] Therefore, if the vector $\mathbf{b}^1$ does not satisfy (4.2), we have then constructed a vector $\mathbf{b}^2 \in W$ that satisfies (4.4) which is stronger than (4.2).

Step 2. Next we proceed to test the inequality

$$f_3 - \delta \le G_{\mathcal{R}(\mathbf{b}^2)}(m_3) \le f_3 + \delta, \tag{4.5}$$

with $\mathbf{b}^2$ the vector constructed in Step 1. If $\mathbf{b}^2$ satisfies (4.5), we set $\mathbf{b}^3 = \mathbf{b}^2$ and we proceed to the next step. If $\mathbf{b}^2$ does not satisfy (4.5), then

$$G_{\mathcal{R}(\mathbf{b}^2)}(m_3) < f_3 - \delta$$

and we proceed as before, now to decrease the value of b_3, the third component of the vector $\mathbf{b}^2$, and construct a vector $\mathbf{b}^3 \in W$ such that

$$f_3 \le G_{\mathcal{R}(\mathbf{b}^3)}(m_3) \le f_3 + \delta,$$

and in particular, (4.5) holds for $\mathbf{b}^3$. Notice that we do not know if the newly constructed vector $\mathbf{b}^3$ satisfies (4.2).

Step 3. Next we proceed to test the inequality

$$f_4 - \delta \le G_{\mathcal{R}(\mathbf{b}^3)}(m_4) \le f_4 + \delta, \tag{4.6}$$

with $\mathbf{b}^3$ the vector from Step 2. If this is true, then we set $\mathbf{b}^4 = \mathbf{b}^3$ and proceed to the next step. Otherwise, we must have

$$G_{\mathcal{R}(\mathbf{b}^3)}(m_4) < f_4 - \delta$$

[4] Notice that for any $a \in [f_2 - \delta, f_1 + \dots, f_N]$, we can pick $b_2^* \in (0, b_2)$ such that $G_{\mathcal{R}(\mathbf{b}^2)}(m_2) = a$.

and we continue in the same way as before now decreasing the fourth component b_4 of $\mathbf{b}^3$ obtaining a new vector $\mathbf{b}^4$ satisfying

$$f_4 \leq G_{\mathcal{R}(\mathbf{b}^4)}(m_4) \leq f_4 + \delta,$$

in particular, (4.6).

Step $N-1$. We proceed to test the inequality

$$f_N - \delta \leq G_{\mathcal{R}(\mathbf{b}^{N-1})}(m_N) \leq f_N + \delta, \tag{4.7}$$

where $\mathbf{b}^{N-1}$ is the vector from Step $N-2$. If this holds we set $\mathbf{b}^N = \mathbf{b}^{N-1}$. Otherwise, we have

$$G_{\mathcal{R}(\mathbf{b}^{N-1})}(m_N) < f_N - \delta,$$

and proceeding as before, by decreasing the Nth-component of $\mathbf{b}^{N-1}$, we obtain a vector $\mathbf{b}^N \in W$

$$f_N \leq G_{\mathcal{R}(\mathbf{b}^N)}(m_N) \leq f_N + \delta.$$

In this way, starting from a fixed vector $\mathbf{b}^1 \in W$, we have constructed intermediate vectors $\mathbf{b}^2, \ldots, \mathbf{b}^N$ all belonging to W and satisfying the above inequalities. Notice that by construction, the ℓth components of $\mathbf{b}^{j-1}$ and $\mathbf{b}^j$ are all equal for $\ell \neq j$. If for some $2 \leq j \leq N, \mathbf{b}^{j-1} \neq \mathbf{b}^j$, then the jth component of $\mathbf{b}^j$ is strictly less than the jth component of $\mathbf{b}^{j-1}$. And so if we needed to decrease the jth component of $\mathbf{b}^{j-1}$ to construct $\mathbf{b}^j$ is because

$$G_{\mathcal{R}(\mathbf{b}^{j-1})}(m_j) < f_j - \delta,$$

and then by construction $\mathbf{b}^j$ satisfies

$$f_j \leq G_{\mathcal{R}(\mathbf{b}^j)}(m_j) \leq f_j + \delta.$$

We therefore obtain from the last two inequalities the following important inequality

$$\delta < G_{\mathcal{R}(\mathbf{b}^j)}(m_j) - G_{\mathcal{R}(\mathbf{b}^{j-1})}(m_j), \quad \text{for intermediate vectors } \mathbf{b}^j \neq \mathbf{b}^{j-1}. \tag{4.8}$$

We now repeat the construction above starting with the last vector $\mathbf{b}_N$. In fact, we start from a vector $\mathbf{b}^{1,1} \in W$ and constructed $N-1$ intermediate vectors $\mathbf{b}^{1,2}, \ldots, \mathbf{b}^{1,N}$ using the procedure described. So we obtain in the first step the finite sequence of vectors

$$\mathbf{b}^{1,1}, \mathbf{b}^{1,2}, \ldots, \mathbf{b}^{1,N}.$$

In the second step we repeat the construction now starting with the vector $\mathbf{b}^{1,N}$ and we get the finite sequence of vectors

$$\mathbf{b}^{2,1}, \mathbf{b}^{2,2}, \ldots, \mathbf{b}^{2,N}$$

with $\mathbf{b}^{2,1} = \mathbf{b}^{1,N}$. For the third step we repeat the process now starting with the last intermediate vector $\mathbf{b}^{2,N}$ obtained in the previous step, obtaining the finite sequence of vectors

$$\mathbf{b}^{3,1}, \mathbf{b}^{3,2}, \ldots, \mathbf{b}^{3,N}$$

with $\mathbf{b}^{3,1} = \mathbf{b}^{2,N}$. Continuing in this way we obtain a sequence of vectors, in principle infinite,

$$\mathbf{b}^{1,1}, \ldots, \mathbf{b}^{1,N}; \mathbf{b}^{2,1}, \ldots, \mathbf{b}^{2,N}; \mathbf{b}^{3,1}, \ldots, \mathbf{b}^{3,N}; \cdots; \mathbf{b}^{n,1}, \ldots, \mathbf{b}^{n,N}; \mathbf{b}^{n+1,1}, \ldots, \mathbf{b}^{n+1,N}; \ldots \tag{4.9}$$

with $\mathbf{b}^{2,1} = \mathbf{b}^{1,N}, \mathbf{b}^{3,1} = \mathbf{b}^{2,N}, \ldots, \mathbf{b}^{n+1,1} = \mathbf{b}^{n,N}, \ldots$. If for some n, the vectors in the nth-stage are equal, i.e., $\mathbf{b}^{n,1} = \mathbf{b}^{n,2} = \cdots = \mathbf{b}^{n,N} := \mathbf{b}^n$, then from the construction

$$|G_{\mathcal{R}(\mathbf{b}^n)}(m_j) - f_j| \leq \delta, \quad \text{for } 2 \leq j \leq N.$$

Furthermore, by conservation of energy, $\sum_{i=1}^{N} G_{\mathcal{R}(\mathbf{b}^n)}(m_i) = \sum_{i=1}^{N} f_i$, so we obtain

$$|f_1 - G_{\mathcal{R}(\mathbf{b}^n)}(m_1)| = \left|\sum_{j=2}^{N} G_{\mathcal{R}(\mathbf{b}^n)}(m_j) - f_j\right| \leq \sum_{j=2}^{N} |G_{\mathcal{R}(\mathbf{b}^n)}(m_j) - f_j| \leq N\,\delta.$$

If we now choose $\delta = \epsilon/N$, then the refractor $\mathcal{R}(\mathbf{b}^n)$ will satisfy (2.8), and the problem is solved.

Therefore, if we show that for some n the intermediate vectors $\mathbf{b}^{1,n}, \mathbf{b}^{2,n}, \ldots, \mathbf{b}^{n,N}$ are all equal, we are done.

4.3. A Lipschitz estimate implies that the process stops

We shall prove that the estimate (5.6) implies that there is an n such that the vectors in the group $\mathbf{b}^{n,1}, \mathbf{b}^{n,2}, \ldots, \mathbf{b}^{n,N}$ are all equal, and we also show an upper bound for the number of iterations.

Suppose we originate the iteration at $\mathbf{b}^0 = (1, b_2^0, \ldots, b_N^0) \in W$. Since by construction the coordinates of the vectors in the sequence (4.9) are decreased or kept constant, the jth coordinate of any vector in the sequence is less than or equal to $b_j^0, 1 \leq j \leq N$. In addition, from (4.1), points in W have all their coordinates bounded below by $1/(1+\kappa)$. Therefore all terms in the sequence (4.9) are contained in the compact box $K = \{1\} \times \prod_{j=2}^{N}[1/(1+\kappa), b_j^0]$. We want to show that there is n_0 such that the intermediate vectors $\mathbf{b}^{n_0,1}, \mathbf{b}^{n_0,2}, \ldots, \mathbf{b}^{n_0,N}$ are all equal. Otherwise, for each n the intermediate vectors $\mathbf{b}^{n,1}, \mathbf{b}^{n,2}, \ldots, \mathbf{b}^{n,N}$ are not all equal. This implies that for each n there are two consecutive intermediate vectors $(1, b_2, b_3, \ldots, b_N)$ and $(1, \bar{b}_2, \bar{b}_3, \ldots, \bar{b}_N)$, that are different. By construction of intermediate vectors, they can only differ in one coordinate, say that $b_j > \bar{b}_j$. Notice that j depends on n, but there is j and a subsequence n_ℓ such that there are two consecutive intermediate vectors $(1, b_2^{n_\ell}, b_3^{n_\ell}, \ldots, b_N^{n_\ell})$ and $(1, \bar{b}_2^{n_\ell}, \bar{b}_3^{n_\ell}, \ldots, \bar{b}_N^{n_\ell})$ in each group $\mathbf{b}^{n_\ell,1}, \ldots, \mathbf{b}^{n_\ell,N}$ such that their jth coordinates satisfy $b_j^{n_\ell} > \bar{b}_j^{n_\ell}$, and all other coordinates are equal. Also notice that since the coordinates are chosen in a decreasing form we have $b_j^{n_\ell} > \bar{b}_j^{n_\ell} \geq b_j^{n_{\ell+1}} > \bar{b}_j^{n_{\ell+1}}$ for $\ell = 1, \ldots$. From (4.8) we then get

$$\delta < G_j\left(1, \bar{b}_2^{n_\ell}, \bar{b}_3^{n_\ell}, \ldots, \bar{b}_N^{n_\ell}\right) - G_j\left(1, b_2^{n_\ell}, b_3^{n_\ell}, \ldots, b_N^{n_\ell}\right) = (*) \tag{4.10}$$

for each $\ell \geq 1$. We write

$$(1, \bar{b}_2^{n_\ell}, \bar{b}_3^{n_\ell}, \ldots, \bar{b}_j^{n_\ell}, \ldots \bar{b}_N^{n_\ell}) = (1, \bar{b}_2^{n_\ell}, \bar{b}_3^{n_\ell}, \ldots, b_j^{n_\ell} + \bar{b}_j^{n_\ell} - b_j^{n_\ell}, \ldots \bar{b}_N^{n_\ell}),$$

and let $t := \bar{b}_j^{n_\ell} - b_j^{n_\ell} < 0$. Since the vectors belong to W, we have $\bar{b}_j^{n_\ell} \geq 1/(1+\kappa)$. Then from (5.6) we obtain

$$(*) \leq -\left(\bar{b}_j^{n_\ell} - b_j^{n_\ell}\right) C_\kappa \left(\sup_{\Omega} g\right)(N-1) := L\left(b_j^{n_\ell} - \bar{b}_j^{n_\ell}\right), \quad \forall \ell. \tag{4.11}$$

On the other hand,

$$\sum_{\ell=1}^{\infty} (b_j^{n_\ell} - \bar{b}_j^{n_\ell}) \leq b_j^0 - \frac{1}{1+\kappa}, \tag{4.12}$$

which contradicts (4.10) and therefore the intermediate vectors $\mathbf{b}^{n_0,1}, \mathbf{b}^{n_0,2}, \ldots, \mathbf{b}^{n_0,N}$ are all equal for some n_0.

Let us now estimate n_0. Consider the groups of vectors (4.9) in the construction. Since the process stops at some n_0, i.e., all vectors in this group are equal. Fix a coordinate $2 \leq j \leq N$. Notice that in each group k, the j coordinate of any vector in the group can decrease *at most only once* and only when passing from a vector $\mathbf{b}^{k,j-1}$ to the vector $\mathbf{b}^{k,j}$. Here k denotes the group and j the location in the group. The coordinate j of all vectors are at most b_j^0 (the j-coordinate of the initial vector), and they are at least $1/(1+\kappa)$ since the vectors belong to W. So the change in the j-coordinate of a vector in the group one to the group n_0, is at

most $b_j^0 - \frac{1}{1+\kappa}$. On the other hand, on each group if the jth coordinate is decreased, then (4.11) is decreased at least $\frac{\delta}{L}$. Having n_0 groups the total decrease of passing from the group one to the group n_0 is at least

$$n_0 \frac{\delta}{L},$$

which is in turn smaller than the total possible decrease, that is, we have

$$n_0 \frac{\delta}{L} \leq b_j^0 - \frac{1}{1+\kappa}.$$

Since this must hold for all the coordinates $2 \leq j \leq N$ we obtain the bound

$$n_0 \leq \frac{L}{\delta} \max_{2 \leq j \leq N} \left(b_j^0 - \frac{1}{1+\kappa} \right) = \frac{C_\kappa \left(\sup_\Omega g \right) (N-1)}{\delta} \max_{2 \leq j \leq N} \left(b_j^0 - \frac{1}{1+\kappa} \right). \tag{4.13}$$

4.4. Limit as $n \to \infty$ of the sequence (4.9)

We will show here that the procedure described always converges in an infinite number of steps, assuming only that $g \in L^1(\Omega)$ with g not necessarily bounded. This can be clearly seen by listing the vectors constructed in the following way:

$$\text{group 1} \left\{ \begin{array}{lcccccc}
\mathbf{b}^{1,1} & \to & 1 & b_2^{1,1} & b_3^{1,1} & b_4^{1,1} & \cdots & b_N^{1,1} \\
 & & \shortparallel & \vee\!\shortparallel & \shortparallel & \shortparallel & \cdots & \shortparallel \\
\mathbf{b}^{1,2} & \to & 1 & b_2^{1,2} & b_3^{1,2} & b_4^{1,2} & \cdots & b_N^{1,2} \\
 & & \shortparallel & \shortparallel & \vee\!\shortparallel & \shortparallel & \cdots & \shortparallel \\
\mathbf{b}^{1,3} & \to & 1 & b_2^{1,3} & b_3^{1,3} & b_4^{1,3} & \cdots & b_N^{1,3} \\
 & & \shortparallel & \shortparallel & \shortparallel & \vee\!\shortparallel & \cdots & \shortparallel \\
\cdots & & & & & & & \\
\mathbf{b}^{1,N-1} & \to & 1 & b_2^{1,N-1} & b_3^{1,N-1} & b_4^{1,N-1} & \cdots & b_N^{1,N-1} \\
 & & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \cdots & \vee\!\shortparallel \\
\mathbf{b}^{1,N} & \to & 1 & b_2^{1,N} & b_3^{1,N} & b_4^{1,N} & \cdots & b_N^{1,N} \\
 & & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \cdots & \shortparallel
\end{array} \right.$$

$$\text{group 2} \left\{ \begin{array}{lcccccc}
\mathbf{b}^{2,1} & \to & 1 & b_2^{2,1} & b_3^{2,1} & b_4^{2,1} & \cdots & b_N^{2,1} \\
 & & \shortparallel & \vee\!\shortparallel & \shortparallel & \shortparallel & \cdots & \shortparallel \\
\mathbf{b}^{2,2} & \to & 1 & b_2^{2,2} & b_3^{2,2} & b_4^{2,2} & \cdots & b_N^{2,2} \\
 & & \shortparallel & \shortparallel & \vee\!\shortparallel & \shortparallel & \cdots & \shortparallel \\
\mathbf{b}^{2,3} & \to & 1 & b_2^{2,3} & b_3^{2,3} & b_4^{2,3} & \cdots & b_N^{2,3} \\
 & & \shortparallel & \shortparallel & \shortparallel & \vee\!\shortparallel & \cdots & \shortparallel \\
\cdots & & & & & & & \\
\mathbf{b}^{2,N-1} & \to & 1 & b_2^{2,N-1} & b_3^{2,N-1} & b_4^{2,N-1} & \cdots & b_N^{1,N-1} \\
 & & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \cdots & \vee\!\shortparallel \\
\mathbf{b}^{2,N} & \to & 1 & b_2^{2,N} & b_3^{2,N} & b_4^{2,N} & \cdots & b_N^{2,N} \\
 & & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \cdots & \shortparallel
\end{array} \right.$$

$$\text{group 3} \left\{ \begin{array}{lcccccc}
\mathbf{b}^{3,1} & \to & 1 & b_2^{3,1} & b_3^{3,1} & b_4^{3,1} & \cdots & b_N^{3,1} \\
 & & \shortparallel & \vee\!\shortparallel & \shortparallel & \shortparallel & \cdots & \shortparallel \\
\mathbf{b}^{3,2} & \to & 1 & b_2^{3,2} & b_3^{3,2} & b_4^{3,2} & \cdots & b_N^{3,2} \\
 & & \shortparallel & \shortparallel & \vee\!\shortparallel & \shortparallel & \cdots & \shortparallel \\
\mathbf{b}^{3,3} & \to & 1 & b_2^{3,3} & b_3^{3,3} & b_4^{3,3} & \cdots & b_N^{3,3} \\
 & & \shortparallel & \shortparallel & \shortparallel & \vee\!\shortparallel & \cdots & \shortparallel \\
\cdots & & & & & & & \\
\mathbf{b}^{3,N-1} & \to & 1 & b_2^{3,N-1} & b_3^{3,N-1} & b_4^{3,N-1} & \cdots & b_N^{1,N-1} \\
 & & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \cdots & \vee\!\shortparallel \\
\mathbf{b}^{3,N} & \to & 1 & b_2^{3,N} & b_3^{3,N} & b_4^{3,N} & \cdots & b_N^{3,N} \\
 & & \shortparallel & \shortparallel & \shortparallel & \shortparallel & \cdots & \shortparallel
\end{array} \right.$$

and continuing in this way we get an infinite matrix having N columns. With the notation $b_k^{i,j}$ we have that $i = \text{group}, j = \text{vector in the group}$, and $k = \text{the component}$. We have

$$b_{j+1}^{i,j} \geq b_{j+1}^{i,j+1}, \quad \text{for } j = 1, \dots, N-1, \text{ and } i = 1, 2, \dots$$

and

$$b_\ell^{i,j} = b_\ell^{i,j+1}, \quad \text{for } \ell \neq j+1.$$

We now look at each of the N columns of the infinite matrix above. Each column has entries in non increasing order (the first column is obviously one), therefore the limit of the entries exists and is a number different from zero because the vectors belong to W and therefore each limiting coordinate is bigger than $1/(1+\kappa)$. Let b_j^∞ be the limit of the entries in the column $j, j \geq 2$. Then the vector

$$\mathbf{b}^\infty = (1, b_2^\infty, b_2^\infty, \dots, b_N^\infty)$$

satisfies

$$f_j - \delta \leq \int_{\mathcal{T}_{\mathcal{R}(\mathbf{b}^\infty)}(m_j)} g(x)\, dx \leq f_j + \delta,\ j = 2, \dots, N. \tag{4.14}$$

In fact, fix $2 \leq j \leq N$, the vector $\mathbf{b}^\infty$ is the limit of the vectors $\mathbf{b}^{i,j}$ as $i \to \infty$. But the vectors $\mathbf{b}^{i,j}$ verify

$$f_j - \delta \leq \int_{\mathcal{T}_{\mathcal{R}(\mathbf{b}^{i,j})}(m_j)} g(x)\, dx \leq f_j + \delta, \quad \text{for } i = 1, 2, \dots.$$

Since the function $\int_{\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_j)} g(x)\, dx$ is continuous as a function of $\mathbf{b}$ for each j, Lemma 3.6ii, taking the limit as $i \to \infty$ we obtain (4.14). As it was shown before, the validity of (4.14) for $j \neq 1$ implies that (4.14) holds with $j = 1$ and with δ replaced by $N\delta$.

5. A Lipschitz estimate of G_i

Consider the map $G : \mathbb{R}_+^N \to \mathbb{R}_{\geq 0}^N$ given by

$$G : \mathbf{b} = (b_1, \dots, b_N) \to (G_1(\mathbf{b}), \dots, G_N(\mathbf{b})) \tag{5.1}$$

where

$$G_j(\mathbf{b}) = G_{\mathcal{R}(\mathbf{b})}(m_j) \tag{5.2}$$

for $j = 1, \dots, N$ and $\mathbb{R}_+^N = \{\mathbf{b} = (b_1, \dots, b_k) : b_j > 0 \text{ for } j = 1, \dots, N\}$.

Let $\mathbf{e}_i$ be the unit vector in $\mathbb{R}^N$ with 1 at the ith position. We shall compute $G_i(\mathbf{b}^t) - G_i(\mathbf{b})$ where $\mathbf{b}^t = (b_1^t, \dots, b_N^t) := \mathbf{b} + t\,\mathbf{e}_i$. From Remark 3.8

$$\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) = \Omega \cap \bigcap_{j=1}^N V_j$$

except possibly on a set of measure zero, with

$$V_j = \left\{ x \in S^2 : \frac{b_i}{1 - \kappa m_i \cdot x} \leq \frac{b_j}{1 - \kappa m_j \cdot x} \right\}, \tag{5.3}$$

where for brevity we have used the notation V_j for V_{ij}. Likewise

$$\mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i) = \Omega \cap \bigcap_{j=1}^N V_j^t$$

where

$$V_j^t = \left\{ x \in S^2 : \frac{b_i^t}{1 - \kappa m_i \cdot x} \leq \frac{b_j^t}{1 - \kappa m_j \cdot x} \right\}. \tag{5.4}$$

We have $V_j^t = V_j = S^2$ for $j = i$. So

$$\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) = \Omega \cap \bigcap_{j \neq i} V_j, \qquad \mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i) = \Omega \cap \bigcap_{j \neq i} V_j^t. \tag{5.5}$$

We prove the following proposition needed to show in Section 4.3 that the algorithm stops in a finite number of steps.

Proposition 5.1. *If g is bounded in Ω, then*

$$0 \leq G_i(\mathbf{b} + t\,\mathbf{e}_i) - G_i(\mathbf{b}) \leq \left(\sup_{\Omega} g \right) \sum_{r \neq i} \frac{C(\kappa, m_i \cdot m_r)}{b_r} (-t), \tag{5.6}$$

for $-b_i < t < 0$ and for each $\mathbf{b} \in \mathbb{R}_+^N$, where the constant $C(\kappa, m_i \cdot m_r)$ depends only on κ and the angle between m_i and m_r.

Proof. We have

$$V_j^t \subset V_j \quad \text{for } t > 0, j \neq i \quad \text{and} \quad V_j \subset V_j^t \quad \text{for } t < 0, j \neq i,$$

so from (5.5)

$$\mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i) \subset \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) \quad \text{for } t > 0$$

and

$$\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) \subset \mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i) \quad \text{for } t < 0.$$

Since

$$G_i(\mathbf{b}^t) - G_i(\mathbf{b}) = \int_{\mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i)} g(x)dx - \int_{\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)} g(x)dx,$$

we obtain

$$G_i(\mathbf{b}^t) - G_i(\mathbf{b}) = \begin{cases} -\displaystyle\int_{\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) \setminus \mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i)} g(x)dx & \text{if } t > 0 \\ \displaystyle\int_{\mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i) \setminus \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)} g(x)dx & \text{if } t < 0. \end{cases}$$

If $t < 0$, then we have

$$\begin{aligned} \mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i) \setminus \mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) &= \Omega \cap \left\{ \left\{ \cap_{j \neq i} V_j^t \right\} \setminus \left\{ \cap_{r \neq i} V_r \right\} \right\} \\ &= \Omega \cap \left\{ \left\{ \cap_{j \neq i} V_j^t \right\} \cap \left\{ (\cap_{r \neq i} V_r)^c \right\} \right\} \\ &= \Omega \cap \left\{ \left\{ \cap_{j \neq i} V_j^t \right\} \cap \left\{ \cup_{r \neq i} V_r^c \right\} \right\} \\ &= \Omega \cap \left\{ \cup_{r \neq i} \left\{ V_r^c \cap \left\{ \cap_{j \neq i} V_j^t \right\} \right\} \right\} \\ &\subset \Omega \cap \left\{ \cup_{r \neq i} \left\{ V_r^c \cap V_r^t \right\} \right\} \\ &\subset \cup_{r \neq i} \left(V_r^t \setminus V_r \right). \end{aligned}$$

On the other hand, if $t > 0$, then

$$\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i) \setminus \mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i) \subset \cup_{r \neq i} \left(V_r \setminus V_r^t \right).$$

We will estimate for $-b_i < t < 0$

$$\begin{aligned} 0 \le G_i(\mathbf{b}^t) - G_i(\mathbf{b}) &= \int_{\mathcal{T}_{\mathcal{R}(\mathbf{b}^t)}(m_i)\setminus\mathcal{T}_{\mathcal{R}(\mathbf{b})}(m_i)} g(x)dx \le \int_{\Omega\cap\{\cup_{r\neq i}\{V_r^c\cap V_r^t\}\}} g(x)\,dx \\ &\le \left(\sup_{\Omega} g\right) \text{area}\left(\cup_{r\neq i}\left(V_r^t \setminus V_r\right)\right) \\ &\le \left(\sup_{\Omega} g\right) \sum_{r\neq i} \text{area}\left(V_r^t \setminus V_r\right). \end{aligned} \tag{5.7}$$

We will calculate the area of $V_r^t \setminus V_r$ for $r \neq i$ and for $-b_i < t < 0$.

Case $V_r = S^2$.

In this case, $V_r^t = S^2$ and so area $(V_r^t \setminus V_r) = 0$.

Case $V_r \neq \emptyset$.

If $t \to -b_i$, then $V_r^t \to S^2$. We will estimate the area measure of $V_r^t \setminus V_r$ when $-b_i < t < 0$. The center of V_r is the point $A_r = \frac{b_i m_r - b_r m_i}{|b_i m_r - b_r m_i|}$. Fix an arbitrary vector u from which we are going to measure the angles θ. Given $0 \le \theta \le 2\pi$ consider the points $\gamma_r(\theta, s)$ along the geodesic originating from A_r and forming an angle θ with the vector u; s denotes geodesic arc length. The point $\gamma_r(\theta, s)$ is on the boundary of V_r if and only if the parameter $s = \tau_r = \cos^{-1}\left(\frac{b_i - b_r}{\kappa|b_i m_r - b_r m_i|}\right)$. Since $V_r \subset V_r^t$, and so the geodesic curve $\gamma_r(\theta, s)$ must intersect the boundary of V_r^t for a unique value of s with $s \ge \tau_r$. Let us denote this value of s by

$$h_r(\theta, t),$$

and so

$$\gamma_r(\theta, s) \in \partial V_r^t \quad \text{if and only if} \quad s = h_r(\theta, t).$$

Let us set

$$x_t = \gamma_r(\theta, h_r(\theta, t)).$$

Since $\gamma_r(\theta, s)$ is a geodesic curve from the point A_r to the point x_t, we have

$$h_r(\theta, t) = \arccos\left(A_r \cdot x_t\right).$$

On the other hand, the boundary of V_r^t is the collection of points where the ellipsoids $E(m_i, b_i + t)$ and $E(m_r, b_r)$ intersect. So x_t satisfies

$$\frac{b_i + t}{1 - \kappa x_t \cdot m_i} = \frac{b_r}{1 - \kappa x_t \cdot m_r},$$

which yields

$$b_r(1 - \kappa\, x_t \cdot m_i) = (b_i + t)(1 - \kappa\, x_t \cdot m_r)$$

which using the definition of A_r yields

$$A_r \cdot x_t = \frac{b_i - b_r}{\kappa\,|b_i m_r - b_r m_i|} + \frac{1 - \kappa\, x_t \cdot m_r}{\kappa\,|b_i m_r - b_r m_i|}\, t = \cos s + \frac{1 - \kappa\, x_t \cdot m_r}{\kappa\,|b_i m_r - b_r m_i|}\, t,$$

where in the last identity we used the definition of $s = \tau_r$. We are now ready to calculate the surface area of $V_r^t \setminus V_r$. Integrating in polar coordinates we obtain

$$\begin{aligned}\operatorname{area}(V_r^t \setminus V_r) &= \int_0^{2\pi} \int_{\tau_r}^{h_r(\theta,t)} \sin s \, ds \, d\theta \\ &= \int_0^{2\pi} \left(\cos \tau_r - \cos h_r(\theta,t)\right) d\theta = (-t) \int_0^{2\pi} \frac{1 - \kappa\, x_t \cdot m_r}{\kappa\, |b_i m_r - b_r m_i|} \, d\theta \\ &\le (-t)\, 2\pi \, \frac{1+\kappa}{\kappa} \, \frac{1}{|b_i m_r - b_r m_i|} \le C(\kappa, m_i \cdot m_r) \, \frac{1}{\max\{b_r, b_i\}} \, (-t), \end{aligned} \tag{5.8}$$

[5]for $-b_i < t < 0$, where $C(\kappa, m_i \cdot m_r)$ is a positive constant depending only on κ and the dot product $m_i \cdot m_r$.

Case when $V_r = \emptyset$.

Let us recall that

$$V_r = \left\{ x \in S^2 : \frac{b_i}{1 - \kappa m_i \cdot x} \le \frac{b_r}{1 - \kappa m_r \cdot x} \right\}$$

and

$$V_r^t = \left\{ x \in S^2 : \frac{b_i + t}{1 - \kappa m_i \cdot x} \le \frac{b_r}{1 - \kappa m_r \cdot x} \right\}.$$

We have that

$$V_r^t = \left\{ x \in S^2 : x \cdot \frac{(b_i + t) m_r - b_r m_i}{|(b_i + t) m_r - b_r m_i|} \ge \frac{b_i + t - b_r}{\kappa\, |(b_i + t) m_r - b_r m_i|} \right\}.$$

Let

$$g(t) = \frac{b_i + t - b_r}{\kappa\, |(b_i + t) m_r - b_r m_i|}, \quad t \in (-b_i, +\infty). \tag{5.9}$$

Since $V_r = \emptyset$ we have

$$g(0) > 1.$$

If we set $\Delta(t) = |(b_i + t) m_r - b_r m_i|$, then by calculation

$$g'(t) = \frac{b_r (b_i + t + b_r)(1 - m_i \cdot m_r)}{\kappa\, \Delta(t)^3},$$

and therefore

$$g'(t) > 0, \quad \forall t \in (-b_i, +\infty).$$

Also

$$g(t) \to -\frac{1}{\kappa}, \quad \text{when } t \to -b_i,$$

and

$$g(t) \to \frac{1}{\kappa}, \quad \text{when } t \to \infty.$$

Therefore there is a unique number $-b_i < t_0 < 0$ such that $V_r^t = \emptyset$ for $t_0 < t < 0$ and $V_r^t \neq \emptyset$ for $-b_i < t \le t_0$; this is the value for which $g(t_0) = 1$. In particular, when $t = t_0$, the set $V_r^{t_0}$ consists only of the center point $A_{r,t_0} = \frac{(b_i + t_0) m_r - b_r m_i}{|(b_i + t_0) m_r - b_r m_i|}$. We need to calculate the area of $V_r^t \setminus V_r = V_r^t$ for $-b_i < t \le t_0$. To do this we

[5] Since $m_i \neq m_r$ and have absolute value one, we have $m_i \cdot m_r \le 1 - \delta$ for some $1 > \delta > 0$. We then have $|b_i m_r - b_r m_i|^2 = b_r^2 - 2 b_r b_i m_r \cdot m_i + b_i^2 \ge b_r^2 - 2 b_r b_i (1 - \delta) + b_i^2 = (b_r - (1 - \delta) b_i)^2 + b_i^2 \delta (2 - \delta) \ge b_i^2 \delta (2 - \delta)$. Similarly, $|b_i m_r - b_r m_i|^2 \ge b_r^2 \delta (2 - \delta)$.

will use the calculation from the previous case with $V_r \rightsquigarrow V_r^{t_0}$. In fact, we now parametrize the boundary of V_r^t from the center of $V_r^{t_0}, A_{r,t_0}$. Fix a arbitrary vector u from which we are going to measure the angles θ. Given $0 \le \theta \le 2\pi$ consider the points $\gamma_r(\theta, s)$ along the geodesic originating from A_{r,t_0} and forming an angle θ with the vector u; s denotes geodesic arc length. The point $\gamma_r(\theta, s)$ is on the boundary of $V_r^{t_0}$ if and only if the parameter $s = \tau_{r,t_0} = \cos^{-1}\left(\frac{(b_i+t_0)-b_r}{\kappa\,|(b_i+t_0)m_r - b_r m_i|}\right) = 0$. Since $t \le t_0, V_r^{t_0} \subset V_r^t$, and so the geodesic curve $\gamma_r(\theta, s)$ must intersect the boundary of V_r^t for a unique value of s with $s > \tau_{r,t_0} = 0$. Let us denote this value of s by

$$h_r(\theta, t),$$

and so

$$\gamma_r(\theta, s) \in \partial V_r^t \quad \text{if and only if} \quad s = h_r(\theta, t).$$

Let us set

$$x_t = \gamma_r(\theta, h_r(\theta, t)).$$

Since $\gamma_r(\theta, s)$ is a geodesic curve from the point A_{r,t_0} to the point x_t, we have

$$h_r(\theta, t) = \arccos\left(A_{r,t_0} \cdot x_t\right).$$

On the other hand, the boundary of V_r^t is the collection of points where the ellipsoids $E(m_i, b_i + t)$ and $E(m_r, b_r)$ intersect. So x_t satisfies

$$\frac{b_i + t}{1 - \kappa x_t \cdot m_i} = \frac{b_r}{1 - \kappa x_t \cdot m_r},$$

which yields

$$b_r(1 - \kappa\, x_t \cdot m_i) = (b_i + t)(1 - \kappa\, x_t \cdot m_r) = (b_i + t_0)(1 - \kappa\, x_t \cdot m_r) + (t - t_0)(1 - \kappa\, x_t \cdot m_r)$$

which using the definition of A_{r,t_0} yields

$$\begin{aligned} A_{r,t_0} \cdot x_t &= \frac{(b_i + t_0) - b_r}{\kappa\,|(b_i + t_0)m_r - b_r m_i|} + \frac{1 - \kappa\, x_t \cdot m_r}{\kappa\,|(b_i + t_0)m_r - b_r m_i|}\,(t - t_0) \\ &= 1 + \frac{1 - \kappa\, x_t \cdot m_r}{\kappa\,|(b_i + t_0)m_r - b_r m_i|}\,(t - t_0), \end{aligned}$$

where in the last identity we used that $\tau_{r,t_0} = 0$. Integrating in polar coordinates as before we obtain for $-b_i < t \le t_0$ that

$$\begin{aligned} \operatorname{area}(V_r^t) &= \int_0^{2\pi} \int_0^{h_r(\theta,t)} \sin s\, ds\, d\theta \\ &= \int_0^{2\pi} \left(1 - \cos h_r(\theta, t)\right)\, d\theta = (t_0 - t) \int_0^{2\pi} \frac{1 - \kappa\, x_t \cdot m_r}{\kappa\,|(b_i + t_0)m_r - b_r m_i|}\, d\theta \\ &\le (t_0 - t)\, 2\pi\, \frac{1 + \kappa}{\kappa}\, \frac{1}{|(b_i + t_0)m_r - b_r m_i|} \\ &\le C(\kappa, m_i \cdot m_r)\, \frac{1}{\max\{b_i + t_0, b_r\}}\, (t_0 - t) \\ &\le C(\kappa, m_i \cdot m_r)\, \frac{1}{b_r}\, (t_0 - t) \le \frac{C(\kappa, m_i \cdot m_r)}{b_r}\, (-t), \end{aligned}$$

since $t_0 < 0$, where $C(\kappa, m_i \cdot m_r)$ is a positive constant depending only on κ and the dot product $m_i \cdot m_r$.

Therefore we obtain in all cases the desired estimate of the measure of $V_r^t \setminus V_r$ and the proposition follows. □

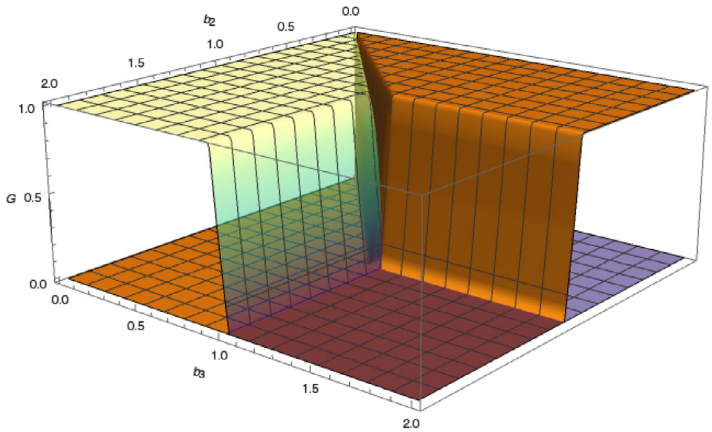


Fig. 1. Graph of the functions $G_2(\mathbf{b}) = G_{\mathcal{R}(\mathbf{b})}(m_2)$ (semitransparent) and $G_3(\mathbf{b}) = G_{\mathcal{R}(\mathbf{b})}(m_3)$ (opaque) in the $[0,2]^2$ square for the case with three unit output directions m_1, m_2, m_3 given by the homogeneous coordinates $[0:0:1]$, $[0:1:5]$, and $[1:0:5]$, respectively.

Remark 5.2. Similar estimates for $G_i(\mathbf{b})$ hold when the increment is in the variables b_r with $r \neq i$. In fact, using arguments similar to the ones used in the proof of the last proposition one can show that

$$0 \leq G_i(\mathbf{b} + t\,\mathbf{e}_r) - G_i(\mathbf{b}) \leq C_\kappa \sup_{\Omega} g \, \frac{1}{\max\{b_i, b_r\}}\, t$$

for all $0 < t < \infty$ and for each $\mathbf{b} \in \mathbb{R}^N_+, r \neq i$. Using these estimates for $r \neq i$, and the fact that in the far field the refractor measure is invariant by dilations, one can also obtain the estimate (5.6).

6. Numerical analysis

In order to see our algorithm in action, we implemented routines in the C/C++ programming language to produce some concrete numerical examples of refractors for a given output image.[6] We will assume that the function g in Definition 2.3 is constant.

Assuming $b_1 = 1$ and conservation of energy $G_1(1, b_2, \ldots, b_N) + G_2(1, b_2, \ldots, b_N) + \cdots + G_N(1, b_2, \ldots, b_N) = \text{constant}$, we have from Remark 3.2 that the map

$$(b_2, \ldots, b_N) \mapsto (G_2(1, b_2, \ldots, b_N), \ldots, G_N(1, b_2, \ldots, b_N)) \tag{6.1}$$

has a highly degenerate Jacobian in a large region of the phase space (e.g. see Fig. 1), that is, in the region $\cup_{i=1}^N F_i$ (with $b_1 = 1$). Notice that the vector $\mathbf{b}$ in (2.8) belongs to $\left(\cup_{i=1}^N F_i\right)^c$.

To evaluate numerically the output intensities $G_i(\mathbf{b}) = G_{\mathcal{R}(\mathbf{b})}(m_i)$ for any fixed $\mathbf{b} = (1, b_2, \ldots, b_N)$, we proceed as follows. We discretize Ω into a finite array of directions A. Fix a direction $\gamma \in A$ and considered the ray, denoted by ℓ_γ, from 0 having direction γ. Now all the ellipsoids $E(m_j, b_j)$ intersect the ray ℓ_γ at some point $P(j, \gamma)$. Then there is a j_γ such that the distance from $P(j, \gamma)$ to the origin is minimum, and we choose this ellipsoid. So for each $\gamma \in A$ we have an index j_γ such that the ellipsoid $E(m_{j_\gamma}, b_{j_\gamma})$ intersects the ray ℓ_γ at the point having minimum distance to the origin. Since the refractor is by definition the minimum of the polar radii of ellipsoids, then, in the direction γ, the refractor refracts into the direction m_{j_γ}. This way we have a map T from each $\gamma \in A$ into a vector m_{j_γ}. Clearly, this map T might not cover all of the

[6] All software used in our numerical investigation and graphical results can be found at http://helios.physics.howard.edu/deleo/Refractor/.

$m_1, \ldots, m_N$. We have

$$G_i(\mathbf{b}) = \frac{\#\{\gamma \in A : T(\gamma) = m_i\}}{\#\{\gamma \in A\}}. \tag{6.2}$$

To reduce computational time in the calculation of $G_i(\mathbf{b})$, it is helpful not only to keep track of how many of the directions γ get refracted in the direction m_i, but also to record $T(\gamma)$ at each $\gamma \in A$. This is because in two consecutive steps of the algorithm described in Section 4 we need to compute the values of $G_i(\mathbf{b})$ and $G_i(\mathbf{b}')$, where $\mathbf{b}$ and $\mathbf{b}'$ are two vectors that differ only in one component. In fact, suppose $\mathbf{b}$ and $\mathbf{b}'$ are successive values in the algorithm differing only in the j_0th component, and we know $T(\gamma) = m_j$ relative to $\mathbf{b}$. To evaluate $T(\gamma)$ relative to $\mathbf{b}'$ and subsequently obtaining the value of $G_i(\mathbf{b}')$, we only need to consider the ellipsoids $E(m_{j_\gamma}, b_{j_\gamma})$ and $E(m_{j_0}, b'_{j_0})$ and the distance from the origin to $P(j, \gamma)$ and $P(j_0, \gamma)$.[7] By doing so we cut the running time by a factor of N.

From (4.13), we know that the number of iterations needed to find the optimal vector $\mathbf{b}$, for which $err = \max_{2 \le i \le N} |f_i - G_i(\mathbf{b})| < \delta$, grows not faster than N/δ. We expect it not to grow slower than this as well, so that we expect a theoretical computational time of order $O(N/\delta)$. In addition, to use smaller values of δ requires increasing the value of K and therefore increasing also the number of directions in A to test (see the end of Section 4.2). Indeed, for any given A, from (6.2) the set of values that $G_i(\mathbf{b})$ takes on is finite. Therefore for δ small enough there is j_0 such that we cannot find a value of b_{j_0} for which $f_{j_0} < G_{j_0}(\mathbf{b}) < f_{j_0} + \delta$. This means that, if we want to find a $\mathbf{b}$ such that $err < \delta$, we need to increase the size of A so that $\#A > 1/\delta$. Since the loop on A leads to a running time proportional to $\#A$, in our implementation we expect a computational time of order $O(N/\delta^2)$.

For the calculations here we choose Ω as the intersection of the upper semi sphere in $\mathbb{R}^3$ with the cone with vertex at the origin and generated by the vectors $(1, 1, 2), (-1, 1, 2), (-1, -1, 2)$ and $(1, -1, 2)$. The set Ω^* is the intersection of the upper semi sphere in $\mathbb{R}^3$ with the cone with vertex at the origin and generated by the vectors $(1, 1, 5), (-1, 1, 5), (-1, -1, 5)$ and $(1, -1, 5)$. This choice of the domains Ω and Ω^* satisfies the condition (2.5) avoiding total internal reflection when $\kappa = 1/2$. Inside Ω^*, we choose the refracted directions $\{m_i\}_{1 \le i \le N}$, with $N = (n + 1)^2$, as

$$\Omega^*_N = \{[r : r' : 5n] : r, r' = -n, -n + 2, \ldots, n - 2, n; \text{ with } r, r' \text{ integers}\};$$

where $[r : r' : 2n]$ denotes the unit vector in the direction $(r, r', 2n)$. We discretize Ω into $K = (2M + 1)^2$ points having the form

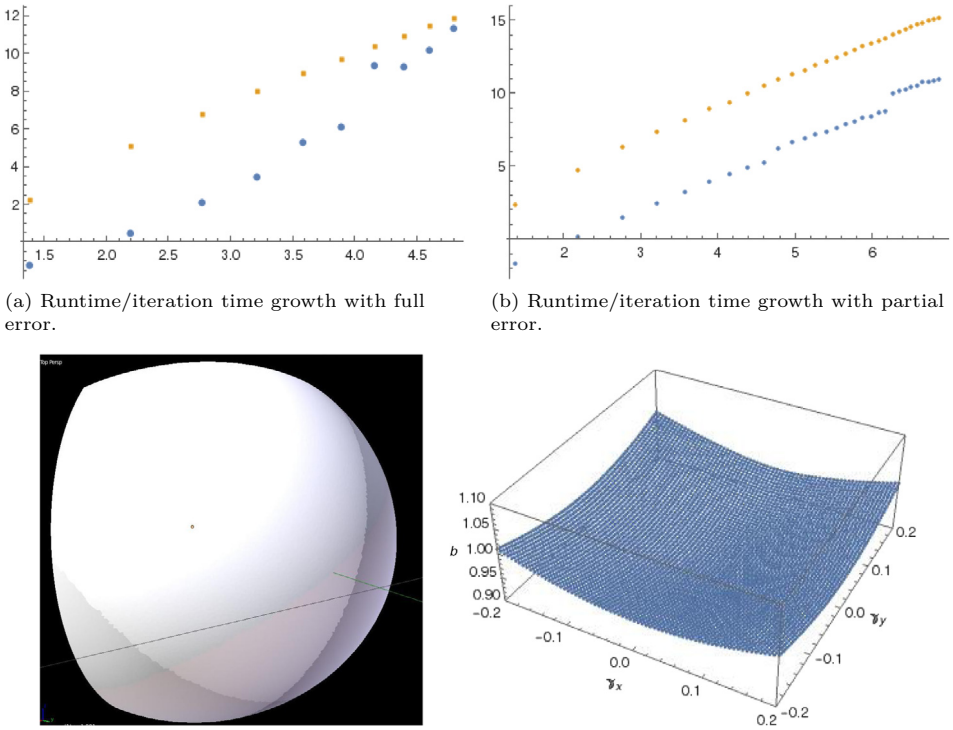
$$\Omega_K = \{[r : r' : 2M] : -M \le r, r' \le M; \text{ with } r, r' \text{ integers}\}.$$

We always start the algorithm in Section 4 with a vector in W, the set of admissible vectors, satisfying $b_1 = 1$ and $b_i = 2$ for $i \ge 2$ to obtain a vector $\mathbf{b}$ satisfying (2.8), with $\epsilon = 1/10N$, and uniform output intensities $f_i = 1/N, 1 \le i \le N$, for the directions $m_i \in \Omega^*_N$. That is, we stop our computations when

$$\max_{1 \le i \le N} \left| G_{\mathcal{R}(\mathbf{b})}(m_i) - \frac{1}{N} \right| \le \epsilon = \frac{1}{10N}.$$

While implementing the algorithm for $1 \le n \le 10$, with $\delta = \epsilon/N = 1/(10N^2)$, as in Section 4.2, our data in Fig. 2(a), show that the number of iterations ν grows roughly as $\nu(N) \simeq 0.3N^{2.8}$; although the exponent appears to decrease towards 2 as N increases. Similarly, for the running times τ we observe that $\tau(N) \simeq 0.003N^3$. These results are in perfect agreement with (4.13), according to which the growth of the number of iterations (and so of all quantities proportional to it) cannot be faster than $O(N^3)$. Note that

[7] First notice that T depends on $\mathbf{b}$. If $\mathbf{b}'$ and $\mathbf{b}$ are as in Lemma 3.3, then if $\gamma \in \mathcal{T}_{\mathbf{b}}(m_j)$, then $\gamma \in \mathcal{T}_{\mathbf{b}'}(m_\ell)$ or $\gamma \in \mathcal{T}_{\mathbf{b}'}(m_j)$ (γ no singular). Because if $\gamma \notin \mathcal{T}_{\mathbf{b}'}(m_\ell)$, then $\gamma \in \mathcal{T}_{\mathbf{b}'}(m_k)$ for some $k \ne \ell$. Then by (3.2), $\gamma \in \mathcal{T}_{\mathbf{b}}(m_k)$, and since γ is not singular, we get $k = j$.



(a) Runtime/iteration time growth with full error.

(b) Runtime/iteration time growth with partial error.

(c) Lens.

(d) Vector $\mathbf{b}$ for $N = 5041$.

Fig. 2. (a) Growth of runtime and number of iterations in our implementation of the algorithm of Section 4.2 when we minimize $|G_{\mathcal{R}(\mathbf{b})}(m_i) - 1/N|$ for *all* m_i. Blue is $(\ln N, \ln \tau(N))$; orange is $(\ln N, \ln \nu(N))$. (b) Same plot when we disregard what happens in the direction m_1, as a function of the number of output directions. (c) Detail of the lens giving rise to Descartes' image with $N = 71^2$ output directions. (d) Plot of the components of the vector $\mathbf{b}$, considered as a map $\Omega_N^* \to \mathbb{R}$, when the refractor $\mathcal{R}(\mathbf{b})$ gives the 71×71 Descartes' picture in Fig. 3(b). The set Ω_N^* is represented as the array of points $(r/5n, r'/5n), r, r' = -n, -n+2, \ldots, n$, inside the square $[-1/5, 1/5]^2$.

the evident jump in the running times when $n \geq 7$ is due to the fact that the values of δ for these cases get so small that for the algorithm to complete successfully it is necessary to use for these cases larger values of K ($M = 200$ for n up to 4, $M = 250$ for $5 \leq n \leq 6, M = 500$ for $n = 7, M = 350$ for $n = 8, M = 800$ for $n = 9$ and $M = 1100$ for $n = 10$). Such a fast growth suggests that, although the algorithm in Section 4 always yields a solution $\mathbf{b}$ after a finite number of iterations, it might take a long computational time for large values of n. For example for $n = 30$, namely $N = 961$, these data predict a running time of at least 34 days.

The running time decreases considerably if we disregard the direction m_1. In fact, in order to be able to use the algorithm in Section 4.2 with higher values of N, we disregard the intensity in the first refracted direction m_1, namely, we stop our computations when

$$\max_{2 \leq i \leq N} \left| G_{\mathcal{R}(\mathbf{b})}(m_i) - \frac{1}{N} \right| \leq \epsilon = \frac{1}{10N}.$$

As it is clear from the discussion in Section 4.2, in order to achieve this result it is enough to take $\delta \leq \epsilon$. This way, omitting $m_1, \delta = \epsilon = 10/N$ decreases only as $O(1/N)$ rather than as $O(1/N^2)$ and, accordingly, the number of iterations and the running time will be bound by $O(N^2)$ rather than $O(N^3)$. In Fig. 2(b) we show the growth of the number of iterations and running time when $1 \leq n \leq 30$, corresponding to $4 \leq N \leq 961$. In this case, the data show a growth in the number of iterations ν roughly quadratic in the number N of the output directions: $\nu(N) \simeq 2.7N^{2.05}$ and, similarly, for the running times τ we observe that $\tau(N) \simeq \alpha N^{1.9}$. This is in perfect agreement with the analytical estimate that the number of directions, with this choice of

δ, increases not faster than $O(N^2)$. Note that the coefficient α in the expression of $\tau(N)$ depends on the value of δ, and so from the duration of every single step in the program's loop to evaluate the map $T(\gamma)$, on the size of the discretization Ω_K (and therefore the number of steps in the loop above), as well as on non mathematical factors like the hardware on which the program runs[8] and the coding details of the algorithm implementation. For $1 \leq n \leq 9$ (see Fig. 2(a)) we use $\delta = 10^{-3}$ and $M = 200$ and find $\alpha \simeq 0.03$ seconds. For $n \geq 10$, the value $\delta = 10^{-3}$ is not small enough for the algorithm to reach a 10% error and so we lower it to $\delta = 2 \cdot 10^{-4}$. This change of course increases α, leading to the visible jump in the (log–log) graph of $\tau(N)$. For $10 \leq n \leq 22$ we find $\alpha \simeq 0.05s$. For $n \geq 23$, a discretization of Ω with $M = 200$ is not fine enough to allow the evaluation of the map $T(\gamma)$. So we increase M to 300, leading to a second visible jump in the graph corresponding to the larger value $\alpha \simeq 0.135$ secs. For example, with this last value of α, we get a lower bound of about 16 days for the running time in the case $N = 5000$, i.e, $n \approx 70$.

The results can be obtained faster combining this algorithm for small values of n with a quasi-Newtonian root-finding algorithm. Such methods are generalizations of the Newton method to find the root of a function without an explicit expression of its Hessian. The problem is that, as for the Newton method, quasi-Newtonian methods require a starting point where the function has a non-degenerate Jacobian and, as we already pointed out at the beginning of the section, the function (6.1) has a degenerate Jacobian in a large portion of its domain. We use the GNU Scientific Library (GSL) implementation of the quasi-Newtonian version of Powell's Hybrid method, since this method does not need an explicit Jacobian.

Therefore, as a first step we use the algorithm from Section 4.2 (disregarding the direction m_1) to find a vector $\tilde{\mathbf{b}}$ inside the region where the Jacobian is non-degenerate. And next use $\tilde{\mathbf{b}}$ as a starting point of the quasi-Newtonian algorithm to find a vector $\mathbf{b}^*$ for which the output intensities $G_i(\mathbf{b}^*)$ are "close enough" to the $f_i = 1/N$. In fact, we start by evaluating a vector $\tilde{\mathbf{b}} = (\tilde{b}_1, \ldots, \tilde{b}_{961})$ which gives *homogeneous light intensity* ($f_i = 1/N$) *in all directions (except* m_1*)* within 10% for the output array Ω^*_{961}, corresponding to $n = 30$. This computation, with $\delta = 10^{-4}$ and $M = 300$, took about 15 h. The vector $\tilde{\mathbf{b}}$ is then used as starting point by any quasi-Newtonian method to find the desired vector $\mathbf{b}^* = (b_1, \ldots, b_{961})$ such that $\max_{1\leq i\leq 961} \left|G_{\mathcal{R}(\mathbf{b})}(m_i) - 1/N\right| < \epsilon$ over the array Ω^*_{961} and any (reasonable) ϵ. With this method, it takes only about 25 min to find, starting from the vector $\tilde{\mathbf{b}}$, a vector $\mathbf{b}^*$ giving rise to a homogeneous distribution of light ($f_i = 1/N$) in *all* the directions of the array Ω^*_{961} within 10%!

We now use the vector $\mathbf{b}^*$ as a pivot in a concrete case; namely, to produce a lens that yields an image of a famous portrait of Descartes by Frans Hals on the array of refracted directions Ω^*_{14641}, corresponding to $n = 120$. The images produced with the lens using LuxRender are shown in Fig. 3 for various resolutions. First of all, we need to extract from a digital version of the original picture the output intensities $f_i, 1 \leq i \leq 14641$. For this purpose we use Imlib2, a general purpose open source C library aimed at images manipulation. Our final goal is finding a vector $\mathbf{b}$ so that the refractor $\mathcal{R}(\mathbf{b})$ satisfies the inequalities

$$\max_{1\leq i\leq 14641} \left|G_{\mathcal{R}(\mathbf{b})}(m_i) - f_i\right| \leq \min_{1\leq i\leq 14641} \{f_i\}/10. \tag{6.3}$$

Note that, since the naked eye cannot usually detect nuances of black within a complex picture, and since for large arrays the amount of light in dark spots is very low, it is actually enough for us that the max and min in (6.3) are taken only over i such that f_i is sufficiently large when N is large (f_i small corresponds with dark spots). Heuristically for this particular case we set this number to be 30% of the total number of refracted directions.

Now we evaluate the coefficients f_i corresponding to Descartes' picture for the array Ω^*_{961}. Next using $\mathbf{b}^*$, calculated in the first step, as a starting point in the quasi-Newtonian algorithm, we find the corresponding $\mathbf{b}$ giving rise to the f_i. It takes about 23 min to get a $\mathbf{b}$ such that all $G_{\mathcal{R}(\mathbf{b})}(m_i)$ are within 10% from the f_i; all but one within 1%, and 96% of them are within.1%. At this point, we consider the array Ω^*_{1681},

[8] All data in Figs. 2 and 3 are produced on an Intel Xeon 2.6 GHz CPU.

(a) Rendering (VTK 41x41). (b) Rendering (CGAL 71x71).

(c) Rendering (VTK 121x121). (d) Rendering (CGAL 121x121).

Fig. 3. Rendering of Descartes' image from 3D models generated by the graphic libraries VTK and CGAL. The rendering has been done via LuxRender, a physically accurate raytracer engine, through the modeling package Blender.

corresponding to $n = 40$, evaluate the f_i's corresponding to Descartes' picture on this array and use a standard interpolation algorithm (in concrete we use an implementation available in the GSL) to interpolate the values of $(b_1, \ldots, b_{961})$ into a new vector $(\tilde{b}_1, \ldots, \tilde{b}_{1681})$ and finally use this as starting point for the quasi-Newtonian code to find a vector $\mathbf{b} = (b_1, \ldots, b_{1681})$ giving rise to the $f_i, 1 \leq i \leq 1681$, within 10%. It takes about 28 min to find a $\mathbf{b}$ such that all $G_{\mathcal{R}(\mathbf{b})}(m_i)$ but three are within 10% from the corresponding f_i (and 98% of them is actually within 1%). From this we move to the array Ω^*_{2601}, interpolate the previous $\mathbf{b} = (b_1, \ldots, b_{1681})$ to a new $\tilde{\mathbf{b}} = (b_1, \ldots, b_{2601})$ and use it as a starting point for the quasi-Newtonian algorithm, that in about 3 h is able to find a $\mathbf{b}$ such that all $G_{\mathcal{R}(\mathbf{b})}(m_i)$ but five are within 10% from the corresponding f_i. We continue with this process by increasing n by 10 at every step until we arrive to $n = 120$, which provides the final $\mathbf{b}$ (see Fig. 3(c),(d)) so that the 70% of the $G_{\mathcal{R}(\mathbf{b})}(m_i)$ are within 10% from the corresponding f_i. The last computational step took about 2 days. The process can be continued to obtain higher resolution pictures.

7. Conclusion

We have obtained a numerical procedure to find far field refractors with arbitrary precision when the target is discrete composed of N directions, and radiation emanates from one source point. The density of the incoming radiation is assumed only bounded away from zero and infinity, and the domains Ω are general subsets of the unit sphere having boundary with surface measure zero. The procedure converges in a finite number of steps and an estimate of this number is given in terms of N, the angles between the different directions in the target, and the required approximation. To show the convergence we prove a Lipschitz estimate of the refractor map. A numerical implementation of the algorithm is carried out by using C/C++

programming language, and concrete examples of refractors for a given output image are provided. The near field case can be treated with similar methods and we will return to this problem in the near future.

Acknowledgments

C. E. G was partially supported by NSF grant DMS–1600578. The numerical part in Section 6 was developed first on the computational cluster of the Italian National Institute for Nuclear Physics (INFN), and lately on the cluster of the College of Arts and Sciences at Howard University.

References

[1] M. Born, E. Wolf, Principles of Optics, Electromagnetic Theory, Propagation, Onterference and Diffraction of Light, seventh (expanded) ed., Cambridge University Press, 1959.
[2] K. Brix, Y. Hafizogullari, A. Platen, Designing illumination lenses and mirrors by the numerical solution of Monge–Ampère equations, J. Opt. Soc. Amer. A 32 (11) (2015) 2227–2236. http://arxiv.org/pdf/1506.07670.pdf.
[3] K. Brix, Y. Hafizogullari, A. Platen, Solving the Monge–Ampere equations for the inverse reflector problem, Math. Models Methods Appl. Sci. 25 (5) (2015) 803–837. https://arxiv.org/pdf/1404.7821v2.pdf.
[4] L.A. Caffarelli, S.A. Kochengin, V. Oliker, On the numerical solution of the problem of reflector design with given far-field scattering data, Contemp. Math. 226 (1999) 13–32.
[5] Pedro Machado Manhães Castro, Quentin Mérigot, Boris Thibert, Far-field reflector problem and intersection of paraboloids, Numer. Math. (2015) 1–23.
[6] C.E. Gutiérrez, Refraction Problems in Geometric Optics, in: Lecture Notes in Mathematics, vol. 2087, Springer-Verlag, 2014, pp. 95–150.
[7] C.E. Gutiérrez, Qingbo Huang, The refractor problem in reshaping light beams, Arch. Ration. Mech. Anal. 193 (2) (2009) 423–443.
[8] C.E. Gutiérrez, H. Mawi, The far field refractor with loss of energy, Nonlinear Anal. TMA 82 (2013) 12–46.
[9] C.E. Gutiérrez, A. Sabra, The reflector problem and the inverse square law, Nonlinear Anal. TMA 96 (2014) 109–133.
[10] C.E. Gutiérrez, F. Tournier, The parallel refractor, Dev. Math. 28 (2013) 325–334.
[11] C.E. Gutiérrez, F. Tournier, Regularity for the near field parallel refractor and reflector problems, Calc. Var. Partial Differential Equations 54 (1) (2015) 917–949.
[12] Jun Kitagawa, An iterative scheme for solving the optimal transportation problem, Calc. Var. Partial Differential Equations 51 (1–2) (2014) 243–263.
[13] Xi-Nan Ma, N. Trudinger, Xu-Jia Wang, Regularity of potential functions of the optimal transportation problem, Arch. Ration. Mech. Anal. 177 (2) (2005) 151–183.