Domains which are integrable close to the boundary and close to the circular ones are ellipses

Illya Koval *

V.N. Karazin Kharkiv National University
November 25, 2021

Abstract

The Birkhoff conjecture says that the boundary of a strictly convex integrable billiard table is necessarily an ellipse. In this article, we consider a stronger notion of integrability, namely integrability close to the boundary, and prove a local version of this conjecture: a small perturbation of an ellipse of small eccentricity which preserves integrability near the boundary, is itself an ellipse. We generalize the result of [6], where integrability was proven only for specific values, proving its main conjecture. In particular, we show that (local) integrability near the boundary implies global integrability. One of the crucial ideas in the proof consists in analyzing Taylor expansion of the corresponding action-angle coordinates with respect to the eccentricity parameter, as well as using irrationality to prove the main non-degeneracy condition.

1 Introduction

A mathematical billiard is a dynamical system, first proposed by G.D. Birkhoff as a playground, where “the formal side, usually so formidable in dynamics, almost completely disappears and only the interesting qualitative questions need to be considered” in [3].

Let $\Omega$ be a strictly convex $C^r$ domain in $\mathbb{R}^2$ with $r > 3$. Let $x$ be a point in the boundary $\partial \Omega$ and $\varphi$ is angle of a direction $V$ with the clockwise tangent to $\partial \Omega$ at $x$. Let $M := \{(x, \varphi) : x \in \partial \Omega, \varphi \in (0, \pi)\}$. Then, one can consider a billiard map $f : M \to M$, where $M$ consists of unit vectors with foot $x$ on $\partial \Omega$ and with inward direction $v$. The map reflects the ray from the boundary of the domain elastically, i.e. the angle of the incidence equals the angle of reflection.

This dynamical system has simple local dynamics, however, its study turns out to be really complex and has many important open questions. One group of "direct" questions is to pick a domains and analyse the properties of the billiard in them. For example, can they be chaotic, have a positive metric entropy, an open set of periodic points, etc. A different way to study the billiards is an indirect one, see e.g. [4]. One can define a length spectrum of

*illyakoval2001@gmail.com
a domain, by looking at perimeters of all periodic orbits. The closure of the union is called a
length spectrum. How much of information is encoded into this spectrum? This question is
studied for example in [13]. It turns out the length spectrum is connected with other spectra
of the domain such as Laplace spectrum. The famous inverse problem of hearing the shape
of a drum [8] in mathematical terms is to determine a domain from its Laplace spectrum.

In this paper we analyse so called integrable billiards. For example, if Ω is an ellipse,
then the billiard map is integrable, meaning all of dynamics can be described in a relatively
simple way. A natural question then arises, the one asked by Birkhoff in [3] and by Poritsky
in [12]. It is then formulated in the following conjecture:

Conjecture 1.1. There are no other examples of integrable billiards.

Despite its simple-looking statement, the question still remains open. Various methods
were developed to attack this problem. For example, in [11] the authors have proven, that
if the curvature of the domain vanishes at one point, then it cannot be integrable.

The strongest non-perturbative result is due to Misha Bialy in [2]:

Theorem 1. ([2]) If the phase space of the billiard ball map is globally foliated by continuous
invariant curves which are not null-homotopic, then it corresponds to a billiard in a disc.

Of course, one should rigorously define what integrable means. Many definitions were
introduced. For example, one can say that the map \( M \) is integrable if there exists a smooth
integral of motion near the boundary.

Here, we study one of the most common definitions of integrability, i.e. preservation of
a smooth foliation by caustics near the boundary. Specifically, we study the preservation of
rational caustics.

Definition 1.1. A smooth convex curve \( \Gamma \subset \Omega \) is called a caustic, if whenever a trajectory
is tangent to it, then it remains tangent after each reflection.

If \( \Omega \) is a disk, then its caustics are concentric circles by a Lemma of Poncelet. For an
ellipse, its caustics are co-focal ellipses. Note, that if one considers tangent directions, a
caucstic defines a natural map on \( \partial \Omega \) onto itself, as such it has a rotation number. We define

Definition 1.2. We say that \( \Gamma \) is an integrable rational caustic for the billiard map in \( \Omega \),
if the corresponding (non-contractible) invariant curve \( \hat{\Gamma} \subset M \) consists of periodic points; in
particular, the corresponding rotation number is rational.

Particularly, the rotation number \( \omega \in (0, 1) \), however we would only consider \( \omega \in (0, 1/2] \)
since others correspond to reverse dynamics on the same caustic. Caustics near the boundary
correspond to small rotation numbers, so we would study those. All rational caustics are
present in a disc, while other ellipses lack a caustic with \( \omega = 1/2 \).

In the recent years, there have been multiple articles on this topic, concerning a local
case, meaning \( \Omega \) is a small deformation of an ellipse. For example, in [1], authors prove that
if locally caustics with rotation numbers \( 1/2 \) for \( q \geq 3 \) are preserved near an ellipse with small
eccentricity, then \( \Omega \) is also an ellipse. Later [9] generalized this, studying ellipses with other
eccentricities. However, these results rely, for example, on caustics with rotation number
1/3 and 1/4, and those are not near the boundary.

Our goal is to study domains with caustics only near the boundary \( \partial \Omega \).
Definition 1.3. Let $q_0 \geq 2$. If the billiard map, associated to $\Omega$ admits integrable rational caustics with rotation numbers $\frac{p}{q}$ for all $0 < \frac{p}{q} < \frac{1}{q_0}$, we say that $\Omega$ is $q_0$-rationally integrable.

$q_0$-rationally integrable domains near ellipses of small eccentricities are studied in details in [6]. Our main result is a proof of their Conjecture 1.9 that such ellipses are rigid.

Theorem 1. For any integer $q_0 \geq 3$, there exist $e_0 = e_0(q_0) \in (0, 1), m = m(q_0), n = n(q_0) \in \mathbb{N}$, such that the following holds. For each $0 < e < e_0$ and $c \geq 0$, there exists $\varepsilon = \varepsilon(e, c, q_0)$, such that any $q_0$-rationally integrable $C^m$-smooth domain $\Omega$, whose boundary is $C^n - \varepsilon$-close to an ellipse $E_{e,c}$, is itself an ellipse.

Here, an estimate on values of $m$ and $n$ can be obtained and computed using Theorem 4.

Note that the condition $\frac{p}{q} < \frac{1}{q_0}$ is a linear bound on $p$ with respect to $q$. However, for some problems this may not suffice. So, in Appendix D, we generalize this result to any polynomial bound on $p$ over $q$, like $\frac{p^2}{q} < \frac{1}{q_0}$, for example.

1.1 Outline of the proof

The main idea behind the proof is the following. Each deformation can be described by a function on a circle. We establish the connection between Fourier harmonics of that deformation function and necessary condition of caustics preservation for ellipses with small eccentricities, deriving explicit formulas. Then we prove that this connection is non-degenerate using irrationality of some coefficients. We consider the following expansion of a deformation in elliptic coordinates (2.1):

$$\partial \Omega = \{(\mu_0 + \varepsilon \mu_1(\varphi) + O(\varepsilon^2), \varphi), \varphi \in [0, 2\pi]\}.$$  \hspace{1cm} (1.1)

Here, if $\mu_0$ is a constant value, it describes an ellipse, while $\mu_1$ is a leading part of perturbation. Then, for each rational caustic with $\omega = p/q$, we can write down a necessary condition that a deformation preserves it as follows:

$$F_{p/q}(\mu_1) = 0,$$  \hspace{1cm} (1.2)

where $F_{p/q}$ is some linear functional of $\mu_1$. Hence, if $\Omega$ is a $q_0$-rationally integrable domain, then $F_{p/q}(\mu_1) = 0$ for all $0 < \frac{p}{q} < \frac{1}{q_0}$. In order to prove the theorem, we need to find a system of functionals that is complete, namely, find $\frac{p_i}{q_i}, i \geq 1$ such that

$$F_{p_i/q_i}(\mu_1) = 0, i \geq 1 \implies \mu_1 \equiv \text{const.}$$  \hspace{1cm} (1.3)

This system would be easier to study in a Fourier basis, so we interpret it as a system of linear equations on the Fourier coefficients of $\mu_1$, that we would call $a_j$ and $b_j$.

A finite dimensional reduction

Let $\mu_1(\varphi) = a_0 + \sum_k a_k \cos(k\varphi) + b_k \sin(k\varphi)$ be the Fourier expansion of $\mu_1$. It can be shown that the system (1.2) can be reduced to essentially a finite-dimensional system. More precisely, for some large $q_1$ and $q > q_1$ it turns out that $F_{p/q}(\mu_1) = 0 \implies a_q \approx b_q \approx 0$ and
has a perturbative expression. Thus, \( F_{p/q}(\mu_1) = 0 \) for all \( q > q_1 \), up to an error annihilates all Fourier coefficients of \( \mu_1 \) with indices \( q \). So, our main goal would be to find a system (1.3) that is complete. For that, after a reduction to a finite-dimensional system we will study its coefficients and using their irrationality prove that the determinant of the system is nonzero, so the system (1.3) is complete.

**A finite dimensional nondegeneracy**

The main difficulty we will face is with Fourier coefficients whose indices \( \leq q_0 \), since vanishing of the other \( a_q, b_q \)'s is closely related to vanishing of the respective linear functionals \( F_{1/q}(\mu_1) = 0 \). This connection is used in [1]. For harmonics with small indices, however, we lack a perturbative relation vanishing of \( F_{1/q}(\mu_1) = 0 \) leading to an almost vanishing of \( a_q \) and \( b_q \), so we are proposing the following method. We study the dependence of other functionals \( F_{p/q} \)'s on the \( q \)-th harmonic. The main idea of this paper is to find a finite collection \( \frac{p_i}{q_i}, i = 1, \ldots, N(q_0) \) and prove that this dependency is non-degenerate, i.e.

\[
F_{p_i/q_i}(\mu_1) = 0,
\]

implies \( a_q = b_q = 0 \) for \( |q| \leq q_0 \), then show that along with \( F_{1/q}(\mu_1) = 0 \) for \( |q| > q_0 \) this concludes (1.3).

After expanding in eccentricity, one can link this dependence to some matrix, whose non-degeneracy we have to study. As stated, we would use their irrationality and the algebraic field theory to prove it. We will discuss this later in the introduction.

We determine a matrix associated with \( F_{p_i/q_i}(\mu_1) = 0 \) for a finite collection of \( i \)'s whose non-degeneracy we need to study. The latter means that we need to study the relation between a caustic’s preservation and elliptic harmonics. However, elliptic coordinates are not the best to study caustics, since every caustic has a coordinate system, where its dynamics is very simple. This coordinate system is called action-angle coordinates. So, we need to express a preservation condition in action-angle harmonics and, hence, to convert them into elliptic harmonics.

To do that, we first obtain important relations between elliptic and action-angle coordinates on an ellipse, corresponding to caustic \( C_\lambda \). We call those \( \varphi \) and \( \theta \), respectively.

These coordinates are formally introduced in the next section, see (2.1). Elliptic coordinates are coordinates \( \mu, \varphi \) on a plane, such that coordinate lines with fixed \( \mu \) are co-focal ellipses, and with \( \varphi \) - hyperbolae. If one fixes \( \mu = \mu_0 \) these coordinates induce a parametrisation \( \varphi \) of the ellipse. Action-angle coordinates, on the other hand, are unique to each caustic. They are introduced, such that the billiard coming from the point \( \theta \) and tangent to the caustic, hits the boundary again at \( \theta + 2\pi \omega \).

Let \( e \) be eccentricity of an ellipse \( \mathcal{E} \) with semi-major axis \( a \). Each co-focal ellipse can be parameterised by a natural parameter \( \lambda \) with \( \lambda = 0 \) being \( \mathcal{E} \), see (2.3). Then there is a Taylor expansion of \( \varphi \) as a function of \( \theta \) in \( e \) as a small parameter

\[
\varphi(\theta, \lambda, e) = \theta + \sum_{j=1}^{N} \varphi_j(\theta) \frac{a^{2j}e^{2j}}{(a^2 - \lambda^2)^j} + O\left(e^{2N+2}\right),
\]

(1.5)
Figure 1: Two coordinates on an ellipse

where

\[ \varphi_j(\theta) = \sum_{l=1}^{j} \beta_{j,l} \sin(2l\theta). \]  

(1.6)

We prove an equality for diagonal coefficients, namely:

**Theorem 2.** For any positive integer \( j \geq 1 \) we have

\[ \beta_{j,j} = \frac{2}{24j^4}. \]  

(1.7)

The proof uses an explicit formula (2.6), that expresses \( \theta \) as an elliptic integral in \( \varphi, \lambda, \) and \( e \). We expand this relation in \( e \) at \( e = 0 \) and invert it. This proof is discussed in Section 3.

We then prove an equality for the change of Fourier coefficients. As we have

**Lemma 1.1.** (Lemma C.2 [6]) Let

\[ \mu(\varphi) = a_0 + \sum_{k=1}^{+\infty} a_k \cos(k\varphi) + b_k \sin(k\varphi) \]  

(1.8)

and \( \mu(\varphi) \in C^m(\mathbb{T}) \). Then, for \( N \leq m - 1 \), the expansion of function \( \mu(\varphi(\theta, \lambda, e)) \) with respect to \( e \) up to order \( O(e^{2N+2}) \) is

\[ \mu(\varphi(\theta, \lambda, e)) = \mu(\theta) + \sum_{j=1}^{N} P_j(\theta) \frac{a^{2j} e^{2j}}{(a^2 - \lambda^2)^j} + O(e^{2N+2}), \]  

(1.9)
where the functions $P_j(\theta)$ are of the form

$$P_j(\theta) = \sum_{k=1}^{+\infty} \sum_{l=-j}^{j} \xi_{j,l}(k)(a_k \cos((k+2l)\theta) + b_k \sin((k+2l)\theta)). \quad (1.10)$$

We obtain a formula for these diagonal coefficients, as well.

**Theorem 3.** Let $j,k \geq 1$, $|l| \leq j$ be a triple of integers, then $\xi_{j,l}(k)$ are polynomials in $k$ of degree at most $j$ and

$$\xi_{j,j}(k) = \frac{1}{2^j j!} k(k+1) \ldots (k+j-1) = 2^{-4j} \binom{k+j-1}{j} \quad (1.11)$$

$$\xi_{j,-j}(k) = (-1)^j \frac{1}{2^j j!} k(k-1) \ldots (k-j+1) = (-1)^j 2^{-4j} \binom{k}{j} \quad (1.12)$$

The system (1.2) expands into (2.8), that contains exactly $\xi_{j,l}(k)$. The diagonal coefficients, however, are the most important ones, since they describe the main dependence between the Fourier coefficients of the deformation of an ellipse and caustic preservation, see (5.9). So, these coefficients arise in the matrix, described above. The anti-diagonal ones can be useful, since they represent the main term of dependence between the $l$-th harmonic of deformation and a rational caustic $\frac{p}{q}$ for $l > q$ (diagonal coefficients represent $l < q$), see (2.8). Using these formulas, one can further study caustic preservation in general, as well as other questions, related to ellipses.

These relations are obtained using Theorem 2 and Lemma 1.1. By substituting some basic harmonics instead of $\mu$, we obtain a formula, connecting $\beta$ and $\xi$. Then, it reduces to a combinatorial equality for 2 described cases. We derive these formulas in the first part. In the second part, we study the matrix of (1.4), and the associated system of equations on harmonics, since we already know all the coefficients. We use irrationality and algebraic field theory to prove non-degeneracy.

### 1.2 Algebraic structure and Vandermonde reduction

Let’s give a couple of examples of motivated by algebraic nature of the matrix (1.4). The first example would be simple, while the second one would be more related to our problem.

**Example 1.** Prove that the matrix

$$\begin{pmatrix} \sqrt{2} & 5 & 2 \\ 4 & 3\sqrt{2} & 7 \\ 2 & 8 & 1 \end{pmatrix} \quad (1.13)$$

is non-degenerate.
Of course, one could just compute the determinant of the matrix approximately and prove it. However, we can do it in a more conceptual way. We substitute $z$ instead of $\sqrt[3]{2}$. Then, we can find the determinant to be the polynomial from $z$ of degree 2 over rationals. If our original matrix had been degenerate, $\sqrt[3]{2}$ would have been a root of this polynomial. This, however, would mean that our polynomial divides the minimal polynomial of $\sqrt[3]{2}$ over rationals. It is impossible, of course, since the said minimal polynomial, $z^3 - 2$, is of degree 3, so it cannot divide polynomial of degree 2. So, the matrix is non-degenerate.

So, in this example we used irrationality of $\sqrt[3]{2}$ and algebraic field theory to prove non-degeneracy of a matrix with rational numbers.

**Example 2.** Let $\alpha_j = e^{2\pi j i}, j = 1, 2, 3, 4$. Prove that the following matrix is non-degenerate:

$$\begin{pmatrix}
3 & 7 & 0 & 0 \\
4 & 1 & 2 & 0 \\
1 & \alpha_1 & \alpha_1^2 & \alpha_1^3 \\
1 & \alpha_2 & \alpha_2^2 & \alpha_2^3
\end{pmatrix}.$$  

(1.14)

This problem is fairly similar to the problems we will soon encounter. In particular, one could interpret the latter two lines in a matrix as representing preservation of caustics with $\omega = 1/5$ and $2/5$, respectively. Moreover, the method of handling this problem is very similar to the method in the main proof.

Here, one could also compute the determinant approximately. Alternatively, one could substitute $\alpha_2 = \alpha_1^2$ into the matrix and use a method from previous example for $\alpha_1$ instead of $\sqrt[3]{2}$:

$$\begin{pmatrix}
3 & 7 & 0 & 0 \\
4 & 1 & 2 & 0 \\
1 & z & z^2 & z^3 \\
1 & z^2 & z^4 & z^6
\end{pmatrix}.$$  

(1.15)

This will run into some problems though, because the resulting polynomial from $z$ will be of degree 8, while the minimal polynomial of $\alpha_1$ over rationals, that is $z^4 + z^3 + z^2 + z + 1$, only has degree 4. Since $8 > 4$, the determinant can divide the minimal polynomial. Let’s propose a viable option.

Recall that a number $m$ is a primitive root module $n$ if $m^j$ through all the residues, except 0 modulo $n$. For example, 2 is a primitive root modulo 5.

**Lemma 1.2.** The matrix (1.15) is non-degenerate, since 2 is a primitive root modulo 5.

**Proposition 1.1.** Let $z = e^{2\pi i/p}$ with $p > 3$ being prime and 2 is a primitive root modulo $p$. Then the matrix (1.15) is non-degenerate.

**Remark 1.** Notice that 2 is a not a primitive root modulo 7 and the method of proof of this proposition does not apply.
To prove both statements we propose a method, which we call a Vandermonde reduction. We will reduce the matrix (1.14) to the Vandermonde matrix (1.21).

**Proof.** The proof is by contradiction. Suppose the determinant is zero. Then, we know that
\[(1, \alpha_2, \alpha_2^2, \alpha_3^2) \in \text{Lin}\left( (3, 7, 0, 0), (4, 1, 2, 0), (1, \alpha_1, \alpha_1^2, \alpha_1^3) \right). \tag{1.16} \]
We already know that the determinant of (1.15) divides \(z^4 + z^3 + z^2 + z + 1\). Since 5 is prime, it has roots at all the unity roots, except \(z = 1\), for example, at \(z = \alpha_2\). Substitute it into (1.15):
\[
\begin{vmatrix}
3 & 7 & 0 & 0 \\
4 & 1 & 2 & 0 \\
1 & \alpha_2 & \alpha_2^2 & \alpha_2^3 \\
1 & \alpha_2^2 & \alpha_2^4 & \alpha_2^6
\end{vmatrix} = \begin{vmatrix}
3 & 7 & 0 & 0 \\
4 & 1 & 2 & 0 \\
1 & \alpha_2 & \alpha_2^2 & \alpha_2^3 \\
1 & \alpha_4 & \alpha_4^2 & \alpha_4^4
\end{vmatrix} = 0. \tag{1.17}
\]
\[(1, \alpha_4, \alpha_4^2, \alpha_4^3) \in \text{Lin}\left( (3, 7, 0, 0), (4, 1, 2, 0), (1, \alpha_2, \alpha_2^2, \alpha_2^3) \right) \subset \text{Lin}\left( (3, 7, 0, 0), (4, 1, 2, 0), (1, \alpha_1, \alpha_1^2, \alpha_1^3) \right). \tag{1.18} \]
Further substituting \(z = \alpha_4\) and so on leads us to
\[(1, \alpha_2^k, \alpha_2^{2k}, \alpha_3^{3k}) \in \text{Lin}\left( (3, 7, 0, 0), (4, 1, 2, 0), (1, \alpha_1, \alpha_1^2, \alpha_1^3) \right) \tag{1.19} \]
Now, since \(2^k\) goes through all the residues modulo 5 (here we use that 2 is a primitive root), we get:
\[(1, \alpha_j, \alpha_j^2, \alpha_j^3) \in \text{Lin}\left( (3, 7, 0, 0), (4, 1, 2, 0), (1, \alpha_1, \alpha_1^2, \alpha_1^3) \right), j = 1, 2, 3, 4. \tag{1.20} \]
This would mean that all these four vectors are linearly dependent on each other. Consequently,
\[
\begin{vmatrix}
1 & \alpha_1 & \alpha_1^2 & \alpha_1^3 \\
1 & \alpha_2 & \alpha_2^2 & \alpha_2^3 \\
1 & \alpha_3 & \alpha_3^2 & \alpha_3^3 \\
1 & \alpha_4 & \alpha_4^2 & \alpha_4^3
\end{vmatrix} = 0. \tag{1.21}
\]
This is of course impossible, since we have a Vandermonde of \(\alpha_1, \alpha_2, \alpha_3, \alpha_4\), and it is nonzero, since all of them are distinct from each other. This means, that the original determinant couldn’t have been zero, so the system is complete.

\[\blacksquare\]

Similar algorithm is described in Section 5 of this paper to prove the main result. The main differences is that instead of 5 we take arbitrary prime number \(q\), instead of roots of unity we have their real parts (cosines) and instead of determinant we study the rank of the matrix.
1.3 Selection of rotation numbers for (1.4)

It can be noted that we used several properties of number 5 in the second example. First, it was important that 5 is a prime number, since otherwise the minimal polynomial would have been different. Moreover, we needed to get all the roots in the Vandermonde matrix, so effectively we have used that 2 is a primitive root modulo 5. For example, if we had chosen 7 instead of 5, we would have only connected 3 roots: $\alpha_1, \alpha_2$ and $\alpha_4$, since $\alpha_4^2 = \alpha_1$. We wouldn’t have a way of proving (1.20) for $j = 3, 5$ and 6. So, in this case the method wouldn’t work.

Note that 2 is a primitive root modulo $q$ if the minimal subgroup of $\mathbb{F}_q^*$, containing $\{1, 2\}$ is $\mathbb{F}_q^*$ itself. Since this example is similar to our problem, we give the following definitions:

**Definition 1.4.** A prime number $q$ is said to be $q_0$-A-good, if the minimal subgroup of $\mathbb{F}_q^*$, symmetrical by negation, and containing elements $1, 2, \ldots, \left[\frac{2}{q_0}\right]$ is $\mathbb{F}_q^*$.

Also define $q_0$-B-good numbers, they will also be of use.

**Definition 1.5.** A prime number $q$ is said to be $q_0$-B-good, if the minimal subgroup of $\mathbb{F}_q^*$, symmetrical by negation, and containing elements $1, 3, \ldots, 2\left[\frac{2}{q_0}\right] - 1$ is $\mathbb{F}_q^*$.

The condition of being symmetrical by negation is due to studying cosines instead of exponents, see Section 5.

These definitions are used to study denominators of caustic rotation numbers. In particular, we receive the following estimates for Theorem 1 for odd $q_0$:

**Theorem 4.** Let $q_0 = 2k_0 - 1$. Denote $M_0$ to be $(k_0 - 1)$-st both $q_0$-A-good and $q_0$-B-good number. Then in Theorem 1 one can choose

\[ n = 2M_0, \quad m = 28M_0 \quad (1.22) \]

We can give an example of possible $m$ and $n$. Let’s say, if $q_0 = 9$ one could pick $n = 58$ and $m = 812$. We provide a table of various $q_0$ and $m$ and $n$ for them near the end of the paper – Table 3.

We also note, that we can lower these estimates for given $q_0$ using direct computation, since we already know all the matrix coefficients, however we will not do this here.

2 Strategy of the proof

The main result of the paper (Theorem 1) is that for any $q_0 \geq 3$ and a small $q_0$-integrable perturbation of an ellipse of small eccentricity is an ellipse. This means that for large $q_0$ we can use only caustics of small rotation numbers $p/q \in (0, 1/q_0)$. This will lead us to study a number theoretic nature of the condition $F_{p/q}(\mu_1) = 0$, (1.2). We start with a review of [6].
2.1 Several claims from Huang-Kaloshin-Sorrentino [6]

This work is based on the structure of action-angle coordinates for ellipses studied in [6]. In order to explain the key ingredients of our contribution, we first need to present several claims from this paper.

In [6], they expand the equation for caustic preservation with respect to the eccentricity $e$ of an ellipse at $e = 0$. These ideas lead to a number of sufficient conditions on a deformation (1.2), specifically its Fourier harmonics in elliptic coordinates, to preserve caustics. Then, these conditions $F_{p/q}(\mu_1) = 0$ are used to prove that Fourier harmonics with small indices are tiny. In [6], though, they use caustics whose rotation numbers $p/q$ are non-small. Showing that Fourier harmonics with small indices are tiny using only caustics with small rotation numbers is the main problem of the paper.

First of all, every ellipse has semi-major and semi-minor axis $a$ and $b$, as well as eccentricity $e = \sqrt{(a^2 - b^2)/a^2}$ and linear eccentricity $c = e/a$. Elliptic coordinates on a plane take the following form, see the left figure of Figure 1:

\[
\begin{align*}
    x &= c \cosh \mu \cos \varphi \\
    y &= c \sinh \mu \sin \varphi
\end{align*}
\]  

(2.1)

When $\mu = \mu_0 = \cosh^{-1}(1/e)$, $\varphi \in [0, 2\pi]$ gives a parametrization of a boundary of an ellipse, called an elliptic one. We will also study a perturbation of a domain using these coordinates and a periodic function $\mu(\varphi)$. From now on, we have that

\[ \partial \Omega = \mathcal{E}_{e,c} + \mu(\varphi). \]  

(2.2)

We also consider a family of caustics – co-focal ellipses $C_\lambda$ parametrized by a parameter $\lambda$

\[
C_\lambda = \left\{ (x, y) \in \mathbb{R}^2 : \frac{x^2}{a^2 - \lambda^2} + \frac{y^2}{b^2 - \lambda^2} = 1 \right\}, \quad 0 < \lambda < b.
\]  

(2.3)

We shall also use another parameterization of caustics $k_\lambda = \frac{a^2 - b^2}{a^2 - \lambda^2}$ and a rotation number $\omega$. We also use elliptic integrals of the first kind, namely

\[
F(\varphi; k) = \int_0^{\varphi} \frac{1}{\sqrt{1-k^2\sin^2 \tau}} d\tau; \quad K(k) = F\left(\pi/2; k\right).
\]  

(2.4)

Then, the following formula holds:

\[
\omega(\lambda, e) = \frac{F(\arcsin(\lambda/b); k_\lambda)}{2K(k_\lambda)}.
\]  

(2.5)

We also write the boundary parametrization induced by caustic $C_\lambda$, denoted by $\theta$, such that the orbit starting at $\theta_0$ and tangent to $C_\lambda$ hits the boundary at $\theta_0 + 2\pi\omega_\lambda$. It is called an action-angle parametrization. Then, we have the following relation:

\[
\theta(\varphi, e, \lambda) = \frac{\pi}{2} \frac{F(\varphi, k_\lambda)}{F(\pi/2, k_\lambda)}
\]  

(2.6)

We shall use the following results from Section 3, [6].
Lemma 2.1. (Lemma 3.2 [6]) There exists $C > 0$ such that for each $e \in [0, \frac{1}{2}]$ and $\omega \in (0, \frac{1}{2})$, we have

$$|\lambda(e, \omega) - b \sin \omega \pi| \leq Ce^2$$

(2.7)

Lemma 2.2. (Proposition 3.6 [6]) Let $0 < \frac{p}{q} \in \mathbb{Q} \cap (0, 1)$ and assume $\Omega$ admits an integrable rational caustic of rotation number $p/q$. Let $N \in \mathbb{Z}_+$ such that $q > 2N$. Then:

$$a_q + \sum_{n=1}^{N} \sum_{l=-n}^{n} \xi_{n,l}(q - 2l)a_{q-2l}(a^2 - \frac{\lambda_p^2}{p/q})^n = \frac{a^{2n}e^{2n}}{(a^2 - \frac{\lambda_p^2}{p/q})^n} = O \left(||\mu||_{C_{N+1}}e^{2N+2} + \lambda^{-1}_{p/q}q^{7}||\mu||_{C^{1}}^2\right).$$

(2.8)

This formula connects harmonics of deformation $a_k$ and caustic preservation, it is (3.8) in [6]. It is used several times in the following sections. One should note here that harmonics $a_k$ and $b_k$ for $k > 2$ are studied separately and identically, so we will just focus on $a_k$ in this paper. Moreover, due to this we will call $a_k$ with even and odd indices ”even” and ”odd harmonics” respectively.

In sections 4, 5, 6 [6] a case of small $q_0$’s is studied. There, they use equations (2.8) to prove the statement about harmonics with small indices. These sections are not really relevant to our proof, but one can use them as an example for the general case.

Section 7 [6] is arguably the most relevant to us. There, the general case is studied and the fact that harmonics with small indices are tiny is proven under some assumptions. The first part is headed by Lemma 7.1, acting as a base for further study. Then, they prove the needed lemma by induction, proving that harmonics with smaller and smaller indices are tiny. This breaks up into two parts: one for odd indices and one for even ones, since equations (2.8) separate them.

In Section 8 [6], there are lemmata covering the change of parametrization of the boundary of an ellipse an its effect on large harmonics.

Finally, Section 9 [6] covers the proof of the main theorem. The strategy behind the proof is discussed, harmonics are classified into several groups, and they are all proven to be small. They also choose the constants $m$ and $n$, representing needed smoothness of the deformation.

Moreover, the paper has also several appendices.

Appendices A and B further study the geometry of an ellipse and will not be expanded upon here.

Appendix C introduces $\xi$ and $\beta$ – objects that would be very important to us. Particularly, we dedicate the first half of this paper to study them. These coefficients turn up in equations (3.8), and so they connect harmonics and caustics.

In Appendix D matrix inversion is studied. This appendix is used in Section 7 of [6], so it would be valuable to us.

2.2 First half of the proof

This part expands on Appendix C of [6]. In particular, we obtain several new formulas, like Theorem 2 and 3. Ideas there are pretty independent from [6], and from other parts of this
work. However, of course Appendix C, and some formulas from Section 3 are still used, in
particular (3.1), (3.6) and Lemma 1.1.

This half is broken up into 2 sections, corresponding for studying $\beta$ and $\xi$ separately. We
first study $\beta$, since $\xi$ is analyzed using it. The main idea of this part would be to follow
these simple steps. First, we write down the definition of a given object, try to connect
it with some explicit formula, calculate some coefficients and then simplify to obtain some
combinatorial identity. This part will also heavily exploit the diagonal nature of the process,
especially in trigonometry.

The results of this half are used in the next one. However, they may have other applica-
tions. For example, one can study other types of caustic rigidity, and these formulas would
be useful, since they connect harmonics of deformation and caustic preservation for small $e$.
Moreover, this part does not rely on the motion of caustics, so it can be of use in other
fields.

\section{Second half}

\subsection{Introduction}

The second part is the main part of the proof. There, we prove Theorem 1. However, whole
sections of our proof of this part are identical or similar to ones in [6], so we need to explain
them. Firstly, we need to address, what problem in [6] do we tackle.

In Section 7 the authors of [6] ran into some difficulties. In particular, they obtained a
system, consisting of equations (2.8) and they needed to prove that this system was complete:
otherwise there could be some deformation preserving all the caustics. However, the matrix
of this system lacked structure and contained $\xi$, that did not have a formula, so it was
impossible to do. So, for their proof to work, they requested for this system to be complete.

Since we obtain the formula for $\xi$ in the first half, we already solved one of the problems.
Then, to prove that the system is complete, we change it a little, and then we use algebraic
field theory, since it has some irrational coefficients. This helps us prove the completeness
of the system and we are able to drop the requirement.

So, we share the same base of proof with [6], however the main part of the proof is created
here. Since ideas concerning this main part are located in Section 7 of [6], we mainly focus
on this section later. However, we also add some ideas in other parts in order for our proof
to work. Let us explain our whole proof here.

\subsection{Proof}

Since sections 2 and 3 of [6] introduce several facts about ellipses, we would also use them
here. Let’s describe them here.

We start with an ellipse $\mathcal{E}$ with an eccentricity $e$. We consider this ellipse to be close to
the circle, meaning $e$ is near 0. As usual, we introduce $a, b, c$ as parameters of an ellipse. We
also consider an elliptic parametrization of the boundary $\varphi$, and caustics $C$. Each caustic
has its own rotation number $\omega, \lambda$, and action-angle parametrization of the boundary $\theta$. We
call rational caustics ones with rational rotation number, we associate it with $\frac{p}{q} = \omega$, where
\( p/q \) is in its lowest terms. We will study rational caustics with \( p/q < 1/q_0 \) for some \( q_0 \). There are connections between these objects, once again explained in [6].

Next, we consider a deformation \( \mu(\varphi) \) of the boundary. This deformation preserves rational caustics with \( p/q < 1/q_0 \), and it is small and smooth. Particularly,

\[ \mu(\varphi) \in C^m(\mathbb{T}), \quad ||\mu(\varphi)||_{C^n} < \varepsilon, \]  

for some \( m,n,\varepsilon(e,c) \). Then, we have that

\[ \mu(\varphi) = a_0 + \sum_{k=1}^{+\infty} a_k \cos(k\varphi) + b_k \sin(k\varphi). \]  

Ideally, we want to show, that every one of these harmonic coefficients is small, that would lead us to contradiction.

After introducing all of that, we follow through Section 3 of [6] and obtain (2.8) for \( N \leq m - 1, q > 2N, p/q \) in its lowest terms, and \( p/q < 1/q_0 \). This equation is an expansion over \( e \) of a condition on preservation of caustic of rotation number \( p/q \).

After all of that, we go to Section 7 of [6]. The ideas of this step are explained in Section 5 of our work. There, we prove (5.4), that is the main result of this part. We start by proving Lemma 5.1, that is a modification of Lemma 7.1 in [6] and has the same idea of the proof. Then, we study odd and even nodes separately. In both cases, we use an induction. In it, we transform (2.8) into (5.9) for odd indices and into (5.54) for even ones. This transformation is done, once again, similarly to [6]. The goal of this lemma is to remove the non-leading terms from the sum, so we can study the terms of the same magnitude over \( e \).

Let’s say a couple of words about the general idea of this part. We have the association between harmonic \( a_q \) and caustic with rotation number \( p/q \). This connection is obvious in (2.8), since the first term of an expansion of condition on preservation of caustic with rotation number \( p/q \) is precisely \( a_q \). So, if we want to preserve caustic \( p/q \) for example, harmonic \( a_q \) needs to be small. We can use this idea for \( q > q_0 \), and as such, we can make every harmonic with large index small. It is not that simple, of course, but the behavior of large index harmonics is studied with great detail in [6].

Harmonics \( a_0,a_1,a_2,b_1,b_2 \) are elliptic harmonics, meaning their alternation results in deformation, that would be close to another ellipse. This result is used to prove that they are small as well. Precisely, we are able to change the starting ellipse a little, so that these harmonics become small. This algorithm is discussed in Section 9, as well as Appendix B of [6].

This leaves us to prove that harmonics \( a_3,a_4,\ldots,a_{q_0} \) are small (\( b_j \) are analogous). Since we do not have a preserved caustic for them, we are forced to seek another way. We study them by looking at high-order terms of (2.8) for caustics with \( p/q < 1/q_0 \), but with \( q \) not greatly larger than \( q_0 \). For example, if \( q_0 = 5 \), then one can consider caustics with rotation number \( \frac{1}{7}, \frac{1}{9}, \frac{1}{11}, \frac{2}{13}, \frac{2}{15} \), as is done in [6]. Then, it can be shown that equations (2.8) result in 6 linear equations, depending on \( a_3,a_5,a_7,a_9,a_{11},a_{13} \). Since the number of equation is the same to the number of variables, one would generally assume that the system is complete, so all these harmonics can be shown to be small.

Of course, to prove that the system is complete, one should study its coefficients and cannot just rely on the number of variables or equations. All the previous ideas of this half
were already present in [6], and now we propose some new ones. One of the systems that we study, has the structure (5.13).

Here, we come to the main part of the proof: using algebraic field theory. As you can see, (5.13) has only rational coefficients, except some cosines of rational multiples of $\pi$. These cosines are not rational, and their algebraic structure is known. We are particularly interested in their minimal polynomials over rationals, so we propose an algorithm, exploiting it, since if the system is not complete, it places heavy restrictions on algebraic structure of the coefficients.

To use this connection, we are forced to study caustics with prime denominators in their rotation number, and this results in considering caustics with higher denominators overall, compared to [6]. Due to it, we need to increase the smoothness requirements of the deformation $m$ and $n$. We control these requirements, by introducing smoothness constants $M_1, M_2, M_3, M_4, M_5$, that are studied later. An algorithm of finding these constants for every $q_0$ is given, as well as the table of them for several values of $q_0$. These constants always exist, and are monotone over $q_0$, so one can use constants for $q_0 = 199$ to study $q_0 = 127$, for example.

Moreover, increasing considered caustic denominators and smoothness requirements results in minor changes to the following lemmata and ideas of [6]’s Section 7. We discuss those changes in Section 5, but since the proof would be analogous, we do not write it down. So, the main result of this part is (5.4), and we move on to the last two sections of [6].

Section 8 of [6] is filled with interesting statements, and accept all of them. We don’t need to change them, because they are still useful to us.

Finally, we come to Section 9, that is the proof of the main result. We follow the same proof. First of all, we pick the closest ellipse to the deformation and prove that it can’t be the one. To do that, we prove that all harmonics of the deformation can be made small. We classify all the harmonics into 4 types: elliptic motions $(a_0, a_1, a_2, b_1, b_2)$, low-order modes $(a_3, a_4, \ldots, a_{q_0})$, intermediate-order modes $(a_{q_0+1}, a_{q_0+2}, \ldots, a_{[N]})$ and high-order $(a_{[N]+1}, \ldots)$.

Elliptic motions are made small, using lemma B.1 of [6], and some change of starting ellipse. We do not change anything concerning them.

To prove that other three types are small, we should introduce new smoothness parameters $m$ and $n$, described in Theorem 1. For the proof to work, they need to satisfy some requirements. Specifically, for intermediate and high-order modes to be made small, one needs the following relations, described in [6]. They say that

$$\frac{2(m - n - 3)}{m - n} \geq \frac{72}{37}, \quad \frac{4(m - n - 3)}{m - 1} \geq \frac{18}{5}. \quad (2.11)$$

For that to happen, they choose $n = 3q_0$, $m = 40q_0$. However, they also need these constants to prove that small-order harmonics are tiny. Specifically, they need

$$a_k \leq C_{q_0}e^{2||\mu_\varepsilon||C^n}; \quad k = 3, \ldots, q_0. \quad (2.12)$$

Of course, since we only have (5.4), we cannot guarantee this to happen with $n = 3q_0$. So, we then choose

$$n = M_1; \quad m = 14M_1. \quad (2.13)$$
This way, the same proof in [6] can be used for small-order harmonics, and since relations (2.11) hold for this choice, same proof could also be used for other 2 groups. With that, we have proven Theorem 1.

So, in our work we share results form Sections 2, 3, 8 of [6], as well as all the appendices. We propose some very minor changes to Section 9, just changing some constants. In Section 7, we add new ideas, and change some statements due to higher smoothness requirements. These ideas are discussed in Section 5 of our work.

3 Formula for $\beta_{j,j}$

3.1 Introduction

Let’s say we have an ellipse. It has a rather simple formula, however it has various parametrizations of the boundary. Of particular interest in this part would be two of them. Our first parametrization is an elliptic one. We call it $\varphi$, according to [6]. There is also another one – the action-angle parametrization $\theta$, corresponding to caustic $C_\lambda$. Both of them go from 0 to $2\pi$. We want to write a function $\varphi(\theta)$, changing coordinates. This would obviously be dependent on the parameters of an ellipse ($e$ and $a$), as well as on $\lambda$. If we want to expand this function by terms of $e$, we will get the following result, according to [6]:

$$\varphi(\theta, \lambda, e) = \theta + \sum_{j=1}^{N} \varphi_j(\theta) \frac{a^{2j} e^{2j}}{(a^2 - \lambda^2)^j} + O(e^{2N+2}),$$

(3.1)

where

$$\varphi_j(\theta) = \sum_{l=1}^{j} \beta_{j,l} \sin(2l\theta).$$

(3.2)

Here, we will give the proof of the last formula and, along with it, Theorem 2.

3.2 Inverse dependency

Now we will carefully prove Theorem 2 and (3.2). In our proof, we will assume that these relations hold for all smaller $j$.

Let’s introduce a few more objects. First of all, we introduce

$$s = k_\lambda^2 = \frac{a^{2j} e^{2j}}{(a^2 - \lambda^2)^j}$$

(3.3)

to simplify (3.1). Next, to avoid confusion (since we have (3.2) in [6]), we denote

$$f(\varphi, s) = \theta(\varphi, e, \lambda), \ g(\theta, s) = \varphi(\theta, e, \lambda).$$

(3.4)

(3.1) takes the form

$$g(\theta, s) = \theta + \sum_{j=1}^{N} \varphi_j(\theta) s^j + O(s^{N+1}).$$

(3.5)
We do not have a formula for \( g(\theta, s) \), however we do have it for \( f(\varphi, s) \). By [6], we have:

\[
f(\varphi, s) = \frac{\pi}{2} F(\varphi, \sqrt{s}) = \frac{\pi}{2} \int_0^\varphi \frac{1}{\sqrt{1-s\sin^2 \tau}} d\tau.
\] (3.6)

We are only working with \( s \) in some neighborhood of 0.

To express the derivative of \( g \), we firstly need to describe them for \( f(\varphi, s) \). Let’s start by differentiating it over \( s \) several times. Since \( f \) is a fraction, we’ll compute the derivatives of the numerator. It is easy to check by induction, that

\[
\frac{\partial^k}{\partial s^k} \int_0^\varphi \frac{1}{\sqrt{1-s\sin^2 \tau}} d\tau|_{s=0} = (2k-1)!! \frac{2}{2^k} \int_0^\varphi (1 - s \sin^2 \tau)^{-\frac{1}{2} - k} \sin^2 k \tau d\tau.
\] (3.7)

Since we are expanding everything at \( e = 0 \), meaning \( s = 0 \), we substitute it here:

\[
\frac{\partial^k}{\partial s^k} \int_0^\varphi \frac{1}{\sqrt{1-s\sin^2 \tau}} d\tau|_{s=0} = (2k-1)!! \frac{2}{2^k} \int_0^\varphi \sin^2 k \tau d\tau.
\] (3.8)

Now we are evaluating the integral on the right-hand side. We use the formula for the cosine of the double-angle, open up brackets and then express the powers of cosine through sum of cosines of multiples using the formula for cosine product:

\[
\int_0^\varphi \sin^2 k \tau d\tau = \int_0^\varphi \left(1 - \cos 2\tau\right)^k d\tau = \\
= \int_0^\varphi \left(D_0 \cos(0\tau) + D_1 \cos(2\tau) + \ldots + D_k \cos(2k\tau)\right) d\tau = \\
= D_0 \varphi + \frac{D_1}{2} \sin(2\varphi) + \frac{D_2}{4} \sin(4\varphi) + \ldots + \frac{D_k}{2^k} \sin(2k\varphi).
\] (3.9)

Here and later, we consider \( D_i \) to be some real constants, that can differ in different expressions.

We have found the general structure of this derivative, that we can use later. Since we are only claiming a formula for diagonal \( \beta \), we don’t need to find exact formulas for all those coefficients \( D \). However, we still need one of those coefficients – namely \( D_k \).

\( \cos(2k\tau) \) can only appear via the coefficient with \( \cos(2\tau)^k \) after opening up brackets. Moreover, using the formula for cosine product \( k - 1 \) times, one gets that it arises with coefficient \( 2^{-k+1} \) from \( \cos(2\tau)^k \).

Bringing it all together, we get

\[
\frac{\partial^k}{\partial s^k} \int_0^\varphi \frac{1}{\sqrt{1-s\sin^2 \tau}} d\tau|_{s=0} = D_0 \varphi + D_1 \sin(2\varphi) + D_2 \sin(4\varphi) + \ldots + D_k \sin(2k\varphi),
\] (3.10)

where \( D_0, \ldots, D_k \) are some coefficients and

\[
D_k = \frac{(2k-1)!!}{2^k} \frac{1}{(-2)^k 2^{k-1} 2^k} = \frac{(-1)^k (2k-1)!!}{2^{3k} k!}.
\] (3.11)
We would also need to differentiate \( f \) over \( \varphi \). So let’s say \( k > 0 \) and \( l = 2r \) is an even natural number (for odd \( l \) everything will be similar). We just differentiate (3.10) \( l \) times. Then,

\[
\frac{\partial^{l+k}}{\partial \varphi^l \partial s^k} \int_0^\varphi \frac{1}{\sqrt{1 - s \sin^2 \tau}} d\tau |_{s=0} = D_0 \sin(0 \varphi) + D_1 \sin(2 \varphi) + D_2 \sin(4 \varphi) + \ldots + D_k \sin(2k \varphi),
\]

where \( D_0, \ldots, D_k \) are some other coefficients and

\[
D_k = (-1)^r \frac{(-1)^k (2k - 1)!!}{2^{3k} k} (2k)^l.
\]

Let’s also say a couple of words about the whole fraction in (3.6). We can rewrite it in the following form:

\[
\frac{\pi}{2} \int_0^\varphi \frac{1}{\sqrt{1 - s \sin^2 \tau}} d\tau \left( \int_0^{\pi/2} \frac{1}{\sqrt{1 - s \sin^2 \tau}} d\tau \right)^{-1}.
\]

We will differentiate it over \( s \) using the product rule. The derivatives of the left part would give rise to the sine sums mentioned above, while the right part will just give rise to functions over \( s \), evaluated at \( s = 0 \). So, the derivatives of this expression over \( s \) would still be of form (3.10).

Also, in order to receive \( \sin(2k \varphi) \) in the sum, one would need to differentiate the left part \( k \) times, and the right part – zero times. Since the right part at \( s = 0 \) is equal to \( \frac{2 \pi}{\pi} \), the value of \( D_k \) will remain the same.

So, we get:

\[
\frac{\partial^k}{\partial s^k} f(\varphi, s) = D_0 \varphi + D_1 \sin(2 \varphi) + D_2 \sin(4 \varphi) + \ldots + D_k \sin(2k \varphi),
\]

where \( D_0, \ldots, D_k \) are some coefficients and

\[
D_k = \frac{(-1)^k (2k - 1)!!}{2^{3k} k}.
\]

Now, consider the following function

\[
f(2\pi, s) - f(0, s).
\]

It is equal to \( 2\pi \) since it is just \( \theta \) of the same point after one rotation. That means that

\[
\frac{\partial^k}{\partial s^k} f(\varphi, s)
\]

is a periodic function over \( \varphi \). Then, \( D_0 = 0 \), so, we get that

\[
\frac{\partial^k}{\partial s^k} f(\varphi, s) = D_1 \sin(2 \varphi) + D_2 \sin(4 \varphi) + \ldots + D_k \sin(2k \varphi).
\]

We also have that for natural \( l = 2r \)

\[
\frac{\partial^{l+k}}{\partial \varphi^l \partial s^k} f(\varphi, s) = D_1 \sin(2 \varphi) + D_2 \sin(4 \varphi) + \ldots + D_k \sin(2k \varphi),
\]

with

\[
D_k = (-1)^r \frac{(-1)^k (2k - 1)!!}{2^{3k} k} (2k)^l.
\]
3.3 Differentiating an expression

3.3.1 Introducing an expression

Now we will use the derivatives of $f$ to find the structure of derivatives of $g$.

First of all, let’s write the following equality:

$$\theta = f(g(\theta, s), s). \quad (3.22)$$

We want to differentiate this equality $j$ times over $s$ and then substitute $s = 0$ to relate derivatives of $g$ and derivatives of $f$. We have that

$$\frac{d^j}{ds^j} f(g(\theta, s), s) = 0. \quad (3.23)$$

If we open up the left-hand side, we will receive a sum of products. Let’s see how they are obtained.

3.3.2 Expression’s terms

At the beginning, we only have $f(g(\theta, s), s)$. Since both parameters of $f$ depend on $s$, we use the rule of differentiating multi-variable function, and we get the sum of two values. If we differentiate $f$ by the second parameter, we will only get

$$\frac{\partial}{\partial s} f(g(\theta, s), s). \quad (3.24)$$

However, when we differentiate by the first one, we get

$$\frac{\partial}{\partial \varphi} f(g(\theta, s), s) \frac{\partial}{\partial s} g(\theta, s). \quad (3.25)$$

If we differentiate it the second time, we will be differentiating a product, so either we will increase the index of derivative of $g$ by 1 and get

$$\frac{\partial}{\partial \varphi} f(g(\theta, s), s) \frac{\partial^2}{\partial s^2} g(\theta, s), \quad (3.26)$$

either we will differentiate $f$. This once again means that we will either increase the degree of derivative of $f$ over $s$ and get

$$\frac{\partial^2}{\partial \varphi \partial s} f(g(\theta, s), s) \frac{\partial}{\partial s} g(\theta, s), \quad (3.27)$$

or we will get a new term by differentiating $f$ by $\varphi$:

$$\frac{\partial^2}{\partial \varphi^2} f(g(\theta, s), s) \frac{\partial}{\partial s} g(\theta, s) \frac{\partial}{\partial s} g(\theta, s). \quad (3.28)$$
Similar things will occur for large $j$. Each time, we will either increase the degree of differentiation over $s$ for some of the terms, or we will introduce a new term with $g$ and increase the degree of differentiation of $f$ over $\varphi$. Then the following values will take part in our sum:

$$\frac{\partial^{j+x_0}}{\partial \varphi^j \partial s^{x_0}} f(g(\theta, s), s) \prod_{i=1}^l \frac{\partial^{x_i}}{\partial s^{x_i}} g(\theta, s).$$

(3.29)

Note that the degree of differentiation over $\varphi$ and the number of terms in the product is the same, since to introduce a new term, we need to differentiate by $\varphi$. Also note, that

$$x_0 + x_1 + \ldots + x_l = j$$

(3.30)

since each time we differentiate over $s$, we increase one of those by one.

Also, in all of this $x_0$ is a non-negative integer, while all the others $x_i$ are natural numbers.

### 3.3.3 Calculating coefficients

Now let’s work out the exact coefficients by those values. Since we can permute $(x_1, \ldots, x_l)$ between themselves, we need to order them somehow, or else we will have to worry about terms with $x_1 = 1$, $x_2 = 3$, when dealing with $x_1 = 3$, $x_2 = 1$. For that reason, let’s assume that the term with $x_1$ was the earliest one (meaning that the first differentiation of $f$ over $\varphi$ produced the term, that would later be differentiated another $x_1 - 1$ times), the term with $x_2$ came second and so on. This way, we remove an ability to permute.

So, let’s compute the coefficient. Of the $j$ differentiations we need to first choose $x_0$ times to differentiate $f$ over $s$. There are exactly $\binom{j}{x_0}$ ways to do it. Then, we know, when we first differentiated $f$ over $\varphi$. So, we already know the time of obtaining the term, that would later become the one with $x_1$. Then, we need to choose another $x_1 - 1$ times from the $j - x_0 - 1$ differentiations left to differentiate the term with $x_1$. It can be done $\binom{j-x_0}{x_1-1}$ ways. We continue with the $x_2$ and so on. So, the coefficient behind the value with $(x_0, x_1, \ldots, x_l)$ is

$$\binom{j}{x_0} \binom{j-x_0-1}{x_1-1} \binom{j-x_0-x_1-1}{x_2-1} \ldots \binom{j-x_0-\ldots-x_{l-1}-1}{x_l-1}.$$  

(3.31)

Rewriting it in factorials, one gets

$$\frac{j!}{x_0!(x_1-1)!(x_2-1)! \ldots (x_l-1)! (j-x_0) (j-x_0-x_1) \ldots (j-x_0-x_1-\ldots-x_{l-1})}.$$  

(3.32)

Now we will substitute known values into this expression, when $s = 0$. We transform (3.29) into

$$\frac{\partial^{j+x_0}}{\partial \varphi^j \partial s^{x_0}} f(\theta, 0) \prod_{i=1}^l \frac{\partial^{x_i}}{\partial s^{x_i}} g(\theta, 0).$$  

(3.33)

Let’s consider an example for $j = 3$. Then, the whole expression will take the following form:

$$f_{ss}(\theta, 0) + 3f_{s\varphi}(\theta, 0)g_{s}(\theta, 0) + 3f_{\varphi s}(\theta, 0)g_{ss}(\theta, 0) + f_{\varphi}(\theta, 0)g_{sss}(\theta, 0) + 3f_{\varphi\varphi}(\theta, 0)g_{ss}(\theta, 0)g_{s}(\theta, 0) + 2f_{\varphi\varphi}(\theta, 0)g_{ss}(\theta, 0)g_{s}(\theta, 0) + f_{\varphi\varphi}(\theta, 0)g_{ss}(\theta, 0)g_{s}(\theta, 0)g_{s}(\theta, 0) + f_{\varphi}(\theta, 0)g_{ss}(\theta, 0)g_{s}(\theta, 0)g_{s}(\theta, 0) = 0.$$  

(3.34)
In this sum, the following terms are present from left to right:

\[(x_0 = 3, l = 0); (x_0 = 2, x_1 = 1, l = 1); (x_0 = 1, x_1 = 2, l = 1); (x_0 = 0, x_1 = 3, l = 1);
(x_0 = 1, x_1 = 1, x_2 = 1, l = 2); (x_0 = 0, x_1 = 2, x_2 = 1, l = 2); (x_0 = 0, x_1 = 1, x_2 = 2, l = 2); (x_0 = 0, x_1 = 1, x_2 = 1, x_3 = 1, l = 3).\] (3.35)

### 3.3.4 \( x_0 \) is natural

First of all, let’s assume that \( x_0 \) is also a natural number. Let’s again only work with even \( l = 2r \) (it would also be similar for odd \( l \)). Then, we can use (3.20) in (3.33) to substitute derivatives of \( f \). Since \( x_0 > 0 \), every other \( x_i < j \), and since

\[
\frac{\partial x_i}{\partial s_i} g(\theta, 0) = \varphi_{x_i}(\theta)(x_i)!, \tag{3.36}
\]

and since we know that (3.2) holds for \( j = x_i \) we get that the value of (3.33) is actually just a product of sums of the type (A.4).

Since \( l \) is even, this product has an odd number of terms, we will open up the brackets and turn the products of sines into the sum of sines, using (A.3). So, this whole product will also be of form (A.4). Let’s find the maximal frequency of sine in this expression. It cannot be greater than the sum of the maximal frequencies in \( \frac{\partial^{l+x_0}}{\partial \varphi^l \partial s_0} f(\theta, 0) \) and in all \( \frac{\partial x_i}{\partial s_i} g(\theta, 0) \). By (3.29) the maximal frequency in \( \frac{\partial^{l+x_0}}{\partial \varphi^l \partial s_0} f(\theta, 0) \) is \( 2x_0 \), whereas by (3.2) the maximal frequency in \( \frac{\partial x_i}{\partial s_i} g(\theta, 0) \) is \( 2x_i \). Since \( 2x_0 + 2x_1 + \ldots + 2x_l = 2j \), we get that

\[
\frac{\partial^{l+x_0}}{\partial \varphi^l \partial s_0} f(\theta, 0) \prod_{i=1}^{l} \frac{\partial x_i}{\partial s_i} g(\theta, 0) = D_1 \sin(2\theta) + D_2 \sin(4\theta) + \ldots + D_j \sin(2j\theta), \tag{3.37}
\]

where \( D_1, \ldots, D_j \) are once again some numbers. We don’t need any of them, except \( D_j \). Let’s combine all of our progress to find it.

First, of all, \( \sin(2j\theta) \) can only be obtained in the sum, when considering only highest frequency terms in every bracket. They are \( \sin(2x_0\theta), \sin(2x_1\theta), \ldots, \sin(2x_l\theta) \). If we consider their product \( \sin(2x_0\theta)\sin(2x_1\theta)\ldots\sin(2x_l\theta) \) and use (A.3) \( r \) times, we will get that \( \sin(2j\theta) \) will arise with coefficient \( \frac{(-1)^r}{4^r} \).

\( \sin(2x_0\theta) \) goes with coefficient

\[
(-1)^r \frac{(-1)^x_0(2x_0 - 1)!!}{2^{3x_0}x_0} (2x_0)^l, \tag{3.38}
\]

due to (3.21), while \( \sin(2x_i\theta) \) go with

\[
\beta_{x_i, x_i}!, \tag{3.39}
\]

due to (3.36) and (3.2) in their respective brackets. Combining it all together, we get

\[
D_j = \frac{(-1)^r}{4^r} \frac{(-1)^x_0(2x_0 - 1)!!}{2^{3x_0}x_0} (2x_0)^l \prod_{i=1}^{l} \beta_{x_i, x_i}!. \tag{3.40}
\]
Now we substitute \( \beta_{x_i,x_i} \) using inductive step and simplify it:

\[
D_j = \frac{(-1)^x (2x_0 - 1)!}{x_0 2^x} 2^{x_0} (2x_0)^t \prod_{i=1}^l (x_i - 1)!. 
\] (3.41)

Now we multiply it by (3.32) to get

**Proposition 3.1.** In the sum of (3.23) for \( s = 0 \) in the term with \((x_0, x_1, \ldots, x_l)\) \(\sin(2j\theta)\) goes with the coefficient

\[
\frac{(-1)^x (2x_0)}{x_0 2^x} 2^{x_0} (2x_0)^t \frac{j!}{(j - x_0)(j - x_0 - x_1) \ldots (j - x_0 - x_1 - \ldots - x_{l-1})} 
\] (3.42)

It will also hold for odd \( l \).

### 3.3.5 \( x_0 \) is zero

Now we move on to \((x_0, x_1, \ldots, x_l)\), where \( x_0 = 0 \). Notice that it means that we differentiate \( f \) only by \( \varphi \) at \( s = 0 \). But if \( s = 0 \) then \( f(\varphi, 0) = \varphi \), so \( \partial f(\theta, 0) / \partial \varphi \) is 0, if \( l > 1 \). Notice that \( l \neq 0 \) in this case. That means, for nonzero result, we need \( l = 1 \). It means we only have \( x_0 \) and \( x_1 \), and for them we have \( x_0 = 0, x_1 = j \). Then (3.33) will take the following form:

\[
\frac{\partial}{\partial \varphi} f(\theta, 0) \frac{\partial^j}{\partial s^j} g(\theta, 0). 
\] (3.43)

\( \partial / \partial \varphi f(\theta, 0) = 1 \), and \( \partial^j / \partial s^j g(\theta, 0) = j! \varphi_j(\theta) \), so it will take the form

\[
 j! \varphi_j(\theta). 
\] (3.44)

Now observe, what (3.23) at \( s = 0 \) gives us. Left hand side has a sum of products, each of whom, except one, takes the form (3.37). That means, that the last one takes the same form. So, we have that

\[
 j! \varphi_j(\theta) = D_1 \sin(2\theta) + D_2 \sin(4\theta) + \ldots + D_j \sin(2j\theta). 
\] (3.45)

This immediately proves (3.2). To find \( \beta_{j,j} \) we only need to add up coefficients with \( \sin(2j\theta) \).

### 3.4 Evaluating a sum

We have that

\[
j! \beta_{j,j} + \sum_{l=0}^{j-1} \sum_{x_0,\ldots,x_l \in \mathbb{N}} (-1)^{x_0} \frac{(2x_0)!}{x_0 2^x} \frac{(2x_0)^l}{(j - x_0)(j - x_0 - x_1) \ldots (j - x_0 - x_1 - \ldots - x_{l-1})} = 0. 
\] (3.46)

So, we only need to prove that

\[
\frac{2}{j} \sum_{l=0}^{j-1} \sum_{x_0,\ldots,x_l \in \mathbb{N}} (-1)^{x_0} \frac{(2x_0)!}{x_0 2^x} \frac{(2x_0)^l}{x_0 (j - x_0)(j - x_0 - x_1) \ldots (j - x_0 - x_1 - \ldots - x_{l-1})} = 0. 
\] (3.47)
Twist this sum a little:

\[
\frac{2}{j} + \sum_{x_0=1}^{j} \frac{(-1)^{x_0}(2x_0)}{x_0(j-x_0)} \sum_{l=0}^{j-1} (2x_0)^l \sum_{x_1,\ldots,x_l \in \mathbb{N}} \frac{1}{(j-x_0-x_1)\ldots(j-x_0-\ldots-x_l)} = 0. 
\]

(3.48)

Now let’s understand the inner sum. Note that the numbers in the denominator are natural, decreasing, and less than \( j - x_0 \). Note that these are the only requirements. That means that

\[
\sum_{x_1,\ldots,x_l \in \mathbb{N}} \frac{1}{(j-x_0-x_1)\ldots(j-x_0-\ldots-x_l)} = \frac{e_{j-x_0-1-l}(1,2,\ldots,j-x_0-1)}{(j-x_0-1)!}.
\]

(3.49)

where \( e \) is an elementary symmetric polynomial. Using the formula of coefficients of polynomial by its roots, we get that (3.48) reduces to

\[
\frac{2}{j} + \sum_{x_0=1}^{j} \frac{(-1)^{x_0}(2x_0)}{x_0(j-x_0)} \frac{(2x_0)(2x_0+1)\ldots(2x_0+(j-x_0-1))}{(j-x_0-1)!} = 0.
\]

(3.50)

This reduces to

\[
\frac{2}{j} + 2 \sum_{x_0=1}^{j} (-1)^{x_0} \frac{(x_0+j-1)!}{x_0!x_0!(j-x_0)!} = 0.
\]

(3.51)

And this to

\[
\sum_{x_0=0}^{j} (-1)^{x_0} \binom{j}{x_0} \binom{x_0+j-1}{j-1} = 0.
\]

(3.52)

That is true, see for example [10].

Then we have proven all the needed facts about the \( \varphi_j(\theta) \) and we can move on to the next step.

### 4 Formula for \( \xi_{j,j}(k) \)

#### 4.1 \( \xi_{j,l}(k) \) definition

Coefficients \( \xi_{j,j}(k) \) (we focus on proving (1.11) here), and in general \( \xi_{j,l}(k) \) arise when one considers change of parametrization of an ellipse with small eccentricity. In particular, they describe how would the Fourier coefficients of the deformation change, when one changes the parametrization. They are introduced in [6] in Lemma [C.1], that is Lemma 1.1 here. We will find coefficients \( \xi_{j,j}(k) \) using this lemma.

Let \( \mu(\varphi) = \cos(k\varphi) \), where \( k \in \mathbb{N} \) (we can scale it to avoid problems with definition). This means that \( a_k = 1 \), all the other \( a_l, b_l \) are equal to 0. Then,

\[
P_j(\theta) = \sum_{l=-j}^{j} \xi_{j,l}(k) \cos((k+2l)\theta).
\]

(4.1)
Notice that the $P_j(\theta)$ is a sum of integer-frequency cosines and the one with highest frequency is $\cos((k+2j)\theta)$ for $l = j$. There are no cosines with the same frequency or with the negative frequency of that, so that cosine is orthogonal to all the others with respect to $\theta \in [-\pi, \pi]$. So,

$$\int_{-\pi}^{\pi} P_j(\theta) \cos((k+2j)\theta) d\theta = \frac{1}{\pi} \int_{-\pi}^{\pi} \xi_{j,j}(k) \cos((k+2j)\theta) d\theta = \xi_{j,j}(k) \int_{-\pi}^{\pi} \cos^2((k+2j)\theta) d\theta = \xi_{j,j}(k) \pi. \tag{4.2}$$

Then, we find

$$\xi_{j,j}(k) = \frac{1}{\pi} \int_{-\pi}^{\pi} P_j(\theta) \cos((k+2j)\theta) d\theta. \tag{4.3}$$

### 4.2 Expanding to get $P_j(\theta)$

Now, we will express $P_j(\theta)$.

First, we use Taylor expansion of cosine:

$$\mu(\varphi(\theta, e)) = \cos(k \varphi(\theta, e)) = \cos(k(\theta + (\varphi(\theta, e) - \theta))) = \sum_{l=0}^{+\infty} \frac{k^l}{l!} \cos^l((\varphi(\theta, e) - \theta)) = \sum_{l=0}^{+\infty} \frac{k^l}{l!} \cos^l(k \theta + \frac{l\pi}{2})(\varphi(\theta, e) - \theta)^l. \tag{4.4}$$

Then, we substitute the expansion for $\varphi(\theta, \lambda, e)$:

$$\mu(\varphi(\theta, e)) = \sum_{l=0}^{+\infty} \frac{k^l}{l!} \cos \left( k \theta - \frac{l\pi}{2} \right) \left( \sum_{i_1=1}^{N} \varphi_{i_1}(\theta) \frac{a_i^2 e^{2i}}{(\lambda^2 - \lambda^2)^l} + O(e^{2N+2}) \right)^l. \tag{4.5}$$

We are interested in $P_j(\theta)$, that is the coefficient by the $\frac{a_i^2 e^{2j}}{(\lambda^2 - \lambda^2)^l}$ in this formula. Let’s open up the $l$-power:

$$\mu(\varphi(\theta, e)) = \sum_{l=0}^{+\infty} \frac{k^l}{l!} \cos \left( k \theta - \frac{l\pi}{2} \right) \sum_{i_1=1}^{N} \ldots \sum_{i_l=1}^{N} \varphi_{i_1}(\theta) \ldots \varphi_{i_l}(\theta) \frac{a_i^2 e^{2l}}{(\lambda^2 - \lambda^2)^l} + O(e^{2N+2}), \tag{4.6}$$

where $I = i_1 + \ldots + i_l$. We need the coefficient with $I = j$, so

$$P_j(\theta) = \sum_{l=0}^{+\infty} \frac{k^l}{l!} \cos \left( k \theta - \frac{l\pi}{2} \right) \sum_{i_1, \ldots, i_l \in N} \varphi_{i_1}(\theta) \ldots \varphi_{i_l}(\theta). \tag{4.7}$$

Substituting it in (4.3) gives:

$$\xi_{j,j}(k) = \frac{1}{\pi} \sum_{l=0}^{+\infty} \frac{k^l}{l!} \int_{-\pi}^{\pi} \cos \left( k \theta - \frac{l\pi}{2} \right) \varphi_{i_1}(\theta) \ldots \varphi_{i_l}(\theta) \cos((k+2j)\theta) d\theta. \tag{4.8}$$
4.3 Integrating sines

Then we will study the value inside the sum. Let’s say that \( l = 2r \). The case for odd \( l \) would be similar. Denote

\[
D = \frac{k^l}{l!} \frac{(-1)^r}{r!} \int_{-\pi}^{\pi} \cos(k\theta)\varphi_{i_1}(\theta) \cdots \varphi_{i_l}(\theta) \cos((k + 2j)\theta) d\theta. \tag{4.9}
\]

Now we use formula for the product of cosines for \( \cos(k\theta), \cos((k + 2j)\theta) \):

\[
D = \frac{k^l}{2l!} \frac{(-1)^r}{r!} \int_{-\pi}^{\pi} \varphi_{i_1}(\theta) \cdots \varphi_{i_l}(\theta) \cos(2(k + j)\theta) d\theta + \frac{k^l}{2l!} \frac{(-1)^r}{r!} \int_{-\pi}^{\pi} \varphi_{i_1}(\theta) \cdots \varphi_{i_l}(\theta) \cos(2j\theta) d\theta. \tag{4.10}
\]

Let’s use Lemma A.1 here. Remember, that \( \varphi_i(\theta) \) is the sum of sines with natural frequency less or equal than \( 2i \). So, if we substitute \( \varphi_i(\theta) \) into the first integral of (4.10), we will get a product of sums of sines. If we then open up all the parentheses, we will get the sum of many integrals of type (A.1). Note, that since \( x_p \leq 2i_p \) and \( \sum_{p=1}^{l} i_p = j \), then \( \sum_{p=1}^{l} x_p < 2(k + j) \). As such, we can use this lemma and say that the first term of (4.10) is zero.

So,

\[
D = \frac{k^l}{2l!} \frac{(-1)^r}{r!} \int_{-\pi}^{\pi} \varphi_{i_1}(\theta) \cdots \varphi_{i_l}(\theta) \cos(2j\theta) d\theta. \tag{4.11}
\]

But the same can be said about this integral. If we substitute \( \varphi_i \) and open up parentheses only one integral will remain (with \( x_p = 2i_p \)). Otherwise \( \sum_{p=1}^{l} x_p < 2j \), and we can use Lemma A.1. We receive:

\[
D = \frac{k^l}{2l!} \int_{-\pi}^{\pi} \beta_{i_1,i_1} \sin(2i_1 \theta) \cdots \beta_{i_l,i_1} \sin(2i_l \theta) \cos(2j\theta) d\theta. \tag{4.12}
\]

\[
D = \frac{k^l}{2l!} \int_{-\pi}^{\pi} \sin(2i_1 \theta) \cdots \sin(2i_l \theta) \cos(2j\theta) d\theta. \tag{4.13}
\]

Now we can evaluate this integral using formula for products of trigonometric functions, like (A.3).

If we substitute (A.3) into the integral, first three terms will vanish by lemma, so effectively we will unite \( i_1, i_2, i_3 \) and multiply the coefficient by \( -\frac{1}{4} \). Doing this \( r - 1 \) times gives us

\[
D = \frac{k^l}{2l!} \beta_{i_1,i_1} \cdots \beta_{i_l,i_1} \left(-\frac{1}{4}\right)^{r-1} \int_{-\pi}^{\pi} \sin(2(j - i_l) \theta) \sin(2i_l \theta) \cos(2j\theta) d\theta =
\]

\[
= -\frac{k^l}{4l!} \beta_{i_1,i_1} \cdots \beta_{i_l,i_l} \left(-\frac{1}{4}\right)^{r-1} \int_{-\pi}^{\pi} \cos(2j\theta) \cos(2j\theta) d\theta =
\]

\[
= \frac{k^l}{l!} \beta_{i_1,i_1} \cdots \beta_{i_l,i_l} \left(-\frac{1}{4}\right)^{r} \pi = \frac{k^l}{l!} \beta_{i_1,i_1} \cdots \beta_{i_l,i_l} \left(\frac{1}{4}\right)^{r} \pi = \frac{k^l}{2l!} \beta_{i_1,i_1} \cdots \beta_{i_l,i_l} \pi. \tag{4.14}
\]

Now we use the formula for \( \beta_{i,i} \)

\[
D = \frac{k^l}{2l!} \frac{2}{i_1 2i_2} \cdots \frac{2}{i_l 2i_l} \pi = \frac{k^l}{2l!} \beta_{i_1,i_1} \cdots \beta_{i_l,i_l} \pi. \tag{4.15}
\]
Substituting it into (4.8), we get

\[ \xi_{j,j}(k) = \sum_{l=0}^{+\infty} \sum_{i_1,\ldots,i_l \in \mathbb{N}} \frac{k^l}{2^{4j}l!} \frac{1}{i_1 \ldots i_l} = \frac{1}{2^{4j}} \sum_{l=1}^{j} \sum_{i_1,\ldots,i_l \in \mathbb{N}} \frac{k^l}{l!} \frac{1}{i_1 \ldots i_l}. \] (4.16)

From here, we see, that \( \xi_{j,j} \) is a polynomial of \( k \). Its degree is \( j \), and all its coefficients are rational positive numbers.

Now we want to simplify the expression for \( \xi_{j,j} \), proving Theorem 3. It would follow from the next Lemma:

**Lemma 4.1.**

\[ \sum_{l=1}^{j} \sum_{i_1,\ldots,i_l \in \mathbb{N}} \frac{k^l}{l!} \frac{1}{i_1 \ldots i_l} = \frac{1}{j!} k(k+1) \ldots (k+j-1) \] (4.17)

**Proof.** For the proof, see [7]. \( \square \)

We have proven (1.11). Now we will study the general case.

### 4.4 General case

We want to obtain an expression for \( \xi_{j,l}(k) \). If we will go the same route, we will encounter our first difficulty in (4.3), since there could be a cosine with the same, but negated frequency in (4.1), that would not be orthogonal to \( \cos(k+2l) \). For that reason, we also consider a sine deformation \( \mu(\varphi) = \sin(k\varphi) \), since then the term with negated frequency will go with a minus. Then, we would have that

\[ \xi_{j,l}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_j^{\cos}(\theta) \cos((k+2l)\theta) \, d\theta + \frac{1}{2\pi} \int_{-\pi}^{\pi} P_j^{\sin}(\theta) \sin((k+2l)\theta) \, d\theta. \] (4.18)

Then, we will find a formula for \( P_j^{\cos}(\theta) \) and \( P_j^{\sin}(\theta) \) using the same expansions. When we will encounter (4.10) we won’t be able to use Lemma A.1. However, the first terms will vanish since they will come with different signs in expansions for \( P_j^{\cos}(\theta) \) and \( P_j^{\sin}(\theta) \). The second term will come with the same signs, so it would double. Then, we receive that

\[ \xi_{j,l}(k) = \sum_{h=0}^{+\infty} \sum_{i_1,\ldots,i_h \in \mathbb{N}} \frac{k^h}{2^{2h}h!} \int_{-\pi}^{\pi} \varphi_{i_1}(\theta) \ldots \varphi_{i_h}(\theta) \cos\left(2l\theta - \frac{h\pi}{2}\right) \, d\theta. \] (4.19)

So, we have proven Theorem 3, since (1.12) is further studied like (1.11), because we can use Lemma A.1.

### 5 Matrix evaluation

#### 5.1 Introduction

Let us now dive into the details of caustic preservation. We will assume the semi-major axis of the ellipse to be equal to 1. Like in [6], let \( q_0 \) be some natural number, and consider
$q_0$-rationally integrable domain $\Omega$, whose boundary is close to an ellipse $E$.

$$\partial \Omega = E + \mu(\varphi). \quad (5.1)$$

Let

$$\mu(\varphi) = \mu_0' + \sum_{k=1}^{\infty} a_k \cos(k\varphi) + b_k \sin(k\varphi), \quad (5.2)$$

and assume

$$||\mu(\varphi)||_{C^1} \leq e^{2M_1}. \quad (5.3)$$

Assume that $q_0$ is an odd integer, such that $q_0 = 2k_0 - 1$. We want to show, that

$$a_k, b_k = O(e^{2||\mu||_{C^1}}), k = 3, \ldots, q_0. \quad (5.4)$$

This will show that these harmonics are small, and it is the main goal of Section 7 of [6] and this section as well. Consider the following lemma, that is just slight deformation of the Lemma 7.1 in [6].

**Lemma 5.1.**

$$a_{2k+1} = O(e^{2(k-k_0)+2||\mu||_{C^{M_2+1}}}), k = k_0, \ldots, k_0 + M_2 - 1$$

$$a_{2k} = O(e^{2(k-k_0)+2||\mu||_{C^{M_3+1}}}), k = k_0, \ldots, k_0 + M_3 - 1 \quad (5.5)$$

for some $M_2, M_3 \geq 1$, assuming $\mu$ is sufficiently smooth.

The proof of this lemma is analogous to the Lemma 7.1 in [6]. The difference between these two lemmas is as follows. In Lemma 7.1 one starts with

$$a_{2k+1} = O(e^{2||\mu||_{C^1}}), k = k_0, \ldots, 4k_0 \quad (5.6)$$

for odd cases. However, we can choose arbitrary large number instead of $4k_0$. If we do it, we would be able to inductive step $M_2 - 1$ times, instead of $k_0 - 1$ times. This will result in higher smoothness requirement, but we will be able to increase the number of harmonics, available to us.

Let’s say a couple of general ideas concerning the proof of this lemma. We prove this step-by-step, using equations (2.8). Each time, we increase the degree of $e$ in the brackets. So, we use (2.8) with increasing $N$. Since the sum there depends on $a_{q-2N}, a_{q-2N+2}, \ldots a_{q+2N-2}, a_{q+2N}$ with powers of $e$, we use previous steps to estimate it. We can describe the process using this tower-like structure:

In this example, we have $q_0 = 5$. Columns are indexed by the number of odd harmonic, while rows are indexed by powers of $e$. A red square represents that we have already proven respective estimate, the green represent the estimates we are proving on this step. On the first step, represented on first picture, we use (2.8) with $N = 0$ to prove for example that $a_7 = O(e^{2||\mu||_{C^1}})$. On the second picture, the third step is represented. On this step, we prove estimates for $a_{11}, a_{13}$ and $a_{15}$. Let’s focus on $a_{11}$, colored purple. Since we are proving estimates with $e^6$, we take $N = 2$. Purple arrows represent the elements of equality (2.8) for $q = 11$ and $N + 2$. It results in $O(e^4)a_7 + O(e^2)a_9 + a_{11} + O(e^2)a_{13} + O(e^4)a_{15} = O(e^6||\mu||_{C^3})$. 

26
Pluses represent powers of $e$ in the respective coefficients. Then, all the other harmonics are estimated using previous steps on the third picture. Blue squares represent additional powers of $e$ from the coefficients. Since all the others harmonics in the sum have at least 3 squares in respective columns, so $a_{11}$ should have also. So, we have proven an estimate, represented by purple square. We continue step-by-step, eventually proving picture 4 (the maximum height of the tower is $M_2$ for odd indices).

To prove the desired result, we will use equations, arising from (2.8). Since these equations contain separately $a$ with odd indices and with even indices, we will also study them separately going forward.

5.2 Odd modes

5.2.1 System construction

Now let’s prove, that for every $1 \leq m \leq k_0$, we have that

$$a_{2k+1} = O\left(e^{2(k-m)+2||\mu||C^{k_0+M_2}}\right); \quad k = m, \ldots, k_0 + M_2 - 1. \tag{5.7}$$

This is analogous to (7.7) in [6].

We want to prove it by induction by $m$, starting with $m = k_0$ and decreasing to $m = 1$. Note, that if we will prove it for $m = 1$, we will prove (5.4). Also note, that in order to do the step, we only need to prove (5.7) for $k = m$, and the rest will follow from the analogue of Lemma 5.1. On each step, we add another layer to the left of tower on Figure 2. Our goal is to extend the base of the tower to $a_3$. Let us prove it for $m - 1$. After denoting

$$N(k) = k - m + 1, \tag{5.8}$$
we would get the following condition on preservation of caustic $\frac{p}{2k+1}$, where $\frac{p}{2k+1}$ is in lowest terms and $\frac{p}{2k+1} < \frac{1}{q_0}$, when $k \leq k_0 + M_2 - 1$:

$$a_{2k+1} + \sum_{j=1}^{N(k)} \xi_{j,j}(2k + 1 - 2j)e^{2j} = O \left( e^{2N(k) + 2\|\mu\|_{C^{k_0 + M_2}}} \right).$$ (5.9)

Note that here we have modified (2.8) a bit. These changes are explained in [6]. We use Lemma 2.1 to substitute $\lambda_{p/q}$ and also use (5.4) for the previous induction step. We remove some terms since they already have big powers of $e$ in them due to the previous lemma and induction step (One can see that on Figure 2 by trying to add a square to the left side of the tower, like (9, 6)). We also use (5.3) to remove the square norm on the right side.

We will regard $a_i$ as variables that we need to find, and our goal would be to show, that the system, consisting of these equations, is complete, then we will use the technique from [6] to prove

$$a_{2m-1} = O \left( e^{2\|\mu\|_{C^{k_0 + M_2}}} \right).$$ (5.10)

We will later define the exact system we are studying.

### 5.2.2 System modification and algorithm description

Now, we would like to modify the equations a bit. First of all, the completeness of the system does not depend on the right-hand side, so we would not consider it. Secondly, we want to substitute the formula for $\xi_{j,j}(k)$ into all the possible equations.

$$a_{2k+1} + \sum_{j=1}^{N(k)} (2k - j)! e^{2j} \cos^{2j} \left( \frac{p\pi}{2k+1} \right) a_{2k+1-2j} = O \left( e^{2N(k) + 2\|\mu\|_{C^{k_0 + M_2}}} \right).$$ (5.11)

Now we remove unnecessary parameters from our possible equations. Specifically, we introduce

$$x_{2k+1} = a_{2k+1} e^{2k}. $$ (5.12)

This reduces our equations to

$$x_{2k+1} + \sum_{j=1}^{N(k)} \frac{(2k - j)!}{j!(2k - 2j)!} \cos^{2j} \left( \frac{p\pi}{2k+1} \right) x_{2k+1-2j}.$$ (5.13)

Note that now the system’s coefficients are independent of $e$ as well as $\mu$. Substituting a formula for binomial coefficient gives:

$$x_{2k+1} + \sum_{j=1}^{N(k)} \frac{(2k - j)!}{\cos^{2j} \left( \frac{p\pi}{2k+1} \right)} x_{2k+1-2j}.$$ (5.14)

Multiplying everything by cosines gives:

$$\cos^{2N(k)} \left( \frac{p\pi}{2k+1} \right) x_{2k+1} + \sum_{j=1}^{N(k)} \binom{2k - j}{j} \cos^{2(N(k) - j)} \left( \frac{p\pi}{2k+1} \right) x_{2k+1-2j}.$$ (5.15)
And now use the formula for the cosine of the double-angle:

\[
\left( \cos\left( \frac{2p\pi}{2k+1} \right) + 1 \right)^{N(k)} x_{2k+1} + \sum_{j=1}^{N(k)} \left( \cos\left( \frac{2p\pi}{2k+1} \right) + 1 \right)^{N(k)-j} x_{2k+1-2j}. \tag{5.16}
\]

We denote the above equation \( v_{p/(2k+1)} \). We also denote \( S_k \) as a system, consisting of all the given \( v_{p/(2l+1)} \) with \( l \leq k \). Also, we may consider \( v_{p/(2k+1)} \) with \( \frac{p}{2k+1} \geq \frac{1}{q_0} \), but we won’t include those into \( S_k \). Note that \( S_k \) has only variables \( x_{2m-1}, x_{2m+1}, \ldots, x_{2k+1} \) in it. Also, denote \( K_k \) to be the kernel of \( S_k \) in the space of \( x_{2m-1}, x_{2m+1}, \ldots, x_{2k+1} \), and let \( \kappa_k = \dim(K_k) \). We need to prove that some \( S_k \) is complete, meaning \( \kappa_k = 0 \). Now let us describe an algorithm of the proof:

1. Start with \( h = k_0 \). Then, \( S_h \) has 1 equation and \( k_0 - m \) variables in it, so \( \kappa_h = k_0 - m - 1 \).

2. Set \( h = h + 1 \), and consider the difference between \( S_h \) and \( S_{h-1} \). We have added a variable \( x_{2h+1} \), and some equations \( v_{p/(2h+1)} \). Since the added equations depend non-trivially on \( x_{2h+1} \), we have \( \kappa_h \leq \kappa_{h-1} \).

3. If \( \kappa_h = 0 \), then \( S_h \) is complete, so we can prove (5.7) using the same techniques of matrix inversion, as in [6].

4. If \( q = 2h + 1 \) is a prime number with some properties (\( q_0 \)-A-good) and \( \kappa_h = \kappa_{h-1} \), we prove that \( \kappa_h = 0 \). So, otherwise the rank should fall at least by one.

5. If \( \kappa_h > 0 \), and \( h < \left( k_0 + M_2 \right)/2 \), return to step 2.

So, each time \( 2h + 1 \) is a prime number with some properties, \( \kappa_h \) falls at least by 1. So, if there would be an infinite amount of such numbers, \( \kappa_h \) would fall indefinitely, until either it hits zero or \( h \) hits \( (k_0 + M_2 - 1)/2 \). Once the latter happens, we won’t be able to continue due to the lack of (5.7) for large harmonics. We can insure the first option by letting \( M_2 \) to be large enough, it would all be discussed later. Note, however, that the algorithm does not depend on the deformation, so the choice of \( M_2 \) would be independent on it.

Let’s discuss the number \( (k_0 + M_2)/2 \) and why we cannot go beyond it. The reason for it is connected with going from (2.8) to (5.9). In (2.8) for odd cases if \( q = 2h + 1 \) and we prove for example for \( m = 2 \), the following variables are involved:

\[
a_3, a_5, a_7, \ldots, a_{2h-1}, a_{2h+1}, a_{2h+3}, \ldots, a_{4h-3}, a_{4h-1}. \tag{5.17}
\]

Then, [6] uses their analogue of (5.7) there, namely formula (7.7). There, they use the previous step of induction to work with those variables. To do that here, we need that every one of those indices lies in the list of \( k \) in (5.7), except the first one. Particularly, it means that

\[
4h - 1 \leq 2(k_0 + M_2 - 1) + 1. \tag{5.18}
\]

This reduces to

\[
h \leq \frac{k_0 + M_2}{2}. \tag{5.19}
\]
Now, if some time $k_h$ becomes 0, that would mean that the system is complete, so we would be able to solve the system and find in particular $a_{2m-1}, a_{2m+1}, \ldots, a_{2k_0-1}$. Then, we have the same situation as in [6], and we follow their algorithm using their Appendix D (see page 38 of [6]). Then, we have the following:

$$a_{2k+1} = O \left( e^{2(k-m+1)+2}\|\mu\|_{C^{k_0+M_2}} \right); \quad k = m - 1, \ldots, k_0 - 1. \quad (5.20)$$

The step of (5.7) follows from here, using similar ideas to Lemma 5.1. We have basically added some squares on the left side of Figure 2 and we can continue building it.

Now we focus on proving the fact in step 4.

### 5.2.3 Field introduction

Let us prove, that $\kappa_h$ would actually decrease for some $h$. We will prove it using algebraic field theory. Let’s say $q = 2h + 1$, is a prime number. Presume $\kappa_h$ did not fall. Let $p_1, p_2$ be some numbers, such that $\frac{p_1}{q} < \frac{1}{q_0}, \frac{p_2}{q} < \frac{1}{q_0}$. Then, note that the angle $\frac{2\pi p_2}{q}$ is some multiple of the angle $\frac{2\pi p_1}{q}$ modulo $2\pi$. As such, we can express $\cos \frac{2\pi p_2}{q}$ through $\cos \frac{2\pi p_1}{q}$ via the formula for cosine of the natural multiple of an angle. Let

$$\frac{2\pi p_2}{q} + 2\pi s = r(p_1, p_2) - \frac{2\pi p_1}{q}, \quad r(p_1, p_2) \in \mathbb{Z}, 1 \leq r(p_1, p_2) \leq q - 1. \quad (5.21)$$

One can also note that if we will consider $p_1, p_2$ as elements of $\mathbb{F}_q$, the following would be true:

$$r(p_1, p_2) = p_2 p_1^{-1}. \quad (5.22)$$

The formula for $r(p_1, p_2)$-multiple cosine will always be a polynomial with rational coefficients. Precisely, let

$$\cos \frac{2\pi p_2}{q} = P \left( \cos \frac{2\pi p_1}{q} \right). \quad (5.23)$$

Now, consider the following system:

$$S_h(p_1, p_2) = \begin{pmatrix} S_{h-1} \\ v_{p_1/(2h+1)} \\ v_{p_2/(2h+1)} \end{pmatrix}. \quad (5.24)$$

Since it is a subsystem of $S_h$, we have

$$K_h \subset \ker(S_h(p_1, p_2)). \quad (5.25)$$

However,

$$\dim \ker(S_h(p_1, p_2)) \leq \kappa_h - 1 = \kappa_h = \dim(K_h). \quad (5.26)$$

So,

$$K_h = \ker(S_h(p_1, p_2)), \quad (5.27)$$

$$\dim \ker(S_h(p_1, p_2)) = \kappa_h - 1. \quad (5.28)$$
That means that by adding 2 new equations and only 1 new variable to $S_{h-1}$, the rank of kernel did not fall. If we write down a condition on that, we will receive that some minors of the matrix of $S_{h}(p_1, p_2)$ vanish.

Let’s describe our following steps. First, we will construct a field, containing some elements of the matrix of $S_{h}(p_1, p_2)$. Out of all coefficients, only $\cos \frac{2\pi p_1}{q}$ will not be present in said field. Then, we will consider a ring of polynomials over the field, depending on some variable $z$. After that, we will substitute $z$ instead of $\cos \frac{2\pi p_1}{q}$ in the matrix and write down the described minors of the matrix. These minors will be polynomials of $z$, and will have a root at $z = \cos \frac{2\pi p_1}{q}$. Then, they will be divisible by the minimal polynomial of $\cos \frac{2\pi p_1}{q}$. We will use it to substitute other roots of the minimal polynomial instead of $z$.

Let’s construct a field $F$. First of all, we will consider the field of rational numbers $\mathbb{Q}$. Next, let $W$ be the lowest common multiple of all the numbers, less than $q$. Then, let $w$ be the primitive root of unity of order $W$. Our field $F$ would be $\mathbb{Q}$ with added element $w$. Now,

$$[F : \mathbb{Q}] = \varphi(W),$$

where $\varphi(W)$ is Euler’s totient function.

Now let’s discuss the element of such field. First of all, rational numbers are obviously present in this field. So, all the binomial coefficients are present. Also, all roots of unity of degree $W$ are present. Then, for every $\tilde{q} < q$, its roots of unity are present. This means that

$$\cos \left( \frac{2\pi \tilde{p}}{\tilde{q}} \right) + i \sin \left( \frac{2\pi \tilde{p}}{\tilde{q}} \right) \in F, \quad \tilde{q} < q.$$  

(5.30)

Since conjugate root is also present, we have that

$$\cos \left( \frac{2\pi \tilde{p}}{\tilde{q}} \right) \in F, \quad \tilde{q} < q.$$  

(5.31)

So, note that every equation in $S_{h}(p_1, p_2)$, except $v_{p_1/(2h+1)}$ and $v_{p_2/(2h+1)}$, has all their coefficients present in $F$. Of those two, the coefficients of $v_{p_1/(2h+1)}$ will have polynomial dependency on $\cos \frac{2\pi p_1}{q}$. For $v_{p_2/(2h+1)}$ this will also be true, after considering (5.23). Now let’s write down the matrix of $S_{h}(p_1, p_2)$:

$$
\begin{pmatrix}
  f_{11} & f_{21} & \cdots & f_{\alpha 1} \\
  f_{12} & f_{22} & \cdots & f_{\alpha 2} \\
  \vdots & \vdots & \ddots & \vdots \\
  f_{1\beta} & f_{2\beta} & \cdots & f_{\alpha \beta} \\
  R_1 \left( \cos \frac{2\pi p_1}{q} \right) & R_2 \left( \cos \frac{2\pi p_1}{q} \right) & \cdots & R_\alpha \left( \cos \frac{2\pi p_1}{q} \right) \\
  \mathcal{R}_1 \left( P \left( \cos \frac{2\pi p_1}{q} \right) \right) & \mathcal{R}_2 \left( P \left( \cos \frac{2\pi p_1}{q} \right) \right) & \cdots & \mathcal{R}_\alpha \left( P \left( \cos \frac{2\pi p_1}{q} \right) \right)
\end{pmatrix}.
$$

(5.32)

Here, $f_{ij}$ denote some elements of $F$, and $R_i$ – some polynomials over $F$. The first $\beta$ lines represent $S_{h-1}$, the second-to-last represents $v_{p_1/(2h+1)}$, and the last one – $v_{p_2/(2h+1)}$. Also note that $P$ is also a polynomial over $F$. Now, introduce new variable $z \in \mathbb{C}$, and put it into
this matrix instead of $\cos \frac{2\pi p}{q}$:

\[
\begin{pmatrix}
  f_{11} & f_{21} & \ldots & f_{\alpha 1} \\
  f_{12} & f_{22} & \ldots & f_{\alpha 2} \\
  \vdots & \vdots & \ddots & \vdots \\
  f_{1\beta} & f_{2\beta} & \ldots & f_{\alpha \beta} \\
  R_1(z) & R_2(z) & \ldots & R_\alpha(z) \\
  R_1(P(z)) & R_2(P(z)) & \ldots & R_\alpha(P(z))
\end{pmatrix}.
\] (5.33)

Now we know, that at $z = \cos \frac{2\pi p}{q}$ some minors of this matrix are zero. Since all the minors of this matrix are polynomials over $F$ from $z$, that means that these polynomials are divisible by the minimal polynomial $\tilde{\Psi}$ of $\cos \frac{2\pi p}{q}$ over $F$. We know, that the minimal polynomial of $\cos \frac{2\pi p}{q}$ over $\mathbb{Q}$ is $\Psi$. We also know, that

\[
\deg(\Psi) = \frac{\varphi(q)}{2} = \frac{q - 1}{2}.
\] (5.34)

The roots of $\Psi$ take the form

\[
\cos \frac{2\pi p}{q}, \ p = 1, \ldots, \frac{q - 1}{2}.
\] (5.35)

Now we will prove, that $\tilde{\Psi} = \Psi$, meaning that by adding new elements to the field, we did not reduce the degree of the minimal polynomial.

**Lemma 5.2.**

\[
\tilde{\Psi} = \Psi
\] (5.36)

**Proof.** Let’s assume it is not true. Then

\[
d = \deg(\tilde{\Psi}) < \deg(\Psi) = \frac{q - 1}{2}.
\] (5.37)

Consider $F_1$ by adding $\cos \frac{2\pi p}{q}$ to $F$. 

\[
[F_1 : F] = d.
\] (5.38)

Now, consider $F_2$ by adding $i \sin \frac{2\pi p}{q}$ to $F_1$ as a solution to $z^2 = \cos \frac{2\pi p}{q} - 1$, if it is not already there. Then,

\[
[F_2 : F_1] \leq 2.
\] (5.39)

So, we have that

\[
[F_2 : \mathbb{Q}] = [F_2 : F_1] [F_1 : F] [F : \mathbb{Q}] \leq 2d \varphi(W).
\] (5.40)

But since $\cos \frac{2\pi p}{q} + i \sin \frac{2\pi p}{q}$ is present in $F_2$, all the other roots of unity of degree $q$ are present there. Since roots of unity of degree $q$ and $W$ are present in $F_2$, the roots of unity of degree $qW$ should be present there, since $q$ and $W$ are co-prime. Since the primitive roots
of unity of degree \( qW \) are present there, the expansion of \( F_2 \) over \( \mathbb{Q} \) should at least have the degree of their minimal polynomial. Then,

\[
[F_2 : \mathbb{Q}] \geq \varphi(qW).
\]

(5.41)

So,

\[
(q - 1)\varphi(W) = \varphi(qW) \leq 2d\varphi(W).
\]

(5.42)

that immediately leads to contradiction. So, \( \tilde{\Psi} = \Psi \).

\[\square\]

5.2.4 Changing roots

So, all the described minors are divisible by \( \Psi(z) \). Then, they have all the roots of \( \Psi(z) \) as their roots. In particular, we can substitute \( z = \cos \frac{2\pi}{q} \). This means, that the following matrix has the same dependencies, as the matrix of \( S_h(p_1, p_2) \):

\[
\begin{pmatrix}
  f_{11} & f_{21} & \cdots & f_{\alpha 1} \\
  f_{12} & f_{22} & \cdots & f_{\alpha 2} \\
  \vdots & \vdots & \ddots & \vdots \\
  f_{1\beta} & f_{2\beta} & \cdots & f_{\alpha\beta} \\
  R_1\left(\cos \frac{2\pi}{q}\right) & R_2\left(\cos \frac{2\pi}{q}\right) & \cdots & R_\alpha\left(\cos \frac{2\pi}{q}\right) \\
  R_1\left(P\left(\cos \frac{2\pi}{q}\right)\right) & R_2\left(P\left(\cos \frac{2\pi}{q}\right)\right) & \cdots & R_\alpha\left(P\left(\cos \frac{2\pi}{q}\right)\right)
\end{pmatrix}.
\]

(5.43)

Now, we know that

\[
P\left(\cos \frac{2\pi}{q}\right) = \cos\left(\frac{2r(p_1, p_2)\pi}{q}\right).
\]

(5.44)

So, (5.43) actually has the following structure

\[
\begin{pmatrix}
  S_{h-1} \\
  v_{1/(2h+1)} \\
  v_{r(p_1, p_2)/(2h+1)}
\end{pmatrix},
\]

(5.45)

since \( R_i\left(\cos \frac{2\pi}{q}\right) \) are just the coefficients of \( v_{p/(2h+1)} \).

It is natural to denote its system as \( S_h(1, r(p_1, p_2)) \). Then,

\[
\dim \ker(S_h(p_1, p_2)) = \kappa_{h-1} \Rightarrow \dim \ker(S_h(1, r(p_1, p_2))) = \kappa_{h-1}.
\]

(5.46)

This means that

\[
\ker(S_h(1, r(p_1, p_2))) = \ker(S_h(1)) = \ker(S_h) = K_h
\]

(5.47)

for a logical definition of \( S_h(1) \).

Let’s understand what we did here. We had two equations, that had \( K_h \) in a kernel, namely \( v_{p_1/(2h+1)} \) and \( v_{p_2/(2h+1)} \).

We did some operations and deduced, that \( v_{r(p_1, p_2)/(2h+1)} \) also has \( K_h \) in a kernel.

Let us now generalize this process. Let \( G \) be the set of all \( p \in \mathbb{F}_q^* \) such that \( v_{p/(2h+1)} \) has \( K_h \) in a kernel. We have proven that if \( p_1 \in G, p_2 \in G \), then \( r(p_1, p_2) = p_2p_1^{-1} \in G \). Note
that since caustic $\frac{1}{q}$ is available to us, that means that $1 \in G$. This immediately proves that $G$ is a subgroup of $\mathbb{F}_q^*$.

Now note that if $\frac{p}{q} < \frac{1}{q_0}$, then $p \in G$(of course here $p$ is a natural number). Also note that $G$ is symmetrical by multiplying by $-1$, since the cosine is an even function.

Now suppose, that for given $q$, these demands force $G$ to be equal to the whole group. We will discuss, for which primes this is true, later, but now notice that this condition depends only on prime number itself and on $q_0$.

If $G$ is the whole group, we will show that $K_h = \{0\}$.

5.2.5 Finishing steps

Let’s count the variables. We have $x_{2m-1}, \ldots, x_{2h+1}$. We have $N(h) + 1$ of them. Notice that, $N(h) + 1 \leq h$. Now consider equations $v_{p/(2h+1)}$ for $p = 1, \ldots, N(h) + 1$, and the system consisting only of them. Since the system has $K_h$ inside its kernel, and if $K_h$ is not zero, then the determinant of its matrix is zero. We will show that it cannot be this way. Write down the equations more precisely (let’s say first row corresponds to $x_{2h+1}$, last – to $x_{2m-1}$)

$$\left(\left(\frac{\cos 2\pi \frac{p_1}{q} + 1}{2}\right)^{N(h)}, \left(\frac{\cos 2\pi \frac{p_1}{q} + 1}{2}\right)^{N(h)-1}, \ldots, \left(\frac{\cos 2\pi \frac{p_1}{q} + 1}{2}\right)^0\right). \quad (5.48)$$

Now the binomial coefficients do not depend on $p$, so elements in the same column have the same coefficients. Since they are non-zero, we can multiply whole columns by their inverses and cancel them. Determinant will remain zero. Consider the rows of the new matrix:

$$\left(\left(\frac{\cos 2\pi \frac{p_1}{q} + 1}{2}\right)^{N(h)}, \left(\frac{\cos 2\pi \frac{p_1}{q} + 1}{2}\right)^{N(h)-1}, \ldots, \left(\frac{\cos 2\pi \frac{p_1}{q} + 1}{2}\right)^0\right). \quad (5.49)$$

Notice that this matrix is just the rotated Vandermonde matrix, and its determinant is nonzero, since

$$\cos 2\pi \frac{p_1}{q} \neq \cos 2\pi \frac{p_2}{q}; \ p_1, p_2 = 1, 2, \ldots, N(h) + 1; \ p_1 \neq p_2. \quad (5.50)$$

So, if $q$ has this property, written above, then the dimension of the kernel should fall at least by one. Since there are infinite number of those primes, the kernel of the system will become zero at some $h_f$. Note that $h_f$ depends only on $q_0$, and does not depend on $e$ or the deformation $\mu$.

$b_{2k+1}$ would be handled the similar way.

5.3 Selection of Primes

Previously, we described a following problem. Let $q$ be a prime, and let $G$ be the minimal subgroup of $\mathbb{F}_q^*$, that contains $1, \ldots, \left[\frac{q}{q_0}\right]$ and is symmetrical under negation. For what $q$ $G = \mathbb{F}_q^*$?

This question is very important, since the required smoothness of the deformation depends on it. These are exactly $q_0$-A-good numbers, introduced in Definition 1.4.
In particular, $M_2$ can be chosen to be equal to $M_4 - k_0$, where $M_4$ is the $k_0 - 1$-st $q_0$-A-good number, due to (5.19). It is true, because we have proven, that every time that $2h + 1$ is a $q_0$-A-good number, the dimension of kernel drops at least by one. Since originally the kernel had a maximal dimension of $k_0 - 1$, at the step $2h_k + 1 = M_4$ the dimension of kernel should be zero. Since we demand (5.19), we have that

$$q \leq k_0 + M_2 + 1. \quad (5.51)$$

Since $q$ never becomes bigger than $M_4$, we can set $M_2$ to be equal to $M_4 - k_0$, and the inequality would always hold.

$M_1$ could be chosen to be $M_2 + k_0 = M_4$. However, since $M_1$ is common between odd and even cases, we need to go through an even case before finally picking it.

Later, we will give a table of these constants for various values of $q_0$.

Now we will give the sufficient condition on number being $q_0$-A-good. Note that $|G|$ should divide $|F_q^*| = q - 1$. Since $G$ is symmetrical, $|G|$ divides 2. Then, $|G|/2$ divides $|F_q^*|/2 = q-1/2$. So, $t = |F_q^*|/|G|$ divides $q-1/2$. However, $t < q_0$. Then, if $q-1/2$ has no prime divisors less than $q_0$, $q$ is $q_0$-A-good.

For example every safe prime $q$ (when $q-1/2$ is a Sophie-Germain prime), greater than $2q_0$, is good. Although the infinity of the number of Sophie-Germain primes is not yet proven, heuristics can estimate the number of them being less then $n$ to be $\Omega\left(\frac{n}{\log^2 n}\right)$. This would give the asymptotic estimate for $M_4 \leq Cq_0^{1+\varepsilon}$.

However, to prove that there are infinite number of $q_0$-A-good numbers, we may use Dirichlet’s theorem on arithmetic progressions. In particular, let $Q$ be a number that gives residues 3 modulo 4 and residue 2 modulo all odd primes, less than $q_0$. Let $D$ be the product of 4 and all those primes. Since $D$ and $Q$ are co-prime, there are infinitely many primes of the form $Q + nD$. They would satisfy the sufficient condition on being $q_0$-A-good.

### 5.4 Even modes

#### 5.4.1 Introduction

The algorithm concerning even indices would be similar to the odd one. Like in [6], one would be able to derive the same equations. Then, it is possible to prove that the dimension of the kernel does not increase. Then one would consider caustics of rotation number $p/q$ for some prime $q$ and odd $p$. Then, we will prove that the dimension of the kernel decreases for some such $q$.

However, in order to proceed with all of this, we first need to obtain some new equations to add to ones that were already obtained in [6].
5.4.2 New equations

Let \( \frac{p}{q} < \frac{1}{q_0} \) and let’s consider a condition of preservation of the caustic \( \frac{p}{q} \) that is shown on the top of page 20 of [6].

\[
\sum_{j=1}^{+\infty} a_{jq} \cos(jq\theta) + b_{jq} \sin(jq\theta) + \\
+ \sum_{n=1}^{N} \sum_{l=-n}^{n} \xi_{n,l}(jq - 2l)(a_{jq-2l} \cos(jq\theta) + b_{jq-2l} \sin(jq\theta)) \frac{a^{2n}e^{2n}}{(a^2 - \lambda_{p/q}^2)^n} = 0
\]

where \( c_{p/q} \) is some constant, and \( N \leq m-1 \). We also demand that \( \frac{p}{q} \) is in lowest terms.

In [6], this equation was multiplied by \( \cos(q\theta) \), and then integrated. All the equations we worked with before originated from this. Now, however we want to multiply the equation by \( \cos(2q\theta) \) and integrate it over \( \theta \in [0, 2\pi] \). We receive the following:

\[
a_{2q} + \sum_{n=1}^{N} \sum_{l=-n}^{n} \xi_{n,l}(2q - 2l)a_{2q-2l} \frac{a^{2n}e^{2n}}{(a^2 - \lambda_{p/q}^2)^n} = 0 \left( ||\mu||_{C^{N+1}} e^{2N+2} + \lambda_{p/q}^{-1} q^7 ||\mu||_{C^1}^2 \right) \quad (5.53)
\]

We will use it with \( q > N \).

We have received an equation about even modes, arising from not necessarily even caustics. Now, we add them to the equations (2.8), already derived in [6] (we substitute 2q in them), also with \( q > N \).

\[
a_{2q} + \sum_{n=1}^{N} \sum_{l=-n}^{n} \xi_{n,l}(2q - 2l)a_{2q-2l} \frac{a^{2n}e^{2n}}{(a^2 - \lambda_{p/2q}^2)^n} = 0 \left( ||\mu||_{C^{N+1}} e^{2N+2} + \lambda_{p/2q}^{-1} q^7 ||\mu||_{C^1}^2 \right) \quad (5.54)
\]

Here, \( \frac{p}{2q} \) should also be in its lowest terms.

5.4.3 Even system of equations

Now, we proceed to analyze even nodes. In particular, we prove, that for every \( 1 \leq m \leq k_0 - 1 \), we have that

\[
a_{2k+2} = O \left( e^{2(k-m)+2} ||\mu||_{C^{k_0+m+3}} \right), \quad k = m, \ldots, k_0 + M_3 - 1.
\]

As with odd nodes, we do that by induction over \( m \), starting with \( m = k_0 - 1 \) and decreasing, until \( m = 1 \). The first case follows from Lemma 5.1. Let’s say we want to prove it for \( m - 1 \). Now we denote

\[
N(k) = k - m,
\]

then, equations (5.54) would reduce to the following, when \( \frac{p}{2k} < \frac{1}{q_0} \), \( \frac{p}{2k} \) is in its lowest terms and \( k < k_0 + M_3 \):

\[
\cos^{2N(k)} \left( \frac{p\pi}{2k} \right) x_{2k} + \sum_{j=1}^{N(k)} \binom{2k - j - 1}{j} \cos^{2(N(k)-j)} \left( \frac{p\pi}{2k} \right) x_{2k-2j}.
\]

(5.57)
Equations (5.53) would similarly reduce, when \( \frac{p}{k} < \frac{1}{q_0} \), \( \frac{p}{k} \) is in lowest terms and \( k < k_0 + M_3 \):

\[
\cos^{2N(k)} \left( \frac{p\pi}{k} \right) x_{2k} + \sum_{j=1}^{N(k)} \binom{2k - j - 1}{j} \cos^{2(N(k)-j)} \left( \frac{p\pi}{k} \right) x_{2k-2j}.
\]

(5.58)

Denote these two \( v_p/(2k) \) and \( v_{(2p)}/(2k) \) respectively. Also, introduce \( S_k, K_k \) and \( \kappa_k \) the same way as with the odd case. We add both types of equations to \( S_k \).

Now we will consider the procedure, similar to the odd case. We start with \( h = k_0 \) and we will increase it by 1 step-by-step. We prove that \( \kappa_h \) would not increase. So, if the rank becomes zero, we would have proven our assumption (5.55).

Similarly to the odd case, let’s prove, that if \( h = q \) is a prime number with some properties, then the rank falls at least by one.

### 5.4.4 Even case field theory

We will first start with \( S_{h-1} \) and add a couple of new equations to it. They will all have \( K_h \) inside their kernel, and we want to study, what happens if the rank does not fall. We separately consider equations (5.57) and (5.58).

Let’s start with (5.57). They take the following form:

\[
\cos^{2N(k)} \left( \frac{p\pi}{2q} \right) x_{2q} + \sum_{j=1}^{N(k)} \binom{2q - j - 1}{j} \cos^{2(N(k)-j)} \left( \frac{p\pi}{2q} \right) x_{2q-2j}.
\]

(5.59)

Then, we will get

\[
\left( \cos \left( \frac{p\pi}{q} \right) + 1 \right)^{N(k)} x_{2q} + \sum_{j=1}^{N(k)} \binom{2q - j - 1}{j} \left( \cos \left( \frac{p\pi}{q} \right) + 1 \right)^{N(k)-j} x_{2q-2j}.
\]

(5.60)

As you can see \( p \) is odd, so we are not getting the same cosines, as we did before (for odd modes we got effectively \( p \) as an even number, because we had a 2 in numerator). However, we can fix that by one simple trick:

\[
\left( -\cos \left( \frac{(q+p)\pi}{q} \right) + 1 \right)^{N(k)} x_{2q} + \sum_{j=1}^{N(k)} \binom{2q - j - 1}{j} \left( -\cos \left( \frac{(q+p)\pi}{q} \right) + 1 \right)^{N(k)-j} x_{2q-2j}.
\]

(5.61)

Now we once again have roots of \( \Psi(z) \) as a coefficients to our equation. Let’s add two these equations to \( S_{h-1} \) and write down the condition of the rank not decreasing. Then, the same algebraic group theory would follow, and we will get a subset \( G \) of the \( F^*_q \) elements that have \( K_h \) inside the kernel of the respective equation.

Precisely, let’s say an equation, corresponding to \( \cos^{2r_1\pi} \frac{q}{q} \) has \( K_h \) in its kernel, and this is also true for \( r_2 \) and \( r_3 \). We add equations, corresponding to \( r_1 \) and \( r_2 \), to \( S_{h-1} \). Then, we express the \( r_2 \) cosine, using \( r_1 \) cosine. Afterwards, we use field theory to transform \( r_1 \) cosine into \( r_3 \) cosine. Then, the cosine that used to be \( r_2 \), transforms into \( r_2 r_3 r_1^{-1} \) cosine. Since the
rank still doesn’t fall and \( r_3 \) row has \( K_h \) inside its kernel, we deduce that \( r_2 r_3 r_1^{-1} \) row has it also inside the kernel. That means that \( r_1, r_2, r_3 \in G \Rightarrow r_2 r_3 r_1^{-1} \in G \). In the odd case we used it with \( r_3 = 1 \), since we had that \( 1 \in G \).

\( G \) would still be symmetrical by negation. However, instead of \( 1, \ldots, \left[ \frac{q}{q_0} \right] \) inside of \( G \) by default, we would get \( \frac{q-1}{2} + 1, \frac{q-1}{2} + 2, \ldots, \frac{q-1}{2} + \left[ \frac{q}{q_0} \right] \), since we rotated everything by \( \pi \) in (5.61).

So, \( G \) has similar structure to the subgroup, and actually becomes a subgroup, if one multiplies it by 2. Then \( G \) would include \( 1, 3, \ldots, 2 \left[ \frac{q}{q_0} \right] - 1 \).

Then one can introduce \( q_0 \)-B-good numbers in the same fashion, as we did in Definition 1.5. For example, the sufficient condition will be the same, and so there would be infinitely many of them. So, \( q \) has to be \( q_0 \)-B-good.

Now, consider equations (5.58). Let’s add two of these to our system instead of (5.57) and write down the condition of rank not falling. These equations have the following form

\[
\cos^{2N(k)} \left( \frac{p\pi}{q} \right)x_{2q} + \sum_{j=1}^{N(k)} \binom{2q-j-1}{j} \cos^{2(N(k)-j)} \left( \frac{p\pi}{q} \right)x_{2q-2j}.
\]  

(5.62)

This case is easier, since we have

\[
\left( \cos \left( \frac{2p\pi}{q} \right) + 1 \right)^{N(k)} x_{2q} + \sum_{j=1}^{N(k)} \binom{2q-j-1}{j} \left( \cos \left( \frac{2p\pi}{q} \right) + 1 \right)^{N(k)-j} x_{2q-2j} 
\]  

(5.63)

and we already have the roots to \( \Psi(z) \). This case is identical to the odd one, so \( q \) would have to be \( q_0 \)-A-good.

So, if \( q \) is both \( q_0 \)-A-good and \( q_0 \)-B-good, we will be able to add all the equations (5.57) and (5.58) without the concern for \( \frac{p}{2q} < \frac{1}{q_0} \) or \( \frac{p}{q} < \frac{1}{q_0} \), respectively.

Let’s count deduced equations. We have equations (5.57) for \( \frac{1}{2q}, \frac{3}{2q}, \ldots, \frac{q-2}{2q} \) - so we have \( \frac{q-1}{2} \) of them. We also have equations (5.58) for \( \frac{1}{q}, \frac{2}{q}, \ldots, \frac{(q-1)/2}{q} \) - \( \frac{q-1}{2} \) of them. In total, we have \( q - 1 \) new equations. Let’s consider a system, consisting of only new equations. This system has \( K_h \) in its kernel. However, one notices that the number of variables is only \( N(q) + 1 = q - m + 1 \leq q - 1 \). Then, using the same ideas realized in the odd case, we transform a matrix of this system into the Vandermonde matrix, so the system has zero kernel, that leads to contradiction.

So, every time \( q \) is prime \( q_0 \)-A-good and \( q_0 \)-B-good, the rank of the kernel drops at least by one. If \( M_3 \) is large enough, it would become 0.

### 5.5 Section of Primes, part 2

So, we need to understand, how many prime \( q_0 \)-A-good and \( q_0 \)-B-good numbers there are. First of all, we sill have the same sufficient condition. So, there are infinitely many of such numbers, and for example every safe prime suffices.
Now let’s pick $M_3$. We have an analogue of (5.19) for even indices. Namely, we have:

$$h \leq \frac{k_0 + M_3}{2}. \quad (5.64)$$

Since $h = q$, this results in

$$q \leq \frac{k_0 + M_3}{2}. \quad (5.65)$$

In particular, $M_3$ can be chosen to be equal to $2M_5 - k_0$, where $M_5$ is $k_0 - 2$nd $q_0$-A-good and $q_0$-B-good number. $M_1$ can be chosen as $M_3 + k_0 = 2M_5$ to satisfy an even case. Overall, we choose $M_1$ to be the maximum of two numbers for even and for odd. We also set $n = M_1$ and $m = 14M_1$. So, we get the following list, after computing $M_5$:

| $q_0$ | $M_4$ | $M_2$ | $M_1^{odd}$ | $M_5$ | $M_3$ | $M_1^{even}$ | $M_1$ | $n$ | $m$ |
|-------|-------|-------|-------------|-------|-------|-------------|-------|-----|-----|
| 3     | 7     | 5     | 7           | -     | -     | -           | 7     | 7   | 98  |
| 5     | 13    | 10    | 13          | 11    | 19    | 22          | 22    | 22  | 308 |
| 7     | 29    | 25    | 29          | 23    | 42    | 46          | 46    | 46  | 644 |
| 9     | 31    | 26    | 31          | 29    | 53    | 58          | 58    | 58  | 812 |
| 11    | 43    | 37    | 43          | 41    | 76    | 82          | 82    | 82  | 1148|
| 19    | 89    | 79    | 89          | 97    | 184   | 194         | 194   | 194 | 2716|
| 29    | 131   | 116   | 131         | 139   | 263   | 278         | 278   | 278 | 3892|
| 49    | 613   | 563   | 613         | 641   | 1232  | 1282        | 1282  | 1282| 17948|
| 99    | 1229  | 1129  | 1229        | 1291  | 2482  | 2582        | 2582  | 2582| 36148|
| 199   | 3407  | 3157  | 3407        | 3593  | 6936  | 7186        | 7186  | 7186| 100604|
| 499   | 7129  | 6629  | 7129        | 7577  | 14654 | 15154       | 15154 | 15154| 212156|

Figure 3: All the smoothness constants for several values of $q_0$

This results in Theorem 4.

### A Important lemmata

**Lemma A.1.** Let $\psi_x(\theta)$ denote either $\cos(x\theta)$ or $\sin(x\theta)$ for $x \in \mathbb{N}$. Let $x_1 + \ldots + x_h < y$, where $x_1, \ldots, x_h, y \in \mathbb{N}$.

Then,

$$\int_{-\pi}^{\pi} \psi_{x_1}(\theta) \ldots \psi_{x_h}(\theta) \psi_y(\theta) d\theta = 0 \quad (A.1)$$

**Proof.** We will prove it by mathematical induction by $h$. For $h = 1$ it just states that trigonometric functions are orthogonal.

If $h > 1$ we can just use the formula for the product of trigonometric functions for $\psi_{x_1}$ and $\psi_{x_2}$. We will get the sum of trigonometric functions for $x_1 + x_2$ and $|x_1 - x_2|$, and then
break this sum in two integrals. In each integral the inequality on $x$ and $y$ can be shown to be true (we exchange $x_1$ and $x_2$ for $x_1 + x_2$ or $|x_1 - x_2|$). Then, both of these integrals will be 0 by induction.

Moreover, notice that if $x_1 + x_2 + \ldots + x_h = y$, then this integral could also be easily calculated. Once again we can use a formula for product of trigonometric functions, and the term with $|x_1 - x_2|$ will vanish by the lemma. So, the integral will simplify to

$$\int_{-\pi}^{\pi} \psi_{x_1+x_2}(\theta)\psi_{x_3}(\theta)\ldots\psi_{x_h}(\theta)\psi_y(\theta)d\theta. \quad (A.2)$$

Also notice that these integrals can express the coefficient in front of $\psi_y(\theta)$ if one would transform $\psi_{x_1}(\theta)\ldots\psi_{x_h}(\theta)$ into the sum of trigonometric functions.

We also would need the following formula:

$$\sin(2i_1\theta)\sin(2i_2\theta)\sin(2i_3\theta) = \frac{1}{4}(\sin(2(-i_1 + i_2 + i_3)\theta) + \sin(2(i_1 - i_2 + i_3)\theta) + \sin(2(i_1 + i_2 - i_3)\theta) - \sin(2(i_1 + i_2 + i_3)\theta)). \quad (A.3)$$

Also, we would frequently use sums of this type:

$$\sum_{l=1}^{k} D_l \sin(2l\theta) \quad (A.4)$$

for some $D_l$ and $k$. As you can see, we already claim $\varphi_j(\theta)$ has this form.

The above statements will also hold if we will substitute (A.4) with $k = x_i$ instead of $\psi_{x_i}(\theta)$. Also note that the product of sums of type (A.4) has also the similar form, if we multiply an odd number of them.

### B Connection to Chebyshev polynomials

Here, we will discuss an interesting connection between the system of equations, particularly (5.57) and Chebyshev polynomials $T_k(x)$. This is not directly related to the main problem, however it gives some structure to the matrix of the system and this results in interesting facts. Here, we would use Theorem 3 and modify the system once more.

We will predominantly focus on the even modes case, although there could be a similar connection for the odd one. We will also only study $m = 2$ (introduced in (5.55)), since it is the lowest value we consider and it gives the most full view of the system. Other cases of $m$ are similar. We prove the following

**Lemma B.1.** For $m = 2$, equations (5.57) and (5.58) reduce to the following

$$\sum_{j=2}^{k} (-2)^j \frac{(k + j - 1)!}{(k - j)! (2j)!} \left(1 + \cos \left(\frac{p\pi}{k}\right)\right)^j \bar{x}_{2j} \quad (B.1)$$

for any $0 < p < k$, with $\bar{x}_{2j}$ being scaled $x_{2j}$, with scaler independent of $k$. 

40
The fact that scaler is independent of $k$ means that this is the valid substitution in the whole system. In this lemma one can see the connection to Chebyshev polynomials of the first kind, since the following holds

$$T_k(z) = k \sum_{j=0}^{k} (-2)^j \frac{(k+j-1)!}{(k-j)!(2j)!} (1-z)^j.$$  

(B.2)

So, we substitute $z = -\cos \frac{p\pi}{k}$ into the polynomial and associate each element of the sum with $\tilde{x}_{2j}$, and we only lack 2 terms.

This is the connection between the sum and our system. One can use it in variety of different ways. For example, it is easy to compute derivatives of Chebyshev polynomials, using (B.2). Notice that we evaluate polynomial at $z = -\cos \frac{p\pi}{k}$ and $T_k(z)$ has local extrema there with $T_k(z) = 1$; $T_k'(z) = 0$ for odd $p$ and $k$. Computing the value and first derivative in (B.2) results in following:

$$\sum_{j=2}^{k} (-2)^j \frac{(k+j-1)!}{(k-j)!(2j)!} \left(1 + \cos \left(\frac{p\pi}{k}\right) \right)^j (j-1) = 0$$  

(B.3)

for odd $p$ and $k$. So, one can set $\tilde{x}_{2j} = j-1$ and get the following:

Lemma B.2. All the equations $v_{p/(2k)}$ with odd $p$ and $k$ share nontrivial element of the kernel. It is associated with

$$a_{2k+4} = (-1)^k C \frac{k+1}{k+2} \left( \frac{e}{2} \right)^{2k}$$  

(B.4)

for every $C \neq 0$.

C Analyticity and greater eccentricities

Of course, there is a natural generalization to our question, concerning ellipses with other eccentricities. One can conjecture the following:

Conjecture C.1. For any integer $q_0 \geq 3$ and $\delta > 0$, there exist $m = m(q_0, \delta), n = n(q_0, \delta) \in \mathbb{N}$, such that the following holds. For each $0 < e < 1 - \delta$ and $c \geq 0$, there exists $\varepsilon = \varepsilon(e,c,q_0,\delta)$, such that any $q_0$-integrable $C^m$-smooth domain $\Omega$, whose boundary is $C^m - \varepsilon$-close to an ellipse $E_{e,c}$, is itself an ellipse.

Here, we exclude the ellipses with eccentricities close to 1, since there are often some irregularities there.

For example, a similar generalization to [1] was attempted in [9]. There, effectively a case $q_0 = 2$ was studied.

This generalization also requires some new methods. However, we propose a bit weaker idea:
Conjecture C.2. For any integer $q_0 \geq 3$ and $\delta > 0$, there exist $m = m(q_0, \delta), n = n(q_0, \delta) \in \mathbb{N}$, such that the following holds. For each, but finite number of $0 < e < 1 - \delta$ and $c \geq 0$, there exists $\varepsilon = \varepsilon(e, c, q_0, \delta)$, such that any $q_0$-integrable $C^m$-smooth domain $\Omega$, whose boundary is $C^n - \varepsilon$-close to an ellipse $E_{e,c}$, is itself an ellipse.

We want to use analyticity arguments to prove it. We give the following plan for this proof, it will heavily rely on the ideas in [9]. The main idea of their proof was that the preservation of a caustic $\frac{1}{q}$ means that the deformation is orthogonal to some functions $c_{1/q}(\varphi), s_{1/q}(\varphi)$ up to some small value. Then, one needs to prove that these functions with various $q$ make up a basis set in $L^2(\mathbb{T})$. This way, the deformation could not be orthogonal to every one of them.

In our case, we can similarly define functions $c_{p/q}, c_{p/(2q)}, c_{(2p)/(2q)}$ and corresponding $s$ for $\frac{p}{q}$ in its lowest terms and $\frac{p}{q} < \frac{1}{q_0}$. The deformation should be orthogonal to every one of them to preserve needed caustics. For example, we would have that

$$c_{p/q}(\varphi) = \frac{\cos \left( \frac{2 \pi q}{4K(k_{p/q})} F(\varphi; k_{p/q}) \right)}{\sqrt{1 - k_{p/q}^2 \sin^2 \varphi}}$$

(C.1)

We want to study its dependency on $e$ for a given $\frac{p}{q}$ and $\varphi$. This function needs to be shown to be real analytic over $e$ in some sub-interval of $(0, 1)$. The main difficulty would be studying elliptic functions and $k_{p/q}(e)$, since it is defined using inverse formulas.

Let’s assume that we have proven it. We cannot directly use analyticity to prove their independence, since there are infinite number of them. We want to make the space finite-dimensional. So, we pick some big $r_0 > 2M_1$, and break up all these functions in two groups: big and small. In the big group, we are only interested in functions of type $c_{1/q}$ and $s_{1/q}$. According to [9], these functions would be independent of each other and form a subspace with a finite codimension.

We will project all the small caustics onto the finite dimensional space. If we show, the projections are still analytic, then we would be able to just compute their determinant. It would still be analytic, so if it isn’t constant zero, all the vectors will make up a basis set for all but finite number of $e$ – the roots of determinant.

To work with projections, one should work with the infinite sums while projecting. Similar work is done in [9] in Appendix D. To do this, special scalar products and coefficients are introduced. But, as in [9], we want to prove that projection coefficients $d_{jk}$ are analytic and decaying exponentially independently of $e$. If the terms in the sum are analytical and decay exponentially, then the convergence is uniform, and the uniform limit of analytic terms is also analytic.

To prove that the determinant is not constant zero, we consider the case of small eccentricity. Since equations, like (2.8) arise the same way, as functions $s_q$ and $c_q$, so it should be possible to prove a basis property, using the completeness of systems like (5.9).

Of course, there are a large number of facts to prove here. We need analyticity and uniform bounds on $d_{jk}$, as well as smoothness requirements and so on. We intend on studying this topic further.
D Adding bounds on $p$

In the main article we have proven Theorem 1 for $q_0$-integrable domains. That means that we only consider caustics with $\frac{p}{q} < \frac{1}{q_0}$. Note that this is a bound on $p$, that is equivalent to $q_0p < q$. However, it is only a linear bound, so it may not suffice for some problems. The question then arises – can we generalize Theorem 1 and get stronger bounds on $q$? In this section, we argue that we can improve those bounds to be polynomial and even discuss a conditional statement that only uses caustics $\frac{p}{q}$ only for $q > q_0$ and $p = 1, 2, 3$.

D.1 Polynomial bounds

Introduce the corresponding definition:

**Definition D.1.** Let $Q_0(x)$ be a real polynomial with a positive leading coefficient. If the billiard map, associated to $\Omega$ admits integrable rational caustics with rotation numbers $\frac{p}{q}$ in its lowest terms for all $0 < \frac{p}{q} < \frac{1}{2}$, $Q_0(p) < q$, we say that $\Omega$ is $Q_0(x)$-rationally integrable.

Note that $q_0$-rationally integrable domains are $Q_0(x)$-rationally integrable for $Q_0(x) = q_0x$. Ellipses are $Q_0(x)$-rationally integrable for every $Q_0(x)$ with a positive leading coefficient. Now we propose the analogue of Theorem 1 for these domains.

**Lemma D.1.** For any real polynomial $Q_0(x)$ with a positive leading coefficient, there exist $e_0 = e_0(Q_0) \in (0, 1)$, $m = m(Q_0), n = n(Q_0) \in \mathbb{N}$, such that the following holds. For each $0 < e < e_0$ and $c \geq 0$, there exists $\varepsilon = \varepsilon(e, c, Q_0)$, such that any $Q_0(x)$-rationally integrable $C^m$-smooth domain $\Omega$, whose boundary is $C^m - \varepsilon$-close to an ellipse $E_{e, c}$, is itself an ellipse.

Denote $\hat{q}_0 = \max(Q_0(1), 3)$. If $q_0$ is even, increase it by 1. Notice that every $Q_0(x)$-integrable domain admits a caustic with rotation number $\frac{1}{q}$ for every $q > \hat{q}_0$. There are of course other caustics that are admitted. However, notice that in our proof of Theorem 1 we used caustics with rotation numbers $\frac{p}{q}, p > 1$, only during the process of Vandermonde reduction in Section 5. The rest of the proof relies only on caustics with rotation number $\frac{1}{q}$.

Hence, if we set $q_0 = \hat{q}_0$, all of our proof for $q_0$-integrable domains will be valid, except the Vandermonde reduction. We constituted the relation $pq_0 < q$ in it using the definitions of $q_0$-good numbers. So, now we just need to extend these definitions to $Q_0(x)$. This is done naturally:

**Definition D.2.** Let $Q_0(x)$ be a real polynomial with positive leading coefficient. A prime number $q > \hat{q}_0$ is said to be $Q_0(x)$-A-good, if the minimal subgroup of $\mathbb{F}_q^*$ symmetrical by negation, and containing elements $1, 2, \ldots , t$ is $\mathbb{F}_q^*$, where $t + 1$ is the first natural number with $Q_0(t + 1) \geq q$, or $\frac{q + 1}{2}$, whichever is smaller.

**Definition D.3.** Let $Q_0(x)$ be a real polynomial with positive leading coefficient. A prime number $q > \hat{q}_0$ is said to be $Q_0(x)$-B-good, if the minimal subgroup of $\mathbb{F}_q^*$ symmetrical by negation, and containing elements $1, 3, \ldots , 2t - 1$ is $\mathbb{F}_q^*$, where $t + 1$ is the first natural number with $Q_0(t + 1) \geq q$, or $\frac{q + 1}{2}$, whichever is smaller.
Note that once we introduce these objects, Vandermonde reduction will work, assuming our prime modulus is $Q_0(x)$-A-good or $Q_0(x)$-B-good, depending on a mode we are studying. Assuming it exists, set $\hat{M}_0$ to be $k_0 - 1$-st $Q_0(x)$-A-good and $Q_0(x)$-B-good number, where $\hat{q}_0 = 2k_0 - 1$. Since the whole proof now works analogously, Lemma D.1 also holds with

$$n = 2\hat{M}_0, \ m = 28\hat{M}_0.$$ (D.1)

To prove Lemma D.1, we just need the existence of $\hat{M}_0$, meaning we need to proof that there are infinite amount of $Q_0(x)$-A-good and $Q_0(x)$-B-good numbers, just like we did $q_0$ ones. This time, however, we need to change the proof, since the one for $q_0$-good numbers relied on the fact that $\frac{q}{t}$ is bounded, in this notation.

First of all, we only care about behavior at infinity, so we will only analyze the case $Q_0(x) < x^l$ for natural $l$ and $x$. Moreover, we will simplify the notion of $Q_0(x)$-good numbers, introducing the following:

**Definition D.4.** Let $l > 2$ be a natural number. A prime number $q > 3$ is said to be $l$-C-good, if the minimal subgroup of $\mathbb{F}_q^*$, symmetrical by negation, and containing elements $1, 3, \ldots, t$ is $\mathbb{F}_q^*$, where $t$ is the first odd number with $t^l \geq q$.

Note, that we can pick $l$ large enough, so that beginning with some $q_{st}$, every $l$-C-good number will be both $Q_0(x)$-A-good and $Q_0(x)$-B-good. This way, we only need to prove that there are infinite amount of $l$-C-good numbers for $l > 2$.

We will accomplish it, using the notion of maximal prime factor. Assume that $1 \leq k \leq q - 1$ is an odd natural number. $k$ is also an element of $\mathbb{F}_q^*$. If the largest prime factor of $k$ is not greater that $t$ or $k = 1$, then $k$ is present in the minimal subgroup. To estimate the size of the minimal subgroup, we denote $\Psi_{odd}(q - 1, t)$ to be the number of such odd $k$. Also, let

$$\Psi(x, y) = \# \{1 \leq n \leq x : P(n) \leq y\}$$ (D.2)

where $P(n)$ denotes the largest prime factor of $n$, with the convention that $P(1) = 1$.

We want to bound the value $\frac{\Psi_{odd}(q - 1, t)}{q - 1}$. We have that

$$\Psi_{odd}(q - 1, t) = \Psi(q - 1, t) - \Psi\left(\frac{q - 1}{2}, t\right)$$ (D.3)

since we can divide every even number by 2. This is useful, since coefficients $\Psi(x, y)$ are studied in great detail in number theory. In [5] the following asymptotic relation is included:

$$\lim_{y \to \infty} \frac{\Psi(y^u, y)}{y^u} = \rho(u)$$ (D.4)

for $u > 0$ and $\rho(u)$ being the Dickman function, that is non-negative and continuous for $u > 0$. Then,

$$\lim_{q \to \infty} \frac{\Psi_{odd}(q - 1, t)}{q - 1} = \lim_{q \to \infty} \frac{\Psi(q - 1, t)}{q - 1} - \lim_{q \to \infty} \frac{\Psi\left(\frac{q - 1}{2}, t\right)}{q - 1}$$ (D.5)

The first limit is obviously $\rho(t)$, while the second can be evaluated the following way:
\[ \rho(l) \leq \lim_{q \to \infty} \frac{\Psi \left( \frac{q-1}{2}, \left( \frac{q-1}{2} \right)^{1/l} \right)}{q-1} \leq \lim_{q \to \infty} \frac{\Psi \left( \frac{q-1}{2}, t \right)}{q-1} \leq \lim_{q \to \infty} \frac{\Psi \left( \frac{q-1}{2}, \left( \frac{q-1}{2} \right)^{1/m} \right)}{q-1} = \rho(m) \]

for \( 2 < m < l \). So,

\[ \lim_{q \to \infty} \frac{\Psi \left( \frac{q-1}{2}, t \right)}{q-1} = \rho(l) \Rightarrow \lim_{q \to \infty} \frac{\Psi_{\text{odd}}(q-1, t)}{q-1} = \rho(l) \]  

This means that starting from some \( q \), the ratio between the size of minimal subgroup halved (due to symmetry by negation and the fact that we considered only odd elements) and the size of \( \mathbb{F}_q^* \) will be not less than \( \frac{\rho(l)}{3} \). So, the inverse of it will be a natural divisor of \( \frac{q-1}{2} \), not greater than \( \left\lceil \frac{3}{\rho(l)} \right\rceil + 1 \). Since there are infinite amount of primes \( q \), such that \( \frac{q-1}{2} \) do not have such divisors, except for 1, for them the minimal subgroup will be \( \mathbb{F}_q^* \).

So, there are infinitely many \( l \)-C-good primes, so Lemma D.1 holds.

### D.2 Conditional statement for \( p = 1, 2, 3 \)

We can achieve even stronger result, assuming some condition. Suppose the following conjecture holds:

**Conjecture D.1.** There are infinitely many Sofie-Germain primes.

Then introduce

**Definition D.5.** Let \( q_0 \geq 7 \). If the billiard map, associated to \( \Omega \) admits integrable rational caustics with rotation numbers \( \frac{p}{q} \) in its lowest terms for all \( q > q_0, p = 1, 2, 3 \), we say that \( \Omega \) is \( q_0 \)-close-rationally integrable.

Then, we have the following:

**Lemma D.2.** Assume Conjecture D.1 holds.

Then, for any integer \( q_0 \geq 7 \), there exist \( e_0 = e_0(q_0) \in (0, 1), m = m(q_0), n = n(q_0) \in \mathbb{N} \), such that the following holds. For each \( 0 < e < e_0 \) and \( c \geq 0 \), there exists \( \varepsilon = \varepsilon(e, c, q_0) \), such that any \( q_0 \)-close-rationally integrable \( C^m \)-smooth domain \( \Omega \), whose boundary is \( C^n - \varepsilon \)-close to an ellipse \( E_{e,c} \), is itself an ellipse.

The idea behind this lemma is identical to the last one. The proof only changes in the definition of \( q_0 \)-good primes. In this case, one can check that similar definition exists. In it, A-good numbers will be associated with the set \{1, 2\} of residues, generating the whole group, while B-good will be associated with the set \{1, 3\}. Since every Sofie-Germain prime will be both A-good and B-good (we discussed it in the proof of Theorem 1), there would be infinite amount of them, and the lemma would hold, assuming Conjecture D.1.
References

[1] Artur Avila, Jacopo De Simoi, and Vadim Kaloshin. “An integrable deformation of an ellipse of small eccentricity is an ellipse”. In: Annals of Mathematics 184.2 (2016), pp. 527–558. issn: 0003486X.

[2] Misha Bialy. “Convex billiards and a theorem by E. Hops.” In: Mathematische Zeitschrift 214.1 (1993), pp. 147–154.

[3] George D. Birkhoff. “On the periodic motions of dynamical systems”. In: Acta Mathematica 50.1 (1927), pp. 359–379.

[4] Eugène Gutkin. “Billiard dynamics: an updated survey with the emphasis on open problems.” In: Chaos 22.2 (2012), p. 026116.

[5] Adolph Hildebrand and Gerald Tenenbaum. “Integers without large prime factors”. In: Journal de Théorie des Nombres de Bordeaux 5.2 (1993), pp. 411–484.

[6] Guan Huang and Alfonso Sorrentino. “Nearly Circular Domains Which Are Integrable Close to the Boundary Are Ellipses”. In: Geometric and Functional Analysis 28 (Apr. 2018).

[7] C. W. Jones. “Calculus of Finite Differences. By C. Jordan. Second edition. Pp. xxii, 654. 1950. (Chelsea Publishing Co., New York)”. In: The Mathematical Gazette 35.314 (1951), 145–146.

[8] Mark Kac. “Can One Hear the Shape of a Drum?” In: The American Mathematical Monthly 73.4 (1966), pp. 1–23. issn: 00029890, 19300972.

[9] Vadim Kaloshin and Alfonso Sorrentino. On the Local Birkhoff Conjecture for Convex Billiards. 2018.

[10] Donald E. Knuth. The Art of Computer Programming, Vol. 1: Fundamental Algorithms. Third. Reading, Mass.: Addison-Wesley, 1997, p. 59. isbn: 0201896834 9780201896831.

[11] John N. Mather. “Glancing billiards”. In: Ergodic Theory and Dynamical Systems 2.3-4 (1982), 397–403.

[12] Hillel Poritsky. “The Billard Ball Problem on a Table With a Convex Boundary–An Illustrative Dynamical Problem”. In: Annals of Mathematics 51.2 (1950), pp. 446–470. issn: 0003486X.

[13] Steve Zelditch. “The inverse spectral problem”. In: Surveys in differential geometry 9 (2004), pp. 401–467.