The Poncelet Theorem and Billiards

Abhishek Pravin Malpure

A dissertation submitted for the partial fulfilment of BS-MS dual degree in Science

Indian Institute of Science Education and Research Mohali

April 2017

Certificate of Examination

This is to certify that the dissertation titled **"The Poncelet Theorem and Billiards"** submitted by **Mr. Abhishek Pravin Malpure** (Reg. No. MS12090) for the partial fulfilment of BS-MS dual degree programme of the Institute, has been examined by the thesis committee duly appointed by the Institute. The committee finds the work done by the candidate satisfactory and recommends that the report be accepted.

Dr. Pranab Sardar

Dr. Mahender Singh

Dr. Krishnendu Gongopadhyay
(Supervisor)

Dated: April 21, 2017

Declaration

The work presented in this dissertation has been carried out by me under the guidance of Dr Krishnendu Gongopadhyay at the Indian Institute of Science Education and Research Mohali.

This work has not been submitted in part or in full for a degree, a diploma, or a fellowship to any other university or institute. Whenever contributions of others are involved, every effort is made to indicate this clearly, with due acknowledgement of collaborative research and discussions. This thesis is a bonafide record of work done by me and all sources listed within have been detailed in the bibliography.

Abhishek Malpure
(Candidate)

Dated: April 21, 2017

In my capacity as the supervisor of the candidates project work, I certify that the above statements by the candidate are true to the best of my knowledge.

Dr. Krishnendu Gongopadhyay
(Supervisor)

Acknowledgment

I would like to express my sincere gratitude to my supervisor, Dr Krishnendu Gongopadhyay for providing me the opportunity to work under his guidance and giving me timely advice to understand my goals. I sincerely thank Sagar Kalane for the helpful discussions and my friend Aditya Vyas for helping me with LATEX. Last but not the least I would like to thank Indian Institute of Science Education and Research Mohali, IISER Mohali informatics centre for providing good facility and services and my family members for continuous support and encouragement.

Finally, I would like to acknowledge that the material presented in this thesis is based on other people's work. If I have made any contribution then it is the selection, presentation and elaboration of the materials from different sources those are listed in the bibliography.

Abhishek Pravin Malpure

List of Figures

Contents

Abstract

Usually billiards are studied in the framework of the theory of dynamical systems. My work actually emphasizes connections to the geometry and to physics, and billiards are treated here in their relation with geometrical optics. For this I studied various types of geometries like affine and projective. The main aim of my project was billiard study, their transformations and the Poncelet theorem which is consequence of the integral billiard map.

Chapter 1

Affine Geometry

1.1 Geometry and transformations

Definition 1.1. *(Euclids Five Axioms of Geometry). Euclid stated five axioms for Euclidean geometry of the plane.*

1. *A straight line can be drawn between any two points.*
2. *A line can be extended indefinitely in either direction.*
3. *Any segment can be described as radius and circle with any point as center.*
4. *All right angles are equal.*
5. *Through a point not on a line there exists a unique line parallel to the given line.*

Definition 1.2. *An isometry is a distance-preserving transformation.*

Every isometry has a translation along a line, a reflection in a line, a rotation about a point, a composite of translations, reflections and rotations in $\mathbb{R}^2$. If we compose two isometries then it will give an isometry and also it forms a group under composition of functions which helps in building transformations in proving Euclidean results.

The most commonly used transformation is an Euclidean transformation. It is either a translation, a rotation, or a reflection.

Definition 1.3. *A **Euclidean transformation** of $\mathbb{R}^2$ is a function $t : \mathbb{R}^2 \to \mathbb{R}^2$ of the form*

$$t(\boldsymbol{x}) = \boldsymbol{U}\boldsymbol{x} + \boldsymbol{a}$$

where $\boldsymbol{U}$ is an orthogonal 2 × 2 matrix and $\boldsymbol{a} \in \mathbb{R}^2$.

We can say that every isometry of the plane is an Euclidean transformation of $\mathbb{R}^2$.

1.2 Affine Transformations

Length and angle measure are preserved by Euclidean transformation. Moreover, the shape of a geometric object will not change. That is, lines transform to lines, planes transform to planes, circles transform to circles, and ellipsoids transform to ellipsoids. Only the position and *orientation* of the object will change. Affine transformations are generalizations of Euclidean transformations. Under affine transformations, lines transforms to lines; but, circles become ellipses. Length and angle are not preserved. In this section we will see geometries in $\mathbb{R}^2$ - *affine geometry* which consists of the space $\mathbb{R}^2$ together with a group of transformations, the *affine transformations*, acting in $\mathbb{R}^2$

Definition 1.4. *An **affine transformation** of $\mathbb{R}^2$ is a function $t : \mathbb{R}^2 \to \mathbb{R}^2$ of the form*

$$t(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \tag{1.1}$$

where $\boldsymbol{A}$ is an invertible 2 × 2 matrix and $\boldsymbol{b} \in \mathbb{R}^2$.

Remark : Since every orthogonal matrix is invertible, every Euclidean transformation of $\mathbb{R}^2$ is an affine transformation of $\mathbb{R}^2$.

1.3 Properties of Affine Transformations

The set of affine transformations forms a group under composition of functions.

Basic Properties of Affine Transformations

The affine transformations:

1. maps a line to a line;
2. maps a line segment to a line segment;
3. preserves the property of parallelism among lines and line segments;
4. preserve ratios of lengths of the two parallel segments.

Images of sets under Affine transformations

We can find the image of a line under affine transformation with the help of the definition of affine transformation. The set of such transformations forms a group, in which the inverse of the transformation t is given by

$$t^{-1}(\mathbf{x}) = A^{-1}\mathbf{x} - A^{-1}\mathbf{b}. \tag{1.2}$$

We use equations (1.1) and (1.2) to find images under t. To avoid the confusion we denote the symbol $\mathbf{x}$ and the coordinates (x, y) for points in the domain of t, and use the symbol $\mathbf{x}'$ and the coordinates (x', y') to denote the image of $\mathbf{x}$ under t. With this notation, we may rewrite equations (1.1) and (1.2) in the form

$$\mathbf{x}' = A\mathbf{x} + \mathbf{b},$$

$$\mathbf{x} = A^{-1}\ \mathbf{x}\ ' - A^{-1}\mathbf{b}.$$

1.4 The Fundamental Theorem of Affine Geometry

In this section we prove that all triangles are affine congruent and they share the same affine properties. Since a triangle is completely determined by its three vertices, the congruence of triangles follows from the *Fundamental Theorem of Affine Geometry* which states that any three non-collinear points can be mapped to any other three non-collinear points by affine transformation.

Theorem 1. ***Fundamental Theorem of Affine Geometry***
Let $\boldsymbol{p}$, $\boldsymbol{q}$, $\boldsymbol{r}$ *and* $\boldsymbol{p}'$, $\boldsymbol{q}'$, $\boldsymbol{r}'$ *be two sets of three non-collinear points in* $\mathbb{R}^2$*. Then :*

1. *there is an affine transformation* t *which maps* p, q *and* r *to* p', q' *and* r'*, respectively;*
2. *the affine transformation* t *is unique.*

Proof We first show special triplet of vectors.

$$\vec{0} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \vec{i} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \vec{j} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

can be mapped by an appropriate affine transformation to arbitrary triplet of vectors.

$$\vec{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}, \vec{q} = \begin{bmatrix} q_1 \\ q_2 \end{bmatrix}, \vec{r} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

corresponds to three non-collinear points.

$$\begin{aligned} p &= t(0,0) = (e,f) & \text{so } (e,f) &= p \\ q &= t(1,0) = (a,c) + (e,f) & (a,c) &= q - p \\ r &= t(0,1) = (b,d) + (e,f) & (b,d) &= r - p \end{aligned}$$

$$\mathrm{A} = \begin{bmatrix} q_1 - p_1 & r_1 - p_1 \\ q_2 - p_2 & r_2 - p_2 \end{bmatrix}$$

$$\vec{b} = \vec{p} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$$

Note columns of A correspond to vectors $\vec{q}$-$\vec{p}$ and $\vec{r}$-$\vec{p}$. Since points (p_1, p_2), (q_1, q_2), (r_1, r_2) are non-collinear, vectors $\vec{q}$-$\vec{p}$ and $\vec{r}$-$\vec{p}$ are non-parallel vectors. Hence determinant is non-zero. Thus A is invertible and

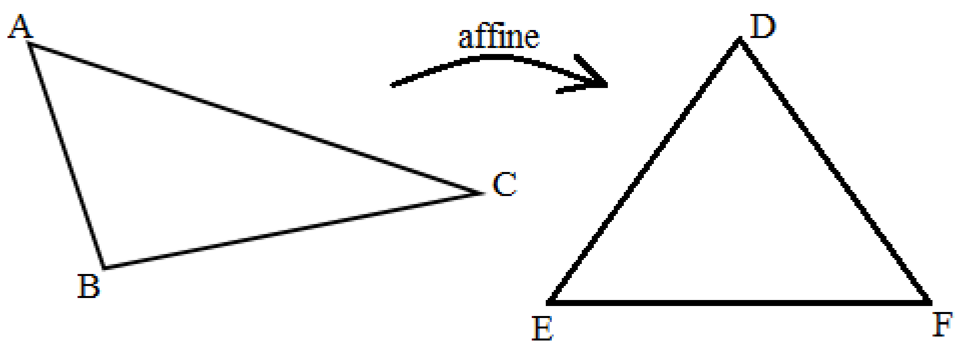
$$t(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{b}.$$

is affine transformation by definition.

1. Let $\vec{p}$, $\vec{q}$, $\vec{r}$ and $\vec{p}\,'$, $\vec{q}\,'$, $\vec{r}\,'$ be two ordered triples of position vectors representing two arbitrary triples of non-collinear points. Using the result we have just proven, there exist affine transformations t_1 and t_2 mapping the special triple $\vec{0}$, $\vec{i}$, $\vec{j}$ to $\vec{p}$, $\vec{q}$, $\vec{r}$ and to $\vec{p}\,'$, $\vec{q}\,'$, $\vec{r}\,'$ respectively. Then $t = t_2 \text{ o } {t_1}^{-1}$ is an affine transformation, and it maps $\vec{p}$, $\vec{q}$, $\vec{r}$ to $\vec{p}\,'$, $\vec{q}\,'$, $\vec{r}\,'$ respectively.

2. Suppose we have affine transformations t and s which maps $\vec{p}$, $\vec{q}$, $\vec{r}$ to $\vec{p}\,'$, $\vec{q}\,'$, $\vec{r}\,'$ respectively. Let t_1 be affine transformation defined in part (1). Then compositions $t \text{ o } t_1$ and $s \text{ o } t_1$ are both affine transformations which map $\vec{0}$, $\vec{i}$, $\vec{j}$ to the points $\vec{p}\,'$, $\vec{q}\,'$, $\vec{r}\,'$ respectively. Since an affine transformation is determined uniquely by its effect on the points $\vec{0}$, $\vec{i}$, $\vec{j}$ it follows that $t \text{ o } t_1 = s \text{ o } t_1$. If we then compose both $t \text{ o } t_1$ and $s \text{ o } t_1$ on right with ${t_1}^{-1}$, it follows that $t = s$. □

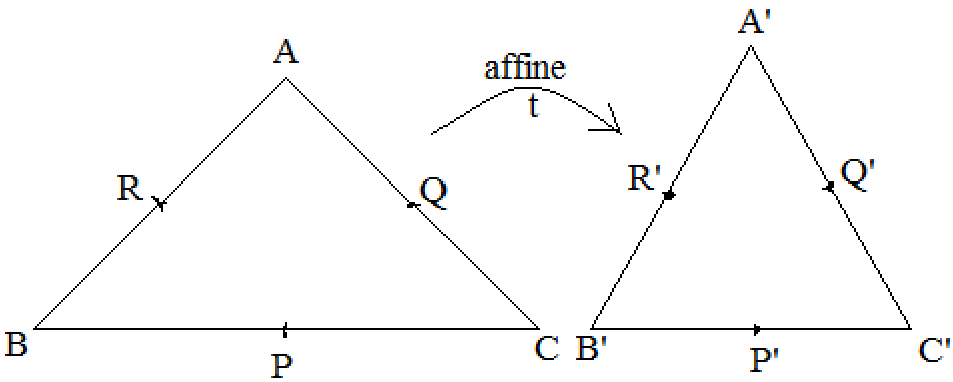
Remark : Two figures are **affine-congruent** if there is an affine transformation which maps one onto the other. For example all triangles are affine-congruent.



1.5 Using Fundamental theorem of Affine Geometry

By fundamental theorem of Affine Geometry there is an affine transformation which maps vertices A, B, C to D, E, F respectively. Since this transformation maps straight lines to straight lines, it must map the sides of Δ ABC to Δ DEF sides. So how we use this theorem to deduce that fact by theorem given below.

Theorem 2. *The medians of any triangle are concurrent from special case that medians of an equilateral triangle are concurrent.*

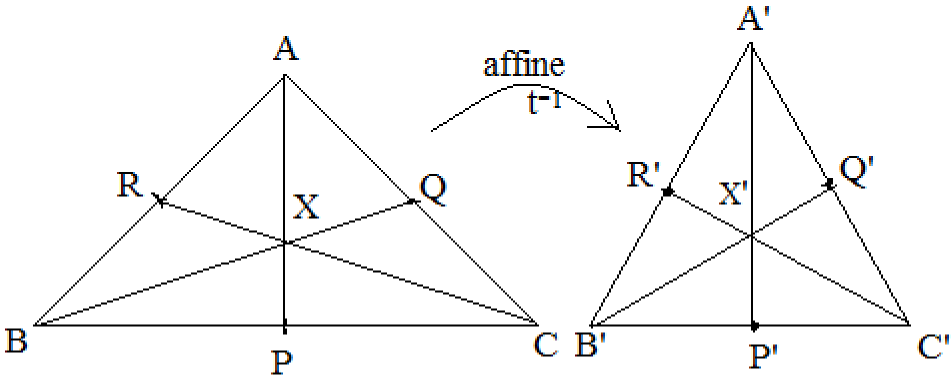


Proof Consider an equilateral triangle ΔABC, with medians AP, BQ, and CR. Since ΔABC. The medians are concurrent if the triangle is equilateral.

If you have to show that the medians of an arbitrary triangle meet at a point, consider arbitrary ΔABC, and let P, Q, and R be the midpoints of BC, CA and AB, which are sides of the triangle respectively. Now, choose a particular equilateral triangle ΔA′B′C′ and also let P′, Q′ and R′ be the midpoints of the sides B′C′, C′A′ and A′B′, respectively.

There is an affine transformation t which maps ΔABC onto ΔA′B′C′. Since affine transformations preserves ratios of lengths along lines, it follows that t maps the midpoints P, Q, R to midpoints P′, Q′ and R′.

From above we know that the medians of any equilateral triangle meet at a point so in particular we know that A′P′, B′Q′ and C′R′ meet at point X'.



Now t has an inverse t^{-1} which also is a affine transformation. then there is a inverse mapping of medians A′P′, B′Q′ and C′R′ back to the medians AP, BQ, and CR of original ΔABC. We can see that X′ lies on all three of the lines A′P′, B′Q′ and C′R′, it follows that t^{-1} maps X′ to an arbitrary point X which is lying on all three of the lines AP, BQ and CR i.e. medians of ΔABC are concurrent. □

Chapter 2

Projective Geometry

In Euclidean geometry, the object sides have lengths, angle are determined by intersecting the lines, and parallel condition is determined if two lines lie in the same plane which never meets. Applying Euclidean transformations these properties are unchanged. Euclidean geometry is actually a subset of what is known as projective geometry which we will discuss in the following sections.

2.1 The Projective plane $\mathbb{RP}^2$

In this section we begin our discussion of projective geometry by first investigating its space of points.

2.1.1 Projective points

If an eye is situated at the origin of $\mathbb{R}^3$ which is looking at a fixed screen. Each point is determined by the ray of light which enters the eye from that point. If we want this idea in mathematical terms, we have a point which is said to be projective point which is a Euclidean line in $\mathbb{R}^3$ passing through the origin.

Definition 2.1. ***A Point*** *(or* ***projective point****) is a line in* $\mathbb{R}^3$ *that passes through the origin of* $\mathbb{R}^3$*. The real projective plane* $\mathbb{RP}^2$ *is the set of all such Points.*

Using an algebraic notation we can specify the points of $\mathbb{RP}^2$. We mention the point $[a, b, c]$ as the point with the homogeneous coordinates $[a, b, c]$.

Definition 2.2. *Let us take a point* (a, b) *in the Euclidean plane. To represent this same point in the projective plane, we simply add a third coordinate of 1 at the*

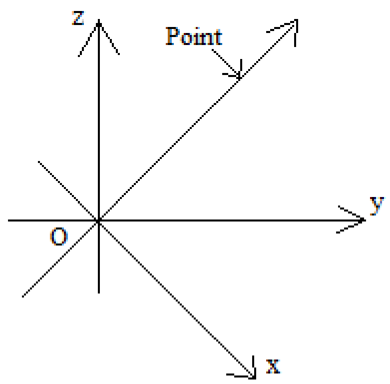


Figure 2.1: Projective point

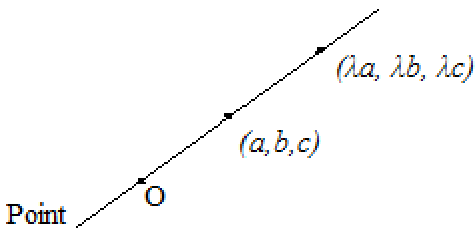
end: $(a, b, 1)$. *Overall scaling is unimportant, so the point* $(a, b, 1)$ *is the same as the point* $(\alpha x, \alpha y, \alpha)$, *for any nonzero* α. *In other words, The expression* $[a, b, c]$, *where* a, b, c *are not all zero represents the point* P *in* $\mathbb{RP}^2$ *which consists of the unique line in* $\mathbb{R}^3$ *that passes through* $(0, 0, 0)$ *and* (a, b, c) *and we refer to* $[a, b, c]$ *as* ***homogeneous coordinates*** *of* P. *If* (a, b, c) *has position vector* $\boldsymbol{v}$. *Then we often denote* P *by* $[\boldsymbol{v}]$ *and we say that* P *can be represented by* $\boldsymbol{v}$.

Notice that there is no uniqueness in homogeneous coordinates of a point. For example, suppose if (a, b, c) is any point on a line passing through the origin, then $(\lambda a, \lambda b, \lambda c)$ also lies on the line where λ is any real number. Moreover, if (a, b, c) is not at the origin and $\lambda \neq 0$. We express this by writing

$$[a, b, c] = [\lambda a, \lambda b, \lambda c], \qquad \text{for any } \lambda \neq 0.$$

Suppose if there is no non-zero real number λ such that $[a, b, c] = [\lambda a', \lambda b', \lambda c']$, then the homogeneous coordinates $[a, b, c]$ and $[a', b', c']$ represent different points of $\mathbb{RP}^2$.

Further, $[a', b', 1] = [a'', b'', 1]$ if and only if $a' = a''$ and $b' = b''$.



We defined projective points above, now we'll define what is a **projective figure**. Just as a figure in Euclidean geometry is defined to be a subset of $\mathbb{R}^2$, so figures in a projective geometry is a subset of $\mathbb{R}^2$. which are sets of lines in $\mathbb{R}^3$ that are passing through the origin. The examples of projective figures is a double cone with a vertex at O and a double square pyramid with a vertex at O.

2.1.2 Projective Lines

Just as we saw before in which we used 'Point' to refer to a 'projective point', just like that we define 'Line' to refer to a 'projective line'. It is actually the completion of the affine line having a projective point, the point at infinity.

Definition 2.3. ***A Line*** *(or* ***projective line****) in $\mathbb{RP}^2$ is a plane in $\mathbb{R}^3$ that passes through the origin. Points in $\mathbb{RP}^2$ are collinear if they lie on a line.*

The general equation of a line in $\mathbb{RP}^2$ is $ax + by + cz = 0$, where a, b, c are real and not all zero.

Collinearity Property of $\mathbb{RP}^2$
Any two distinct points of $\mathbb{RP}^2$ lie on a unique Line.
Strategy Let us find out an equation for the line in $\mathbb{RP}^2$ through the Points $[d, e, f]$ and $[g, h, k]$:

1. we write down an equation as $\begin{vmatrix} x & y & z \\ d & e & f \\ g & h & k \end{vmatrix} = 0$

2. We have to obtain the equation which is needed in the form $ax + by + cz = 0$. This can be done by expanding the determinant in terms of the entries in the first row.

It is also possible to calculate line equation through two points without using the determinant.
Strategy We have to find whether the three points $[a, b, c]$, $[d, e, f]$ and $[g, h, k]$ are collinear or not :

1. Calculate the determinant $\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & k \end{vmatrix}$;

2. If the determinant is zero then only the Points $[a, b, c]$, $[d, e, f]$ and $[g, h, k]$ are collinear.

The points $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$ are known as the triangle of reference. The point $[1, 1, 1]$ is called the unit point.

Incidence Property of $\mathbb{RP}^2$.
Any two distinct Lines in $\mathbb{RP}^2$ intersect in a unique point of $\mathbb{RP}^2$.

Let π be any plane in $\mathbb{R}^3$ not passing through the origin O. Then there is one-one correspondence between the points of π and those points of $\mathbb{RP}^2$ piercing π. Those points of $\mathbb{RP}^2$ not piercing π are called the **ideal Points** for π.
The set of ideal Points for π is actually a plane through O which is parallel to π, called the **ideal Line** for π.

So an **Embedding plane** is a plane together with the set of all Ideal points for π, not passing through the origin. **Standard embedding plane** is the plane in $\mathbb{R}^3$ with equation $z = 1$. The mapping of $\mathbb{RP}^2$ into the standard embedding plane is called the standard embedding of $\mathbb{RP}^2$.

Remark : As there is no parallel lines dependence on the embedding planes choice, the parallel lines concept is meaningless.

2.2 Projective Transformations

We are known with the idea that a geometry consists of a space of points together with a group of transformations acting on that space. As we have introduced the space of projective points $\mathbb{RP}^2$, now we will describe the transformations of $\mathbb{RP}^2$.

Recall that a point of $\mathbb{R}^3$ (other than the origin) on an embedding plane π (that does not pass through the origin) has coordinates $\mathbf{x}$ = (x,y,z) with respect to the standard basis of $\mathbb{R}^3$, and homogeneous coordinates of the corresponding Point $[\mathbf{x}]$ in $\mathbb{RP}^2$ are $[\lambda x, \lambda y, \lambda z]$ for some real $\lambda \neq 0$. Since the points of $\mathbb{RP}^2$ are just lines through the origin of $\mathbb{R}^3$ onto the lines through the origin of $\mathbb{R}^3$.

Definition 2.4. ***A Projective transformation*** *of* $\mathbb{RP}^2$ *is a function* $t : \mathbb{RP}^2 \to \mathbb{RP}^2$ *of the form*

$$t : [\boldsymbol{x}] \mapsto [\boldsymbol{Ax}]$$

where $\boldsymbol{A}$ *is an invertible 3* $\times$ *3 matrix. Here* $\boldsymbol{A}$ *is a matrix associated with* t*. P(2) is the set of all projective transformations of* $\mathbb{RP}^2$.

If $\mathbf{A}$ is a matrix associated with t, so is $\lambda\mathbf{A}$ for any non-zero number λ.

This P(2) is the set of projective transformations forming a group under the operation of composition of functions. In particular, if t_1 and t_2 are projective transformations with associated matrices A_1 and A_2 respectively, then so are t_1 o t_2 and ${t_1}^{-1}$ projective transformations with associated matrices A_1A_2 and ${A_1}^{-1}$.

Strategy Composing two projective transformation t_1 and t_2 :

1. Writing the matrices A_1 and A_2 associated with t_1 and t_2;
2. Calculate the matrices A_1A_2;
3. write down the composition t_1 o t_2 with which the matrices A_1A_2 is associated.

Strategy Now in finding the inverse of a projective transformation t :

1. Writing the matrix $\mathbf{A}$ associated with t;
2. calculate the matrices ${A_1}^{-1}$;
3. write down the inverse t^{-1} with which the matrices ${A_1}^{-1}$ is associated.

Strategy If we want to find the image of a Line in the form $ax + by + cz = 0$ under a projective transformation $t : [\mathbf{x}] \mapsto [\mathbf{Ax}]$:

1. writing the equation of the Line in the form $\mathbf{Lx} = 0$, where $\mathbf{L}$ is the matrix (abc);
2. finding a matrix $\mathbf{B}$ associated with t^{-1};
3. writing the equation of the image as $\mathbf{(LB)x} = 0$.

2.3 The Fundamental theorem of Projective Geometry

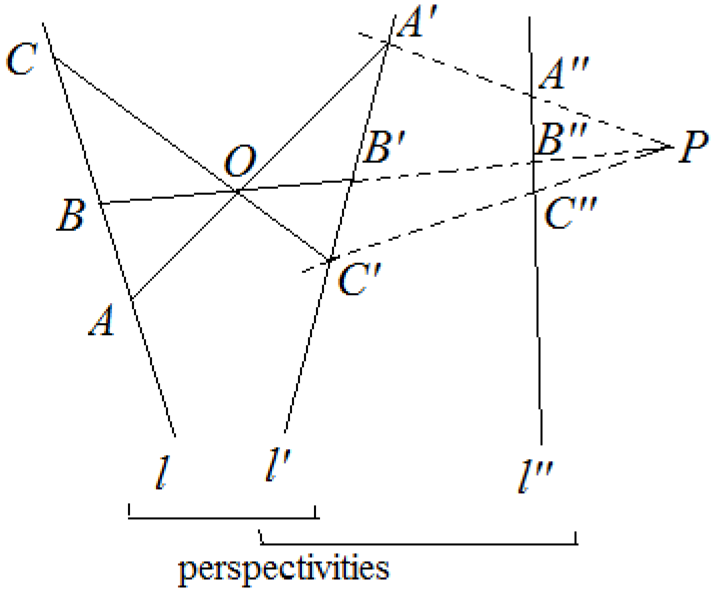
Let us first look at two dimensional situation where we are only talking about lines and perspectivities between the lines.

Here is the point O which is called the centre of perspectivity to the coordinate points on a line l onto line l'.

So we connect A to A', B to B' and we continue for all the points on these line l associating points on line l'. That association of points on l to points on l' is called perspectivity.

Now we have another line l'' with perspectivity point P (say). which means we draw lines from P. So A' corresponds to A'', B' to B'' and all the points corresponds to l'' from l'. That's another perspectivity in which one has centre as O and other has centre P.

So projectivity is the sequence of perspectivities. We can have many perspectivities followed by one other. Here we have one perspectivity followed by another perspectivity and that gives projectivity which is defined as association of points on line l to points on l''.



We can find a projective transformation which maps any set of four points to any set of four points. The only constraint is that no three of the Points in either set can be collinear. In the following statement of the Fundamental Theorem we express that each of the four sets of points lie at the vertices of some quadrilateral, where the quadrilateral is defined as a set of four points A, B, C, D (no three of which are collinear), together with the Lines AB, BC, CD, and DA.

Theorem 3. ***Fundamental Theorem of Projective Geometry***
Let $ABCD$ and $A'B'C'D'$ be two quadrilaterals in $\mathbb{RP}^2$. Then :

1. *there is a projective transformation t which maps A to A', B to B', C to C', D to D';*
2. *the projective transformation t is unique. [BEG99]*

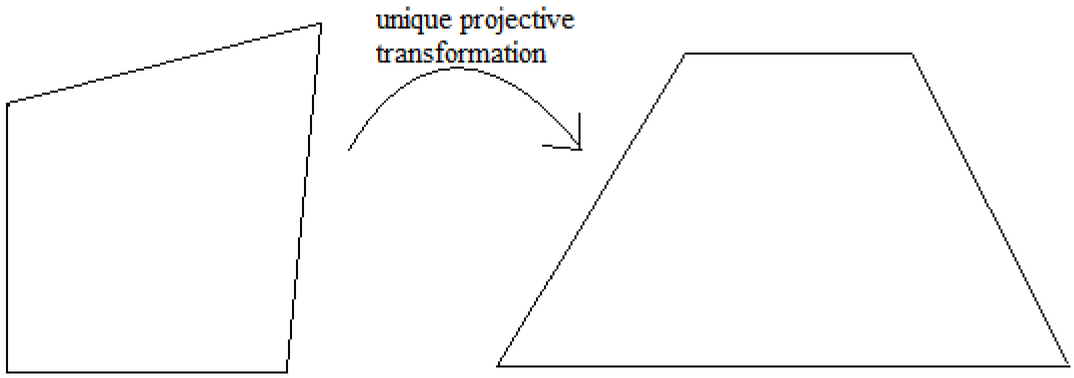
Proof There is a projective transformation which maps the points $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$, $[1, 1, 1]$ to the points A, B, C, D respectively. Similarly there is a projective transformation which maps the points $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$, $[1, 1, 1]$ to the points A', B', C', D' respectively.
The composite $t = t_2 o {t_1}^{-1}$ is then a projective transformation which maps A to A', B to B', C to C', D to D'.
In checking the uniqueness of t, we check that the identity transformation is the only projective transformation which maps each of the points $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$, $[1, 1, 1]$ to themselves.
Next suppose we have two projective transformations t and t' satisfying the conditions of the theorem. Then the compositions ${t_2}^{-1}$ o t o t_1 and ${t_2}^{-1}$ o t' o t_1 must both be projective transformations which mapping points $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$, $[1, 1, 1]$ to themselves. Since we can see that both compositions are equal to the identity, so we deduce that

$${t_2}^{-1} \text{ o } t \text{ o } t_1 = {t_2}^{-1} \text{ o } t' \text{ o } t_1.$$

Composing both sides of this equation with t_2 on the left and with ${t_1}^{-1}$ on the right, we obtain $t = t'$ as needed. □

The Fundamental Theorem tells us that there is a projective transformation which maps any given quadrilateral onto any other given quadrilateral. So we have the following corollary.

Corollary 2.1. *All quadrilaterals are projective-congruent.*



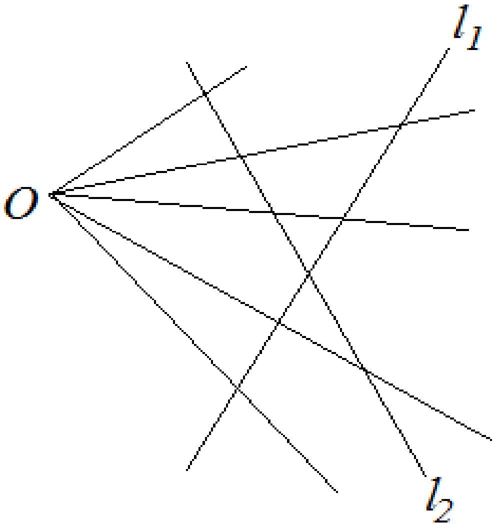
Projective Geometry is the study of figures in $\mathbb{RP}^2$ which has property that they are preserved by projective transformations. There are two important properties of projective geometry which is collinearity and incidence.

Collinearity property : Any two distinct points lie on a unique line.
Incidence property : Any two distinct Lines meet in a unique point.
We can obtain one property from the other by interchanging the words 'Point' and 'Line'. We say that each statement is the dual of other statement. Triangles are the self dual figures.

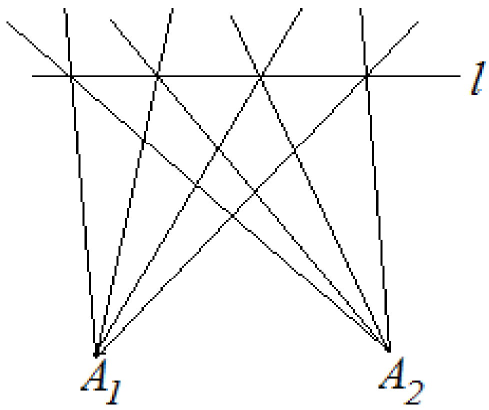
We talked about the perspectivity.

1. **Perspectivity between the two ranges :** Suppose if we have line l_1 and line l_2, then take a point O connecting lines through O to line l_1 and line l_2. That gives association of all points on l_1 with all points on l_2.

2. **Perspectivity between two pencils :** There is complete duality between the above perspectivity and perspectivity between the pencils. Now we need line of perspectivity. Pencil of lines usually means the straight lines that are incident with one point.



(1)



(2)

Chapter 3

Billiards and Geometry

Mathematical billiards describes the motion of a mass point in a particular domain containing elastic reflections off the boundary. In differential geometry, the billiard flow is the geodesic flow on a manifold with boundary.

3.1 Introduction to Mathematical Billiards

A billiard table is a Riemannian manifold M having a piecewise smooth boundary. The billiard dynamical system in manifold M is described by the free motion of a mass-point which we can say it as a billiard ball containing elastic reflection in the boundary such that the point moves along a straight line in M with a unit speed until it hits the boundary. In two dimension this collision is described with the familiar law : the angle of incidence equals the angle of reflection.

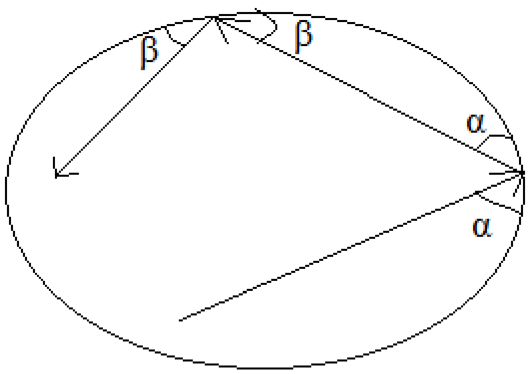


Figure 3.1: Billiard Reflection

As soon as the reflection happens, the ball (point) continues its free motion with the new velocity until it hits the boundary again. Thus there are many features in common between the theory of billiards and the theory of geometrical optics. We assume that the reflection occurs at the smooth point of the boundary. Suppose if the billiard ball hits a corner of the table, its further motion is not defined and the motion of the ball terminates right there.

Now let us discuss another source of motivation for the study of billiards, geometrical optics. According to the **Fermat principle**, light propagates from point A to point B in the least possible time. In Euclidean geometry this means that the light chooses the straight line AB.

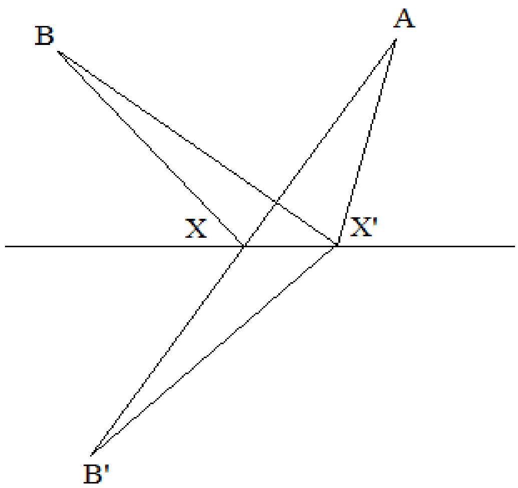


Figure 3.2: Flat mirror reflection

As a consequence of the Fermat principle, billiard reflection law is obtained.

The billiard transformation acts on those pairs $(x; v)$ with $x \in \mathrm{M}$; $v \in T_x\mathrm{M}$ whose trajectories undergo finitely many reflections in the boundary avoiding the corners on the time interval $[0,\ t]$. The phase space of billiard is the unit tangent bundle to M and the configuration space is the manifold M.

3.2 Billiard Ball Map and Invariant measure

We will discuss here billiard ball map. The billiard transformation has an invariant area form. First of all we deal with discrete time system vs continuous time system. In billiard situation if one has a point it moves inside is continuous time system but then there is a discontinuity when we hit the boundary and reflect. For various reasons, it is better to avoid this discontinuity but to treat continuous time systems for discrete time systems. For that purpose we can have argument in two ways. One way is to consider only those position of the billiard ball which are on the boundary curve.

Let M be a bounded plane billiard table. Consider a point on the boundary and unit tangent vectors $(x;\, v)$ with the inward direction v. A vector $(x;\, v)$ moves along the straight line through x in the direction of v to the next point of its intersection x'. We are describing a map which takes unit tangent vector on the foot point of the boundary to the other. We call it as a **billiard ball map**. It is a discrete time system but the configuration space where this x is two dimension, one degree of freedom for the point and other for the direction. Whereas if we want to do a continuous time system, it is three dimension containing two degrees of freedom inside and one for the direction.
We define the billiard transformation T as $T(x; v) = (x'; v')$.
A fundamental property of the billiard ball map is the existence of an invariant area form.

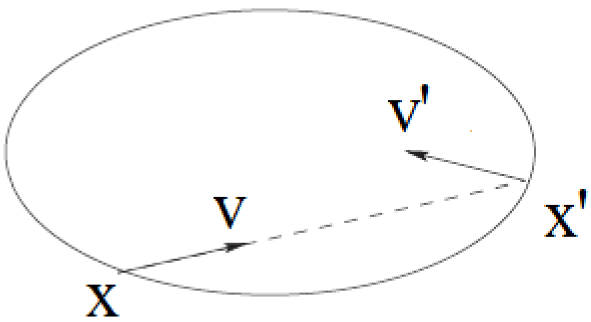


Figure 3.3: Billiard ball map

Theorem 4. *The billiard ball map has a T-invariant form $\omega = sin\ \alpha\ d\alpha \wedge dt$.*

Proof The proof is an application of understanding of billiard ball map. Let us introduce coordinates in figure 3.3. One is arc length parameter of the curve t, where t varies from 0 to L (L is the length of the curve) and second is the angle made by the positive direction of the curve α where $(\alpha \in \pi)$.
Let $T(t; \alpha) = (t_1; \alpha_1)$. To prove its invariance, let us consider a function $f(t, t_1)$, where $f(t, t_1)$ is the distance between the points $x(t)$ and $x(t_1)$. The partial derivatives of

the function with respect to its ordinates t and t_1. Now let us see the point t_1. The perpendicular component of the point doesn't change the length while the tangential component changes the length. So, only tangential component matters and it is given by $cos\ \alpha_1$. Therefore,

$$\frac{\partial f}{\partial t_1} = cos\ \alpha_1 \qquad \text{and} \qquad \frac{\partial f}{\partial t} = -cos\ \alpha.$$

Hence $$df = \frac{\partial f}{\partial t} dt + \frac{\partial f}{\partial t_1} dt_1 = -cos\ \alpha\ dt + cos\ \alpha_1\ dt_1.$$

and hence,

$$0 = d^2 f = sin\ \alpha\ d\alpha \wedge dt - sin\ \alpha_1\ d\alpha_1 \wedge dt_1.$$

This means that ω is a T -invariant area form. □

3.3 Billiard Transformation of the Space of Rays in the Plane

It would be more in the spirit of geometrical optics to deal with space of rays of light. So in dimension two a ray of light is just an oriented line. It can be characterized by choosing the positive direction of reference, an angle ϕ, and its signed distance p from the origin O (the sign of p is given by the right hand rule and depends on the orientation of the frame).

We identify the space of lines with an infinite cylinder $(S^1 \times \mathbb{R}^1)$ (ϕ varies on the circle and p is real number). We have the area form on the space of lines which is $\Omega = dp \wedge d\phi$.

Suppose if we change the origin O, how much of this coordinates depend upon the choice of the origin i.e. if we move the origin O by some vector we don't change the coordinate φ, we change the coordinate p by the linear combination of first harmonic *sin* and *cosine* φ.

Change $O \leftrightarrow p \mapsto p + a\ cos\ \varphi + b\ sin\ \varphi$ and $\varphi' = \varphi$. Another characterizations of this area form is invariance under motions of the plane.

- The area form Ω is invariant under the motions of the plane. Every orientation preserving motion is a composition of a rotation about the origin and a parallel translation.

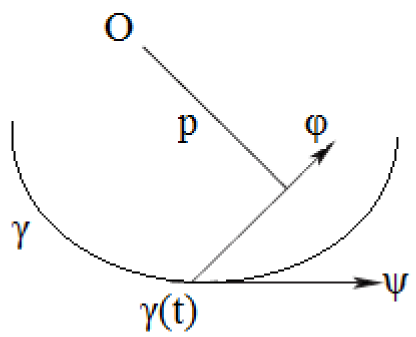


Figure 3.4: Relating two area forms

- The next claim is that the area form Ω coincides with the area form ω.

Proposition 3.3.1 : $\Omega = \omega$.

Proof Previously when we wrote the formula for ω as $sin\ \alpha\ d\alpha \wedge dt$, it is definitely dependent on the billiard table where as the Ω doesn't depend on anything only on the metric in the space where the action takes place.
The two spaces, M and N, are related by the map $\Phi : M \to N$. Let $(t,\ \alpha)$ be the coordinates in M and $(\varphi,\ p)$ the respective coordinates in N. Denote by $\Psi(t)$ the direction of the positive tangent line to the curve γ at point $\gamma(t)$, and let γ_1 and γ_2 be the two components of the position vector γ that associates the oriented line with a unit vector. We have few relations and one of them is :

$$\varphi = \psi + \alpha \quad \text{and} \quad \begin{aligned} p &= \gamma \times\ (cos\ \varphi,\ sin\ \varphi) \\ &= \gamma_1\ sin\ \varphi - \gamma_2\ cos\ \varphi \end{aligned}$$

Differentiating we get,

$$d\varphi = d\alpha + \psi'\ dt \quad \text{and} \quad dp = (\ \gamma_1{}'\ sin\ \phi - \gamma_2{}'\ cos\ \varphi\ )\ dt + (\ \gamma_1\ cos\ \varphi + \gamma_2\ sin\ \varphi\ )\ d\varphi$$

By taking wedge product it follows that,

$$d\varphi \wedge dp = (\ \gamma_1{}'\ sin\ \phi - \gamma_2{}'\ cos\ \varphi\ )\ d\alpha \wedge dt.$$

Since $(\gamma_1{}',\ \gamma_2{}') = (cos\ \varphi,\ sin\ \varphi)$, we have $\gamma_1{}'\ sin\ \phi - \gamma_2{}'\ cos\ \varphi = sin\ \alpha$ and therefore

$$d\varphi \wedge dp = sin\ \alpha\ d\alpha \wedge dt.$$ □

3.4 Elliptical Billiards and confocal conics in the Euclidean plane

Let us see another notion of geometrical optics used in Mathematical billiards. The simplest billiard table is a circular one in which each trajectory makes a constant angle with the boundary and remains tangent to the concentric circle. The transformation which is induced on this tangent circle is a rotation through a fixed angle, which is a translation. Let us first see the concept.

Definition : The *elliptical billiard* is a dynamical system of the unit mass moving under inertia, or in other words, with a constant velocity inside an ellipse and obeying the reflection law at the boundary. [Tab95]

Any segment of a given elliptical billiard trajectory is tangent to the same conic, confocal with the boundary.

Theorem 5 (Chasles Theorem)**.** *The theorem states that each segment of a given billiard trajectory is tangent to a fixed conic that is confocal to the boundary. This conic is called the* ***caustic*** *of the given trajectory. [Figure 3.5]*

Moreover, for any pair of conic, there is a projective transformation of coordinates such that conics become confocal in the new coordinates. Then the polygonal lines inscribed in one of the conics and circumscribed about the other conic will become billiard trajectories.

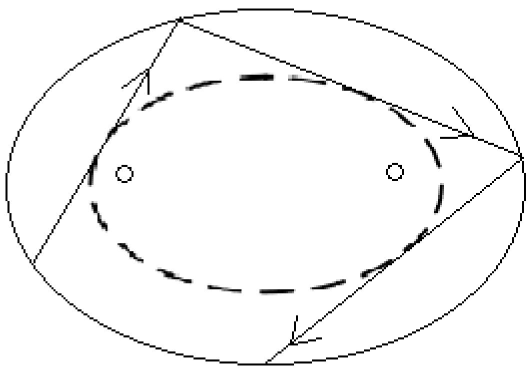
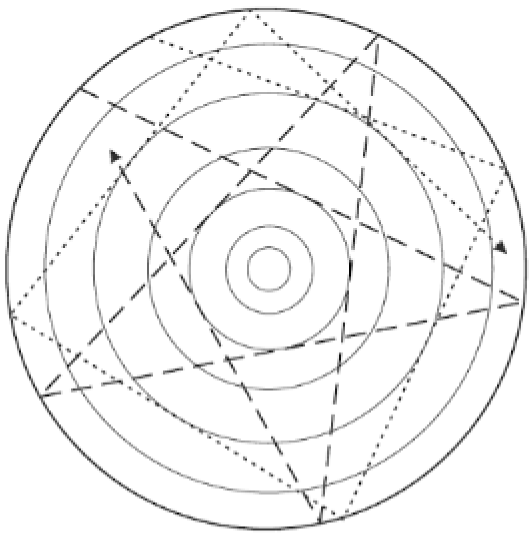


Figure 3.5: Caustic of the billiard trajectory

Suppose we have Billiard table which has round circle, then by symmetry every concentric circle is a caustic simply because of rotation of symmetry.



Elliptical Billiards : The next case to consider is that of conics. Recall that an ellipse consists of points whose sum of distances to two given points F_1 and F_2 is fixed which are called the foci of an ellipse. An ellipse can be constructed using a string, whose ends are fixed at the foci (the method carpenters and gardeners actually use). See Figure 3.6.

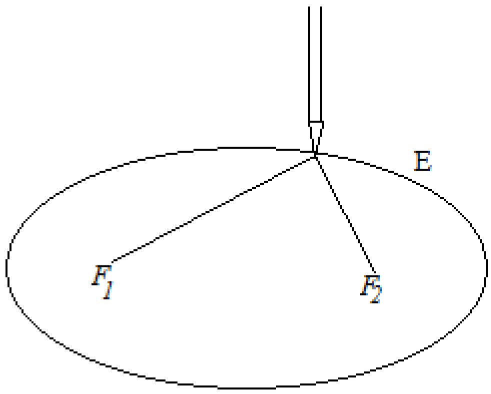


Figure 3.6: Gardener's construction for an ellipse

Similarly, the sum of distances replaced by the absolute value of their difference is an hyperbola and a parabola is the set of points at equal distances from a given point (the focus) and a given line (the directrix). Ellipses, hyperbolas and parabolas all have second order equations in Cartesian coordinates.

The construction of an ellipse with given foci has a parameter, the length of the string. A family of conics all of which share the same foci is called confocal. An equation that describes a confocal family that includes both ellipses and hyperbolas is

$$\frac{x_1^2}{a_1^2 + \lambda} + \frac{x_2^2}{a_2^2 + \lambda} = 1 \tag{3.1}$$

where λ is a parameter.

Fix foci F_1 and F_2. For a generic point X in the plane there exist a unique ellipse and a unique hyperbola with foci F_1 and F_2 that passes through X. The ellipse and the hyperbola are orthogonal to each other: this follows from the fact that the sum of two unit vectors is perpendicular to their difference.

Let us now discuss geometry of billiard caustics.
Let Γ be a strictly convex closed billiard curve. Figure 3.3 is the configuration plane where the billiard table lives. But there is another plane which is called phase plane (Figure 3.7) of the billiard ball map T which consists of oriented lines that intersect Γ. We will use the same coordinates (t, α) where $t \in [0, L]$ as it is a circular coordinate where L being the perimeter length of the billiard curve and $\alpha \in [0, \pi]$

Now we want to see what does the caustics in figure 3.3 corresponds in the phase plane. We define an invariant curve where an invariant circle of the billiard ball map is a simple closed invariant curve δ that makes one turn around the phase cylinder. We can think δ as a smooth one-parameter family of oriented lines which intersects the billiard table if it is a smooth curve.

Since we are dealing with billiards, this is an important result due to Birkhoff's theorem.

Theorem 6 (**Birkhoff's Theorem**). *If a billiard ball map has an invariant curve δ, then this curve is the graph $\alpha = f(t)$ of a continuous function f*

This theorem holds in a more general setup, so what we need here is the area preserving twist maps of a cylinder. The twist condition for a map $T : (t, \alpha) \mapsto (t_1, \alpha_1)$ means that $\partial t_1/\partial\alpha > 0$. This condition clearly holds for the billiard ball map in a convex billiard.

Geometrically this implies that a caustic, if it exists lies strictly inside our table then we have a lemma.

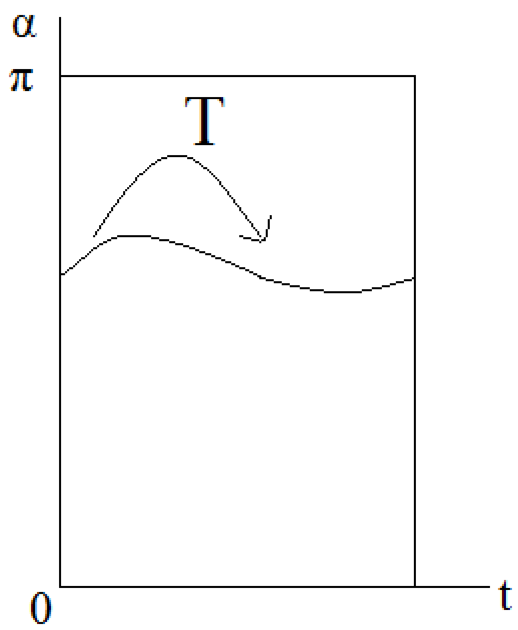


Figure 3.7: Phase portrait

Lemma 3.1. *Let γ be the caustic corresponding to an invariant circle δ of the billiard ball map inside a convex curve Γ. Then by Birkhoff's theorem γ lies inside Γ.*

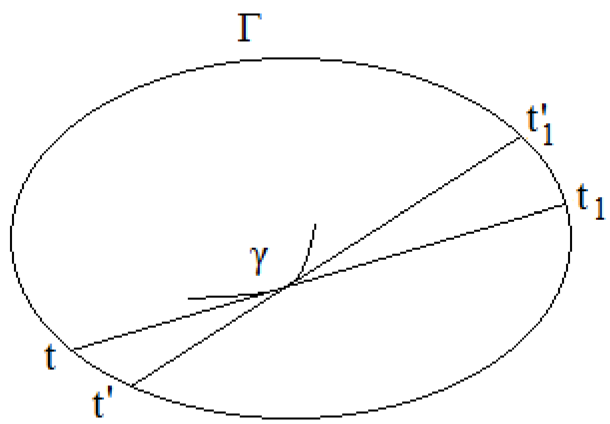


Figure 3.8: Caustics lie inside the billiard table

More general for circles and ellipses, caustics do exist. According to KAM (Kolmogorov-Arnold-Moser) theory, what we have known is caustics always exist. It is not one particular theory rather it is collection of methods. So the statement is for every billiard which is strictly convex and smooth in arbitrary small neighbourhood of the boundary, there exists the caustic. But they don't make foliation, there are gaps between them. If we measure transversely, the relative measure goes to 1 across the boundary then there are many caustics which exist near the boundary. [Tab93]

Now if the billiard table is smooth and convex but not strictly convex where strictly convex means the curvature is positive everywhere. Suppose that there is one point where the curvature is zero then there are no caustics at all. Let us see the converse part.

Let Γ be a billiard curve and γ a caustic. Suppose if we erase the billiard curve, and only the caustic remains. Then the question is can we recover the curve Γ from γ which is a elementary problem and the solution is given by the **String Construction** which produces a one-parameter family of billiard curves.

Lemma 3.2. *Wrap a closed inelastic string around γ, pull it tight at a point and move the point around γ to construct the billiard curve Γ.*

Proof Pick a reference point y on curve γ. For a point $x \in \Gamma$, let $f(x)$ and $g(x)$ be the distances from x to y by going around γ on the right and left tangentially, respectively. Then Γ is a level curve of the function $f + g$.

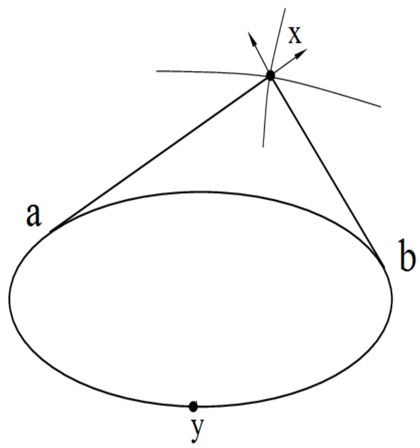


Figure 3.9: String Construction

Now let us see the gradient of these two functions. The point x is not restricted so we can move it in any direction. So there are two gradient unit vectors going in the tangential directions. Its gradient at x is the sum of two unit vectors in the directions of the tangent lines to γ. Hence the gradient makes equal angles with these tangent lines, and therefore, so does Γ. $\square$

Thus we saw the converse to construct curve from the caustic by this **String Construction** method. Note that the string construction provides a one-parameter family of billiard curves Γ : the parameter is the length of the string.

Now we consider phase portrait of billiards inside the circle (Figure 3.10). As we mentioned earlier, for circles one has a one-parameter family of caustics which are concentric circles which is obvious from symmetry and each caustic corresponds to the invariant curve and the whole phase space is foliated by the curves which are parallel in the coordinate system. That's why we labelled it as (t, α), each invariant curve is just a constant angle.

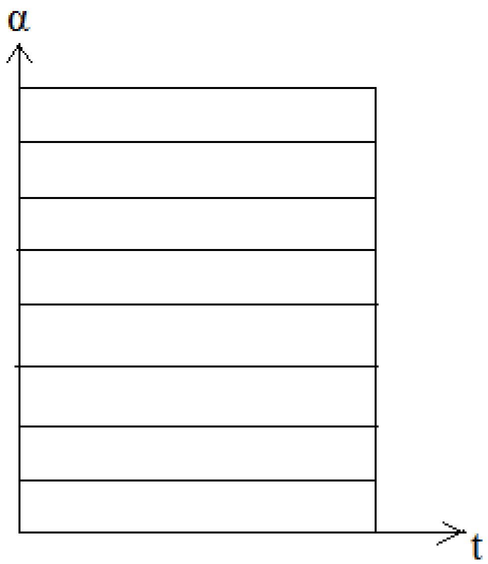


Figure 3.10: Phase portrait of the billiards in circle

Now for an ellipse we have the canonical equation as :

$$\frac{x_1^2}{a_1^2} + \frac{x_2^2}{a_2^2} = 1.$$

There is a one parameter family of caustics which consists of confocal ellipses. These ellipses are given by a string construction which is fixed at the two points (foci) and pulled tightly shown in the figure 3.11. There is a segment which connects the two foci. Now there is another family of caustics which consists of confocal hyperbolas. Hyperbolas are given by a similar geometrical condition that if the locus of points in which sum of distances of two foci is fixed then the confocal hyperbolas is a locus of points such that its difference of length to foci is fixed and suppose the segment of billiard trajectories is tangent to one of these confocal hyperbolas then the reflected one is tangent to same confocal hyperbola.

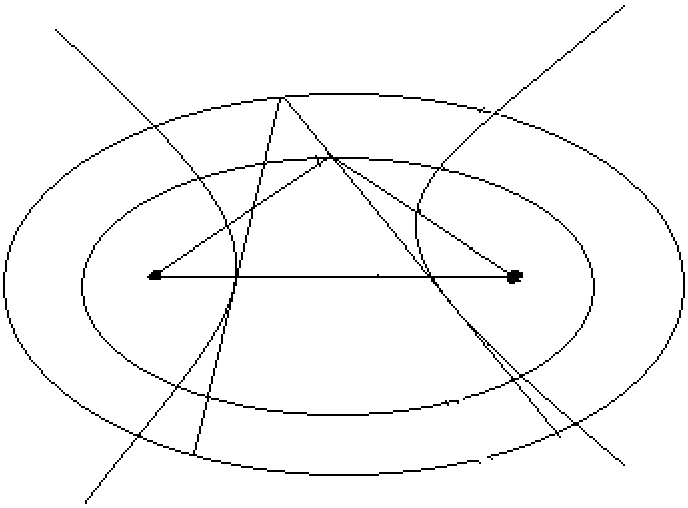


Figure 3.11: Confocal Hyperbola and ellipse.

So in particular, if a ray goes between two foci then the reflected ray will also cross the segment and it will continue the same way. On the other hand if the ray does not cross the segment then all the reflected rays will not cross it ever.

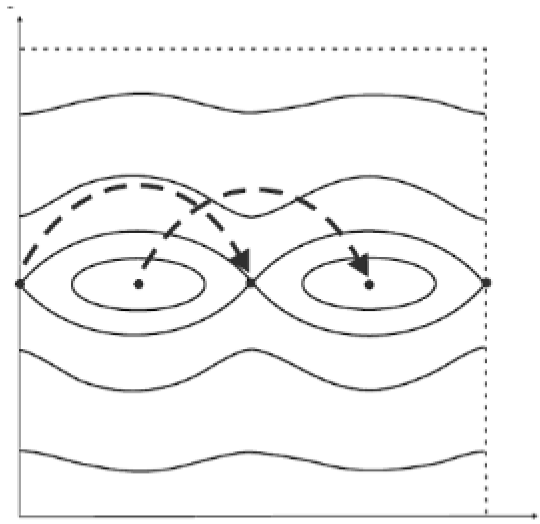


Figure 3.12: Phase portrait of the billiards in Ellipse.

Figure 3.3 is the configuration plane while in the phase plane (Figure 3.12), caustic is one parameter family invariant curve of the billiard ball map T which corresponds to caustics of confocal ellipses. Points in phase plane are oriented lines in the figure.

So the curve in phase portrait is one parameter line in the figure. This singular curve (eye shaped) corresponds to a particular family of rays which pass through foci. If we shoot ray of light from one focus of ellipse, it reflects to another focus and then the process goes on. It is totally different from the Poncelet theorem which we will discuss later.

Figure 3.10 and Figure 3.12 are definitely different from each other. In the theory of billiards there is a famous problem which is called as Birkhoff's conjecture. It concerns billiards which has the phase portraits in the figures. Birkhoff's conjecture states that the curve is an ellipse if the neighborhood of strictly convex smooth billiard curve is foliated with the invariant curves (caustics). So far, this conjecture remains open. There were many false proofs given. The best result was given by M. Bialy who proved that figure 3.10 happens only when the table is circle. So if the whole space is foliated by invariant circles then it must be a circle.

Now working towards punch line, looking to the figure 3.12, we have a phase cylinder which has an area form having a billiard ball map which is area preserving which also has invariant curves. So every point of the curve is sent to another point on the curve by billiard ball map. There is a way to introduce the measure of coordinate on each invariant curve such that that the map will become parallel translation $x \mapsto x + c$. We define an important construction.

Let us choose a function $f = c$ whose level curves are the invariant curves. We introduce some kind of length on each invariant curve. Consider a nearby curve γ_ϵ given by $f = c + \epsilon$. By taking a strip between the two invariant curves i.e. interval of the curve I and considering the area $w(I, \epsilon)$ between γ and γ_ϵ over the interval I. "Length of the curve I" is defined as

$$\lim_{\epsilon \to 0} \frac{w(I, \epsilon)}{\epsilon} \tag{3.2}$$

This is the value which we assign to this interval of invariant curve and it is almost well defined. The function f is not well-defined while level curves are well defined. The length of an element is dx by choosing a coordinate x and this coordinate is well defined. We define an affine structure which is that invariant curve has a canonical coordinate system which is defined up to an affine reparametrization $x \mapsto ax + b$. If the invariant curves are closed, then T is a parallel translation in the respective affine coordinate.

Corollary 3.1. *Let T be an area preserving integrable map and assume that invariant curves are closed. If an invariant curve γ contains a k-periodic point, then every point of γ is k-periodic.*

Proof In an affine coordinate we have $T(x) = x + c$. If $T^k(x) = x$, then $kc \in \mathbb{Z}$, and therefore $T^k = id$. $\square$

Next we will discuss the Integral ball map inside the ellipse. This means that there is a smooth function on the phase space, called an integral, which is invariant under T. We have the equation of an ellipse so let us consider symmetric 2×2 matrix with entries

$$B = \begin{pmatrix} \frac{1}{a_1^2} & 0 \\ 0 & \frac{1}{a_2^2} \end{pmatrix}$$

Then in terms of this matrix the ellipse is given by $Bx.x = 1$. Let (x, v) be the foot point of the boundary while the vector direction is inward (Refer figure 3.13) and (x', v').

Claim : $Bx.v = Bx'.v'$.

Proof The proof consists of two parts. Billiard ball map has many similar maps. It is composition of two involutions. In the figure below one involution interchanges points without changing the vector i.e. $(x, v) \mapsto (x', v)$ and other one fix the point and change the vectors i.e $(x', v) \mapsto (x', v')$.

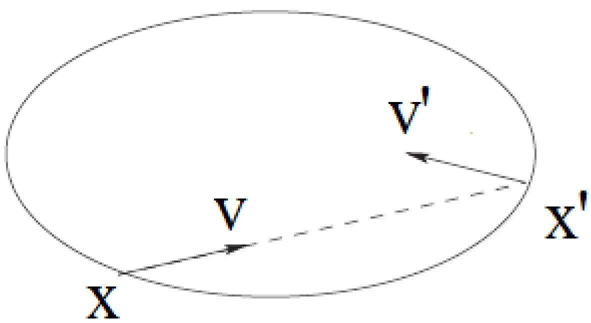


Figure 3.13: Billiard ball map.

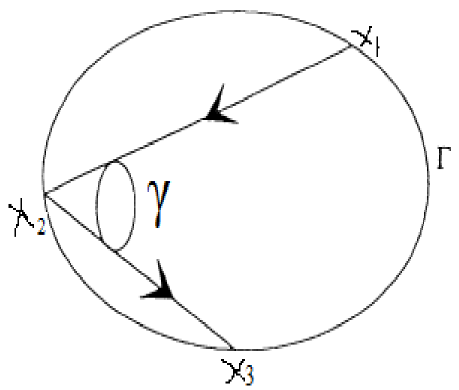
(i) For the first involution we write the identity $B(x' + x).(x' - x) = 0$, which follows from the fact that x and x' belong to the ellipse and B is symmetric. $(x' - x)$ is collinear with v which means that $Bx.v = -Bx'.v$

(ii) For the second involution to use the fact that the sum of two vectors v and v' both being unit is tangent to the ellipse at point x'. The condition for a vector to be a tangent to a conic given by the equation $B\ x.x = 1$ at a point is $B\ x.v = 0$. Thus here we have $Bx'.(v + v') = 0$ which implies that $Bx'.v = -Bx'.v'$.
It follows that $Bx.v = Bx'.v'$ which is an integral of the billiard ball map. $\square$

3.5 The Poncelet porism

The integrability of the billiard ball map in an ellipse which we discussed has an interesting consequence.

If we have two conics. We can take any pair of conics. They need not be nested ellipses. Let us consider two confocal ellipses γ and Γ. Choose any point $x_1 \in \Gamma$. We can see there can be two tangent lines through x_1. So we choose one of those two tangent lines to γ through x_1. There exists a polygon inscribed in Γ and circumscribed about γ. Here polygon means cyclically ordered collection of lines $(l_1, l_2, ..., l_n)$ and points $(x_1, x_2, ..., x_n)$ respectively. Configuration is that all the points $x_i \in \Gamma$ and L_i are tangent lines to γ.



Now L_1 will meet the curve γ at a second point and we have no choice, that has to be next vertex of the polygon. So we have to take x_2 as other point of $L_1 \cap \Gamma$. Assume that this is n-periodic trajectory, i.e. it closes up after n steps. Now repeat this construction choosing another starting point . It follows from Corollary 3.1 that the respective billiard trajectory closes up after n steps as well. Indeed, the family of lines tangent to γ is an invariant curve of the billiard ball map in Γ. In fact, the assumption that Γ and γ are confocal is not necessary at all for the conclusion of the closure theorem to hold. We have the following Poncelet theorem.

Theorem 7 (The Poncelet Theorem). *Suppose we have two plane conics γ and Γ and there is a polygon inscribed in Γ and circumscribed about γ. Then there are infinitely many such polygons, and all of them have the same number of sides. [DR14]*

However the above theorem is a special case of Poncelet porism, it implies the general version. This is because a generic pair of nested ellipses is projectively equivalent to a pair of confocal ones. We start with a Poncelet polygon and it better be a big one and also not too small and what we see is bunch of lines and bunch of points and all this system of lines and points is Poncelet Grid. [Sch07]

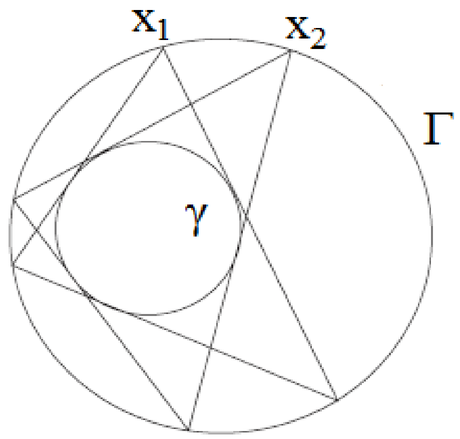


Figure 3.14: Poncelet closure theorem.

Let us draw a heptagon (7-gon) which contains ellipse inside so that all the lines are tangent to it and there is another ellipse so that it becomes a Poncelet polygon which is inscribed into one conic and circumscribed about the other. We want to view is the collection of lines by labelling the lines in general 1 to n. The figure below is referred to as Poncelet grid and there are two ways to organize the collection of points, we can view it as concentric finite sets.

Now we see the definition, we have the lines l_i and consider cyclic sets which we denote it by P_k where k are the levels.The Poncelet grid consists of $\frac{n(n+1)}{2}$ points $l_i \cap l_j$.

Define the sets :

$$P_k = \bigcup_{i-j=k} l_i \cap l_j \tag{3.3}$$

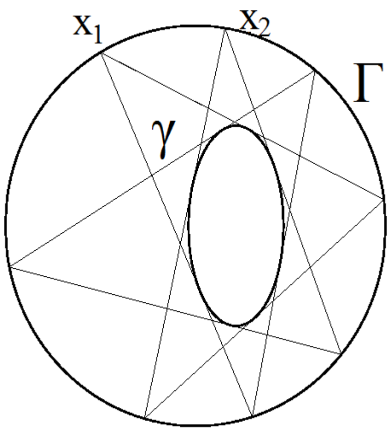


Figure 3.15: 7-gon inside the ellipse

$$Q_k = \bigcup_{i+j=k} l_i \cap l_j \tag{3.4}$$

The cases of odd and even n differ somewhat,we assume that n is odd. There are $\frac{(n+1)}{2}$ sets P_k, each containing n points, and n sets Q_k, each containing $\frac{(n+1)}{2}$ points. [Tab05]

Lemma 3.3. *A confocal family of conics consists of all conics tangent to four fixed lines.*

Proof A curve, projectively dual to a conic, is a conic. The 1-parameter family of conics, dual to the confocal family (3.1), is given by the equation

$$(a_1^2 + \lambda)x_1^2 + (a_1 + \lambda)x_2^2 = 1.$$

This is an equation of a pencil, a 1-parameter family of conics that pass through four fixed points; these are the intersections of the two conics, $a_1^2x_1^2 + a_2^2x_2^2 = 1$ and $x_1^2 + x_2^2 = 1$. Projective duality interchanges points and tangent lines; applied again, it yields a 1-parameter family of conics sharing four tangent lines. □

Bibliography

[BEG99] David A. Brannan, Matthew F. Esplen, and Jeremy J. Gray, *Geometry*, Cambridge University Press, Cambridge, 1999. MR 1694139

[DR14] Vladimir Dragović and Milena Radnović, *Bicentennial of the great Poncelet theorem (1813–2013): current advances*, Bull. Amer. Math. Soc. (N.S.) **51** (2014), no. 3, 373–445. MR 3196793

[Sch07] Richard Evan Schwartz, *The Poncelet grid*, Adv. Geom. **7** (2007), no. 2, 157–175. MR 2314815

[Tab93] Serge Tabachnikov, *Poncelet's theorem and dual billiards*, Enseign. Math. (2) **39** (1993), no. 3-4, 189–194. MR 1252062

[Tab95] ______, *Billiards*, Panor. Synth. (1995), no. 1, vi+142. MR 1328336

[Tab05] ______, *Geometry and billiards*, Student Mathematical Library, vol. 30, American Mathematical Society, Providence, RI; Mathematics Advanced Study Semesters, University Park, PA, 2005. MR 2168892