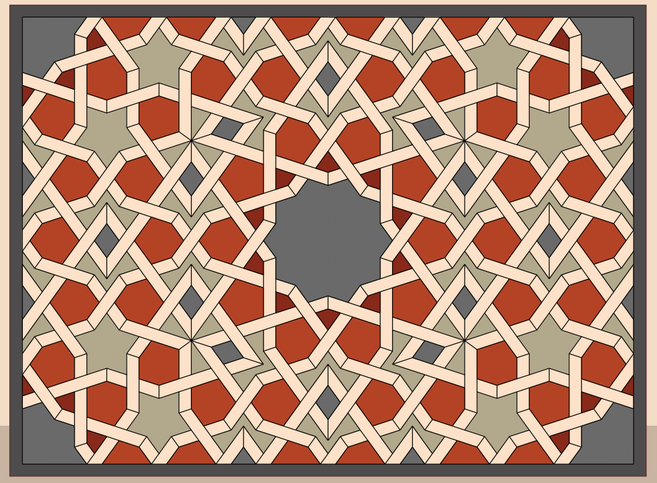


Research Journal of Mathematics and Technology, Volume 14, Number 1, 2025

Research Journal of Mathematics & Technology

ISSN 2163-0380

Vlume 14, Number 1, 2025

Research Journal of Mathematics & Technology

Volume 14, Number 1

ISSN 2163-0380

June 2025

RJMT Editorial Board

Web site: http://rjmt.mathandtech.org/
Published by:

Mathematics and Technology, LLC
PO Box 215
Welcome, NC27374-0215, USA

Foreword

We hope that you, as friends and colleagues, along with your families and your loved ones, are all in good health and good spirits.

The Research Journal of Mathematics and Technology (RJMT) is a printed forum for the publication of selected papers from the Electronic Journal of Mathematics and Technology (eJMT: http://ejmt.mathandtech.org/). One of eJMT's goals is to publish peer-reviewed papers demonstrating how "technology" can be utilized to make mathematics and its applications fun (F), accessible (A), challenging (C), and theoretical (T).

The first four papers of this issue are from the February 2025 eJMT issue. The second four papers are from the October 2024 eJMT issue. As you read these eight papers, you will see discussions of different areas of mathematics using various technologies. There are several papers exploring the innovative use of technological tools in discovering mathematics. We hope these creative ways of experimenting with mathematics will be beneficial to you and your students. As you do so, we encourage you to write about your experiences for future issues of eJMT. Instructions for preparing and submitting papers to eJMT can be found online at https://ejmt.mathandtech.org/SubmissionGuidelines.html.

As this edition of RJMT is being prepared, we are planning the 30th ATCM (Asian Technology Conference in Mathematics: https://atcm.mathandtech.org/) to be held in hybrid format from December 13-16, 2025, which is to be held at ATENEO DE MANILA UNIVERSITY, Quezon City, Philippines. Thank you all for your continued support of eJMT, RJMT, and ATCM.

Mirosław Majewski
Guest Editor

Wei-Ch Yang
Editor-in-chief

TABLE OF CONTENTS

Topology of Quartic Loci in 2D and 3D Inspired by A College Entrance Exam

Wei-Chi Yang[1] Zoltán Kovács[2] T. Dana-Picard[3]

[1] Radford University, VA, USA
wyang@radford.edu

[2] The Private University College of Education of the Diocese of Linz, Austria
zoltan@geogebra.org

[3] Jerusalem College of Technology and Jerusalem Michlala College, Israel
ndp@jct.ac.il

Abstract

We explore with technological tools on some problems that originated from a college practice entrance question (see [12]). The problems involve an affine transformation whose image is a locus based on lines passing through a fixed point (see Figure 1(a) or 1(b)), which has been explored in ([17], [18], [19], [20]) by using parametric equations. We are interested in the topological structures for the shapes of image curves or surfaces when the scaling factor s vary. We found the value for s which affects the topological structures for the loci in 2D first, and we extend the results to 3D based on the arguments in 2D. Finally, we include an exploration on the topological structure for the locus ellipsoid when scaling factor s and the fixed point A approach to infinity.

1 A College Entrance Exam Practice Problem from China

We explore a problem that originated from a college practice entrance problem (see [12]).

Explorations: *We are given a fixed circle and a fixed point A in the interior of the circle $(x-a)^2+(y-b)^2 = r^2$. A line passes through A and intersects the circle at C and D respectively, and the point E is the midpoint of CD. Find the locus of E when C moves along the circle.*

This problem has been discussed using parametric equation approaches in 2D and 3D, see ([17], [18], [19], [20]). In this paper, we will use the implicit equation of the locus and assume the fixed $A = (a_1, a_2)$ to be outside the circle or the ellipse. We are given a fixed point $A = (a_1, a_2)$ and the lines passing through this fixed point to intersect a conic $c : x^2 + qy^2 = 1$ at C and D $(q > 0)$. The point E lies on CD and satisfies

$$\overrightarrow{ED} = s \cdot \overrightarrow{CD}, \tag{1}$$

where s is a given real number. For convenience, we sometimes write this as

$$E = sC + (1 - s)\, D. \tag{2}$$

Since D is one of the points of intersection between the line passing through A and the conic c, sometimes we call D as the antipodal point of the point C. We remind the readers that a generalized plane quartic curve is a bivariate quartic equation (see [16]). Our aim is to study the topology of the obtained parametrized quartic curve $\mathcal{C}(a_1, a_2, q, s; x, y)$, and determine the value(s) of s which will change the topological structures for the quartic curve $\mathcal{C}(a_1, a_2, q, s; x, y)$, and we extend the results to 3D accordingly.

2 Preliminary

A topological space is a non-empty set together with a topological structure that is given by the open sets of the topology. A non-empty set X can be equipped with different topological structures, and any subset $A \subset X$ can be endowed with an inherited topological structure, namely, the *relative topology*: the open sets in the relative topology on A are exactly the subsets of the form $A \cap O$ for some open sets O of X.

For the spaces $\mathbb{R}^n$, with coordinates $(x_1, \ldots, x_n)$, the usual topological structure is given by a base of open balls, which is defined by the metric space structure of $\mathbb{R}^n$. Recall that the *metric function* $d(x, y) = \sqrt{(x_1 - y_1)^2 + \cdots (x_n - y_n)^2}$, for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ makes $(\mathbb{R}^n, d)$ a *metric space*. This metric gives rise to open balls $B_\varepsilon(x) = \{y \in \mathbb{R}^n \mid d(x, y) < \varepsilon\}$, where ε is a positive real number, which are the basic open sets for the usual topology on $\mathbb{R}^n$.

Thus, any curve or any hyper-surface in $\mathbb{R}^n$ can be viewed as topological spaces with their respective relative topologies, acquired from the space in which they are embedded. For the cases we are interested in, the environment spaces will be simply $\mathbb{R}^2$ or $\mathbb{R}^3$; thus, we can view these curves and surfaces as relative topological spaces on which we can apply topological tools that are introduced in what follows.

Now, let us recall a fundamental concept in topology, namely that of *continuity*. Let (X, τ_X) and (Y, τ_Y) be two topological spaces, with τ_X and τ_Y their respective topologies. A function $f : X \longrightarrow Y$ is *continuous* if for any open set $O \subset Y$ it holds that $f^{-1}(O) \subset X$ is open in X. Here, the notation $f^{-1}(U)$ denotes the inverse image, in X, of $U \subset Y$.

When f is bijection, then it has an inverse $f^{-1} : Y \longrightarrow X$ defined by: for any $y \in Y$, $f^{-1}(y) = x \in X$ if and only if $f(x) = y$. We say that f is a *homeomorphism* if it is a continuous bijection and with f^{-1} also being continuous. In this situation, X and Y are said to be *homeomorphic* topological spaces.

Homeomorphisms between topological spaces are also called *continuous transformations*, or *topological transformations*. Intuitively, these transformations can be thought of as functions that transform points of one space, that are arbitrarily close to each other, onto points of another space that are also arbitrarily close. Spaces that are related in this way are said to be *topologically equivalent*. In a technical sense, we have to be careful with the notion of closeness since it only makes sense when we can measure distances in a space, so we need to have a metric structure. Nevertheless, that is what we meant by "intuitively". It is intuitively evident that all simple closed curves in the plane and all closed polygons are topologically equivalent to a circle. Similarly, all closed cylinders, closed cones, convex polyhedra, and other simple closed surfaces are topologically equivalent to a sphere.

For the problem described in the"**Exploration**", now we are interested in a fixed point A outside of a given conic. For example, the loci in red are shown in Figures 1(a) and (b), when the fixed point A is outside a circle, $x^2 + y^2 = r^2$. More specifically, we encourage the readers

to use their favorite DGS (Dynamic Geometry System) tools to explore the followings:

1. When $A = (1.598024, 0.6045678)$, $s = 0.25$ and $r = 2$, we provide such case in the following screenshot using [2] in Figure 1(a).

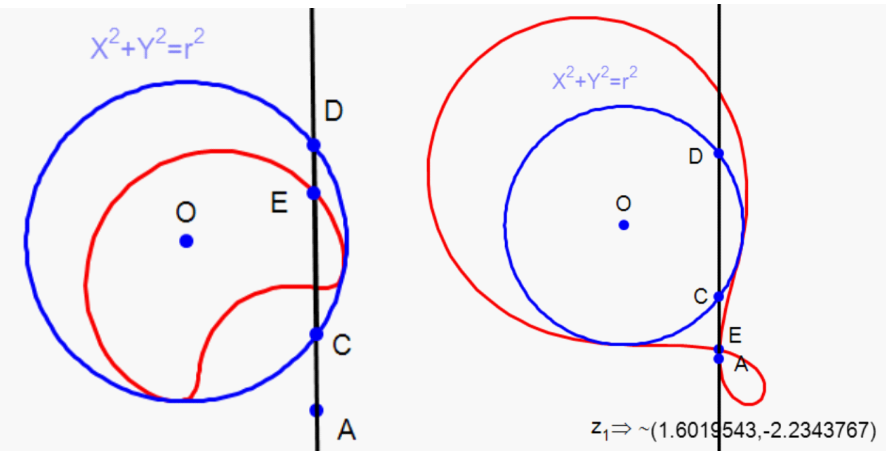


Figure 1(a) (left). The fixed point A is outside the circle and $s = 0.25$
Figure 1(b) (right). The fixed point A is outside the circle and $s = 1.43$

2. If $A = (1.6019543, -2.2343767)$ is outside the circle, and $s = 1.43$, the locus is shown in Figure 1(b).

Among other techniques, we used the automated GeoGebra Discovery (see [3]) to obtain the accurate locus by computing the locus equation for the investigated geometric setups automatically. The following table explains how a possible construction protocol can be given by the user to get a quartic polynomial equation for the case $s = -1.5$, $A = (2, 1)$, if the unit circle is used. Recent versions of GeoGebra Discovery allow the use of the Dilate tool in a symbolic environment, therefore it is expected to get the output locus equation very quickly. As of today, version 2024Jun29 can compute and plot the corresponding quartic curves at about 9.9 frames per second.

No.	Name	Icon	Description	Definition	Value
1	Number s				$s = -1.5$
2	Point A				$A = (2, 1)$
3	Point O		Intersection of xAxis and yAxis	$Intersect\,(xAxis, yAxis)$	$O = (0, 0)$
4	Point U		Point on xAxis	$Point\,(xAxis)$	$U = (1, 0)$
5	Circle c		Circle through U with center O	$Circle\,(O, U)$	$c : x^2 + y^2 = 1$
6	Point D		Point on c	$Point\,(c)$	$D = (0.53, 0.85)$
7	Line l		Line A, D	$Line\,(A, D)$	$l : 0.15x - 1.47y = -1.16$
8	Point C		Intersection of c and l	$Intersect\,(c, \ell, 2)$	$C = (-0.7, 0.71)$
9	Point E		C dilated by factor s from D	$Dilate\,(C, s, D)$	$E = (2.39, 1.04)$
10	Implicit Curve eq1		LocusEquation(E, D)	$LocusEquation\,(E, D)$	$eq1 : x^4 - 4x^3 + 2x^2y^2 - 2x^2y + 4x^2 - 4xy^2 - 60xy + 64x + y^4 - 2y^3 + 49y^2 + 32y = 80$

2.1 The special case of a circle

We consider the circle when $a_1 = 0, a_2$ is any real number, $q = 1$, and by using symmetry, the fixed point $A = (0, a_2)$ on the y-axis can be assumed. For example, we use [4] to demonstrate such a scenario in the following Figure 1(c).

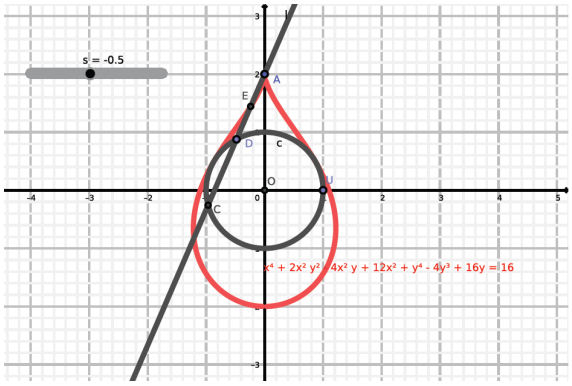


Figure 1(c). Locus in red

Intuitively, we see that the topological structures for the loci, shown in red color, in Figures 1(b) and (c) are different. The Figure 1(b) looks like a figure 8 with one point of self-intersection, and Figure 1(c) is not and has a cusp. Therefore, we say that the Figure 1(b) is not topologically equivalent to the Figure 1(c). In this exploration, we study the topology of the obtained family of parametrized quartic curves, which is denoted by $\mathcal{C}(a_1, a_2, q, s; x, y)$. We shall prove that the topology is changing at

$$s = \frac{1 \pm \sqrt{a_1^2 + q a_2^2}}{2}.$$

We proceed with the following steps:

1. We look for the algebraic equations for points on a circle, collinearity and dilation.
2. We use elimination, by choosing a sensible polynomial factor of the generator polynomial p of the elimination ideal, so a parametrized curve $\mathcal{C}$ can be obtained.
3. The solutions of the equation system $\{\mathcal{C} = 0, \frac{d\mathcal{C}}{dx} = 0, \frac{d\mathcal{C}}{dy} = 0\}$ can help to find the critical points of all possible curves.

2.2 System of equations for the special case of a circle

We consider the system of equations for $\mathcal{C}(a_1 = 0, a_2, q = 1, s; x, y)$, where a_2 is any real number:

$$
\begin{aligned}
&e_1 := c_1^2 + c_2^2 = 1, \quad e_2 := d_1^2 + d_2^2 = 1,\\
&e_3 := \det \begin{pmatrix} c_1 & c_2 & 1 \\ d_1 & d_2 & 1 \\ 0 & a_2 & 1 \end{pmatrix} = 0,\\
&e_4 := x - d_1 = s \cdot (c_1 - d_1),\\
&e_5 := y - d_2 = s \cdot (c_2 - d_2),\\
&I := \langle e_1, e_2, \ldots, e_5 \rangle \cap \mathbb{Q}[a_2, s, x, y], \quad \langle p \rangle := I, \quad \mathcal{C} \mid p,\\
&S := \{\mathcal{C} = 0, \frac{d\mathcal{C}}{dx} = 0, \frac{d\mathcal{C}}{dy} = 0\}.
\end{aligned}
$$

The following **Giac code** (see [5]) solves the system:

```
e1:=d1^2+d2^2=1
e2:=c1^2+c2^2=1
e3:=det([c1,c2,1],[d1,d2,1],[0,a2,1])
e4:=x-d1=s*(c1-d1)
e5:=y-d2=s*(c2-d2)
I:=eliminate([e1,e2,e3,e4,e5],[c1,c2,d1,d2])
p:=I[0]
c:=(factor(p))[2]
dcx:=diff(c,x)
dcy:=diff(c,y)
S:=solve([c,dcx,dcy],[x,y,s])
```

In view of the intermediate step, when computing the principal ideal $I = \langle p \rangle$, is of the form

$$
p = \left(x^2 + y^2 - 1\right) \begin{pmatrix} 4a_2^2x^2s^2 - 4a_2^2x^2s + a_2^2x^2 - 4a_2^2s^2 + 4a_2^2s + a_2^2y^2 - a_2^2 \\ -2a_2x^2y + 8a_2s^2y - 8a_2sy - 2a_2y^3 + 2a_2y \\ +x^4 - 4x^2s^2 + 4x^2s + 2x^2y^2 - x^2 - 4s^2y^2 + 4sy^2 + y^4 - y^2 \end{pmatrix}, \qquad (3)
$$

we note that the first factor corresponds to a **degenerate case**. Therefore, we only need the second factor in the next intermediate step: `c:=(factor(p))[2]`.

Solution (with Giac), computed via Gröbner bases and other techniques:

$$
\begin{aligned}
(x, y, s) = \{ & \\
& (0, a_2, 0)\,, \\
& (0, a_2, 1)\,, \\
& \left(\pm\sqrt{-y^2 + a_2 y}, y, \frac{1}{2}\right), \\
& \left(0, a_2, \frac{1 \pm a_2}{2}\right), \\
& \left(0, a_2, \frac{1}{2}\right), \\
& (0, a_2, s) \\
& \}.
\end{aligned}
$$

Geometrically, only the solution $x = 0$ makes sense. Assuming a special case $a_2 = 2$, the candidates for s are: $0, 1, \frac{3}{2}, -\frac{1}{2}, \frac{1}{2}$. We shall see when $s = \frac{3}{2}$ or $-\frac{1}{2}$ (which produces the same locus), is the boundary where the topology of the locus changes. We check the algebraic solutions geometrically as follows:

Example 1 *We consider the circle when $a_1 = 0, a_2 = 2, q = 1$, the fixed point $A = (0, a_2)$, and $s = 0, \frac{1}{2}, 1, \frac{3}{2}, -\frac{1}{2}$. The loci are depicted in red in Figure 2(a)-2(e) respectively using [4]. In addition, we include the corresponding GeoGebra file in [S1] for further explorations.*

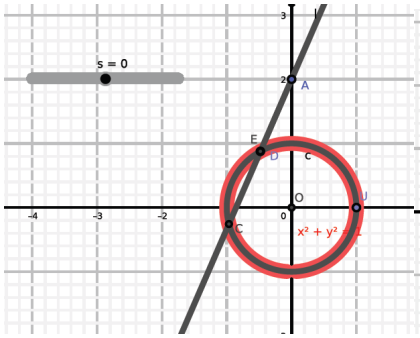


Figure 2(a). Locus when $s = 0$

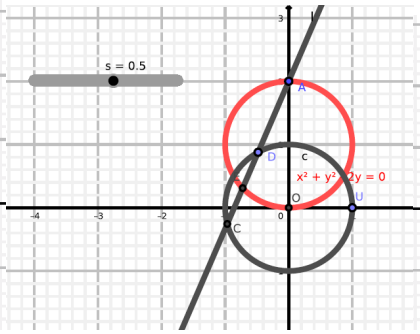


Figure 2(b). Locus when $s = 0.5$

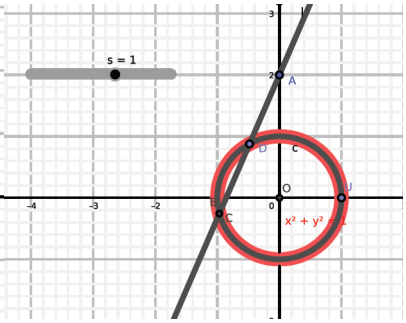


Figure 2(c). Locus when $s = 1$

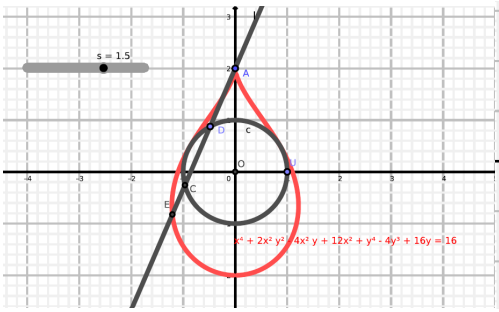


Figure 2(d). Locus when $s = 1.5$

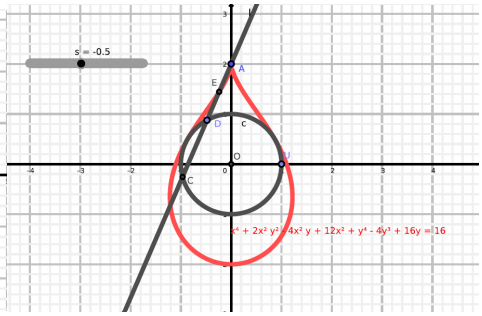


Figure 2(e). Locus when $s = -0.5$

3 Generalization to an ellipse and a special external point A

We consider the general ellipse of $x^2 + qy^2 = 1$, a quartic curve $\mathcal{C}(a_1 = 0, a_2, q > 0, s; x, y)$. We have the following equations:

$$\begin{aligned}
e_1 &:= c_1^2 + qc_2^2 = 1,\\
e_2 &:= d_1^2 + qd_2^2 = 1,\\
e_3 &:= \det\begin{pmatrix} c_1 & c_2 & 1\\ d_1 & d_2 & 1\\ 0 & a_2 & 1\end{pmatrix} = 0,\\
e_4 &:= x - d_1 = s\cdot(c_1 - d_1),\\
e_5 &:= y - d_2 = s\cdot(c_2 - d_2),\\
I &:= \langle e_1, e_2, \ldots, e_5\rangle \cap \mathbb{Q}[a_2, s, x, y],\\
\langle p\rangle &:= I, \quad \mathcal{C} \mid p,\\
S &:= \{\mathcal{C} = 0, \tfrac{d\mathcal{C}}{dx} = 0, \tfrac{d\mathcal{C}}{dy} = 0\}.
\end{aligned}$$

During an intermediate step we yield the following output of a principal ideal $I = \langle p\rangle$, which is an extension of the Equation (**??**).

$$p = \left(qy^2 + x^2 - 1\right)\begin{pmatrix} q^2a_2^2y^2 + 4a_2^2qx^2s^2 - 4a_2^2qx^2s + a_2^2qx^2 - 4a_2^2qs^2 + 4a_2^2qs + a_2^2y^2 - a_2^2q\\ -2a_2qx^2y + 8a_2qs^2y - 8a_2qsy - 2a_2q^2y^3 + 2a_2qy\\ +x^4 - 4x^2s^2 + 4x^2s + 2qx^2y^2 - x^2 - 4qs^2y^2 + 4qsy^2 + qy^4 - qy^2\end{pmatrix}.$$

Similar to the circle case, the first factor corresponds to a degenerate case, where we identify the polynomial defining the ellipse. Therefore, we need only the second factor in the next step: `c:=(factor(p))[2]`.

Solution: Geometrically, only the solution $x = 0$ makes sense. Therefore, candidates for s are:

$$0, 1, \frac{1 \pm a_2\sqrt{q}}{2}, \frac{1}{2}.$$

3.1 Checking $s = \frac{1 \pm a_2\sqrt{q}}{2}$ geometrically

Example 2 *Consider the ellipse of* $c : x^2 + qy^2 = 1$. *If we set the fixed point* $A = (0, a_2)$ *with* $a_2 = 2$ *and* $q = 4$, *then* $s = 2.5$ *or* $s = -1.5$ *(which produces the same locus), is the boundary where the topology of the locus changes. We demonstrate such a scenario in Figure 3 below. We also include the GeoGebra file for readers to explore in [S2].*

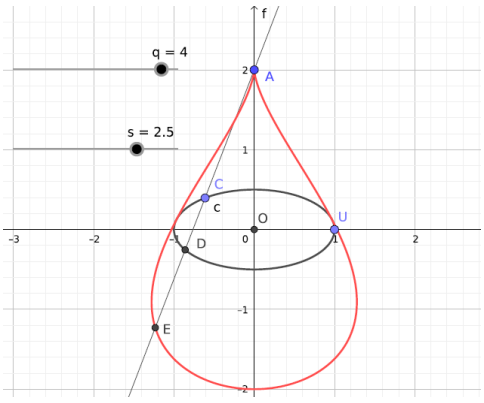


Figure 3. Locus changes its topology when $s = 2.5$ *or* -1.5

3.2 Full generalization

1. Computation of the Gröbner bases is too difficult (it leads to exceeding memory resources of more than 130 GB). We kindly thank RISC (Research Institute of Symbolic Computation, Hagenberg, Austria) for providing us with their "big memory server" to help us attempt computing the result in the general case.

2. We used Gröbner cover instead of traditional Gröbner basis computation, by using the computer algebra system **Singular** (See Appendix section for its input and output), and the package `grobcov`. See [15] for more details.

3. When performing the computation, there are lots of non-geometric and degenerate pieces of information. We searched for polynomials that contain variables a_1, a_2, q, s.

4. We factorized the polynomials and tried to find solutions for s that is a formula of a_1, a_2, q (but not of x or y).

5. With several attempts, we get a simple factorized form:

$$s \cdot (s-1) \cdot (2s-1)^2 \cdot (qa_2^2 + a_1^2 - 4s^2 + 4s - 1)^2 \cdot (y - a_2),$$

and this is equal to 0 if, and only if, s is equal to $0, 1, \frac{1}{2}, \frac{1 \pm \sqrt{a_1^2 + qa_2^2}}{2}$ or $y = a_2$. (We are interested in s, so the last result does not play a role.)

Theorem 3 *(Algebraic Interpretation) If* $a_1^2 + a_2^2 q - 1 \neq 0$ *and* $a_1^2 + a_2^2 q \neq 0$ *and* $3a_1^2 - a_2^2 q \neq 0$ *and* $q \neq 0$ *and* $a_2 \neq 0$ *and* $a_1 \neq 0$*, then the locus curve* $\mathcal{C}(a_1, a_2, q, s; x, y)$ *may change its topology at* $s = \dfrac{1 \pm \sqrt{a_1^2 + q a_2^2}}{2}$.

Theorem 4 *(Geometric Interpretation) If* $A \notin C$ *and* $a_1/a_2 \neq \sqrt{3}/q$ *and* A *is outside of the axes, then the locus curve* $\mathcal{C}(a_1, a_2, q, s; x, y)$ *may change its topology at* $s = \frac{1 \pm \sqrt{a_1^2 + q a_2^2}}{2}$.

4 Is this still just a conjecture?

We actually obtained an algebraic statement. Its geometric counterpart should be verified! How can this be achieved for such a large number of input cases ($a_1, a_2 \in \mathbb{R}$, $q \in \mathbb{R}^+$)?

4.1 A proof based on affine transformations

1. Let us consider the general case $\mathcal{C}^*(a_1^*, a_2^*, q^* > 0, s; x, y)$, and we let $a_1 = a_1^*$, $a_2 = a_2^* \sqrt{q^*}$.
2. Let us stretch the whole construction to get $q = 1$.
   This results in a curve $\mathcal{C}\,(a_1, a_2, 1, s; x, y)$ because the dilation factor remains the same. The stretching ratios are $1 : \sqrt{q^*}$ with respect to $x : y$.
3. Since we have a circle now, it is allowed to rotate the whole construction about the origin to have $a_1' = 0$. So we obtain a second curve $\mathcal{C}'\,(0, a_2', 1, s; x, y)$ which can be studied via elimination without any difficulties.
4. Since we used stretching and rotation, the topology of the curves did not change.
5. By conclusion, we learn that the general case $\mathcal{C}(a_1^*, a_2^*, q^*, s; x, y)$ can be traced back to the case of a circle and a special point on the y-axis.

4.2 3D proof

By following the ideas from 2D:

1. We start with a fixed point $A = (0, a_2, 0)$ on the y-axis, and the ellipsoid whose equation is
$$x^2 + q y^2 + q z^2 = 1, \tag{4}$$
which is obtained by rotating the ellipse $x^2 + q y^2 = 1$ around the y-axis.
2. Using the same argument about the ellipse for $x^2 + q y^2 = 1$, as explained in Section 4.1, we prove that the topology for (4) changes at
$$\frac{1 \pm a_2 \sqrt{q}}{2}. \tag{5}$$
3. For the general case of the topological problem for the ellipsoid of the form
$$x^2 + q_1 y^2 + q_2 z^2 = 1, \tag{6}$$
we may apply an affine transformation, and the problem for Equation (6) can be traced back to the simple case of (4) and a special point on the y-axis.

4.3 Checking $s = \frac{1 \pm a_2\sqrt{q}}{2}$ geometrically for 3D

Example 5 *Consider the ellipsoid*

$$x^2 + qy^2 + qz^2 = 1, \tag{7}$$

with a fixed point $A = (0, a_2, 0)$ *on the y-axis. We set* $a_2 = 2$ *and* $q = 4$ *then it follows from the corresponding 2D formula (5) that* $s = 2.5$ *(or* $s = -1.5$*), is the boundary point where the locus changes its topology. We illustrate such scenarios in Figures 4(aa)-4(d) respectively when* $s = 2.25,\ 2.5, 2.7$ *and* 2.9.

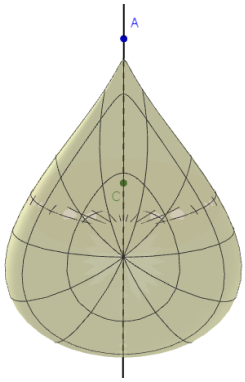


Figure 4(aa). Locus when $s = 2.25$

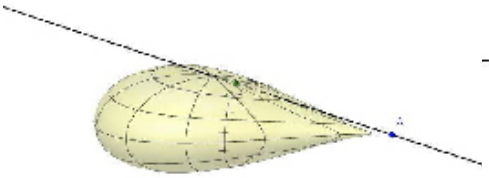


Figure 4(ab). Second view when $s = 2.25$

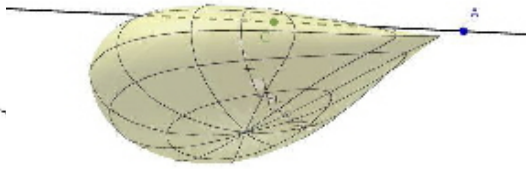


Figure 4(ac). Third view when $s = 2.25$

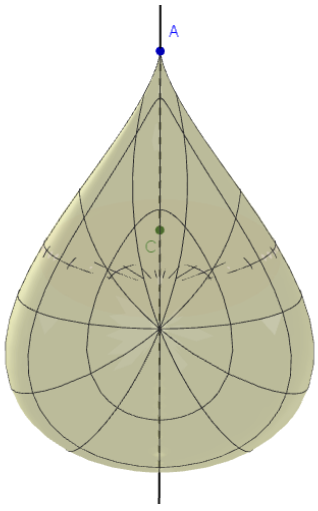


Figure 4(b). Locus when $s = 2.5$

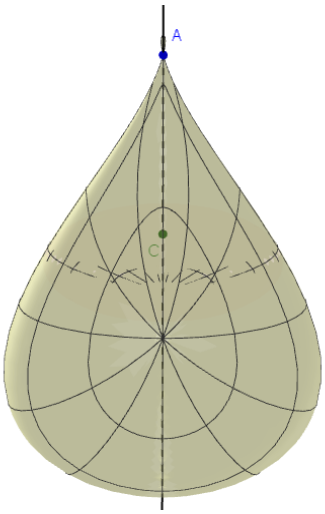


Figure 4(c). Locus when $s = 2.7$

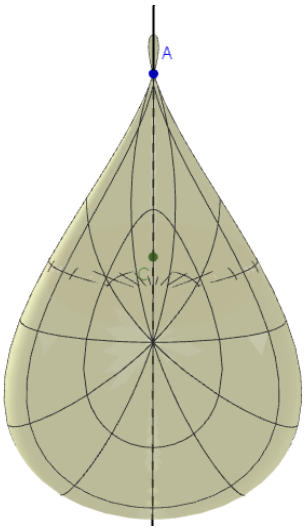


Figure 4(d). Locus when $s = 2.9$

5 Locus Surfaces and Linear Transformations

In earlier sections, we discussed how topological structures change with respect to the parameter s when the fixed point A is given at a finite distance from the origin. Now we summarize

the scenario when the fixed point A is at an infinity, which has been discussed in [19]. For completeness, we summarize it here.

It is known that the image of an ellipsoid under a linear transformation is another ellipsoid. We proved this observation directly in [19] when the fixed point A is at an infinity. We quote two main results from [19] as follows:

Proposition 6 *If Σ is the ellipsoid $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1$, $A_{\inf}(u_0,v_0)$ be the fixed point at an infinity in the direction of $(\cos u_0 \sin v_0, \sin u_0 \sin v_0, \cos v_0)$, $C \in \Sigma$ and D_{inf} be the "antipodal" point of C corresponding to $A_{\inf}(u_0,v_0)$ satisfying $E_{\inf} = sC + (1-s)D_{\inf}$, then there exists a matrix*

$$L_E^e = sI + (1-s)L_D^e \tag{8}$$

such that $L_E^e C = E_{\inf}$, and therefore, the locus surface $\Delta_{\inf}(s,u_0,v_0)$ is the image of Σ under the linear transformation given by the matrix L_E^e.

Proposition 7 *For $s \in \mathbb{R} \setminus \{1\}$, the ellipsoid Σ and locus ellipsoid $\Delta_{\inf}(s,u_0,v_0)$ intersect themselves tangentially at an elliptical curve.*

Since the locus surface $\Delta_{\inf}(s,u_0,v_0)$ is an ellipsoid and since $\Sigma \subset \Delta_{\inf}(s,u_0,v_0)$ and two surfaces are tangent to each other. The locus surface $\Delta_{\inf}(s,u_0,v_0)$ becomes a longer ellipsoid containing Σ as s increases. It is natural to imagine what will happen when $s \to \infty$. As one would expect that the locus ellipsoid $\Delta_{\inf}(s,u_0,v_0)$ in this case would be blown up to a cylinder, which changes its topological structure, and it will not be topologically equivalent to an ellipsoid. We use the following example from [19] for demonstration.

Example 8 *We depict the locus elliptical cylinder (yellow) when $a=5, b=4, c=3$, with $u_0=\frac{\pi}{6}, v_0=\frac{\pi}{3}$ together with the original ellipsoid (blue) in Figure 5(a). The two surfaces are tangent to an elliptical curve as seen in Figure 5(b).*

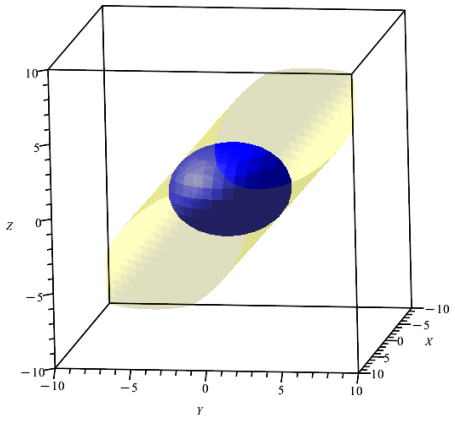


Figure 5(a). The original ellipsoid and its locus when $s \to \infty$.

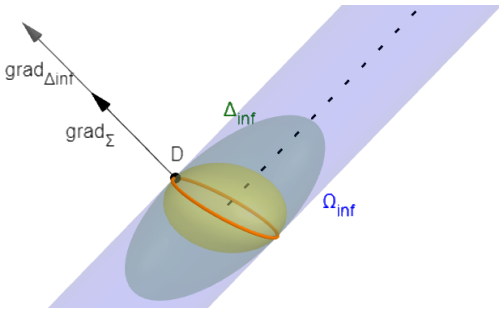


Figure 5(b). Ellipsoid and its locus are tangent at an elliptical curve..

6 Conclusion

The implementation of technological tools into research and teaching has changed the approach to old questions and also enabled the exploration of new situations, which until now seemed out of reach. Dynamic Geometry is the most popular tool among teachers and High School students, but some decades ago, software has been developed for more abstract domains, such as Number Theory, Group Theory, etc.. The traditional progression *definition-theorem-proof* with examples at some places has been replaced by *exploration-discovery-conjecture-proof*. Many mathematical domains became experimental.

Technology provide the researcher and the learner with crucial intuitions, before an attempt to find more rigorous analytical solutions. After experimentation with software, the question may be : "what happens here? what can I prove here?"

In this paper, we have gained geometric intuitions while using a DGS. College entrance exams can lead to challenging problems. Dynamic geometry software with computer algebra extensions may suggest valid conjectures. Most of the time, a DGS leads to discovery and conjecture, and a rigorous proof is obtained using the symbolic algebraic abilities of a CAS. The respective roles of DGS and CAS may be versatile, depending on the topic. Generally the DGS provides numerical support and the CAS symbolic exact work, but not always. GeoGebra Discovery has symbolic complements to numerical commands which pre-existed in GeoGebra. In [9], the CAS had to be used for numerical experimentation, not only the DGS. See also [8]. The working environment has changed, and the technological literacy is crucial. M. Artigue [1] points out that technological knowledge is an integral part of the new mathematical knowledge.

We extend systems with new functionalities (like symbolic `Dilation` for GeoGebra Discovery) that can speed up the investigation substantially. Free computer algebra systems (like Giac) can be comparable to commercial systems.

Evolving technological tools have made mathematics fun and accessible on the one hand, but they also allow the exploration of more challenging and theoretical mathematics. We hope that when mathematics is made more accessible to students, more students can be attracted to the field, and inspired to investigate problems ranging from the simple to the more challenging.

Moreover, the usage of technology may encourage greater collaborations, between humans, and between man and machine. We give here also an example where different kinds of software have to collaborate with humans. Communication between machines is still to be developed. E. Roanes-Lozano [11] expressed often this wish, some progress has been made [13], but is still quite slow. Improving technological tools among researchers in mathematics and mathematics education, and developing all kinds of communication and collaboration, is a crucial task. Following Shulman [14], Loewenberg et al. [10] analyze and explain the importance of changing the teacher's knowledge for teaching. This requires also analyzing the didactic transition [7], from mathematical knowledge of taught knowledge ("du savoir savant" au "savoir enseigné"). We claim that working in the way we propose in this paper will not only spark new areas of research but also increase future teachers' content knowledge at the same time.

7 Supplementary Electronic Materials

[S1] GeoGebra file for Example 1: https://www.geogebra.org/m/fjb9bzz4.

[S2] GeoGebra file for Example 2: https://www.geogebra.org/m/xeykgv7t.

[S3] GeoGebra file for Example 5: https://www.geogebra.org/m/pkv54ee.

References

[1] Artigue, M. Learning Mathematics in a CAS Environment: The Genesis of a Reflection about Instrumentation and the Dialectics between Technical and Conceptual Work. International Journal of Computers for Mathematical Learning 7, 245–274 (2002). https://doi.org/10.1023/A:1022103903080

[2] Geometry Expressions, see http://www.geometryexpressions.com/.

[3] GeoGebra Discovery download: https://github.com/kovzol/geogebra-discovery.

[4] GeoGebra: See https://www.geogebra.org/.

[5] Giac/Xcas, a free computer algebra system: https://www-fourier.ujf-grenoble.fr/~parisse/giac.html.

[6] Maple, A product of Maplesoft, see http://Maplesoft.com/.

[7] Chevallard, Y., Bosch, M. (2020). Didactic Transposition in Mathematics Education. In: Lerman, S. (eds) Encyclopedia of Mathematics Education. Springer, Cham. https://doi.org/10.1007/978-3-030-15789-0_48

[8] Th. Dana-Picard and Z. Kovács (2021): Networking of technologies: a dialog between CAS and DGS, The electronic Journal of Mathematics and Technology (eJMT) 15 (1), 43-59, 2021.

[9] Dana-Picard, T. Networking between kinds of software: geometric loci and exploration of simple models of space curves, Preprint 2024.

[10] Loewenberg Ball, D., Thames, M. H., & Phelps, G. (2008). Content Knowledge for Teaching: What Makes It Special? Journal of Teacher Education, 59(5), 389-407. https://doi.org/10.1177/0022487108324554

[11] Roanes-Lozano, E. and Roanes-Macias, E. (2003). *A Bridge between Dynamic Geometry and Computer Algebra*, Mathematical and Computer Modelling **37**, 1005-1028.

[12] Shi Gi Gin Bang. "Strategies for High School Mathematics Complete Review". In: ed. by Cuiun Guan Zhiming. Century Gold. Yanbian University Press, 2015.

[13] Kovács, Z., Parisse, B., Recio, T., Vélez, M.P, Yu, J.H. (2024). The `ShowProof` command in *GeoGebra Discovery*: Towards the automated ranking of elementary geometry theorems, ACM Commun. Comput. Algebra **57**, 27-30.

[14] Shulman, L. Knowledge and Teaching:Foundations of the New Reform Harvard Educational Review (1987) 57 (1): 1–23. https://doi.org/10.17763/haer.57.1.j463w79r56455411

[15] Montes, A. 'The Gröbner Cover', 2018, Springer

[16] Quartic plane curve, https://en.wikipedia.org/wiki/Quartic_plane_curve.

[17] Yang, W. -C., Exploring Locus Surfaces Involving Pseudo Antipodal Points, at the Proceedings (refereed) of the 25th Asian Technology Conference in Mathematics, 2020, Mathematics and Technology, LLC, Scopus indexed, ISSN 1940-4204 (online version).

[18] Yang, W.-C. & Morante, A. (February 2021). 3D Locus Problems of Lines Passing Through A Fixed Point. The Electronic Journal of Mathematics and Technology, 15(1), ISSN 1933-2823, pp1-22.

[19] Yang, W.-C. & Morante, A. (February 2022). Locus Surfaces and Linear Transformations when a Fixed Point is at an Infinity. The Electronic Journal of Mathematics and Technology, 16(1), ISSN 1933-2823, pp1-24.

[20] Yang, W.-C., Ellipsoid is Tangent to its Locus under a Linear Transformation, Isometries and Sheared Maps, The Proceedings of the 27th Asian Technology Conference in Mathematics, ISSN 1940-4204, https://atcm.mathandtech.org/EP2022/invited/21949.pdf.

8 Appendix

- Singular-input: LIB "grobcov.lib";

ring R=(0,a1,a2,q),(x,y,s),lp;

```
poly c=a2^2*q^2*y^2+4*a2^2*q*x^2*s^2-4*a2^2*q*x^2*s+a2^2*q*x^2
 -4*a2^2*q*s^2+4*a2^2*q*s-a2^2*q-2*a2*q^2*y^3-2*a2*q*x^2*y
 -8*a2*q*x*y*s^2*a1+8*a2*q*x*y*s*a1+8*a2*q*y*s^2-8*a2*q*y*s
 +2*a2*q*y+q^2*y^4+2*q*x^2*y^2-2*q*x*y^2*a1+4*q*y^2*s^2*a1^2
 -4*q*y^2*s^2-4*q*y^2*s*a1^2+4*q*y^2*s+q*y^2*a1^2-q*y^2
 +x^4-2*x^3*a1-4*x^2*s^2+4*x^2*s+x^2*a1^2-x^2+8*x*s^2*a1
 -8*x*s*a1+2*x*a1-4*s^2*a1^2+4*s*a1^2-a1^2;
poly dcx=4*a2^2*q*2*x*s^2-4*a2^2*q*2*x*s+a2^2*q*2*x-2*a2*q*2*x*y
 -8*a2*q*y*s^2*a1+8*a2*q*y*s*a1+2*q*2*x*y^2-2*q*y^2*a1+4*x^3
 -2*3*x^2*a1-4*2*x*s^2+4*2*x*s+2*x*a1^2-2*x+8*s^2*a1-8*s*a1+2*a1;
poly dcy=a2^2*q^2*2*y-2*a2*q^2*3*y^2-2*a2*q*x^2-8*a2*q*x*s^2*a1
 +8*a2*q*x*s*a1+8*a2*q*s^2-8*a2*q*s+2*a2*q+q^2*4*y^3+2*q*x^2*2*y
 -2*q*x*2*y*a1+4*q*2*y*s^2*a1^2-4*q*2*y*s^2-4*q*2*y*s*a1^2
+4*q*2*y*s+q*2*y*a1^2-q*2*y;
ideal I=c,dcx,dcy;
short=0;
grobcov(I); // the input is less than 900 characters
```

- Singular-output file can be found in this link:

https://ejmt.mathandtech.org/Contents/v19n1p1/singular-output.pdf.

A web-based application for the visual exploration of regression outliers, high leverage points, and influential points

Christopher J. Casement
casementc@gmail.com
Department of Mathematics
Fairfield University
Fairfield, Connecticut, 06824
USA

Abstract

Regression outliers, high leverage points, and influential observations are commonly taught in statistics courses in which regression methods are covered – from the introductory to the advanced levels. When covering these topics, it is typical to discuss various methods that exist for determining whether a point is one of these types – such as Cook's distance or DFFITS. Yet while these topics are routinely covered, it can be difficult for instructors to graphically illustrate whether observations are outliers, high leverage points, or influential observations based on the methods and their associated criteria. To address this shortcoming, in this article, a free, web-based app created by the author, which focuses on influential and related observations, is described, with intended usage by both faculty and students. Faculty feedback, which highlights the accessibility, ease of use, and usefulness of the app, is also discussed.

1. Introduction

Regression outliers, high leverage points, and influential observations are topics commonly taught in statistics courses that cover regression. A point is deemed a regression outlier if it falls outside of the overall trend or pattern of the data (particularly in the y-direction), and a high leverage point if it is extreme in the x-direction (typically relative to the mean of the x-values). A point is then considered to be potentially influential if it is both a regression outlier and a high leverage point. Various measures and cutoffs are used to determine if a value is one of the aforementioned types. For instance, the standard deviation of the residuals and studentized residuals are commonly used to assess whether a point is a regression outlier; the leverage statistic (also known as the hat value) is used for high leverage points; and Cook's distance, DFFITS, and DFBETAS can be used to examine whether a point is potentially influential.

While teaching regression outliers, high leverage points, and influential points is common, it is perhaps most beneficial to pair them with a graphical tool when explaining them in a simple linear regression setting. This is due to the fact that it has been found that visualizations can boost the ability of students to understand new concepts [2, 4, 11]. To facilitate these efforts, the author of the article created a free, web-based application that implements the aforementioned methods for detecting potential points of the three types.

The article proceeds as follows. In Section 2, the use of real data in the classroom is discussed, as are recommended characteristics for evaluating statistical tools. Then, in Section 3, a description and example of the app is provided. Next, in Section 4, a survey given to faculty who teach statistics regularly is described and the results discussed. Lastly, the article is summarized in Section 5.

2. Background

Numerous articles have focused on the importance of using real data, technology, and active learning in the classroom. For instance, Garfield and Ben-Zvi [6], Garfield and Everson [7], and Neumann, Hood, and Neumann [10] focus on statistical reasoning in introductory statistics courses, while others, including Rumsey [15] and Singer and Willett [18] have discovered that students participate more when real world situations and data are utilized.

Of course, when real world data is involved, oftentimes it is not as 'clean' as one might hope for. For instance, the violation of assumptions for a particular statistical method (e.g., due to the presence of outliers) can lead to the potential for misleading or even invalid conclusions to be drawn without proper care being taken. As far as regression is concerned, regression outliers, high leverage points, and/or influential points can lead to such issues. To this end, anyone running a regression analysis should fully understand these types of data points and how they can produce such issues.

A plethora of tools exist that enable users to run regression analyses, including point-and-click programs and programming languages. Point-and-click programs include JMP, SPSS, JASP [8], StatCrunch [12], and Rguroo [17], among others. While some of these programs enable the calculation of certain statistics for assessing whether a point might be a regression outlier, high leverage point, or influential point, the programs typically calculate the measure of interest and report its values, such as in the form of a new column in the original dataset, rather than displaying them graphically. Or, if the program does plot the values, it might do so in a separate plot. Some programming languages also enable users to run regression analyses, such as R, Python, MATLAB, and Stata (while not a programming language itself, it has programming languages built in), among others. While the calculation of the measures of interest is not always particularly burdensome using these languages, a specialized knowledge of the languages is required in order to accomplish various things that the proposed app has built in automatically, such as:

- interactively moving points in a scatterplot (which is not even doable using some languages) and viewing the potential impact on regression statistics,
- adding statistic values (e.g., Cook's distance or DFFITS values) to a scatterplot, and
- quickly and easily identifying which observations in a dataset are potential outliers, high leverage points, or influential points based on the statistic of interest.

In fact, none of the aforementioned tools – point-and-click programs or programming languages – (to the author's knowledge) provides a 'ready-to-go' tool for the visual exploration of potential regression outliers, high leverage points, or influential points via the commonly-used statistics discussed in Section 1, all in a single scatterplot. Of course, one could create an interactive tool (like the author did) using certain programming languages such as R, but that would require specialized knowledge of the language, which is not expected of the vast majority of students learning regression or instructors teaching it.

In order to evaluate technological tools, Biehler [1], McNamara [9], and Repenning [14] provide attributes to consider. In particular, McNamara focuses on ten key characteristics for evaluating tools for statistical computing. However, due to the proposed app's focus on the plotting and fitting of data, the six characteristics included in Table 1 – all of which are possessed by the app – are of particular interest. The table also explains how the app meets each characteristic.

Table 1. The proposed application in relation to characteristics recommended by McNamara [9].

Characteristic	How the App Possesses the Characteristic
"Be accessible"	The app is free to use and simply requires a web browser.
"Provide easy entry"	The app has a point-and-click interface and does not require any programming background.
"Privilege data as a first-order object"	The app requires users to input data, with the resulting scatterplot and regression output focused on that data.
"Support exploratory and confirmatory analysis"	The app enables users to explore their data via the scatterplot and associated regression statistics. It also provides users with desired statistics for determining whether a point might be an outlier, high leverage point, or influential point.
"Allow for flexible plot creation"	The app automatically creates a scatterplot for users and allows users to customize the axes.
"Be interactive"	The app utilizes the *plotly* [16] package in R [13], enabling users to manually move points displayed in the scatterplot.

3. Description and example of the application

A description of the *Influential Points* application is now provided. The app was created using the *shiny* package [5] in R [13] and is free to access at https://educationapps.shinyapps.io/InfluentialPoints. The app utilizes various R packages, but most notably it uses *rio* [3] for flexible data importing and *plotly* [16] for interactive plotting. Strengths of the app are now discussed, followed by an example of its usage.

3.1 Strengths of the app

The app possesses various strengths, including the following:

- The app is accessible via the internet. In fact, it can be accessed via any device with internet access (e.g., laptops, tablets, and even smartphones).
- The app is free to use by both faculty and students.
- Users have the option to either manually input data or upload a dataset using files of various common types – e.g., .csv, .xlsx, .sav, and .txt, among others.
- The app displays values for common methods of detecting regression outliers, high leverage points, and influential points, including the standard deviation of the residuals, studentized residuals, leverage values (i.e., hat values), Cook's distance, DFFITS, and DFBETAS.
- Users can manually move individual points on the scatterplot to examine the potential impact of outliers, high leverage points, or influential points on regression output. For instance, they can see how the slope, y-intercept, correlation coefficient, r-squared, standard deviation of

the residuals, and the p-value (when testing the slope) do or do not change as a result of one of these types of points.

3.2 Example using the app

An example of the app's usage is now provided using real-world data collected in an introductory statistics course. The dataset contains data on houses, including the selling price, house size, number of bedrooms, and number of bathrooms, among other variables. The focus of the example is on regressing selling price (measured in U.S. dollars) on house size (measured in square feet).

When the user opens the app, they are immediately taken to the 'Home' tab, which displays instructions for how to use the app, as can be seen in Figure 1.

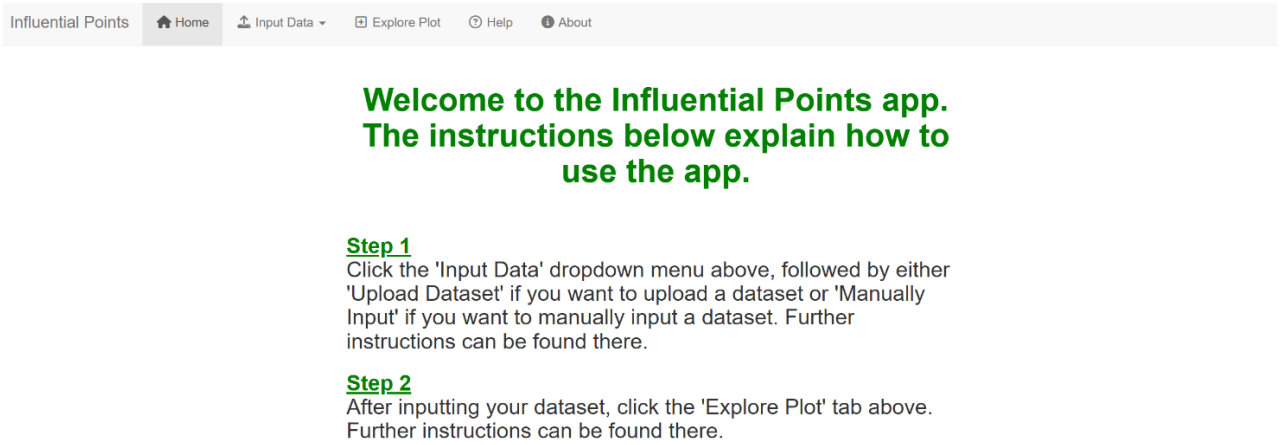


Figure 1. After opening the app, the user starts in the 'Home' tab, which contains instructions for how to navigate the app.

After reading the instructions, the user should click on the 'Input Data' dropdown menu at the top of the app, where they can either upload a dataset or manually input data. In this case, the user can quickly and easily upload their existing dataset. As Figure 2 displays, the user has clicked the "Browse…" button to select the dataset from their computer and then clicked the blue 'Store Dataset' button to store the dataset in the app for further usage.

Once they have uploaded their dataset, the user then navigates to the 'Explore Data' tab found at the top of the app. There, they start by selecting the two variables of interest (X and Y) from the respective dropdown menus, and they then click the blue 'Make Plot' button to make a scatterplot of the selected data. Figure 3 displays the resulting scatterplot when regressing selling price on house size, as well as the regression output, which includes the y-intercept and slope of the regression line, the correlation coefficient, r-squared, the standard deviation of the residuals, and the p-value for testing the slope of the regression line. Additionally, users have the option of customizing the x-axis and y-axis labels, as is done in Figure 3.

When examining the scatterplot, the user should look for potential regression outliers, high leverage points, and influential points, as those types of observations are the main focus of the app.

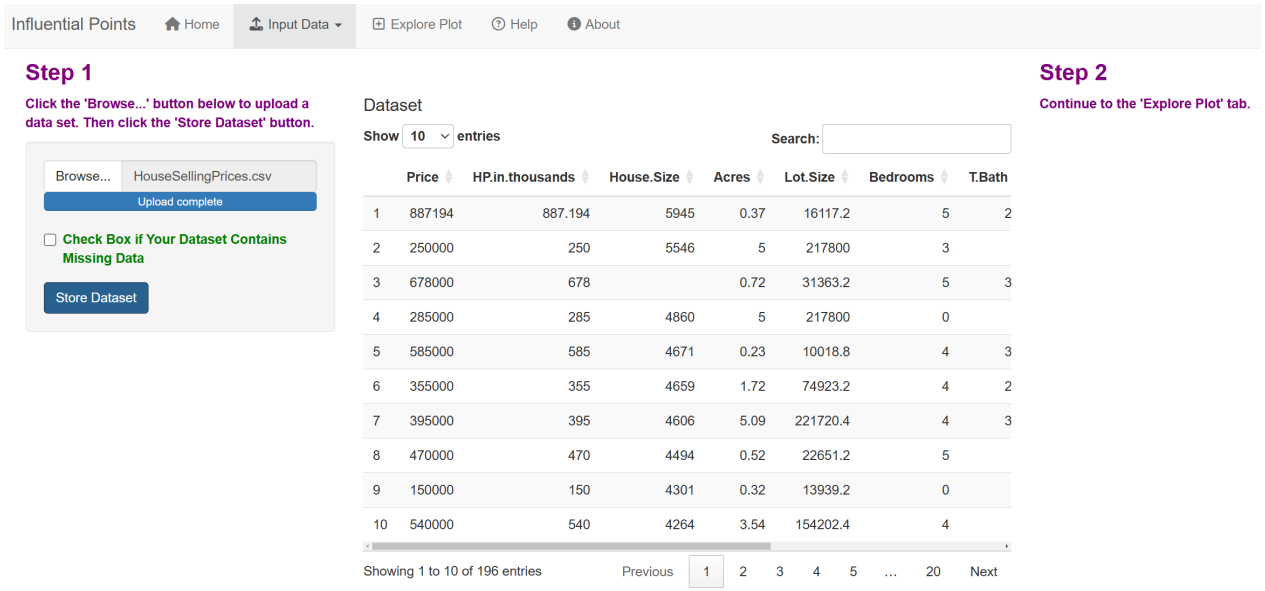


Figure 2. The user can upload a dataset in the 'Input Data' tab. After the user clicks the 'Store Dataset' button, the uploaded dataset is displayed in the app.

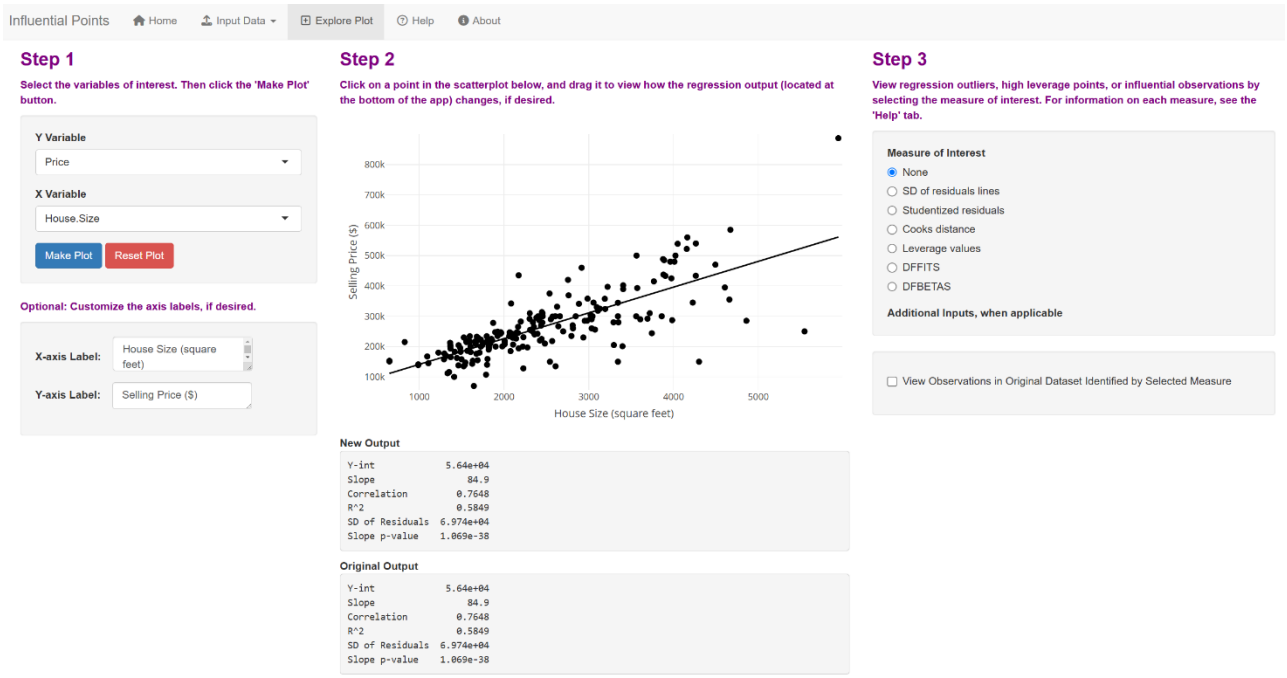


Figure 3. A scatterplot of the user's data can be found in the 'Explore Plot' tab. Below the plot, regression statistics such as the y-intercept and slope of the regression line, correlation coefficient, r-squared, standard deviation of the residuals, and p-value (when testing the slope of the line) are printed. Additionally, the user has the option of customizing the axis labels, as is done here.

However, while a visual assessment can be helpful, determining whether an observation is a regression outlier, high leverage point, or influential point in this way is a subjective approach. For a more objective method, people analyzing data oftentimes turn to statistical measures. For instance, when assessing whether an observation is a regression outlier, two commonly-used statistics are the standard deviation of the residuals and studentized residuals, both of which are built into the app. To visually assess observations using the standard deviation of the residuals, the user should click the button corresponding to 'SD of residuals lines' for the measure of interest found in 'Step 3.' After the user has made this selection, lines are plotted in the scatterplot at values that are two (orange) and three (red) standard deviations (of the residuals) from the regression line. Any observation that falls beyond the preferred set of lines (i.e., above the orange/red lines located above the regression line or below the orange/red lines found below the regression line) is a potential regression outlier. Of course, users likely want to be able to easily determine *which* observations in the dataset are potential outliers (i.e., the row numbers). In fact, the app provides users with this option. The user can simply check the checkbox labeled 'View Observations in Original Dataset Identified by Selected Measure,' which is found in the bottom-right-hand corner of the app. Figure 4 displays the scatterplot of the housing data with the orange and red lines added, as well as the row numbers of the observations that are deemed potential outliers by falling more than two standard deviations from the regression line. While not shown in Figure 4, the app also prints the observations that are more than three standard deviations from the regression line in the bottom-right-hand corner.

Figure 4. When the user clicks the 'SD of residuals lines' button in 'Step 3,' lines are superimposed at two (orange) and three (red) standard deviations (of the residuals) from the regression line. Here, 11 observations are potential regression outliers, as they lie more than two standard deviations from the regression line. Users can view which observations those are in the bottom-right-hand corner.

The other common statistic used for regression outliers is the studentized residual. To work with this statistic in the app, the user simply clicks the button corresponding to ‘Studentized residuals’ for the measure of interest. They then input a value for the desired cutoff, which represents the value such that any observation that takes on a statistic value (e.g., studentized residual) of that number or larger is identified in the scatterplot, along with the value. Figure 5 displays the output after inputting a value of two for the cutoff, with purple line segments drawn from individual points to the regression line (and studentized residual values added to the plot) when points are deemed regression outliers based on the studentized residual cutoff of two specified by the user.

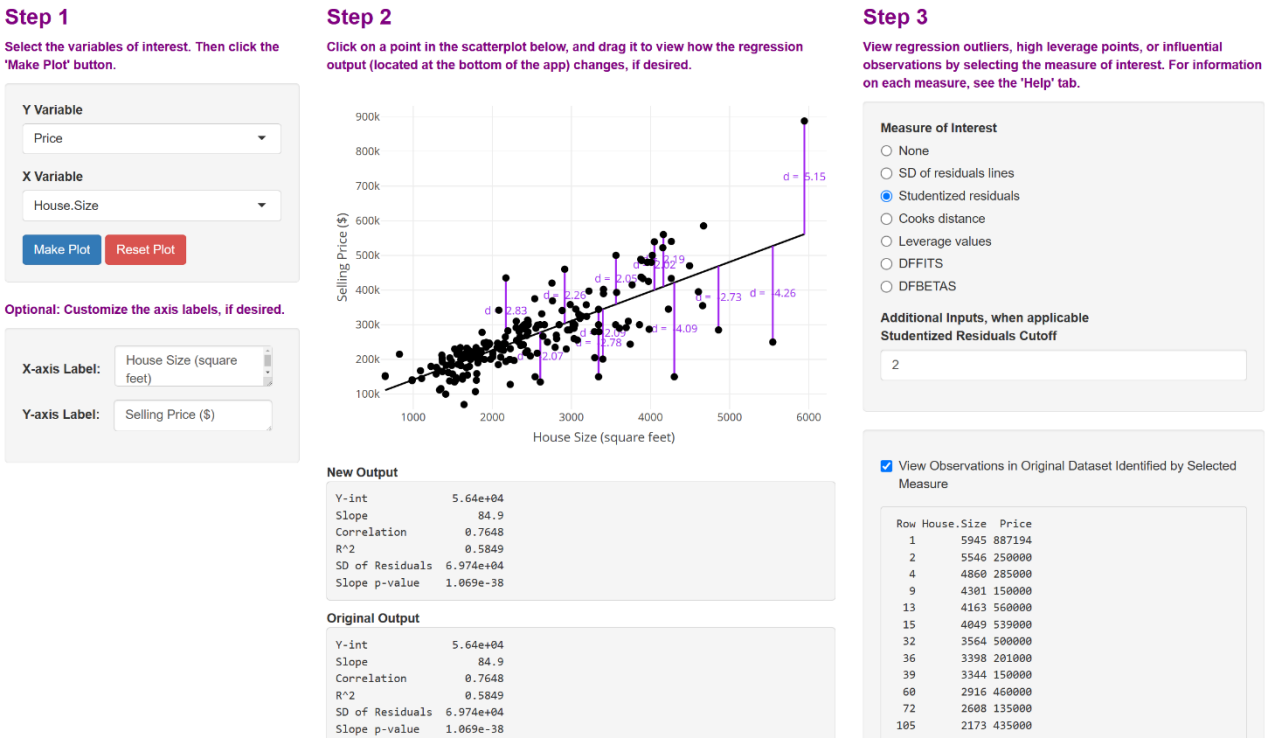


Figure 5. When the user clicks the ‘Studentized residuals’ button and inputs a cutoff (here, two) for detecting regression outliers using this statistic, the observations with studentized residual values equal to or above the cutoff can be identified graphically in the scatterplot, with the respective studentized residual values displayed in purple. The row numbers corresponding to those points are printed in the bottom-right-hand corner of the app. In this case, with the specified cutoff of two, there are 12 potential regression outliers. Of course, this cutoff is quite conservative, and there would be fewer outliers if a cutoff of three were used instead.

When assessing whether an observation is influential, three typical statistics one can use are Cook’s distance, DFFITS, and DFBETAS. To work with Cook’s distance in the app, the user simply needs to click the ‘Cooks distance’ button and input a cutoff value. Figure 6 displays the output for the housing example based on a conservative Cook’s distance cutoff of 0.5 (where a smaller cutoff

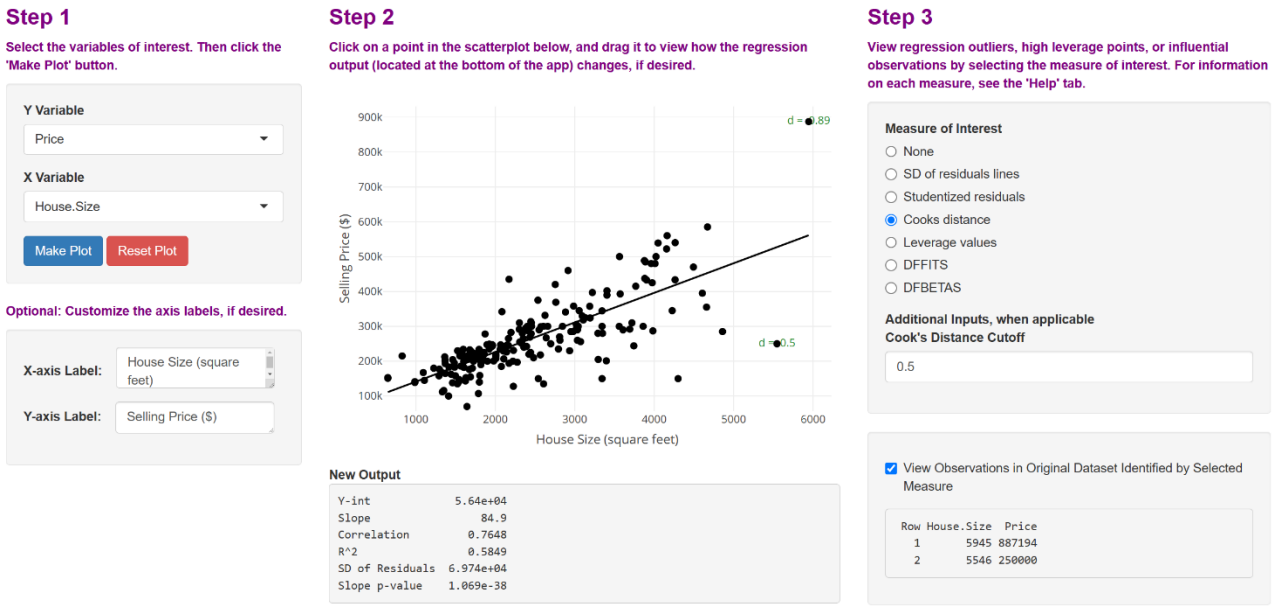


Figure 6. After the user clicks the 'Cooks distance' button and inputs a Cook's distance cutoff (0.5 here), any observations with a Cook's distance of the specified cutoff or larger is potentially influential. Here, two observations might be influential based on the user's selections.

would be less conservative). Based on the output, two observations are classified as potential influential observations based on Cook's distance with a cutoff of 0.5.

DFFITS and DFBETAS are also built into the app when exploring potential influential observations, as is leverage when assessing whether any points have high leverage. While the author has chosen to leave out screenshots of the app when working with these statistics, users should note that the steps are similar to those when working with Cook's distance. The user simply selects the button corresponding to the statistic of interest and inputs a cutoff, after which the app displays (in the scatterplot) the statistic values for the observations that meet the criteria. Users also have the option to view the row numbers of the identified observations, in the same way as before.

In addition to providing users with various options for detecting potential regression outliers, high leverage points, and influential points, the app enables users to interactively explore these types of observations. Users can do so by clicking and dragging individual points in the scatterplot. This method of exploration is aimed at helping statistics students better understand the differences among regression outliers, high leverage points, and influential points. When moving a point in the scatterplot, users can view the potential effect of individual observations on the regression output. To enable users to assess the potential impact of moving a point on regression statistics (e.g., the regression coefficients or r-squared), both the original regression output and the new regression output (i.e., the output after moving a point) are displayed, as are any updated statistic values (e.g., studentized residuals or Cook's distance).

Returning to the housing example, the focus moving forward will be on the point located at coordinates of approximately (3,000, 350,000), as seen in Figures 3-6. Suppose the user moves that point to coordinates of approximately (3,000, 700,000), as Figure 7 shows. The new point is a potential regression outlier due to its location outside of the overall trend of the data in the vertical

direction, but it is not a high leverage point because its location is not extreme in the horizontal direction relative to the other data. Since the point does not have high leverage, it is not influential. In fact, a comparison of the original and new regression output found below the scatterplot supports the lack of influence of this regression outlier, as the regression statistics across the two sets of output are quite similar. Additionally, the Cook's distance for that point is under 0.5 (since it is not identified in the plot as having a Cook's distance of at least 0.5), adding even more support to the point's lack of influence.

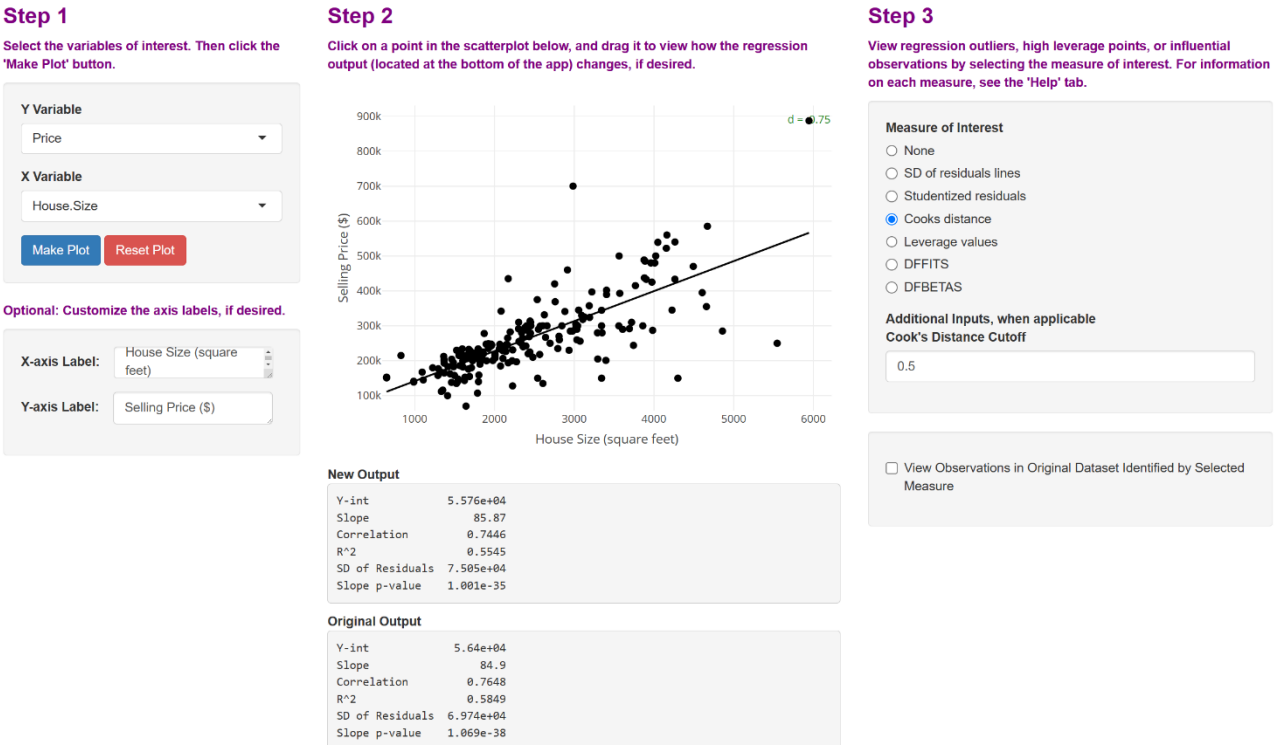


Figure 7. In the scatterplot, the point originally at coordinates of approximately (3,000, 350,000) has been moved to coordinates of approximately (3,000, 700,000). After the move, the observation is a potential regression outlier, but not a high leverage point, and thus is not an influential observation (which is also confirmed by the observation's small Cook's distance).

After resetting the plot by clicking the red 'Reset Plot' button, suppose the user now moves the same point of interest (i.e., the one originally located at coordinates of roughly (3,000, 350,000)) to coordinates of roughly (6,000, 350,000), as seen in Figure 8. The new point is a potential high leverage point due to its extreme horizontal location relative to most of the other data values, but it is not a regression outlier because it is not particularly extreme in the vertical direction. As a result, the point is not influential. A comparison of the original and new regression output found below the scatterplot supports the lack of influence of this point, as the regression statistics across the two sets of output are quite similar. The observation's Cook's distance also supports the lack of influence.

After resetting the plot once again by clicking the 'Reset Plot' button, suppose the user now moves the same point of interest (i.e., the one originally located at coordinates of roughly (3,000,

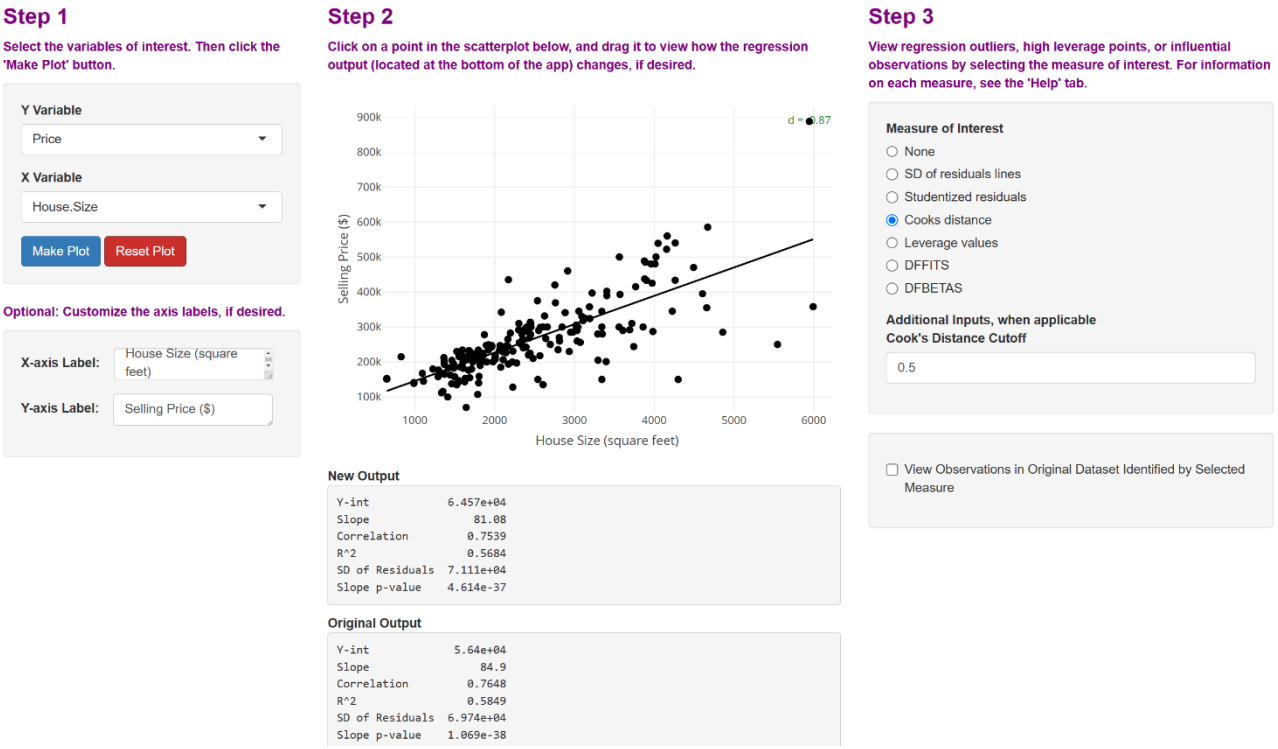


Figure 8. In the scatterplot, the point originally at coordinates of approximately (3,000, 350,000) has been moved to coordinates of approximately (6,000, 350,000). After the move, the observation is a potential high leverage point, but not a regression outlier, and is therefore not an influential observation (which is also confirmed by the observation's small Cook's distance).

350,000)) to coordinates of roughly (5,000, 1,600,000), as seen in Figure 9. The new point is potentially both a regression outlier and a high leverage point due to its falling outside of the overall pattern of the data as well as its extreme horizontal location relative to most of the other data values. As a result, the point is potentially influential. A comparison of the original and new regression output found below the scatterplot indicates the influence of this point, as the regression statistics across the two sets of output are noticeably different. In fact, the observation's Cook's distance of 2.17 is substantially larger than 0.5, which also supports the influence of the observation on the analysis.

While the previous examples of moving a point throughout the scatterplot can be used to explain potential regression outliers, high leverage points, and influential points, users could additionally have the app calculate other measures that were discussed previously in this section (e.g., studentized residuals and leverage). By doing so, instructors and students could then fully tie together all of the ideas discussed in this article.

A final important note regarding moving points in the scatterplot is that the author is in no way suggesting users modify their data when running a regression analysis in practice. The app was developed specifically for pedagogical purposes: to assist (1) faculty when teaching regression outliers, high leverage points, and influential points, and (2) students when learning these important concepts.

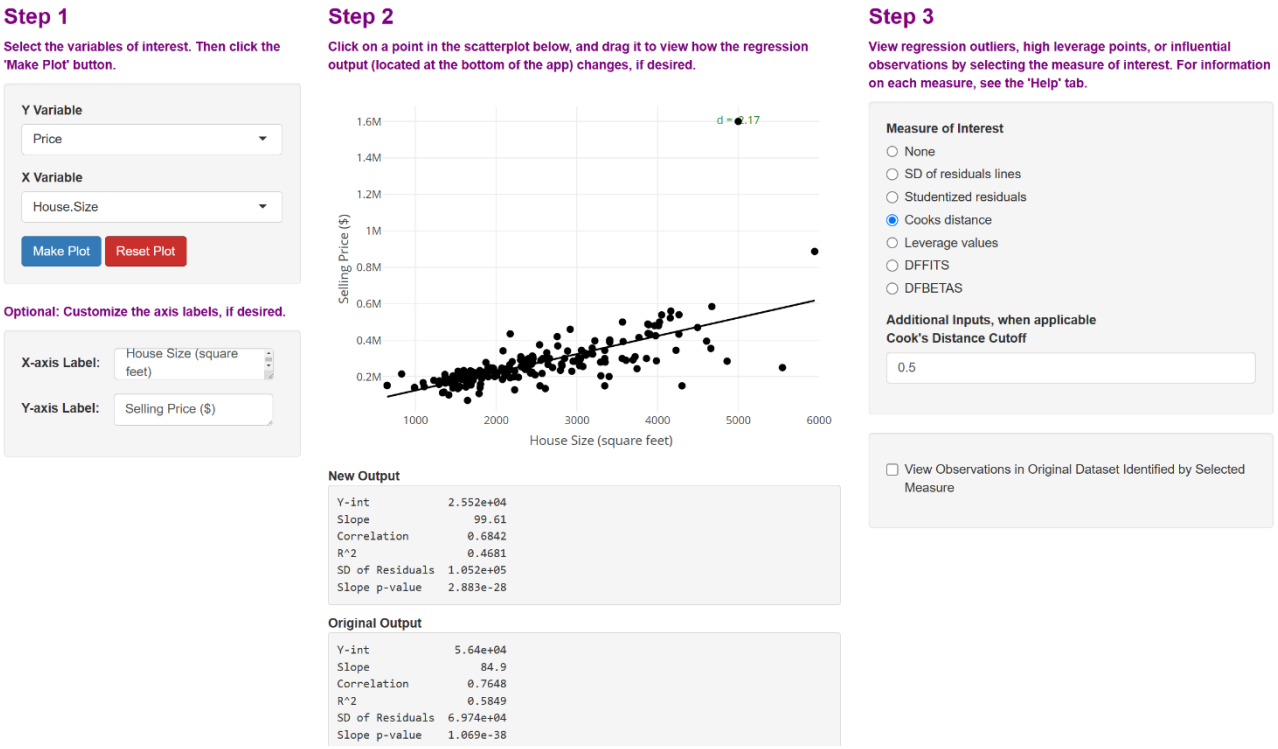


Figure 9. In the scatterplot, the point originally at coordinates of approximately (3,000, 350,000) has been moved to coordinates of approximately (5,000, 1,600,000). After the move, the observation is potentially influential, as it is appears to be both a regression outlier and a high leverage point. The observation's large Cook's distance also supports its influence.

4. Assessment of the application

An online survey focused on obtaining feedback from faculty after exploring the app described in this article is now discussed. The survey, which was created by the author and conducted using Qualtrics, was emailed to ten faculty members at Fairfield University, a comprehensive university with more than 6,000 undergraduate and graduate students, during the Fall 2024 semester. All ten of the faculty members regularly teach statistics courses that cover regression. The survey was exempt from review by the Institutional Review Board at Fairfield University and included a required question asking if participants give their consent for their responses to be used anonymously for the study. Of the ten faculty contacted, four responded to the survey, with each of the four giving their consent for their responses to be used anonymously for the study.

While the full survey can be found in the Appendix, the statements and questions contained within it focused on the accessibility, ease of use, and usefulness of the app, with Likert-scale response options provided for each of the statements and open-ended text box responses provided for the questions. The responses to each of the Likert-scale statements are summarized in Table 2. For each statement, 100% of the respondents provided positive responses about the app – consisting of either "Agree" or "Strongly Agree" for the positively worded statements (with a single exception of "Neutral" for one statement), and either "Disagree" or "Strongly Disagree" for the negatively worded statement.

Table 2. Faculty responses to Likert-scale statements about the *Influential Points* application.

Statement: Main Idea	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
App easy to access	0	0	0	1	3
App difficult to use	2	2	0	0	0
App includes all desired implementations of methods for assessing regression outliers and related points	0	0	1	0	3
Dragging individual points in the scatterplot helpful	0	0	0	2	2
Plan on using app when teaching these concepts	0	0	0	0	4

The survey then included open-ended questions asking what the respondents found most useful about the app, whether or not there are any features the app does not provide that the faculty member would find useful, and what they did not like about the app, along with suggestions for improvement. When responding to what they found most useful about the app, faculty said:

- "It is easy and straightforward to use the app and students will surely benefit from getting a visual representation of how outliers in responses and/or predictors affect the linear regression fit."
- "Availability of [a] 'missing value' check box; points are easy to move and manipulate; stores the original output so making comparisons becomes easier; allows me to use different metrics for high leverage/[influential] points."
- "The ability to quickly switch between methods to identify high leverage points."
- "Ease of use, and the wide variety of tools it has preprogrammed."

When asked if there are any features the app does not provide that they would find useful, two respondents indicated there are not, while one indicated they would like to see a table below the original regression output that lists the row information for all of the data points that meet the different criteria. Further, when asked what they did not like about the app, none of the faculty had anything negative to say. When asked about anything that could be improved, three did provide suggestions, all of which were minor in nature:

- "Since there are several methods to visualize high leverage points, perhaps a way for a user to access a definition of each method."
- "The instruction[s] could mention the location of the 'explore plot'."
- "I wonder if it was possible to add x-axis and y-axis labels on the plot."

Additional positive feedback about the app was left in response to the final question, including:

- "It was easy to use. No changes to suggest."

- "I liked all the aspects of the app, including the fact that the plot can be downloaded as a png."

In response to the constructive feedback provided by the faculty members, the author made the following changes to the app:

- The addition of a 'Home' tab that contains instructions for how to navigate the app.
- The addition of a table, located in the 'Explore Plot' tab, that lists the row information for each data point that meets the criteria specified by the user.
- The ability to add custom x-axis and y-axis labels to the scatterplot in the 'Explore Plot' tab.
- The addition of a 'Help' tab that contains URL links to websites that provide details of the methods implemented.

5. Conclusion

In this article, a free, web-based application (created by the author) for use by statistics instructors and students when covering regression outliers, high leverage points, and influential observations is described. Feedback provided by faculty members who tested the app highlights the ease of access, ease of use, and usefulness of the app. The author's hope is that the app will be helpful in both faculty instruction and student comprehension of regression outliers, high leverage points, and influential points at all levels of the statistics curriculum and in far-reaching disciplines. Potential future work includes an evaluation of the tool by statistics students, a formal assessment of the effectiveness of the tool in improving student comprehension of the concepts implemented within it, and the development of another app (or an extension of the current app) focused on these types of observations in multiple linear regression settings.

Acknowledgements

The author offers his gratitude to the editor of eJMT, the anonymous reviewers, and his colleagues for their helpful comments that strengthened both this article and the application described within it.

Appendix

The online survey sent to faculty regarding the app can be found below.

1. Do you permit the use of your responses in a study that will be submitted for publication in a journal? Note that no identifying information will be included, and your responses will remain anonymous. [required question]

    a. Yes, I give my consent for my responses to be used anonymously for the study as described.
    b. No, I do not give my consent for my responses to be used anonymously for the study as described.

Please respond to the following statements and questions about the Influential Points application, making sure to read each carefully.

2. I found the app easy to access.

    a. Strongly Disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly Agree

3. I found the app difficult to use.

    a. Strongly Disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly Agree

4. The app includes implementations of all of the methods I teach for determining whether or not a point is a regression outlier, high leverage point, and/or influential point.

    a. Strongly Disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly Agree

5. I found dragging individual points in the scatterplot helpful for visualizing the potential impact of regression outliers, high leverage points, and influential points on a regression analysis.

    a. Strongly Disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly Agree

6. I plan on using the app when I teach regression outliers, high leverage points, and/or influential points.

    a. Strongly Disagree
    b. Disagree
    c. Neutral
    d. Agree
    e. Strongly Agree

7. What did you find most useful about the app?

[text response]

8. Are there any features that the app does **not** provide that you would find useful when you teach regression outliers, high leverage points, and/or influential points?

   [text response]

9. Overall, what did you **not** like about the app? Is there anything you think could be improved? Please explain briefly.

   [text response]

References

[1] Biehler, R., Frischemeier, D., Reading, C., and Shaughnessy, M. (2018). Reasoning About Data In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), International Handbook of Research in Statistics Education (pp. 139-192): Springer International.

[2] Bobek, E., & Tversky, B. (2016). "Creating Visual Explanations Improves Learning," *Cognitive Research: Principles and Implications*, 1, 1-14.

[3] Chan, C., Leeper, T., Becker, J., & Schoch, D. (2023). *rio: A Swiss-Army Knife for Data File I/O*. R package version 1.2.2.

[4] Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). "The Role of Technology in Improving Student Learning of Statistics," *Technology Innovations in Statistics Education*, 1:1, 1-26.

[5] Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., & Borges, B. (2024). *shiny: Web Application Framework for R*. R package version 1.9.1.

[6] Garfield, J., & Ben-Zvi, D. (2007). How Students Learn Statistics Revisited: A Current Review of Research on Teaching and Learning, *International Statistical Review*, 75:3, 372-396.

[7] Garfield, J., & Everson, M. (2009). Preparing Teachers of Statistics: A Graduate Course for Future Teachers, *Journal of Statistics Education*, 17:2.

[8] JASP Team (2024). JASP. Version 0.19.1. Available at https://jasp-stats.org/.

[9] McNamara, A. (2018). Key Attributes of a Modern Statistical Computing Tool, *The American Statistician*, 1-30.

[10] Neumann, D., Hood, M., & Neumann, M. (2013). Using Real-life Data When Teaching Statistics: Student Perceptions of This Strategy in an Introductory Statistics Course, *Statistics Education Research Journal*, 12:2, 59-70.

[11] Pea, R. D. (1987). "Cognitive Technologies for Mathematics Education," In A. Schoenfeld, *Cognitive Science and Mathematics Education* (pp. 89-122). Erlbaum.

[12] Pearson Education (2024). *StatCrunch*. Pearson Education, London, UK. Available at https://www.statcrunch.com.

[13] R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[14] Repenning, A., Webb, D., & Ioannidou, A. (2010). "Scalable game design and the development of a checklist for getting computational thinking into public schools." *SIGCSE'10*, https://dl.acm.org/citation.cfm?id=1734357.

[15] Rumsey, D. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses, *Journal of Statistics Education*, 10:3.
[16] Sievert, C. (2020). *Interactive Web-Based Data Visualization with R, plotly, and shiny*. Chapman and Hall/CRC, Florida.
[17] Soflytics Corp. (2024). *Rguroo*, Encino, CA. Available at https://www.rguroo.com.
[18] Singer, J., & Willett, J. (1990). Improving the Teaching of Applied Statistics: Putting the Data Back into Data Analysis, *The American Statistician*, 44:3, 223-230.

Billiards: At the intersection of Physics, Geometry and Computer Algebra Systems

José A. Vallejo
jvallejo@mat.uned.es
Departmento de Matemáticas Fundamentales
Universidad Nacional de Educación a Distancia
Spain

Abstract

We study the properties of billiards (mainly elliptical) as dynamical systems, in particular, their integrability, and the existence and computation of periodic orbits. An implementation of the involved computations in a free CAS (Maxima) is also presented, as a tool for visualization and experimentation, which can be useful in web-based courses and distance learning.

1 Introduction

In a realistic setting, even the simplest question about the motion in a billiard can be difficult to solve. For instance, starting with a ball at (x_0, y_0), kicked with initial velocity (v_x, v_y), we may ask: 'Will the ball hit the table at the fixed but otherwise randomly chosen position (x, y) before 10^6 rebounds?' In general, the answer will be 'No', because there is a loss of energy when rolling and impacting on the table, so after a few rebounds the ball will stop, probably without reaching the goal point. Thus, we must assume that energy is conserved, so the motion of the ball continues indefinitely. Now the problem is that the only way we can answer, is following the ball's trajectory from the starting point, maybe along a really large number of rebounds. It is not feasible to try to answer armed just with pencil and paper, so we go for a computer and write a program to construct the trajectory given the data (x_0, y_0) and (v_x, v_y), using Newton's equations. We are then modeling the billiard as a dynamical system. Here we face the following problem: The computer can store the intermediate values of the position only up to a certain precision. This fact will influence the answer depending on whether the system is chaotic or not. In the first case, sensitivity on initial conditions will render our answer useless: Our prediction will depend on the precision used to describe the dynamics, and each rebound will amplify the initial errors committed in determining positions. Therefore, we need to be sure about the chaotic character of our billiard before we start to play.

It turns out that the chaotic character of a billiard depends on its shape. To be precise, rectangular and elliptical billiards determine integrable dynamical systems, hence their dynamics are regular. But other polygons or smooth boundaries, give rise to non-integrable, chaotic systems (indeed, there is

a conjecture by Birkhoff [2] stating that among all the billiards whose shape is given by a smooth, convex curve, only the elliptic case is integrable). We will explore what can be said about the motion in this case, and what kind of questions about it make sense, trying to keep the mathematics (and the physics) at a level as elementary as possible. In doing so, we will see that billiards provide an excellent environment for blending Classical Mechanics, Mathematics (Calculus, Geometry) and the use of Computer Algebra Systems (CAS). The possibility of numerically and visually experimenting with a concrete physical system, makes this an interesting tool, well suited for distance learning and self-study.

In this paper, we use the CAS Maxima (`https://maxima.sourceforge.io`), but everything here can be easily adapted to any other CAS. Some general references for dynamical systems are [1], [9], [12], for billiards see [7], [11], [13]. To save space, this paper does not include the source code of all the functions used in the text, only the main functions are shown in Section 6 (full code is available upon request).

2 Dynamical Systems and their integrability

For us, a dynamical system will be any physical system Σ that can be modeled using Newton's equations. A *configuration* will be any point in $\mathbb{R}^N$, where $N \in \mathbb{N}$. A given system will not be able to attain any arbitrary configuration. The allowed configurations fill up a submanifold $\Gamma \subset \mathbb{R}^N$ called the *configuration space*, and the dimension $r = \dim \Gamma$ is called the number of *degrees of freedom* of the system.

The time evolution of Σ will be described by a smooth map $\mathbf{x} : I \subset \mathbb{R} \to \Gamma \subset \mathbb{R}^N$, carrying t to $\mathbf{x}(t)$, where I is an interval containing $0 \in \mathbb{R}$. Newton's equation are then written as $m\frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}t^2}(t) = \mathbf{F}(\mathbf{x}(t))$, where $\mathbf{F} : \Gamma \subset \mathbb{R}^N \to \mathbb{R}^N$ is the *force field* (notice that, unless otherwise explicitly stated, we are going to consider only *autonomous*, velocity-independent forces). Defining the *linear momentum* $\mathbf{p} = m\mathrm{d}\mathbf{x}/\mathrm{d}t$ this is also equivalent to $\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t}(t) = \mathbf{F}(\mathbf{x}(t))$.

Example 1 *Here $N = 1$ and $\Gamma = \mathbb{R}$. We assume that the motion is determined by a potential $V(x)$, so $x : I \subset \mathbb{R} \to \mathbb{R}$ satisfies Newton's equation $m\ddot{x} = -\nabla V(x) = -V'(x)$ (In the sequel, we will use points over a letter to denote derivatives with respect to time t). More precisely, for each $t \in I \subset \mathbb{R}$ we have a configuration $x(t) \in \mathbb{R}$, and the curve $t \mapsto x(t)$ satisfies the ordinary differential equation $m\ddot{x} = -V'(x(t))$. This is an example of one-dimensional motion.* △

We now introduce two important classes of systems: the conservative and the integrable ones. They both have in common the existence of conserved quantities of the motion.

2.1 Conservative systems

Consider the case of a time-independent force which comes from a potential function $V \in \mathcal{C}^\infty(\Gamma)$. Thus, we have

$$m\ddot{x}(t) = -\nabla V(\mathbf{x}(t))\,. \tag{1}$$

These systems form a particular class, called *conservative systems*. The reason is that a certain function, called the *energy*, is conserved along their motion.

To explain this fact, we need to introduce some more terminology. Recall that the allowed configurations form the configuration space $\Gamma \subset \mathbb{R}^N$, now we are going to introduce a space containing configurations *and* momenta, the so called *phase space*. Generally speaking[1] this space is just $M = \Gamma \times \mathbb{R}^N \subset \mathbb{R}^{2N}$ (but see the footnote at the beginning of the next subsection). Each trajectory on configuration space, $\mathbf{x} : I \subset \mathbb{R} \to \Gamma \subset \mathbb{R}^N$, determines another curve on phase space, called its *lifting*, denoted $\tilde{\mathbf{x}}$ and acting as $\tilde{\mathbf{x}}(t) = (\mathbf{x}(t), \mathbf{p}(t))$.

Remark 2 *The points on phase space $M \subset \mathbb{R}^{2N}$ are called the* states *of Σ. They can be identified with the set of initial conditions for Newton's equations; thus, each state uniquely determines the evolution of the system (by the existence and uniqueness theorems for ordinary differential equations).*

For a conservative system, there exists a function on phase space $E : M \to \mathbb{R}$ such that its composition with the lifting of the actual trajectory $\mathbf{x}(t)$ is a constant mapping. This function, called the *total energy* of Σ, is defined as $E(\mathbf{x}, \mathbf{y}) = \frac{1}{2} m \langle \mathbf{y}, \mathbf{y} \rangle + V(\mathbf{x})$, where $\langle \cdot, \cdot \rangle$ is the Euclidean scalar product on $\mathbb{R}^N$, and $V : \Gamma \subset \mathbb{R}^N \to \mathbb{R}$ is the potential.

Theorem 3 (Conservation of Energy) *For any trajectory $\mathbf{x} : I \subset \mathbb{R} \to \Gamma \subset \mathbb{R}^N$ (that is, a solution of Newton's equations (1)), we have $\frac{\mathrm{d}}{\mathrm{d}t}(E \circ \tilde{\mathbf{x}})(t) = 0$.*

The proof is just a direct computation using the chain rule and equations (1).

Example 4 (One-dimensional motion) *For any* conservative *system in one dimension, the trajectory $x : I \subset \mathbb{R} \to \Gamma \subset \mathbb{R}$ must satisfy $\varepsilon = E(x, m\dot{x}) = \frac{1}{2} m \dot{x}(t)^2 + V(x(t))$ for some constant $\varepsilon \in \mathbb{R}$. Then, we can solve for $\dot{x}$ to obtain $\dot{x}(t) = \sqrt{2(\varepsilon - V(x))/m}$. Moreover, recalling the inverse function theorem, applied to $t = t(x)$, we get (in the region $\varepsilon > V(x)$ bounded by the* turning points*, that is, those configurations for which the velocity is zero, $\dot{x}(t) = 0$): $\frac{\mathrm{d}t}{\mathrm{d}x} = 1/\sqrt{\frac{2}{m}(\varepsilon - V(x))}$. Therefore, we can compute $t = t(x)$ by means of an integration:*

$$t = \int \frac{\mathrm{d}x}{\sqrt{\frac{2}{m}(\varepsilon - V(x))}},$$

and once the relation $t = t(x)$ is explicitly determined, we can invert it to get $x = x(t)$ if needed. Thus, we get the following important fact: for one-dimensional systems, the existence of one conserved quantity (the energy) leads to the solution of the equations of motion by a quadrature (the computation of an integral). We say that one-dimensional conservative systems are integrable. △

2.2 Integrable systems

We have just seen that one-dimensional conservative systems are integrable. How about higher dimensional systems? To answer this question, think of the motion not taking place in configuration space, but on phase space M. This is a $2r-$dimensional space, where r is the number of degrees of freedom[2]. If we have a set of n functions *on phase space* which are conserved quantities, $f_1, \ldots, f_n$,

[1] That is, in absence of non-holonomic constraints.

[2] We have $r = \dim \Gamma$ and M defined as $M = \Gamma \times \mathbb{R}^N$, but the factor $\mathbb{R}^N$ can actually be replaced by $\mathbb{R}^r$ because velocities must be tangent to $\Gamma \subset \mathbb{R}^N$.

then the motion will occur on the intersection $M(c_1,\dots,c_n) = M \cap f_1^{-1}(\{c_1\}) \cap \cdots \cap f_n^{-1}(\{c_n\})$, where $c_1,\dots,c_n$ are some real constants.

The dimension of this intersection (as a submanifold of M) is $2r-n$. There are some possibilities here. For instance, in the most favorable case in which $n = 2r-1$, the dimension of this intersection (as a submanifold of M) will be exactly 1, that is, we will have the motion restricted to a curve which must be the solution we are looking for. In Example 4, $r = 1 = N$, so the phase space was $2-$dimensional. The existence of just one conserved quantity defined on phase space, the energy, restricts the motion to the solution curve $x(t)$. This kind of systems, with r degrees of freedom and $n = 2r-1$ independent conserved quantities defined on phase space (also called *first integrals* or simply *integrals of the motion*)[3], are called *maximally superintegrable*.

When the system Σ has r degrees of freedom and exactly $n = r$ independent first integrals, we say that it is *integrable*, and *superintegrable* if the number n of independent first integrals satisfies $r < n < 2r-1$.

Example 5 (Central forces) *A force acting on a particle is called a* central force *if its supporting line (the line along which it acts on bodies) passes through a fixed point O, which is then called the* center *of the force. Thus, if we denote by $\mathbf{r}(t)$ the position of the particle at instant $t \in I \subset \mathbb{R}$ as measured from O, so we get the curve in configuration space $\mathbf{r}(t) = \overline{O\mathbf{x}(t)}$, we will have $\mathbf{F}(\mathbf{r}(t)) = f(\mathbf{r}(t))\mathbf{e_r}$, with $f : \Gamma \subset \mathbb{R}^N \to \mathbb{R}$ a smooth function and $\mathbf{e_r}$ the unit vector in the direction of $\overline{O\mathbf{x}(t)}$ (we omit the time dependence to avoid overloading the notation).*

When describing the motion under the action of a central force, the notion of angular momentum is crucial. The angular momentum *with respect to a point $P \in \mathbb{R}^N$ of a particle of mass m, described by the trajectory $\mathbf{x}(t)$, is defined as the function* on phase space *given by $\mathbf{L}_P(\mathbf{x},\mathbf{y}) = \overline{P\,\mathbf{x}} \times \overline{P\,\mathbf{y}}$. Thus, if we consider the lifting of the curve $\mathbf{x}(t)$ to phase space, and evaluate the angular momentum on it, writing $\mathbf{L}_P(t) = \mathbf{L}_P(\mathbf{x}(t), m\dot{\mathbf{x}}(t))$ for the composition, we get $\mathbf{L}_P(t) = m\overline{P\mathbf{x}(t)} \times \overline{P\dot{\mathbf{x}}(t)}$. In the case of central forces we obviously take $P = O$, yielding the more common expression $\mathbf{L}(t) = m\mathbf{r}(t) \times \dot{\mathbf{r}}(t)$.*

A straightforward computation shows that, when the force is central, $\dot{\mathbf{L}}(t) = 0$, that is: the angular momentum is a first integral in any central-force system. Hence, central-force systems (with $r = 3$ degrees of freedom) are superintegrable (as the three components of $\mathbf{L}$ are conserved independently)[4]. △

The motion of integrable systems is *regular* and *stable*. Under very mild conditions, precisely stated in the so-called Arnold-Liouville-Mineur (ALM) theorem, the trajectories on phase space are bounded and either periodic or quasi-periodic. In any case, small deviations from the initial conditions yield new trajectories remaining close to the original ones for some time (actually, two initially close trajectories will diverge linearly in time [4]). This feature survives when projecting down to configuration space: no chaos is possible in integrable systems.

More precisely, a geometric interpretation of the ALM theorem is this: When a system is integrable, it admits a special class of coordinates, called *action-angle coordinates*; when the system is

[3] Some additional assumptions are needed in order to guarantee that these integrals are functionally independent, that is, the intersections above effectively define a submanifold with reduced dimension, but we will not care about these details here, and we will simply assume that all the required conditions are satisfied.

[4] In fact, these systems are maximally superintegrable, due to the existence of an additional first integral provided by the Laplace-Runge-Lenz vector, which only has one independent component, giving a total of $5 = 2 \cdot 3 - 1$ integrals of motion.

expressed in terms of these, its phase space becomes foliated by tori. Each torus is labeled by a particular value of the action coordinate J (for $r = 2$), and on that torus, the motion is linear, parameterized by the angle coordinate θ. Figure 1 (taken from [8]) illustrates the aspect that phase space adopts in a typical integrable system.

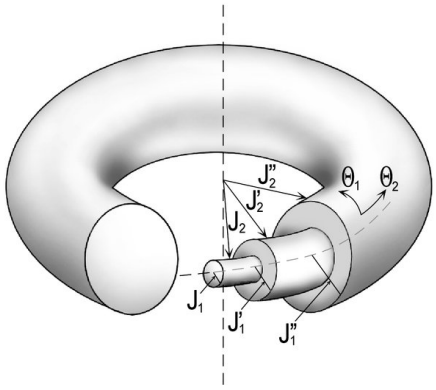


Figure 1: Foliation of phase-space by tori

3 Billiards

There are some properties common to all billiards considered as dynamical systems. Fist of all, as mentioned in the Introduction, we consider that the collisions of the ball on the boundary occur without exchanging energy (*elastic collisions*), so we have a conserved quantity. Being a two-dimensional system, every time there exists another first integral the billiard under consideration will be integrable.

The next observation is that the impact force acts along the normal to the boundary at the point of impact, in absence of friction. This property alone, allows us to deduce many important consequences that characterize the simplest cases of rectangular and circular billiards. One of them is the reflection law.

Theorem 6 (Reflection law) *In a billiard collision, the angle of reflection equals the angle of incidence.*

Example 7 (A rectangular billiard) *In this case, the hypothesis of elastic collisions means that energy (equivalently, v^2) is conserved, and the reflection law, taking into account the changes of orientation occurring at the boundary, implies that the absolute value of the horizontal component of velocity, $|v_x|$, is conserved too (in fact, both $|v_x|$ and the vertical component $|v_y|$ are conserved). The rectangular billiard is integrable, hence not chaotic.* △

3.1 Circular billiards. The Poncelet porism

The next case to consider is that of a circular billiard. Here, the forces along the normal of the boundary always point radially, that is, they are directed toward the geometric center of the circumference: We have a central force problem. Thus, aside from the energy we have another conserved quantity, angular momentum $\mathbf{L}$ (recall Example 5). This system is also integrable, not chaotic.

For circular billiards, some interesting phenomena begin to appear. Take a Cartesian coordinate system centered at the geometric origin of the billiard O, and consider the ball departing from a point $\mathbf{z}$ inside the billiard with a certain velocity, and reaching the boundary at point $\mathbf{P}$. Some point $\mathbf{x}_0$ on the segment $\overline{\mathbf{zP}}$ will be the orthogonal projection of the center O, that is, the line $\overline{\mathbf{zP}}$ will be tangent to the circle with center O and radius $\|\mathbf{x}_0\|$. Also, if the ball incides on $\mathbf{P}$ forming an angle θ_0 with respect to the line containing $\overline{O\mathbf{P}}$, it will rebound with an angle $-\theta_0$ with respect to the same line $\overline{O\mathbf{P}}$; thus, after running a distance $\|\overline{\mathbf{x}_0\mathbf{P}}\|$ along the new direction, it will touch again the inner circumference with center O and radius $\|\mathbf{x}_0\|$, $S_{\mathbf{x}_0}(O)$. This reasoning applies without changes after each rebound, so the trajectory of the ball is such that each straight segment part of it, is tangent to the circumference $S_{\mathbf{x}_0}(O)$ (see Figure 2, where the trajectory starting from a point on the bisector line, with coordinates $(2r/9, 2r/9)$ and initial angle 0.469π, r being the radius of the circumpherence, is shown after 50 rebounds).

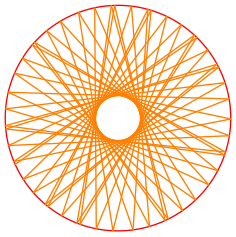


Figure 2: Caustics in a circular billiard

Due to this property, that circumference is called the *caustic* of the trajectory (geometrically, it is just the envelope).

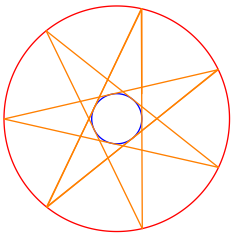


Figure 3: Closed trajectories in a circular billiard

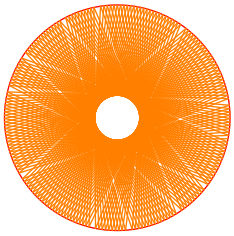


Figure 4: Non-closed trajectories in a circular billiard

The trajectory in Figure 2 is not closed. One could wonder whether there exists closed trajectories or not, so let us try with a different initial angle. Figure 3 shows the result of taking the same initial point but with initial angle (with respect to the horizontal) $\pi/2$, after 50 iterations. The caustic is drawn in blue. For a generic trajectory, though, we get a non-closed orbit; the first example considered, that of an initial angle 0.469π, gives the graph in Figure 4 after 200 iterations.

Notice, however, that none of these trajectories is *dense*: the inner region bounded by the caustic is never crossed. It turns out that there exists a close relation between caustics and closed orbits on any billiard with perimeter determined by a conic, due to a classical result on Projective Geometry discovered by J. V. Poncelet (1822).

Theorem 8 (Poncelet porism) *Consider two confocal conics $\mathcal{C}_{inn}$ and $\mathcal{C}_{out}$. Take a point $\mathbf{p}_0$ on the outer conic $\mathcal{C}_{out}$ and draw a line from it, tangent to the inner conic $\mathcal{C}_{inn}$. Determine the point of intersection of this line with the outer conic, and call it $\mathbf{p}_1$. Repeat this procedure with $\mathbf{p}_1$ as the starting point, instead of $\mathbf{p}_0$, and consider the resulting trajectory. If it is closed after n rebounds, then any other trajectory similarly constructed will be closed after n rebounds. In other words, given that a single closed polygonal trajectory exists, all other trajectories sharing the same caustic will be closed, with the same period.*

There are many proofs of this theorem available. Perhaps the most accessible is the one in [6]. The reasoning given at the beginning of this subsection proves the theorem for the particular case of concentric circles. In 3.2 we will present the proof of a weaker result valid when the outer conic is an ellipse.

Remark 9 *The existence of the caustic, and its forbidden interior, can be physically interpreted in terms of the conservation of angular momentum. Recall that $\mathbf{L} = \mathbf{r} \times \mathbf{v}$. Hence, its norm is $L = rv \sin\alpha$, where α is the angle between $\mathbf{r}$ and $\mathbf{v}$. As $\mathbf{L}$ lies on the normal to the billard's plane, its direction is constant and so the fact that is conserved translates into the constancy of L as well. We can write, then $\sin\alpha = L/rv = C/r$, with $C > 0$ constant. The trigonometric bound $\sin\alpha \leq 1$ then yields $r \geq C$. We have considered the degenerate case of a circular billiard (where the caustics are also circles), but as we will see, Poncelet porism can also be physically interpreted in the elliptic case.*

3.2 Elliptic billiards

In this section all vectors will be considered three-dimensional. To this end, consider that the ellipse describing the billiard lies on the $z = 0$ plane.

Continuing with more complicated shapes, we now analyze elliptic billiards. Suppose the boundary is given by an ellipse of major semiaxis a, and minor semiaxis b. Let $\mathbf{f} = (\sqrt{a^2 - b^2}, 0, 0)$ be the position of one of the foci, and $-\mathbf{f}$ that of the other. If $\mathbf{r}(t)$ denotes the position of the ball at the instant t, define the focal distances (see Figure 3) $\mathbf{r}_1(t) = \mathbf{r}(t) - \mathbf{f}$ and $\mathbf{r}_2(t) = \mathbf{r}(t) + \mathbf{f}$.

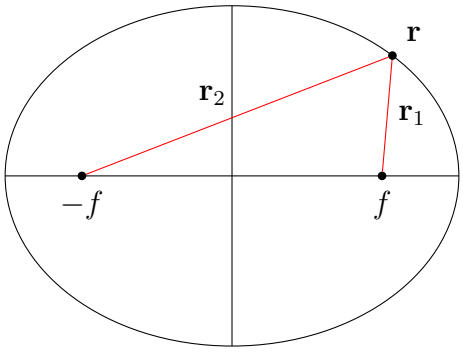


Figure 5: Ellipse geometry

The following result implies (because we already have a first integral, the energy) that an elliptic billiard is an integrable system.

Theorem 10 *The function $I = \langle \mathbf{L}_1, \mathbf{L}_2 \rangle$, where $\mathbf{L}_i(t) = \mathbf{r}_i(t) \times \mathbf{v}(t)$ (for $1 \leq i \leq 2$) and $\mathbf{v}(t) = \dot{\mathbf{r}}(t)$ is the velocity, is a first integral of the motion.*

The proof is a long but straightforward computation, which is best done introducing elliptic coordinates (ξ, η) in the plane of the ellipse, defined in terms of the Cartesian coordinates (x, y) as $x = f \cosh \xi \cos \eta$, and $y = f \sinh \xi \sin \eta$.

The computation of the angular momenta $\mathbf{L}_1$, $\mathbf{L}_2$ from $\mathbf{r}_1 = (x - f, y, 0)$, $\mathbf{r}_2 = (x + f, y, 0)$ and $\mathbf{v}$, leads to that of

$$I = \langle \mathbf{L}_1, \mathbf{L}_2 \rangle = f^4(\cosh^2 \xi - \cos^2 \eta)(-\sin^2 \eta \cdot \dot{\xi}^2 + \sinh^2 \xi \cdot \dot{\eta}^2)\,. \tag{2}$$

It is clear that this expression does not change under the transformation occuring at the point of impact, $(\dot{\xi}, \dot{\eta}) \mapsto (-\dot{\xi}, \dot{\eta})$, hence the theorem follows.

The next (purely geometric) result, taken from [11], tells us what the caustics of an elliptic billiard are.

Theorem 11 *A billiard trajectory inside an ellipse forever remains tangent to a fixed confocal conic. More precisely, if a segment of a billiard trajectory does not intersect the segment $\overline{F_1F_2}$, then all the segments of this trajectory do not intersect $\overline{F_1F_2}$ and are all tangent to the same ellipse with foci F_1 and F_2; and if a segment of a trajectory intersects $\overline{F_1F_2}$, then all the segments of this trajectory intersect $\overline{F_1F_2}$ and are all tangent to the same hyperbola with foci F_1 and F_2.*

Proof. Let $\overline{A_0A_1}$ and $\overline{A_1A_2}$ be consecutive segments of a billiard trajectory, see Figure 6. Assume that $\overline{A_0A_1}$ does not intersect the segment $\overline{F_1F_2}$ (the other case is dealt with similarly). It follows from the reflection law (Theorem 6), that the angles made by segments $\overline{F_1A_1}$ and $\overline{F_2A_1}$ with the ellipse are equal. Likewise, the segments $\overline{A_0A_1}$ and $\overline{A_2A_1}$ make equal angles with the ellipse. Hence the angles $\angle A_0A_1F_1$ and $\angle A_2A_1F_2$ are equal. Reflect F_1 in $\overline{A_0A_1}$ to point F_1', and F_2 in $\overline{A_1A_2}$ to F_2'. Let B be the intersection point of the lines $\overline{F_1'F_2}$ and $\overline{A_0A_1}$, and C of the lines $\overline{F_2'F_1}$ and $\overline{A_1A_2}$. Consider

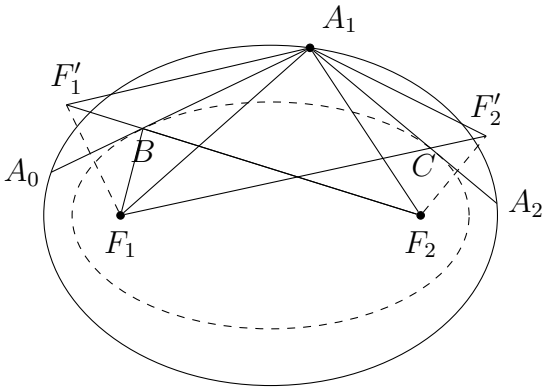


Figure 6: Caustics in an elliptic billiard

the ellipse $\mathcal{E}_1$ with foci F_1 and F_2 that is tangent to the line $\overline{A_0A_1}$. Since the angles $\angle F_2BA_1$ and $\angle F_1'BA_0$ are equal, and so are the angles $\angle F_1'BA_0$ and $\angle F_1BA_0$, the angles $\angle F_2BA_1$ and $\angle F_1BA_0$ are equal. Again by the reflection law, the ellipse $\mathcal{E}_1$ touches $\overline{A_0A_1}$ at the point B. Likewise the ellipse

$\mathcal{E}_2$ with foci F_1 and F_2 touches $\overline{A_1A_2}$ at the point C. We want to show that these two ellipses coincide or, equivalently, that $|\overline{F_1B}|+|\overline{BF_2}| = |\overline{F_1C}|+|\overline{CF_2}|$, which is reduced to $|\overline{F_1'F_2}| = |\overline{F_1F_2'}|$. Notice that the triangles $\triangle F_1'A_1F_2$ and $\triangle F_1A_1F_2'$ are congruent. Indeed, $|\overline{F_1'A_1}| = |\overline{F_1A_1}|$ and $|\overline{F_2A_1}| = \overline{F_2'A_1}|$ by symmetry. In addition, the angles $\angle F_1'A_1F_2$ and $\angle F_1A_1F_2'$ are equal: the angles $\angle A_0A_1F_1$ and $\angle A_2A_1F_2$ are equal, hence so are the angles $\angle F_1'A_1F_1$ and $\angle F_2'A_1F_2$, and adding the common angle $\angle F_1A_1F_2$ implies that $\angle F_1'A_1F_2 = \angle F_1A_1F_2'$. Equality of the triangles $\triangle F_1'A_1F_2$ and $\triangle F_1A_1F_2'$ implies that $|\overline{F_1'F_2}| = \overline{F_1F_2'}|$, which had to be proven. ■

As an immediate consequence of this theorem, we have for elliptic billiards the same relation between caustics and closed orbits that we found in the circular case, namely: Given that a single closed polygonal trajectory exists, all other trajectories sharing the same caustic will be closed, with the same period.

There is also an interpretation of Theorem 11 in terms of integrals of motion. Let our elliptic table be given by its semiaxes a, b. The constant-coordinates grid defined by the elliptic coordinates can be described by the one-parameter set of confocal conics

$$\frac{x^2}{a^2+\lambda} + \frac{y^2}{b^2+\lambda} = 1 \tag{3}$$

where the original ellipse corresponds to $\lambda = 0$, for values $\lambda \in]-b^2, +\infty[$ we get a family of ellipses sharing their foci, and for $\lambda \in]-a^2, b^2[$ we have a family of confocal hyperbolae orthogonal to that of the ellipses.

Suppose now that the initial segment of a trajectory is tangent to one of these constant-coordinate curves. Can we tell which one is going to be? If the segment is described by the equation $y = mx+c$, the tangency condition means that there exists a point (x_0, y_0) belonging to both curves, the line $y = mx + c$ and a certain member of the family (3), yielding the system

$$\begin{cases} \dfrac{x_0^2}{a^2+\lambda} + \dfrac{y_0^2}{b^2+\lambda} = 1 \\ y_0 = mx_0 + c\,. \end{cases}$$

After some algebraic manipulations, we get $(m^2+1)\lambda = c^2 - b^2 - m^2a^2$. Thus, the value of λ determining the tangent confocal conic is

$$\lambda = \frac{c^2 - (m^2a^2 + b^2)}{1+m^2}\,. \tag{4}$$

Notice that, because of Theorem 11, all the subsequent segments of the trajectory are tangent to the same conic, and the preceding reasoning applies to any of them. In other words, λ is a first integral of motion.

Thus, we have two known integrals, $I = \langle \mathbf{L}_1, \mathbf{L}_2 \rangle$ and a new one λ. They are not independent, though (again, the proof is a long but straightforward computation).

Proposition 12 *The integrals I and λ are related by*

$$\lambda = I - b^2\,.$$

4 Periodic orbits

In this section, following Sieber [10], we determine the conditions that a trajectory on an elliptic billiard must satisfy in order to be periodic (notice how this approach is different to the one used in previous works [3, 5] where periodic orbits are found by numerical brute force and trial and error, following a trajectory until it passes again through the intial point, within a certain tolerance). For notational reasons, we will introduce a rescaling of the invariant $I = \langle \mathbf{L}_1, \mathbf{L}_2 \rangle$ by the energy (which is also an invariant):

$$\alpha = \frac{I}{4E}\,. \tag{5}$$

Our main tool to determine periodic orbits is the Arnold-Liouville-Mineur theorem, and this requires computing the action coordinates of the elliptic billiard. As these are given by integrals of the momenta, and they turn out to be related to the first integrals of the system, it will be convenient to write the momenta in terms of the integrals E, α (we could as well take $E, I = L_1 L_2$, but the expressions in terms of E, α are cleaner, as we will see). Some long and boring computations lead to

$$p_\eta^2 = E(f^2 \sin^2\eta + \alpha)\,, \tag{6}$$

and

$$p_\xi^2 = E(f^2 \sinh^2\xi - \alpha)\,. \tag{7}$$

Now we are ready to build the action integrals. These are

$$I_\xi = \frac{1}{2\pi}\oint p_\xi \,\mathrm{d}\xi = \frac{\sqrt{E}}{\pi}\int_{\xi_0}^{\xi_1} \sqrt{f^2\sinh^2\xi - \alpha}\,\mathrm{d}\xi \tag{8}$$

$$I_\eta = \frac{1}{2\pi}\oint p_\eta \,\mathrm{d}\eta = \frac{2\sqrt{E}}{\pi}\int_{\eta_0}^{\eta_1} \sqrt{f^2\sin^2\eta + \alpha}\,\mathrm{d}\eta\,, \tag{9}$$

where the limits of integration are given by

- For $\alpha > 0$, $\xi_0 = \operatorname{asinh}\frac{\sqrt{|\alpha|}}{f}, \xi_1 = \operatorname{asinh}\frac{b}{f}, \eta_0 = 0, \eta_1 = \frac{\pi}{2}$.
- For $\alpha < 0$, $\xi_0 = 0, \xi_1 = \operatorname{asinh}\frac{b}{f}, \eta_0 = \arcsin\frac{\sqrt{|\alpha|}}{f}, \eta_1 = \frac{\pi}{2}$.

Not surprisingly at all, the integrals in (8) and (9) can be solved in terms of elliptic functions:

$$I_\xi = \begin{cases} \dfrac{\sqrt{E}}{\pi}\left[\dfrac{a}{b}\sqrt{b^2-\alpha} - \dfrac{f}{\kappa}E\left(\arcsin\sqrt{\dfrac{b^2-\alpha}{b^2}}, \kappa\right)\right] \\ \dfrac{\sqrt{E}}{\pi}\left[\dfrac{ab}{\sqrt{b^2-\alpha}} - \dfrac{\alpha}{f}F\left(\arcsin\sqrt{\dfrac{b^2}{b^2-\alpha}}, \dfrac{1}{\kappa}\right) - fE\left(\arcsin\sqrt{\dfrac{b^2}{b^2-\alpha}}, \dfrac{1}{\kappa}\right)\right] \end{cases},$$

(for $\alpha > 0$ and $\alpha < 0$, respectively) and

$$I_\eta = \begin{cases} \dfrac{2\sqrt{E}}{\pi}\dfrac{f}{\kappa}E\left(\kappa\right) \text{ , for } \alpha > 0 \\ \dfrac{2\sqrt{E}}{\pi}\left[\dfrac{\alpha}{f}K\left(\dfrac{1}{\kappa}\right) + fE\left(\dfrac{1}{\kappa}\right)\right] \text{ , for } \alpha < 0\,. \end{cases}$$

where the modulus is

$$\kappa = \frac{f}{\sqrt{f^2+\alpha}}\,,$$

and F, E denote, respectively, the elliptic incomplete integrals of the first and second kind, while K is the *complete* elliptic integral of the first kind.

The periodic orbits live on the tori given by the Arnold-Liouville-Mineur theorem, and in order to be closed trajectories on these tori, the quotient of the angular frequencies w_ξ, w_η must be rational. Thus, we have the condition that, for some $m, n \in \mathbb{N}$,

$$\frac{w_\xi}{w_\eta} = \frac{\left.\frac{\partial E}{\partial I_\xi}\right|_{I_\eta}}{\left.\frac{\partial E}{\partial I_\eta}\right|_{I_\xi}} = -\left.\frac{\partial I_\eta}{\partial i_\xi}\right|_E = -\frac{\left.\frac{\partial i_\eta}{\partial \alpha}\right|_E}{\left.\frac{\partial I_\xi}{\partial \alpha}\right|_E} = \frac{n}{m}\,, \tag{10}$$

where, by the properties of elliptic integrals,

$$\left.\frac{\partial I_\xi}{\partial \alpha}\right|_E = \begin{cases} -\dfrac{\sqrt{E}\kappa}{2\pi f} F\left(\arcsin\sqrt{\dfrac{b^2-\alpha}{b^2}}, \kappa\right) \text{, for } \alpha > 0 \\ -\dfrac{\sqrt{E}}{2\pi f} F\left(\arcsin\sqrt{\dfrac{b^2}{b^2-\alpha}}, \dfrac{1}{\kappa}\right) \text{, for } \alpha < 0\,, \end{cases}$$

and

$$\left.\frac{\partial I_\eta}{\partial \alpha}\right|_E = \begin{cases} \dfrac{\sqrt{E}\kappa}{\pi f} K(\kappa) \text{, for } \alpha > 0 \\ \dfrac{\sqrt{E}}{2\pi f} K\left(\frac{1}{\kappa}\right) \text{, for } \alpha < 0\,. \end{cases}$$

Two possibilities that can appear, depending on the sign of the invariant α. For instance, when $\alpha > 0$ (the case $\alpha < 0$ can be similarly analyzed) the condition for periodic orbits (10) reads

$$F\left(\arcsin\sqrt{\frac{b^2-\alpha}{b^2}}, \kappa\right) = \frac{2m}{n} K(\kappa)\,,$$

or, by applying elliptic inverses,

$$\sqrt{\frac{b^2-\alpha}{b^2}} = \mathrm{sn}\left(\frac{2m}{n} K(\kappa)\right)\,, \tag{11}$$

which has a solution for all $n \in \mathbb{N}, n \geq 3$, and $1 \leq m < n/2$. The integers n, m have the interpretation of number of rebounds of the orbit and its rotation number[5], respectively.

The values of κ (hence, α) that satisfies (11) must be determined numerically but, fortunately, there exist very fast and stable algorithms for dealing with elliptic functions.

[5]Roughly, how many times the trajectory makes a whole loop around.

4.1 Practical computations

Let us see how the knowledge of the first integral α (as numerically computed in (11)) determines the slope of the first segment of a closed orbit. We will make use of the invariant λ, which appeared in subsection 3.2. There, we saw that λ parameterizes a family of confocal ellipses in such a way that $\lambda = 0$ corresponds to the boundary ellipse of the billiard, and we could determine the value of λ corresponding to the confocal ellipse to which the first segment of a trajectory is tangent. If that segment belongs to the line $y = mx + c$, we found in (4) that

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

Suppose now we want to compute a closed $n-$orbit starting at a point (x_0, y_0) on the border of the billiard. Let $y = mx + c$ be the line containing its initial segment. As it starts at (x_0, y_0), we must have $y_0 = mx_0 + c$, or $c = y_0 - mx_0$. Substituting in the expression for λ above, we get a relation between the slope m and λ: $\lambda = ((y_0 - mx_0)^2 - (m^2a^2 + b^2))/(1 + m^2)$. Moreover, in Proposition 12 we found a relation between λ and the invariant I. Let us make the assumption that $E = 1/4$, for simplicity[6]. In this case, we have $\alpha = I$, and also $\lambda = \alpha - b^2$. By considering these expressions for λ together, we get a system of equations which leads to

$$\alpha = \frac{(y_0 - mx_0)^2 - m^2(a^2 - b^2)}{1 + m^2}\,.$$

This is the expression used in the Maxima code, but we can turn it into an explicit relation between m and α: First, write $(1+m^2)\alpha = (y_0 - mx_0)^2 - m^2(a^2 - b^2)$, then factor terms containing powers of m, $(\alpha+x_0^2+a^2-b^2)m^2+2x_0y_0m-y_0^2 = 0$, and finally get $m = \left(x_0y_0 \pm \sqrt{\alpha + a^2 - b^2 + 2x_0y_0}\right)/(\alpha + a^2 - b^2 + x_0^2)$.

Although the computations can be made simbollically, for large orbits this would cause a buffer overflow, hence the code converts everything to bfloats, thus introducing small rounding errors that, if so desired, can be avoided substituting the `bfloat` commands by `radcan` ones.

5 Simulations in a CAS

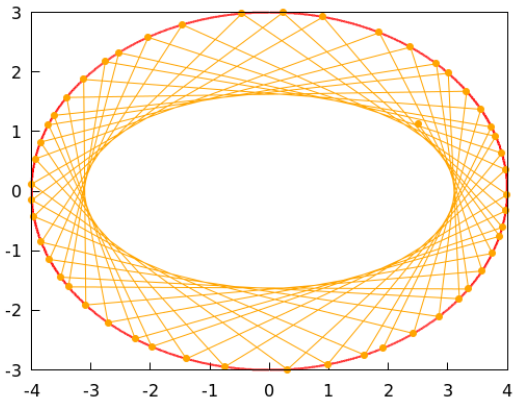
5.1 Elliptic billiards

The command `elliptic_billiard(a,b,xinit,yinit,phi0,N)` simulates the behavior of an elliptic billiard table, where `a,b` are the major and minor semiaxes of the bounding ellipse, `xinit,yinit` are the coordinates of the initial point (it can lay in the interior of the ellipse), `phi0` is the initial angle (measured counterclockwise in radians with respect to the horizontal), and the positive integer `N` is the number of rebounds.

Let us see some examples of their use:

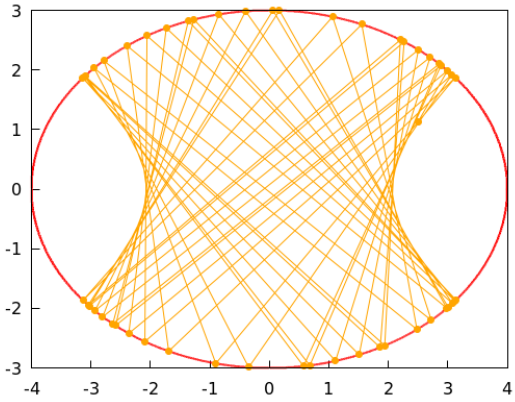
```
(%i1) elliptic_billiard(4,3,2.5,1.13,0.7*%pi,50);
```

[6]That just means that we will be selecting a particular $n-$orbit among the many possible.



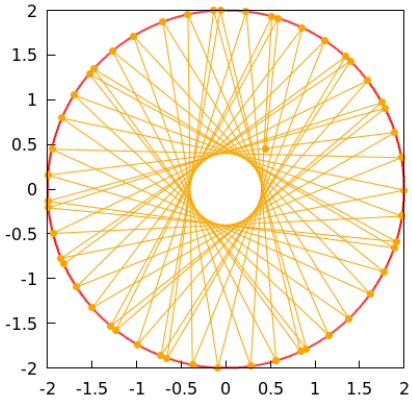
The same initial point but with a different initial slope:

```
(%i2) elliptic_billiard(4,3,2.5,1.13,%pi/3,50);
```



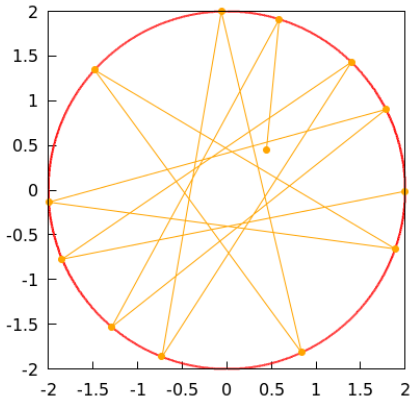
Here we see the degenerate case of a circular billiard:

```
(%i3) elliptic_billiard(2,2,4/9,4/9,0.469*%pi,50);
```



An animation of this last example can be produced with the previous command suitably modified:

```
(%i4) elliptic_billiard_animated(2,2,4/9,4/9,0.469*%pi,50);
```



For the sake of completeness, let us comment on some of the computations required in the function `elliptic_billiard` (see Section 6). First, assume that we already know a point of impact with the boundary, (x_n, y_n), and the velocity after the rebound, (u_n, v_n), so we want to determine the next point of impact (x_{n+1}, v_{n+1}). The parametric equations of the line $\mathbf{r}(t)$ joining these two points are $r_1(t) = x_n + tu_n$ and $r_2(t) = y_n + tv_n$. This line meets the ellipse at two points, determined by the condition $r_1^2/a^2 + r_2^2/b^2 = (x_n + tu_n)^2/a^2 + (y_n + tv_n)^2/b^2 = 1$. Developing this expression and taking into account that (x_n, y_n) lies on the ellipse (so $x_n^2 + y_n^2 = 1$) we get

$$\left(\frac{2x_nu_n + tu_n^2}{a^2} + \frac{2ty_nv_n + tv_n^2}{b^2}\right) t = 0\,,$$

which has two solutions: $t = 0$ (corresponding to (x_n, y_n)), and the one of interest for us: $t = -2(x_nu_nb^2 + y_nv_na^2)/(b^2u_n^2 + a^2v_n^2)$. By substituting in the equations above, we find the coordinates of the point of intersection:

$$\begin{cases} x_{n+1} = x_n - 2\dfrac{x_nu_nb^2 + y_nv_na^2}{b^2u_n^2 + a^2v_n^2}u_n \\ y_{n+1} = y_n - 2\dfrac{x_nu_nb^2 + y_nv_na^2}{b^2u_n^2 + a^2v_n^2}v_n\,. \end{cases}$$

Next, we need to compute the velocity vector after the impact, (u_{n+1}, v_{n+1}). This vector is the reflection of the incoming velocity (u_n, v_n) across the normal line of the ellipse at (x_{n+1}, y_{n+1}). The inward normal[7] (N_1, N_2) is easily found:

$$(N_1, N_2) = -\frac{\left(\frac{2x_n}{a^2}, \frac{2y_n}{b^2}\right)}{\left\|\left(\frac{2x_n}{a^2}, \frac{2y_n}{b^2}\right)\right\|} = \frac{-1}{\sqrt{b^4x_{n+1}^2 + a^4y_{n+1}^2}}(b^2x_{n+1}, a^2y_{n+1})\,.$$

[7] We need to take the inward normal, so the ball bounces *into* the ellipse.

The reflection of (u_n, y_n) across the line determined by (x_{n+1}, y_{n+1}) and (N_1, N_2) is given by $(u_{n+1}, v_{n+1}) = (u_n, v_n) - 2\,\langle (u_n, v_n), (N_1, N_2)\rangle\,(N_1, N_2)$, which leads to the expressions used in the code (compare with [3]):

$$\begin{cases} u_{n+1} = u_n - \dfrac{2(b^2 x_{n+1} u_n + a^2 y_{n+1} v_n)}{b^4 x_{n+1}^2 + a^4 y_{n+1}^2} b^2 x_{n+1} \\ v_{n+1} = v_n - \dfrac{2(b^2 x_{n+1} u_n + a^2 y_{n+1} v_n)}{b^4 x_{n+1}^2 + a^4 y_{n+1}^2} a^2 y_{n+1}\,. \end{cases}$$

The preceding reasonings can *not* be applied to the first segment of the orbit, as (x_0, y_0) does not necessarily lie on the ellipse[8]. Thus, we start with the initial point (x_0, y_0) and the initial velocity, which is the unitary vector (u_0, v_0) along the direction given by the slope $m_0 = \tan\phi_0$, that is, $(u_0, v_0) = (1, \tan\phi_0)\,/\sqrt{1 + \tan^2\phi_0}$ (the particular cases $\phi_0 = \pi/2$ and $\phi_0 = 3\pi/2$, for which (u_0, v_0) are $(0, 1)$ and $(0, -1)$, respectively, are treated separatedly).

We want an explicit formula for (x_1, y_1) depending on these data alone. The initial segment of the orbit is the line $y = m_0 x + c_0$, so $c_0 = y_0 - m_0 x_0$. Substitution into the ellipse's equation yields $(x/a)^2 + ((m_0 x + c_0)/b)^2 = 1$, so the x coordinate of the new point must satisfy $(b^2 + a^2 m_0^2)x^2 + 2a^2 m_0 c_0 x + a^2(c_0^2 - b^2) = 0$, a quadratic equation with two solutions, depending only on (a, b, m_0, c_0). Let us label them $x_+ \geq x_-$. A geometric interpretation of $y = m_0 x + c_0$ and the relative positions of $(x_0, y_0), (x_1, y_1)$, makes clear that we must choose x_+ when $u_0 > 0$ (so the line $y = m x_0 + c_0$ is traveled left to right and the abcissa x_1 is to the right of x_0), and choose x_- when $u_0 < 0$ (so the line $y = m x_0 + c_0$ is traveled right to left and the abcissa x_1 is to the left of x_0).

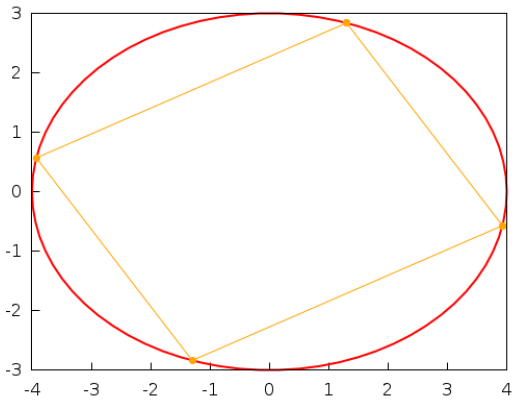
5.2 Periodic orbits

5.2.1 Case $I > 0$ (elliptic caustic)

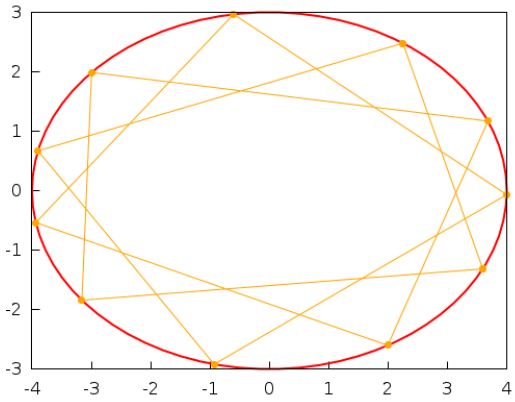
The function is `norbits_elliptic(a,b,x0,N,m)`. Its syntax should be clear, the only caveat is that N represents the number of points in the orbit (number of bounces or rebounds), while m represents the rotation number (roughly speaking, how many times the closed orbit winds around), which must satisfy the restriction $m < N/2$. As the starting point is taken on the ellipse, only the initial coordinate `xinit` is needed (the corresponding `yinit` is determined from the ellipse equation). Here are some examples, starting with a $4-$orbit on an ellipse of semiaxes $a = 4$, $b = 3$, and rotation number $N = 1$ (the initial point is taken at $x_0 = 1.3$):

```
(%i5) norbits_elliptic(4,3,1.3,4,1);
```

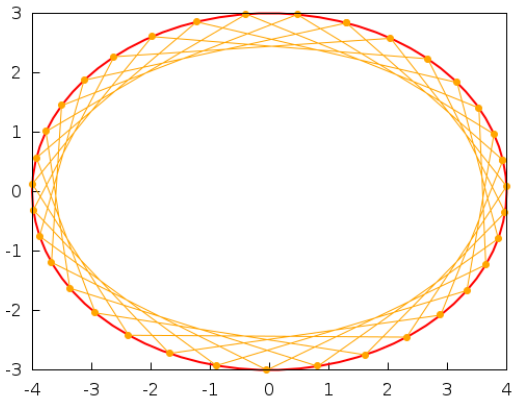
[8] In the formulae for (u_{n+1}, v_{n+1}) we use the expressions for (x_{n+1}, y_{n+1}), which were obtained under the assumption that (x_n, y_n) was on the ellipse.




```
(%i6) norbits_elliptic(4,3,2.25,11,3);
```



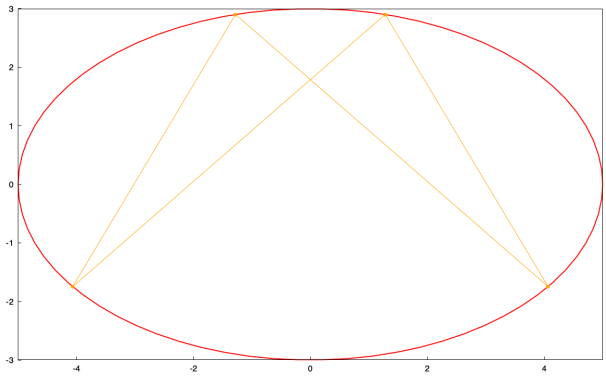

```
(%i7) norbits_elliptic(4,3,1.3,35,6);
```



5.2.2 Case $I < 0$ (hyperbolic caustic)

The command and syntax are completely analogous to the previous one. It will suffice to give an example of its use:

```
(%i8) norbits_hyperbolic(5,3,1.28,4,1);
```



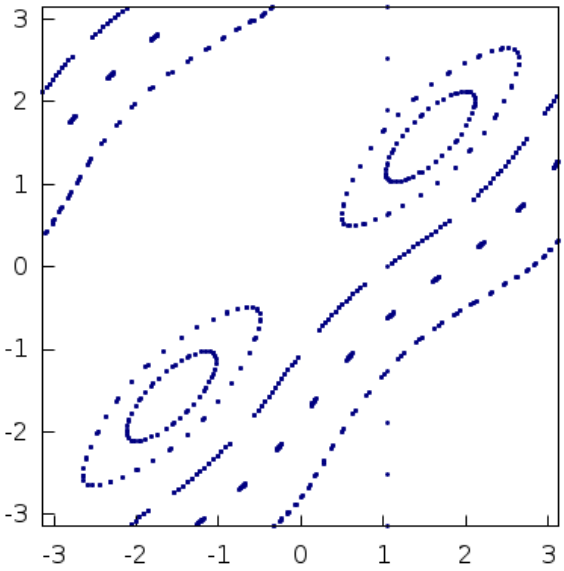
As a particular feature of this case, there are some conditions that must be satisfied in order to get an orbit. If that does not occur, a message will be issued.

```
(%i9) norbits_hyperbolic(4,3,1.28,6,1);
(%o9) With these parameters, the motion takes place on a complex
torus. Please, select them so the condition b/a<sin(m*%pi/N) is
satisfied
```

5.3 Poincaré sections

It is also possible to analyze Poincaré sections nof elliptic billiards, which are the basic tool to study the onset of chaos in a dynamical system [9],[12]. This is done by the function below, whose arguments are the semiaxes a, b, a list of initial conditions of the form $[[\theta_0, \phi_0], \ldots, [\theta_k, \phi_k]]$ (such that the initial point on the ellipse is $(x_0, y_0) = (a\cos\theta_0, b\sin\theta_0)$ and the slope of the initial trajectory is $m_0 = \tan\phi_0$), and the number of iterations N. As an example, with the following command we compute the Poincaré section of an elliptic billiard of semi-axes $a = 4$, $b = 3$, following the trajectories of 11 points given by initial conditions of the form $[\theta_j, \phi_j] = [\pi/3, -\pi + 2\pi j/10]$, along 100 rebounds:

```
(%i10) eb_poincare(4,3,makelist([%pi/3,-%pi+2*j*%pi/10],j,0,10),100);
```



Notice how these Poincaré sections are, basically, deformations of that of a harmonic oscillator, reflecting the integrability of elliptic billiards.

6 Code example

As an example of the practical implementation of the ideas exposed in the paper, here is the code of the Maxima function `elliptic_billiard` and the auxilliary `norbits_elliptic_angle` used to simulate the behavior of an elliptic billiard table. The syntax is

```
elliptic_billiard(a,b,xinit,yinit,phi0,N)
```

where `a,b` are the major and minor semiaxes of the bounding ellipse, `xinit,yinit` are the coordinates of the initial point (it can lay in the interior of the ellipse), `phi0` is the initial angle (measured counterclockwise in radians with respect to the horizontal), and the positive integer `N` is the number of rebounds.

```
(%i1)      norbits_elliptic_angle(a,b,N,m):=block(
           [foc:sqrt(a^2-b^2),eeqqnn,xx],
           if is(m>=N/2) then
               return("The rotation number must satisfy m<N/2"),
           eeqqnn:bfloat(sqrt(b^2-xx)/(b)-jacobi_sn((2*m/N)*elliptic_kc(foc^2/(foc^2+xx)),foc^2/(foc^2+xx))),
           bf_find_root(eeqqnn,xx,0.001,b^2)
           )$
```

```
(%i2)     elliptic_billiard(a,b,xinit,yinit,phi0,N):=
          block(
          [ratprint:false,solns,xp,xm,segments],
          local(x,y,u,v,m,c),
          if is(a<b) then
              return("The semiaxes must satisfy b<a"),
          if (is(not(integerp(N))) or is(N<1)) then
              return("The number of rebounds must be a positive integer"),
          if is(bfloat((xinit/a)^2+(yinit/b)^2-1)>0) then
              return("The initial point does not lie on the ellipse or its interior"),
          if (is(phi0=%pi/2) or is(phi0=0.5*%pi))
              then (x[0]:bfloat(xinit),y[0]:bfloat(yinit),
              x[1]:x[0],y[1]:b*sqrt(1-(x[1]/a)^2),u[0]:0,v[0]:1)
          elseif (is(phi0=3*%pi/2) or is(phi0=1.5*%pi))
              then (x[0]:bfloat(xinit),y[0]:bfloat(yinit),
              x[1]:x[0],y[1]:-b*sqrt(1-(x[1]/a)^2),u[0]:0,v[0]:-1)
          else (
          m[0]:bfloat(tan(phi0)),
          x[0]:bfloat(xinit),
          y[0]:bfloat(yinit),
          u[0]:1/sqrt(1+(m[0])^2),
          v[0]:m[0]/sqrt(1+(m[0])^2),
          c[0]:y[0]-m[0]*x[0],
          solns:bfloat(map('rhs,solve(
          (b^2+a^2*(m[0])^2)*x^2+2*a^2*m[0]*c[0]*x+a^2*((c[0])^2-b^2),x))),
          xp:lmax(solns),
          xm:lmin(solns),
          if is(u[0]<0) then x[1]:xm
          elseif is(u[0]>0) then x[1]:xp,
          y[1]:m[0]*x[1]+c[0]
              ),
          u[1]:u[0]-2*(b^2*x[1]*u[0]+a^2*y[1]*v[0])*b^2*x[1]/(b^4*(x[1])^2+a^4*(y[1])^2),
          v[1]:v[0]-2*(b^2*x[1]*u[0]+a^2*y[1]*v[0])*a^2*y[1]/(b^4*(x[1])^2+a^4*(y[1])^2),
          for j:1 thru N do (
          x[j+1]:x[j]-2*(x[j]*u[j]*b^2+y[j]*v[j]*a^2)*u[j]/(b^2*(u[j])^2+a^2*(v[j])^2),
          y[j+1]:y[j]-2*(x[j]*u[j]*b^2+y[j]*v[j]*a^2)*v[j]/(b^2*(u[j])^2+a^2*(v[j])^2),
          u[j+1]:u[j]
          -2*(b^2*x[j+1]*u[j]+a^2*y[j+1]*v[j])*b^2*x[j+1]/(b^4*(x[j+1])^2+a^4*(y[j+1])^2),
          v[j+1]:v[j]
          -2*(b^2*x[j+1]*u[j]+a^2*y[j+1]*v[j])*a^2*y[j+1]/(b^4*(x[j+1])^2+a^4*(y[j+1])^2)
          ),
          segments:makelist([x[j],y[j]],j,0,N),
          wxdraw2d(proportional_axes=xy,
          color=red,line_width=2,nticks=75,
          parametric(a*cos(t),b*sin(t),t,0,2*%pi),
          color=orange,line_width=1,
          points_joined=true,
          point_type=filled_circle,
          points(segments))
          )$
```


Acknowledgements

The author express his gratitude to Prof. Aldo Boiti for many useful suggestions and detailed comments on the contents of this paper.

References

[1] V. I. ARNOLD, *Mathematical methods of classical mechanics*, vol. 60 of Graduate Texts in

Mathematics, Springer-Verlag New York, 1989.

[2] G. BIRKHOFF, *Dynamical Systems*, AMS publishing, Providence, USA, 1927.

[3] A. BOITI, *Exact Orbits of Light Rays Reflected Inside Ellipses, Traced by Means of Rational Formulae with the Help of the CAS Derive™ 6*, The Electronic Journal of Mathematics and Technology, 17-2 (2023), pp. 125 – 137.

[4] G. CASATI, B. V. CHIRIKOV, AND J. FORD, *Marginal local instability of quasi-periodic motion*, Physics Letters A, 77 (1980), pp. 91 – 94.

[5] G. DÁVILA-RASCÓN AND W.C. YANG, *Investigating the Reflections of Light Rays Inside Ellipses with GeoGebra, Maxima and Maple*, The Electronic Journal of Mathematics and Technology, 13-3 (2020), pp. 190 – 218.

[6] L. HALBEISEN AND N. HUNGERBÜHLER, *A simple proof of Poncelet's theorem (on the occasion of its bicentennial)*, The American Mathematical Monthly, 122 (2015), pp. 537–551.

[7] V. V. KOZLOV AND D. V. TRESHCHEV, *Billiards: A Genetic Introduction to the Dynamics of Systems with Impacts*, vol. 89 of Translations of Mathematical Monographs, American Mathematical Society, 1991.

[8] J. MASOLIVER AND A. ROS, *Integrability and chaos: the classical uncertainty*, European Journal of Physics 32 2 (2011) 431–458.

[9] E. OTT, *Chaos in Dynamical Systems*, Cambridge University Press, 2 ed., 2002.

[10] M. SIEBER, *Semiclassical transition from an elliptical to an oval billiard*, Journal of Physics A: Mathematical and General, 30 (1997) 4563–4596.

[11] I. G. SINAI, *Introduction to Ergodic Theory*, Princeton University Press, 1977.

[12] S. STROGATZ, *Nonlinear Dynamics and Chaos*, CRC Press, 2 ed., 2015.

[13] S. TABACHNIKOV, *Geometry and Billiards*, vol. 30 of Student Mathematical Library, American Mathematical Society, 2005.

Automated exploration of geometric loci: the example of dynamic constructions of hyperbolisms of plane curves

Thierry (Noah) Dana-Picard
ndp@jct.ac.il
Jerusalem College of Technology
and Jerusalem Michlala College
Israel

Abstract

Hyperbolism of a given curve with respect to a point and a line is an interesting construct, a special kind of geometric locus, not frequent in the literature. While networking between two different kinds of mathematical software, we explore various cases, involving quartics, among them the so-called Külp quartic and topologically equivalent curves, and also an example with a sextic and a curve of degree 12. By a similar but different way, we derive a new construction of a lemniscate of Gerono. First, parametric equations are derived for the curve, then we perform implicitization Gröbner bases packages and using elimination. The polynomial equation which is obtained enables to check irreducibility of the constructed curve.

1 Introduction

1.1 The needed dialog between two kinds of mathematical software

Prior to the development of Computer Algebra Systems (CAS) and Dynamic Geometry Software (DGS), mechanical devices were used to draw specific curves [29]; spirographs were among the most popular and enabled to draw epitrochoids and hypotrochoids[1] and other related curves. Here, instead of mechanical devices, we use software, both a DGS (GeoGebra, version 5.2.871.0-d, released on December 10th, 2024) and a CAS (Maple 2024). The first one enables to construct and explore the curves, their shape and topology, sometimes providing polynomial equations but not always. We use then the CAS in order to perform algebraic computations and derive polynomial equations for the curves under study. Afterwards it is possible to copy-paste the formulas to the DGS and to analyze the curves with its dynamical features. This CAS-DGS collaboration provides a useful environment for the exploration of new constructs, as in [15]. Such a dialog between the two kinds of software has been used for example in [12, 9], and for years has been wished to be more automatic [25].

[1] See https://mathcurve.com/courbes2d.gb/epitrochoid/epitrochoid.shtml and https://mathcurve.com/courbes2d.gb/hypotrochoid/hypotrochoid.shtml.

In this work, the curve are determined first by parametric equations. Implicitization is important, as a polynomial presentation may provide information whether the curve is irreducible or not (see [21], chap. 1). The algebraic part of the work is based on the theory of Gröbner bases and on Elimination; see [7, 28]. We may refer also to [2], in particular for the non familiar reader who can see how to work out elementary examples by running the algorithms "by hand". Note that Maple has an **implicitize** routine in its package *algcurves*, based on [6]. Irreducibility is also checked with automated methods; see Section 2.

Mathematical objects cannot be grasped with hands, and are approached using numerous registers of representations [17]; the classical registers for plane curves are graphical, numerical and algebraic. Parametric representations and implicit representations as two subregisters of the algebraic one. The study is made rich and efficient by switching between registers, but switching from parametric to implicit and from implicit to parametric are non trivial tasks; see [22, 27, 28, 7]. In some cases, the switch is impossible.

Switching between different representations of plane curves is an important issue, with numerous applications in computer aided design and other fields. In [30], Wang emphasizes the role in computer aided geometric design and modeling. He develops "an extremely simple method that converts the rational parametric equations for any curve or surface into an implicit equation" (we recommend also the vast bibliography there in the paper). He uses Gröbner bases, resultants, etc. In our work here, we transform the obtained parametric equations into parametric rational presentations, then into polynomial equations. We use Maple's *PolynomialIdeals* package and elimination to derive implicit equations. It is often easier to analyze the topology of a curve using an implicit presentation than a parametric presentation. Anyway, both enable the study and classification of singular points.

1.2 Plane curves defined as geometric loci

Plane algebraic curves are a classical topic, to which numerous books have been devoted, such as [32]. Websites are devoted to curves (and surfaces) such as Mathcurve (`http://mathcurve.com`). Full catalogues of curves of degree 2,3 and 4 exist, and partial catalogues for degree 6. Some of them are constructed as geometric loci. The bifocal definition of ellipses and hyperbolas is generally the first example met by students. Cassini ovals (also called spiric curves) and Cayley ovals are more advanced examples. Recently, some octic curves (curves defined by polynomials of degree 8) have been described as geometric loci in relation with a classical theorem of plane geometry, namely Thales second theorem [13]. The present paper shows a bunch of curves, more or less classical, appearing as geometric loci by construction of *hyperbolisms*.

Our concern is the construction and study of plane curves as hyperbolism of classical curves, according to the definition in the Mathcurve website [20]:

Definition 1 *The hyperbolism of a curve Γ_0 with respect to a point O and a line (D) is the curve Γ, locus of the point M defined as follows: given a point M_0 on Γ_0, the line (OM_0) cuts (D) at P; M is the projection of P on the line parallel to (D) passing by M_0.*

Remark 2 *Analytically, if the line (D) is given by the equation $x = a$, the transformation of Γ_0 into Γ can be written $(x, y) \to \left(x, a\frac{y}{x}\right)$; it is quadratic, so an algebraic curve of degree n is transformed into an algebraic curve of degree at most $2n$. The 2nd coordinate $a\frac{y}{x}$ makes the connection with hyperbolas, whence their name* hyperbolism.

By definition, hyperbolisms of curves are a subtopic of geometric loci, which are a classical topic, from middle school to university. The last decades have seen numerous works devoted to automated methods in Geometry, such as [1, 3, 4, 18] and [23], where loci are one of the main topics for which automated methods have been developed, and implemented in GeoGebra-Discovery[2]. A large number of versions of automated commands exist, we use here only a few of them, mostly GeoGebra's **Locus(<Point Creating Locus>,<Point>)** and **Locus(<Point Creating Locus>,<Point>)**. The place holders are called respectively the *Tracer* and the *Mover* in the above mentioned papers. The automated command provides a plot of the curve, sometimes also an implicit equation. For this it uses numerical methods. In the companion package GeoGebra-Discovery, this has been supplemented by symbolic algorithms yielding more precise answers. For the algebraic work, we switched to the Maple software, especially for implicitization, but not only.

The notion of a hyperbolism has been presented to a small group of in-service mathematics teachers, learning towards an advanced degree M.Ed. (Master of Education). These teachers had previous knowledge including the perpendicular bisector of a segment, a circle, conics, etc. as geometric loci, but almost neither CAS nor DGS literacy. The outcome of their work was double: the development of new perspectives in geometry and acquisition of technological skills. The feedback was very positive, and true curiosity at work.

2 First easy examples

2.1 Hyperbolism of a circle centered at the point O and the line is tangent to the circle

We consider the circle $\mathcal{C}$ centered at the origin O with radius r. The line D has equation $x = r$ and is thus tangent to the circle $\mathcal{C}$. A point $M_0 \in \mathcal{C}$ is given by $(x, y) = (r\cos t, r\sin t)$ where $t \in [0, 2\pi]$. The line OM_0 has thus equation $y = x\tan t$ and intersects the line D at $P(r, r\tan t)$. It follows that the point M has coordinates

$$(x, y) = (r\cos t, r\tan t) \tag{1}$$

Equation (1) is a parametric presentation of the geometric locus that we are looking for. Figure 1 shows a screenshot of a GeoGebra session for this question. The requested curve has been obtained using the **LocusEquation(**M, M_0**)** command. It could have been obtained also with the **LocusEquation(<Point Creating Locus Line >,< Slider >)**, after entering directly the parametrization (1), but in this case the new construct is independent of what has been done previously. The plots corresponding to the 2 last rows in the algebraic window overlap each other, therefore only one plot is viewed. Nevertheless, they are considered by the software as 2 different objects.

In this session, the radius may be changed, but the equation of the geometric locus remains of the same form. In Figure 1, $r = 4$ and the geometric locus has equation

$$x^2y^2 + 16x^2 = 256, \tag{2}$$

[2] A freely downloadable companion to GeoGebra, available from https://github.com/kovzol/geogebra-discovery. We used the November 5^{th}, 2024 version; a new version has been released on December 21^{st}, 2024.

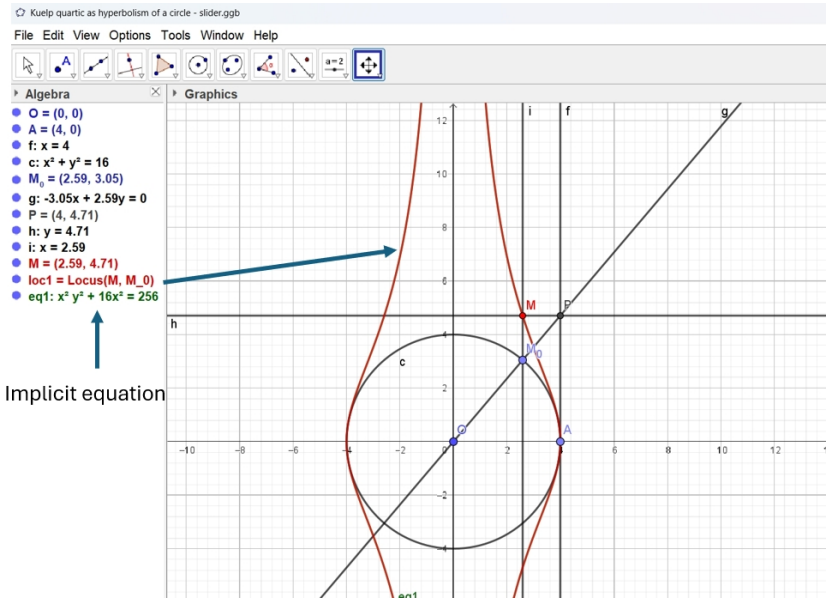


Figure 1: Hyperbolism of a circle with respect to its center and a tangent

which is the equation of a *Külp quartic*[3]. A GeoGebra session[4] may help to visualize the construct for various values of the radius. Slightly different ways to construct the hyperbolism with the software may be chosen, and for some of them an implicit equation is not obtained. The user has to modify his protocol to have it "more geometric" (v.i. Section 4). It happens that GeoGebra-Discovery displays a message telling that the construction involves steps which are not supported by the command **LocusEquation**, which gives an indication that the construction has to be more geometric (e.g. not dependent on a slider). In our applet, the tangent has not been constructed directly with its equation, as in Figure 1, but using the **Tangent** command of the software, which is important as part of a geometric construct; see Figure 2.

In order to derive from Equation (1) a polynomial equation, we use the following substitution, as in [12]:

$$\forall t \in \mathbb{R},\ \exists u \in \mathbb{R} \text{ such that } \begin{cases} \cos t = \frac{1-u^2}{1+u^2} \\ \sin t = \frac{2u}{1+u^2} \end{cases} \tag{3}$$

We apply the following Maple code:

```
xk := r*cos(t); yk := r*sin(t)/cos(t);
xkrat := subs(cos(t) = (-u^2 + 1)/(u^2 + 1), xk);
ykrat := subs(cos(t) = (-u^2 + 1)/(u^2 + 1),
         subs(sin(t) = 2*u/(u^2 + 1), yk));
p1 := x*denom(xkrat) - numer(xkrat);
p2 := y*denom(ykrat) - numer(ykrat);
J := <p1, p2>;
JE := EliminationIdeal(J, {r, x, y});
```

[3] A curve studied in 1878 by Külp; see https://mathcurve.com/courbes2d/kulp/kulp.shtml

[4] https://www.geogebra.org/m/md4f8aaa

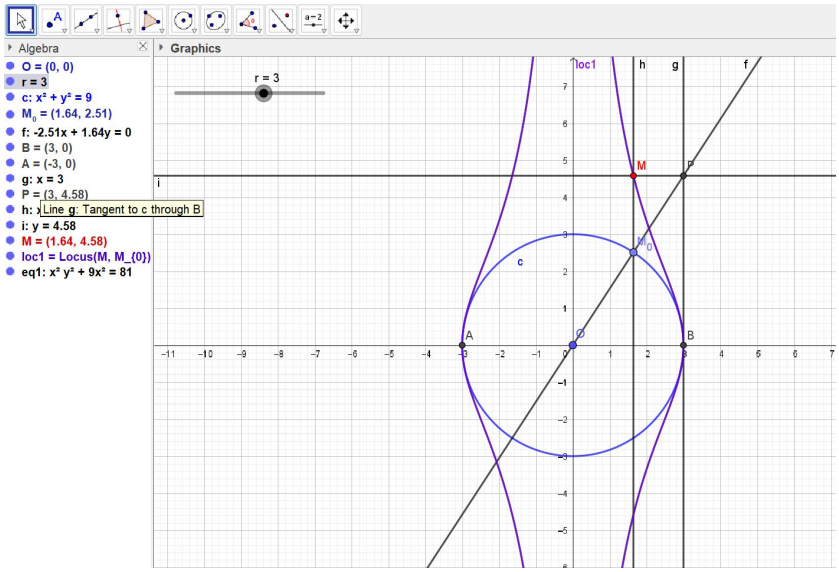


Figure 2: Hyperbolism of a circle with respect to its center and a tangent - variable radius

The output is an ideal with a unique generator, providing the following equation (for any real r):

$$r^2x^2 + x^2y^2 - r^4 = 0. \tag{4}$$

Note the dependence of the coefficients in Equation (4) on the radius of the circle. Using Maple's command **evala(AFactor(...)**, we check that the left hand side in Equation (4) is an irreducible polynomial. This means that the 2 components of the curve described by this equation cannot be distinguished by algebraic means, i.e. they are not 2 distinct components of the curve (see [21], p.17-18).

2.2 Hyperbolism of an ellipse

We generalize slightly the situation of subsection 2.1 and consider now an ellipse $\mathcal{C}$ whose equation is $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, where a and b are positive parameters[5]. Here too, the point O is the origin. The line D has equation $x = a$, i.e. is tangent to the ellipse $\mathcal{C}$ and parallel to the $y-$axis. A point $M_0 \in \mathcal{C}$ is given by $(x, y) = (a\cos t, b\sin t)$ where $t \in [0, 2\pi]$. The line OM_0 has thus equation $y = x\frac{b}{a}\tan t$ and intersects the line D at $P(a, b\tan t)$. It follows that the point M has coordinates

$$(x, y) = (a\cos t, b\tan t) \tag{5}$$

Remark 3 *The segment c in the upper left corner of Figure 3 comes instead of a slider for a parameter which is involved in the implicit equation of the ellipse. This, in order to have the ellipse dependent on a geometric construct. Actually, the ellipse itself could have been constructed as a geometric loci, but in such a case GeoGebra's command **LocusEquation** may provide a plot and an implicit equation, but not the possibility to use the command **Point on Object**.*

[5]GeoGebra applets are available at https://www.geogebra.org/m/eqywkwcw and https://www.geogebra.org/m/tfnrwaqh

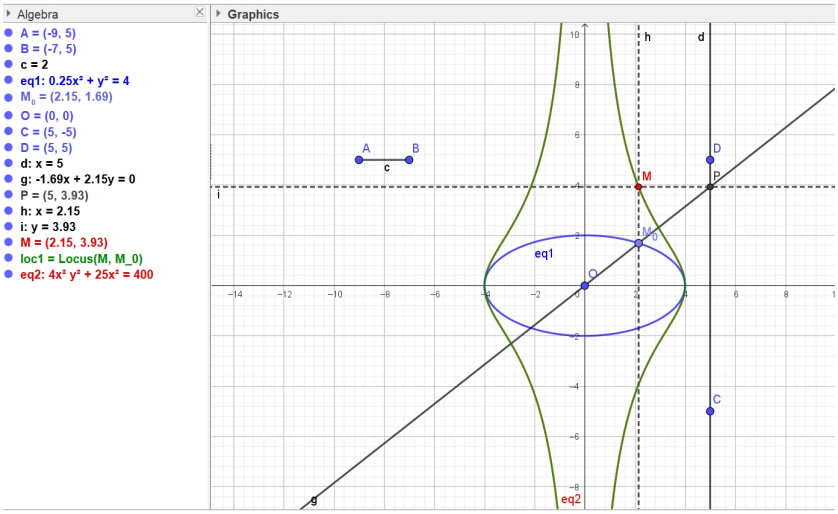


Figure 3: Hyperbolism of an ellipse

We apply a Maple code similar to the code in the previous subsection[6]. The output provides an equation for the hyperbolism of the ellipse:

$$b^2x^2 + x^2y^2 - a^2b^2 = 0. \tag{6}$$

Using the same tool as in previous section, we show that the obtained curve is irreducible. Note that it is not a Külp quartic (the coefficient of x^2 is not the square of the free coefficient). The equations are quite similar, but not identical. Figure 4 shows how a GeoGebra applet may help to understand that, for various values of the axes of the ellipse, the obtained curves have the same topology. The exploration leads to proving that two curves in this family are obtained from each other by an affinity whose axis is the $y-$axis and the direction is perpendicular to it. Moreover, this may be an opportunity to explore in class the similarities and the differences of these quartics, Külp quartic, and also the Witch of Agnesi (which is a cubic).

Remark 4 *Denote $F(x,y) = b^2x^2 + x^2y^2 - a^2b^2$. We have: $F_x(x,y) = 2x(b^2+y)$ and $F_y(x,y) = 2x^2y$. It is easy to show that the system of equations $F_x = F_y = 0$ has no solution (actually $(0,0)$ solves the system, but the origin does not belongs to the curve). Therefore the curve has no singular point.*

Remark 5 *Assuming that at inflexion points the curvature is equal to 0, it is possible to look for candidates using the following code:*

```
with(Student[VectorCalculus]);
Curvature(<a*cos(t), b*tan(t), t>, t);
simplify(%);
infl := solve(% = 0, t);
allvalues(infl[1]);
```

[6]Maple's **implicitize** command did not always provide an answer.

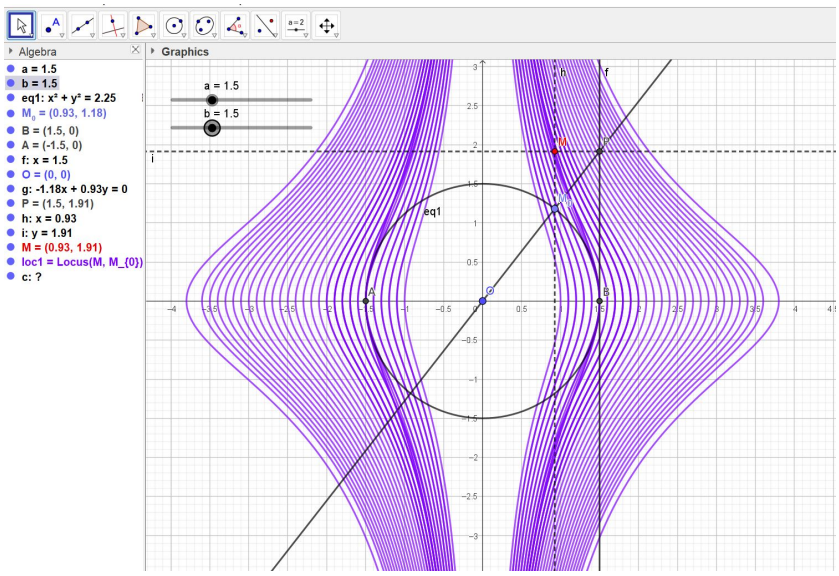


Figure 4: Showing that all the curves have the same topology

Different issues may appear. First the meanings of $< \ldots, \ldots >$ *in the two packages* PolynomialIdeals *and* VectorCalculus *are different, therefore we suggest to introduce the 2nd one only after the first one has been used. Second, the answer for general parameters* a *and* b *may be heavy; it may be wiser to use them with specific values. Finally, this provides values of the parameter* t *which can correspond to points of inflexion, but not only. More verifications are needed.*

Remark 6 *The given ellipse is the image of the circle whose equation is* $x^2 + y^2 = a^2$ *by the affine transformation* $(x, y) \longmapsto \left(x, \frac{b}{a}y\right)$. *It is easily proven that the quartic obtained here is the image by the same transformation of the Külp quartic found in the previous subsection.*

3 Hyperbolism of a circle with respect to a line secant to the circle

3.1 The circle is centered at the origin

Other quartics can be obtained as hyperbolisms of circles. We consider a circle $\mathcal{C}$ centered at the origin with radius r and the line D with equation $x = b$, where $b \neq r$. Figure 5 is a screenshot of a GeoGebra applet[7]); the line D is a secant to the circle $\mathcal{C}$. We perform the same construction as in subsection 2.1, and obtain a plot of the geometric locus, but cannot obtain an implicit equation with the command **LocusEquation**. Therefore, we need to make the algebraic computation using the CAS. Note that, even when changing the value of r by changing the length of the corresponding segment AB, the implicit equation is not obtained.

```
f := -r^2 + x^2 + y^2;
```

[7] https://www.geogebra.org/m/pwrx2uzf; a related applet is available at https://www.geogebra.org/m/x22etuhc.

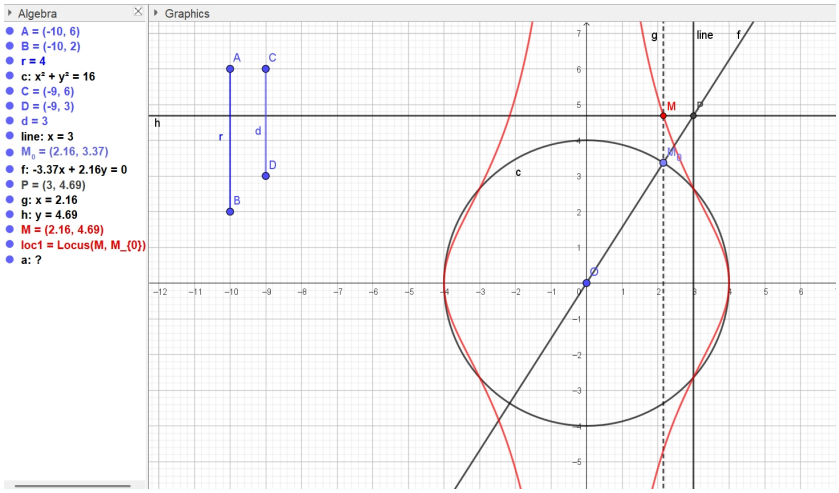


Figure 5: The line is not tangent to the circle

```
li := y = r*sin(t) + sin(t)*(x - r*cos(t))/cos(t);
yP := subs(x = d, rhs(li));
xM := r*cos(t);
yM := yP;
p1 := x - xM;
p2 := y - yM;
p1 := subs(cos(t) = (-u^2 + 1)/(u^2 + 1), p1);
p2 := subs(cos(t) = (-u^2 + 1)/(u^2 + 1), subs(sin(t) = 2*u/(u^2 + 1), p2));
p1 := simplify(p1*denom(p1));
p2 := simplify(p2*denom(p2));
J := <p1, p2>;
JE := EliminationIdeal(J, {d, r, x, y});
```

The output provides an implicit equation for the hyperbolism. After simplification, it reads as follows:

$$x^2y^2 + d^2x^2 - d^2r^2 = 0 \tag{7}$$

As expected, for $d = r$, we have Equation (4). Figure 6 shows hyperbolisms of the same circle for 4 different lines whose respective equations are $x = 1/2$, $x = 1$, $x = 2$ and $x = 3$, from the innermost to the outermost curve.

Their differences are emphasized in Figure 7, in an orthogonal but not orthonormal system of coordinates (therefore, the circle does not "look like" a circle) . These are screenshots of an animation programmed with Maple, with the **animate** command, as follows:

```
ac := animate(plot, [[r*cos(t), subs(d = A, yP), t = 0 .. 2*Pi]],
       A = 0.5 .. 4, frames = 50, color = red, thickness = 3)
```

Here too, as in subsection 2.2, the animation provides a visualization of the fact that for different values of the parameters, the obtained curve has the same topology.

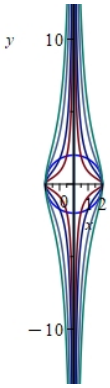


Figure 6: The line is not tangent to the circle - several plots

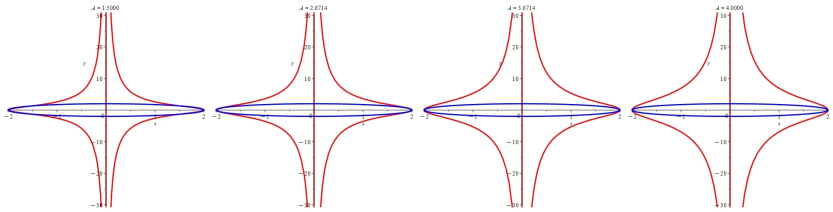


Figure 7: 4 screenshots of the Maple animation

3.2 The circle passes through the origin

We consider now a circle $\mathcal{C}$ passing through the origin and centered on the $x-$axis. The circle $\mathcal{C}$ has equation $\left(x-\frac{a}{2}\right)^2+y^2=\left(\frac{a}{2}\right)^2$. We take a line $D: x=b$, where $b<a$ and the point O is the origin. The construction is displayed in Figure 8. The segments on the left are used instead of sliders.

A plot of the geometric locus of the point M (the Tracer) when M_0 (the Mover) runs over the circle is obtained by the **Locus** command. Not as in the previous section, the command **LocusEquation** does not provide an answer. Therefore, algebraic computations have to be performed using the CAS.

First, we derive a parametric presentation for the circle $\mathcal{C}$, using the intersection of the circle and of a line through the origin with slope t (pay attention that this will be the generic line OM_0)

$$\begin{cases} x=\frac{a}{t^2+1} \\ y=\frac{at}{t^2+1} \end{cases} \tag{8}$$

For the point P: if $x_P=b$, then $y_P=tb$ and M has coordinates

$$\begin{cases} x=\frac{a}{t^2+1} \\ y=tb \end{cases}. \tag{9}$$

We define polynomials

$$P_1(x,y)=x(t^2+1)-a \qquad \text{and} \qquad P_2(x,y)=-b*t+y. \tag{10}$$

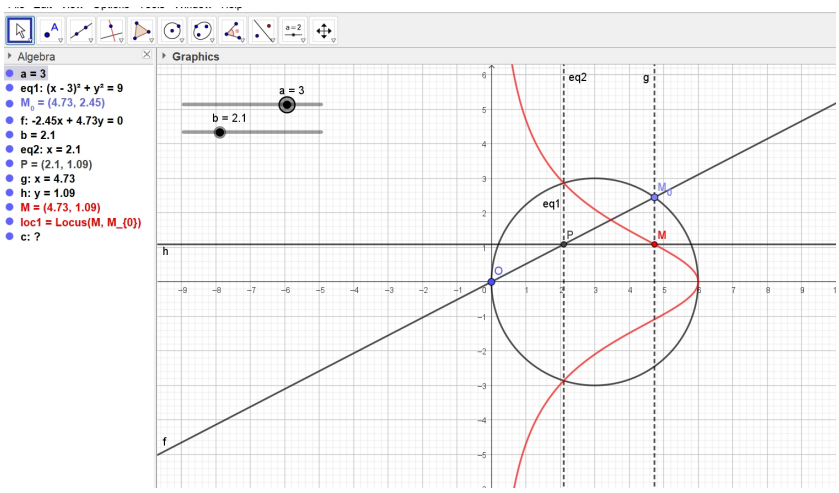


Figure 8: The line is a secant of the circle

Let $J =< P_1, P_2 >$. By elimination of the parameter t, we obtain the following implicit equation of degree 3:

$$xy^2 + b^2x - ab^2 = 0. \tag{11}$$

Note that for $b = a$, i.e. the line is tangent to the circle at the end point of a diameter passing by the origin, the curve is a Witch of Agnesi.

4 Piriform quartics

4.1 Hyperbolism of a piriform quartic curve

We refer to [19] for the implicit and parametric presentation of the curve. A piriform curve is a quartic $\mathcal{P}$ whose equation is

$$b^2y^2 = x^3(a - x), \tag{12}$$

where a and b are positive parameters. In Figure 9 (a screenshot of a GeoGebra applet[8]), these parameters are determined by 2 segments, in order to have a purely geometric construction. This figure correspond to the case $(a, b) = (3, 2)$. A plot is obtained and, simultaneously, an implicit equation, which reads as follows:

$$x(9x^2 - 54x + 4y^2) = 0 \tag{13}$$

The factor x determines the $y-$axis, emphasized in the Figure. This component is irrelevant to the geometric question; it appears because issues related to Zariski topology. Such issues are discussed in [11], and are beyond the scope of the present article. The 2nd factor determines an ellipse, whose equation can be written as follows:

$$\frac{(x-3)^2}{9} + \frac{y^2}{\frac{81}{4}} = 1 \tag{14}$$

[8] https://www.geogebra.org/m/smcehrzy

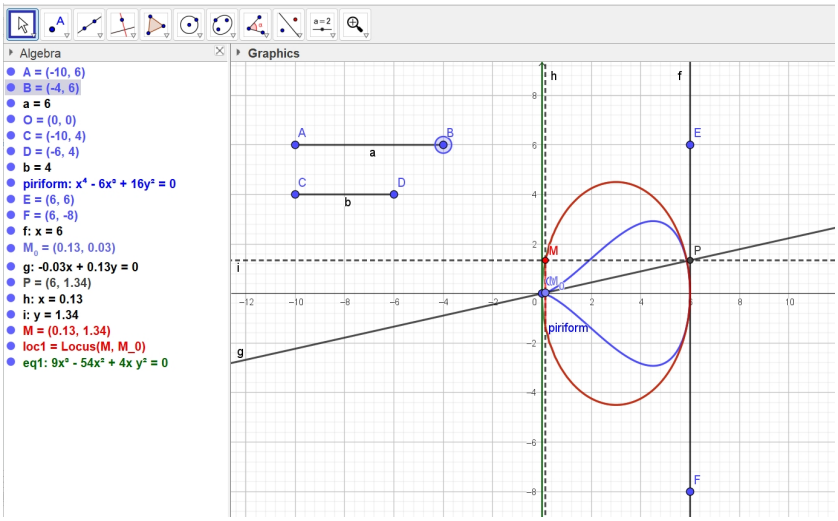


Figure 9: Hyperbolism of a piriform curve

enabling to find the geometric characteristic elements of the ellipse. This is the desired hyperbolism of the piriform curve.

In order to perform the algebraic computations, it is worth to begin with a parametric presentation of the curve $\mathcal{P}$. We may use a trigonometric parametrization[9]:

$$\begin{cases} x = \frac{a}{2}(1+\cos t) \\ y = \frac{a^2}{8b}(\sin 2t + 2\sin t) \end{cases} \tag{15}$$

but we have to transform it into a rational parametrization. We choose a different way. Any line but the $y-$axis through the origin intersects again the curve $\mathcal{P}$. Let L be the line whose equation is $y = tx,\ t \in \mathbb{R}$. We use the following Maple code:

```
pear := b^2*y^2 - x^3*(a - x);
l := -t*x + y;
solve({l = 0, pear = 0}, {x, y});
par := allvalues(%[2]);
```

The output gives two components, given by:

$$(x,y) = \left(\frac{a}{2} + \frac{\sqrt{-4b^2t^2+a^2}}{2},\ t\left(\frac{a}{2} + \frac{\sqrt{-4b^2t^2+a^2}}{2}\right)\right) \tag{16}$$

and

$$(x,y) = \left(\frac{a}{2} - \frac{\sqrt{-4b^2t^2+a^2}}{2}, t\left(\frac{a}{2} - \frac{\sqrt{-4b^2t^2+a^2}}{2}\right)\right) \tag{17}$$

We work now with the first component, the 2nd one can be treated exactly in the same way.

[9]It is given at https://mathcurve.com/courbes2d.gb/piriforme/piriforme.shtml.

Denote $x_P = a,\ y_P = ta$. By construction, we have:

$$\begin{cases} x_M = \frac{a}{2} + \frac{1}{2}\sqrt{-4b^2t^2 + a^2} \\ y_M = ta \end{cases} \tag{18}$$

Now we use the following Maple code:

```
p1 := x - a/2 =sqrt(-4*b^2*t^2 + a^2)/2;
p1 := p1^2;
p1 := lhs(p1) - rhs(p1);
p2 := -a*t + y;
J:=<p1,p2>;
JE := EliminationIdeal(J, {a, b, x, y})
G := Generators(JE)[1];
```

whose output is the polynomial

$$G(x, y) = a^2x^2 - a^3x + b^2y^2 \tag{19}$$

We have

$$G(x, y) = a^2(x^2 - ax) + b^2y^2 = a^2\left(x - \frac{a}{2}\right)^2 + b^2y^2 - \frac{a^4}{4}. \tag{20}$$

Rewriting the equation under the form

$$\frac{\left(x - \frac{a}{2}\right)^2}{b^2} + \frac{y^2}{a^2} = \frac{a^2}{4b^2}, \tag{21}$$

which is equivalent to

$$\frac{\left(x - \frac{a}{2}\right)^2}{\left(\frac{a}{2}\right)^2} + \frac{y^2}{(2b)^2} = 1, \tag{22}$$

we identify that the requested curve is an ellipse.

Remark 7 *The obtained hyperbolism has been identified as an ellipse by algebraic means. This appears also in the GeoGebra applet, but a slight modification may induce a big change, and the implicit equation may not be obtained. An experimental way to check that the curve is an ellipse consists in marking 5 arbitrary points on the curve (with the* ***Point on Object*** *command) and determine a conic by 5 points on it (there is a button driven command for this). This is a numerical checking, not a symbolic proof, but may enable students to proceed further. Of course, the algebraic computations that we performed with the CAS are a must.*

4.2 A piriform curve as antihyperbolism of a circle

The choice of the original curve in previous subsection incites to have a look at piriform curves from another point of view; see [20, 19].

Definition 8 *With the notations of Definition 1, the inverse transformation* $(x, y) \to \left(x, \frac{xy}{a}\right)$ *is called* antihyperbolism.

Screenshots of a GeoGebra applet[10] are displayed in Figure 10. Exploration using the sliders show that for $|b| > a$, the piriform curve lies inside the circle and is tangent to it at A. For $-a < b < a$, the curve is part inside and part outside of the circle, and is still tangent to the circle at A, but from outside.

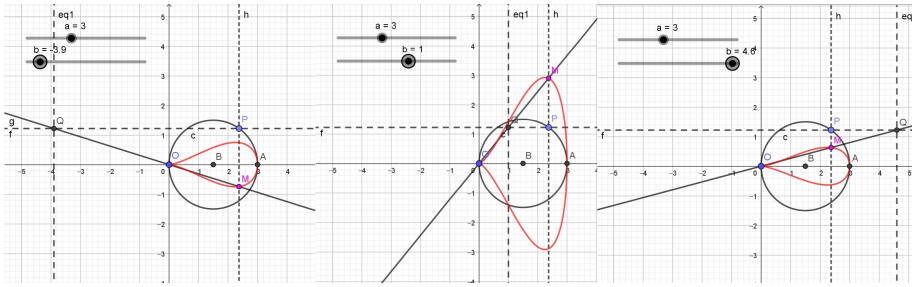


Figure 10: A piriform quartic constructed from a circle - exploration with the sliders

5 Hyperbolism of a nephroid

Let $\mathcal{C}$ be a circle whose center is at the origin. The envelope of the family of circles centered on $\mathcal{C}$ and tangent to the $y-$ axis is a sextic called a nephroid [16]. By that way the curve has been constructed in Figure 11, which is a screenshot of a GeoGebra applet[11]. We denote the nephroid by $\mathcal{N}$. Later, this figure will enable to compare the hyperbolism of $\mathcal{N}$ with a hyperbolism of a circle, as described in Subsection 2.1.

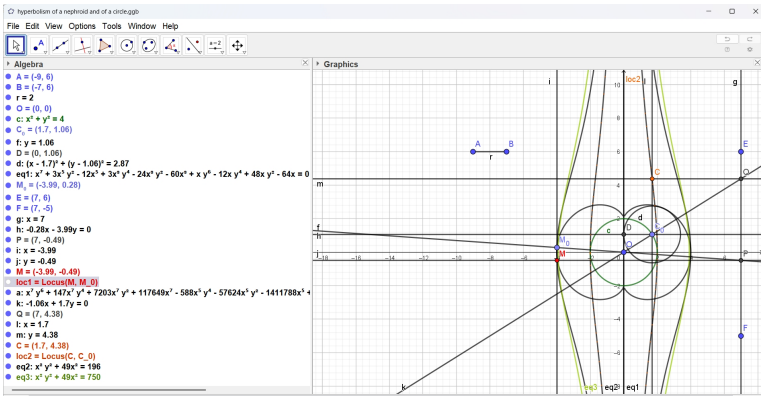


Figure 11: Comparison between hyperbolisms of a nephroid and of a circle

A general implicit equation for a nephroid is

$$4(x^2 + y^2 - a^2)^3 = 27a^4x^2, \tag{23}$$

[10] https://www.geogebra.org/m/yfq7wmqz
[11] https://www.geogebra.org/m/pyhk9qvr

where a is a positive parameter. The curve $\mathcal{N}$ can also be described by a parametric presentation:

$$\begin{cases} x = 2a\sin^3 t \\ y = a(1+\sin^2 t)\ \cos t \end{cases} \tag{24}$$

To construct a hyperbolism of $\mathcal{N}$ with respect to the origin and to a vertical line, we follow the same path as in previous sections. This begins with deriving from Equation (24) a rational parametrization. Applying the substitution in Equation (3), we obtain:

$$\begin{cases} x_0 = \frac{-u^6-9u^4+9u^2+1}{(u^2+1)^3} \\ y_0 = \frac{16au^3}{(u^2+1)^3} \end{cases} \tag{25}$$

The coordinates of P are thus

$$\begin{cases} x_P = b \\ y_P = \frac{(-u^6-9u^4+9u^2+1)b}{16au^3} \end{cases}$$

and the coordinates of M are

$$\begin{cases} x_M = \frac{16au^3}{u^2+1)^3} \\ y_M = \frac{(-u^6-9u^4+9u^2+1)b}{16au^3} \end{cases} \tag{26}$$

We run now the following Maple code, similar to what we did already:

```
p1 := x*denom(xM) - numer(xM);
p2 := y*denom(yM) - numer(yM);
J := <p1, p2>;
JE := EliminationIdeal(J, {a, b, x, y});
G := Generators(JE)[1];
evala(AFactor(G));
```

The last command is intended to check that the obtained polynomial $G(x,y)$ is irreducible. We have here a polynomial of degree 12, which fits Remark 3:

$$\begin{aligned} G(x,y) = 4a^6x^6y^6 - 12a^6b^2x^4y^4 + 12a^4b^2x^6y^4 + 12a^6b^4x^2y^2 - 24a^4b^4x^4y^2 \\ +12a^2b^4x^6y^2 - 4a^6b^6 - 15a^4b^6x^2 - 12a^2b^6x^4 + 4b^6x^6. \end{aligned} \tag{27}$$

Figure 12 shows the curve $\mathcal{N}$ for $a = 1$ and hyperbolisms with respect to the origin and the line whose equation is $x = b$ for $b = 1/2, 2, 3$.

6 A construction of a lemniscate

The construction proposed in this section is slightly different from a hyperbolism. Studying at least one case may be an appeal to explore more cases.

We consider a circle $\mathcal{C}$ whose center is the origin and radius r and a line D whose equation is $x = b$. Take a point M_0 on the circle $\mathcal{C}$; the point P is the intersection of D with the horizontal line through M_0. Then we define M to be the point of intersection of the line (OP) with the vertical line through M_0. The geometric locus of M when M_0 runs on $\mathcal{C}$ is a lemniscate. This is illustrated in

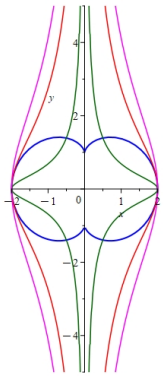


Figure 12: Three hyperbolisms of the same nephroid with respect to the origin

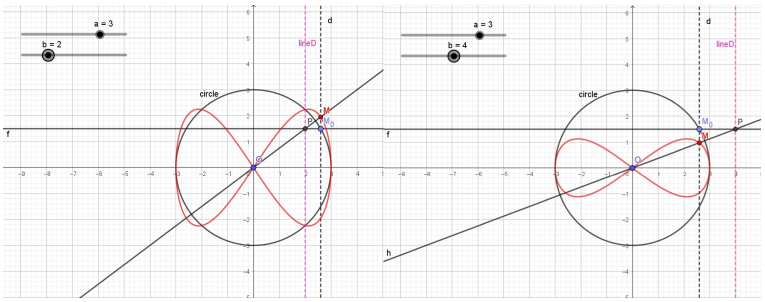


Figure 13: The circle is centered at the origin and the line intersects it.

Figure 13. The figure shows screenshots of a GeoGebra applet[12]; using the sliders may help to study the family of curves.

For $b < r$, the lemniscate lies part out of the circle. For $b > r$, the lemniscate is bounded by the circle. In both cases, the lemniscate is tangent to the circle at its vertices, i.e. its points on the $x-$axis different form the origin.

GeoGebra could not give an implicit equation for the lemniscate, we will look for such an equation via algebraic computations. We choose the same parametrization as above for the circle, i.e. the coordinates of M_0 are given by:

$$(x_0, y_0) = \left(a \cdot \frac{1-u^2}{1+u^2},\ a \cdot \frac{2u}{1+u^2} \right). \tag{28}$$

Then the coordinates of P are

$$(x_P, y_P) = \left(b, a \cdot \frac{2u}{1+u^2} \right).$$

The equation of the line (OP) is $y = \frac{2au}{b(1+u^2)} \cdot x$, whence the coordinates of M:

$$(x_M, y_M) = \left(a \cdot \frac{1-u^2}{1+u^2},\ \frac{2a^2u(1-u^2)}{b(1+u^2)^2} \right). \tag{29}$$

[12] https://www.geogebra.org/m/zkgmxume

We apply the following code:

```
p1 := x*denom(xM) - numer(xM);
p2 := y*denom(yM) - numer(yM);
J := <p1, p2>;
JE := EliminationIdeal(J, {a, b, x, y});
G := Generators(JE)[1];
```

and obtain the polynomial

$$G(x,y) = a^2x^4 - a^4x^2 + a^2b^2y^2. \tag{30}$$

The vanishing set of $G(x,y)$ in the real plane is called a *lemniscate of Gerono*.

7 Special features of the curves which have been obtained

In Section 2, the obtained hyperbolisms show common features:

1. The curve has two disjoint components. On the one hand, these components cannot be distinguished by algebraic means as the defining polynomial is irreducible over the field of real numbers. This can be proven using Maple command **evala(AFactor(...)**; the need for that command, and not the ordinary **Factor** command, has been analyzed in [13].

2. The geometric basis of the construction induces easily that the two components are symmetric about the $y-$axis.

3. From the exploration, the $y-$axis seems to be an asymptote to both. Actually, this is a consequence of the construction: when the point M_0 gets arbitrarily close to the $y-$axis, the slope of the line (OM_0), and thus the $y-$coordinate of M, tend to infinity.

Actually the origin does not belong to the hyperbolism of the original curve. It can be obtained only if M_0 is at the origin, but in this case the line (OM_0) is not defined.

8 Discussion

In this paper, we considered a question in plane geometry, which can look rather simple. Its translation using technology revealed a more complicated situation that foreseen. The 2 first examples may be treated by hand, with students having learnt an elementary course in the spirit of [2], but quickly the computations become heavier. The problem can be understood by a regular High-School student[13], but the algebraic machinery beneath is then out of reach, as it belongs to Computer Algebra, and uses algorithms from Ring Theory. Moreover, we deal here with curves of higher degree than what High-School students learn, as explained in section 1.

Along the paper, implicitization has been performed using polynomial rings and elimination. Maple has an **implicitize** command, but we preferred to have our computations more understandable, and not to use this command as a blackbox. The user has no access to the algorithms, but via

[13]Decades ago, such constructions were performed in High-School by hand using paper, pencil and ruler. Only in easy cases, an equation was derived.

reading the paper in reference on the help page of the command. On the one hand, an educator has to make a decision between low-level and high level commands (see [14]). On the other hand, what can be done depends on the students' theoretical background; sometimes it is necessary to use the CAS in order to bypass a lack of knowledge (see [8]).

Moreover, respective affordances of the two kinds of software used here have been analyzed in the past, and the importance of networking between them has been emphasized, expressing a wish to see development of tools for an automatic dialog between the kinds of software [25, 26, 12]. See also [31] for more recent developments with GeoGebra. Here the strengths of both kinds appeared sometimes different from what we were accustomed. Anyway, it is only natural that for different questions, the respective abilities and strengths of the different kinds of software which are used may vary.

Finally, the lemniscate of Gerono in Section 6 shows that classical constructions can be imitated with slight modifications and lead to software's broader experience. After all, curiosity is the main engine for exploration and discovery. The 4 C's of 21st Century Education [24] are well-known; they are Collaboration, Communication, Critical Thinking and Creativity. Nevertheless, a 5th C has to be added, namely Curiosity, without which the 2 last C's cannot appear. As in [10], we illustrate this here with the construction and the exploration of plane curves. Of course, this claim is true in much larger domains.

References

[1] Abánades, M., Botana, F., Montes, A. and Recio, T. (2014). *An algebraic taxonomy for locus computation in dynamic geometry*, Computer Aided Design **56**, 22–33. DOI: https://doi.org/10.1016/j.cad.2014.06.008

[2] Adams, W. and Loustaunau, P. (1994). *An Introduction to Gröbner Bases*, Graduate Studies in Mathematics 3, American Mathematical Society. DOI: https://doi.org/10.1090/gsm/003

[3] Botana, F. and Abánades, M., (2014). *Automatic Deduction in (Dynamic) Geometry: Loci Computation*, Computational Geometry **47** (1), 75-89. DOI: https://doi.org/10.1016/j.comgeo.2013.07.001

[4] Blazek, J. and Pech, P. (2017). *Searching for loci using GeoGebra*, International Journal for Technology in Mathematics Education **27**, 143–147. DOI: http://dx.doi.org/10.1564/tme_v24.3.06

[5] Bruce, J.W. and Giblin, P.J. (2012). *Curves and Singularities*, Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139172615

[6] Corless, R., Giesbrecht, M., Kotsireas, I. and Watt, S. (2001). *Numerical implicitization of parametric hypersurfaces with linear algebra*. AISC'2000 Proceedings, Madrid, Spain. LNAI 1930, Springer. DOI: https://doi.org/10.1007/3-540-4499

[7] Cox, D.Little, J. and O'Shea, D. (1992). *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, Undergradu-

ate Texts in Mathematics, NY: Springer, (1992). DOI: https://doi.org/10.1007/978-0-387-35651-8

[8] Dana-Picard, T. (2005). *Technology as a bypass for a lack of theoretical knowledge*, International Journal of Technology in Mathematics Education 11 (3), 101-109.

[9] Dana-Picard, T. (2020). *Safety zone in an entertainment park: Envelopes, offsets and a new construction of a Maltese Cross*, Electronic Proceedings of the Asian Conference on Technology in Mathematics (ACTM 2020), 31-50. https://atcm.mathandtech.org/EP2020/invited/21794.pdf

[10] Dana-Picard, T. (2024). *Networking between kinds of software: geometric loci and exploration of simple models of space curves*, Electronic Journal of Technology in Mathematics 18 (3). https://ejmt.mathandtech.org/Contents/eJMT_v18n3p1.pdf.

[11] Dana-Picard, T. and Kovács, Z. (2018). *Automated determination of isoptics with dynamic geometry*, Calculemus, CICM-11 (Conference on Intelligent Computer Mathematics), RISC (Hagenberg, Austria). DOI: https://doi.org/10.13140/RG.2.2.23002.24003

[12] Dana-Picard, T. and Kovács, Z. (2021). *Networking of technologies: a dialog between CAS and DGS*, The electronic Journal of Mathematics and Technology (eJMT) 15 (1), 43-59. https://ejmt.mathandtech.org/Contents/eJMT_v15n1p3.pdf

[13] Dana-Picard, T. and Recio, T. (2023). *Dynamic construction of a family of octic curves as geometric loci*, *AIMS Mathematics* 8 (8), 19461-19476. DOI: https://doi.org/10.3934/math.2023993

[14] Dana-Picard, T. and Steiner, J. (2004). *The importance of "low level" CAS commands in teaching Engineering Mathematics*, European Journal of Engineering Education 29 (1), 139 -146.

[15] Dana-Picard, T., Kovács, Z. and Yang, W.C. (2023). *Topology of Quartic Loci Resulted From Lines Passing through a Fixed Point and a Conic*, CGTA 2023 (Conference on Geometry: Theory and Applications), Kefermarkt, Austria. DOI: https://doi.org/10.13140/RG.2.2.24291.94249

[16] Dana-Picard, T. and Zehavi, N. (2016). *Revival of a classical topic in Differential Geometry: the exploration of envelopes in a computerized environment*, International Journal of Mathematical Education in Science and Technology 47(6), 938-959.

[17] Duval, R. (2018). *Understanding the Mathematical Way of Thinking – The Registers of Semiotic Representations*, Springer. DOI: http://dx.doi.org/10.1007/978-3-319-56910-9

[18] Ferrarello, D., Mammana, M.F., Pennisi, M. and Taranto, E. (2017). *Teaching Intriguing Geometric Loci with DGS*. In: Aldon, G., Hitt, F., Bazzini, L., Gellert, U. (eds) Mathematics and Technology. Advances in Mathematics Education. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-51380-5_26

[19] Ferréol, R. (2017). *Piriform Quartic*, https://mathcurve.com/courbes2d.gb/piriforme/piriforme.shtml (retrieved November 2024).

[20] Ferréol, R. (2017). *Hyperbolism and Antihyperbolism of a Curve - Newton Transformation*, https://mathcurve.com/courbes2d.gb/hyperbolisme/hyperbolisme.shtml (retrieved November 2024).

[21] Fischer, G. (2001). *Plane Algebraic Curves*, Student Mathematical Library 15, American Mathematical Society.

[22] Gao, X.S. and Chou, S.C. (1992). *Implicitization of rational parametric equations*, Journal of Symbolic Computation 14 (5), 459-470. DOI: https://doi.org/10.1016/0747-7171(92)90017-X.

[23] Kovács, Z.:, Recio, T. and Vélez, M.P. (2022). *Automated reasoning tools with GeoGebra: What are they? What are they good for?*, In: P. R. Richard, M. P. Vélez, S. van Vaerenbergh (eds): Mathematics Education in the Age of Artificial Intelligence: How Artificial Intelligence can serve mathematical human learning. Series: Mathematics Education in the Digital Era, Springer,23-44. DOI: https://doi.org/10.1007/978-3-030-86909-0_2

[24] National Research Council. (2010). *Exploring the Intersection of Science Education and 21st Century Skills: A Workshop Summary*. Margaret Hilton, Rapporteur. Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academy Press.

[25] Roanes-Lozano, E., E. Roanes-Macías, E. and Villar-Mena, M. (2003). *A bridge between dynamic geometry and computer algebra*, Math. Comput. Model. **37**, 1005–1028. DOI: https://doi.org/10.1016/S0895-7177(03)00115-8

[26] Roanes-Lozano, E., Van Labeke, N. and Roanes-Macías, E. (2010). *Connecting the 3D DGS Calques3D with the CAS Maple*, Mathematics and Computers in Simulation 80 (6), 1153–1176. DOI: https://doi.org/10.1016/j.matcom.2009.09.008

[27] Sederberg, T. Anderson, D.C and Goldman, Ronald (1984). *Implicit representation of parametric curves and surfaces*, Computer Vision, Graphics, and Image Processing 28, 72-84. DOI: https://doi.org/10.1016/0734-189X(84)90140-3

[28] Sendra, R. and Winkler, F. (2008). *Rational Algebraic Curves: A Computer Algebra Approach*, Algorithms and Computation in Mathematics 22, Springer.

[29] Taimina, D. (2007). *Historical Mechanisms for Drawing Curves*, in (Amy Shell-Gellasch, edt) *Hands on History: A Resource for Teaching Mathematics*, 89 - 104, Cambridge University Press. DOI: https://doi.org/10.5948/UPO9780883859766.011

[30] Wang, D. (2004). *A simple method for implicitizing rational curves and surfaces*, Journal of Symbolic Computation **38**, 899–914. DOI: https://doi.org/10.1016/j.jsc.2004.02.004

[31] Weinhandl, R., Lindenbauer, E., Schallert-Vallaster, S., Pirklbauer, J. and Hohenwarter, M. (2024). *GeoGebra, a Comprehensive Tool for Learning Mathematics*, in (A. Gegenfurtner and I. Kollar, edts) Designing Effective Digital Learning Environments, London: Routledge. DOI: https://doi.org/10.4324/9781003386131

[32] Yates, R. (1947). *A Handbook On Curves And Their Properties*, Ann Arbor: J. Edwards.

Networking between kinds of software: geometric loci and exploration of simple models of space curves *

Thierry (Noah) Dana-Picard
ndp@jct.ac.il
Jerusalem College of Technology
and Jerusalem Michlala College
Israel

Abstract

Geometric loci are an important topic in geometry. We explore constructs of plane curves and of space curves inspired by space trajectories (bicircular and tricircular motions). The exploration relies on joint usage of Dynamic Geometry Software (DGS) and of a Computer Algebra System (CAS), and involves automated methods to determine loci and sometimes dichotomy methods. The computer assisted work reveals some surprises, in particular the respective roles of the two kinds of software may be different from what previous works revealed. If previously the roles were quite distinct, DGS for exploration and numerical approach, and CAS for symbolic proof, here the CAS is needed for numerical methods also. The curves are described by trigonometric parametrization, and implicitization is performed using elimination, for 2D and 3D models. Finally, we discuss approaches to developments using models and elaborate on the importance to develop automatic dialog between kinds of software.

1 Introduction

1.1 A motivation: frequent presentation of transfer orbits of space probes.

A modern way to attract students to study mathematics consists in proposing activities around items from the daily news: for example, events related to the International Space Station (ISS), the almost simultaneous launching of three Mars-bound spacecrafts in February 2021 and their arrival 6 months later, the James Webb space telescope launched 10 months later, etc. Since then, all of these provide big titles. More recently, the race to the Moon, American Artemis lunar program, the Chinese mission to the far side, attract much attention. The general audience discovered that a spacecraft's trajectory from the Earth to Mars is not straight but curved (Figure 1).

*Supported by the CEMJ Chair at JCT

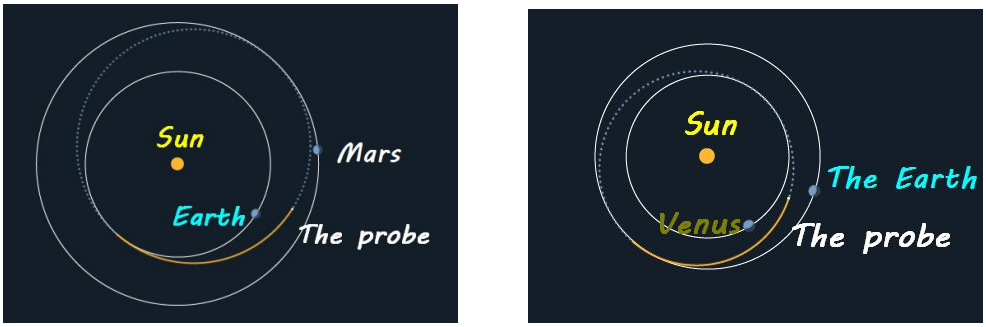


Figure 1: Transfer orbits between the Earth and another planet

Roughly speaking, each of these trajectories is the union of arcs of ellipses. The shape of the trajectory and the velocity of the spacecraft are ruled by the so-called Kepler laws (Karttunen et al., 2008). Kepler's 1st law reads that the trajectory of an object P orbiting an object S under the influence of gravitation only is an ellipse with S at one of the foci. Kepler's 2nd law explains how the velocity changes along the trajectory (see Figure 3): flying from C to D or from E to F takes the same time if, and only if, the curved triangles SCD and SEF have equal areas.

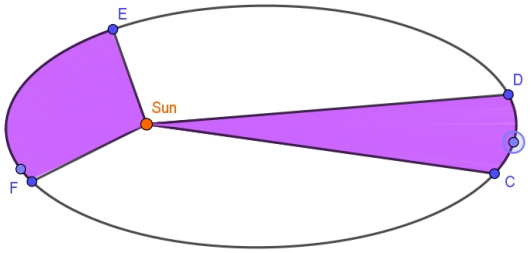


Figure 2: Kepler laws.

The Israeli Moon-bound probe Beresheet was planned to move along ellipses. At specific times, thrust has been exploited to increase eccentricity, keeping the Earth as a focus, until the ellipse englobes the Moon. Then the probe has been slowed down to be captured by the Moon's gravitational field. At that time the trajectory is an ellipse with the Moon at a focus. The probe orbits te Moon in the original direction. Figure 3 shows a diagram of the planned trajectory.

Artemis's trajectory seems more complicated (Figure 4). The diagram on the left shows the points were thrust is operated to change the geometric characteristics of the trajectory. After the engines stop, the motion is again ruled by Kepler laws, and the trajectory follows an arc of a new ellipse with the Sun at a focus. At some time, the probe is slowed down to be captured by the gravitation field of the Moon. From now on, the orbit is an ellipse with the Moon as a focus. The figure on the left seems static, but the figure on the right expresses the fact that actually the Moon moves on its orbit around the Earth, pulling the spacecraft with it. Simple models in 2D and 3D (taking into account the fact that the Sun also moves) have been displayed in [13]

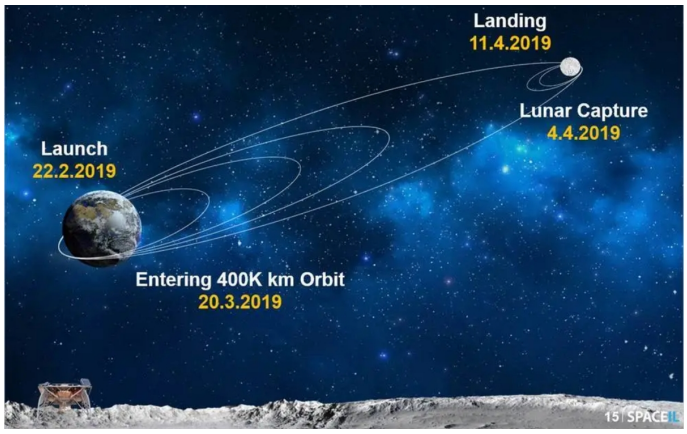
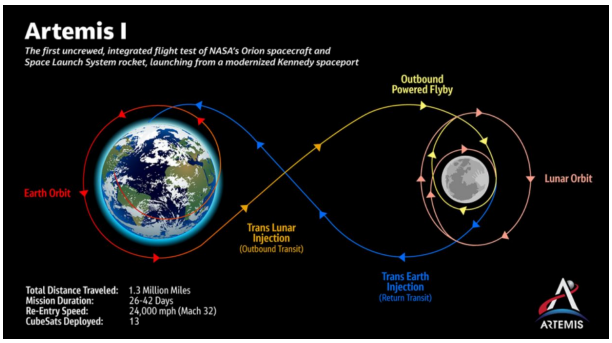


Figure 3: Beresheet's trajectory (Credit: NASA - SpaceIL)



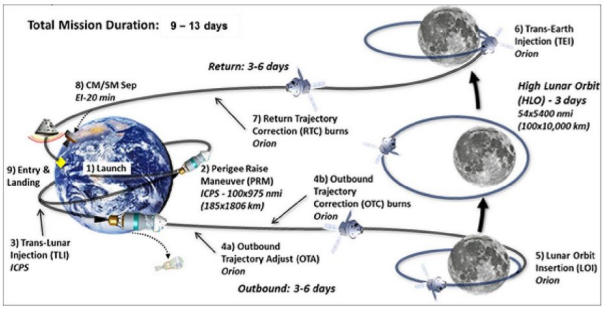


Figure 4: Diagrams of Artemis trajectory (Credit: NASA)

These two examples explain why we explore bicircular and tricircular motions, (a) when all the components are oriented in the same direction(Section 2) and (b) when one component is retrograde (Section 3).

All this is a little bit caricatural. In reality, more than one space object influence the trajectory, and the axis of a planet and/or a natural satellite may make some angle with the orbital plane. For example, the axis of the Earth makes an angle of $23^o27'$ with the ecliptic plane. So, the ecliptic plane and the equatorial plane of the moon may be different. This influences the geometric description of the trajectories. Moreover, for travelling from the Earth to the Moon, engineers have to take into account the gravitational influence of Mars also. As we are interested in simple models accessible to either upper-high school or to beginning undergraduates, we simplify all this.

According to Kepler laws, the planetary motion is well described by an equation in polar coordinates, namely

$$r = \frac{p}{1 + \varepsilon\theta},$$

where r is the distance from the planet to the Sun, ε is the eccentricity of the ellipse and θ is the angle to the planet's current position from its position closest to the Sun (called aphelion). Early undergraduates' literacy with polar coordinates is generally not so high. Therefore, we chose to work with parametric presentations for plane curves, which yielded the construction of, sometimes, classical curves such as epitrochoids, epicycloids, etc. The orbits of the planets of the Solar System are ellipses whose eccentricity is very close to 0, therefore for a 1st approximation, suitable for beginners, the orbits have been modeled by circles centered at the Sun, and velocity is constant, encoded in the angular velocity. Other simplifications were as follows:

- The mean distance Sun-Earth is equal to 1 (astronomical unit, A.U.), whence the definition of E at the beginning of Section 2.
- The period of the Earth on its orbit is 1 (same remark as above).
- The orbital data for another planet is given in proportion to these data. We caplin in next section why we decided not to use actual orbital data.

Actually, the only "true" orbital data we can be interested in consists in distance between Sun and planets, or between planets and their satellites, either natural or artificial. At the beginning of next section, we explain why this is irrelevant to the present work. The author teaches in an advanced course about satellites and Earth observation. There, the "true" shape of orbits, (whence variable velocity, etc.), inclination with respect to the equator, are considered. Anyway, the intuition acquired with the simple models presented here helps.

1.2 What are we doing, what are the educational goals?

The ratio between the distances Sun-Earth and Earth-Moon is close to 413. No hope to present on a screen the Sun and trajectories, neither of the Moon around the Earth, nor of an object orbiting the Moon. Therefore, we build the activities with arbitrary values of the parameters, not with genuine orbital data. Nevertheless, these choices provide constructions of interesting curves, from which the student can imagine what happens in space. Most of the examples in [13] are planar, even when taking the Sun's motion into account: the Sun, the planet and the satellite move in the same plane (in

Section 2). Animations and computations are performed with a Dynamic Geometry Software (DGS, here GeoGebra) and a Computer Algebra System (CAS, here Maple 2024), and networking between them. For one example, we compute an implicit equation for the trajectory, using Elimination. We provide Maple session's code.

In Section 4, we consider a model of a satellite moving around the planet on an orbit not included in the plane of the planet's orbit, here a polar orbit. In this last case, we begin working with parametric presentations for the orbits and implicitize the presentation. As for any space curve, an implicit presentation shows it as the intersection of two surfaces (i.e. we look for 2 polynomial equations). This provides a new opportunity to compare the affordances of plotting with a Dynamical Geometry Software (DGS) and with a Computer Algebra System (CAS), comparing parametric plots and implicit plots.

The activities that we propose are based on exploration and discovery in a technology-rich environment. The outcome is multi-faced: plots based on numerical data, search for exact (symbolic) expressions, implicitization, etc. Some of the activities with DGS provide some knowledge, but further exploration has to be performed with the CAS, among other issues, analyzing numerical solutions of a trigonometric equation (Section 3). The core of the search is based on thinking and computing in a way, which has been learnt by undergraduates in a 1st Calculus course.

2 Tricircular motion: a satellite on a direct orbit around the Moon in the SUN-Earth-Moon plane

We consider a moving point E given by $(x_E, y_E) = (\cos t, \sin t)$. An object M orbits E according to

$$(x_M, y_M) = (\cos t + r\cos(ht), \sin t + r\sin(ht)), \tag{1}$$

where r is the distance from E and h encodes the orbital period[1].

Now, we define an object Sat by the following formula:

$$(x_{Sat}, y_{Sat}) = (x_M, y_M) + r/2(cos(3hu), sin(3hu)). \tag{2}$$

The orbit of Sat looks more complicated in any case. A GeoGebra applet [S1] is available at https://www.geogebra.org/m/mnsdv3fz; the reader is invited to explore with the sliders. The orbits can be visualized (a) with the command **Locus** and (b) running an animation with **Trace On**. The locus of E is a circle and the orbit of M is an epicycloid[3]. Examples are the dotted curves in Figures 5,6 and 7.

Figure 5 shows 3 cases, for $r = 0.2$ and $h = 3, 5, 6, 8, 9$ (from left to right).

For for $r = 0.4$ and $h = 3, 4, 5, 6, 8$ (from left to right), we have Figure 6, and for $r = 0.8$ and $h = 3, 4, 5, 6, 9$ (from left to right), we have Figure 7:

At first glance, we conjecture that for odd h, both the $x-$axis and the $y-$axis are axes of symmetry, but for even h, only the $x-$axis is such. Of course, this requests a symbolic proof. Rotational symmetries are not very apparent, but may exist. Broader exploration can be made with the above

[1] For the sake of simplification, we consider here only integer values for h. For non integer values, determining the motion periodicity is harder. Such examples will be studied in a f[2]urther work.

[3] See https://mathcurve.com/courbes2d.gb/epicycloid/epicycloid.shtml.

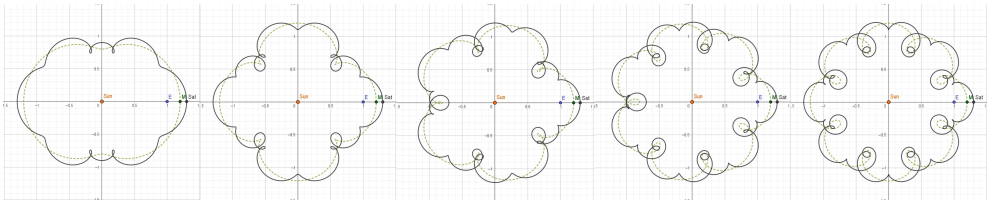


Figure 5: Three coplanar movements in the same direction - $r = 0.2$

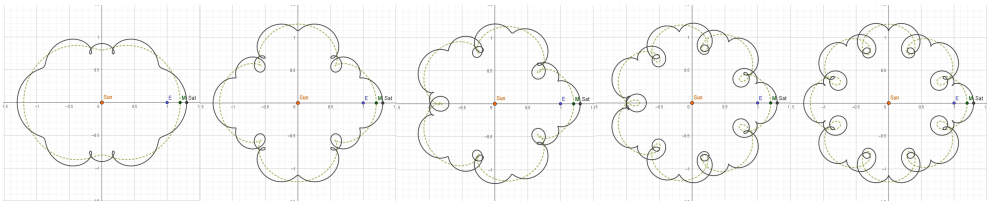


Figure 6: Three coplanar movements in the same direction - $r = 0.4$

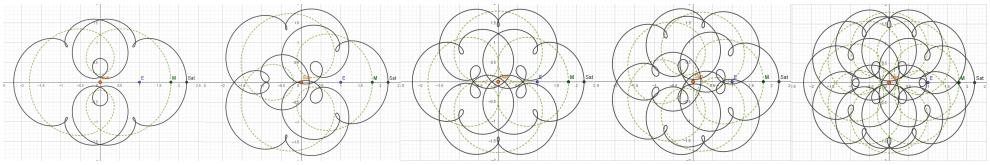


Figure 7: Three coplanar movements in the same direction - $r = 0.8$

mentioned applet.Note that both trajectories, M viewed from the Sun and Sat viewed from E, are epicycloids. The composition of the motions yield the complicated curves.

The comparison of two different points of views is interesting. On the basis of Tycho Brahe's observations, Kepler drew a sketch of the orbit of Mars viewed from the Earth [22]; the curve that he plotted (Figure 8) is very similar to curves obtained in other situations (see [7]). In an other direc-

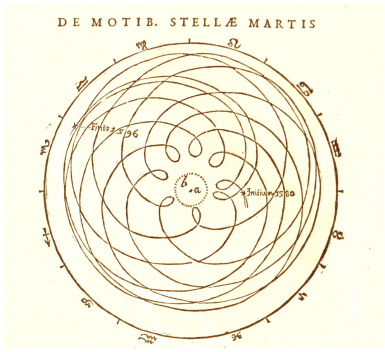


Figure 8: Kepler's sketch of Mars's orbit when viewed from the Earth

tion, the website https://mathcurve.com/courbes2d.gb/courbes2d.shtml shows constructions of epicycloids by means of a circle rolling on another circle. We can construct the complicated curves obtained in this section using a circle rolling on a circle rolling on a 3rd circle. This is beyond the scope of the present paper. By the time these rows are written, the author is waiting for the outcome of students' activities.

3 A satellite on a retrograde lunar orbit in the SUN-Earth-Moon plane

3.1 Equations and first examples

Here too, 3 different coplanar motions compose the satellite motion, when viewed from outside: the planet E, its moon M and the satellite, all of them described by trigonometric parametric presentation. The difference with the previous section is that here the satellite moves in retrograde direction. Recall Artemis trajectory, where the probe orbits the Moon in retrograde direction (Figure 4). Models for the composition of two direct motions and a retrograde 3rd motion have already been described in [13]. It provides non classical curves, as shown in Figure 10. The satellite moves according to the equation

$$(x, y, z) = (\cos u, \sin u) + r(\cos au, \sin au) + s(\sin bu, c\cos bu); \quad u \in [0, 2\pi]. \tag{3}$$

WLOG, we chose to work with coefficients 1/3 and 1/2, other choices could provide curves with similar properties. For the mathematical work, the given interval can be replaced by the whole of $\mathbb{R}$, then the parametrization determines infinitely many copies of the same geometric locus. But working with software requests a closed interval. With a suitable choice of the interval, and with the *increasing* option for animation with a slider, GeoGebra mimics infinitely many periods on the orbit.

The retrograde motion of the 3rd object is encoded by the last term in the sum. A GeoGebra applet [S2] is available at https://www.geogebra.org/m/p7uaewbc.

For $a = 1$, the motion is actually bicircular and the curve is a hypotrochoid[4].Examples with $r = 1/3$ and $s = 1/2$ are displayed in Figure 9, for $b = 3, 4, 5, 6, 9$.

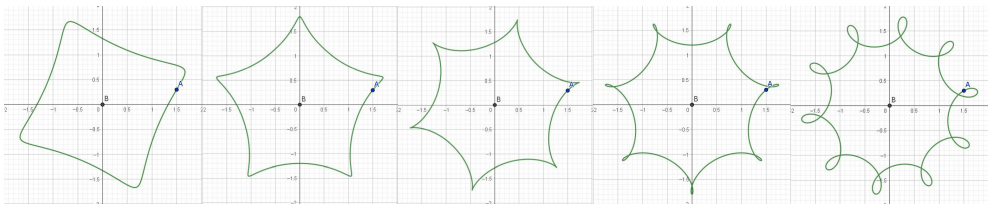


Figure 9: Three movements, one in retrograde direction - classical curves

The reader is invited to play with the 5 parameters involved and explore the existence of self-intersections and of cusps. In [13], we show various examples, including the so-called Mystery Curves, and apply them for creating Math Art and 3D printed objects[5]. Another point of view on the Mystery Curves can be found in [4]. Figure 10 displays the output for $r = 1/3$, $s = 1/2$ and $(a, b) = (4, 8), (6, 14), (8, 14), (8, 17)$ from left to right. The two curves on the right show also a partial view on how the animation is displayed.

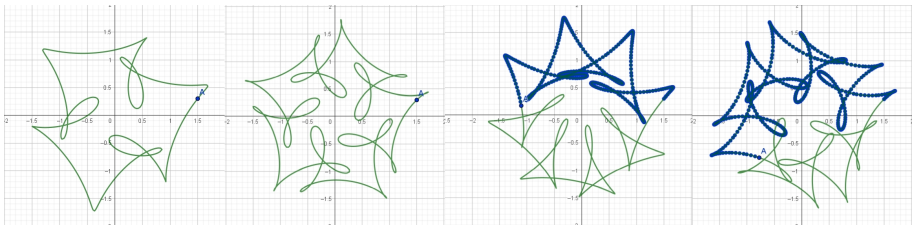


Figure 10: Three movements, one in retrograde direction

3.2 Exploration of the shape of the curves

Using the different sliders to make the values of the parameters vary, provides activities with different kinds of focus, according to teacher's and students' creativity. One of them is to study the symmetries of the curves. Rotational symmetries can appear, but the greater the parameters a and b, the harder to understand are the curves and rotational symmetries may remain unidentified. Generally the curves present one or two axial symmetries, but not more. Rotational symmetries of order 5 appear for other choices of the parameters, e.g., for $(a, b) = (6, 9)$. All this is purely experimental, but in this last case, computations with a CAS may be possible. Note that the explorations described here may provide

[4]See the Mathcurve page at https://mathcurve.com/courbes2d.gb/hypotrochoid/hypotrochoid.shtml.

[5]Animations can be found in [14].

conviction, but it is not a proof (for detailed analysis of the difference, and of what should be done, refer to the book [19]).

Let us study one example. Figure 12 corresponds to $(r, s) = (0.5, 0.3)$ $(a, b) = 11, 4)$. The curve seems to have a rotational symmetry of order 5. This can be explored in various ways. One of them can be to prove that $u \in \mathbb{R}$, $Sat(u + 2\pi/5) = Sat(u)$, but this may be unilluminating, actually only numerical, as there is no simple closed formula for these sines and cosines.

3.3 A 1st exploration - symmetry

On the basis of the conjecture that there is a 5-fold rotational symmetry, we consider one arc of the curve, given by

$$(x, y, z) = (\cos u, \sin u) + \frac{1}{3}(\cos au, \sin au) + \frac{1}{2}(\sin bu, c\cos bu); \quad u \in \left[0, \frac{2\pi}{5}\right]. \tag{4}$$

The corresponding arc can be plotted with GeoGebra using the parametric presentation and the command **Locus(Point creating the locus, slider)**. Then the arc should be rotated about the origin by an angle of $2\pi/5$. As the locus appearing on the screen is not recognized by the software as a geometric construct, the automatic command for rotation does not work. The point A defined by Equation (4) has to be rotated, using GeoGebra's button for rotations, and is automatically identified as depending on the parameter (the slider) u. Now the **Locus** command works. The process has to be iterated 4 times. The arcs are shown in Figure 11 with different colors and dots. A GeoGebra applet [S5] is

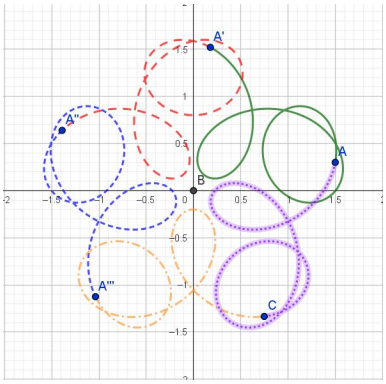


Figure 11: Exploration of the 5-fold symmetry of one of the curves

available at https://www.geogebra.org/classic/vhdkpfgk.

This process can be adapted to every case where a rotational symmetry seems to appear. Exploration with variable parameters (a, b) provides numerous examples with rotational symmetries of exotic order (we mean not the simple cases generally shown in class, but order 7,11, 13 etc.). Extension of such constructs may provide interesting mathematical art, as in [7, 13].

3.4 A 2nd exploration - circumcircle

Here, the exploration is visual only. The Locus command provides a plot based on numerical data, and the locus on the screen is not a geometrical construct. Therefore, the button and the intersection

commands do not work here. Until now, we do not have a symbolic equation either. As the coordinates are sums of sines and cosines, they are bounded. Is it possible top find a circumcircle? The situation is illustrated in Figure 12.

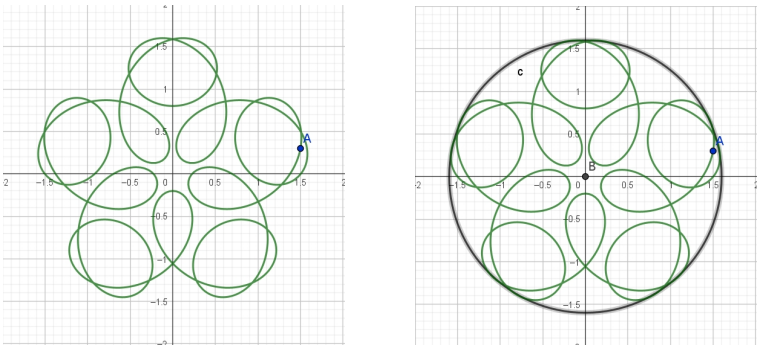


Figure 12: The curve has a rotational symmetry of order 5

The number of points of intersection of the curve with a circle centered at the origin can give an interesting indication. The exploration yields a radius approximatively equal to 1.6 fro a circumcircle. In order to check this, it is necessary to solve the following equation

$$x(t)^2 + y(t)^2 = 1.6^2 \tag{5}$$

for the unknown t. Hand-made computations are unilluminating, whence the importance to use a CAS. Even with technology this is not trivial. After more than one hour on a recent computer with a strong CPU, Maple did not provide an answer. Further attempts have been performed for $(a, b) = (6, 9)$. Experimentation with GeoGebra suggests that the curve is circumscribed by a circle centered at the origin with a radius R between 1.7 and 1.8. Further experimentation has been made with Maple, solving equations similar to Eq. (5) for variable radius. For each tested value of the radius, Maple gave an immediate answer.

Experimentation provide the following output:

- For $1.7 \leq R \leq 1.73$, Equation (5) has 10 real solutions and numerous complex non real solutions. This corresponds to 10 points of intersection of the circle with the curve $\mathcal{C}$. The GeoGebra plot reveals that the circle is not a circumcircle to $\mathcal{C}$.
- For $R \geq 1.74$, Equation (5) has only non real solutions. The circle does not intersect the curve $\mathcal{C}$. A approximation of the radius can then be found numerically by a dichotomy method (similar to the dichotomy method learnt by undergraduates in a 1st Calculus course[6])

A similar process can be applied with GeoGebra. The radius of the circle has to be defined by a slider, and the increment of the slider can be defined as smaller than the default one. The precision of the work depends here on te user's hand.

[6]This is a good opportunity to enhance the fact that learning ways of thinking is as important as learning new notions and theorems.

4 A satellite on a polar lunar orbit

4.1 With "fixed Sun"

Because of the Earth revolution about its axis, a satellite orbiting the Earth on a polar orbit may provide imaging of the entire Earth surface, whence its importance for Earth observation. Details are to be found in [5]; an online simulator is available at https://observablehq.com/@jake-low/satellite-ground-track-visualizer. Here, for visualization and animations, the model requests usage of a software for 3D geometry. We use GeoGebra's 3D Graphics package. The Sun is placed, as usual, at the origin. A "planet" E is defined by the following equations (the coefficient 2 has be introduced in order to have "reasonable" proportions on the plot):

$$\begin{cases} x_E(t) = 2\cos t \\ y_E(t) = 2\sin t \\ z_E(t) = 0 \end{cases} . \tag{6}$$

and a moon M in the plane of the orbit of E by

$$\begin{cases} x_M(t) = 2\cos t + 0.4\cos at \\ y_M(t) = 2\sin t + 0.4\sin at \\ z_M(t) = 0 \end{cases} , \tag{7}$$

where a is a positive parameter. The GeoGebra commands are as follows:

```
E=2(cos(u),sin(u),0)
M=E+0.4 (cos (au),sin(au),0)
Sat=M+0.4(cos(bu),sin(bu),0)
```

Sliders for a and for b have to be defined; here, we did it for integer values (otherwise, the periodicity issue becomes hard to deal with - we intend to addres this issue in a further work). The coefficients 0.4 and 2 have been chosen in order to have a readable plot. We mentioned already that the ratio between true distances between Sun, Earth, Moon, and artificial satellites in the Solar System do not enable to display faithful models on a computer screen.

Now we define a satellite Sat around M orbiting in a plane perpendicular to the planes of the orbits of E and M.

$$\begin{cases} x_{Sat}(t) = 2\cos t + 0.4\cos at \\ y_{Sat}(t) = 2\sin t + 0.4\sin at \\ z_{Sat}(t) = 0.4\cos bt \end{cases} . \tag{8}$$

Figure 13 shows the curves obtained for different values of the parameters. These are snapshots of a GeoGebra applet [S3], available at https://www.geogebra.org/m/guwguzgk. On the left are displayed the orbits of E and M, on the right the 3D situation is shown, including the orbit of Sat. It is clear that the projection of the orbit of Sat on the $x, y-$plane is the orbit of M. This one is a epicycloid, a well-know curve (see https://mathcurve.com/courbes2d/epicycloid/epicycloid.shtml). For example, $(a, b) = (3, 2)$ gives a nephroid.

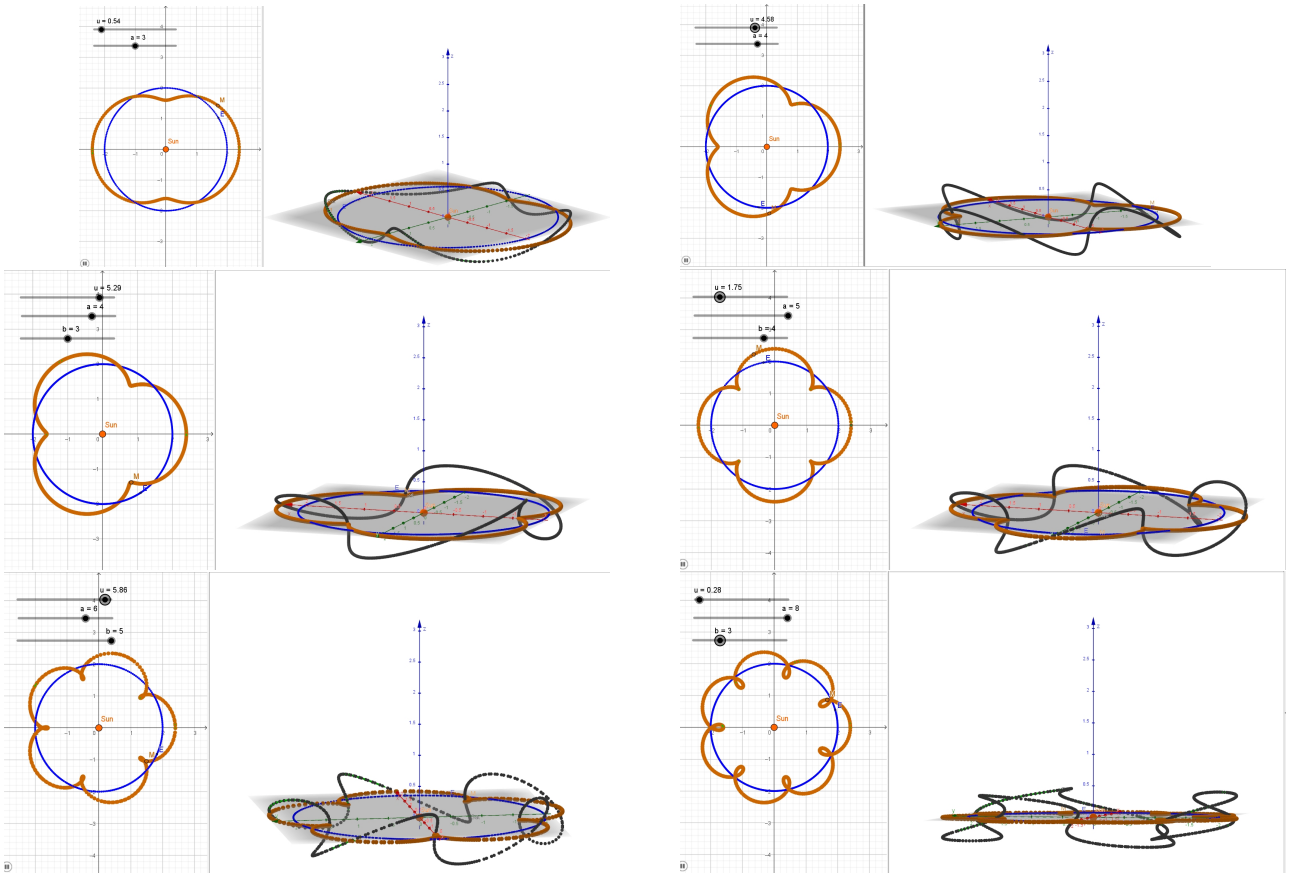


Figure 13: Polar orbits around a moon

4.2 With ”moving Sun”

For the sake of simplicity, we consider a ”Sun” moving according to the parametric presentation $(x, y, z) = (0, 0, u);\ u \in \mathbb{R}$. The only modifications to the commands are:

```
Sun=(0,0,u)
E=Sun+2(cos(u),sin(u),0)
```

The other points are changed automatically. Figure 14 shows two cases, corresponding to $(a, b) = (3, 2)$ on the left and to $(a, b) = (5, 6)$ on the right. The plane containing the orbits of E and M has been made apparent, and moves according to the Sun.

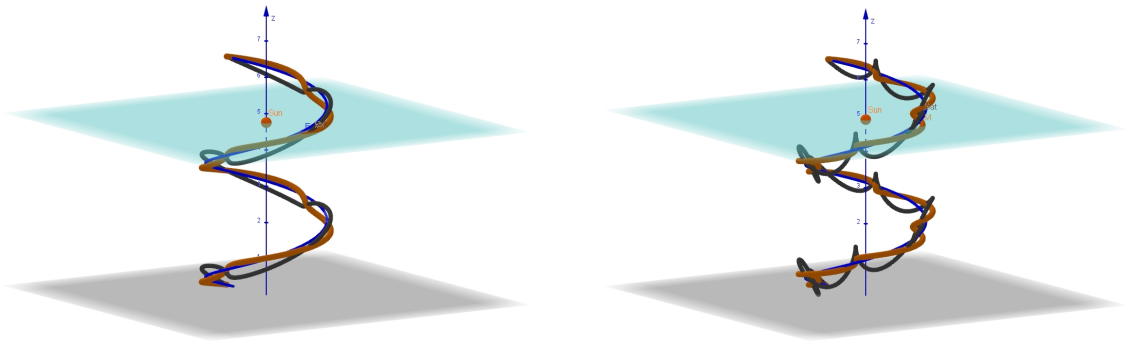


Figure 14: Polar orbits around a moon - a dynamic view with moving Sun

5 Implicitization

5.1 The process and the implicit equation – a planar situation

All the situations and activities presented above are based on trigonometric parametric equations for the different curves. For graphics, these provide accurate plots, as explained in [31]. For other issues, it may be better to have implicit equations. Therefore, books such as [30, 25, 29] and online encyclopedias of curves display both parametric and implicit equations (also polar equations), when possible. Implicitization is not an easy task, and requires heavy algebraic machinery. For our trigonometric parametric equations, sines and cosines have to be substituted by rational expressions, then polynomials can be defined. They generate ideals, for which algorithms from Gröbner bases theory and Elimination are generally used. For example, this has been done extensively in [6].

The situation given here is more complicated, as we deal with tricircular motion. As we saw in Equations (8), sines and cosine appear with 3 different angular velocities, which makes the substitution more intricate. We show here only one example.

The curve presented in Figure 12 has parametric equations

$$(x, y) = (\cos(u) + 0.5\cos(11u) + 0.3\sin(4u), \sin(u) + 0.5\sin(11u) + 0.3\cos(4u)) \qquad (9)$$

It is well-known that for every $u \in \mathbb{R}$, it exists a $t \in \mathbb{R}$ such that

$$\begin{cases} \cos u = \frac{1-t^2}{1+t^2} \\ \sin u = \frac{2t}{1+t^2} \end{cases} . \tag{10}$$

On the one hand, using De Moivre formula, we have:

$$(\cos u + i\ \sin u)^4 = \cos 4u + i\ \sin 4u.$$

On the other hand, Newton binomial yields:

$$(\cos u + i\ \sin u)^4 = \cos^4 u + 4i\ \cos^3 u \sin u - 6\cos^2 u \sin^2 u - 4i \cos u \sin^3 u + \sin^4 u.$$

Comparing real parts and imaginary parts in bote equations, we obtain:

$$\begin{cases} \cos 4u = \cos^4 u - 6\cos^2 u \sin^2 u + \sin^4 u \\ \sin 4u = 4\cos^3 u \sin u - 4\cos u \sin^3 u \end{cases} . \tag{11}$$

Substituting here Equations (10) and simplifying, we obtain:

$$\begin{cases} \cos 4u = \frac{t^8-28t^6+70t^4-28t^2+1}{t^2+1)^4} \\ \sin 4u = \frac{-8t^7+56t^5-56t^3+8t}{(t^2+1)^4} \end{cases} \tag{12}$$

By the same method we obtain:

$$\begin{cases} \cos 11u = -11\cos u \sin^{10} u + 165\cos^3 u \sin^8 u - 462\cos^5 u \sin^6 u + 330\cos^7 u \sin^4 u \\ \qquad -55\cos^9 u \sin^2 u + \cos^{11} u \\ \sin 11u = -\sin^{11} u + 55\cos^2 u \sin^9 u - 330\cos^4 u \sin^7 u + 462\cos^6 u \sin^5 u \\ \qquad -165\cos^8 u \sin^3 u + 11\cos^{10} u \sin u \end{cases} \tag{13}$$

and finally[7]

$$\begin{cases} (t^2+1)^{11}\cos 11u = -t^{22} + 231t^{20} - 7315t^{18} + 74613t^{16} - 319770t^{14} \\ \qquad +646646t^{12} - 646646t^{10} + 319770t^8 - 74613t^6 + 7315t^4 - 231t^2 + 1 \\ (t^2+1)^{11}\sin 11u = 22t^{21} - 1540t^{19} + 26334t^{17} - 170544t^{15} + 497420t^{13} \\ \qquad -705432t^{11} + 497420t^9 - 170544t^7 + 26334t^5 - 1540t^3 + 22t \end{cases} \tag{14}$$

Now we substitute into Equations (9):

$$\begin{cases} 10(t^2+1)^{11}x = -15t^{22} - 24t^{21} + 1065t^{20} - 36925t^{18} + 504t^{17} + 372315t^{16} + 1536t^{15} \\ \qquad -1599750t^{14} + 1680t^{13}3232810t^{12} - 3232810t^{10} - 1680t^9 + 1599750t^8 - 1536t^7 \\ \qquad -372315t^6 - 504t^5 + 36925t^4 - 1065t^2 + 24t + 15 \\ 10(t^2+1)^{11}y = 3t^{22} + 130t^{21} - 63t^{20} - 7500t^{19} - 315t^{18} + 132570t^{17} - 273t^{16} \\ \qquad -850320t^{15} + 990t^{14} + 2491300t^{13} \qquad 2730t^{12} - 3522120t^{11} \\ \qquad +2730t^{10} + 2491300t^9 + 990t^8 - 850320t^7 - 273t^6 + 132570t^5 - 315t^4 - 7500t^3 - 63t^2 \\ \qquad 130t + 3 \end{cases} \tag{15}$$

[7]The equations are presented as they are for technical reasons.

Denote by $P_1(x, y, t)$ (resp. $P_2(x, y, t)$) the polynomial obtained by subtracting right hand side of 1st (resp. 2nd) equation from its left hand side. These polynomials generate an idela $I =< P_1, P_2 >$ in $\mathbb{R}[x, y, t]$. By elimination we obtain (quickly) one polynomial in x, y of degree 22. It is too long for being written here, we obtained it within a Maple session using the *PolynomialIdeals* package.

Here is the Maple code [S6] for the session. Note that the coefficients 0.5 and 0.3 have been entered as 1/2 and 3/10 respectively, as a priori the algorithms in *PolynomialIdeals* do not like floating point.

```
restart; with(PolynomialIdeals);
trig4 := (cos(u) + sin(u)*I)^4;
cos4 := evalc(Re(trig4));
sin4 := evalc(Im(trig4));
# rational expressions for cosine and sine
cos4rat := subs(cos(u) = (-t^2 + 1)/(t^2 + 1), subs(sin(u) = 2*t/(t^2 + 1), cos4));
cos4rat := simplify(cos4rat);
sin4rat := subs(cos(u) = (-t^2 + 1)/(t^2 + 1), subs(sin(u) = 2*t/(t^2 + 1), sin4));
sin4rat := simplify(sin4rat);
trig11 := (cos(u) + sin(u)*I)^11;
cos11 := evalc(Re(trig11));
sin11 := evalc(Im(trig11));
cos11rat := simplify(subs(cos(u) = (-t^2 + 1)/(t^2 + 1),
             subs(sin(u) = 2*t/(t^2 + 1), cos11)));
sin11rat := simplify(subs(cos(u) = (-t^2 + 1)/(t^2 + 1),
             subs(sin(u) = 2*t/(t^2 + 1), sin11)));
expand(numer(sin11rat));
# the output was long - that is a way to have a shorter expression
xsat := (-t^2 + 1)/(t^2 + 1) + (1/2)*cos11rat + (3/10)*sin4rat;
xsat := simplify(xsat);
ysat := 2*t/(t^2 + 1) + (1/2)*sin11rat + (3/10)*cos4rat;
ysat := expand(ysat);
ysat := simplify(ysat);
# definition of polynomials - pay attention to capital and lowcase letters
p1 := 10*(t^2 + 1)^11*xsat;
P1 := 10*(t^2 + 1)^11*x - p1;
p2 := 10*(t^2 + 1)^11*ysat;
P2 := 10*(t^2 + 1)^11*y - p2;
K := <P1, P2>; # definition of an ideal
KE := EliminationIdeal(K, {x, y});
Gen := Generators(KE);
Gen[1];
Gen[2];
evala(Factors(Gen[1]));
```

All the commands give an almost immediate answer, elimination requests a few seconds. The command Gen[2] returns an error message; it verifies that the ideal KE is generated by one polynomial. This had to be checked, as the output is dispatched on about 30 rows. The last command checks that this polynomial is irreducible.

5.2 A non planar situation

Here too, we illustrate the process with an example. We consider the first case shown in Figure 13. We have:

$$(x, y, z) = \left(2\cos t + \frac{2}{5}\cos 3t, 2\sin t + \frac{2}{5}\sin 3t, \frac{2}{5}\cos 2t\right). \tag{16}$$

With the same method as above, based on Equations (10), and obtain the following rational parametrization for the curve:

$$\begin{cases} x = \frac{4(-3t^6+5t^4-5t^2+3)}{5(t^2+1)^3} \\ y = \frac{32t(t^4+1)}{5(t^2+1)^3} \\ z = \frac{2(t^4-6t^2+1)}{5(t^2+1)^2} \end{cases} \tag{17}$$

We transform equations (17) into polynomials

$$\begin{aligned} P_1(x, y, z) &= 5(t^2+1)^3 x + 12t^6 - 20t^4 + 20t^2 - 12 \\ P_2(x, y, z) &= 5(t^2+1)^3 y - 32t(t^4+1) \\ P_3(x, y, z) &= 5z(t^2+1)^2 - 2t^4 + 12t^2 - 2 \end{aligned}$$

which generate an ideal $J =< P_1, P_2, P_3 >$. After elimination of the parameter t, we obtain an elimination ideal JE generated by 2 polynomials $F_1(x, y, z)$ and $F_2(x, y, z)$, such that:

$$\begin{aligned} F_1(x, y, z) &= 25x^2 + 25y^2 - 100z - 104 \\ F_2(x, y, z) &= 125z^3 + 25y^2 + 250z^2 + 60z - 72 \end{aligned}$$

It was expected to find 2 polynomials, as a space curve is defined as the intersection of two surfaces, each one determined by an equation. The 2 surfaces are displayed in Figure 15. Note that

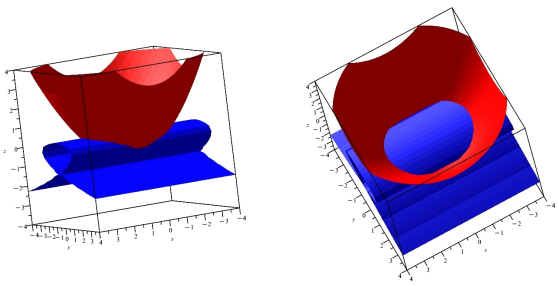


Figure 15: The trajectory for $(a, b) = (3, 2)$ with implicit equations

the intersection curve is difficult to identify. To have a good idea, the plot has to be rotated on the screen using the mouse, but even so the visual impression is not perfect. Part of the problem is the choice of the mesh to visualize surfaces in 3D space (see [31]). The plot can be enhanced if adding the parametric plot of the space curve. Figure 16 shows the respective plots obtained with the rational parametrization (on the left) and the trigonometric parametrization (on the right). The plots have been obtained using the following commands:

(a) Rational parametrization (using the same notations as above):

```
cratspace := spacecurve([xsatspace, ysatspace, zsatspace], t = -15 .. 15,
              numpoints = 5000, thickness = 3, color = black);
```

(b) Trigonometric parametrization (one period is enough):

```
ctrigplot := spacecurve([2*cos(t) + 2/5*cos(3*t),
                    2*sin(t) + 2/5*sin(3*t), 2/5*cos(2*t)],
                t = 0 .. 2*Pi, color = black, thickness = 7)
```

Note that the trigonometric parametrization gives a complete plot, but something is missing when using a rational parametrization. Filling the gap is not always possible because of issues related to limits at infinity of the involved rational functions.

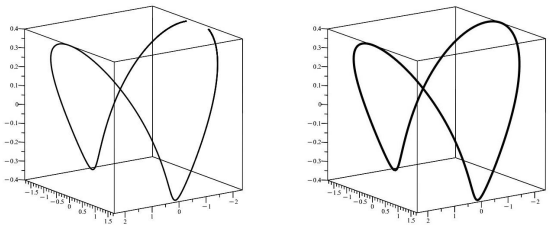


Figure 16: Plots with two parameterizations of the same curve

Figure 17 shows the surface and enhances the intersection curve.

6 Conclusions

6.1 Networking between technologies

The exploration and determination of geometric loci is an important topic in geometry. A few examples are provided by [1, 3, 2]. Our study relies strongly on the usage of two kind of software: a Dynamic Geometry Software (DGS), here GeoGebra, and a Computer Algebra System (CAS), here Maple 2024. The DGS enabled to explore the curves dynamically. It is natural to represent the dynamics of mobile points using a slider, and the trajectories are displayed dynamically (using **Trace On**, without this the points move but the user's eyes do not keep a lasting impression). Moreover, the automated command **Locus** provides complete pictures of the trajectories. Joint usage of both helps to have a graphical understanding of the phenomenon.

Nevertheless, for some parts of the study, the abilities of the DGS were not enough. Actually, it is well known that for algebraic symbolic computations, a stronger software is needed. Therefore we used Maple. In this study, we did it at two places:

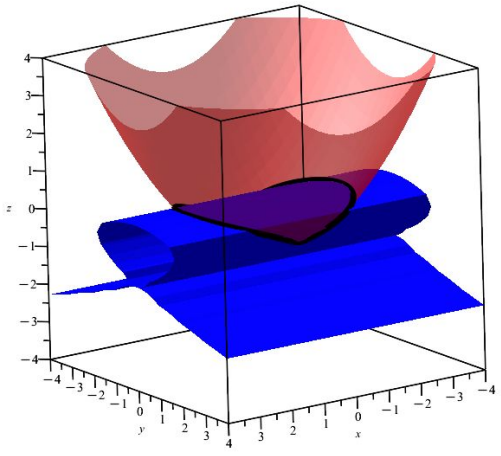


Figure 17: Visualization of the curve as intersection of two surfaces

- Implicitization involves a large amount of computations (generally transparent to the user, until the output is displayed). We have a detailed example in subsection 5.1. This is not a new phenomenon, for example it appears already in [6, 10, 11]. The computation of the elimination ideal computes actually a projection. For topological reasons (work is done according to Zariski topology), it is important to check whether the obtained polynomial is irreducible or not. It happens that irrelevant components appear, but this is not the case.

- In subsection 3.2, the exploration process is similar the dichotomy method taught to undergraduates for solving equations numerically. The DGS is mostly based on numerical methods, whence the need to add other algorithms which can check the results symbolically. This is done in the GeoGebra-Discovery companion, but not for the study here, even if we tried a more geometric construction of the mobile point Sat. Here we have an originality: generally switching from DGS to CAS is motivated by the need for symbolic computations, but here one of the switches was motivated by the need to make a numerical exploration, which could not be done totally with the other software..

In any case, we have here a new example of the importance of networking between different kinds of software. Even when both kinds provide similar features, they may have different affordances. Duval [16, 17]) explained that mathematical objects are abstract and can be approached and understood only using different registers of representation. Traditionally, the identified registers are numerical, symbolic, graphical, etc. The activities described here show that even within the same register, but with different tools, the object is presented in a different way.

An important feature of a DGS is interactivity: the possibility to drag points and the availability of sliders are central. They allow animations, with possible modifications in real time. Animations are also available with a CAS, but they require a priori programming, and the possibility to interact

with the animation in real time is limited.

For a long time, wishes to develop automated networking between CAS and DGS have been expressed clearly [27, 6, 12]. In the past, it existed for certain questions between Maple and MatLab, but not on a very large scale. It is possible to export the outputs in a different language: Maple allows to export the session to rtf or LaTeX format; this is useful, but not for the sake of computations and explorations. It allows export to STL format; this is important for exploration, as this is the language used by certain 3d printers [13]. This has also limitations, and for certain problems, we had to switch to another language, without automated export [15]. Advances have been recently announced with GeoGebra-Discovery [24].

Generally, a DGS and a CAS provide different registers of representation (in the sense of Duval [16, 17]): numerical (this was almost all what was available a few decades ago), graphical, symbolic, etc. A central issue was switching between registers, most CAS enabling to switch from symbolic to graphical (generally via the numerical, but this may be transparent), sometimes in reversed direction, from graphical to numerical (in DGS), more rarely from graphical to symbolic.

6.2 Modeling and Creativity

The 4 C's of 21st Century Education are already well-known and documented, as defined in OECD documents [26, 28]: Collaboration, Communication, Critical Thinking and Creativity. The present paper is anew contribution to we emphasize two of theses C's: Communication and Creativity. As discussed in the previous subsection, Communication is understood here between machines, the man-and-machine communication being minor, and between humans not discussed at all. As a follow-up of [7] (where we explained being on the verge of mathematical art), we have here a new contribution to explore Creativity, on the basis of simple models of trajectories in space.

We enhanced the exploratory aspect of the work, which has been enabled by the sliders of the applets. It is clear that exploration request, first of all, the 5th C, namely Curiosity[8].

Actually, the initial motivation was to describe/model trajectories in space (Section 1.1). Because of the complexity of space reality, we had to switch to simple models. Generally, models are intended to provide tools for understanding the concrete situation [20]. Here we went to another direction, as "true modeling" is out of reach of the visualization abilities of our computers. Choosing arbitrary parameters may provide some understanding of concrete situations, but it leads to create exotic curves beyond what is usually presented in courses. The regular usage of modelling appears in the left part of Figure 18 (Source: [8]), we went more towards the right part.

Finally, we wish to mention that Duval [16, 17] explained that in opposition to other scientific domains, where the objects of study are present on the table or graspable, mathematical objects are abstract and can be studied only via representations, that he classifies into registers. Switching between registers enable to master the object. We did it between numerical, symbolic and graphical registers. In [13], a new register is proposed, namely 3D printed objects, which are really graspable. Their pros and cons are discussed there, such as the necessity to thicken the curve to have them 3d printed. This register is beyond the scope of the present paper.

Interactive exploration is intended to educate students to be active and more involved in their learning process than with the traditional segment definition-theorem-proof. The teacher is a mentor

[8] As we mentioned initially that we have been inspired by news from the space exploration, we cannot omit to recall that one of the NASA robots on Mars is called Curiosity.

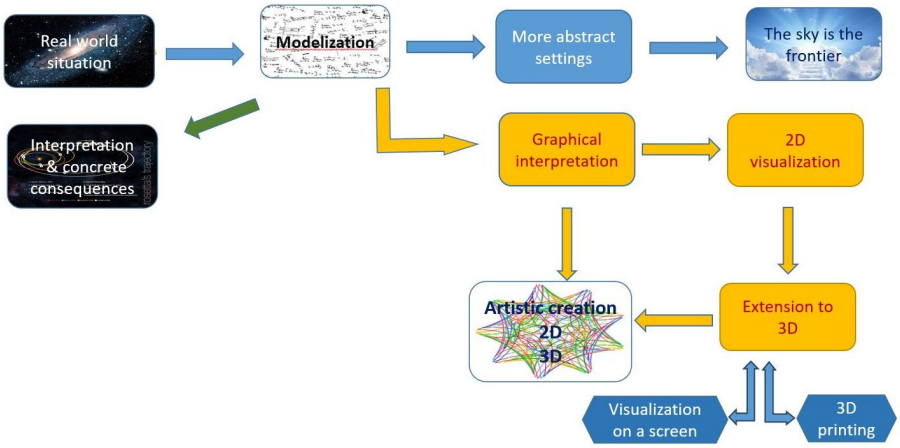


Figure 18: Modelling and creativity

and is here to facilitate the exploration, to help to understand, not to convey knowledge ex-cathedra. Nowadays, starting from real examples, taken from everyday life, from the students' culture, or from the news, not only motivates them to learn here and now, but also may improve their preparation for professional life and long life learning.

7 Supplementary Electronic Material

[S1] A GeoGebra applet for tricircular motion, in Section 2.

[S2] A GeoGebra applet for tricircular motion, one object moving in retrograde direction, in section 3.

[S3] A GeoGebra applet for a lunar polar orbit, in subsection 4.1.

[S4] A GeoGebra applet for a lunar polar orbit, in section 5.2

[S5] A GeoGebra applet to check 5-fold symmetry of a trajectory, in section 3.3.

[S6] A Maple worksheet for implicitization, in subsection 5.1.

References

[1] Abánades M., Botana, F., Montes, A. and Recio, T. (2014). *An algebraic taxonomy for locus computation in dynamic geometry*, Computer-Aided Design **56**, 22-33.

[2] Blazek, J. and Pech, P. (2017). *Searching for loci using GeoGebra*, International Journal for Technology in Mathematics Education **27**, 143–147.

[3] Botana, F. and Abánades, M. (2014). M. *Automatic Deduction in (Dynamic) Geometry: Loci Computation*, Computational Geometry **47 (1)**, 75-89.

[4] Campuzano, J. (2024). *Tracing Closed Curves with Epicycles: A Fun Application of the Discrete Fourier Transform*, North American GeoGebra Journal 11(1), 1-14. https://mathed.miamioh.edu/index.php/ggbj/article/view/197

[5] Capderou M. (2014) Handbook of Satellite Orbits: From Kepler to GPS, Springer Cham: Heidelberg.

[6] Dana-Picard, Th. (2020). *Safety zone in an entertainment park: Envelopes, offsets and a new construction of a Maltese Cross*, Electronic Proceedings of the Asian Conference on Technology in Mathematics (ACTM 2020); Mathematics and Technology.

[7] Dana-Picard, Th. (2021). *The loci of virtual points constructed with elementary models of planetary orbits*, Electronic Proceedings of the Asian Conference on Technology in Mathematics (ACTM 2021); Mathematics and Technology.

[8] Dana-Picard, Th. (2023). *Modelling from reality to mathematics and further creation*, Conference on Design & Visualization in STEAM Education, JKU, Linz, Austria. https://doi.org/10.13140/RG.2.2.34764.01920

[9] Dana-Picard, Th. & Hershkovitz, S. (2023). *From Space to Maths and to Arts: Virtual Art in Space with Planetary Orbits*, International Journal for Technology in Mathematics Education 30(4), 257-264. https://doi.org/10.1564/tme_v30.4.8

[10] Dana-Picard, Th. and Recio, T. (2023). *Dynamic construction of a family of octic curves as geometric loci*, AIMS Mathematics 8 (8), 19461-19476.

[11] Dana-Picard, Th. and Recio, T. (2023). *Second Thales Theorem: a basic but involved example on the need for DGS and CAS cooperation for geometric locus computation*, Boletín de la Sociedad Puig Adam 116, 44-59.

[12] Dana-Picard, Th. and Recio, T. (2023). *Automated computation of geometric Loci in Mathematics Education*, Applications of Computer Algebra (ACA 2023 conference), Warsaw, Poland, 2023. https://iit.sggw.edu.pl/wp-content/uploads/sites/18/2023/07/ACA2023_Program_Abstracts-1.pdf?x58870, 45-46. https://doi.org/10.13140/RG.2.2.35666.53444.

[13] Dana-Picard, Th., Tejera, M. and Ulbrich, E. (2024). *3D space trajectories and beyond: abstract art creation with 3D printing*, Electronic Proceedings in Theoretical Computer Science EPTCS 398, 142–152. https://doi.org/10.4204/EPTCS.398.17

[14] Dana-Picard, Th., Tejera, M. and Ulbrich, E. (2023). *3D Space Trajectories and beyond: Abstract Art Creation with 3D Printing*, Presentation at ADG 2023 Conference. https://www.researchgate.net/publication/374757703_3D_space_trajectories_and_beyond_abstract_art_creation_with_3D_printing.

[15] Dana-Picard, Th., Tejera, M. and Ulbrich, E. (2024). *Art and Maths: a STEAM Approach with 3D Printing to the Study of Exotic Surfaces in Space*, ICME-15, Sydney, Australia. https://doi.org/10.13140/RG.2.2.19984.01284

[16] Duval, R. (1995). *Sémiosis et pensée humaine.* Bern: Peter Lang.

[17] Duval, R. (2006). *A cognitive analysis of problems of comprehension in a learning of mathematics*, Educational Studies in Mathematics **61**, 103–131. https://doi.org/10.1007/s10649-006-0400-z.

[18] Hanna, G. (2000). *Proof, explanation and exploration: An overview*, Educational Studies in Mathematics 44(1), 5–23.

[19] Hanna, G. and de Villiers, M. (edts) (2012). *Proof and Proving in Mathematics Education, The 19th ICMI Study*, Springer.

[20] Jablonski, S. (2023). *Is it all about the setting? -A comparison of mathematical modelling with real objects and their representation*, Educational Studies in Mathematics 113(2).

[21] Karttunen, H., Kröger, P., Oja, H., Poutanen, M., Donner, K.J. (Eds.) (2008). Fundamental Astronomy, Springer.

[22] Kepler, J. (1609) Astronomia Nova, Heidelberg.

[23] Kovács, Z. and Parisse, B. (2015). *Giac and GeoGebra – improved Gröbner basis computations*, in Gutierrez, J., Schicho, J., Weimann, M. (eds.), Computer Algebra and Polynomials, *Lecture Notes in Computer Science* 8942, 126-138, Springer.

[24] Kovács, Z,, Parisse, B., Recio, T., Vélez, M.P. and Yu, J. (2024). *ShowProof in GeoGebra Discovery: Towards automated ranking of elementary geometry theorems*, ISSAC 2024, https://doi.org/10.13140/RG.2.2.22680.81928.

[25] Lawrence, J.D. (2014). A Catalog of Special Plane Curves, Dover Books on Mathematics.

[26] National Research Council. (2010). *Exploring the Intersection of Science Education and 21st Century Skills: A Workshop Summary*. Margaret Hilton, Rapporteur. Board on Science Education, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academy Press.

[27] Roanes-Lozano, E. and Roanes-Macias, E. (2003). *A Bridge between Dynamic Geometry and Computer Algebra*, Mathematical and Computer Modelling **37**, 1005-1028.

[28] Saimon, M., Lavicza, Z., and Dana-Picard, Th. (2022). *Enhancing the 4 C's among College Students of a Communication Skills Course in Tanzania through a project-based Learning Model*, Education and Information Technologies. https://doi.org/10.1007/s10639-022-11406-9.

[29] Xah, L. (2024). Visual Dictionary of Special Plane Curves, (retrieved July 2024) http://xahlee.info/SpecialPlaneCurves_dir/specialPlaneCurves.html

[30] Yates, R. (1947). A Handbook on Curves and their Properties, J.W. Edwards, MI: Ann Arbor.

[31] Zeitoun, D. and Dana-Picard, Th. (2010). *Accurate visualization of graphs of functions of two real variables*, International Journal of Computational and Mathematical Sciences 4(1), 1-11.

Transcendental Function Meditations

William C. Bauldry, bauldrywc@gmail.com
Michael J. Bossé, bossemj@appstate.edu
William J. Cook, cookwj@appstate.edu
Department of Mathematical Sciences
Appalachian State University
Boone, North Carolina 28608
USA

Abstract

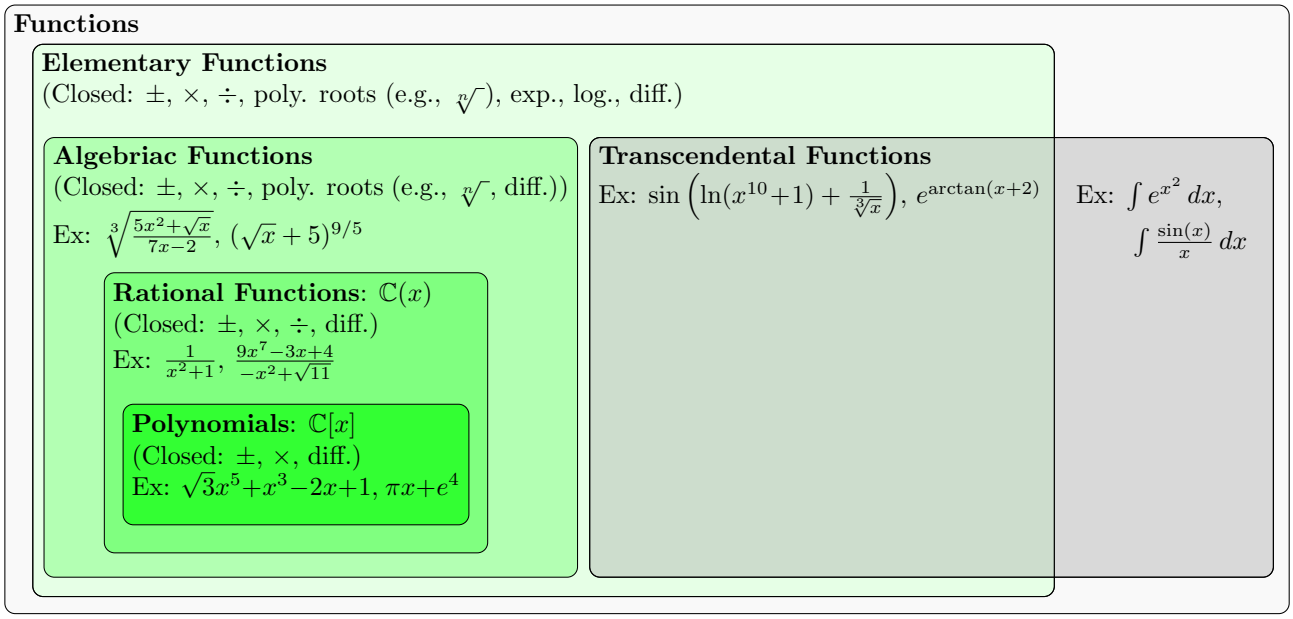
What is a transcendental function? Whether they realize it or not, most students are quite familiar with these functions. In this paper, we first seek to unravel and explain the terms "algebraic" and "transcendental" in the context of numbers and then in the context of functions. We then prove some basic results establishing that exponential, logarithmic, trigonometric, and inverse trigonometric functions are transcendental. We also briefly discuss the nature of an "elementary function.". The paper concludes with a series of reader investigations exploring these concepts in a computer algebra system such as Maple.

1 Why Ponder the Transcendental?

There are many types of functions, including polynomial, rational, algebraic, transcendental, and elementary functions (many of which seem far from *elementary*). We deal with these kinds of functions every day in ordinary algebra and calculus courses. In terms of symbolic manipulation, there is a sizable gulf between algebraic (e.g., $\sqrt{x}$) and transcendental (e.g., e^x) functions. Given the ubiquity of such functions, it seems wise to understand what makes them different and how they arise. Paralleling operational closures leading to the development of various familiar number systems, we seek to elucidate such collections of functions and indicate how they arise when seeking various forms of closure.

We begin this discussion with a figure that sums up the culmination of our investigation. Here we show nested collections function types with some indications of the kinds of closures these systems possess. This is intended to motivate interest among readers who may wish to visually organize polynomial, rational, transcendental, elementary, and others into subset and superset relationships based on notions of

closure. We hope this diagrammatic representation will encourage others to investigate a hierarchical relationship among these functions. Notably, this document's contents will explain much of the symbolism in the diagram.



2 A Numerical Preamble

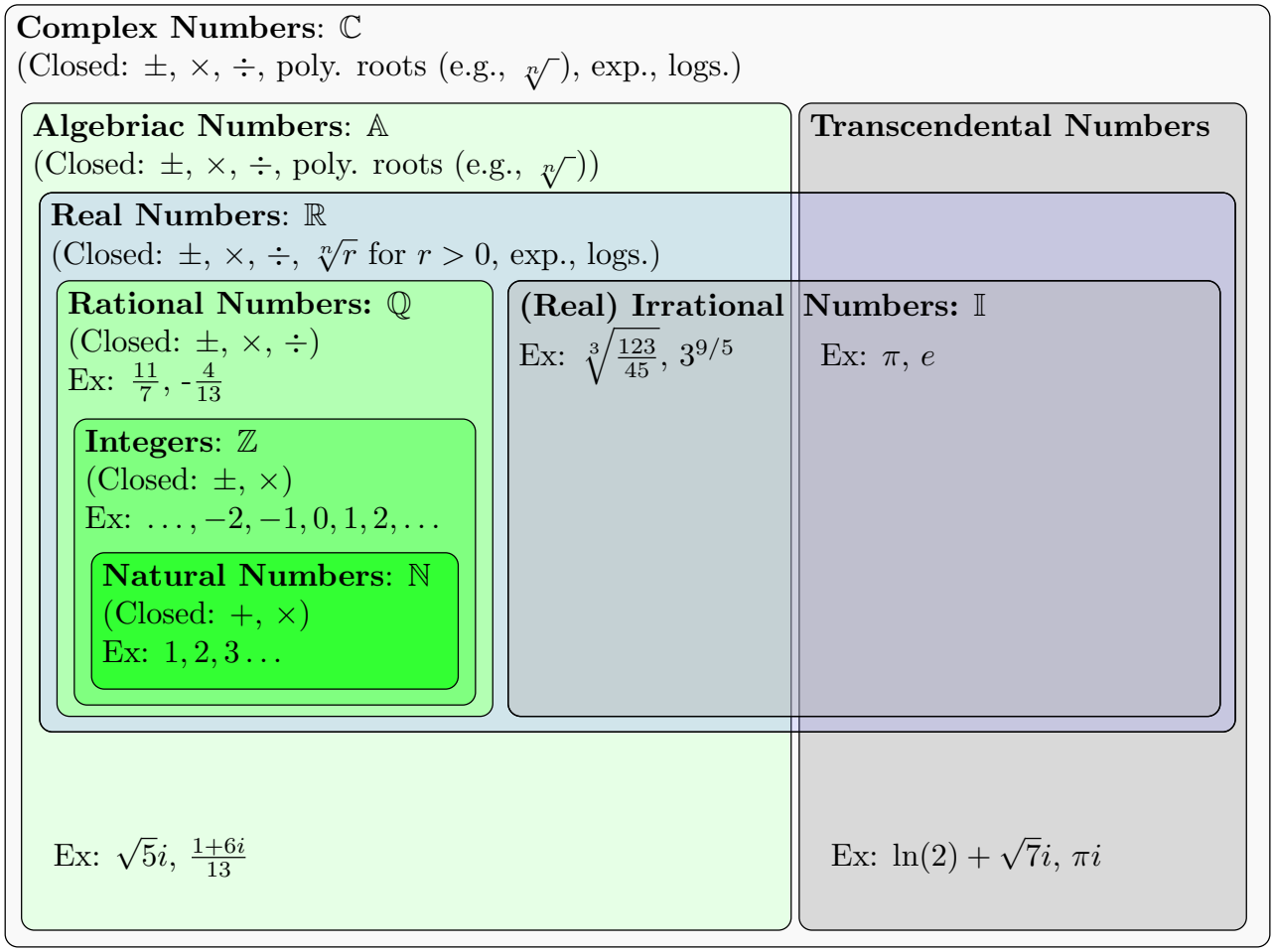
Before diving into the world of functions, we seek to gain a concrete sense of *algebraic* versus *transcendental* in the context of familiar number systems. For a more detailed exploration of various number systems, we recommend [2].

2.1 Number Systems and Operational Closure

The story of number systems can be viewed through the lens of various closures. We understand that the natural numbers (i.e., positive integers), $\mathbb{N} = \{1, 2, 3, 4, \dots\}$, are closed under addition and multiplication and include the multiplicative identity. If we wish to include the additive identity, we must consider the whole numbers (i.e., non-negative integers), $\mathbb{W} = \{0, 1, 2, 3, 4, \dots\}$. Suppose we wish to include closure for subtraction. We must then step up to the integers, $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$, which include additive inverses.[1] Demanding closure under division (as always, except by zero) leads us to the rational numbers, $\mathbb{Q} = \{\frac{a}{b} \mid a, b \in \mathbb{Z} \text{ and } b \neq 0\}$, which include multiplicative inverses.

[1] For those with some abstract algebra background, in this paper, we will only consider rings with characteristic zero (i.e., repeatedly adding a non-zero element to itself will never yield zero). Every such ring (with multiplicative identity) contains an isomorphic copy of the integers (its *prime subring*) and thus gives us a natural starting point.

While this might seem like the end of the road, the rational numbers ($\mathbb{Q}$) are not closed under root taking (e.g., $\sqrt{2} \notin \mathbb{Q}$) as well as many other operations. This leads to the need for additional number systems such as the real numbers ($\mathbb{R}$), the real irrational numbers ($\mathbb{I}$), the algebraic numbers ($\mathbb{A}$), and complex numbers ($\mathbb{C}$). A nesting of some number systems is provided in the following figure.



2.2 Caressing (*Lightly Touching*) Algebraic Structures

We recall names for systems with various closure properties. Without giving all the formal details (see [4]), a *commutative ring* (with unity) is a non-empty collection equipped with operations of addition, subtraction, and multiplication. A good example of such is the integers ($\mathbb{Z}$). If we also want division (except by zero), we have a system called a *field*. The rational numbers ($\mathbb{Q}$), real numbers ($\mathbb{R}$), and complex numbers ($\mathbb{C}$) should be familiar examples of fields.

The real numbers include irrational numbers like $\sqrt{2}$ and $\sqrt[3]{5}$ but also wilder numbers such as $\pi \approx 3.14159265$ and $e \approx 2.7182818$. The real numbers are not only closed under root taking (for non-negative

real numbers) but also closed under a kind of limiting. In fact, one can characterize the real numbers as being the only ordered field closed under taking greatest lower bounds and least upper bounds (i.e., it the the only complete ordered field up to order preserving isomorphism [5]).

2.3 Wildly Irrational Numbers

Two considerations arise from the above discussion. First, what about roots of negative numbers? If we demand root taking for all numbers, we must introduce i where $i^2 = -1$ (i.e., the imaginary root).[2] Gauss, in his dissertation, was able to prove that the complex numbers are *algebraically closed* [8]. This means that $\mathbb{C}$ is not only closed under addition, subtraction, multiplication, division, and root taking but also under extracting roots of any polynomial relation.

Second, what exactly makes π wilder than $\sqrt{2}$? This concern is essentially the main consideration of this paper. The difference between π and $\sqrt{2}$ is that π transcends algebraic (i.e., polynomial) relationships among rational numbers, whereas $\sqrt{2}$ satisfies such a relation (e.g., $(\sqrt{2})^2 = 2$). In particular, we define the notion of an algebraic number:

Definition 2.1 *A complex number $z \in \mathbb{C}$ is* **algebraic** *if there are rational numbers $a_n, \ldots, a_0 \in \mathbb{Q}$, not all zero, such that $a_n z^n + \cdots + a_1 z + a_0 = 0$. In other words, a number is algebraic if it is the root of a non-zero polynomial with rational coefficients. In contrast, any number that fails to be algebraic is called* **transcendental**.

After briefly considering the above definition, it should be apparent why showing a number to be algebraic is often pretty easy. For example, $\sqrt[3]{5}$ is algebraic because it is a root of $x^3 - 5$ (end of proof). On the other hand, showing a number is transcendental is almost always very challenging – challenging but not impossible. In 1873, Charles Hermite showed that Euler's number e is transcendental, and in 1882, Ferdinand von Lindemann proved that π is transcendental [3]. To explore a large collection of transcendental numbers with patterned decimal expansion, see [1]. It should be noted that algebraic numbers are not always easily recognized. For example, $\cos(\frac{\pi}{9})$ is algebraic since it is a root of the cubic polynomial $8x^3 - 6x - 1$, but this is not exactly obvious.

2.4 Planting Some Ideas

This section comes with the following disclaimer: *Here, we provide an overview of ideas that will be addressed in more detail in future sections. While we acknowledge the immediate topical jump in complexity, we ask the reader to persevere and trust that explanations are coming.*

[2] We eschew the notation $\sqrt{-1}$ for the imaginary root i. Why? For example, $\sqrt{4}$ denotes *the* non-negative number that squares to be 4. Thus $\sqrt{4} = 2$. If we just carelessly state that $\sqrt{4}$ is *the* number that squares to 4, we are left with ambiguity for both ± 2 accomplish this. Now if one writes $\sqrt{-1}$, then this should be the *positive* number that squares to be -1. However, there is no way to linearly order the complex numbers in a way compatible with their arithmetic, so neither $\pm i$ can be considered *positive*. Another reason to avoid this notation is that it misleads one into making fallacious arguments such as: $-1 = (\sqrt{-1})^2 = \sqrt{(-1)^2} = \sqrt{1} = 1$.

One must be careful about the general notions of algebraic and transcendental. They are defined relative to some base system. When someone refers to algebraic or transcendental numbers without mentioning a base system, they typically mean they are working over the rational numbers ($\mathbb{Q}$).

More generally, if $\mathbb{K}$ is a field[3] containing another field $\mathbb{F}$ (i.e., $\mathbb{K}$ is an *extension field* of $\mathbb{F}$, or equivalently $\mathbb{F}$ is a *subfield* of $\mathbb{K}$), then we say an element $\alpha \in \mathbb{K}$ is *algebraic* over $\mathbb{F}$ if it is a root of some nonzero polynomial with coefficients in $\mathbb{F}$: $c_n\alpha^n + \cdots + c_1\alpha + c_0 = 0$ for some $c_n, \ldots, c_1, c_0 \in F$ where not all c_i's are zero. Any element $\alpha \in \mathbb{K}$ which fails to be algebraic over $\mathbb{F}$ is said to be *transcendental* over $\mathbb{F}$. This means that while an algebraic element in $\mathbb{K}$ might not belong to $\mathbb{F}$, it is not *too far* away in terms of algebraic entanglements. In contrast, a transcendental element has no algebraic relationship (in terms of adding, subtracting, multiplying, dividing) to the base system!

We emphasize that *the base system matters*. Notice that while π is transcendental over $\mathbb{Q}$, it is algebraic over $\mathbb{R}$ since π is the root of $x - \pi$ (a polynomial whose coefficients 1 and $-\pi$ belong to $\mathbb{R}$). In fact, working over $\mathbb{R}$, every element of $\mathbb{C}$ is algebraic, whereas working over $\mathbb{Q}$, most of the elements of $\mathbb{C}$ are transcendental.[4] Trivially, if $c \in \mathbb{F}$, then c is the root of $x - c$ whose coefficients (i.e., 1 and $-c$) belong to $\mathbb{F}$. Thus, unsurprisingly, every element of $\mathbb{F}$ is algebraic over $\mathbb{F}$ itself.

3 A Brief Account of *Some* Mathematical Field Theory

We note that algebraic numbers possess some closure properties. Consider some field $\mathbb{K}$ extending another field $\mathbb{F}$ with $\alpha, \beta \in \mathbb{K}$. If α and β are algebraic over $\mathbb{F}$, then so are $\alpha + \beta$, $\alpha - \beta$, $\alpha \cdot \beta$, and α/β (if $\beta \neq 0$). In fact, roots of algebraic numbers are still algebraic. Loosely speaking, doing algebraic stuff to algebraic elements yields algebraic elements (working over some fixed base field $\mathbb{F}$). For example, since $\sqrt{2}$ and $\sqrt[3]{5}$ are algebraic (over $\mathbb{Q}$), we have $\sqrt{2} - \sqrt[3]{5}$ and $\sqrt[9]{\sqrt[3]{5} + 123\sqrt{2}}$ are algebraic as well.[5]

Let us briefly explain why the statements above hold and simultaneously introduce some useful notation and concepts. In linear algebra, we learn that a vector space is a collection of elements equipped with a (vector) addition and scalar multiplication. Introductory courses generally restrict scalars to the fields of real ($\mathbb{R}$) or sometimes complex ($\mathbb{C}$) numbers. However, one can work more generally with any field $\mathbb{F}$ as their field of scalars. Notably, a vector space V working over a field of scalars $\mathbb{F}$ always has a basis, and any two bases have the same number of elements. This common size of a basis is called the *dimension* of the vector space and is often denoted $\dim_{\mathbb{F}}(V)$ (or leaving out the $\mathbb{F}$ subscript if our field of scalars is understood). For example, $\dim_{\mathbb{F}}(\mathbb{F}^3) = 3$ if $\mathbb{F}^3$ denotes ordered triples with entries in $\mathbb{F}$. For example, we could use $\{(1,0,0),(0,1,0),(0,0,1)\}$ (a set of 3 elements) as a basis.

Given a field $\mathbb{K}$ extending a field $\mathbb{F}$, we can view $\mathbb{K}$ as a vector space with scalars in $\mathbb{F}$ (we certainly

[3] Recall that a field is an abstract system closed under addition, subtraction, multiplication, and division.

[4] Given $z = a + bi \in \mathbb{C}$ where $a, b \in \mathbb{R}$, we have $(x - z)(x - \overline{z}) = (x - a - bi)(x - a + bi) = x^2 - 2ax + (a^2 + b^2)$ is a real polynomial where z is a root. Thus z is algebraic over $\mathbb{R}$. On the other hand, one can show that there are only *countably* many algebraic numbers, whereas there are *uncountably* many real and complex numbers. Thus "most" of these numbers must be transcendental.

[5] Good luck guessing a rational polynomial with such roots! There are algorithms that do this. Computer algebra systems such as Maple can find such polynomials for us if we would like.

can add elements of $\mathbb{K}$ and multiply elements of $\mathbb{K}$ by elements of $\mathbb{F}$). In mathematical field theory, one defines $[\mathbb{K} : \mathbb{F}] = \dim_{\mathbb{F}}(\mathbb{K})$ to be the *degree* of the extension of $\mathbb{K}$ working over $\mathbb{F}$. In other words, the degree of an extension is the dimension of that field viewed as a vector space over the base field. For example, every complex number can be uniquely expressed as $a + bi$ where $a, b \in \mathbb{R}$. This means $\{1, i\}$ is a basis for $\mathbb{C}$ viewed as a real vector space and so $[\mathbb{C} : \mathbb{R}] = 2$. On the other hand, $[\mathbb{C} : \mathbb{C}] = 1$ (it is always true that $[\mathbb{F} : \mathbb{F}] = 1$) and $[\mathbb{C} : \mathbb{Q}]$ is infinite.[6]

We need some notations for extending a number system to a larger one. First, given a commutative ring (with multiplicative identity) R, the notation $R[t]$ denotes the smallest ring accommodating all of R and the "new" element t. A general element of $R[t]$ looks like $c_n t^n + \cdots + c_1 t + c_0$ where $c_n, \ldots, c_1, c_0 \in R$ (i.e., a general element is a polynomial in t with coefficients drawn from R). Next, $R(t)$ denotes the smallest ring accommodating all of R and the new element t and its multiplicative inverse $1/t$. A general element in $R(t)$ looks like $f(t)/g(t)$ where $f(t)$ and $g(t)$ are elements of $R[t]$ (with $g(t) \neq 0$). In other words, these are rational functions in t with coefficients in R. If R is a field, then $R(t)$ is the smallest field containing all of R as well as t. For example, since even powers of $\sqrt{2}$ are of the form 2^k and odd powers of $\sqrt{2}$ are of the form $2^k\sqrt{2}$, it turns out that elements of $\mathbb{Q}[\sqrt{2}]$ are of the form $a + b\sqrt{2}$ for some $a, b \in \mathbb{Q}$. Moreover, by the conjugate trick we learned in our youth,

$$\frac{a + b\sqrt{2}}{c + d\sqrt{2}} = \frac{a + b\sqrt{2}}{c + d\sqrt{2}} \cdot \frac{c - d\sqrt{2}}{c - d\sqrt{2}} = \frac{ac - 2bd}{c^2 - 2d^2} + \frac{-ad + bc}{c^2 - 2d^2}\sqrt{2},$$

ratios of elements in $\mathbb{Q}[\sqrt{2}]$ are still elements in $\mathbb{Q}[\sqrt{2}]$. Thus $\mathbb{Q}(\sqrt{2}) = \mathbb{Q}[\sqrt{2}]$. Similarly, $\mathbb{C} = \mathbb{R}[i] = \mathbb{R}(i)$. When x is an indeterminate, $R[x]$ simply denotes the ring of polynomials with coefficients in R and $R(x)$ denotes rational functions with coefficients in R. For example, $\dfrac{\pi x^2 - \sqrt{3}}{-x^3 + 5x + \sqrt[5]{11}}$ belongs to the field of real rational functions, $\mathbb{R}(x)$.

3.1 Algebraic and Transcendental Elements

Now, we return to algebraic versus transcendental elements. Using our new notation: Given a field $\mathbb{K}$ extending a field $\mathbb{F}$, recall that $\alpha \in \mathbb{K}$ is algebraic over $\mathbb{F}$ if $f(\alpha) = 0$ for some nonzero polynomial $f(x) \in \mathbb{F}[x]$. It is useful to collect all polynomials with root α together into a collection (called an *ideal*). Indeed, given any algebraic (over $\mathbb{F}$) element α, there is a unique monic (i.e., its leading coefficient is 1) polynomial $m(x) \in \mathbb{F}[x]$ of lowest degree such that $m(\alpha) = 0$. Such a polynomial is irreducible in $\mathbb{F}[x]$ (i.e., it does not factor in a non-trivial way in $\mathbb{F}[x]$), and if $f(x) \in \mathbb{F}[x]$ is such that $f(\alpha) = 0$, then $f(x)$ is a multiple of $m(x)$. We call $m(x)$ the *minimal polynomial* of α. For example, $x^2 - 2$ is the minimal polynomial of $\sqrt{2}$ working over $\mathbb{Q}$. Notice that $\sqrt{2}$ is also a root of $x^4 - 4$. So, $x^4 - 4 = (x^2 + 2)(x^2 - 2)$ is a multiple of $x^2 - 2$. Once again, the base field matters. The minimal polynomial of $\sqrt{2}$ working over $\mathbb{R}$ is just $x - \sqrt{2}$. Notice that $x^2 - 2 = (x + \sqrt{2})(x - \sqrt{2})$ factors over $\mathbb{R}$ but not over $\mathbb{Q}$.

When α is algebraic over $\mathbb{F}$, $\mathbb{F}[\alpha]$ is always a field (i.e., $\mathbb{F}[\alpha] = \mathbb{F}(\alpha)$) and $[\mathbb{F}[\alpha] : \mathbb{F}]$ is precisely the degree of α's minimal polynomial. For example, since every element of $\mathbb{Q}[\sqrt{2}]$ can be uniquely written

[6] Technically $[\mathbb{C} : \mathbb{Q}] = 2^{\aleph_0} = \mathfrak{c}$ since $\mathbb{C}$ requires a basis of continuum cardinality when working over the rationals.

as $a + b\sqrt{2}$ for some $a, b \in \mathbb{Q}$, we have that $\{1, \sqrt{2}\}$ is a basis for $\mathbb{Q}[\sqrt{2}]$ (as a vector space over $\mathbb{Q}$), so $[\mathbb{Q}[\sqrt{2}] : \mathbb{Q}] = \dim_{\mathbb{Q}}(\mathbb{Q}[\sqrt{2}]) = 2$. Alternatively, we see that $[\mathbb{Q}[\sqrt{2}] : \mathbb{Q}] = \deg(x^2 - 2) = 2$.

On the other hand, if α is transcendental over $\mathbb{F}$, our element α behaves like an indeterminate. In this case, $\mathbb{F}[\alpha]$ is always a proper subset of $\mathbb{F}(\alpha)$. Also, both $[\mathbb{F}[\alpha] : \mathbb{F}]$ and $[\mathbb{F}(\alpha) : \mathbb{F}]$ are infinite.[7] For example, $\mathbb{Q}[\pi]$ behaves just like the ring of polynomials with rational coefficients, $\mathbb{Q}[x]$. Putting this all together, α is algebraic over $\mathbb{F}$ if and only if $\mathbb{F}[\alpha] = \mathbb{F}(\alpha)$ if and only if $[\mathbb{F}[\alpha] : \mathbb{F}]$ is finite where the degree of this extension is precisely the degree of our algebraic element's minimal polynomial. On the other hand, α is transcendental over $\mathbb{F}$ if and only if $\mathbb{F}[\alpha] \subsetneq \mathbb{F}(\alpha)$ if and only if $[\mathbb{F}[\alpha] : \mathbb{F}]$ is infinite.

If the degree $[\mathbb{K} : \mathbb{F}]$ is finite, then we say $\mathbb{K}$ is a finite extension of $\mathbb{F}$. The above paragraph tells us that $\mathbb{F}[\alpha]$ is a finite extension of $\mathbb{F}$ exactly when α is algebraic over $\mathbb{F}$. One can easily show if $\mathbb{K}$ is a finite (degree) extension of $\mathbb{F}$, then all the elements of $\mathbb{K}$ are algebraic over $\mathbb{F}$. The converse is not true. It is possible to have an infinite (degree) extension containing only algebraic elements.

The *degree formula* is an important tool when working with these concepts. It states that given a field $\mathbb{L}$ extending a field $\mathbb{K}$ which in turn extends a field $\mathbb{F}$, we have

$$[\mathbb{L} : \mathbb{F}] = [\mathbb{L} : \mathbb{K}] \cdot [\mathbb{K} : \mathbb{F}].$$

Consequently, finite extensions of finite extensions are still finite extensions. This also implies that if α is algebraic over $\mathbb{K}$ and $\mathbb{K}$ is an algebraic extension of $\mathbb{F}$, then α is algebraic over $\mathbb{F}$. For example, $\mathbb{Q}[\sqrt[3]{5}]$ is algebraic over $\mathbb{Q}$ since $[\mathbb{Q}[\sqrt[3]{5}] : \mathbb{Q}] = \deg(x^3 - 5) = 3$ is finite. Notice that $\sqrt{1 + \sqrt[3]{5}}$ is a root of $x^2 - (1 + \sqrt[3]{5}) \in (\mathbb{Q}[\sqrt[3]{5}])[x]$. Thus $\sqrt{1 + \sqrt[3]{5}}$ is algebraic over $\mathbb{Q}[\sqrt[3]{5}]$ and thus also algebraic over $\mathbb{Q}$.

All algebraic elements in a field $\mathbb{K}$ (working over a base field $\mathbb{F}$) collected together form a field (a subfield of $\mathbb{K}$ and an extension field of $\mathbb{F}$). In other words, adding, subtracting, multiplying, and dividing (except by zero) algebraic elements yields an algebraic element. Therefore, if t is transcendental and $a \neq 0$ is algebraic, then $a \cdot t$ must be transcendental (otherwise $t = (a \cdot t) \cdot \frac{1}{a}$ is a product of algebraic elements and thus algebraic). Likewise, if t is transcendental and a is algebraic, then $t+a$ is transcendental (otherwise, $t = (t + a) + (-a)$ is the sum of algebraic elements and thus algebraic). This means that numbers like $\sqrt{2} + \pi$ and $5^{1/3} \cdot e$ are transcendental (over $\mathbb{Q}$).

4 Fields of Functions

Recall that when someone says a real or complex number is algebraic (respectively transcendental) without referencing a base field, we understand that they mean it is algebraic (respectively transcendental) over the rational numbers ($\mathbb{Q}$). When we move the realm of functions, we let our default base field be the field of rational functions with complex coefficients.

Definition 4.1 *We let x denote an indeterminate (essentially our real or complex variable) and then $\mathbb{C}(x) = \left\{ \frac{f(x)}{g(x)} \;\middle|\; f(x), g(x) \in \mathbb{C}[x] \text{ and } g(x) \neq 0 \right\}$ is the field of rational polynomials with complex coefficients. Given a complex function $h(x)$, we say $h(x)$ is* **algebraic** *(respectively* **transcendental***) if it is algebraic (respectively transcendental) over $\mathbb{C}(x)$.*

[7] More precisely, both $[\mathbb{F}[\alpha] : \mathbb{F}]$ and $[\mathbb{F}(\alpha) : \mathbb{F}]$ are $\aleph_0$ (countably infinite dimensional).

Why is $\mathbb{C}(x)$ a natural base field? To explain this, we first note why $\mathbb{Q}$ was a natural base field for numbers. We noted that if we want a system possessing the number 1 with closures under addition, subtraction, and multiplication, we must include all of the integers ($\mathbb{Z}$). If we also demand closure under division (except by zero), we must include all rational numbers ($\mathbb{Q}$). For functions, we wish to start off with all "numbers" (i.e., constant functions), and so $\mathbb{C}$ is a starting point. We should also include the so-called identity function $f(x) = x$. If we wish to accommodate all constants ($\mathbb{C}$), the identity function, and also require closures under addition, subtraction, and multiplication, then we must include all complex coefficient polynomials ($\mathbb{C}[x]$). If we also demand closure under division (except by zero), we are forced to include all complex coefficient rational functions ($\mathbb{C}(x)$).[8]

Since x is now playing the role of our variable, we will use a different symbol, Y, as our indeterminate when referring to polynomial relations. We restate our main definition: we have $h(x)$ is an algebraic (complex or real) function if there exists some non-zero polynomial $m(Y) \in (\mathbb{C}(x))[Y]$ such that $m(h(x)) = 0$. In other words, we have $\frac{f_n(x)}{g_n(x)}(h(x))^n + \cdots + \frac{f_1(x)}{g_1(x)}h(x) + \frac{f_0(x)}{g_0(x)} = 0$ for some complex polynomials $f_n(x), \ldots, f_1(x), f_0(x), g_n(x), \ldots, g_1(x), g_0(x) \in \mathbb{C}[x]$ where $g_i(x) \neq 0$ for all $i = 0, 1, \ldots, n$ and at least one $f_j(x) \neq 0$.

To make life easier we notice that we could clear the denominators in our above relation and get the equivalent definition: $h(x)$ is an algebraic function if there exists complex polynomials $c_n(x), \ldots, c_1(x)$, $c_0(x) \in \mathbb{C}[x]$ (not all zero) such that $c_n(x)(h(x))^n + \cdots + c_1(x)h(x) + c_0(x) = 0$. In other words, $h(x)$ is an algebraic function if and only if it is a root of some non-zero polynomial with complex polynomial coefficients: $C(Y) = c_n(x)Y^n + \cdots + c_1(x)Y + c_0(x) \in (\mathbb{C}[x])[Y]$. For example, the square root function $h(x) = \sqrt{x}$ is algebraic. Notice $(\sqrt{x})^2 - x = x - x = 0$. In particular, we have $h(x) = \sqrt{x}$ is a root of the non-zero polynomial $C(Y) = Y^2 - x \in (\mathbb{C}[x])[Y]$.

Building on our previous discussion about how collections of algebraic elements are closed under algebraic operations such as addition, subtraction, multiplication, division, and root taking, we should not be surprised to learn that functions such as $h(x) = \sqrt[5]{\dfrac{x^3 - \sqrt{x}}{\sqrt[3]{x^2 + 1} - 11}}$ are algebraic.[9]

4.1 Transcendental Functions

The main goal of this paper is to demonstrate that many familiar functions are transcendental. Interestingly, even though proving numbers such as π and e are transcendental (over $\mathbb{Q}$) is rather intricate and challenging (see [3]), proving trigonometric, exponential, and logarithmic functions are transcendental is rather elementary. We rely on the following simple theorem:

[8] Since we aim to study the algebraic versus the transcendental, we sweep issues of domain and analysis under the rug. Considering function domains, branch cuts, and polynomial functions versus formal polynomials would take us too far afield. The careful reader should be comforted to know that (when working in characteristic 0) there is algebraically essentially no difference between polynomial functions (treating x as a variable) versus the formal polynomials (treating x as an indeterminate).

[9] See §6 Reader Investigations #5.

Theorem 4.2 *Just like polynomial functions, the zero function is the only algebraic function with infinitely many zeros.*

Proof: Suppose $h(x)$ is algebraic. Therefore, there exists a non-zero polynomial $C(Y) \in (\mathbb{C}[x])[Y]$ such that $C(h(x)) = 0$, say $C(Y) = c_n(x)Y^n + \cdots + c_1(x)Y + c_0(x)$ where $c_n(x), \ldots, c_1(x), c_0(x) \in \mathbb{C}(x)$ and at least one $c_j(x) \neq 0$. Without loss of generality, we may assume $c_n(x) \neq 0$. Next, consider the greatest common divisor (GCD) of the coefficients of the powers of Y in $C(Y)$, say $d(x)$ is the GCD of the polynomials $c_n(x), \ldots, c_1(x), c_0(x)$. Then, $\frac{1}{d(x)}C(Y)$ is still a non-zero element of $(C[x])[Y]$ with root $h(x)$. Thus, without loss of generality, we also assume that the GCD of $c_n(x), \ldots, c_1(x), c_0(x)$ is 1.

Next, suppose that $h(x)$ has infinitely many zeros, say $z_1, z_2, \ldots$ are distinct complex numbers such that $h(z_1) = h(z_2) = \cdots = 0$. We can take our polynomial relation $0 = C(h(x)) = c_n(x)(h(x))^n + \cdots + c_1(x)h(x) + c_0(x)$ and evaluate it at any of these numbers. For any $k = 1, 2, \ldots$, we get $0 = C(h(z_k)) = c_n(z_k)(h(z_k))^n + \cdots + c_1(z_k)h(z_k) + c_0(z_k) = c_n(z_k) \cdot 0 + \cdots + c_1(z_k) \cdot 0 + c_0(z_k)$. Therefore, $c_0(z_k) = 0$ for all $k = 1, 2, \ldots$ so that $c_0(x)$ is a complex polynomial with infinitely many roots. Thus $c_0(x) = 0$ (i.e., $c_0(x)$ is the *zero polynomial*). We now have $C(Y) = c_n(x)Y^n + \cdots + c_1(x)Y + 0 = (c_n(x)Y^{n-1} + \cdots + c_1(x))Y$. Let $D(Y) = C(Y)/Y = c_n(x)Y^{n-1} + \cdots + c_1(x)$. Either $h(x)$ is a root of $D(Y)$ or $h(x)$ is a root of Y. The latter would mean $h(x) = 0$ (and so we are done). Otherwise, we have $h(x)$ is the root of a polynomial (in Y) of one degree less: $D(Y) = c_n(x)Y^{n-1} + \cdots + c_1(x)$. Since the GCD of $c_n(x), \ldots, c_1(x), c_0(x) = 0$ is 1, the GCD of $c_n(x), \ldots, c_1(x)$ is still 1. Thus, $D(Y)$ satisfies the same criterion as $C(Y)$ did but is of one degree lower.

Continuing in this fashion, we eventually must have that $h(x)$ is a root of $P(Y)Y$ where $P(Y)$ is a polynomial in $(\mathbb{C}[x])[Y]$ of degree zero whose coefficients have a GCD of 1. In other words, $h(x)$ is the root of $p_0(x)Y$ for some $p_0(x) \in \mathbb{C}[x]$ where the GCD $p_0(x)$ (by itself) is 1. Thus, $p_0(x)$ is a (non-zero) constant polynomial. Without loss of generality, we may assume $p(x) = 1$. Thus, $h(x)$ is a root of Y. Therefore, we cannot escape that $h(x)$ must be identically 0. ♦

Armed with the above theorem, we can easily get that many familiar functions are transcendental functions.

Corollary 4.3 *Any non-constant periodic function is transcendental.*

Proof: Essentially any non-constant periodic function appropriately shifted by a constant yields a non-constant function with infinitely many zeros and thus must be transcendental by our theorem above.

In more detail, suppose $f(x)$ is non-constant and periodic with some period p. In other words, $f(z) = f(z+p) = f(z+2p) = \cdots = f(z+np)$ for any z in the domain of $f(x)$ and integer $n \in \mathbb{Z}$. Pick some z_0 in $f(x)$'s domain and define $c = f(z_0)$. Then the function $h(x) = f(x) - c$ has infinitely many zeros since given $z_k = z_0 + pk$ for $k \in \mathbb{Z}$, we have $h(z_k) = f(z_0 + pk) - c = f(z_0) - c = c - c = 0$. Also, since $f(x)$ is non-constant, $h(x) = f(x) - c \neq 0$. Thus, $h(x)$ is transcendental, and so $f(x)$ must be as well. ♦

All of our familiar trigonometric and exponential functions are periodic – at least when working over the complex numbers.

Corollary 4.4 *The trigonometric functions:* $\sin(x)$, $\cos(x)$, $\tan(x)$, $\sec(x)$, $\csc(x)$, $\cot(x)$, *the hyperbolic trigonometric functions:* $\sinh(x)$, $\cosh(x)$, $\tanh(x)$, $\operatorname{sech}(x)$, $\operatorname{csch}(x)$, $\coth(x)$, *and the exponential function* e^x *are all transcendental.*

Proof: The trigonometric functions have periods 2π or π. The hyperbolic trigonometric functions and the exponential function have periods $2\pi i$ or πi. ♦

This also allows us to see that b^x is transcendental for any fixed base $b > 0$ (and $b \neq 1$).[10] Next, we could ask about inverse functions. Here, we have a very clear answer.

As a note, we have been fairly careless about considering domains of functions. Since the notion of being algebraic versus transcendental is about satisfying polynomial relations in an indeterminate, we should see that the specific domains of definition are not really that consequential. Along these lines, we refer to "an" inverse. Technically, we are referring to some branch of an otherwise multivalued inverse. This is in step with a typical Calculus sequence level treatment of functions. For example, when working over the real numbers, we refer to $\sqrt{x}$ as *the* inverse of x^2, but in reality we have two potential branches: $\sqrt{x}$ and $-\sqrt{x}$. Likewise, we refer to $\arcsin(x)$ as *the* inverse of the sine function, but this is just one of infinitely many possible branches of sine's "inverse". Working in the context of the complex numbers muddies these waters even more. Again, we sweep such issues under the rug and focus on algebraic relations. We will call $h^{-1}(x)$ an *inverse* of $h(x)$ if $h(h^{-1}(x)) = x$ for any x where the composition is defined.

Proposition 4.5 *An inverse of an algebraic function is algebraic. Therefore, an inverse of a transcendental function is transcendental.*

Proof: Let $h(x)$ be algebraic with inverse $h^{-1}(x)$ (i.e., $h(h^{-1}(x)) = x$). Since $h(x)$ is algebraic, we have $h(x)$ is the root of some non-zero polynomial $C(Y) = c_n(x)Y^n + \cdots + c_1(x)Y + c_0(x) \in (\mathbb{C}[x])[Y]$. We can expand each coefficient $c_j(x) = b_{m_j j}x^{m_j} + \cdots + b_{1j}x + b_{0j}$. Expanding out our polynomial and thinking about it as a polynomial in both Y and x, we have $B(x, Y) = \sum_{i,j} b_{ij}x^iY^j$.

We have that $B(x, h(x)) = \sum_{i,j} b_{ij}x^i(h(x))^j = \sum_j b_j(x)(h(x))^j = 0$. We now make the substitution $x = h^{-1}(t)$ (using a different variable t for clarity) so that $h(x) = h(h^{-1}(t)) = t$. Thus, $0 = B(x, h(x)) = B(h^{-1}(t), t) = \sum_{i,j} b_{ij}(h^{-1}(t))^i t^j$.

Briefly, we have $B(x, h(x)) = 0$ implies $B(h^{-1}(x), x) = 0$. In other words, $Y = h^{-1}(x)$ is a root of the non-zero polynomial $B(Y, x) = \sum_{k,\ell} c_{\ell k}x^kY^\ell \in (\mathbb{C}[x])[Y]$. Thus $h^{-1}(x)$ is algebraic.

Next, if $h^{-1}(x)$ is algebraic, then $(h^{-1})^{-1}(x) = h(x)$ must be algebraic too. Therefore, either both $h(x)$ and $h^{-1}(x)$ are algebraic, or neither are (i.e., either both are algebraic or both are transcendental). ♦

To illustrate the above theorem, consider $\sqrt{x}$ is a root of $Y^2 - x$, and so its inverse x^2 should be a root of $x^2 - Y$ (which it is).

[10] Notice $b^x = e^{x\ln(b)} = e^{x\ln(b)+2\pi i} = e^{(x+2\pi i/\ln(b))\ln(b)} = b^{x+2\pi i/\ln(b)}$, so these (non-constant) functions have period $2\pi i/\ln(b)$.

Corollary 4.6 *Inverse trigonometric functions:* $\arcsin(x)$, $\arccos(x)$, $\arctan(x)$, $\operatorname{arcsec}(x)$, $\operatorname{arccsc}(x)$, $\operatorname{arccot}(x)$, *inverse hyperbolic trigonometric functions:* $\operatorname{arcsinh}(x)$, $\operatorname{arccosh}(x)$, $\operatorname{arctanh}(x)$, $\operatorname{arcsech}(x)$, $\operatorname{arccsch}(x)$, $\operatorname{arccoth}(x)$, *and logarithm functions:* $\ln(x)$ *(as well as* $\log_b(x)$ *for positive base* $b \neq 1$*) are transcendental.*

5 Differential Algebra

Working with complex functions opens a door to consider yet another kind of closure. We might ask if we have closure under derivatives. Notice that the derivative of a polynomial (respectively a rational function) is still a polynomial (respectively a rational function). Thus both $\mathbb{C}[x]$ and $\mathbb{C}(x)$ are closed under differentiation. How about algebraic and transcendental functions?

Proposition 5.1 *The derivative of an algebraic function is an algebraic function.*

Proof: Let $h(x)$ be an algebraic function. In particular, suppose $h(x)$ is a root of $C(Y) = c_n(x)Y^n + \cdots + c_1(x)Y + c_0(x)$ where $c_n(x), \ldots, c_1(x), c_0(x) \in \mathbb{C}[x]$. Also, assume $C(Y)$ is a non-zero polynomial of lowest possible degree (i.e., it is a $\mathbb{C}[x]$-multiple of the minimal polynomial of $h(x)$).

We can use the product, power, and chain rules to help differentiate the relation $0 = C(h(x))$ and get

$$0 = c_n'(x)(h(x))^n + c_n(x)n(h(x))^{n-1}h'(x) + \cdots + c_1'(x)h(x) + c_1(x)h'(x) + c_0'(x).$$

Therefore, $-(c_n'(x)(h(x))^n + \cdots + c_1'(x)h(x) + c_0'(x)) = (nc_n(x)(h(x))^{n-1} + \cdots + c_1(x))h'(x)$. We note that $C(Y)$ had minimal degree such that $h(x)$ was a root, so $nc_n(x)(h(x))^{n-1} + \cdots + c_1(x)$ cannot be identically zero. Therefore, we can solve for $h'(x)$:

$$h'(x) = -\frac{c_n'(x)(h(x))^n + \cdots + c_1'(x)h(x) + c_0'(x)}{nc_n(x)(h(x))^{n-1} + \cdots + c_1(x)}.$$

Finally, notice that $h(x)$ and polynomials ($c_n'(x)$ and $nc_n(x)$ etc.) are algebraic. Also, algebraic functions form a field (i.e., closed under adding, subtracting, multiplying, and dividing by non-zero elements). Therefore, $h'(x)$ must be algebraic too. ♦

We note that the above proof not only shows the derivative of an algebraic function must be algebraic, but it also gives a formula for its derivative in terms of our algebraic function's minimal polynomial. For example, $h(x) = \sqrt{x}$ satisfies $Y^2 - x$. Thus, $h'(x) = -\dfrac{0(h(x))^2 - 1}{2h(x)} = \dfrac{1}{2\sqrt{x}}$ as we know from Calculus.

In contrast to algebraic functions, transcendental functions are not closed under differentiation. For example, we already know that $\arctan(x)$ is transcendental. However, its derivative, $\dfrac{1}{1+x^2}$, is algebraic. On the other hand, sometimes derivatives of transcendental functions do remain transcendental. For example, $\sin(x)$ is transcendental, as is its derivative $\cos(x)$.

From the results above and our experiences in Calculus, we are led to believe that differentiation leads us from more complicated functions to simpler ones. From an analytic perspective, this is definitely not the case![11] However, from a symbolic perspective, there is some truth. Differentiation moves us from more exotic to less exotic collections. Repeatedly differentiating polynomials eventually yields constants and then zero. Differentiating our inverse trigonometric and log functions yields rational functions or square roots of rational functions.

5.1 Elementary Functions

In Calculus, we want access to certain algebraic and exponential, logarithmic, and trigonometric functions. Functions built from these pieces are referred to as *elementary functions*. First, we note that trigonometric and inverse trigonometric functions are redundant when working in the world of complex variables. For example, $\sin(x) = \frac{1}{2i}(e^{ix} - e^{-ix})$ and $\arcsin(x) = -i\ln(ix + \sqrt{1-x^2})$. Thus, up to some algebraic stuff, we can just consider exponentials and logarithms.

One more carefully defines an elementary function to be a function obtained by finitely many steps of exponentiating, taking a logarithm, and extracting a root of a polynomial. For example,

$$h(x) = \sqrt{\frac{e^{\sqrt[3]{x^5+1}} + x^{10} - 1}{\ln(x^3 + \sqrt{x^9 - 4}) + x^3 + \ln(e^{1/x} + 11)}}$$

is an elementary function.

One can show that while elementary functions are closed under our usual algebraic operations and differentiation, they are not closed under anti-differentiation (i.e., integration). For example, $\int e^{x^2}\,dx$ and $\int \frac{\sin(x)}{x}\,dx$ are not elementary functions [7]. A special part of the study of Differential Algebra known as Differential Galois Theory gives a more complete picture of how all of these things fit together. For example, Andy Magid [6] tells us that elementary functions are exactly those that live in an abelian differential field extension of $\mathbb{C}(x)$. Without explaining exactly what this means, we assure the reader that elementary functions (as non-elementary as they might seem) are a mathematically natural collection of functions for one to consider when working in a field of study such as Calculus.

6 Reader Investigations

The following is a set of investigatory exercises that use the computer algebra system Maple™. These problems can be easily modified for other computer algebra systems such as TI Nspire™, Sage, PocketCAS™, Mathematica™, or Wolfram Alpha™. A Maple™worksheet containing these investigations can be found at: https://BillCookMath.com/papers/Transcendental-Investigations.mw

1. We'll use Maple to explore some minimal polynomials for simple radicals. Start with:
   > *MP := PolynomialTools:-MinimalPolynomial*:

[11] Differentiation destroys. Integration smooths.

> *p3* := *MP*($\sqrt{3}, x$);

$$p3 := x^2 - 3$$

> *p5* := *MP*($\sqrt{5}, x$);

$$p5 := x^2 - 5$$

Now try:

> *p3p5* := *expand*(*p3*·*p5*);

$$p3p5 := x^4 - 8x^2 + 15$$

> *p15* := *MP*($\sqrt{15}, x$);

$$p15 := x^2 - 15$$

> *solve*(*p3p5*);

$$\sqrt{3}, -\sqrt{3}, \sqrt{5}, -\sqrt{5}$$

> *solve*(*p15*);

$$\sqrt{15}, -\sqrt{15}$$

Explain how combining the radicals affects the minimal polynomials and their roots.

Let's try:

> *MP*($\sqrt{3} + \sqrt{5}$, *x*);
solve(%);

$$x^4 - 16x^2 + 4$$
$$\sqrt{5} - \sqrt{3}, -\sqrt{5} + \sqrt{3}, \sqrt{3} + \sqrt{5}, -\sqrt{3} - \sqrt{5}$$

> *p3* + *p5*;
solve(%);

$$2x^2 - 8$$
$$2, -2$$

What can we conclude about the minimal polynomials of combinations of radicals?

2. Let's consider a different combination of simple radicals. What does the following sequence of commands tell us?

> *MP* := *PolynomialTools:-MinimalPolynomial*;
> *r2* := $2^{1/2}$;
> *r8* := $8^{1/2}$;
> *minpolyr2* := *MP*(*r2*, *x*);
> *minpolyr8* := *MP*(*r8*, *x*);
> *roots2* := *solve*(*minpolyr2*);
> *roots8* := *solve*(*minpolyr8*);

3. Investigate minimal polynomials of constants and of simple polynomials using the interactive Maple document "Exploring Simple Minimal Polynomials" hosted on the Maple.cloud.

4. A minimal polynomial can be viewed as a level curve of a surface.
Execute the following in Maple.

> *with(plots):*
> *plotOpts := (style = surfacecontour, shading = XY, transparency = 0.5, thickness = 5):*
> *p2 := (x, Y)* $\rightarrow Y^2$ *- x;*
> *plot3d(p2(x, Y), x = -3 .. 3, Y = -3 .. 3, plotOpts, contours = [1]);*
> *display(%, orientation = [-90, 0, 0], axes = normal);*
What can we conclude from the graphs?

> *p3 := (x, Y)* $\rightarrow Y^3$ *- x;*
> *plot3d(p3(x, Y), x = -3 .. 3, Y = -3 .. 3, plotOpts, contours = [1]);*
> *display(%, orientation = [-90, 0, 0], axes = normal);*
What can we conclude from this pair of graphs?

Now interpret the result of:
> *p31 := (x, Y)* $\rightarrow 5 \cdot Y^3 - 3 \cdot Y \cdot x$*;*
> *plot3d(p31(x, Y), x = -3 .. 3, Y = -3 .. 3, plotOpts, contours = [1]);*
> *display(%, orientation = [-90, 0, 0], axes = normal);*
Identify the common feature from all of these graphs?

5. Do the following Maple commands verify that the function

$$h(x) = \sqrt[5]{\frac{x^3 - \sqrt{x}}{\sqrt[3]{x^2+1} - 11}}$$

given at the end of §4's introduction is algebraic or do they show that $h(x)$ is actually transcendental?
> *mp := evala@Minpoly:*
> $h := x \rightarrow ((x^3 - x^{1/2})/((x^2+1)^{1/3} - 11))^{1/5}$*;*
> *Ph := mp(h(x), Y);*
> *subs(Y = h(x), Ph);*
> *radsimp(%);*

6. Showing the negative result "There is no polynomial such that..." is very difficult. Execute:
> *mp := evala@Minpoly:*
> *Psin := mp(*$\sin(x)$*, Y);*

   (a) Is $f(x) = \sin(x)$ algebraic or transcendental? Explain your response.

   (b) What does this result tell us about the capabilities of Maple's *Minpoly* command?

7. Let's look at nesting roots and what happens to the minimal polynomials. Execute the following in Maple:
> *mp := evala@Minpoly:*

> *r1 :=* $\sqrt{2}$*;*
> **for** i **from** 2 **to** 4 **do** $r||i := sqrt(2 + r||(i-1))$ **end do**;
Now compute the minimal polynomials and then find the roots.
> **for** i **to** 4 **do** $p||i := mp(r||i, Y)$ **end do**;
> **for** i **to** 4 **do** $[solve(p||i = 0, Y)]$ **end do**;

(a) Do you see any patterns in the minimal polynomials?

(b) Do the polynomials share any common roots?

8. Proposition 4.5 concerns the inverse of an algebraic function. Define the function $g(x) = x^2 + \sqrt{x}$, then find its minimal polynomial.
> *mp := evala@Minpoly:*
> $f := x \to x^2 + \sqrt{x}$;
> *mp(f(x), Y);*
> *minpoly := sort(%, Y);*

Now verify that f is a root of the polynomial.
> *subs(Y = f(x), minpoly);*
> *expand(%);*

Compute an inverse for f.
> *solve(Y = f(x), x);*
> *finv := %;*

Proposition 4.5 says the inverse function f^{-1} satisfies the minimal polynomial of f with the variables interchanged ($Y \leftrightarrow x$).
> *subs(x = finv, minpoly);*
> *simplify(%);*
Did it work? Explain.

9. The (real-valued) Lambert-W function is a *special function* that is defined as the solution of the equation $ye^y = x$ for y. This function has import applications in quantum physics, biochemistry, and applied mathematics.
Define the function $f(x) = x \cdot e^x$ in Maple.
> $f := x \to x \cdot \exp(x)$;

Find the inverse function.
> *finv := solve(y = f(x), x);*

This inverse is analogous to $\arctan$ in that it has an infinite number of "branches".
Let's check it.
> *'f'(x) = f(x);*
> *LambertW(f(x)):*
> *simplify(%) assuming* $x > 0$:

> *%% = %;*

The other direction:
> *f(LambertW(y));*
> *simplify(%) assuming* $y > 0$*;*

To know more than you ever wanted, enter
> *FunctionAdvisor(LambertW);*

10. Proposition 5.1 states that if $f(x)$ is algebraic, then $f'(x)$ is also algebraic.

    (a) In Maple, define the absolute value $A(x) = |x|$ and the signum $S(x) = \begin{cases} 1 & x > 0 \\ 0 & x = 0 \\ -1 & x < 0 \end{cases}$ functions.

    (b) Plot the two functions together with

    $$> plot([A(x), S(x)], x = -2..2, -2..2, discont = [symbol = solidcircle]);$$

    How do the functions appear to be related from looking at the graph?

    (c) $A(x)$ is algebraic as it satisfies the polynomial $p(x, Y) = Y^2 - x^2$. However $S(x)$ is *not* algebraic! (An algebraic function is continuous on its domain.) Explain how this does not violate Proposition 5.1. Or does it? (Hint: Ask Maple to $convert(diff(|x|, x), piecewise)$.)

11. Examine the trigonometric functions written in terms of e^x and $\ln(x)$ for patterns by making charts in Maple. Which of the functions in the charts are *elementary functions*?

    (a) Trigonometric functions chart:
    $> T := \langle \cos(x), \sin(x), \tan(x), \sec(x), \csc(x), \cot(x) \rangle:$
    $> Te := map(convert, T, expln):$
    $> \langle T \,|\, \langle \text{`}\leftrightarrow\text{`}\$6 \rangle \,|\, Te \rangle;$

    (b) Inverse trigonometric functions chart:
    $> AT := \langle \arccos(x), \arctan(x), \arcsin(x), \operatorname{arcsec}(x), \operatorname{arccsc}(x), \operatorname{arccot}(x) \rangle:$
    $> ATl := map(convert, AT, expln):$
    $> \langle AT \,|\, \langle \text{`}\leftrightarrow\text{`}\$6 \rangle \,|\, ATl \rangle;$

References

[1] W. Bauldry, M. Bossé, & W. Cook, Patterned Transcendental Numbers, EJMT, Vol. 18 (2024) No. 1 (February).

[2] W. Bauldry, M. Bossé, & W. Cook, Number Systems Tower, Journal of Humanistic Mathematics, Vol. 13 Issue 2 (July 2023), pages 398-426.

[3] E. Burger and R. Tubbs, *Making Transcendence Transparent*, Springer, New York, NY 2004.

[4] J. A. Gallian, *Contemporary Abstract Algebra*, 9^{th} edition, Cengage Learning, Belmont, CA 2017.

[5] E. V. Huntington, Complete sets of postulates for the theory of real quantities, Trans. Am. Math. Soc. vol. 4, No. 3 (1903) 358–370.

[6] A. Magid, *Lectures on Differential Galois Theory*, AMS University Lecture Series, Vol. 7, 1994.

[7] D. Mead, Integration, Am. Math. Monthly Vol. 68 No. 2 (Feb. 1961) 152–156.

[8] J. Stillwell, *The Story of Proof: Logic and the History of Mathematics*, Princeton University Press, Princeton, NJ 2022.

Blending Media Arts with Mathematics: Insights and Innovations in STEAM Education

Sahar Aghasafari
e-mail: s.aghasafari@sc.edu
University of South Carolina Lancaster, SC, US

Mark Malloy
e-mail: mm232@email.sc.edu
University of South Carolina Lancaster, SC, US

Abstract

This study explores the integration of media arts with mathematics education through a focused investigation of the Pythagorean theorem, aiming to elucidate how this interdisciplinary approach can enhance student engagement and conceptual understanding. Recognizing the transformative potential of technology in education, we specifically examine the use of digital media arts tools to teach this foundational mathematical concept. Our research is motivated by the goal of leveraging technology to create more engaging and effective learning experiences in mathematics, thereby addressing challenges in teaching abstract concepts. We investigate the pedagogical benefits of integrating media arts into mathematics education through a qualitative case study involving undergraduate research assistants. The selection of the Pythagorean theorem as our focal point is justified by its critical importance in geometry and its wide applicability across STEM fields. Our findings reveal that media arts foster deeper engagement and facilitate a more comprehensive understanding of mathematical principles among students. This paper contributes to the field by providing evidence-based insights into the synergistic potential of combining media arts with mathematics education, underscoring the value of innovative approaches in enhancing the learning and teaching of mathematics.

1. Introduction

Integrating science, technology, engineering, arts, and mathematics (STEAM) into educational curricula is not merely a trend but a vital evolution in teaching and learning methodologies. The inclusion of media arts within mathematics education emerges as a transformative method poised to redefine student engagement and learning outcomes. Media arts, encompassing digital arts, animation, game design, and more, offer unique avenues for creativity, expression, and exploration. Their application in mathematics education can significantly elevate student engagement, comprehension, retention, and practical application of complex mathematical concepts [1,2,3,4, 5].

The Pythagorean theorem, a fundamental principle in geometry, serves as an ideal case study for this integration. Despite its critical importance in numerous STEM fields, its abstract nature often challenges learners. Applying digital media arts as a pedagogical tool promises a more accessible, engaging, and comprehensive approach to this and similar mathematical concepts. Recent studies indicate that students exposed to STEAM approaches, particularly those incorporating media arts, show improved engagement and a deeper understanding of mathematical principles [5,6]. These findings highlight the potential of media arts to make abstract mathematical concepts more tangible and relatable, thereby fostering a more engaging and effective learning environment.

Our investigation is further motivated by the evolving role of technology in education and the imperative to leverage such advancements to enhance traditional teaching methodologies. This paper aims to contribute to the growing body of evidence supporting innovative STEAM education strategies by integrating media arts with the Pythagorean theorem study. Through a qualitative analysis of educational interventions that blend media arts with mathematics, we offer insights into

how such integrative approaches can enrich student engagement and facilitate a more nuanced understanding and appreciation of mathematical concepts. This forward-looking approach to STEAM education addresses the urgent need for engaging and effective educational strategies. By incorporating media arts into mathematics education, we leverage diverse creative expression and exploration avenues, enhancing student engagement and facilitating a deeper connection to mathematical concepts. This paper examines the integration process and outcomes, focusing on the Pythagorean theorem to highlight the abstract challenges often presented to students and proposing a multifaceted approach to overcome these barriers, thereby contributing valuable insights into the enrichment of STEAM education and underscoring the importance of interdisciplinary methods in fostering a comprehensive understanding of mathematics for the 21st century.

2. Theoretical Background

The integration of media arts into mathematics education is an evolving area of pedagogical research, increasingly recognized for its potential to enrich student engagement, enhance conceptual understanding, and bolster knowledge retention. This theoretical foundation underpins our study, drawing upon various sources that illustrate this interdisciplinary approach's multifaceted benefits.

2.1 Student Engagement

Numerous studies have demonstrated increased student engagement and motivation to learn mathematical concepts by incorporating creative and body-based learning activities. Jeff et al. [7] found that engaging students in creative body-based learning activities significantly improved their attitudes toward mathematics. The authors noted that "students who previously showed little interest in mathematics became intensely engaged" [7, p. 70]. This indicates that leveraging students' physical and creative expressions can make studying abstract mathematical ideas more appealing and approachable.

Similarly, Tracy et al. [8] revealed enhanced participation and focus when students were involved in activities that integrated movement and creativity into mathematics lessons. They found that such approaches engaged students and supported long-term retention of mathematical concepts. Tracy et al. [8] emphasized that "the hands-on process of creative activities was highly engaging and helped solidify students' understanding of geometric concepts" [p. 170]. Robyne and MacGill [9] explored the role of creative and body-based learning in fostering inclusion and found that these methodologies significantly increased motivation and comprehension among diverse student groups. They highlighted that "artistic aspects activate emotions and an aesthetic sense of mathematical beauty," thereby enhancing student engagement and interest in mathematics [9, p. 1230].

Furthermore, An et al. [10] examined the impact of integrating music into mathematics lessons and found that this interdisciplinary approach improved elementary students' disposition towards mathematics. The study showed that "students' attitudes towards mathematics became more positive when music was used as a medium for teaching mathematical concepts" [10, p. 15]. This body of research collectively underscores the critical role of student engagement in effective math education and points to the potential of creative and body-based learning integration to provide engaging, creative gateways for students to interact with and find meaning in mathematical ideas and processes.

2.2 Conceptual Understanding Enhanced Through Media Arts

The integration of media arts into mathematics education transcends mere student engagement, creating a powerful bridge between conceptual understanding and mastery of mathematical concepts. An enlightening 18-month ethnographic study by Peppler [2] illuminates this transformative potential, showcasing urban youth engaging with mathematics through media arts projects, from quilting to music production and breakdancing. This approach not only made

mathematics more accessible and meaningful but also demonstrated significant improvements in students' spatial reasoning, proportional logic skills, and pattern recognition—core competencies in mathematical thinking.

Echoing Peppler's findings, Bequette and Bequette [1] emphasize the critical role of integrating arts into STEM education. They argue that art and design provide unique contexts that stimulate an innovative and inquiry-based approach to learning, which is essential for the holistic understanding of STEM subjects, including mathematics. This foundational perspective suggests that the arts are not peripheral but integral to fostering a deep, conceptual grasp of mathematical principles through creativity and design thinking.

Further supporting this notion, Laura and Kesteren [3] highlight the effectiveness of STEAM approaches in bridging the gap between the arts and STEM. They point out that STEAM education, which combines the analytical strengths of STEM disciplines with the creative and critical thinking fostered by the arts, leads to a more engaged learning experience. This engagement is pivotal for developing a nuanced understanding of mathematics, as it encourages students to apply mathematical concepts in diverse and creative ways, thereby deepening their understanding and mastery of the subject.

Moreover, Aibhín and Tangney's [5] work on digital art integration reinforces the idea that media art projects enhance mathematical understanding. Their research reveals that digital art projects engage students and challenge them to apply mathematical concepts in innovative contexts, leading to a more profound comprehension of the material. This approach aligns with the broader objectives of STEAM education by marrying mathematics's abstract and logical aspects with the expressive and imaginative facets of the arts.

In essence, the fusion of media arts with mathematics education represents a paradigm shift in how we conceive of teaching and learning mathematics. It underscores the power of contextual, applied learning in deepening conceptual understanding and showcases the remarkable potential of media arts as a conduit for transformative educational experiences. By integrating media arts into the mathematics curriculum, educators can unlock new dimensions of understanding and appreciation for mathematics among students, fostering a generation of learners equipped with both the creative and analytical skills necessary for the challenges of the 21st century.

2.3 Knowledge Retention

The multimodal nature of media art production also seems to have benefits for encoding math concepts into memory. Studies by An et al. [10] found that having students physically manipulate materials to construct fractal geometry models out of paper or 3D print mathematical shapes improved knowledge retention over time. They posit that tapping into diverse sensory experiences beyond purely visual processing activates more brain areas when learning new mathematical information. Other research has also shown positive links between embodied cognition theory and STEAM education practices [11].

As this review outlines, integrating media arts activities into the math curriculum is strongly grounded in the latest educational research. Outcomes include heightened student engagement, improved conceptual mastery, and increased retention of mathematical knowledge over time. As Bequette and Bequette [1] advocate, “Artistic pursuits provide an excellent venue for reinforcing mathematical concepts and opening conversations that unravel the mystery of mathematics” [1, p. 40]. Further research is still needed to develop specific best practices for STEAM integration models. However, the evidence thus far points to substantial benefits from blending creative arts, such as media production, with teaching mathematical topics and problem-solving.

3. Research Method

3.1 Research Objective and Rationale

The primary objective of this study is to investigate how integrating media arts into mathematics education, specifically focusing on the Pythagorean theorem, can enhance student engagement, comprehension, and retention of mathematical concepts. The rationale behind this research is grounded in the potential of digital media arts to transform traditional mathematics teaching methods, making abstract concepts more accessible and engaging through creative and interactive digital tools. The Pythagorean theorem was chosen as the focal point of this study due to its fundamental role in mathematics and potential for creative exploration. Its selection was driven by the desire to explore a universally recognized and applicable concept across various fields, providing a robust foundation for integrating media arts. This decision also stems from preliminary discussions with participants, who identified the theorem as a mathematical concept that often presents challenges in understanding, thereby offering a valuable opportunity for innovation in teaching methods.

3.2 Participants

Our study engaged five undergraduate research assistants from the University of Southeastern US, each selected for their unique blend of expertise and interest in mathematics, arts, and technology. These individuals were involved in a specialized project to develop and pilot mathematics learning modules that intricately weave media arts into the curriculum centered around the Pythagorean theorem. Their participation in this project was entirely voluntary, driven by a shared interest in exploring the intersection of digital media arts and mathematics education. This engagement offered them a unique opportunity to contribute to innovative educational strategies outside the traditional classroom setting. Their tasks within the project spanned from the conceptual design of learning modules to the hands-on implementation of these educational tools, providing them with a comprehensive experience in the practical application of integrating media arts into mathematics education.

3.3 Data Collection Methods

The data collection process was meticulously designed to capture the essence of media arts integration into mathematics education. A collection of students' works—including digital graphics, animations, and multimedia presentations—served as artifacts, evidencing the interplay between media arts and mathematics. Field notes taken during the project's development phases detailed the decision-making process, including platform selection and instructional design. These notes were complemented by team meeting recordings and brainstorming sessions, providing insight into the collaborative and innovative environment fostered among the participants. Focus group interviews were structured around specific questions aimed at understanding participants' experiences, such as:

- How did the integration of media arts influence your understanding of the Pythagorean theorem?
- What challenges and opportunities did you encounter while using digital platforms for mathematical exploration?

3.4 Platform Selection Criteria

In our study, we meticulously selected eight digital platforms from an initial review of more than twenty based on carefully defined criteria aligned with our pedagogical objectives. The platforms were evaluated using a comparison matrix in Table 1 to ensure they met the following criteria:

- User-Friendliness: Ensuring the platforms were intuitive and simple for students to navigate.
- Accessibility: Guaranteeing that students at different levels of technological proficiency could easily access and use the tools.
- Creative Potential: Favoring platforms offering extensive creative expression features specifically tailored to mathematical concepts.

- Relevance: Prioritizing platforms capable of facilitating a meaningful exploration of the Pythagorean theorem in innovative and pedagogically sound ways.

Platform	User-Friendliness	Accessibility	Creative Potential	Relevance to Pythagorean Theorem
Adobe Creative Cloud Express	High	High	High	High
Video Poem	Medium	Medium	High	High
The Wick Editor	High	High	High	High
Code Scratch	High	High	High	High
Piskel App	Medium	Medium	High	Medium
AutoDraw	High	High	Medium	Medium
Adobe Voice Animator	High	High	Medium	Medium
Digital Krita	Medium	Medium	High	Medium

Table 1*: Comparison Matrix of Selected Digital Platforms*

3.5 Data Analysis

The thematic analysis [14] of data from field notes, student artifacts, and interview transcripts followed a structured approach to identify patterns and themes related to integrating media arts and mathematics education. This method was chosen for its flexibility and effectiveness in exploring complex qualitative data, enabling a comprehensive understanding of participants' experiences and the educational impact of the project.

4. Implementation

In our innovative quest to integrate media arts with mathematics education, we focused specifically on the Pythagorean theorem. This endeavor led us beyond traditional teaching methodologies, aiming to highlight computational aspects of the theorem's foundational proofs, enriching students' conceptual understanding. Our comprehensive review of over twenty online media art platforms resulted in selecting eight distinguished for their pedagogical value, accessibility, creative potential, and user-friendliness: Adobe Creative Cloud Express, Video Poem, The Wick Editor, Code Scratch, Piskel App, AutoDraw, Adobe Voice Animator, and Digital Krita. These platforms were chosen for their exceptional potential to render abstract mathematical concepts into tangible and creative expressions, thereby elucidating the conceptual underpinnings of the Pythagorean theorem.

The selected digital tools were meticulously aligned with our learning objectives, aimed at fostering an interactive and immersive exploration of the Pythagorean theorem. Adobe Creative Cloud Express in Figure 1, for instance, was utilized to craft visually engaging infographics elucidating the theorem’s principles, whereas Video Poem in Figure 2 facilitated students in articulating the theorem’s real-world applications through narrative storytelling. Similarly, The Wick Editor in Figure 3 and the Piskel App in Figure 4 enabled the creation of animations that contextualize the theorem in real-life scenarios, enhancing the accessibility and relatability of these abstract mathematical concepts.

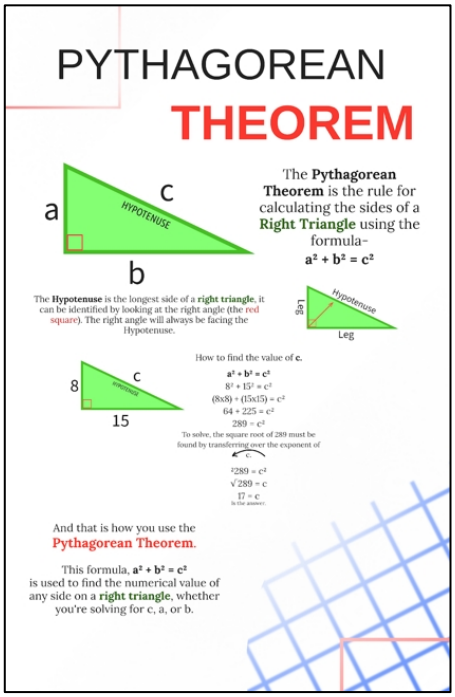


Figure 1: *Adobe Creative Cloud Express*

Figure 2: *Video Poem, visit* *https://drive.google.com/file/d/1hvxhT24xTUTEyBuKFr2PzPb6iqmQZ7tb/view?usp=share_link*

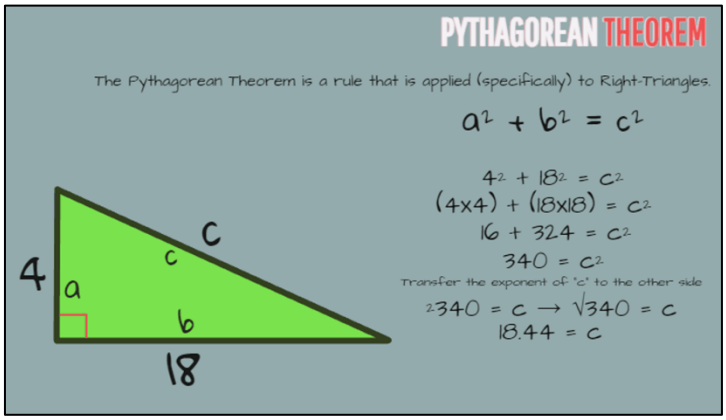


Figure 3: *Wick Editor, visit*

https://drive.google.com/file/d/1l6GxzZBMZjP4gcsRTmr5tckNyWMq1l5j/view?usp=share_link

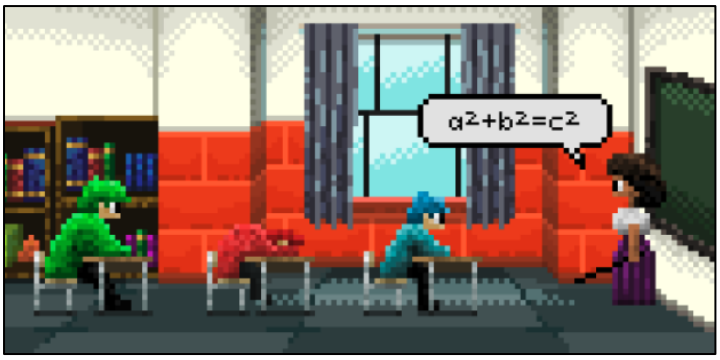


Figure 4: *Piskel App, visit*
https://drive.google.com/file/d/1KDANicTYruywiF7VhzMF3YCX_ZMjjJkH/view?usp=share_link

Further, platforms like Code Scratch in Figure 5 and Adobe Voice Animator in Figure 6 were pivotal in developing interactive simulations and animated stories depicting the Pythagorean theorem in action, promoting a hands-on learning experience. AutoDraw in Figure 7 and Digital Krita in Figure 8 provided unique avenues for students to depict the theorem's geometric principles through digital art, seamlessly merging art and mathematics.

To maximize the efficacy of these tools, scaffolded activities were meticulously designed, guiding students from foundational introductions to more elaborate projects. Adobe Creative Cloud Express was used to create infographics that visually broke down the theorem's components and principles, serving as a foundation for further exploration. In Video Poem, students created narrative storytelling projects that tied the theorem's principles to real-world scenarios, such as navigation and construction, enhancing their understanding through creative expression. The Wick Editor allowed students to create interactive animations, enabling them to manipulate geometric shapes to form right-angled triangles and visually prove the theorem, promoting hands-on exploration. Code Scratch involved students in developing a game that required players to construct proofs of the Pythagorean theorem, encouraging critical thinking about geometric relationships and properties. Piskel App and Adobe Voice Animator were used to create pixel art and animated stories that depicted the theorem's principles in action, enhancing understanding through visual and interactive engagement. AutoDraw helped students sketch and label diagrams of right-angled triangles, experimenting with different configurations to internalize the theorem's logic. Digital Krita facilitated detailed visualizations of the theorem's proof, supporting reflective and exploratory activities.

These activities were instrumental in encouraging students to actively engage with the Pythagorean theorem—creating posters, animations, games, and digital artwork—thus deepening their understanding and practical application of the mathematical concepts. This pedagogical strategy aimed to transform students from passive learners to active creators, employing media arts as a medium to visualize, explore, and internalize mathematical ideas.

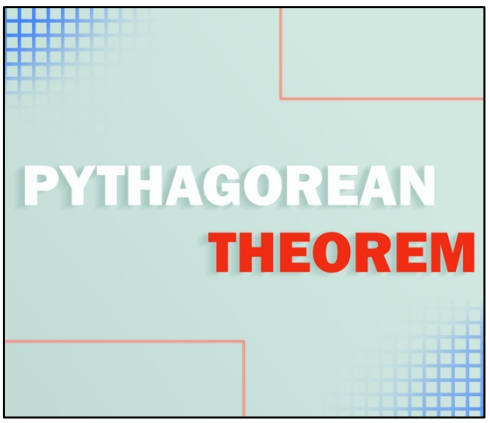


Figure 5: *Code Scratch, visit pythag-theo.netlify.app*

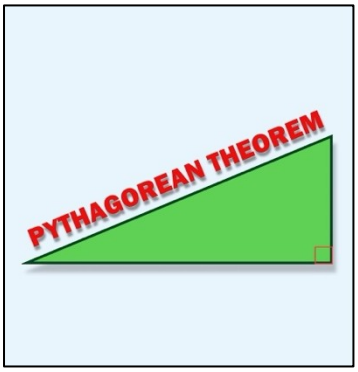


Figure 6: *Adobe Voice Animator, visit https://drive.google.com/file/d/1Kvpw0_q6GlmXLpz8q9qUyvR62oeCg8rL/view?usp=sharing*

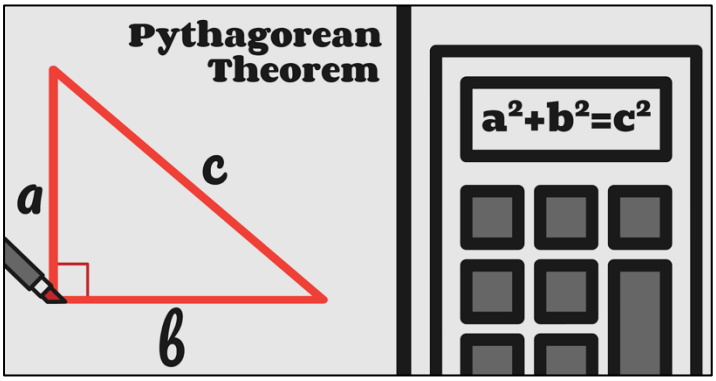


Figure 7: *AutoDraw*

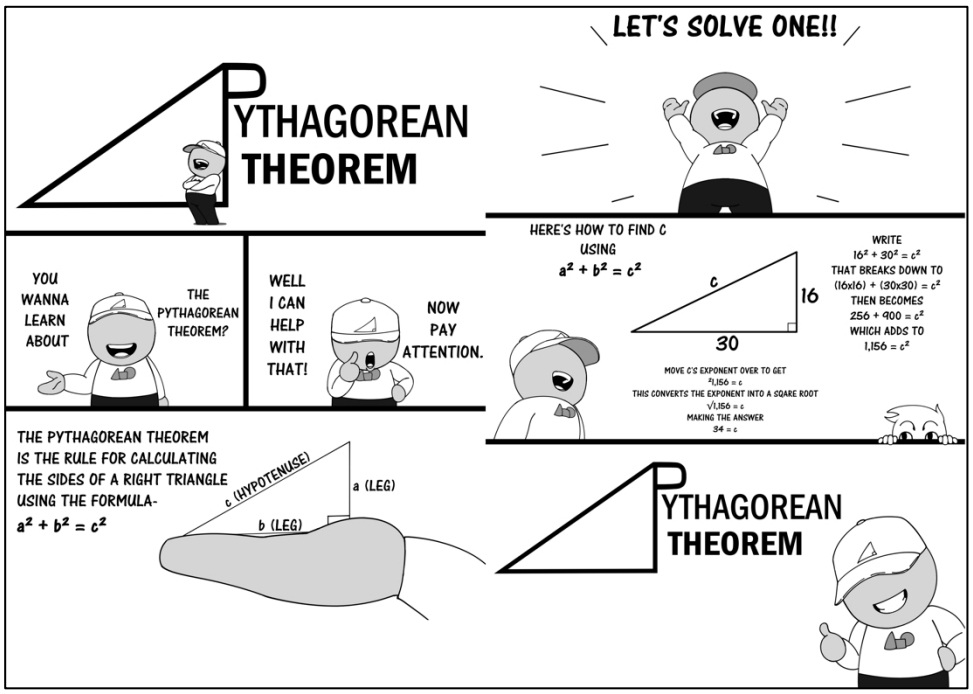


Figure 8: *Digital Krita*

Our approach to integrating media arts into mathematics education aimed to bolster students' comprehension of the Pythagorean theorem, instilling a sense of ownership and enthusiasm for learning. By offering diverse and innovative outlets for expression, we encouraged students to venture beyond traditional mathematical boundaries, fostering a deeper, more personal connection with the subject. This innovative methodology demonstrates that when students are equipped with creative tools to express and manipulate mathematical concepts, their engagement, understanding, and retention of these concepts are significantly enhanced.

Figures 1 through 8 serve as a testament to this approach, showcasing student-created artifacts that perform calculations and explore the theorem's proof, embodying higher-order thinking. These artifacts underscore the viability of media arts integration in mathematics, paving the way for future applications beyond the Pythagorean theorem and contributing to the evolving discourse on STEAM education.

5. Findings

Our exploration into integrating media arts with the Pythagorean theorem uncovered insights that transcend traditional learning outcomes. Through the qualitative analysis of observations, student artifacts, and focus group interviews, we discovered that our approach significantly bolstered engagement, comprehension, and the retention of mathematical concepts.

5.1 Enhanced Visualization and Interactivity

Notably, our investigation revealed that students' engagement with digital platforms did more than bring the theorem's concepts to life; it also facilitated a deeper understanding of its proof. An illustrative student project used The Wick Editor in Figure 9 to depict the theorem in a real-world context and animate the step-by-step logic of the proof itself, offering a tangible insight into its foundational truth.

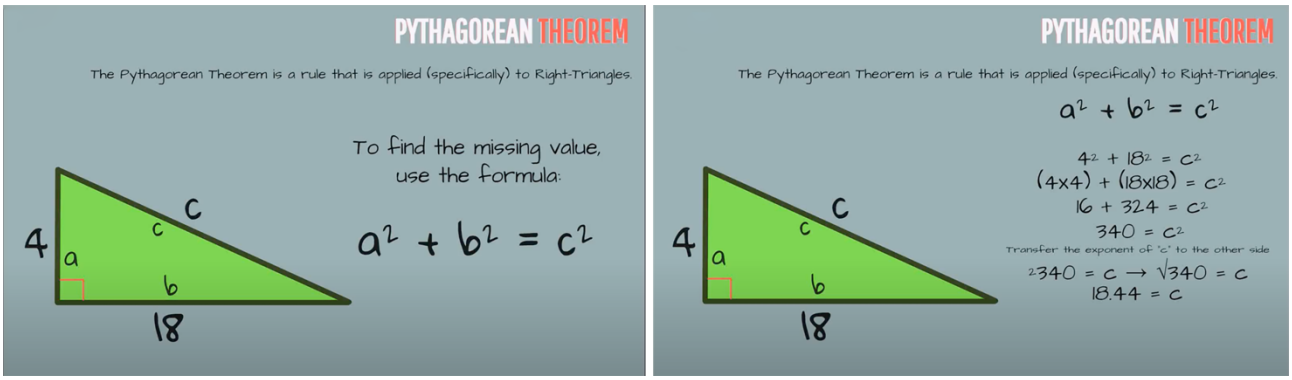


Figure 9: *Wick Editor, Animating the Step-by-Step Proof of the Pythagorean Theorem*

5.2 Creativity and Personalized Learning

The integration of media arts also allowed students to examine the theorem's proofs creatively. A standout video poem in the Figure 10 project tied the theorem's principles to architectural design, encouraging students to think critically about the geometric proof underlying their creative expressions.

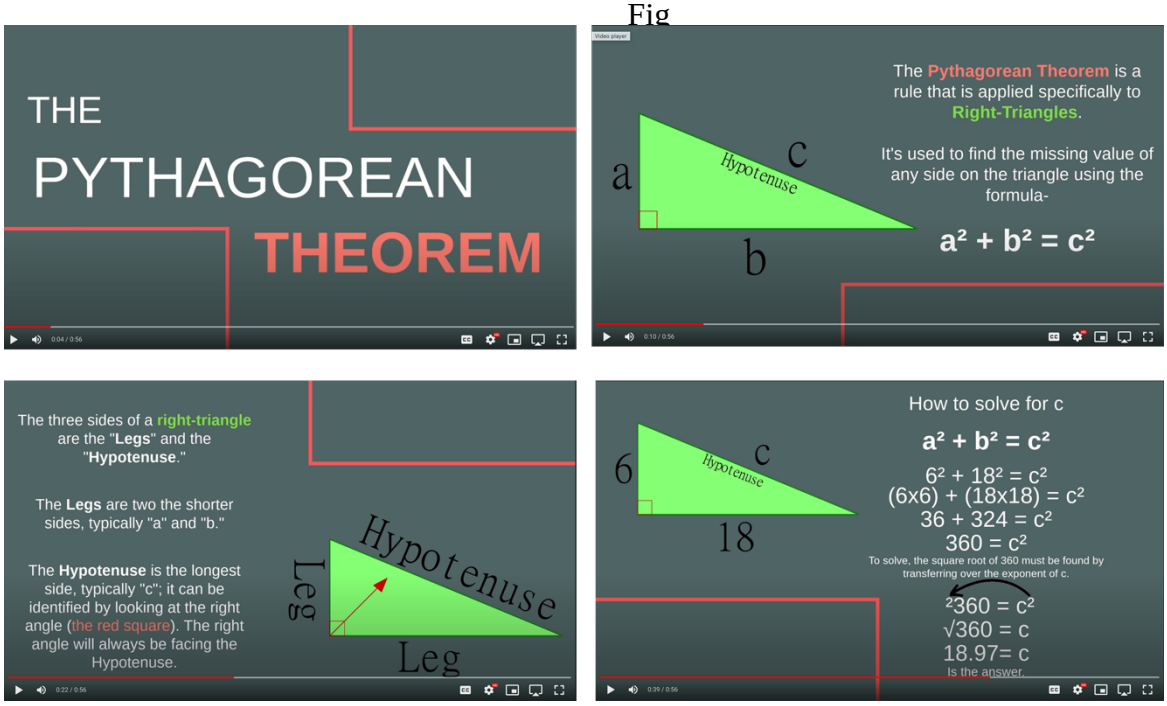


Figure 10: *Video Poems, Narrative Story of the Pythagorean Theorem*

5.3 21st Century Skill Building

The creation of media art projects focusing on the Pythagorean theorem's proof significantly enhanced students' technological literacy, critical thinking, and problem-solving abilities. A collaborative project in Code Scratch, illustrated in Figure 11, involved students using a structured process to create representations of the Pythagorean theorem and develop practice exercises to answer related questions. The process included configuring a concept through brainstorming and research, sketching and selecting a design, creating and merging required assets, moving these assets to Code Scratch, inputting appropriate commands and codes, and finalizing and rendering the game. The resulting game, which challenged users to construct proofs of the theorem, exemplified how digital tools could be harnessed to reinforce mathematical reasoning skills. This interactive

approach reinforced their understanding of the theorem and enhanced their problem-solving skills, ensuring that students thoroughly understood the content by applying their knowledge in various contexts.

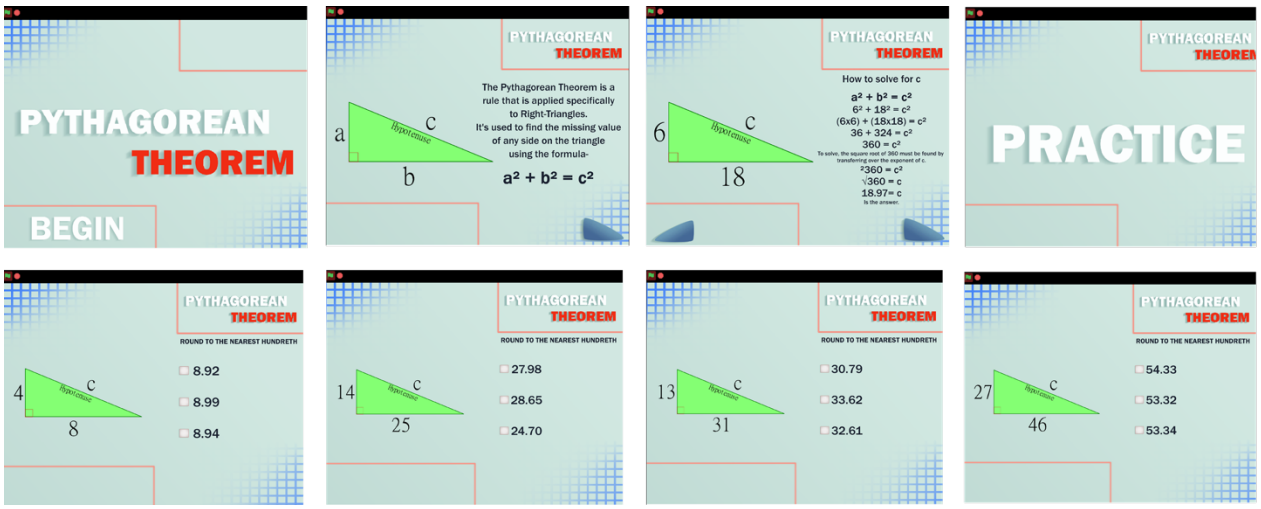


Figure 11: *Code Scratch for Understanding and Applying the Pythagorean Theorem*

5.4 Metacognition and Conceptual Understanding

By creating personalized representations that highlighted the theorem's proof, students underwent a reflective process that deepened their conceptual understanding. A digital art piece using Digital Krita, which visually dissected the proof, exemplified how students could explore and internalize the geometric principles underlying the Pythagorean theorem.

These findings underscore the significant potential of integrating media arts into mathematics education. By providing avenues for students to creatively express and critically engage with mathematical concepts, especially through the exploration of proofs, our study demonstrates the value of innovative and hands-on learning experiences in fostering a deeper understanding of mathematics. This approach not only enhances student engagement and understanding but also cultivates a profound appreciation for the beauty and logic of mathematical proofs, marking a meaningful contribution to the field of STEAM education.

6. Discussion

Our investigation into integrating media arts within the Pythagorean theorem's educational framework marks a significant advancement in the interdisciplinary STEAM dialogue. Through a focused application of digital media arts, this research not only aligns with but also extends the foundational insights provided by Bequette and Bequette [1], Peppler [2], and others [3,4] on the fusion of arts and STEM education. Our study underscores the unique pedagogical value of employing digital media to foster a deeper engagement and understanding of mathematical concepts, notably the Pythagorean theorem.

Echoing the sentiments of Laura and Kesteren [3] and Pavel [4] on the positive impacts of arts integration, our findings reveal that media arts play a crucial role in demystifying complex mathematical ideas. This is further supported by the empirical evidence from Aibhín and Tangney [5] and Teaching Mathematics through Art Using Digital Technologies [6], highlighting digital media's capability to significantly bolster mathematical engagement and comprehension. Our research diverges by providing a granular look at how specific digital tools captivate students and enhance their grasp of mathematical fundamentals through creative exploration.

The diversity in media arts applications in mathematics education, ranging from creative body-based learning [7] to fostering motivation and engagement through qualitative teacher-student interactions [8], illustrates the wide spectrum of creative approaches that can augment learning outcomes. Our study contributes novel insights into this array by showcasing the successful application of creative computing [16] and digital storytelling [15] to illuminate the Pythagorean theorem, reinforcing the utility of such innovative pedagogical strategies. Additionally, integrating creative and body-based learning methods supports diverse learners and fosters inclusion in educational settings [9].

7. Conclusion

By integrating media arts into the teaching of the Pythagorean theorem, our study illuminates a pathway for transforming mathematics education. Our empirical evidence corroborates the existing advocacy for blending arts and technology [1,2,3] and illuminates specific digital media arts tools' roles in deepening mathematical understanding. Our findings extend beyond the realm of the Pythagorean theorem, suggesting a flexible and impactful methodology for incorporating media arts across various mathematical topics.

This adaptability is particularly crucial for meeting the educational demands of the 21st century, where bridging traditional disciplinary divides and leveraging technological advancements becomes essential. Therefore, our investigation contributes to the ongoing evolution of STEAM education and opens new avenues for research. By demonstrating the practical application of media arts to clarify and engage with fundamental mathematical concepts, we underline the effectiveness of creative pedagogical strategies in enhancing educational experiences and outcomes.

Inspired by the groundwork laid in this study, future research promises to further broaden the scope of STEAM education. By exploring the integration of media arts in other mathematical areas, we can continue to innovate and enrich teaching and learning practices, equipping students with the critical, creative, and technological skills necessary for their future success.

Our study enriches the STEAM education discourse by offering detailed insights into the application of media arts within mathematics education. It aligns with the burgeoning body of literature advocating for the arts and technology merger and charts new territories for how this integration can be specifically tailored to enhance mathematics teaching and learning.

8. References

[1] J. Bequette and M. Bequette, A place for art and design education in the STEM conversation, *Art Education*, vol. 65, no. 2, pp. 40-47, 2012.
[2] K. Peppler, Media arts: Arts education for a digital age, *Teachers College Record*, vol. 112, no. 8, pp. 2118-2153, 2010.
[3] B. Laura, and J. Van Kesteren. The Art of Math. *Ontario Mathematics Gazette*, vol. 57, no. 2, pp. 21–24, Dec. 2018.
[4] S. Pavel. Bringing More STEAM to Mathematics Education. *International Journal for Technology in Mathematics Education*, vol. 24, no. 4, pp. 191–98, Oct. 2017.
[5] B. Aibhín, and B. Tangney, Enhancing Student Engagement through the Affordances of Mobile Technology: A 21st Century Learning Perspective on Realistic Mathematics Education. *Mathematics Education Research Journal*, vol. 28, no. 1, pp. 173–97, 2016.
[6] Teaching Mathematics through Art Using Digital Technologies: A Research-Action on Complementary Distance Learning in Schools. (English). *Open Education: The Journal for Open & Distance Education & Educational Technology*, vol. 15, no. 1, pp. 136–59, Jan. 2019.
[7] M. Jeff et al., Creative Body-Based Learning: Not Just Another Story about the Arts and Young People, *Childhood Education*, vol. 95, no. 6, pp. 66–75, Nov. 2019.
[8] D. Tracy et al., Motivation and Engagement in Mathematics: A Qualitative Framework for Teacher-Student Interactions. Mathematics Education Research Journal, vol. 29, no. 2, pp. 163–81,

June 2017.
[9] G. Robyne, and B. MacGill. Fostering Inclusion in School through Creative and Body-Based Learning *International Journal of Inclusive Education*, vol. 25, no. 11, pp. 1221–35, Sept. 2021.
[10] S. An et al., Fostering Elementary Students' Mathematics Disposition through Music-Mathematics Integrated Lessons. *International Journal for Mathematics Teaching & Learning*, pp. 1–19. Sept. 2014.
[11] K. Guyotte et al., STEAM as social practice: Cultivating creativity in transdisciplinary spaces, *Art Education*, vol. 67, no. 6, pp. 12-19, 2014.
[12] R. Yin, *Applied social research methods series: Case study research design methods*, Sage, 2009.
[13] R. Yin, *Case Study Research: Design and Methods* (5th ed.). Thousand Oaks, CA: Sage, 2014.
[14] V. Braun, V. Clarke, Using thematic analysis in psychology, *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77-101, 2006.
[15] S. Andreja Istenic et al., Engaging Preservice Primary and Preprimary School Teachers in Digital Storytelling for the Teaching and Learning of Mathematics. *British Journal of Educational Technology*, vol. 47, no. 1, pp. 29–50. Jan. 2016.
[16] L. Henrikson, Creative computing and STEAM: The role of creativity in integrating the arts and STEM, Paper presented at the annual STEAM conference, Vancouver, BC, 2014.

The affordances of mathematical tasks in a learning management system

Henri Heiskanen
e-mail: henri.heiskanen@uef.fi
University of Eastern Finland
Joensuu, FI

Lasse Eronen
e-mail: lasse.eronen@uef.fi
University of Eastern Finland,
Joensuu, FI

Pasi Eskelinen
e-mail: pasi.eskelinen@uef.fi
University of Eastern Finland,
Joensuu, FI

Laura Hirsto
e-mail: laura.hirsto@uef.fi
University of Eastern Finland,
Joensuu, FI

Abstract

In this study, 400 randomly selected mathematics tasks in a widely used LMS in Finland designed to function as electronic learning platform for mathematics for primary and secondary school pupils were analysed using conceptual and procedural knowledge-based emphases. The analysis showed that approximately 29% of the selected 400 math problems emphasized conceptual knowledge. The results of this study suggest that the LMS consists of more math tasks emphasizing procedural knowledge than conceptual knowledge at primary and secondary school levels. These results are similar to the former findings of studies on Finnish mathematics textbooks. It can be concluded that mathematics tasks in the LMS promote the learning of procedural knowledge, but the criticisms regarding the materials and their versatility are still relevant. Understanding how conceptual and procedural knowledge is emphasized in different LMS helps current and future teachers to select and adapt teaching and learning materials and methods to better meet the needs of their students.

1. Introduction

A learning environment is often conceptualized as an environment where learning takes place. Typical dimensions of a learning environment include physical, social, technological, and pedagogical perspectives (e.g. [1], [2]). As teachers work in these learning environments, they must plan how to design environments to support students' learning as well as how to utilize the learning environments that they are provided. From this perspective, a theoretical framework known as The Activity-Centred Analysis and Design (ACAD) has been developed, which approaches the learning environments from the perspective of affordances (see e.g. [3], [4]). These various affordances are classified into three distinct categories: social, physical, and epistemic [4]. The idea in the framework is that teachers design the affordances for the learning environments before the teaching

sessions, and during the session through active use of these different affordances the emerging learning environment is formed. Within this framework, social affordances refer to various social formations available in the classroom or online platform. Physical or set affordances include all the tools, artifacts, resources, and technologies. Epistemic affordances, for their part, are primarily focused on the tasks, sequences, and pace of tasks and content.

In this study, LMS is defined with reference to the research by Chaubey and Bhattacharya [9] as follows: LMS is a software application or web-based technology used to plan, implement, and evaluate a specific learning process [9]. The LMS platform of this study provides teachers with versatile tools for learning management and teaching [7]. It provides students and teachers opportunities for interactive and flexible mathematics learning, but it also helps students to improve their calculation skills and induce motivation to study mathematics [8], [10]. The LMS provides teachers with ready-made task repositories and learning paths that are already designed to suit, for example, a certain grade level. This LMS system is primarily used in the classroom or at home in addition to other study material. The LMS examined in this study considers the perspective of assessment by providing learning analytics data to both students and teachers about the completion of tasks, time use, and the correctness of answers. In this LMS system, teachers can also modify existing task repositories, create their own tasks, and choose the tasks they consider suitable for their students. For teachers to modify existing task repositories, they should consider what kind of epistemic affordances the tasks in the LMS system and repositories generally contain. The LMS of this study is widely used in Finland primary and secondary schools.

A key question that arises in this context is how to ensure the availability of high-quality mathematics teaching and learning materials in the digital learning environments and LMS because PISA results have suggested that transferring pencil-and-paper mathematical tasks as they are to electronic environments may decrease students' performance [11]. The goal of this study is to shed light on how conceptual and procedural knowledge is emphasized in mathematics tasks by primary and secondary school levels. Specifically, information is obtained about the type of content included in the tasks of the LMS system.

2. The need to assess conceptual and procedural knowledge of the mathematics tasks

The quality of mathematics tasks can be viewed from the perspective of mathematical content, which involves using theory to classify mathematical knowledge. A well-known theory in this regard separates procedural and conceptual aspects of mathematical knowledge [12], [13]. After an analysis of researchers' views, Haapasalo and Kadijevich [14] suggested the following characterization that fits viable theories of teaching and learning: Conceptual knowledge can be defined as an understanding of how concepts and problems form a network with each other. Haapasalo and Kadijevich [14] referred to the concepts in this semantic network as nodes and the connections between them as links. According to them, conceptual knowledge also includes the ability to navigate this network flexibly and to find concepts and their properties that are suitable for each situation or problem. Conceptual knowledge included the concepts, principles and relationships underlying the mathematical subject area and their understanding [12], [15]. More specifically, this included interpretation and construction of concepts and their attributes, procedures, functions, and perspectives [14]. Procedural knowledge can be defined as the ability to use and represent rules, algorithms, and procedures [12]. Haapasalo and Kadijevich [14] saw the foundation of procedural knowledge as processes formed by sequentially executed operations. Procedural knowledge referred to the dynamic and purposeful execution of rules, methods, or algorithms using certain presentation methods [14]. Therefore, procedural knowledge meant the ability to recognize the steps of a process and understand how the next step is carried out.

According to Haapasalo and Kadijevich [14], [16], [17] solving a math task utilization of conceptual knowledge typically required conscious thinking and knowledge of why the task is being solved, while procedural knowledge may consist of automated and subconscious steps. When solving a math task, the utilization of procedural information required the execution of automated calculation routines and knowledge of how to solve the task efficiently and accurately [12], [15], [16], [17]. In this case, there was a dynamic relationship between the two types of knowledge, and a clear boundary between them cannot be defined. Haapasalo [18] suggested that when solving a problem relying on procedural knowledge, two related rules can be successfully combined without understanding the underlying justifications. In general, conceptual, and procedural knowledge are strongly interconnected, and both are needed in solving many problems.

Research literature helps to group mathematics tasks into procedurally or conceptually focused areas [12], [15], [16], [19], [20] [21]. The complexity of tasks can be influenced by the number of procedures or concepts required to solve the task [21]. For example, Authors [22] presented a four-category classification tool that considers the complexity of tasks when evaluating procedural and conceptual knowledge within them. Haapasalo [20] also presented in his work a way to classify mathematics tasks utilizing conceptual and procedural knowledge. Additionally, Phuong [21] introduced a PCK taxonomy that can be utilized for classifying tasks and used in assessing students' mathematics learning.

Star [23] aimed to challenge the notion that procedural knowledge is shallow while conceptual knowledge is deep. For instance, she offered the concept of deep procedural knowledge, which included procedural flexibility that allows an individual to choose the most suitable method from various options in a problem-solving situation, such as solving an equation. On the other hand, knowledge of the relationships between concepts can be quite deficient and superficial, based on rote memorization rather than true comprehension.

When examining mathematics teaching practices, whether teachers should teach mathematics for procedural knowledge, conceptual knowledge, or a combination of the two is a question that arises [13]. This question can be examined through learning materials and mathematical tasks. The procedural–conceptual nature of mathematical tasks is revealed to each student individually based on the student’s existing knowledge and skills [24], [25], [26], [27], although mathematical task can be designed to be conceptually or procedurally oriented [15], [16], [18], [20], [21], [28]. From the point of view of primary school teachers’ professional relevance, mathematics textbooks are central because they are one of the most common epistemic affordances of mathematics in classroom [4] and the tasks given in textbooks shape learners’ perceptions of mathematical knowledge [29], [30].

The construction of connections between concepts can also be supported by other educational means, such as by using functional teaching and learning methods, increasing the use of illustration tools [31], or strengthening the role of interactions in learning mathematics [32]. In teaching arrangements, technology can create opportunities for dialogue about the emphasis on mathematical knowledge [33]. Many LMS systems enable the creation and introduction of conceptual tasks as part of the study of mathematics, making it possible to balance the dialogue around conceptual and procedural emphases in the mathematics learning process. According to Hurrell [34], mathematics teaching and learning materials should emphasize conceptual knowledge as this would allow learners to develop mathematical competence. In addition, mastering a calculation method without a conceptual understanding behind it does not produce permanent or very sustainable knowledge of the matter.

Research suggests that both conceptual and procedural knowledge are important for learning mathematics, and that the balance between the two are crucial for student achievement [7], [36]. A balanced approach to mathematics instruction should emphasize the development of both conceptual and procedural knowledge. Studies have suggested that students who have a strong conceptual knowledge of mathematics tend to perform better on math assessments and are better

able to transfer their learning to new situations than students who lack this conceptual knowledge [36]. Procedural knowledge is also important for success on math assessments, especially for tasks that involve routine calculations or basic algorithms [15].

The effectiveness of mathematics tasks in promoting a balance between conceptual and procedural knowledge depends on several factors, such as the quality of tasks, teacher, and the background knowledge and motivation of the students. Well-designed mathematics tasks provide students with opportunities to develop both conceptual and procedural knowledge and encourage them to make connections between the two [37]. However, it should also be noted that other factors, such as how teachers use mathematics curriculum materials, may be more important in influencing student achievement than the balance of conceptual and procedural knowledge in mathematics learning materials [38], [39].

The findings of researchers on the biases in textbooks are causing serious concern in Finland. In Joutsenlahti's and Vainiopää's [19], [40] analysis studies, more than 80 percent of the tasks in Finnish elementary school mathematics textbooks turned out to be procedurally weighted. According to these researchers, the examination of mathematical concepts in textbooks was superficial and did not support the formation of connections between concepts. According to Viholainen et al. [41], Finnish secondary school mathematics textbooks emphasized mathematical procedures more than concepts and the connections between them.

Therefore, teachers should be made aware of contemporary research and literature on procedural and conceptual knowledge so that they can make informed decisions when selecting appropriate epistemic affordances of mathematics, such as the math tasks and the learning environments [45]. In addition, teachers' understanding about how the conceptual and procedural types of knowledge are emphasized in math tasks is essential when an LMS is used in the math classroom. This is especially true when the implemented technology offers the opportunities to choose whether to focused on the fluency of calculations and the execution of algorithms, which further contributed to develop students' understanding of mathematical concepts, or to focus on understanding concepts, which serves to develop calculations and the execution of algorithms [13], [18], [34].

3. Research Questions and Methods

This study examined epistemic affordances in a widely used LMS in Finland based on procedural and conceptual knowledge. An awareness of the procedural and conceptual knowledge involved in mathematical tasks can provide teachers with valuable information for choosing tasks when creating teaching and learning environments. As such, two research questions were formulated for this study:

Q1. What is the epistemic affordance of the LMS in primary school mathematics task examined from the perspective of conceptual and procedural knowledge?

Q2. What is the epistemic affordance of the LMS in secondary school mathematics task examined from the perspective of conceptual and procedural knowledge?

In this study, the epistemic affordance was formed from 400 electronic math tasks in the LMS which were randomly selected. The tasks were chosen based on a simple random sampling. In all, 220 of 881 tasks were selected from *Mathematics, Task Bank Primary School 2022–2023* in LMS, and 180 of 872 tasks were selected from *Mathematics, Task Bank Secondary School 2022–2023.* Both task banks are freely available to all teachers who have credentials to use the LMS and teachers can use the tasks as resources to develop their own courses in the LMS.

The analysis framework used in this study was built within the research group and has been used in two peer-reviewed articles [22], [42]. In this study, the analysis was conducted in three stages. In the first stage, the first author performed preliminary analysis by categorizing 400 mathematics tasks using the analysis framework [22], [42]. In the second stage, 120 of the 400 tasks (30%) were randomly selected for analysis by two other authors to enhance the inter-rater reliability of the categorization by Krippedorff's alpha values [43]. These 120 tasks were distributed into 66 elementary school tasks and 54 middle school tasks. Discrepancies in classification were resolved through consensus discussions to strengthen classification principles. Krippendorff's alpha values were calculated for the primary school ($\alpha = 0.83$) and secondary school ($\alpha = 0.90$) task classifications. Both Krippendorff's alpha values considered acceptable ($\alpha > 0.80$) [53]. In the third stage, the first author corrected the results of the preliminary analysis according to the established classification principles, and the analysis was completed. The results are reported in Tables 2 and 3. The tasks were analysed and classified into four categories in following way. Tasks requiring only symbolic representations and successive executions of the same algorithm or calculation to produce solutions fell into the categories of simple or complex procedural tasks. Task that can be solved with a one calculation and only require symbolic representations was classified the category of *simple procedural* tasks (see Table 1 for examples). When task solutions required more complex algorithms, or the execution of several algorithms or calculations was necessary, the task was classified into the category of the *complex procedural* task (see Table 1 for examples).

Tasks requiring connections between multiple representations, not only the symbolic form, fell into the category of simple or complex conceptual tasks. The task was classified into the category of *simple conceptual* tasks if it required identifying and ordering the different stages of the calculation. Typically, in the analysis, the solution required connections between multiple representations, not only symbolic form. (See Figure 5 and 6 for examples). Tasks were identified into the category of *complex conceptual* tasks when the concepts were presented in an open problem-solving format. The emphasis was on making decisions about the concepts or parts needed for the solution, as well as on organizing the connections between the parts and different types of representations to produce the solution (see Figure 7 and 8 for examples).

4. Results

This section first presents the basis for classifying tasks into four categories with the help of examples. After this, research questions 1 (Table 2) and 2 (Table 3) are examined using tabulation. Table 1 shows examples of procedurally weight tasks that emphasize procedural knowledge at Primary and Secondary school levels.

Table1: Examples of procedurally weighted tasks at Primary School and Secondary School

Procedural task types	Example
1. A Simple Procedural Mathematics Task at the Primary School Level	$\frac{3}{3} - \frac{1}{3} =$
2. A Simple Procedural Mathematics Task at the Secondary School Level	$-4-(2) =$
3. A Complex Procedural Mathematics Task at the Primary School Level	$\begin{array}{r} 3,9 \\ -\ 1,1 \\ \hline \end{array}$
4. A Complex Procedural Mathematics Task at the Secondary School Level	Calculate, when $x = 7$. $10 \cdot x + 8$

Table 1 shows that the solution to tasks 1 and 2 is obtained by performing one subtraction operation. Therefore, these two tasks are examples of simple procedural tasks. For task 3 and 4, the solution is obtained by performing more than one calculation operation. Thus, these two tasks are examples of complex procedural tasks.

In each of the tasks below (Figures 1 and 2), generating a solution requires identifying and ordering the stages involved in the calculation. In Figure 1, subtraction calculation skills are not necessarily needed due to the pictorial representation provided. In Figure 2, equations and non-equations need to be recognized and classified. Thus, these two tasks are examples of simple conceptual tasks.

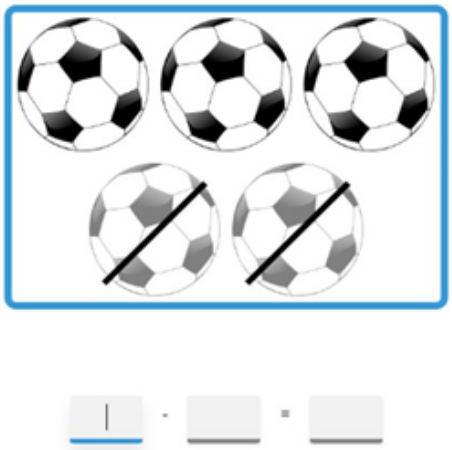


Figure 1: *A Simple Conceptual Mathematics Task at the Primary School Level*

3-(x+1)	x+x	x-(1-x)=5	2x+1
5x	x+13	x+5=10	5x=5

Equations	Non-equations

Figure 2: *A Simple Conceptual Mathematics Task at the Secondary School Level.*

The tasks below (Figures 3 and 4) are both presented in a problem-solving format, and decisions need to be made about the operations required to produce the solutions. Figure 3 requires identifying the rule behind the sequence and producing the solution. In Figure 4, a decision needs to be made about the order and kinds of calculations to be used to produce the solution. Thus, these two tasks represent complex conceptual tasks.

60	57				45	42

Figure 3: *Complex conceptual mathematics task (primary school)*

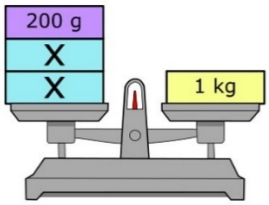


Figure 4. *Solve x. Complex conceptual mathematics task (secondary school)*

Based on the content analysis, 400 electronic mathematical tasks were classified into four categories. As shown in Table 2, 285 tasks emphasized procedural knowledge, and 115 emphasized conceptual knowledge. The greatest percent of tasks were procedurally complex mathematical tasks, and the least percentage of tasks were conceptually complex tasks. The number of tasks emphasizing procedural knowledge was two and a half times the number of tasks emphasizing conceptual knowledge. Interestingly, there were more complex procedural tasks than simple procedural tasks. In contrast, there were almost four times as many simple conceptual tasks compared to complex conceptual tasks.

Table 2. 400 Math Tasks Classified Based on Conceptual and Procedural Knowledge and Difficulty (Simple or Complex)

Task type	Simple	Complex	Total
Procedural	123 (30 %)	162 (41 %)	285 (71 %)
Conceptual	91 (23 %)	24 (6 %)	115 (29 %)

Table 3 shows how the 400 mathematical tasks were classified into four categories based on school level. In comparison to examining the set as a whole, there are several differences. For example, the number of procedural tasks in the primary school is almost triple that of the conceptual tasks. However, that same comparison is only double at the secondary school. Additionally, there are over 8 times as many simple conceptual tasks at the primary level, while a little over twice as many at the secondary level.

Table 3. 400 Math Tasks Classified Based on Conceptual and Procedural Knowledge and Difficulty by School Level (Primary and Secondary)

Task type	Primary school	Secondary school
Simple procedural tasks	76 (35 %)	47 (26 %)
Complex procedural tasks	88 (40 %)	74 (41 %)
Simple conceptual tasks	50 (22 %)	41 (23 %)
Complex conceptual tasks	6 (3 %)	18 (10 %)

5. Conclusions and Discussion

The focus of this study was on analyzing freely available mathematics tasks in the LMS as epistemic affordance of primary and secondary school mathematics based on whether the task emphasized conceptual or procedural knowledge and the difficulty of the tasks. The results of this study showed that the tasks as an epistemic affordance consist of more math tasks emphasizing procedural knowledge than conceptual knowledge at both school levels. The most tasks were categorized to the category of complex procedural tasks at both school levels. It can be noted that when comparing primary school tasks to secondary school tasks, the number of simple procedural tasks decreased, and the number of complex conceptual tasks increased. However, the percentage of complex procedural and simple conceptual tasks were comparable between the two school levels.

This study has some factors related to its reliability and generalizability, which should be considered when interpreting the results and conclusions. First, the research is repeatable using the analysis framework and task banks since the tasks are easily accessible. Second, the randomly selected task sample was large enough (approx. 25% of all tasks), but this must be considered when generalizing the research results. The findings of this study cannot be generalized to all materials obtained from the LMS. It is possible that some other materials in the LMS, such as grade-specific mathematics study paths, contain tasks that emphasize conceptual knowledge more than the materials selected for this study do. Third, the analysis framework used in this study is scientifically reported. In addition, the reliability of this study and classification was confirmed by calculating Krippendorff's alpha values, but it was only done for 120 tasks. In addition, the discriminating ability of the analysis framework and the classification model can be considered good, but placing the task in one of the four categories requires the researcher's own interpretations and decision-making, which requires good theoretical knowledge. Thus, caution should be exercised when generalizing the results of this study.

The results that there are more procedurally weighted tasks than conceptually weighted tasks are nearly similar to the findings results of Finnish mathematics textbooks [19], [40], [41]. Based on the results (Table2), 71 percent of tasks still seemed to focus on procedural knowledge, and complex procedural tasks were the most common tasks at both school levels. It can be concluded that mathematics tasks in the LMS emphasized the learning of procedural knowledge [7], [8], [10]. For this reason, criticisms regarding the materials and their versatility, which emerged based on earlier research, are still relevant. When mathematics tasks from textbooks are transferred to the LMS without adaptation, their conceptual–procedural emphasis does not change, only the format changes. In such a case, the possibilities of the LMS as an epistemic affordance to promote the exploration and manipulation of mathematical concepts to achieve conceptual knowledge will be forgotten [5], [7].

Results reveal that mathematics tasks in LMS also promote the learning of conceptual knowledge at both school levels and this trend can be considered good. Based on the distribution of tasks, the situation compared to the mathematics textbooks seems to have improved in that the primary school LMS and the secondary school LMS appear to integrate slightly more conceptual tasks, but the primary school LMS more closely aligns with the previous textbook studies [19], [40], [41]. The lower number of conceptual tasks in the sample may explain that not all tasks are suitable for automatic checking, and it is easier and faster to perform repetitive tasks aimed at fluency in counting.

From the perspective of teachers and teacher education, this research provides information about each LMS and its potential in the planning, implementation, and evaluation of mathematical tasks. Recommendations for the use of various electronic learning environments and the LMS come from the Finnish National Core Curriculum for Basic Education 2014, which underlines the significant role of ICT in mathematics across all grade levels [44]. For this reason, a variety of technologies are increasingly being used in mathematics classrooms. It is possible to balance epistemic affordance in the LMS by analysing and reviewing the conceptual and procedural emphases of mathematical tasks with the framework used in this study [22], [42].

Differentiating conceptual and procedural knowledge from each other is both a theoretically and methodologically demanding task [14]. The teacher’s knowledge, skills, and abilities in selecting and assigning mathematics tasks in the LMS, utilizing both conceptual and procedural knowledge, become necessary [34]. For teachers to effectively select tasks based on knowledge emphasis, it is crucial for them to recognize how conceptual and procedural knowledge are emphasized in various mathematics tasks [33], [34]. Understanding how conceptual and procedural knowledge is emphasized in different LMS tasks will help current and future teachers select and adapt teaching materials and methods to meet the needs of their students.

It is notable that there are not any components in the two LMS and task repositories to help teachers meet their classroom needs of procedural and conceptual knowledge, but the LMS is flexible and facilitates quick material updates. For this reason, its potential to promote learning of conceptual knowledge can be viewed from two perspectives of epistemic affordance. Firstly, the epistemic affordance of tasks emphasizing conceptual knowledge in the LMS can be increased. Secondly, the existing epistemic affordance of the LMS can be improved by adding information about the classification of the conceptual and procedural knowledge of the tasks.

8. References

[1] Manninen, J., Burman, A., Koivunen, A., Kuittinen, E., Luukannel, S., Passi, S., et al. (2007). *Environments supporting learning: Introduction to learning-environment-thinking.* Helsinki: Finnish National Board of Education. 1

[2] Radcliffe, D. (2008). *A pedagogy-space-technology (PST) framework for designing and evaluating learning places*. In D. Radcliffe, W. Wilson, D. Powell, & B. Tibbetts (Ed.), *Learning spaces in higher education: Positive outcomes by design* (Proceeding of the next generation learning spaces 2008 colloquium; p. 11–16). St Lucia, Queensland: The University of Queensland.

[3] Goodyear, P., Carvalho, L., & Yeoman, P. (2021). *Activity-Centred Analysis and Design (ACAD): Core purposes, distinctive qualities and current developments*. Educational Technology Research & Development, 69, 445–464. https://doi.org/10.1007/s11423-020-09926-7

[4] Carvalho, L., Castaneda, L., & Yeoman, P. (2023). *The "Birth of Doubt" and "The Existence of Other Possibilities": Exploring How the ACAD Toolkit Supports Design for Learning*. Journal of New Approaches in Educational Research, 12(2), 340-. https://doi.org/10.7821/naer.2023J.1494

[5] Akçay, A. O. ., Karahan, E., & Bozan, M. A. (2021). *The Effect of Using Technology in Primary School Math Teaching on Students' Academic Achievement: A Meta-Analysis Study*. FIRE: Forum for International Research in Education, 7(2), 1–21.

[6] Wiest, L. R. (2001). *The role of computers in mathematics teaching and learning*. Computers in the Schools, 17(1-2), 41-55. 10.1300/J025v17n01_05

[7] Kurvinen, E., Dagienė, V. & Laakso, M. (2018). *The Impact and Effectiveness of Technology Enhanced Mathematics Learning*. Constructionism, Computational Thinking and Educational Innovation: Conference Proceedings. Constructionism 2018. Vilna.

[8] Mononen, R., Aunio, P. Väisänen, E., Korhonen, J. & Tapola, A. (2017). *Matemaattiset oppimisvaikeudet*. PS-kustannus.

[9] Chaubey, A., & Bhattacharya, B. (2015). Learning management system in higher education. International Journal of Science Technology & Engineering, 2(3), 158-162

[10] Laakso, M., Kaila, E. & Rajala, T. (2018). *ViLLE – collaborative education tool: Designing and utilizing an exercise-based learning environment*. Education and Information Technologies, 23. 10.1007/s10639-017-9659-1.

[11] Julin, S., & Rautopuro, J. (2016). *Läksyt tekijäänsä neuvovat: perusopetuksen matematiikan oppimistulosten arviointi 9. vuosiluokalla 2015*. Kansallinen koulutuksen arviointikeskus. Julkaisut / Kansallinen koulutuksen arviointikeskus, 2016, 20. https://karvi.fi/app/uploads/2016/04/KARVI_2016.pdf

[12] Hiebert, J., & Lefevre, P. (1986). *Conceptual and procedural knowledge In mathematics: An introductory analysis*. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (p. 1-27). Lawrence Erlbaum.

[13] Rittle-Johnson, B. & Schneider, M. (2015). *Developing Conceptual and Procedural Knowledge of Mathematics*. In R. Cohen Kadosh & A. Dowker (Ed.), *Oxford library of psychology. Oxford handbook of numerical cognition* (p. 1118-1134). Oxford University Press. DOI: 10.1093/ox-fordhb/9780199642342.013.014.

[14] Haapasalo, L. & Kadijevich, D. (2000). *Two types of mathematical knowledge and their relation*. Journal Für Mathematikdidaktik 21 (2), 139-157. 10.1007/BF03338914.

[15] Gilmore, C. K., Keeble, S.,Richardson, S. & Cragg, L. (2019). *The Interaction of Procedural Skill, Conceptual Understanding and Working Memory in Early Mathematics Achievement*. Journal of Numerical Cognition, 3, 400-416. https://doi.org/10.5964/jnc.v3i2.51

[16] Kadijevich, D. M. (2018). *Relating Procedural and Conceptual Knowledge*. Journal Teaching of Mathematics, 21(1), 15–28

[17] Baroody, A. J. (2003). *The development of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge*. In A. J. Baroody & A. Dowker, *The Development of Arithmetic Concepts and Skills: Constructive Adaptive Expertise* (p. 1-33). Lawrence Erlbaum Associates.

[18] Haapasalo, L. (2004). *Pitäisikö ymmärtää voidakseen tehdä vai pitäisikö tehdä voidakseen ymmärtää?* In P. Räsänen, P. Kupari, T. Ahonen & P. Malinen (Ed.) *Matematiikka –näkökulmia opettamiseen ja oppimiseen* (p. 50–83). Niilo Mäki Instituutti.

[19] Joutsenlahti, J. & Vainionpää, J. (2008). *Oppikirja vai harjoituskirja? Perusopetuksen luokkien 1–6 matematiikan oppimateriaalin tarkastelua MOT-projektissa*. In A. Kallioniemi (Ed.) *Uudistuva ja kehittyvä ainedidaktiikka. Ainedidaktinen symposiumi 8.2.2008 Helsingissä*. Tutkimuksia. (p. 547–558). Helsingin yliopisto.

[20] Haapasalo, L. 2011. *Oppiminen, tieto & ongelmanratkaisu*. Medusa-Software.

[21] Phuong, M. T. H. (2019). *On the Procedural-Conceptual Based Taxonomy and Its Adaptation to the Multi-Dimensional Approach SPUR to Assess Students' Understanding Mathematic*. American Journal of Educational Research, 7(3), 212-218. DOI:10.12691/education-7-3-4

[22] Heiskanen, H., Eronen, L., Eskelinen, P., & Väisänen, P. (2021). *Eri tiedonalapainotteiset tehtävätyypit luokanopettajaopiskelijoiden omaehtoisessa matematiikan opiskelussa*. FMSERA Journal, 4(1), 16–30. Retrieved from https://journal.fi/fmsera/article/view/95438

[23] Star, J. R. (2005). *Reconceptualizing procedural knowledge*. Journal for Research in Mathematics Education, 36, 404-411.

[24] Frade, C. & Borges, O. (2006). *The tacit-explicit dimension of the learning of mathematics: an investigation report*. International Journal of Science and Mathematics Education, 4, 293–317. https://doi.org/10.1007/s10763-005-9008-5

[25] Nogueira de Lima, R. & Tall, D. (2008). *Procedural embodiment and magic in linear equations.* Educational Studies in Mathematics, 67. DOI: 10.1007/s10649-007-9086-0.

[26] Tall, D. (2004a). *Introducing three worlds of mathematics*. For the Learning of Mathematics, 23(3), 29-33.

[27] Tall, D. (2004b). *Thinking through three worlds of mathematics*. Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education, 4, 281–288.

[28] Lauritzen, P. (2012). *Conceptual and Procedural Knowledge of Mathematical Functions*. Publications of the University of Eastern Finland Dissertations in Education, Humanities, and Theology, 34. University of Eastern Finland.

[29] Niemi, E. K. (2004). *Perusopetuksen oppimistulosten kansallinen arviointi ja tulosten hyödyntäminen koulutuspoliittisessa kontekstis*sa. Perusopetuksen matematiikan oppimistulosten kansallinen arviointi 6. vuosiluokalla vuonna 2000. Turun yliopiston julkaisuja C 216. Turun yliopisto.

[30] Lehtonen, D. 2022. Now I get it: Developing a Real-World Design Solution for Understanding Equation-Solving Concepts. Kasvatustieteiden ja kulttuurin tiedekunta. Väitöskirja. Tampereen yliopisto. https://trepo.tuni.fi/handle/10024/136918

[31] Heinonen, J.-P. (2005). *Opetussuunnitelmat vai oppimateriaalit – peruskoulun opettajien käsityksiä opetussuunnitelmien ja oppimateriaalien merkityksestä opetuksessa*. Tutkimuksia 257. Helsingin yliopisto.

[32] Hannula-Sormunen, M., Mattinen, A., Räsänen, P. & Ruusuvirta, T. 2018. *Varhaisten matemaattisten taitojen perusta: synnynnäiset valmiudet, tietoinen toiminta ja vuorovaikutus*. In J. Joutsenlahti, H. Silfverberg & P. Räsänen (toim.) *Matematiikan opetus ja oppiminen* (p. 294–305). Niilomäki instituutti.

[33] Eronen, L. (2019). *Quasi-systematic minimalism within socio-constructivist learning of mathematics*. The Electronic Journal of Mathematics and Technology, 13(1), 25–60.

[34] Hurrell, D. P. (2021). *Conceptual knowledge OR Procedural knowledge OR Conceptual knowledge AND Procedural knowledge:Why the conjunction is important for teachers*. Australian Journal of Teacher Education, 46(2). http://dx.doi.org/10.14221/ajte.2021v46n2.4

[35] National Council of Teachers of Mathematics. (2014). Principles to actions: ensuring mathematical success for all. Reston, VA.

[36] National Research Council. (2001). *Adding it up: Helping children learn mathematics*. Washington, DC: National Academy Press.

[37] Boaler, J., & Staples, M. (2008). *Creating mathematical futures through an equitable teaching approach: The case of Railside School*. Teachers College Record, 110(3), 608-645. https://ed.stanford.edu/sites/default/files/boaler__staples_2008_tcr.pdf

[38] Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … & Tsai, Y. M. (2010). *Teach'rs' mathematical knowledge, cognitive activation in the classroom, and student progress*. American educational research journal, 47(1), 133-180. https://doi.org/10.4324/9781410607218

[39] Remillard, J. T. (2005). *Examining key concepts in research on teach'rs' use of mathematics curricula*. Review of educational research, 75(2), 211-246. https://doi.org/10.3102/00346543075002211

[40] Joutsenlahti, J. & Vainionpää, J. (2007). *Minkälaiseen matemaattiseen osaamiseen peruskoulussa käytetty oppimateriaali ohjaa?* In K. Merenluoto, A. Virta, P. Carpelan (toim.) *Opettajankoulutuksen muuttuvat rakenteet: Ainedidaktinen symposium 9.2.2007*. (p. 184–191). Turun yliopiston kasvatustieteiden tiedekunnan julkaisuja B 77.

[41] Viholainen, A., Partanen, M., Piiroinen, J., Asikainen, M. & Hirvonen, P. E. (2015). *The role of textbooks in Finnish upper secondary school mathematics: theory, examples, and exercises*. Nordic Studies in Mathematics Education, 20(3–4), 157–178.

[42] Eronen, L., Eskelinen, P., Heiskanen, H., Juvonen, A., & Väisänen, P. (2022). *Luokanopettajaopiskelijoiden motivaation yhteys matematiikan tehtävien suorittamiseen ViLLE-oppimisympäristössä*. LUMAT: International Journal on Math, Science and Technology Education, 10(1), 319–342. https://doi.org/10.31129/LUMAT.10.1.1731

[43] Krippendorff, K. (2019). *Content Analysis: An Introduction to Its Methodology* (4th Ed.). SAGE Publications. https://doi.org/10.4135/9781071878781Perusopetuksen opetussuunnitelman perusteet 2014. Opetushallitus. Helsinki. https://www.oph.fi/sites/default/files/documents/perusopetuksen_opetussuunnitelman_perusteet_2014.pdf