ADG 2018

Proceedings of the 12th International Conference on Automated Deduction in Geometry

In Memory of Wen–tsün Wu,
Founder of the Chinese School of Mathematics Mechanization

Hongbo Li (Editor)

Proceedings of the 12th International Conference on Automated Deduction in Geometry

Nanning, China, September 11–14, 2018

EDITED BY
Hongbo Li

http://adg2018.cc4cm.org/

Invited Speakers

Jacques Fleuriot
Ilias S. Kotsireas
Dingkang Wang
Xiao-Shan Gao
Pascal Schreck
Chuanming Zong

General Chair

Dongming Wang

Program Committee Chair

Hongbo Li

Program Committee

Francisco Botana
Xiaoyu Chen
Lei Huang
Tetsuo Ida
Na Lei
Julien Narboux
Tomás Recio
Thomas Sturm
Chee K. Yap
Falai Chen
Yong Feng
Hoon Hong
Predrag Janičić
ZhuoJun Liu
Pedro Quaresma
Ileana Streinu
Dingkang Wang

Local Arrangements

Xiaoyu Chen
Jing Yang

Webmaster & Publicity Chair

Fadoua Ghourabi

Sponsors of ADG 2018

- SMS International, Guangxi University for Nationalities
- Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis

Invited Talks

Interactive Theorem Proving in Geometry: From Foundations to Applications
Jacques D. Fleuriot 1

Wen-Tsün Wu and Mathematics Mechanization
Xiao-Shan Gao 2

Hard Combinatorial Problems via SAT
Ilias S. Kotsireas 3

Around Dandelin-Gallucci Theorem
Pascal Schreck 4

Comprehensive Gröbner Systems and Discovering Geometric Theorems Mechanically
Dingkang Wang 5

Can You Pave the Plane Nicely with Identical Tiles
Chuanming Zong 6

Regular Papers

Automated Geometer, a Web-based Discovery Tool
Francisco Botana, Zoltán Kovács and Tomás Recio 7

Computation of GCD Chain over the Power of an Irreducible Polynomial
Xavier Dahan 14

Geometric Search in *TGTP*
Yannis Haralambous and Pedro Quaresma 19

Mining Geometry Theorems in a Regular Polygon
Zoltán Kovács 26

Towards a Mechanisation in Isabelle of Birkhoff's Ruler and Protractor Geometry
Imogen I. Morris and Jacques D. Fleuriot 46

Mechanising an Independent Axiom System for Minkowski Space-time
Jake Palmer and Jacques D. Fleuriot 64

n-dimension Area Method Using Exterior Products
Vincent Pavan 80

User Interface and Operations of GeometryTouch on Small and Large Touchscreens
Wei Su, Chuan Cai and Jinzhao Wu 93

A System for Automated Deduction in Engineering Mechanics
Philip Todd 97

On *n*-sectors of the Angles of an Arbitrary Triangle
Dongming Wang, Bo Huang and Xiaoyu Chen 117

Interactive Theorem Proving in Geometry: From Foundations to Applications

Jacques Fleuriot

University of Edinburgh, UK

Abstract. In this talk, I'll argue that interactive theorem proving is an effective tool for the systematic investigation of geometric problems, ranging from axiomatic foundations to formal verification. This type of mechanization is usually carried out within the settings of proof-assistants such as Isabelle, which provide both a rich language for formalizing non-trivial concepts, e.g. higher-order geometric axioms or even inductive definitions, and an array of powerful automated tools, e.g. first order theorem provers and decision procedures, that can help the user in their quest for a (readable) proof.

I will illustrate the above by discussing some of the achievements from the past twenty years, a number of which were originally presented at Automated Deduction in Geometry. Time permitting, I'll also talk about some potential mechanization challenges and avenues for collaborative proof efforts in geometry.

Wen-tsün Wu and Mathematics Mechanization

Xiao-Shan Gao

AMSS, Chinese Academy of Sciences, China

Abstract. In this talk, I will give a review of some of the major advances in the field of mathematics mechanization coined by Wen-tsün Wu. These include the Ritt-Wu characteristic set method for symbolic solution of algebraic, differential, and difference polynomial equation systems; methods for automated proving and discovering geometry theorems; and applications in computer aided geometric design, computer vision, robotics, etc.

Hard Combinatorial Problems via SAT

Ilias S. Kotsireas

Wilfrid Laurier University, Canada

Abstract. The area of boolean satisfiability and SAT solvers has seen dramatic advances in the past two decades. A recent trend in SAT solving is an attempt to combine the strengths of symbolic computation tools with the power of SAT solvers, in order to improve their effectiveness and to build custom-tailored SAT solvers for hard combinatorial problems. We will describe our work in this context, with a focus on some particularly hard combinatorial problems, described via autocorrelation of finite sequences.
Based on joint work with Vijay Ganesh (University of Waterloo) and Curtis Bright (University of Waterloo) in the context of the Horizon 2020 EU project "Satisfiability Checking and Symbolic Computation" (SC^2).

Around Dandelin-Gallucci Theorem

Pascal Schreck

University of Strasbourg, France

Abstract. This talk is about mechanization of projective incidence geometry. I present some parts of a work which began in 2004 with a proposal of Dominique Michelucci about the so-called hexamys and a study around combinatorial proofs in incidence geometry. Most of this presentation relates to joint works with other members of the Strasbourg team and especially David Braun and Nicolas Magaud. After a brief description of projective incidence geometry, and after recalling the importance of both Desargues and Pappus theorems, I explain how projective incidence geometries correspond to a certain class of matroids. Then, I present Dandelin-Gallucci theorem with several proofs using very different approaches: combinatoric algebra, synthetic geometry, matroids, ...

Comprehesive Gröbner Systems and Discovering Geometric Theorems Mechanically

Dingkang Wang

AMSS, Chinese Academy of Sciences, China

Abstract. For many geometric theorems, the hypotheses can be represented by a system of parametric polynomial equations and the conclusion can be represented by a parametric polynomial equation. An important problem concerning proving geometric theorems is to determine whether a geometric statement is valid under a specialization of parameters. Comprehensive Gröbner system is an important tool to solve the problem related to parametric polynomial system. We will review the progresses in comprehensive Gröbner systems and then use it to discover geometric theorems mechanically, i.e., we can find out complementary conditions on the parameters such that the geometric statement becomes true or true on components.

Can You Pave the Plane Nicely with Identical Tiles

Chuanming Zong

Tianjin University, China

Abstract. Everybody knows that identical regular triangles or squares can tile the whole plane. Many people know that identical regular hexagons can tile the plane properly as well. In fact, even the bees know and use this fact! Is there any other convex domain which can tile the Euclidean plane? Of course, there is a long list of them! To find the list and to show the completeness of the list is a unique drama in mathematics, which has lasted for more than one century and the completeness of the list has been mistakenly announced not only once! Up to now, the list consists of triangles, quadrilaterals, fifteen types of pentagons, and three types of hexagons. In 2017, Michael Rao announced a computer proof for the completeness of the list. Meanwhile, Qi Yang and Chuanming Zong made a series of unexpected discoveries in multiple tilings in the Euclidean plane. For examples, besides parallelograms and centrally symmetric hexagons, there is no other convex domain which can form any two-, three- or four-fold translative tiling in the plane. However, there are two types of octagons and one type of decagons which can form nontrivial five-fold translative tilings. Furthermore, a convex domain can form a five-fold translative tiling of the plane if and only if it can form a five-fold lattice tiling. In this talk we will report these progresses.

Automated Geometer, a Web-based Discovery Tool (extended abstract)

Francisco Botana[1], Zoltán Kovács[2], and Tomás Recio[3]

[1] Dept. of Applied Mathematics I, University of Vigo
Campus A Xunqueira, 36005 Pontevedra, Spain
fbotana@uvigo.es
[2] The Private University College of Education of the Diocese of Linz
Salesianumweg 3, 4020 Linz, Austria
zoltan@geogebra.org
[3] Universidad de Cantabria
Avenida de los Castros, s/n, 39071 Santander, Spain
tomas.recio@unican.es

Abstract. The goal of this communication to ADG 2018 is to report on-going work by the authors towards the automated and systematic finding of properties on a given geometric construction. Our *Automated Geometer* is being implemented on top of *GeoGebra*, of a dynamic geometry system with millions of users at high schools and universities. It exploits *GeoGebra* recently added functionalities regarding automated reasoning tools, providing rigorous, symbolic-driven, proofs of geometric facts. In the talk we will illustrate and describe some basic facts about the system we are developing. An expanded version of this document has been submitted to the AISC 2018 Conference.

Keywords: Automated discovery · Automated theorem proving · C-omputer algebra · *GeoGebra*

1 Problem

Half a century ago Lenat's *AM* [18] introduced a rule based system able to successfully discover (or rediscover) non-trivial mathematical results in number theory. It tried to replicate human approach for "doing mathematics". Our aim, somehow similar, but in the realm of automated discovery in geometry, has been inspired by the strategy reported in [13, p. 44]. Roughly, it consists on using automatic reasoning tools for checking mechanically produced statements involving some elements of a geometric construction, both concerning those elements that are actually present in the given construction and those elements that could be automatically generated by the program, or added by the user, starting from the input data.

For example, assume we are given a triangle and a point on its plane. Then we would like the system to develop some elementary operations between the given point and the vertices (or sides) of the triangle. These operations, such as drawing perpendicular lines, adding midpoints of sides and lines from vertices to midpoints, ..., could be imposed by the user or suggested somehow by the system. Finally, the goal is to have a program that will systematically and autonomously verify the truth/failure of mechanically produced statements concerning collinearity, parallelism, ... of the different elements, given or generated, in the final construction.

Moreover, we would like our system to be built on top of *GeoGebra*, exploiting its recent abilities on automatic reasoning tools *ART* in geometry [4]. The *Automated Geometer*, *AG*, (also meaning *Amateur Geometer*) intends to be a *GeoGebra* module where pure automatic discovery is performed. It includes a generator of further geometric elements from those of a given construction, and a set of rules for producing conjectures on the whole set of elements, that would be decided by the already present automatic proving tools.

Finally, let us add that our ultimate goal is not just accomplishing this automatic and systematic exploration of the space of possible conjectures, but attempting to moderate its foreseeable combinatorial explosion, by mimicking human thought when doing geometry, as described by the former president of ICMI, Miguel de Guzmán in [13]. Unfortunately, we can not report, yet, on this more ambitious goal.

2 Motivation

We consider that *GeoGebra*'s automated reasoning capabilities can help students to do mathematics better or faster, just as we think it is beneficial to have an electronic calculator to compute the square root of a number much faster than using the traditional, mechanical method by hand (which, not surprisingly, is no longer part of most curricula).

Moreover, we think that the mere existence and availability of a well disseminated, easily accessible *Automated Geometer*, could mean a drastic change in the teaching and learning of school geometry: human and machine collaborative settled to explore, jointly, the geometric context. In this way we could argue that the use of *AG* is not just to do the same kind of mathematics better and easier, but to do, in some sense, "a better kind of mathematics". Let us borrow Kaput's visionary words, cited by Balacheff: instead of *Doing (old) Things Better* we should focus on *Doing Better Things* [3].

Several didactical reflections, and the analysis of some on-going experiences, concerning the classroom use of *AG* can be consulted in [17], [21], [14].

Other motivations could be the use of *AG* for the analysis of geometric properties of real life objects (say, for automatically getting augmented reality on math trails, with the help of a Hough transform, see [5]) and the possible role of *AG* to dictate geometric properties of objects to blind persons.

3 State of the Art

This line of work concerning automated discovery (i.e. finding statements holding in a given figure) in geometry was initiated, to the best of our knowledge, in [2]. There, the authors developed a system with a generator of constructions where a systematic search is performed to find new conjectures which are then proved through Wu's algebraic method. A related proposal, but this time using fixpoint reasoning and deductive database methods, able to discover all properties of a construction given a set of rules, was developed in the program Geometry Expert [11], [10], [8]. Finally, a report on discovering properties from scanned images has been described in [7]. Some strategies are used to generate conjectures involving the image translation to a geometric figure, and algebraic computations return their truth or falsity.

4 Contribution and Main Idea

In our contribution the main idea is to use *GeoGebra*'s Automated Reasoning Tools [16] extended by automatized JavaScript code inside a web browser.

GeoGebra, from its first versions, incorporates a *Relation* tool that returns the results of some basic checks (for instance, incidence, parallelism, perpendicularity, equal length, ...) between a pair of selected elements. This command is not exclusive of *GeoGebra*: it also existed in previous dynamic software like *Cabri*, but till the inclusion of automated reasoning tools all these approaches in widespread environments were based on numerical checking. A paradigmatic example of this numerical checking is *OK Geometry* [19]: this tool detects relevant facts in a construction by slightly moving its free points, checking which relations among them remain then invariant, and filtering the results through a library of well-known properties.

The *GeoGebra ART* module does not perform numerical checking. Rather, all facts are symbolically managed by means of the `Prove` command [4]. The module runs on modern browsers, thus providing universal accessibility, and it is controlled by the Javascript API [12].

Currently, our on-going Automated Geometer AG program is already able to accept a user defined construction (that could be also the result of loading a preexistent one) and it searches for meaningful relations between the construction elements. All possible relations are listed on a combinatorial basis, and those classified as *generally true* by the prover algorithm are returned. So, AG only outputs certified true properties in constructions.

Furthermore, since it is built on top of *GeoGebra*, it can reach a millionaire audience in mathematics students. This fact could be, indeed, the most differential characteristic regarding previous work on the same direction.

From the technical perspective we highlight that the AG module currently runs in a web browser and it implicitly uses a precompiled version of the *Giac* [15] computer algebra system as a piece of JavaScript or WebAssembly [6] code. To our knowledge, these kinds of technologies ensure the users the quickest performance to obtain results on heavy computations in a popular, user-friendly way.

For instance, for implicitly compute Gröbner bases of ideals in several variables, or to eliminate some variables from a given ideal, two time and space consuming operations that are required to run the algorithm in the `Prove` command in our method, since it is based on some algebraic geometry approach to automatic proving, see [20].

5 Examples

The AG module can be freely tested at `http://htmlpreview.github.io/?https://github.com/kovzol/ag/blob/master/automated-geometer.html`, and its development is shown at `https://github.com/kovzol/ag`.

Figure 1 shows part of the default screen of the web application. There, a simple construction is displayed, involving three user-defined points A, B, C, their midpoints D, E, F and a fourth midpoint G between D and E. The user is requested to select, among a list of possible choices, the type of generic properties to be tested. Currently, the choices are: collinearity or equality of distances between three points, and perpendicularity or parallelism of segments defined by two points.

In Figure 2 the two free points A and B are given, and the midpoint C of the segment determined by them, and the circle centered at C and going through one of the free points, finally a semifree point D in the this circle. AG discovers that the AD and BD are perpedicular.

In Figure 3 three free points A, B, C are given, then parallels to AB through C and to AC through B, intersecting at D; finally E is the intersection of lines AD and BC. The AG tool has discovered, for instance, that E is the midpoint of the diagonals.

Acknowledgement

Authors partially supported by the grant MTM2017-88796-P from the Spanish MINECO and the ERDF (European Regional Development Fund).

References

1. Abánades, M., Botana, F., Kovács, Z., Recio, T., Sólyom-Gecse, C.: Development of automatic reasoning tools in GeoGebra. ACM Communications in Computer Algebra 50 (2016) 85–88
2. Bagai, R., Shanbhogue, V., Zytkow, J.M., Chou, S.C.: Automatic theorem generation in plane geometry. Lecture Notes in Artificial Intelligence 689 (1993) 415–424
3. Balacheff, N. ICME 8, TG19 Followup Report. Computer-Based Learning Environments: "CBILE",`http://mathforum.org/mathed/seville/followup.html`, 1997.
4. Botana, F., Hohenwarter, M., Janicic, P., Kovács, Z., Petrovic, I., Recio, T., Weitzhofer, S.: Automated theorem proving in GeoGebra: Current achievements. Journal of Automated Reasoning 55 (2015) 39–59

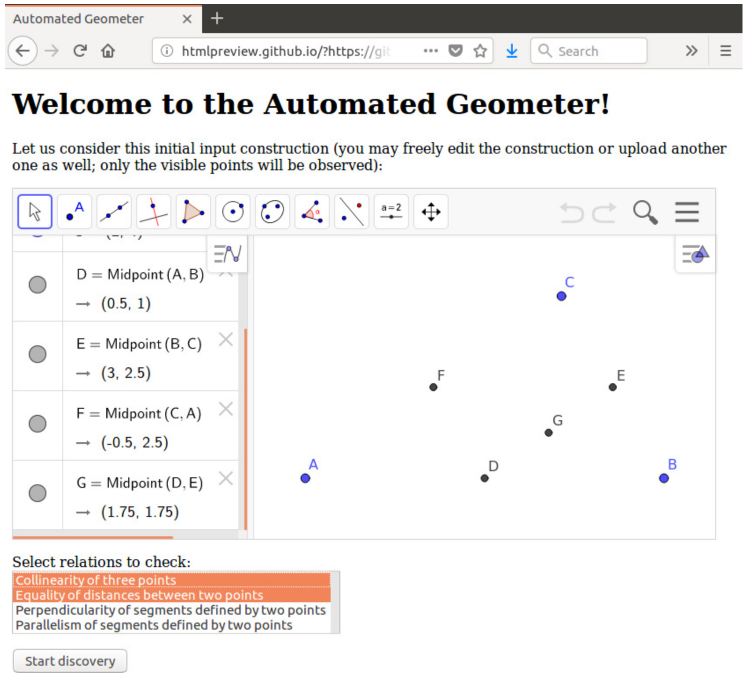


Fig. 1. The default construction in Automated Geometer. By default, on a simple run, Automated Geometer returns the following 16 true statements: $D \in AB, F \in AC, E \in BC, G \in BF, G \in DE, AD = BD, AD = EF, AF = CF, AF = DE, BD = EF, BE = CE, BE = DF, BG = FG, CE = DF, CF = DE, DG = EG$.

5. Botana, F., Kovács, Z., Martínez-Sevilla, A., Recio, T.: Automatically augmented reality with GeoGebra, *to appear*
6. Bright, P.: The Web is getting its bytecode: WebAssembly. Condé Nast (2015)
7. Chen, X., Song, D., Wang, D.: Automated generation of geometric theorems from images of diagrams. Annals of Mathematics and Artificial Intelligence 74 (2015) 333–358
8. Chou, S.C., Gao, X.S., Zhang, J.Z.: A deductive database approach to automated geometry theorem proving and discovering. Journal of Automated Reasoning 25 (2000) 219–246
9. Gao, H., Goto, Y., Cheng, J.: A set of metrics for measuring interestingness of theorems in automated theorem finding by forward reasoning: A case study in NBG set theory. Lecture Notes in Computer Science 9243 (215) 508–517
10. Gao, X.S., Zhang, J.Z., Chou, S.C.: Geometry Expert (in Chinese), Nine Chapters, Taipei, 1998.
11. Geometry Expert. `http://www.mmrc.iss.ac.cn/gex/`
12. The GeoGebra Team: Reference: GeoGebra Apps API. `https://wiki.geogebra.org/en/Reference:GeoGebra_Apps_API`, accessed 5/10/2018
13. de Guzmán, M.: La experiencia de descubrir en geometría. Nivola, Madrid (2002).

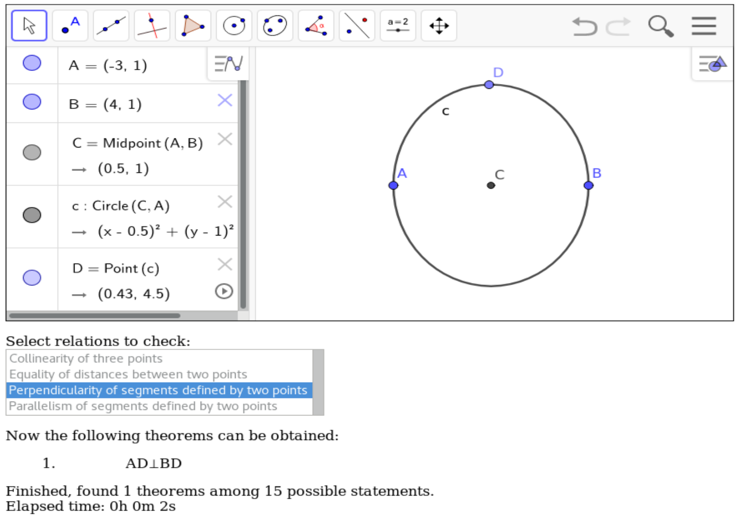
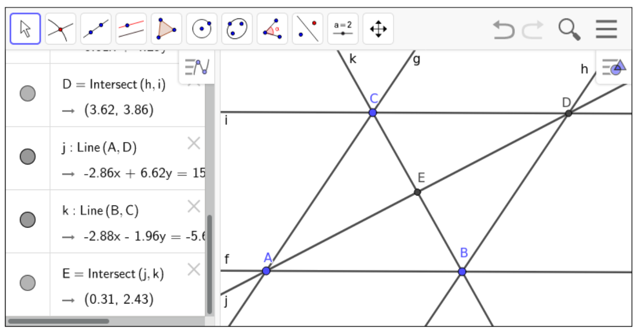


Fig. 2. *AG* discovers Thales' circle theorem

14. Hohenwarter, M., Kovács, Z., Recio, T.: Using GeoGebra Automated Reasoning Tools to visualize and explore geometric statements and conjectures. *to appear*
15. Kovács, Z., Parisse, B.: Giac and GeoGebra – Improved Gröbner basis computations. Lecture Notes in Computer Science 8942 (2015) 126–138
16. Kovács, Z., Recio, T., Vélez, M.P.: GeoGebra Automated Reasoning Tools. A tutorial. `https://github.com/kovzol/gg-art-doc/` (2017)
17. Kovács, Z., Recio, T., Richard, P., Vélez, M.P.: GeoGebra Automated Reasoning Tools: A tutorial with examples. Communication at the 13th International Conference on Technology in Mathematics Teaching – ICTMT 13, Lyon, France, 2017. `https://ictmt13.sciencesconf.org/148350/document`
18. Lenat, D.B.: Automated theory formation in mathematics. Contemporary Mathematics 29 (1984) 287–314
19. Magajna, Z.: OK Geometry. `http://z-maga.si/index?action=article&id=40`, accessed 5/3/2018
20. Recio, T., Vélez, M.P.: Automatic discovery of theorems in elementary geometry. Journal of Automated Reasoning 23 (1999) 63–82
21. Recio, T., Richard, P., Vélez, M.P.: Designing tasks supported by GeoGebra Automated Reasoning Tools for the development of mathematical skills. Communication to the Conference on Digital Tools in Mathematics Education (7th CADGME 2018). Coimbra, Portugal. 2018. `https://www.uc.pt/en/congressos/cadgme2018/generalinformation/listTalks`



Select relations to check:

Collinearity of three points
Equality of distances between two points
Perpendicularity of segments defined by two points
Parallelism of segments defined by two points

Now the following theorems can be obtained:

1. E∈AD 2. E∈BC 3. AB=CD 4. AC=BD 5. AE=DE 6. BE=CE

Finished, found 6 theorems among 55 possible statements.
Elapsed time: 0h 0m 6s

Fig. 3. *AG* discovers some simple theorems.

Computation of GCD Chain over the Power of an Irreducible Polynomial (extended abstract)

Xavier Dahan[1]

Ochanomizu university, Faculty of General Educational Research & Dept. of Mathematics
xdahan@gmail.com

Abstract. A notion of gcd chain has been introduced by the author at ISSAC 2017 for two univariate monic polynomials with coefficients in a ring $R = k[x_1, \ldots, x_n]/\langle T \rangle$ where T is a *primary* triangular set of dimension zero. A complete algorithm to compute such a gcd chain remains challenging. This work treats the case of a triangular set $T = (T_1(x))$ in one variable, namely a power of an irreducible polynomial. This seemingly easy case is nonetheless essential for considering the general case

1 Introduction

Computing gcd is without a doubt one of the most fundamental algorithm in computer algebra and computational aspects have been continuously studied extensively, till today. In [2] is introduced the concept of *gcd chain* to bring a similar notion of the classical gcd of polynomials of one variable over a field, to the case of over a ring $k[x_1, \ldots, x_n]/\langle T \rangle$ where T is a primary triangular set of dimension zero. Such a ring has nilpotent elements, and non nilpotent elements are invertible. Some attempts to treat this case in prior [2] have concluded in somewhat unsatisfactory solutions. Indeed for polynomials of one variable over an ideal that is "almost" maximal, here a primary ideal, a desirable fundamental property of gcd is an ideal equality $\langle a, b \rangle = \langle g \rangle$. While if a and b have coefficients in a ring of type $R = k[x_1, \ldots, x_n]/\langle T \rangle$ where T is a *primary* triangular set, it is well-known that a polynomial g does not exist in general, the outcome of [2] being to present a strategy to circumvent this impediment by "iterating" somehow an Euclidean Remainder Sequence. On the algorithmic front, this raises several challenging questions, even in the seemingly "easy" case of a primary triangular set $T = (T_1(x_1))$ of *one* variable. As shown in this extended abstract, this case is already not simple. And it is important since it builds the framework to tackle the case of several variables, which likely uses the same ingredients.

2 Motivation

An early motivation in computing gcds over triangular sets come from the *triangular-decomposition* algorithm to solve polynomial (commutative or differential) systems [10,1]. This set of computational methods traces back to the early work of Ritt [9], and the major computational advances realized later by Wu-Wen Tsu [11]. This has lead to several new directions of researches, followed by many researchers. In term of algorithms, only *pseudo-divisions* were initially used. In 1993, [4] Kalkbrener introduced a "gcd"-point of view to realize the decomposition, and the elimination (See also the notes [3]). This point of view has later been significantly developed by M. Moreno-Maza in particular with the implementation of the library `RegularChains` in the software Maple.

However, such a gcd does not handle "faithfully" polynomials having multiplicities; this question was raised as early as 1995 [7] and later studied furthermore in [6], but without a satisfactory general answer. In this regard, the present work situates in the realm of triangular decomposition as initiated by Wu-Wen Tsu.

The gcd chain has the following **geometric** interpretation. The underlying triangular set can be thought as modeling some algebraic constraints, over which one may want to compute further data modeled by polynomials, that is over the solutions of the constraints *only*. It may happen that the solution (a constraint), is *multiple*. One may think of $t_1(x) = x^3$ for example, in the case where constraints are modeled by a polynomial of one variable like in this work. When computing *over* t_1, this allows to consider Taylor expansions at order 2 (for example to control the first and second order derivative of the state as well).

In Example 1 below, we want to compute the constraints defined by both polynomials a and b with coefficients in $\mathbb{R}[x]/\langle t_1 \rangle = \mathbb{R}[x]/\langle x^3 \rangle$. We obtain three cases, as computed by the algorithm of this paper.

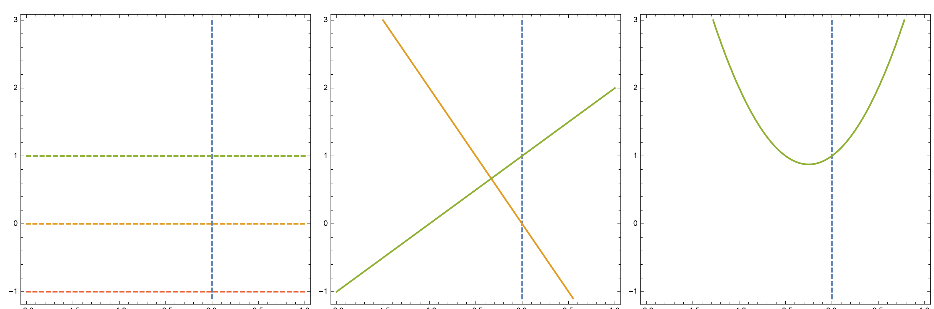


Fig. 1. Precision x (left): Three points intersection of $y = 0, -1, 1$ with $x = 0$. Precision x^2 (middle): Two lines, expanded from $y = 0, -1$. Precision x^3 (right): One parabola, expanded from $y = 1$

3 Definitions

A primary triangular set in one variable is just a power of an irreducible polynomial p: $T = (T_1(x)) = (p(x)^e)$. Then the ring $R = k[x]/\langle T \rangle$ is local of maximal ideal $\mathfrak{m} = \langle p \rangle$.

Definition 1. *Given two monic polynomials $a, b \in R[y]$, a* gcd chain *of a, b is a sequence $(g_i, p^{e_i})_{i=1,\dots,s}$ such that:*

- $e_1 < \ldots < e_s \le e$ *and* $\deg_y(g_1) > \cdots > \deg_y(g_s)$.
- g_{i+1} *divides* g_i *modulo* p^{e_i}.
- *defining* $G_i := g_i/g_{i+1}$ *all* $i = 2, \ldots, s$ *and* $G_1 := g_1$, *the isomorphism holds:*

$$R[y]/\langle a, b\rangle \simeq (k[x]/\langle p^{e_1}\rangle)[y]/\langle G_1\rangle \times \cdots \times (k[x]/\langle p^{e_s}\rangle)[y]/\langle G_s\rangle \qquad (1)$$

where the r.h.s is a direct product of rings.

We reproduce the example of [2, Ex. 3.5, Ex. 5.2].

Example 1. Let $p(x) = x$, $T = (T_1(x)) = (x^3) = (p^3)$. Define the two monic polynomials a and b as follows:

$$\begin{aligned} a &= y^4 + \left(2x^2 + 3x + 1\right) y^3 + \left(-x^2 - x - 1\right) y^2 \\ &\quad + \left(-13x^2 - 4x - 1\right) y - 7x^2 - 2x \\ b &= y^3 + (3x^2 + 3x)y^2 + (-3x^2 - 3x - 1)y - 10x^2 - 2x \end{aligned}$$

Then a gcd chain is given by:

$$[((y-1)y(y+1), x)\ ,\ ((y-1-x)(y+2x), x^2)\ ,\ (y-1-x-2x^2, x^3)],$$

and yields the following isomorphism according to (1).

$$\begin{aligned} (k[x]/\langle x^3\rangle)[y]/\langle a, b\rangle \simeq (k[x]/\langle x\rangle)[y]/\langle y+1\rangle \times (k[x]/\langle x^2\rangle)[y]/\langle y+2x\rangle \\ \times (k[x]/\langle x^3\rangle)[y]/\langle y-1-x-2x^2\rangle \quad (2) \end{aligned}$$

However, Section 5 of [2] dealing with algorithms is more an indication of directions for future work, than a complete and definitive exposition. This is what the present work does, treating the case of one variable completely.

4 Results

The main outcome is the following main routine, in particular the introduction of Weierstrass factorization at Lines 7 and 10, as well as Hensel lifting at Line 11

The gcd chain Algo. 2 simply iterates over the main subroutine "largest-Factor" (Algo. 1) to compute the gcd chain as defined in Definition 1. Indeed, at each iteration, the algorithm outputs one block of the gcd chain (line 17) as shown in Equality (1), starting with the block (g_1, p^{e_1}) (e_1 the smallest precision, $\deg(g_1)$ the largest degree gcd in the gcd chain). The novelty brought compared to Section 5 of [2] is:

- The introduction of (a variant of) Weierstrass preparation theorem [5, Theorem 9.2] to get rid of the overoptimistic Assumption (C) in [2, page 115]. In the output at Line 7, the polynomials A is monic and verify $\langle A, T\rangle = \langle S_{r_j}, T\rangle$. At Line 10, we have similarly $\langle B', T/p^{e_1}\rangle = \langle S, T/p^{e_1}\rangle$.

Input: Monic polynomials a, b with $\deg_y(a) \geq \deg_y(b)$, power of an irreducible polynomial $T = (p(x)^e)$.
Output: p^{e_1}, A, B where A is monic and $\langle A, p^{e_1}\rangle = \langle a, b, p^{e_1}\rangle$;
and $B =$"*end*" ; or $B \in k[y]$ monic, $\deg_y(A) \geq \deg_y(B)$

1 **if** $b = 0$ **then** **// Finished: No iteration necessary**
2   **return** $T, a,$"*end*"
3 **if** $\deg(b) = 0$ **then** **// Finished: No iteration necessary**
4   **return** $T, b,$"*end*"
5 Compute an extended subresultant sequence modulo T: $\mathbf{S} = [S_{r_0} = a, S_{r_1} = b, S_{r_2}, \dots, S_{r_t}]$, and $[U_{r_1}, V_{r_1}], \dots, [U_{r_t}, V_{r_t}]$, $\quad U_{r_\ell} a + V_{r_\ell} b \equiv S_{r_\ell} \mod \langle T\rangle)$
6 $j \leftarrow \mathrm{indexLastNonNil}(\mathbf{S}, T)$ **//** S_{r_j} **is the last non-nilpotent subresultant**
7 $A \leftarrow \mathrm{WeierstrassMonic}(S_{r_j}, T)$ **// Put** S_{r_j} **in monic form**
8 $p^{e_1} \leftarrow \mathrm{nilpotentFactor}(S_{r_{j+1}}, p^{\ell})$ **// Extract nilpotent part**
9 $S \leftarrow S_{r_{j+1}}/p^{e_1}$ **//** S **is no more nilpotent**
10 $B' \leftarrow \mathrm{WeierstrassMonic}(S, T/p^{e_1})$ **// Put** S **in monic form**
11 $B \leftarrow \mathrm{HenselLift}(S, A, b, T, p^{e_1})$ **// Recover precision loss**
12 **return** p^{e_1}, A, B

Algo **1:** The Largest Common Factor (largestFactor)

Input: Power of an irreducible polynomial $T = (p(x)^e)$
Univariate polynomials a and b in $R[y]$ where $R = k[x]/\langle p^e\rangle$
Output: gcd chain $[(g_1, p^{e_1}),\ (g_2, p^{e_2}), \dots,\ (g_s, p^{e_s})]$

13 $B_{next} \leftarrow 0$; $\quad A_{next} \leftarrow 0$; $\quad T' \leftarrow T$
14 $C \leftarrow [\,]$
15 **while** $B_{next} \neq$ "*end*" **do** **// Algo. 1 ended Line 2 or 4**
16     $T', A_{next}, B_{next} \leftarrow \mathrm{largestFactor}(a, b, T)$ **// Iteration of Algo. 1**
17     $C \leftarrow C$ cat $[(A_{next}, T')]$ **// Add the block** (A_{next}, T') **to the gcd chain**
18     $a \leftarrow A_{next}$; $\quad b \leftarrow B_{next}$ **// Update values for the iteration**
19 **return** C

Algo **2:** The gcd-chain algorithm

- The use of Hensel lifting (Line 11) in order to recover precision loss entailed by the division by the nilpotent part p^{e_1} at Line 9. Without this step, the result would be computed at precision p^{e-e_1} instead of p^e.
- A complete proof of correctness.

As already mentioned, each of these steps is a non trivial improvement above [2] that is necessary to tackle the more difficult case of a triangular set T in several variables, hence the case of one variable, although apparently artificial, is essential.

A related work [8] computes a truncated resultant, which is a resultant of a and b in $(k[x]/\langle x^e\rangle)[y]$. There are some similarities with the algorithm presented above, in the case where the irreducible polynomial $p(x) = x$, but there are also fundamental distinctions. The first one being that treating $T(x) = p(x)^e$ makes the overall computation more difficult. Second, the resultant is a well-defined object which has received considerable attention since decades, whereas the gcd chain is more subtle, requiring more care. Finally the algorithm presented there is focused on improving an asymptotic complexity, not giving a practical implementation. The algorithm in [8] given is based on the half-gcd, being notoriously difficult to implement, yet alone to make timings reflect the theoretical running-time complexity. Here, our aim is to devise as simple as possible routines to plan their extension to the case a triangular set in several variables.

References

1. C. Chen and M. Moreno-Maza. Algorithms for computing triangular decomposition of polynomial systems. *Journal of Symbolic Computation*, 47(6):610–642, 2012.
2. Xavier Dahan. Gcd modulo a primary triangular set of dimension zero. In *Proceedings of the 2017 ACM on International Symposium on Symbolic and Algebraic Computation*, ISSAC '17, pages 109–116, New York, NY, USA, 2017. ACM.
3. É. Hubert. Notes on triangular sets and triangulation-decomposition algorithms. I. Polynomial systems. volume 2630 of *Lecture Notes in Comput. Sci.*, pages 1–39. Springer, Berlin, 2003.
4. M. Kalkbrener. A generalized Euclidean algorithm for computing triangular representations of algebraic varieties. *J. Symbolic Comput.*, 15(2):143–167, 1993.
5. Serge Lang. *Algebra (revised third edition)*, volume 211 of *Graduate Texts in Mathematics*. Springer Science and Media, 2002.
6. X. Li, M. Moreno-Maza, and W. Pan. Computations modulo regular chains. In *Proceedings of the 2009 international symposium on Symbolic and algebraic computation*, pages 239–246. ACM, 2009.
7. M. Moreno Maza and R. Rioboo. Polynomial gcd computations over towers of algebraic extensions. In *Proc. AAECC-11*, pages 365–382. Springer, 1995.
8. Guillaume Moroz and Éric Schost. A fast algorithm for computing the truncated resultant. In *Proceedings of the ACM on International Symposium on Symbolic and Algebraic Computation*, pages 341–348. ACM, 2016.
9. J. F. Ritt. Differential equations from an algebraic standpoint. *Colloquiium publications of the AMS*, 14, 1932.
10. D. Wang. *Elimination Methods*. Texts & Monographs in Symbolic Computation. Springer Vienna, 2012.
11. W. T. Wu. On zeros of algebraic equations - an application of Ritt principle. *Kexue Tongbao*, 5:1–5, 1986.

Geometric Search in *TGTP*

Yannis Haralambous[1], Pedro Quaresma[2]

[1] IMT Atlantique Bretagne-Pays de la Loire, Computer Science Department
UMR CNRS 6285 Lab-STICC, Technopôle Brest-Iroise CS 83818,
29238 Brest Cedex 3, France
`yannis.haralambous@imt-atlantique.fr`

[2] CISUC / Department of Mathematics, University of Coimbra
Apartado 3008, EC Santa Cruz, 3001-501 Coimbra, Portugal
`pedro@mat.uc.pt`

Abstract. In this age of information the importance of retrieve the knowledge from the many sources of information is paramount. In Geometry, apart from textual approaches, common to other areas of mathematics, there is also the need for a geometric search approach, i.e. semantic searching in a corpus of geometric constructions.
The Web-based repository of geometric problems Thousands of Geometric problems for geometric Theorem Provers (*TGTP*) has, from the start, some text search mechanisms. Since version 2.0 an implementation of the geometric search mechanism is integrated in it. Using a dynamic geometry system it is possible to build a geometric construction and then semantically search in the corpus for geometric constructions that are super-sets of the former, with regard to geometric properties.
It should be noted that this is a semantic check, the selected constructions may not look like the query construction, but they will possess similar sets of geometric properties.

Keywords: Geometric automated theorem proving · Repository of geometric problems · Common formats · Geometric search · Conceptual graphs · Typed sub-graph isomorphism

1 Introduction

Having repositories of information, one of the first question to solve is how to browse the information contained within. Regarding repositories of geometric information we should add to the usual text searches, geometric searches, i.e. we should be able to provide a geometric construction and look for similar constructions.

Searching the *TGTP* repository can be done in three ways: a simple textual query, a more comprehensive textual search, and a geometric search [7].

The simple textual query is done using *MySQL* regular expressions queries [6], over the `name` attribute of the `Conjectures` table, it will provide the list of conjectures with names similar to the query. Another, more powerful, textual query mechanism is available, using the *full-text search* of *MySQL* [6]. The attributes `name, description, shortDescription, keyword` of the `theorems` and `keywords` tables are used, allowing, for a given input sentence, to get the list of most similar sentences in any attribute of the different problem descriptions.

Based on some preliminary work on geometric search [4] we developed a geometric search mechanism. The queries are constructed using a dynamic geometry system (*GeoGebra*[3]) and the constructed figure is semantically compared with the figures in the repository.

2 Geometric Search in *TGTP*

2.1 Conceptual Graphs

Knowledge representation provides techniques for describing objects in a knowledge domain, using concepts and relations defined by consensus in a community of users. In the case of Euclidean geometry the choice of concepts and relations is quite straightforward: we will use points, segments, lines (see 2.2). Once the signature decided, there are several mathematical structures for building knowledge bases. We have chosen to use *conceptual graphs* [1] rather than OWL ontologies based on RDF triples because the former allow relations of arbitrary arity and because conceptual graphs can be processed with graph theory algorithms.

To give an example, the (trivial) geometric figure of a single line segment AB is represented by a CG of four concepts and three relations (see Fig. 1): two concepts of type point, with markers A and B, one concept of type segment, with marker S_{AB} and one concept of type line, with marker L_{AB}; the three relations are: between A and S_{AB} as well as between B and S_{AB} there are binary relations I "is incident to," and between S_{AB} and L_{AB} there is a binary relation C "contained in":

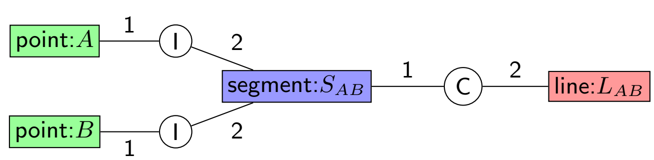


Fig. 1. Conceptual Graph, Single Line Segment AB

The semantics are as follows: point:A and point:B are distinct[4] extremities of segment:S_{AB}, which, in turn, is contained in line:L_{AB}. When converting a

[3] `https://www.geogebra.org/`

[4] We consider that every concept of the graph represents a distinct geometric object. Whenever inference reveals that two concepts represent the same object, they are merged.

geometric figure from some other representation to conceptual graphs, geometric constraints of the figure (as in [5, §2.2]) become conceptual graph relations. Furthermore, geometric inference rules become conceptual graph λ-rules. We repeatedly apply inference rules until inferential closure is obtained. By doing this a search will succeed in finding a figure even if it has been originally described in a different but geometric equivalent way. For example, if a figure has been converted into a conceptual graph as a *triangle with three equal sides* and the user searches a *triangle with three equal angles* (which is geometrically equivalent), the search will be successful because—thanks to inferential closure—the property that angles are equal will be already part of the graph (see Fig. 2).

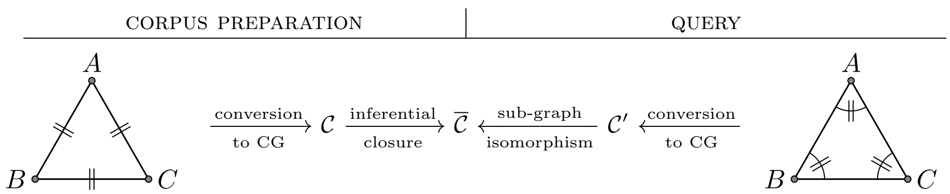


Fig. 2. Geometric Query Using Conceptual Graphs

For each construction in *TGTP* the conceptual graphs is found, then its inferential closure is calculated. When a query is done the conceptual graph is found, then a intermediate representation (see Section 2.4) is used as a filter to found a list of potential candidates and the (still to be implemented), using a sub-graph isomorphism the matches would be found.

2.2 Figures Represented as Conceptual Graphs

We use concepts point, segment, line, circle, angle, ratio. The model semantics of this vocabulary are the corresponding geometric notions (interpretations of the former four concepts take their values in $\mathbb{R}^2$, the interpretation of angle is a full-angle ([2, p. 44–50]) and interpretation of ratio is a real number). We also use three constants: angle:0 and angle:1 are angle-type individuals corresponding to full-angles $\angle[0]$ and $\angle[1]$ (in Chou notation), and ratio:1 is a ratio individual of value 1. The relations of our vocabulary are the following (see Table 1):

In conceptual graphs representing geometric figures, every segment has to be connected to a single line by an is_contained_in relation and every pair of lines ℓ, ℓ' is connected to angle concepts $\angle[\ell, \ell']$ and $\angle[\ell', \ell]$ by is_angle_of relations. These angle concepts are interconnected by the is_negative_of relation. When two lines are geometrically parallel, their angle is the individual angle:0; when two lines are geometrically perpendicular, their angle is the angle:1 individual. Every pair of segments is connected to a ratio by an is_ratio_of$_s$ relation; when they are geometrically congruent, then their ratio is the ratio:1 individual. If point:C is the geometric midpoint of segment:AB then three segments AB, AC

relation	arity	signature	relation	arity	signature
is_incident_to$_s$	2	(point,segment)	is_negative_of	2	(angle,angle)
is_incident_to$_c$	2	(point,circle)	is_ratio_of$_s$	3	(ratio,segment,segment)
is_contained_in	2	(segment,line)	is_ratio_of$_a$	3	(ratio,angle,angle)
is_center_of	2	(point,circle)	is_equal_to$_a$	2	(angle,angle)
is_angle_of	3	(angle,line,line)	is_equal_to$_r$	2	(ratio,ratio)
is_summit_of	2	(point,angle)			

Table 1. Geometric Relations

and BC have to be provided in the graph and the ratio of AC and BC is ratio:1. As for is_equal_to relations, they are only used between angles or between ratios. Consistency checking algorithms continuously verify that equality is transitive and that for every path containing an even number of is_negative_of relations there will an equality relation. Integrity checking algorithms verify that every segment is connected to a single line, that lines with zero angle and a common point are merged, that every circle has a single center point, etc.

2.3 Inferential Closure

We have implemented the inference rules as Python functions. They are then repeatedly applied until the CG remains unchanged, which means inferential closure has been attained. We have adapted rules specific to full-angles to this format, as well as rules D1–D75 of [3, p. 242].

The fact that we use CGs allows us to be independent of predicate argument order: for example, the sole purpose of rules D14–D17 in [3] is to ensure that the predicate cyclic(A, B, C, D) is true for any order of arguments A, B, C, D. In our case, we get four point concepts connected to the same circle concept via a is_incident_to relation, without any order. This allows us to significantly reduce the number of inference rules to apply, compared to [3].

Applying an inference rule is finding a pattern in the graph (i.e., a CG λ-rule, see [1, Ch. 10]) and transforming the graph in a specific way (by adding or merging vertices and/or edges). To avoid unnecessary use of the sub-graph isomorphism algorithm, the system calculates global trail distributions of inference rule patterns so that they are applied only if there is a chance that they will indeed match some sub-graph and transform it.

2.4 Global Trail Distributions

The inferential closure and the sub-graph isomorphism algorithms are heavy CPU consumers, we have developed a strategy for finding sets of potential matches, so that the set of figures to which the algorithms has to be applied is as small as possible.

To allow easier searching of match candidates in the corpus, a sequence of numeric values, called *global trail distribution*, is calculated out of the query graph.

A *global trail distribution*[5] is a sequence of key/value pairs partially describing the query conceptual graph. It has the important property that if the query graph is indeed contained in some corpus graph, then (a) *all keys of the query graph must also be keys of the corpus graph*, and (b) *the value of every key of the query must be smaller or equal to the value of the same key in the corpus graph*. Verification of graph compatibility is straightforward: we check whether all query keys also belong to the corpus graph key set, if not then the corpus graph is removed from the candidate list. Once this test is passed we compare values of query keys and corpus keys and check whether inequality is verified in all cases.

Keys are obtained in the following way: we take *all trails* of the CG. Let $p = (e_1, \ldots, e_n)$ be a trail (and hence $e_i \neq e_j$ for all $1 \leq i < j \leq n$), where e_i are concept nodes, relation nodes and relation edge labels. If we replace concepts by their types and relations by their symbols, we get a sequence $(t_1, \ldots, t_n)$ of concept types, relation symbols and relation edge labels. We replace t_is by numeric identifiers and call the obtained numeric sequence $\tau(p)$. Sequence $\tau(p)$ describes the specific trail p but also all other trails containing similar concepts and relations in the same order. The fact is that if we want the query graph to be contained in the corpus graph, then the latter must also contain at least the same number of trails of the same type sequence as the former. Let $\#\tau(p) = |\{p' \in \mathrm{Trails}(G) : \tau(p) = \tau(p')\}|$ where $\mathrm{Trails}(G)$ is the set of trails of G. We use $\tau(p)$ sequences as keys and $\#\tau(p)$ as values, i.e., the *global trail distribution* $\mathrm{GTD}(G)$ *of graph* G is defined as:

$$\mathrm{GTD}(G) = \{(\tau(p), \#\tau(p)) : p \in \mathrm{Trails}(G)\}$$

3 Geometric Search Implementation in *TGTP*

The implementation of the geometric search in *TGTP* is divided in two steps with several sub-steps:

1. corpus preparation (to be done once for each figure in *TGTP*)[6]:
   (a) convert the corpus into conceptual graphs. This conversion is very efficient, $0.54s \leq t \leq 5.55s$, for the examples timed, with an average value of $1.34s$;
   (b) obtain the inferential closure of each figure in the corpus. This can be very heavy time consuming process, ranging from $0.91s$ to $> 100000s$. This step is still not completed, not all the figures have a corresponding conceptual graphs, inferential closure.
2. the query (to be done for every query):
   (a) use a DGS (*GeoGebra*, JavaScript applet) to make the query;
   (b) convert the query into a conceptual graph (very efficient, see 1a);

[5] In graph theory, a trail is a path with no repeated edges.

[6] It should be noted that some geometric conjectures can be in *TGTP* without a corresponding figure, For those cases the geometric search is not applicable

(c) compare the global trail distribution of the query with those in the corpus, obtaining a set of candidates. Using a *MySQL right outer join* query [6] this is done very quickly (few seconds);
(d) apply typed sub-graph isomorphism algorithm to candidates. This step is still to be implemented;
(e) if the algorithm succeeds, return a list of corpus graphs as standard geometry figures representations. Using the same DGS used for making the query, the *TGTP*'s user should be able to browse the list of results. For each geometric construction in the list it will be possible to visualise it, with the part matching the query highlighted.

In *TGTP* the queries are constructed using *GeoGebra*, the global trail distributions for that construction are calculated and then matched against the ones in the corpus. This provides a very fast mechanism to build a list of similar constructions that is made available to the *TGTP* user making the query.

4 Future Work and Conclusion

The *TGTP* system has already fulfilled many of the goals specified at the beginning of the project. The geometric search mechanism is, in our opinion, a very interesting addition to the platform. Nevertheless there are still many improvements to be done.

For now all the searches are independent of each others, the user should be able to combine them, e.g. after a given full-text search, run a geometric search in the resulting list.

Stopping the query at the global trail distribution matching level, is a very fast way to get results, but the price to pay is uncertainty about precise matching and impossibility to highlight the query match inside the corpus graph. The next steps, still to be implemented in *TGTP*, are: to improve the filtering of the result list (step 2c) with a deep learning generic graph representation learning framework [8]; after having the result list of problems we should be able to apply the typed sub-graph isomorphism, the result could then be visualised using *GeoGebra*, with colours marking the query construction as a sub-construction of the corpus constructions.

References

1. Chein, M., Mugnier, M.L.: Graph-based Knowledge Representation. Computational Foundations of Conceptual Graphs. Advanced Information and Knowledge Processing Series, Springer (2009)
2. Chou, S.C., Gao, X.S., Zhang, J.Z.: Machine Proofs in Geometry. World Scientific (1994)
3. Chou, S.C., Gao, X.S., Zhang, J.Z.: A deductive database approach to automated geometry theorem proving and discovering. Journal of Automated Reasoning 25, 219–246 (2000)

4. Haralambous, Y., Quaresma, P.: Querying geometric figures using a controlled language, ontological graphs and dependency lattices. In: et al., S.W. (ed.) CICM 2014. LNAI, vol. 8543, pp. 298–311. Springer (2014)
5. Mathis, P., Thierry, S.E.B.: A formalization of geometric constraint systems and their decomposition. Formal Aspects of Computing 22(2), 129–151 (Mar 2010), https://doi.org/10.1007/s00165-009-0117-8
6. Oracle: MySQL 5.5 Reference Manual. Oracle, 5.5 edn. (January 2011), revision: 24956
7. Quaresma, P.: Thousands of Geometric problems for geometric Theorem Provers (TGTP). In: Schreck, P., Narboux, J., Richter-Gebert, J. (eds.) Automated Deduction in Geometry, Lecture Notes in Computer Science, vol. 6877, pp. 169–181. Springer (2011)
8. Rossi, R.A., Zhou, R., Ahmed, N.K.: Deep inductive network representation learning. In: Companion Proceedings of the The Web Conference 2018. pp. 953–960. WWW '18, International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, Switzerland (2018), https://doi.org/10.1145/3184558.3191524

Mining Geometry Theorems in a Regular Polygon

Zoltán Kovács

The Private University College of Education of the Diocese of Linz
Salesianumweg 3, 4020 Linz, Austria
zoltan@geogebra.org

Abstract. We demonstrate a systematic, automated way of discovery of geometry theorems on regular polygons.

Keywords: Automated theorem proving · Computer algebra · Regular Polygons · WebAssembly · *GeoGebra*

1 Introduction

Obtaining interesting mathematical theorems automatically is a usual dream of many mathematicians. By defining a formal language (with its logical axioms) on a research field, and a set of (non-logical) axioms, one can deduce various statements only by repeating the axioms. In principle, proofs for all propositions in a research field can be traced back to consecutive uses of the axioms.

Several axiomatizations are available for many research fields in mathematics, however, interesting theorems (with proofs) are more difficult to find. One problem is that combining the axioms consecutively usually produces an unmanageable big database of propositions, including trivial or uninteresting ones. There is already remarkable work done in this field, including [1] which is one of the first reports, and, in particular, producing proofs in elementary planar geometry, we refer to [2–4] where the combinatorial explosion is also addressed. The other problem is to identify which propositions are interesting enough to call them theorems, here we refer the reader to [5, 6].

In this paper we limit our considerations to planar Euclidean geometry, namely to find interesting properties in a *regular polygon*. The literature on listing such properties is, actually, huge, including constructible polygons (by compass and straightedge or origami, for example). In fact, from the very start of the availability of computer algebra systems (CAS) and dynamic geometry software (DGS), namely, the 1990s, non-constructible polygons can also be better observed, either numerically or symbolically.

In this study we limit the available axioms to very simple operations on a regular n-gon. Its diagonals (including the sides) will be taken, two pairs of them will determine a pair of intersection points which define a segment. By considering all possible segments defined in this way, we will compute the lengths of them symbolically, and, depending on the "simplicity" of the symbolic result we classify the segment either as "interesting" or "not interesting". This is, of course, somewhat subjective, but this approach can be slightly modified by allowing other results interesting enough, or by defining some other points as well for the domain of interest.

The paper consists of the following parts: In section 2 the mathematical background is explained on computing an appearing segment symbolically. Section 3 presents some manually obtained new results. Section 4 demonstrates how the mathematical computations can be automated by using the tool `RegularNGons`. Finally, section 5 depicts some future ideas.

We remark that the "geometry theorems" we obtain in this article are related to *lengths* appearing in regular polygons. Therefore these results may also be considered as "algebraic theorems" because the lengths are always expressed by roots of algebraic equations. On the other hand, the method we use can be easily extended to focus on "more geometric" properties like perpendicularity or parallelism of the obtained segments. Also, combining some "algebraic theorems" we can even conclude congruency of triangles (see the third property in Theorem 2 in section 2), among others.

2 Mathematical Background

In this section we discuss the mathematical background on a possible method to handle regular polygons with means in algebraic geometry.

2.1 Constructibility

Algebraization of the setup of a planar geometry statement is a well known process since the revolutionary book [7] of Chou's. It demonstrates on 512 mathematical statements how an equation system can describe a geometric construction, and by performing some manipulations on the equation system, a mechanical proof can be obtained. Chou's work was one of the first publicly available applications of Wu's algebraic geometry method [8]. It focuses mainly on constructible setups, that is, mostly on such constructions that can be created only by using the classic approach, namely by compass and straightedge. There is, however, a proof on Morley's trisector theorem presented which assumes a non-Euclidean, cubic way of being constructed, however, the explicit way of construction is successfully avoided, therefore the theorem is manageable.

It is well known (Gauß, 1801, Wantzel, 1837, see [9,10]) that a regular n-gon is constructible by using compass and straightedge if and only if n the product of a power of 2 and any number of distinct Fermat primes (including none). We recall that a Fermat prime is a prime number of the form $2^{2^m}+1$. By using this theorem the list of the constructible regular n-gons are:

$$n = 3, 4, 5, 6, 8, 10, 12, 15, 16, 17, 20, 24, \ldots$$

A generalization of this result (Pierpont, 1895, see [11]) by allowing an angle trisector as well (for example, origami folding steps), is that a regular n-gon is constructible if and only if

$$n = 2^r \cdot 3^s \cdot p_1 \cdot p_2 \cdots p_k,$$

where $r, s, k \geq 0$ and the p_i are distinct primes of form $2^t \cdot 3^u + 1$ [12]. The first constructible regular n-gons of this kind are

$$n = 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 24, \ldots$$

From this second list the cases $n = 11$, $n = 22$ and $n = 23$ are missing again, and, as a natural consequence, there are much less scientific results known on regular 11-, 22- and 23-gons than for n-gons appearing in the lists. Later we will show some—to our best knowledge—new results on the cases $n = 11$ and $n = 23$, among other ones.

2.2 An Algebraic Formula for the Vertices

From now on we assume that $n \geq 1$. The cases $n = 1, 2$ have no geometrical meaning, but they will be useful from the algebraic point of view.

In the algebraic geometry approach the usual way to describe the points of a construction is to assign coordinates (x_i, y_i) for a given point P_i $(i = 0, 1, 2, \ldots)$. When speaking about a polygon, in many cases the first vertices are put into coordinates $P_0 = (0, 0)$ and $P_1 = (1, 0)$, and the other coordinates are described either by using exact rationals, or the coordinates are expressed as possible solutions of algebraic equations.

For example, when defining a square, $P_2 = (1, 1)$ and $P_3 = (0, 1)$ seem to make sense, but for a regular triangle two equations for $P_2 = (x_2, y_2)$ are required, namely $x_2^2 + y_2^2 = 1$ and $(x_2 - 1)^2 + y_2^2 = 1$. It is easy to see that this equation system has two solutions, namely $x_2 = \frac{1}{2}, y_2 = \frac{\sqrt{3}}{2}$ and $x_2 = \frac{1}{2}, y_2 = -\frac{\sqrt{3}}{2}$. It is well known that there is no way in the algebraic geometry approach to avoid such duplicates, unless the coordinates are rational. In other words, if both minimal polynomials of the coordinates are linear (or constant), then the duplicates can be avoided, otherwise not. Here, for x_2 we have $2x_2 - 1 (= 0)$, but for y_2 the minimal polynomial is $4y_2^2 - 3 (= 0)$. We remark that the minimal polynomials are irreducible over $\mathbb{Z}$.

Clearly, minimal polynomials of a regular n-gon with vertices $P_0 = (0, 0)$ and $P_1 = (1, 0)$ can play an important role here. The paper [13] suggests an algorithm to obtain the minimal polynomial $p_c(x)$ of $\cos(2\pi/n)$, based on the Chebyshev polynomials $T_j(x)$ of the first kind (see Algorithm 1).

Algorithm 1 Computing the minimal polynomial of $\cos(2\pi/n)$

procedure COS2PIOVERNMINPOLY(n)
    $p_c \leftarrow T_n - 1$
    for all $j \mid n \ \wedge\ j < n$ **do**
        $q \leftarrow T_j - 1$
        $r \leftarrow \gcd(p_c, q)$
        $p_c \leftarrow p_c / r$
    return SquarefreeFactorization(p_c)

Clearly, adding the equation $p_c(x)^2 + p_s(y)^2 = 1$ to the equation system, we have managed to describe a polynomial $p_s(y)$ such that $p_s(\sin(2\pi/n)) = 0$. Table 1 shows the minimal polynomials for $n \leq 7$.

Table 1. List of minimal polynomials of $\cos(2\pi/n)$, $n \leq 7$

n	Minimal polynomial of $\cos(2\pi/n)$
1	$x - 1$
2	$x + 1$
3	$2x + 1$
4	x
5	$4x^2 + 2x - 1$
6	$2x - 1$
7	$8x^3 + 4x^2 - 4x - 1$

It is clear, that—not considering the cases $n = 1, 2, 3, 4, 6$—the number of roots of p_c is more than one, therefore the solution of the equation system $\{p_c(x) = 0, p_s(x) = 0\}$ is not unique. The number of solutions for $p_c(x) = 0$ depends on the degree of p_c, and—not considering the cases $n = 1, 2$—the number of solutions for $p_s(x) = 0$ is two for each root of $p_c(x)$, therefore the number of solutions for $\{p_c(x) = 0, p_s(y) = 0\}$ is usually $2 \cdot \deg(p_c)$. As a result, the point

$$P = (\cos(2\pi/n), \sin(2\pi/n))$$

can be exactly determined by an algebraic equation in the algebraic geometry approach only in case $n = 4$, as shown in Table 2.

Table 2. Degree of ambiguity for $(\cos(2\pi/n), \sin(2\pi/n))$, $3 \leq n \leq 13$

n	3	4	5	6	7	8	9	10	11	12	13
Degree	2	2	4	2	6	4	6		4	10		4	12

It seems to make sense that the degree of ambiguity (not considering the case $n = 4$) can be computed with Euler's totient function, that is, the degree equals to $\varphi(n)$. Later we will give a short proof on this.

Now we are ready to set up additional formulas to describe the coordinates of the vertices of a regular n-gon, having its first vertices $P_0 = (0, 0)$ and $P_1 = (1, 0)$, and the remaining vertices $P_2 = (x_2, y_2), \ldots, P_{n-1} = (x_{n-1}, y_{n-1})$ are to be found. By using consecutive rotations and assuming $x = \cos(2\pi/n), y =$

$\sin(2\pi/n)$, we can claim that

$$\begin{pmatrix} x_i \\ y_i \end{pmatrix} - \begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix} = \begin{pmatrix} x & -y \\ y & x \end{pmatrix} \cdot \left(\begin{pmatrix} x_{i-1} \\ y_{i-1} \end{pmatrix} - \begin{pmatrix} x_{i-2} \\ y_{i-2} \end{pmatrix} \right)$$

and therefore

$$x_i = -xy_{i-1} + x_{i-1} + xx_{i-1} + yy_{i-2} - xx_{i-2}, \tag{1}$$

$$y_i = y_{i-1} + xy_{i-1} + yx_{i-1} - xy_{i-2} - yx_{i-2} \tag{2}$$

for all $i = 2, 3, \ldots, n-1$.

3 Manual Results on Regular 5- and 11-gons

In this section we present a well-known statement on a regular 5-gon that can be obtained by using the formulas from the previous section. Also, we list some properties of a regular 11-gon, obtained with the same approach.

3.1 Some Properties of a Regular Pentagon

Theorem 1. *Consider a regular pentagon (Fig. 1) with vertices $P_0, P_1, \ldots, P_4$. Let $A = P_0$, $B = P_2$, $C = P_1$, $D = P_3$, $E = P_0$, $F = P_2$, $G = P_1$, $H = P_4$. Let us define diagonals $d = AB, e = CD, f = EF, g = GH$ and intersection points $R = d \cap e, S = f \cap g$. Now, when the length of P_0P_1 is 1, then the length of RS is $\frac{3-\sqrt{5}}{2}$.*

This result is well-known from elementary geometry, but here we provide a proof that uses the developed formulas from section 2. We will use the variables x_0, x_1, x_2, x_3, x_4 for the x-coordinates of the vertices, y_0, y_1, y_2, y_3, y_4 for the y-coordinates, and x and y for the cosine and sine of $2\pi/5$, respectively. Points P_0 and P_1 will be put into $(0, 0)$ and $(1, 0)$.

By using Table 1 and Equations (1) and (2), we have the following hypotheses:

$$\begin{aligned}
h_1 &= 4x^2 + 2x - 1 = 0, \\
h_2 &= x^2 + y^2 - 1 = 0, \\
h_3 &= x_0 = 0, \\
h_4 &= y_0 = 0, \\
h_5 &= x_1 - 1 = 0, \\
h_6 &= y_1 = 0, \\
h_7 &= -x_2 + -xy_1 + x_1 + xx_1 + yy_0 - xx_0 = 0, \\
h_8 &= -y_2 + y_1 + xy_1 + yx_1 - xy_0 - yx_0 = 0, \\
h_9 &= -x_3 + -xy_2 + x_2 + xx_2 + yy_1 - xx_1 = 0, \\
h_{10} &= -y_3 + y_2 + xy_2 + yx_2 - xy_1 - yx_1 = 0, \\
h_{11} &= -x_4 + -xy_3 + x_3 + xx_3 + yy_2 - xx_2 = 0, \\
h_{12} &= -y_4 + y_3 + xy_3 + yx_3 - xy_2 - yx_2 = 0.
\end{aligned}$$

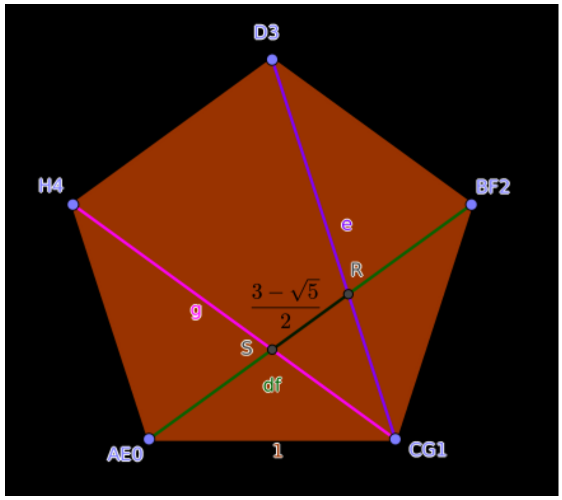


Fig. 1. A well-known theorem on a regular pentagon (for convenience we use only the indices of the points in the figure, that is, $0, 1, \ldots, n-1$ stand for $P_0, P_1, \ldots, P_{n-1}$, respectively)

Since $R \in d$ and $R \in e$, we can claim that

$$h_{13} = \begin{vmatrix} x_0 & y_0 & 1 \\ x_2 & y_2 & 1 \\ x_r & y_r & 1 \end{vmatrix} = 0, h_{14} = \begin{vmatrix} x_1 & y_1 & 1 \\ x_3 & y_3 & 1 \\ x_r & y_r & 1 \end{vmatrix} = 0,$$

where $R = (x_r, y_r)$. Similarly,

$$h_{15} = \begin{vmatrix} x_0 & y_0 & 1 \\ x_2 & y_2 & 1 \\ x_s & y_s & 1 \end{vmatrix} = 0, h_{16} = \begin{vmatrix} x_1 & y_1 & 1 \\ x_4 & y_4 & 1 \\ x_s & y_s & 1 \end{vmatrix} = 0,$$

where $S = (x_s, y_s)$. Finally we can define the length $|RS|$ by stating

$$h_{17} = |RS|^2 - \left((x_r - x_s)^2 + (y_r - y_s)^2\right) = 0.$$

From here we can go ahead with two methods:

1. We directly prove that $|RS| = \frac{3-\sqrt{5}}{2}$. As we will see, this actually does not follow from the hypotheses, because they describe a different case as well, shown in Fig. 2. That is, we need to prove a weaker thesis, namely that $|RS| = \frac{3-\sqrt{5}}{2}$ or $|RS| = \frac{3+\sqrt{5}}{2}$, which is equivalent to

$$\left(|RS| - \frac{3-\sqrt{5}}{2}\right) \cdot \left(|RS| - \frac{3+\sqrt{5}}{2}\right) = 0.$$

Unfortunately, this form is still not complete, because $|RS|$ is defined implicitly by using $|RS|^2$, that is, if $|RS|$ is a root, also $-|RS|$ will appear. The correct form for t is therefore

$$t = \left(|RS| - \frac{3-\sqrt{5}}{2}\right) \cdot \left(|RS| - \frac{3+\sqrt{5}}{2}\right) \cdot \left(-|RS| - \frac{3-\sqrt{5}}{2}\right) \cdot \left(-|RS| - \frac{3+\sqrt{5}}{2}\right) = 0,$$

that is, after expansion,

$$t = (|RS|^2 - 3|RS| + 1) \cdot (|RS|^2 + 3|RS| + 1) = |RS|^4 - 7|RS|^2 + 1 = 0.$$

Proving the thesis $t = 0$ can be done by contradiction: we insert $t \cdot z - 1 = 0$ into the equation system $\{h_1, h_2, \ldots, h_{17}\}$ and get a contradictory equation system. This approach is based on the Rabinowitsch trick, introduced by Kapur in 1986 (see [14]).

2. We can also discover the exact value of $|RS|$ by eliminating all variables from the ideal $\langle h_1, h_2, \ldots, h_{17}\rangle$, except $|RS|$. We will follow this second method, suggested by Recio and Vélez in 1999 (see [15]).

Let us emphasize that the first method can be used only *after* one has a conjecture already. In contrast, the second method can be used *before* having a conjecture, namely, to find a conjecture *and* its proof at the same time.

For the first method we must admit that in Wu's approach there is no way to express that the length of a segment is $\frac{3-\sqrt{5}}{2}$. Instead, we need to use its minimal polynomial, having integer (or rational) coefficients. Actually, $|RS|^2 - 3|RS| + 1$ is a minimal polynomial of both $\frac{3-\sqrt{5}}{2}$ and $\frac{3+\sqrt{5}}{2}$, and $|RS|^2 + 3|RS| + 1$ is of $-\frac{3-\sqrt{5}}{2}$ and $-\frac{3+\sqrt{5}}{2}$. In fact, given a length $|RS|$ in general, we need to prove that the equation $t = t_1 \cdot t_2 = 0$ is implied where t_1 and t_2 are the minimal polynomials of the expected $|RS|$ and $-|RS|$, respectively. Even if geometrically $t_1 = 0$ is implied, from the algebraic point of view $t_1 \cdot t_2 = 0$ is to be proven.

Also, we remark that $|RS|$ always appears to an even power in t.

Finally, when using the second method, by elimination (here we utilize computer algebra), we will indeed obtain that

$$\langle h_1, h_2, \ldots, h_{17}\rangle \cap \mathbb{Q}[|RS|] = \left\langle |RS|^4 - 7|RS|^2 + 1 \right\rangle.$$

3.2 Star-regular Polygons

Before going further, we need to explain the situation with the star-regular pentagon in Fig. 2. Here we need to mention that the equation $h_1 = 4x^2+2x-1 = 0$ describes not only $\cos(2\pi/5)$ but also $\cos(2\cdot 2\pi/5)$, $\cos(3\cdot 2\pi/5)$ and $\cos(4\cdot 2\pi/5)$, however, because of symmetry, the first and last, and the second and third values

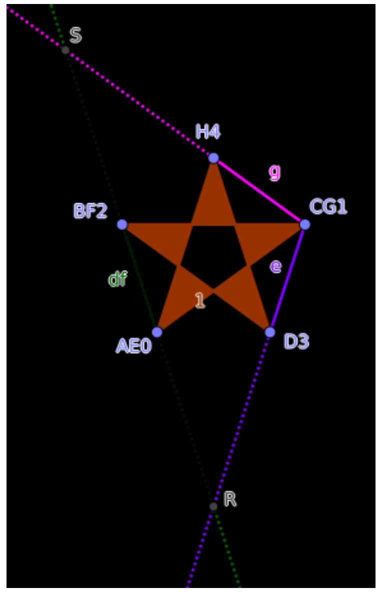


Fig. 2. A variant of the theorem in a star-regular pentagon

are the same. (We can think of these values as the projections of z_1, z_2, z_3, z_4 on the real axis, where

$$z_j = (\cos(2\pi/5) + i\sin(2\pi/5))^j = \cos(j \cdot 2\pi/5) + i\sin(j \cdot 2\pi/5),$$

$j = 1, 2, 3, 4$.)

That is, in this special case (for $n = 5$) h_1 is a minimal polynomial of $\operatorname{Re} z_1 (= \operatorname{Re} z_4)$ and $\operatorname{Re} z_2 (= \operatorname{Re} z_3)$. By considering the formulas (1) and (2) we can learn that the rotation is controlled by the vector (x, y), where $2\pi/n$ is the external angle of the regular n-gon. When changing the angle to a double, triple, ..., value, we obtain star-regular n-gons, unless the external angle describes a regular (or star-regular) m-gon ($m < n$).

This fact is well-known in the theory of regular polytopes [16], but let us illustrate this property by another example. When choosing $n = 6$, we have $h_1' = 2x - 1 = 0$ that describes $\cos(2\pi/6) = \cos(5 \cdot 2\pi/6)$. Now by considering $z_1', z_2', z_3', z_4', z_5'$ where

$$z_j' = \cos(j \cdot 2\pi/6) + i\sin(j \cdot 2\pi/6),$$

$j = 1, 2, 3, 4, 5$, we can see that z_2' can also be considered as a generator for $\cos(1 \cdot 2\pi/3)$ (when projecting it on the x-axis) since $2 \cdot 2\pi/6 = 1 \cdot 2\pi/3$. That is, $z_2' (= \overline{z_4'})$ is not used when generating the minimal polynomial of $\cos(2\pi/6)$ (it occurs at the creation of the minimal polynomial of $\cos(2\pi/3)$), and this is the case also for z_3' (because it is used for the minimal polynomial of $\cos(2\pi/2)$).

An immediate consequence is that z'_j is used as a generator in the minimal polynomial of $\cos(2\pi/6)$ if and only if j and 6 are coprimes, but since $\cos(2\pi/6) = \cos(5 \cdot 2\pi/6)$, only the first half of the indices j play a technical role. In general, when n is arbitrary, the number of technically used generators are $\varphi(n)/2$ (the other $\varphi(n)/2$ ones produce the same projections).

Finally, when considering the equation $x^2 + y^2 = 1$ as well, if $n \geq 3$, there are two solutions in y, hence the hypotheses describe *all* cases when j and n are coprimes (not just for the half of the cases, that is, for $1 \leq j \leq n/2$). Practically, the hypotheses depict not just the regular n-gon case, but also *all* star-regular n-gons. It is clear, after this chain of thoughts, that the number of cases is $\varphi(n)$ (which is the number of positive coprimes to n, less than n). From this immediately follows that the degree of ambiguity for $(\cos(2\pi/n), \sin(2\pi/n))$ is *exactly* $\varphi(n)$.

Also, it is clear that there exists essentially only one regular 5-gon and one star 5-gon (namely, $\{5/2\}$, when using the Schäfli symbol, see [16]). But these are just two different cases. The other two ones, according to $\varphi(5)$, are symmetrically equivalent cases. The axis of symmetry is the x-axis in our case.

On the other hand, by using our method, it is not always possible to distinguish between these $\varphi(n)$ cases:

1. $t = |RS|^2 - c$ where c is a rational. In this case clearly $|RS| = \sqrt{c}$ follows.
2. Otherwise, the resulting polynomial t is a product of two polynomials $t_1, t_2 \in \mathbb{Q}[|RS|]$, and the half of the union of their roots are positive, while the others are negative. On the other hand, the positive roots can be placed in several combinations in t_1 and t_2 in general:
   (a) In our concrete example there are two positive roots in t_1 and two negative ones in t_2. When considering similar cases, the positive roots can always occur in, say t_1, and the negative roots then in t_2. Albeit the elimination delivers the product $t = t_1 \cdot t_2$, clearly t_2 cannot play a geometrical role, therefore t_1 can be concluded.
   However, if t_1 contains more than one (positive) root, those roots cannot be distinguished. This is the case in our concrete example as well.
   (b) In general, t_1 may contain a few positive solutions, but t_2 may also contain some other ones. In such cases the positive solutions in t_1 and t_2 cannot be distinguished from each other.
   Such an example is the polynomial $t = t_1 \cdot t_2$ where $t_1 = |RS|^2 - |RS| - 1$ and $t_2 = |RS|^2 + |RS| - 1$. It describes the length of the diagonal of a regular (star-) pentagon, namely both lengths $\frac{\sqrt{5}\pm1}{2}$. Here t_1 contains one of the positive roots, namely $\frac{\sqrt{5}+1}{2}$, while t_2 the other one, $\frac{\sqrt{5}-1}{2}$. At the end of the day, only t can be concluded, none of its factors can be dropped because both contain geometrically useful data.

3.3 Lengths in a Regular 11-gon

In section 2 we mentioned that scientific results on a regular 11-gon are not very well-known because it is not constructible by typical means. Here we show some—for us, previously unknown—results that have been obtained by our method, implemented in the free dynamic geometry tool *GeoGebra*.

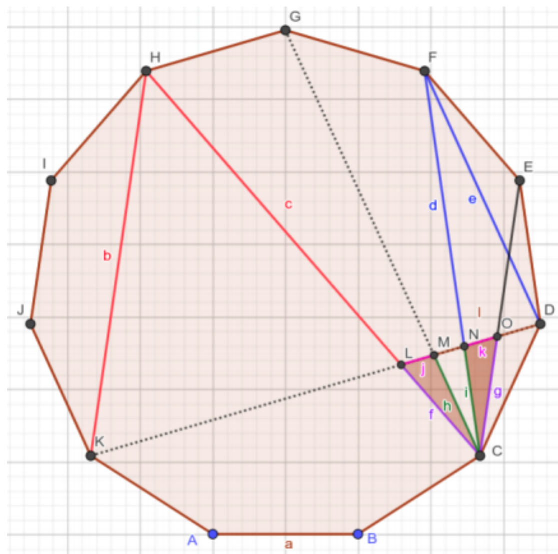


Fig. 3. Some properties of a regular 11-gon

Theorem 2. *A regular 11-gon is defined by points $A, B, C, \ldots, J, K$. Diagonals CE, CF, CG, CH, DF, DK and HK are drawn. Then intersection points L, M, N and O are defined as shown in Fig. 3. The following properties hold:*

- $b = c$,
- $d = e$,
- *triangles CLM and CON are congruent,*
- $a = l$ *(that is, $AB = DL$).*

Proof. By using the method described above, all of these statements can be mechanically proved in a straightforward way.

4 Automated Discovery of Theorems

Obtaining beautiful new results randomly is one of the possible aims when observing regular polygons. But, luckily, this kind of discovery can be systematic when the different setups $\mathcal{S}$ are numbered consecutively. If there is a bijective map

$$S : \{0, 1, 2, \ldots, s-1\} \to \mathcal{S},$$

clearly there are some programmatic benefits for the processing of the cases:

1. A database $D : \{0, 1, 2, \dots, s-1\} \to \{\text{true}, \text{false}\}$ can be maintained. Here for each $k \in \mathbb{N}_0, k < s$ there is an explicitly defined setup $S(k) \in \mathcal{S}$, and it can be saved as a database entry $D(k)$ if the check was already performed or not. If the computation loop needs to be suspended or stopped due to the high amount of computations for a given k, it can be restarted at the same value k in a next loop, independently from the first run.
2. This also supports parallel or distributed computing. The number of cases k can be then split and the setups can be divided among several processors or computers.
3. The distributed computation can also be controlled via a centralized Internet application that communicates with the clients, assigns the task to them, collects the results, and updates the central database. Of course, not only the success of the performed computations should be stored, but also their results, by using a map $D' : \{0, 1, 2, \dots, s-1\} \to \dots$ that has a sophisticated output data structure.

This idea is well-known from various public projects, including the Great Internet Mersenne Prime Search (GIMPS, [17]), available since 1996. Today, also, harnessing the idle time of the user's processor is very popular in mining, for example, bitcoins [18], directly (on the user's own decision) or indirectly (by programs that abuse the available resources, as hidden applications on malicious websites and other malware, see [19]). This kind of technology is, however, well-tested and very successful. CPU time to find new Mersenne primes. A success story of Jonathan Pace's, a GIMPS volunteer who contributed for over 14 years, is that he discovered the 50th known Mersenne prime in December 2017, $2^{77,232,917} - 1$, and won \$3,000 reward [20].

4.1 A Bijective Mapping

In our approach we assume that a regular n-gon is to be studied. It has $\binom{n}{2}$ diagonals (including the sides). From these we select two different ones, d and e (the order of selection does not matter) to designate their intersection point R. That is, the number of possible selections are $\binom{\binom{n}{2}}{2}$. On the other hand, to designate another intersection point S from another combination of the diagonals, we finally have

$$\binom{\binom{\binom{n}{2}}{2}}{2} \tag{3}$$

different selections for the segment RS. When expanding the formula (3) we learn that the number of cases is

$$\frac{n^8 - 4n^7 + 2n^6 + 8n^5 - 15n^4 + 12n^3 + 12n^2 - 16n}{128} \sim \frac{n^8}{128},$$

that is, s is equal to $n^8/128$ asymptotically.

It would be useful to find a formula for $S(k)$ to compute RS quickly. For the first step we will construct another map

$$c : \{0, 1, 2, \ldots, \binom{m}{2} - 1\} \to \binom{\{0, 1, 2, \ldots, m-1\}}{2}$$

where $\binom{\{0,1,2,\ldots,m-1\}}{2}$ stands for the set of 2-combinations of the set $\{0, 1, 2, \ldots, m-1\}$. Here we will assume that

$$\begin{array}{llll} c(0) = \{0,1\}, & c(1) = \{0,2\}, & c(2) = \{0,3\}, \ldots, & c(m-2) = \{0, m-1\}, \\ c(m-1) = \{1,2\}, & c(m) = \{1,3\}, & c(m+1) = \{1,4\}, \ldots, & c(2m-4) = \{1, m-1\}, \\ c(2m-3) = \{2,3\}, & & \ldots, & \end{array}$$

..., and finally $c\left(\binom{m}{2} - 1\right) = \{m-2, m-1\}$. To compute c quickly, we consider the inverse map c^{-1}. It is clear that $c^{-1}(k, k+1) = (m-1)+(m-2)+\ldots+(m-k)$, that is, $\frac{(m-1)+(m-k)}{2} \cdot k = -\frac{1}{2}k^2 + k \cdot \frac{2m-1}{2} = p$.

Let us now assume that p is given, and k is to be computed. Clearly $-\frac{1}{2}k^2 + k \cdot \frac{2m-1}{2} - p = 0$, and using the quadratic equation solver formula,

$$k = \frac{\frac{1-2m}{2} \pm \sqrt{\left(\frac{2m-1}{2}\right)^2 - 2p}}{-1} = m - \frac{1}{2} \mp \sqrt{\left(m - \frac{1}{2}\right)^2 - 2p}.$$

Here obviously the subtraction should be chosen. By some further simple calculations finally we obtain the formula $c(p) = \{k, l\}$ where

$$k = \left\lfloor m - \frac{1}{2} - \sqrt{\left(m - \frac{1}{2}\right)^2 - 2p} \right\rfloor, \tag{4}$$

$$l = \frac{2p + k^2 - (2m-3) \cdot k}{2} + 1. \tag{5}$$

This formula can be used then multiple times for $m = \binom{\binom{n}{2}}{2}$, $m = \binom{n}{2}$ and $m = n$.

Example Let $n = 5$, then $s = \binom{\binom{\binom{5}{2}}{2}}{2} = 990$. We are interested in, say, the 678th case when observing all possible segments RS.

1. First we compute $\binom{\binom{5}{2}}{2} = 45 = m_1$. That is, we search for $c(678)$. By using formulas (4) and (5), we get $k = 19$ and $l = 33$.
2. Now we search for the 19th and 33th combinations of a set with $\binom{5}{2} = 10 = m_2$ elements. Using the same formulas, we get $k = 2, l = 5$ and $k = 4, l = 8$ values for $p = 19$ and $p = 33$, respectively.
3. Finally we search for the 2nd, 5th, 4th and 8th combinations of a set with $5 = m_3$ elements. Using the same formulas again, we get $k = 0, l = 3$, $k = 1, l = 3$, $k = 1, l = 2$ and $k = 2, l = 4$ values for $p = 2, 5, 4$ and 8, respectively.

Lastly we conclude that the 678th case describes when $A = P_0$, $B = P_3$, $C = P_1$, $D = P_3$, $E = P_1$, $F = P_2$, $G = P_2$, $H = P_4$.

4.2 An Implementation

This automated "mining" algorithm has been recently implemented in the software tool `RegularNGons` [21].

The following input parameters can be used the fine tune its output:

- $n = \ldots$ defines the number of vertices in the regular polygon.
- s and e define the starting and ending cases (both are non-negative integers, less than the formula (3)).
- By adding $m = \ldots$ or $M = \ldots$ the minimal and maximal degrees of outputs can be controlled, respectively. By default $m = 1$ and $M = 2$, that is, either linear results or quadratic surds are mined.
- The parameter u will force searching for results given as parameters. For example, $u = 2$ considers only the outputs that are of $|RS| = 2$.
- The option $S = 0$ tries to avoid checking cases that were already checked in a symmetrically equivalent position. When this is set, only the $A = 0$, $B \leq n/2$ cases will be checked. (Here, and from now on, we will use the indices of the points, that is, 0 stands for P_0, 1 for P_1, and so on.)
- When using $f = 1$, once a length is found, no more results will be printed that have the same length.
- The user may request to find lengths that are close to a given decimal number, but they are just approximately the same. The parameter $a = \ldots$ is to be set to the sought decimal. (See subsection 4.5 for some examples.)

The software tool runs in a modern web browser, for example, Google Chrome 64. It uses the Giac computer algebra system to compute eliminations (its *WebAssembly* [22] version is used in an embedded way), and GeoGebra to visualize the obtained results on-the-fly—finally (or during the run) the results can be saved as a GeoGebra file.

The timing for a complete run for a given n-gon depends on the magnitude of n. For smaller n values the complete run can be performed in seconds or minutes. For bigger n values, a complete run may take several hours, or days, or even more. Some, yet unresolved memory issues in Giac may require multiple runs for bigger n values.

A typical partial output of `RegularNGons` is the following, when using inputs $n = 7$, $S = 0$ and $f = 1$:

```
Welcome to RegularNGons (https://github.com/kovzol/RegularNGons)...
Starting with n=7, s=0
s can be incremented until 21945
n=7, s=4: A=0, B=1, C=0, D=2, E=0, F=1, G=1, H=2: {RS^2-1}, {{RS=1}}
n=7, s=124: A=0, B=1, C=0, D=2, E=1, F=3, G=2, H=6: {RS^2-2}, {{RS=(√2)}}
n=7, s=2113: A=0, B=1, C=2, D=3, E=0, F=5, G=1, H=6: {RS^2-4}, {{RS=2}}
Elapsed time: 0h 28m 40s
11627 cases were not checked to ignore symmetry
```

This result will be recalled later in Theorem 6.

4.3 Some Results

We will find the following definition useful when presenting the statements that can be mined by using `RegularNGons`.

Definition 1. – ***Points of the first kind*** *of a regular n-gon are its vertices. We denote this set by $\mathcal{P}_1$.*
– ***Segments of the first kind*** *of a regular n-gon are its sides and diagonals. We denote this set by $\mathcal{S}_1$.*
– ***Points of the k-th kind*** *of a regular n-gon are the intersection points of its segments of the $(k-1)$-th kind. We denote this set by $\mathcal{P}_k$.*
– ***Segments of the k-th kind*** *of a regular n-gon are the segments defined by its points of the (k)-th kind. We denote this set by $\mathcal{S}_k$.*

By using this notion, in this paper we consider *segments of the second kind* of a regular n-gon. We remark that it makes sense to study segments of higher kinds in a regular n-gon. It is easy to see that a recursive formula can be given to determine the number of possible cases for the various kinds of points and segments of a regular n-gon:

Proposition 1. – $|\mathcal{P}_1| = n$.
– $|\mathcal{S}_1| = \binom{|\mathcal{P}_1|}{2}$.
– $|\mathcal{P}_k| = \binom{|\mathcal{S}_{k-1}|}{2}$.
– $|\mathcal{S}_k| = \binom{|\mathcal{P}_k|}{2}$.

Proof. By construction, these formulas are obvious.

Now we can present some results:

Theorem 3. *Given a regular 7-gon, there are 42 segments of its second kind that are of length 2, shown in Fig. 4.*

Proof. By exhausting all $|\mathcal{S}_2| = 21945$ cases, there exist exactly the cases as presented. (The running time on a modern PC was about 1 hour and 15 minutes.)

The 42 different cases can be classified into 3 substantially different groups, shown in green, red and magenta in Fig. 4. Because of symmetry, each substantially different segment have 6 rotated copies and a mirrored copy with 6 other rotated copies. In total there are $7+7=14$ elements of the groups. In the figure only 2 representatives are colored in each group (they are mirror images), the others are all blue.

Theorem 4. *Given a regular 7-gon, and consider the segment $|RS| = 1$ of its second kind. Then:*

1. *There is a side AE of the 7-gon such that AE and RS are parallel such that $EARS$ is a parallelogram;*
2. *for this AE, the lines AS and ER are parallel diagonals of the 7-gon,*

unless RS is chosen from the red segments in Fig. 5.

Proof. Again, by exhaustion.

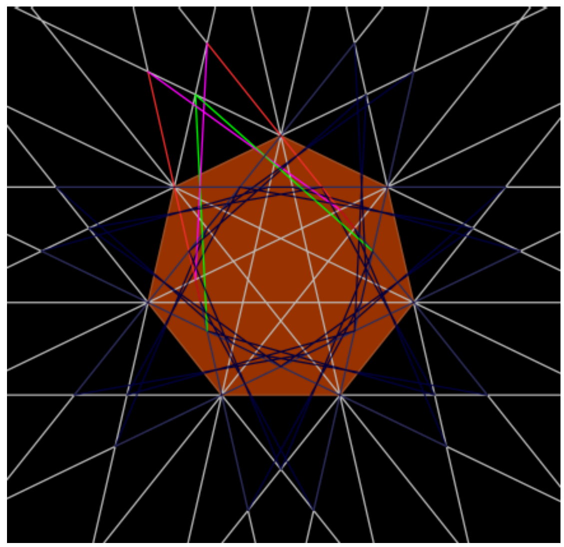


Fig. 4. Some properties of a regular heptagon

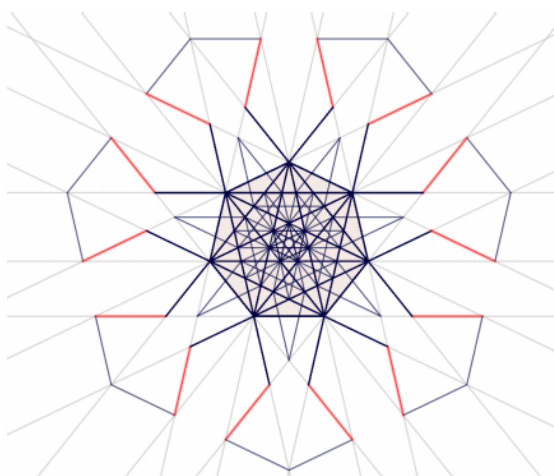


Fig. 5. Unit lengths appearing in a regular heptagon

It is easy to see that the converse of this theorem holds in all regular polygons, independently of n. That is, the following simple theorem can be stated:

Theorem 5. *Given a regular n-gon. Let us consider any side AE, and parallel diagonals $d = AB$ and $f = EF$; and, in addition, the diagonal $e = CD = g = GH$ which is parallel to AB. Now by choosing $R = d \cap e$, $S = f \cap g$, $|RS| = 1$.*

Proof. Due to parallelism, $EARS$ is clearly a parallelogram, and therefore $|RS| = |AE| = 1$.

By using elementary combinatorics, the number of possible cases can easily be counted. Also, taking an arbitrary diagonal instead of side AE in Theorem 5 we obtain similar theorems on lengths that are not unit long, but the same as the length of some diagonal.

4.4 Algebraic Numbers

In this subsection we give some statements without proofs. Rigorous proving of these statements is subject to future work.

We claim that the elimination process always produces a univariate polynomial t (of $|RS|$) that is a multiple of a minimal polynomial of $|RS|$, and therefore $|RS|$ is an algebraic number. In this case t must be a polynomial that contains all roots for all star-regular polygon cases as well.

As already mentioned in section 2, $|RS|$ is always described by RS^2 among the hypotheses, therefore if t has an arbitrary root $|RS|$, $-|RS|$ must also be a root. This implies that all possible $|RS|$ values appear together with their negative associates.

We know that there are exactly $\varphi(n)/2$ indistinguishable cases among the regular n-gons, including the star-regular ones. It is possible that some cases produce different $|RS|$ values, but some of them may be the same. If all of them are different, we have $\varphi(n)/2$ different roots, each with meaningful geometry. Because of all roots appear with their negative associates, the minimal polynomial t is of degree $\varphi(n)$ in the case of maximal number of roots.

With further observations it can be conjectured that $\deg t$ divides $\varphi(n)$ (see also Example 4 below).

Also, as seen before in section 2, Theorem 1, $|RS|$ always appeared to an even power in t. This is true in general as well (see [23, chapter 7, §2, Example 12]).

Example 1. By considering $n = 24$, case 48, the following output is mined by `RegularNGons` (note that $\varphi(24) = 8$): $A = 0$, $B = 1$, $C = 0$, $D = 2$, $E = 0$, $F = 1$, $G = 2$, $H = 8$,

$$t = 4|RS|^8 - 72|RS|^6 + 288|RS|^4 - 324|RS|^2 + 81,$$

and the possible positive $|RS|$ values are

$$\frac{-\sqrt{3}+\sqrt{6}+3}{2}, \frac{\sqrt{3}-\sqrt{6}+3}{2}, \frac{\sqrt{3}+\sqrt{6}+3}{2}, \frac{\sqrt{3}+\sqrt{6}-3}{2}.$$

Example 2. When checking $n = 23$, case 70, the outputs are (note that $\varphi(23) = 22$): $A = 0$, $B = 1$, $C = 0$, $D = 2$, $E = 0$, $F = 1$, $G = 3$, $H = 13$,

$$\begin{aligned} t =&|RS|^{22} - 228|RS|^{20} + 5618|RS|^{18} - 52167|RS|^{16} + 221675|RS|^{14} + \\ &- 490131|RS|^{12} + 590069|RS|^{10} - 378575|RS|^{8} + 117198|RS|^{6} + \\ &- 13963|RS|^{4} + 503|RS|^{2} - 1, \end{aligned}$$

and the possible positive $|RS|$ values are 0.0459, 0.2424, 0.3734, 0.7426, 1.0002, 1.1919, 1.3209, 1.4892, 3.0158, 3.2263, 14.1901. Notably, the 5th value is very close to 1. This result (among many others) supports creating new, tricky tasks on disproving facts that are visually not decidable. On the other hand, by searching for good approximations some remarkable numerical results can also be achieved, say, by finding close values to non-algebraic numbers like e or π (see subsection 4.5 for some other examples).

Also, in the current example, when considering the factorization $t = t_1 \cdot t_2$,

$$\begin{aligned} t_1 =&|RS|^{11} - 24|RS|^{10} + 174|RS|^{9} - 543|RS|^{8} + 703|RS|^{7} - 5|RS|^{6} + \\ &- 861|RS|^{5} + 679|RS|^{4} - 34|RS|^{3} - 107|RS|^{2} + 17|RS| + 1, \\ t_2 =&|RS|^{11} + 24|RS|^{10} + 174|RS|^{9} + 543|RS|^{8} + 703|RS|^{7} + 5|RS|^{6} + \\ &- 861|RS|^{5} - 679|RS|^{4} - 34|RS|^{3} + 107|RS|^{2} + 17|RS| - 1, \end{aligned}$$

we learn that the 11 positive roots take place in such a way that the 1st, 3rd and 5th one (3 roots) are present in t_2, and the other ones (8 roots) are in t_1.

Example 3. Since the situation in Example 2 is difficult to show geometrically, we go back to an easier case in the regular heptagon. In Fig. 6 we computed the length of the diagonal of two opposite vertices (case $n = 7$, case 42: $A = 0$, $B = 1$, $C = 0$, $D = 2$, $E = 0$, $F = 3$, $G = 1$, $H = 3$, which stands for the $\{7\}$, $\{7/2\}$ and $\{7/3\}$ polytopes). The polynomials $t_1 = |RS|^3 - 2|RS|^2 - |RS| + 1$ and $t_2 = |RS|^3 + 2|RS|^2 - |RS| - 1$ can be obtained with positive roots 0.5549, 2.2469 for t_1 and 0.8019 for t_2. This shows that all positive roots have a geometrical meaning. Actually, by considering the absolute values of only one of the factors—that is, either t_1 or t_2—we also get all solutions for $|RS|$.

In this way, however, we admit that there is no real possibility to isolate the different positive roots of t. We need to observe all of them as one, by accepting the fact that our method cannot really distinguish between them.

Example 4. Let us consider the regular heptagon again, case 124: $A = 0$, $B = 1$, $C = 0$, $D = 2$, $E = 1$, $F = 3$, $G = 2$, $H = 6$. Here all variants in the star-regular heptagons result in the same $|RS| = \sqrt{2}$. We note that the degree of t is here just 2 ($t = |RS|^2 - 2$), less than 6, but it is a divisor of it.

Finally we present a result which can be proven by exhaustion:

Theorem 6. *– In a regular heptagon the only rational lengths in $\mathcal{S}_2$ are* 1 *and* 2*, and the only quadratic surd is* $\sqrt{2}$*.*

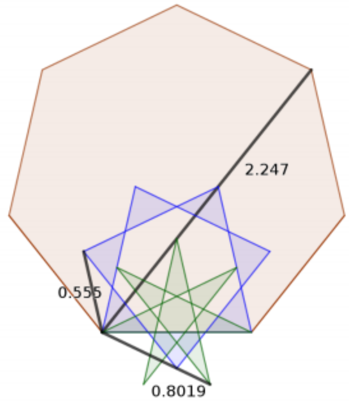


Fig. 6. Lengths of some diagonals in a regular and two star-regular heptagons

– *In a regular nonagon the only rational lengths in* $\mathcal{S}_2$ *are* 1, 2 *and* 3, *and the only quadratic surds are* $\sqrt{3}$ *and* $\sqrt{7}$.
– *In a regular 11-gon the only rational lengths in* $\mathcal{S}_2$ *are* 1, 2, *and the only quadratic surd is* $\sqrt{3}$.

4.5 Approximate Results in Regular 11- and 12-gons

By using the $a = \ldots$ option in `RegularNGons`, one may obtain some "almost"-results that can be interesting when creating tricky problems. Here some results are listed—all involve star-regular polygons.

Example 5. Let us consider the case $n = 11$. By observing case 30781, we get $A = 0$, $B = 1$, $C = 2$, $D = 5$, $E = 4$, $F = 6$, $G = 8$, $H = 10$ that produces $|RS|^{10} - 53|RS|^8 + 732|RS|^6 - 2807|RS|^4 + 3073|RS|^2 - 947$. It has a root 0.9990910 which is near 1.

Let us consider case 31507 now: $A = 0$, $B = 1$, $C = 2$, $D = 6$, $E = 1$, $F = 3$, $G = 6$, $H = 10$. Here the polynomial $|RS|^{10} - 81|RS|^8 + 1465|RS|^6 - 4142|RS|^4 + 2825|RS|^2 - 67$ can be obtained that has a root 1.0003614 which is closer to 1.

In addition, in case 50867: $A = 0$, $B = 1$, $C = 4$, $D = 6$, $E = 2$, $F = 7$, $G = 5$, $H = 8$ yields $|RS|^{10} - 64|RS|^8 + 1029|RS|^6 - 6085|RS|^4 + 13831|RS|^2 - 8713$ which has a root 1.0001111, even closer to 1.

Example 6. Again, assuming $n = 11$, case 40220, we get $A = 0$, $B = 1$, $C = 3$, $D = 5$, $E = 1$, $F = 2$, $G = 6$, $H = 9$: $|RS|^{10} - 130|RS|^8 + 886|RS|^6 - 2147|RS|^4 + 2116|RS|^2 - 727$. One of the roots is 1.66665066 which is very close to 5/3.

Example 7. Let $n = 12$, case 43261, $A = 0$, $B = 1$, $C = 2$, $D = 3$, $E = 0$, $F = 10$, $G = 6$, $H = 9$: Here the polynomial $9|RS|^4 - 240|RS|^2 + 1492$ has a root 3.141533338 which is a 4-digits approximation of π. Note that this is exactly the same value that was given by the Polish Jesuit priest Kochański in 1685 [24], namely $\sqrt{120 - 18\sqrt{3}}/3$, and our approximation can also be constructed by compass and straightedge [25].

4.6 Other Examples

Some other results can be found at `https://www.geogebra.org/m/AXd5ByHX`. The software tool `RegularNGons` can be launched on-line at `http://prover-test.geogebra.org/~kovzol/RegularNGons/`. An example run can be started to request solving the case $n = 5$ by invoking the URL `http://prover-test.geogebra.org/~kovzol/RegularNGons/?n=5`.

5 Conclusion and Future Work

We presented an automated way on obtaining various new theorems on regular polygons, based on the work of [7], [8], [13] and [15]. Enumerating the possible cases was an important detail in our work, we mapped the first non-negative numbers to the possible cases bijectively, however, some cases in our definitions still yield the same segment RS. This case occurs when R or S, or both, are among the vertices of the n-gon. This should be addressed later.

Further theorems can be developed by considering segments of higher kinds, not just of the second. The number of cases to check—according to Proposition 1—grows rapidly. For the third kind, it is asymptotic to $n^{16}/2^{15}$, and is more than 119 billions for $n = 5$. That is, there can be lots of new theorems to explore, even if not all of them are of interest.

The high number of cases calls for distributed computing. Our further plan is to extend our software tool to be a centralized system that assigns interesting tasks to the contributors' computers. By this way the idle computer time could be used to "mine" new geometry theorems.

Acknowledgments

The author was partially supported by a grant MTM2017-88796-P from the Spanish MINECO (Ministerio de Economia y Competitividad) and the ERDF (European Regional Development Fund). Many thanks to Gábor Ancsin, Tomás Recio, Francisco Botana and the GeoGebra Team for their valuable comments and suggestions.

References

1. Lenat, D.: Automated theory formation in mathematics. Contemporary Mathematics **29** (1984) 287–314
2. Chou, S., Gao, X., Zhang, J.: A deductive database approach to automated geometry theorem proving and discovering. Journal of Automated Reasoning **25** (2000) 219–246

3. Bagai, R., Shanbhogue, V., Zytkow, J., Chou, S.: Automatic theorem generation in plane geometry. In: Lecture Notes in Artificial Intelligence 689. (1993) 415–424
4. Chen, X., S.D.W.D.: Automated generation of geometric theorems from images of diagrams. Annals of Mathematics and Artificial Intelligence **74** (2015) 333–358
5. Gao, H., Goto, Y., Cheng, J.: A set of metrics for measuring interestingness of theorems in automated theorem finding by forward reasoning: A case study in nbg set theory. In: Lecture Notes in Computer Science. Volume 9243. (2015) 508–517
6. Colton, S., Bundy, A., Walsh, T.: On the notion of interestingness in automated mathematical discovery. International Journal of Human-Computer Studies **53** (2000) 351–375
7. Chou, S.C.: Mechanical Geometry Theorem Proving. Springer Science + Business Media (1987)
8. Wu, W.T.: On the Decision Problem and the Mechanization of Theorem-Proving in Elementary Geometry (1984)
9. Wantzel, P.: Recherches sur les moyens de reconnaître si un problème de géométrie peut se résoudre avec la règle et le compas. Journal de Mathématiques Pures et Appliquées **1** (1837) 366–372
10. Sethuraman, B.: Rings, Fields, and Vector Spaces: An Introduction to Abstract Algebra via Geometric Constructibility. Springer (1997)
11. Pierpont, J.: On an undemonstrated theorem of the disquisitiones arithmeticæ. Bulletin of the American Mathematical Society **2** (1895) 77–83
12. Gleason, A.M.: Angle trisection, the heptagon, and the triskaidecagon. The American Mathematical Monthly **95** (1988) 185–194
13. Watkins, W., Zeitlin, J.: The minimal polynomial of $\cos(2\pi/n)$. The American Mathematical Monthly **100** (1993) 471–474
14. Kapur, D.: Using Gröbner bases to reason about geometry problems. Journal of Symbolic Computation **2** (1986) 399–408
15. Recio, T., Vélez, M.P.: Automatic discovery of theorems in elementary geometry. Journal of Automated Reasoning **23** (1999) 63–82
16. Coxeter, H.S.M.: Regular Polytopes. 3. edn. Dover Publications (1973)
17. Mersenne Research, Inc.: The Great Internet Mersenne Prime Search (1997) `https://www.mersenne.org`.
18. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system (2009) `https://bitcoin.org/bitcoin.pdf`.
19. Tassev, L.: Hackers target 400,000 computers with mining malware (2018) `https://news.bitcoin.com/hackers-target-400000-computers-with-mining-malware/`.
20. Mersenne Research, Inc.: GIMPS project discovers largest known prime number: $2^{77,232,917-1}$ (2018)
21. Kovács, Z.: RegularNGons. A GitHub project (2018) `https://github.com/kovzol/RegularNGons`.
22. Haas, A., Rossberg, A., Schuff, D.L., Titzer, B.L., Gohman, D., Wagner, L., Zakai, A., Bastien, J., Holman, M.: Bringing the web up to speed with WebAssembly. In: Proceedings of the 38th ACM SIGPLAN Conference on Programming Language Design and Implementation. Association for Computing Machinery (2017) 185–200
23. Cox, D., Little, J., O'Shea, D.: Ideals Varieties, and Algorithms. Springer New York (2007)
24. Kochański, A.A.: Observationes cyclometricae adfacilitandam praxin accomodatae. Acta Eruditorum **4** (1685) 394–398
25. Kovács, Z.: Constructing π from a regular star-12-gon. GeoGebra (2018) `https://www.geogebra.org/m/jnZSeBnq`.

Towards A Mechanisation in Isabelle of Birkhoff's Ruler and Protractor Geometry

Imogen I. Morris and Jacques D. Fleuriot

The University of Edinburgh, Edinburgh, Scotland
s1402592@ed.ac.uk jdf@inf.ed.ac.uk

Abstract We report a work-in-progress formalisation of Birkhoff's axioms for metric Euclidean geometry in the interactive theorem prover Isabelle. Because his axioms are strong, it gives us a head start in proving high-level theorems, without needing to build up as far from the foundations as in the axiom systems given by Euclid, Hilbert or Tarski. Our formalisation begins with the axioms on line measure, includes theorems on angle measure and finally shows that the measure of a straight angle is π. Birkhoff's presentation is sometimes hard to follow so we use instead a more precise rewriting of his axioms given by Brossard in 'Birkhoff's Axioms for Space Geometry'.

Keywords: G. D. Birkhoff · Metric geometry · Isabelle · Archimedean Axiom

1 Introduction

Birkhoff's paper 'A Set of Postulates for Plane Geometry based on Scale and Protractor' lays out a system of axioms for metric Euclidean geometry [2]. He gives both a metric on lines ('distance') and on angles ('angle measure'). Because his axioms are strong, it gives us a head start in proving high-level theorems, without needing to build up as far from the foundations as in the axiom systems given by Euclid [5], Hilbert [6] or Tarski [15]. We formally verify results from Birkhoff's paper in the interactive theorem prover Isabelle [12]. Isabelle has theorems, lemmas, definitions and proofs, as well as some automatic proof methods, and all of this can be written in a structured proof language Isar [11], which is designed to read similarly to a traditional pen-and-paper proof. Our formalisation begins with his axioms on line measure, includes theorems on angle measure and finally shows that the measure of a straight angle is π.

Unfortunately, Birkhoff's presentation is sometimes hard to follow (e.g. he seems to be missing some axioms on lines, has unnecessary undefined notions and omits proofs). So we follow instead a more precise rewriting of his axioms given by Brossard [3]. Although Brossard states that his motivation for this paper was to allow 3 dimensional constructions, the main difference for us is that he gives a more detailed treatment and more formal statements of the axioms. A set of axioms, building on Birkhoff's, but adding some postulates and discussing separation of the plane, is given by MacLane [7]. We use MacLane's treatment to inform and explain our mechanisation. We also explore the relation between Birkhoff's and Brossard's Continuity Axioms and the Archimedean Axiom.

1.1 Birkhoff's Axioms

In the 1920s Birkhoff came up with a way of representing metric Euclidean geometry in terms of the ruler and protractor, as opposed to the straight-edge and compass of the Ancient Greek constructions. Birkhoff, with his collaborator Beatley, saw that this approach was perfectly suited to teaching demonstrative geometry with the purpose of introducing formal proof to students. They noticed that

The traditional approach to demonstrative geometry involves careful study of certain theorems which the beginner is eager to accept without proof and which he might properly be led to take for granted as assumptions or postulates. Such an approach obscures at the very outset the meaning of "proof" and "demonstration". The employment of superposition in the proof of some of these theorems is even more demoralising. [1, p. 3]

Their geometry, which comes with the real numbers embedded in from the beginning, gives a strong enough system that the students get a head start, allowing them to prove all theorems without needing to skip proofs or use unjustified methods such as superposition, which enables any figure to be placed on top of any other. Yet the axioms retain a certain simplicity and intuitiveness coming from the practical representation of the two main axioms as a ruler and protractor.

Their axiom system was later expanded by the School Mathematics Study Group for use in several schools across America. Through this Birkhoff and Beatley's work influenced the use of the 'number line' (their 'ruler postulate') to connect geometry with numbers [16]. The concept of a number line is still used in schools around the world today. Therefore to understand it more deeply is important for its own sake, since it will help us to comprehend the geometry that most people have been taught.

1.2 Related Mechanisations

Much geometry has been mechanised, for example Hilbert's 'Grundlagen' [4,9,14] and Tarski's axioms [8,10], however, little metric geometry has been mechanised from axioms upwards and we know of none based on Birkhoff's system. To show the limit of mechanised metric geometry we note that there is no formally verified definition of sine and cosine from geometry. That is one planned application of this work.

2 Lines and the Ruler Postulate

Birkhoff and Brossard choose to represent lines as point sets. Brossard states the axioms about lines as follows:

L_1. *There exist at least two distinct points.*
L_2. *If A and B are two distinct points, then there exists one and only one line containing A and B.*
L_3. *There exist points not all on the same line.*

Birkhoff misses out the L_1 and L_3 axioms, perhaps believing that his readers will assume them nonetheless. He names his version of L_2 the 'point-line postulate'. We formalise L_1 - L_3 as

```
locale Lines =
 fixes isLine :: " 'p set ⇒ bool"
 assumes
 brossard_line1: "∃ (A::'p) B. A ≠ B"
 and point_line_brossard_line2: "A ≠ B ⟹ ∃! l. isLine
   l ∧ A ∈ l ∧ B ∈ l"
 and brossard_line3: "∃ A B C. ¬(∃ l. isLine l ∧ A ∈ l
   ∧ B ∈ l ∧ C ∈ l)"
```

We formalise the postulates in a **locale**. This is a mechanism in Isabelle for treating a set of statements as conditions for the theorems in the context, as opposed to accepting them as truths in the way of axioms. This is useful for two reasons. Firstly, if a contradiction arises as a result of the statements, it will only permeate the parts of the theory dependent on the locale and will not introduce any inconsistency at the theory level. Secondly, it allows instantiation of the locale conditions to an existing theory. For example, we could instantiate our lines and points to 'tables and beer mugs' as Hilbert suggested[1], or to Euclid's theory of lines and points.

Note the `'p` which is a type variable, here representing the type of points. This allows points to be instantiated to any kind of object, e.g. pairs of real coordinates, so long as it satisfies the locale axioms.

The formalisation of `brossard_line3` is a literal translation of what Brossard wrote: it could be shortened to the equivalent `isLine l ⟹∃A. A ∉ l`.

Brossard and Birkhoff discuss 'the line AB' only when the points A and B are distinct. However, if we defined the line on two points using this condition, it would mean that all the lemmas proved about it would become conditional also. So, we define it to be some line through the points A and B, using the Hilbert choice operator.

```
definition "line A B ≡ (SOME l. isLine l ∧ A ∈ l)"
```

Given some expression `SOME` $b.\ P\,b$ where P is a predicate, this refers to an arbitrarily chosen object b which satisfies the condition P. Unless the object is unique, we cannot tell which object is chosen. If there is no such object, then `SOME` $b.\ P\,b$ is undefined. If A and B are distinct, the uniqueness of the line is guaranteed by the axiom `point_line_brossard_line2`. If $A = B$, then `line` $A\,B$ is an arbitrary line through that point. Since there is more than one

[1] Actually this would only be possible for our theory if tables happened to be sets of beer mugs.

point in the plane by axiom `brossard_line1`, we know that there is at least one line through any point. Thus, whether or not $A = B$, we can prove the same generic facts about `line` $A\,B$, e.g.

```
lemma line_bestdef: shows "(line A B) ∈ Line"
   and "A ∈ (line A B)" and "B ∈ (line A B)"
```

and

```
lemma line_sym: "line A B = line B A"
```

Instead of using the predicate `isLine` we use the set `Line` defined by this predicate.

```
definition "Line = {l. isLine l}"
```

This simplifies expressions of the form $\exists l.$ `isLine` $l \wedge P\,l$ to $\exists l \in$ `Lines.` $P\,l$ which is more concise.

2.1 Ruler Postulate

Both Birkhoff and Brossard have a postulate corresponding to the ruler, however Brossard expresses it differently. Birkhoff takes the distance between two points A, B as a primitive $d(A, B)$ and gives the line measure axiom as

Postulate I. *The points A, B, ... of any line I can be put into* $(1, 1)$ *correspondence with the real numbers* x *so that* $\mid X_B - X_A \mid = d(A, B)$ *for all points A, B.*

Birkhoff follows his axiom with a throwaway statement that any second system of numeration is a shift and optional inversion of the first. This gives us some trouble to prove in Brossard's system later. Brossard gives the line measure axiom as

CL_1. *There exists associated with each line L, a nonempty class X of one-to one mappings* x *of L onto the field* $\mathbb{R}$ *of real numbers. If* x_i *is a member of X and if* x_j *is any one-to-one mapping of L onto R, then* x_j *is a member of X if and only if for all* $A \in L$ *and for all* $B \in L$, $\mid x_i(A) - x_i(B) \mid = \mid x_j(A) - x_j(B) \mid$.

We interpret each 'coordinate function' x as being a bijection since x is 'one-to-one' and 'onto' according to Brossard, and a '$(1, 1)$ correspondence' in Birkhoff's words.

Brossard's treatment is equivalent to Birkhoff's but avoids the unnecessary primitive d and emphasises the fact that lines have (uncountably) many systems of measurement. Brossard's class X of coordinate functions acknowledges that we could choose different starting points (and even different directions of incrementation) for the coordinate system. We know from experience of Euclidean geometry that we could also choose a different line segment to have unit length, but that would lead to an altogether different class X of coordinate functions. Brossard's axiom does not describe that case.

As we will find later (see Section 5), Brossard has a symmetry in his formulations of the line measure and angle measure axioms, via the class of coordinate functions and the bundle of half lines. The symmetry is less obvious in Birkhoff's statements. We formalise the ruler postulate as

```
locale Line_Measure = Lines isLine
  for isLine ::"'p set ⇒ bool" +
  fixes Coord  :: "'p set ⇒ ('p ⇒ real) set"
  assumes
  brossard_line_measure1:
  "l ∈ Line ⟹ ∃ x. x ∈ Coord l"
  and brossard_line_measure2:
  "⟦l ∈ Line; x ∈ Coord l⟧
  ⟹ bij_betw x l (UNIV::real set)"
  and brossard_line_measure3:
  "l ∈ Line
  ⟹ ⟦x_i ∈ Coord l ; bij_betw x_j l (UNIV::real set)⟧
  ⟹ ((x_j ∈ Coord l) =  ∀ A ∈ l. ∀ B ∈ l.
  ¦x_i A - x_i B¦ = ¦x_j A - x_j B¦ )"
```

Here, `(UNIV::real set)` refers to $\mathbb{R}$, the 'universe of real numbers'. Notice that we have split the single postulate into as many parts as possible. This makes it easier to read and reuse. The predicate `bij_betw` A B is true iff there exists a bijection between A and B.

We can now simply define distance as `¦x A - x B¦` for each coordinate function `x`. This is initially relative to the coordinate function, but the ruler postulate allows us to prove its independence.

```
definition "x ∈ Coord (line A B)
⟹ distance_rel x A B ≡ ¦x A - x B¦"
```

This is following Brossard: we saw above that Birkhoff takes distance as a primitive.

3 Betweenness

Birkhoff and Brossard define the notion of betweenness for three points on a line. Their definition is of strict betweenness, so none of the points may coincide. Betweenness is in terms of the coordinate functions on a line (in Brossard's terminology, which we follow, or in terms of the 'system of numeration in the line' as Birkhoff puts it). Here we see that Brossard's definition is less redundant, since it is both necessary and possible to prove that betweenness is independent of the particular coordinate function chosen for the line, whereas in Birkhoff this generality is only implicit.

Again, betweenness is initially defined relative to the coordinate function of the line.

```
definition "between_rel x A B C ≡
(if (C ∈ (line A B) ∧ x ∈ Coord (line A B))
then (x A < x B ∧ x B < x C ∨ x C < x B ∧ x B < x A)
else False)"
```

The proof that betweenness is independent of the coordinate function is more interesting. Again, this is only proved by Brossard and not by Birkhoff. Birkhoff's reasoning is quoted: (Given three points A, B and C on a line)

If B is between A and C with respect to a coordinate function x_i then

$$(1)\ x_i(A) < x_i(B) < x_i(C) \ \textit{or else}\ (2)\ x_i(C) < x_i(B) < x_i(A)$$

Let x_i be an arbitrary coordinate function for the same line. Axiom CL_1 implies that

$$\begin{aligned}
&(3)\ x_i(A) - x_i(B) = x_j(A) - x_j(B) \ \textit{or else}\\
&(4)\ x_i(A) - x_i(B) = x_j(B) - x_j(A),\\
&(5)\ x_i(A) - x_i(C) = x_j(A) - x_j(C) \ \textit{or else}\\
&(6)\ x_i(A) - x_i(C) = x_j(C) - x_j(A),\\
&(7)\ x_i(B) - x_i(C) = x_j(B) - x_j(C) \ \textit{or else}\\
&(8)\ x_i(B) - x_i(C) = x_j(C) - x_j(B).
\end{aligned}$$

If (1) is valid then all left members of (3), (4), (5), (6), (7), (8) are negative and if (2) is valid the same left members are all positive. Equations (3), (5), and (7) are valid or else equations (4), (6), and (8) are valid because any two equations in a group implies the third one. Consequently with respect to x_j, B is also between A and C. [3, p. 594].[2]

In our formal proof we could easily prove this by going through all eight cases, but it would be time-consuming, unreadable and would not capture Brossard's reasoning. We see that his reasoning here could apply to more than just this example. We would like to capture in our formal proof what Brossard means by saying that one whole group of equations must be true, or the other group must be true 'because any two equations in a group implies the third one'. We notice it would be true for any set of six propositions which come in pairs of the form P and $\neg P$, and for which any two in a group imply the third, as Brossard says. We formulate the lemma as follows, indexing the propositions by natural numbers up to 3 so that 'any two in a group imply the third' can be formalised more abstractly than simply listing all of the implications.

[2] Section numbers relevant only to Brossard's paper have been omitted.

```
lemma two_imply_third:
assumes "(P::nat ⇒ bool) 0 ∧ P 1 ⟶ P 2" and
                        "P 1 ∧ P 2 ⟶ P 0" and
                        "P 2 ∧ P 0 ⟶ P 1" and
                     "¬P 0 ∧ ¬P 1 ⟶ ¬P 2" and
                     "¬P 1 ∧ ¬P 2 ⟶ ¬P 0" and
                     "¬P 2 ∧ ¬P 0 ⟶ ¬P 1"

shows "(∀n.(0≤n∧n≤2)⟶ P n) ⟷ ¬ (∀n.(0≤n∧n≤2)⟶ ¬P n)"
```

The proof of this proposition requires six cases, however, the advantage of proving it separately is still clear. It is cleaner to prove than the version in terms of the six equations because it is stripped to only the essentials. Also when we prove the independence of betweenness we are not distracted by any messy cases. Finally, the statement of the proposition expresses Brossard's reasoning: which in itself is interesting to represent formally. We instantiate `P 0` as `x_i A - x_i B = x_j A - x_j B`, which is equation 3 above, and `P 1` as equation 4 above etc. and hence prove the theorem.

Now that we have proved that betweenness is independent of coordinate function, we can define it independently.

```
"between ≡
(SOME between. ∀ A B C. ∀x ∈ Coord (line A B).
between A B C = between_rel x A B C)"
```

4 Angles and the Protractor Postulate

4.1 Half Lines

Before we define angles, we need the notion of half-lines or rays. Again, Birkhoff and Brossard mention the half line defined by point A and endpoint X. We wanted to define this concept to agree with the previous definition of a line through two points. Thus the half line $X\ A$ needs to be defined even in the degenerate case $X = A$ and since we would like to show `halfline` $XA \subset$ `line` XA we need to define the arbitrarily chosen half line to agree with the arbitrarily chosen line.

```
definition
"halfline X A ≡
(if X ≠ A then {P. ¬between A X P ∧ P∈line X A}
else SOME h. ∃B∈line X A. B ≠ X
∧ h = {P. ¬between B X P ∧ P∈line X A})"
```

We cannot reduce this definition to the single case of some half-line with a particular endpoint and through a particular line (as we did for lines) because there are in fact two half-lines with the same endpoint on each line. Now we can proceed to the definition of angles in terms of two half-lines.

5 Angle Measure Locale

5.1 Modular Arithmetic in the Reals

There is some ambiguity in the word 'angle' because it can be taken to mean either the object composed of two half lines, or the size of the object. For clarity, we call the second 'measured angles'. Measured angles are defined to take values in $[0, 2\pi)$. They form a cyclic group under addition because adding angles up to a full turn gives you an angle of measure 0. Thus we need to use modular arithmetic to describe their values. We take the measure of a full angle, 2π, to be 4, partly because this is an integer and easier to prove certain theorems about modular arithmetic and partly because this ties in with some of our ongoing work on defining sine and cosine from the reals to reals axiomatically where $\frac{\pi}{2}$ is taken as 1. Another motivation for choosing 4 as the measure of a full angle is that it divides angles into four quadrants which is often a useful distinction e.g. when evaluating sine and cosine or the argument of a complex number.

Modular arithmetic on the integers already has some defined notions and theorems in Isabelle. However, none of this is defined for the reals. So we create a separate theory in which we develop arithmetic modulo 4 on the reals. It would be more useful to develop modular arithmetic on the reals for any real number r, not just 4, however it was non-obvious how to formulate and prove the following theorem for any r.

```
lemma int_consec4div:"(∃n. (k::int) = 4 ∗ n) ∨
(∃n. k + 1 = 4∗n)∨ (∃n. k+2 = 4∗n) ∨ (∃n. k+3 = 4∗n)".
```

This theorem is used to prove that there is a unique canonical representative of the equivalence class modulo 4.

Equality modulo 4 is defined as

```
 definition eq4 :: "real → real → bool" (infix "=4" 61)
where "((a::real) =4 b) = (∃(n::int). a=b+(4::int)∗n)"
```

We prove that it is an equivalence relation. We also have the function `mod4` which takes any real number to the canonical representative of the equivalence class modulo 4. It would be possible to define this constructively using the Euclidean algorithm, but here we just formulate it in terms of the condition that the canonical representative must satisfy.

```
  definition mod4 :: "real → real"
where "(mod4 a) = (THE b. 0≤b ∧ b < 4 ∧ a =4 b)"
```

Given some expression `THE b. P b` where `P` is a predicate, this refers to the unique object `b` which satisfies the condition `P`. If there is no unique object, then `THE b. P b` is undefined.

Since any theorem which holds for `eq4` has a version which can be written in terms of `mod4`, many of the theorems are duplicated e.g.

```
 lemma mod4_2_inv[simp]: "mod4( - 2) = 2"
```

and

```
lemma eq4_2_inv: "- 2 =4 2".
```

It would be convenient and more concise if this could be formalised as a general principle rather than repeating the proofs for each of these theorems.

Much of the reasoning in Isabelle concerning the reals is integrated into automatic proof methods or sets of simplification theorems which makes it easier to use and reduces the need to work out which theorems are required and in which order. However, most the reasoning concerning arithmetic modulo 4 has to be done explicitly, since for now, only limited theorems have been added to the simplifier.

5.2 Angle Measure and Bundles

Brossard (but not Birkhoff) defines the notion of bundles as ‘certain subclasses of the class of all half-lines with the same endpoint’. These are useful later in Brossard’s paper since each bundle defines a plane, and he discusses three dimensional geometry. He gives an axiom concerning the uniqueness of the bundle.

CB_1. *If l and m are two noncollinear half-lines with the same end-point O, then there exists one and only one bundle B_0 containing these half-lines.*

Our locale captures the axiom, as `brossard_bundle1`, as well as the definition of the bundle, as `brossard_bundle2`.

```
 locale Bundles = Line_Measure isHalfLine
  for isHalfLine ::"'p set ⇒ bool" +
  fixes isBundle  :: "('p set) set ⇒ bool"
  assumes
  brossard_bundle1:
"⟦l≠m;¬(∃ L ∈ Line. l ∪ m = L);
l ∈ HalfLine;m ∈ HalfLine;endpoint l = endpoint m⟧
⟹ ⟦ ∃!B. isBundle B ∧ l∈B ∧ m∈B"
brossard_bundle2:⟦ isBundle B; l∈B ; m∈B⟧⟹
endpoint l = endpoint m ∧ l ∈ HalfLine ∧ m ∈ HalfLine"
```

where we define

```
definition "endpoint h = (THE O.∃ A. h = halfline O A)"
```

Given the notion of bundles, Brossard is then able to state the axiom corresponding to the protractor.

CB_1. *There exists, associated with each bundle B, a nonempty class Φ of one-to-one mappings ϕ of B_0 onto the equivalence classes of real numbers modulo 2π. If ϕ_i is a member of Φ and if ϕ_j is any one-to-one mapping of B_0 onto the equivalence classes of real numbers modulo 2π, then ϕ_j is a member of Φ if and only if for all $l \in B_0$ and for all $m \in B_0$.*

$$| \phi_i(l) - \phi_i(m) | \equiv | \phi_j(l) - \phi_j(m) |,$$

where $\mid \phi_i(l) - \phi_i(m) \mid \equiv \mid \phi_j(l) - \phi_j(m) \mid$ *stands for*

$$\phi_i(l) - \phi_i(m) = (\phi_j(l) - \phi_j(m))(\textit{modulo } 2\pi)$$

or

$$\phi_i(l) - \phi_i(m) = (\phi_j(m) - \phi_j(l))(\textit{modulo } 2\pi).$$

In Birkhoff's angle measure postulate he assumes that all half-lines with the same endpoint can be put into correspondence with the real numbers in this way. So Brossard's concept of bundles allows his axiom more generality. Explicitly using bundles in his angle measure postulate also emphasises the symmetry of the angle and line measure postulates: bundles correspond with lines, points with half-lines, equality with equivalence modulo 2π (or modulo 4 in our formalisation) and all else is the same. Similarly we call the functions in the class Φ, coordinate functions.

We break the angle measure axiom down in a similar way to the line measure axiom as we put it in a locale.

```
locale Angle_Measure = Bundles Coord
  for Coord :: "'p set ⇒ ('p ⇒ real) set"  +
  fixes Acoord
  :: "('p set) set ⇒ ('p set ⇒ real) set"
  assumes
  brossard_angle_measure1:
  "B ∈ Bundle ⟹ ∃ ϕ. ϕ ∈ Acoord B"
  and brossard_angle_measure2:
  "⟦B ∈ Bundle; ϕ ∈ Acoord B⟧
             ⟹ bij_betw ϕ B {x. 0≤x ∧ x<4}"
  and brossard_angle_measure3:
  "B ∈ Bundle
 ⟹ ⟦ϕ_i ∈ Acoord B ; bij_betw ϕ_j B {x. 0≤x ∧ x<4}⟧
 ⟹ ((ϕ_j ∈ Acoord B)
 ⟷ ∀ l∈B. ∀ m∈B. ¦ϕ_i l - ϕ_i m¦ =4 ¦ϕ_j l - ϕ_j m¦)"
```

Again we define angle measure relative to the coordinate function (for angles).

```
definition
 "⟦l ∈ HalfLine;m ∈ HalfLine;endpoint l = endpoint m;
  ϕ ∈ Acoord B; l ∈ B; m ∈ B; B ∈ Bundle⟧
  ⟹ ((ameasure_rel ϕ l m)::real)
  = min (mod4 (ϕ l - ϕ m)) (mod4 (ϕ m - ϕ l))"
```

Notice the use of `min`: this is because given two half-lines, there are two possible angle measures between them: the larger and the smaller. Using `min` ensures we always take the smaller, i.e. the proper angle.

We prove its independence, but this is a little more complicated than for line measure[3]. We need to prove first that the angle measure is independent of the particular coordinate function chosen for the bundle

```
lemma measure_independence: assumes "B_O ∈ Bundle"
"(halfline O A) ∈ B_O" "(halfline O B) ∈ B_O"
shows "∃ameasure. ∀φ∈Acoord B_O. ameasure =
ameasure_rel φ (halfline O A) (halfline O B)"
```

and then that it is independent of the bundle which it is in

```
lemma measure_independent_of_bundle:
assumes "B ∈ Bundle" "C  ∈Bundle" "l ∈ B" "m ∈ B"
"l ∈ C" "m ∈ C""φ ∈ Acoord B" "ψ ∈ Acoord C"
shows "ameasure_rel φ l m = ameasure_rel ψ l m"
```

There is usually a unique bundle for each angle, i.e. for each pair of half-lines, by Brossard's axiom. However, when the half-lines are collinear, there may be more than one bundle. Since we can prove that the angle between coinciding half-lines is zero, the independence is also proven for this case. In the case that the pair of half-lines form a line, the angle is always π no matter which bundle is chosen. However, to prove this final piece of the independence theorem, we need one more axiom.

6 The Archimedean Axiom and Continuity

Here we discuss the meaning of the final axiom in our formalisation and how it relates to Birkhoff's paper. Brossard introduces an axiom which he calls the Continuity Axiom.

Axiom 1 (Brossard's Continuity Axiom) *If B_O is a bundle of vertex O, and if A, B are distinct nonvertex points of noncollinear half-lines of the bundle, then to every point P on the segment AB, there exists a half-line OC of B_O containing P such that*

$$[\angle AOP + \angle POB] = [\angle AOB].$$

Conversely if a half-line OC of the bundle B_0 is such that $[\angle AOC + \angle COB] = [\angle AOB]$, then there exists a point P belonging simultaneously to the half-line OC and to the segment AB.[4]

[3] The independence of line measure is easier to deal with because there is a unique line through both points except in the case that the points coincide, and that case is simple.

[4] The second half of the Continuity Axiom (which begins 'Conversely...') is not yet used in our proofs: it will be interesting to note what can be proved without it.

We formalise this as

```
 locale Continuity = Angle_Measure  isLine
  for isLine ::"'p set ⇒ bool" +
  assumes brossard_continuity:
  "⟦B_O  ∈ Bundle;l ∈ B_O;m ∈ B_O;A≠(endpoint l);
  B≠(endpoint m);A≠B;A∈l;B∈m; ϕ ∈ (Acoord B_O);
  ¬(∃ L ∈ Line. l ∪ m = L); between A P B ∨ P=A∨P=B⟧
  ⟹ ∃C. (halfline (endpoint l) C)∈B_O
  ∧ P ∈ (halfline (endpoint l) C)
  ∧
ameasure_rel ϕ
   (halfline (endpoint l) A) (halfline (endpoint l) P)
+
ameasure_rel ϕ
   (halfline (endpoint l) P) (halfline (endpoint l) B)
=4
ameasure_rel ϕ
   (halfline (endpoint l) A) (halfline (endpoint l) B)"
```

This does not relate directly to any axiom given by Birkhoff. However, as part of his Postulate III, Birkhoff mentions something which appears to be what Brossard based his continuity axiom on.

POSTULATE I. *The half-lines l, m, . . ., through any point O can be put into* $(1,1)$ *correspondence with the real numbers $a\,(mod\ 2\pi)$, so that, if $A \neq O$ and $B \neq O$ are points of l and m respectively, the difference $a_m - a_l\,(mod\ 2\pi)$ is* $\angle AOB$. Furthermore, if the point B on m varies continuously in a line r not containing the vertex O, the number a_m varies continuously also.

We will name the last sentence of Postulate III as Birkhoff's Continuity Axiom thereafter. Birkhoff appends the following footnote to it.

More precisely, $\lim a_m = a_l$ *if* $\lim_{B=A} d(B,A) = 0$ *for points B, A of such a line m (see Fig. 6). It is the second part of Postulate III that excludes "non-Archimedean" possibilities ...*

His figure is reproduced here as Figure 1.

His mention of non-Archimedean possibilities is intriguing. One possibility that comes to mind is non-Archimedean fields such as the hyperreals but it is not clear how half-lines could form a field. However, there is also the Archimedean axiom in geometry which is sometimes called the Continuity Axiom [13, p. 991]. This can be stated for line segments as follows.

Given two segments AB and CD then there is an $n \in \mathbb{N}$ such that laying CD contiguously n times along the segment AB will eventually exceed AB.

However, because of the analogy between line segments and angles, we can restate it for angles.

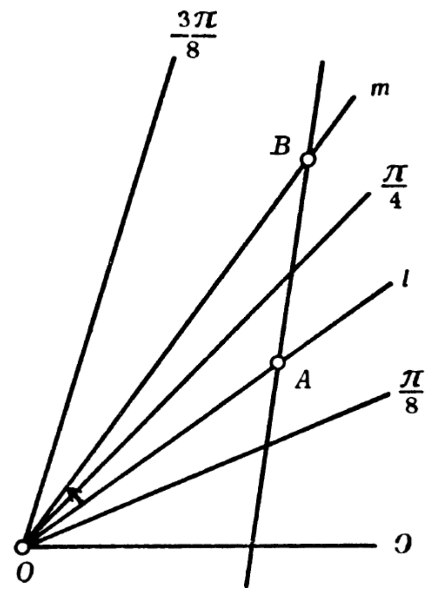


Figure 1. The figure illustrating Birkhoff's remark

Given two angles $\angle AOB$ *and* $\angle COD$ *then there is an* $n \in \mathbb{N}$ *such that laying* $\angle COD$ *contiguously* n *times within the* $\angle AOB$ *will eventually exceed* AB.

This does not yet make sense for the axiom systems of Brossard and Birkhoff (e.g. what does it mean to place angles contiguously?) thus we restate it once again more formally. This formal version could be considered slightly arbitrary as there is more than one sense in which the axiom could be interpreted but at least it can be related to the original axiom. We also choose the form of the axiom to match as far as possible Brossard's version of the Continuity Axiom.

Given two angles $\angle AOB$ *and* $\angle COD$*, there exists a half-line* OQ *such that* $\angle AOQ$ *has measure* $n\angle AOB$ *for some* $n \in \mathbb{N}$ *and* $n\angle AOB > \angle AOC$

We have a pen-and-paper proof that this version follows from Birkhoff's Continuity Axiom. It is also possible to prove that Brossard's Continuity Axiom follows from Birkhoff's (MacLane's paper gives a clue to this effect [7, p. 550]). A formal proof of this is ongoing work.

7 Measure of a Straight Line

After the Continuity Axiom is stated, we finally come to Brossard's first theorem.

Theorem I. *The measure of an angle is* π *if and only if this angle is straight.*

We have mechanised Brossard's proof of this and discovered some interesting points. As Brossard points out [3, p. 597], this is the final condition we need for the measure of an angle to be independent of the bundle in which it is measured. We already know that, given a bundle, any coordinate function for that bundle gives the same angle measure. We also know that, given an angle formed of non-collinear half-lines, Brossard's Bundle Uniqueness axiom guarantees that there will be a unique bundle containing the angle. So, in the case of non-collinear half-lines, the angle-measure is trivially independent of the bundle. In the case of a degenerate angle formed by coinciding half-lines, the measure is 0 no matter which bundle is chosen. The only remaining case is when the union of the half-lines forms a line, i.e. the angle is straight. If we show the measure is always π, then we have shown that it does not depend on the bundle.

7.1 Proof of the Theorem

π Implies Straight We formalise the statement as

```
lemma pi_imp_straight: assumes "B_0 ∈ Bundle" "l ∈ B_0"
"m ∈ B_0" "φ ∈ (Acoord B_0)" "ameasure_rel φ l m = 2"
shows "∃ L ∈ Line. l ∪ m = L"
```

Brossard first proves that if an angle has π for measure, then it is straight. The idea behind the proof is to assume that the angle is not straight and find the difference between that and a straight angle. The main results that are used are the first half of the Continuity Axiom, which gives the initial sum for the angle in terms of the straight angle, and to evaluate the sum we use the result that an angle has measure 0 iff it is degenerate.

Use of Continuity Axiom in Proof To all appearances, Brossard uses the Continuity Axiom only once in his proof. He says

... in the unique bundle B_0 containing the half-lines OA, OB, OP the continuity axiom implies that $[\angle AOP + \angle POB] = [\angle AOB]$.

When we try to formally prove this, a problem becomes apparent. There is a unique bundle which OA and OB are contained in by Brossard's Bundle Uniqueness axiom. There is also a unique bundle which OA and OP are contained in by Brossard's Bundle Uniqueness axiom. However, we can't get a unique bundle which OA, OB and OP are contained in by using just that axiom. Another axiom which mentions bundles and half-lines is the axiom on the coordinate functions of bundles. This is also unhelpful since it would only allow us to conclude a half-line was in the bundle if we could show that it was measurable by a coordinate function of the bundle, which we could only conclude if we knew that the half-line was in the bundle, thereby getting us nowhere. The final axiom concerning bundles and half-lines is the Continuity Axiom. This has two halves. The second half allows us to show that a specific half-line is contained in a bundle ... but only if we first prove $[\angle AOP + \angle POB] = [\angle AOB]$ which is exactly the reason why we wanted to prove that OP was in the bundle, thus it does not help. All that remains is the first half of the Continuity Axiom. At first sight, this doesn't allow us to show that a specific half-line is contained in a bundle. It only gives us the existence of a third half-line in the bundle. But because that half-line is required to be through a point of our choosing, we are able to show that the half-line it gives us is exactly the one that we wanted to have in the bundle. So we end up using the Continuity Axiom (see Section 6) once to give us that the three lines are in a unique bundle, and a second time to actually show that the identity holds.

Obtaining a Contradiction Remember Brossard assumes $\angle AOP = \pi$ and $\angle AOP$ is not straight, in order to obtain a contradiction, and so obtains the equation $[\pi + \angle POB] = [\angle AOB]$. Brossard considers possible values of $\angle POB$, and using the result that an angle has measure 0 iff it is degenerate, he finds that it cannot be equal to either π or 0. This is a contradiction since we have defined the angle-measure so that it always gives us an angle up to π in size (since given a pair of half-lines we consider the angle they make to be the smaller angle).

Straight Implies $\boldsymbol{\pi}$ We formalise the statement as

```
lemma straight_imp_measure_pi:
assumes "B_X ∈ Bundle" "l ∈ B_O" "m ∈ B_O"
"ϕ ∈ (Acoord B_O)" "∃ L ∈ Line. l ∪ m = L"
shows "ameasure_rel ϕ l m = 2"
```

The idea of the proof of the other direction is to show that the straight angle lies on the same line as an angle which is chosen to have measure π. Here, the main difference between Brossard's proof and the mechanised proof is the complication of the reasoning in terms of modular arithmetic. It requires showing the following lemmas combining `mod4` and `min` which allow manipulation of the angle measure (recall this is defined relative to the coordinate function as `((ameasure_rel ϕl m)::real)= min (mod4 (ϕl - ϕm))(mod4 (ϕl - ϕm))` where l and m are the half-lines which form the angle ϕ is the coordinate function of the bundle in which l and m are contained.)

```
lemma mod4_min_projection_property:
"0 ≤ (min (mod4 (x-y)) (mod4 (y-x))) ∧
 (min (mod4 (x-y)) (mod4 (y-x))) ≤ 2"

lemma min_mod4_difference_bounds:
"0 ≤ (min (mod4 (x-y)) (mod4 (y-x))) ∧
 (min (mod4 (x-y)) (mod4 (y-x))) ≤ 2"
```

8 Various Other Theorems

It is easy to underestimate the number of theorems which need to be proven for a formal representation of Birkhoff's geometry. Pen-and-paper mathematicians often skip such theorems believing they are obvious or trivial. This may be true much of the time and is why we have not spent much time outlining them, but these invisible theorems form the main part of our theory. Therefore we will provide a few examples to give the feeling of the detail that is required.

8.1 Betweenness Theorems

```
lemma between_sym: "between A B C = between C B A"

lemma between_otherside:
ssumes "A ≠ X" shows "∃B ∈ line X A. between A X B"

lemma sameside_eq_notbetween: assumes "between A X B"
"between A X P"
 shows "¬ between B X P"

lemma collinear_choice_of_between: assumes
"collinear A X P" "A≠X" "A≠P" "P≠X"
  shows "between A P X ∨ between A X P ∨ between P A X"
```

8.2 Theorems on halflines

```
lemma halfline_independence:assumes "B ∈ halfline X A"
   "B ≠ X" shows "halfline X A = halfline X B"

lemma line_built_from_halflines:
assumes "between A X P"
      shows "line A P = halfline X A ∪ halfline X P"

lemma between_imp_eq_halflines: assumes
"between X A P" shows "halfline X A = halfline X P"
```

8.3 Miscellaneous

```
lemma angle_at_origin: assumes "B ∈ Bundle"
"ϕ ∈ Acoord B"  "g∈B" "h∈B"
shows "∃f∈B. ϕ g - ϕ h =4 ϕ f"

lemma coordfn_preserves_distinctness:
assumes "l ∈ Line" and  "x∈ Coord l" and "A ∈ l" and "B
    ∈ l"
shows "(A=B) =(x A = x B)"
```

9 Future and Ongoing Work

Ongoing work on this mechanisation includes formal proofs of the equivalence of the various continuity axioms and proofs of the theorems on triangles. Then moving beyond Brossard's paper, but possibly following his book [1], theorems of circles could be formalised and finally this could be applied to a definition of sine and cosine from geometry.

10 Conclusion

In our mechanisation we mainly followed Brossard's rewriting of Birkhoff's axioms since he had less primitives, greater symmetry between the protractor and ruler postulates and in general was more formal and detailed. We used locales to formalise the axioms which could allow instantiation to equivalent systems. We expanded Brossard and Birkhoff's definitions of lines, half-lines and bundles to cover degenerate cases so that less case-splits were required in proofs involving these concepts. We mechanised the proof of the independence of betweenness and avoided a long case split by formalising the general form of Brossard's reasoning.

As there was previously no suitable theory of modular arithmetic over the reals, we developed our own for arithmetic modulo 4. Reasoning with this theory is rather tedious because it has not been integrated with the automated proof methods available in Isabelle.

Finally, we mechanised the proof of the independence of angle measure. The main lemma of this proof was that an angle has measure π iff it is straight. The proof of this lemma required one more application of the Continuity Axiom than Brossard explicitly states.

Overall, we found that there were almost a hundred lemmas which needed substantial proof in Isabelle, but were simply assumed by Birkhoff or Brossard (some of these were stated in Section 8).

References

1. Birkhoff, G., Beatley, R.: Basic Geometry. AMS Chelsea Publishing Series, AMS Chelsea Pub. (2000), https://books.google.co.uk/books?id=TB6xYdomdjQC
2. Birkhoff, G.D.: A set of postulates for plane geometry, based on scale and protractor. Annals of Mathematics **33**(2), 329–345 (1932), http://www.jstor.org/stable/1968336
3. Brossard, R.: Birkhoff's axioms for space geometry. The American Mathematical Monthly **71**(6), 593–606 (1964), http://www.jstor.org/stable/2312318
4. Dehlinger, C., Dufourd, J.F., Schreck, P.: Higher-order intuitionistic formalization and proofs in hilbert's elementary geometry. In: Richter-Gebert, J., Wang, D. (eds.) Automated Deduction in Geometry. pp. 306–323. Springer Berlin Heidelberg, Berlin, Heidelberg (2001)
5. Euclid, (ed. John Casey): The First Six Books of the Elements of Euclid. Dublin University Press, Dublin, 3rd edn. (1885)
6. Hilbert, D.: The Foundations of Geometry. The Open Court, La Salle, Illinois, 2nd edn. (1950), translator E. J. Townsend
7. MacLane, S.: Metric postulates for plane geometry. The American Mathematical Monthly **66**(7), 543–555 (1959), http://www.jstor.org/stable/2309851
8. Makarios, T.J.M.: The independence of tarski's euclidean axiom. Archive of Formal Proofs (Oct 2012), formal proof development: http://isa-afp.org/entries/Tarskis_Geometry.shtml. Thesis: http://hdl.handle.net/10063/2315
9. Meikle, L., Fleuriot, J.: Formalizing hilbert's grundlagen in isabelle/isar. In: Basin, D., Wolff, B. (eds.) Theorem Proving in Higher Order Logics. LNCS, vol. 2758, pp. 319–334. Springer, Heidelberg (2003)

10. Narboux, J.: Mechanical theorem proving in tarski's geometry. In: Botana, F., Recio, T. (eds.) Automated Deduction in Geometry. pp. 139–156. Springer Berlin Heidelberg, Berlin, Heidelberg (2007)
11. Nipkow, T.: A tutorial introduction to structured isar proofs. Tech. rep., Technische Universität München (2009), http://isabelle.in.tum.de/website-Isabelle2009-1/dist/Isabelle/doc/isar-overview.pdf
12. Paulson, L.C.: Isabelle: A Generic Theorem Prover. Springer-Verlag, Berlin, 1st edn. (1994)
13. Scanlan, M.: Who were the american postulate theorists? Journal of Symbolic Logic **56**(3), 981?1002 (1991). https://doi.org/10.2307/2275066
14. Scott, P.: Mechanising Hilbert's Foundations of Geometry in Isabelle. Msc, University of Edinburgh (2008)
15. Tarski, A., Givant, S.: Tarski's system of geometry. The Bulletin of Symbolic Logic **5**(2), 175–214 (1999), http://www.jstor.org/stable/421089
16. Weaver, J.F.: The school mathematics study group project on elementary-school mathematics. The Arithmetic Teacher **8**(1), 32–35 (1961), http://www.jstor.org/stable/41186496

Mechanising an Independent Axiom System for Minkowski Space-time

Jake Palmer and Jacques D. Fleuriot

School of Informatics, University of Edinburgh, Edinburgh, UK
s1673264@sms.ed.ac.uk jdf@inf.ed.ac.uk

Abstract. We describe a work in progress that takes the first steps in implementing and investigating an axiomatisation of Minkowski space-time whose primitive undefined basis consists of a set of events, a set of paths consisting of events, and a ternary betweenness relation. There are 15 independent axioms in total: 6 so-called axioms of order, 7 axioms of incidence, the axiom of symmetry, and the axiom of continuity. We describe Minkowski space-time and how it relates to special relativity, formalise and correct some of the axioms presented by the original author of the system, develop his proofs, and fill in some gaps. Ultimately the purpose of this work is to try to push and explore the boundaries of automated reasoning in physics. The result is a starting point for a new formal, mechanised foundations for Minkowski space-time in Isabelle/HOL.

Keywords: Minkowski space-time · Geometry · Interactive theorem proving · Isabelle/HOL · Physics

1 Introduction

A formalisation of a physical theory provides four main benefits: it exposes the basic, underlying assumptions and primitive concepts of the theory[1]; it provides a very high degree of confidence in the theory; it provides a means to explore subtly different theories by small modifications of the base assumptions and concepts; and perhaps–in the future–combining or analysing two incompatible, formalised physical theories may help to show which assumptions it is that they make which don't agree.

The aim of this work is to help make possible some of these benefits within Minkowski space-time and special relativity by formalising the 15 axioms presented in Schutz's axiomatic system [8] and provide the beginnings of a body of proofs using this system. A formal set of axioms alone is not very useful without a body proofs which can give us some confidence that the system is sound and which can be inherited for free for anything proven to be a model, such as the usual coordinate model for Minkowski space-time.

[1] Whether those basic assumptions are insightful or not is a debate we will not have here; that is a question best answered by physicists with training in formal logic.

So far we have formalised the first 13 of the 15 axioms and their associated definitions along with a few minor simplifications which do not fundamentally change the original system. With regards to proofs, we have mechanised Schutz's first 10 theorems and lemmas. We use the structured proof language Isar to stick as close to the original prose as possible so that the book remains a useful reference as the development progresses, and to investigate more closely what the original presentation gets wrong or omits.

Proving the independence of the system and providing concrete interpretations in order to prove its consistency or to demonstrate the correctness of the implementation are currently outside of the scope of this work.

2 Background

The text we are working with is *Independent axioms for Minkowski space-time* [8] by John Schutz[2]. The major goals he laid out in the book are:

- To provide a set of independent axioms for Minkowski space-time.
- To prove that the theory is consistent as long as mathematical logic, set theory, and the reals are consistent.
- To show that all models of the axiomatic system are isomorphic, and that the models are isomorphic to the coordinate system as it is usually understood.

and if the work is broadly correct it accomplishes all three. It is for this reason that having a working formalisation of this theory would be so exciting; Schutz's system has great potential for being the de facto formal representation of Minkowski space-time.

Its predecessors include work by Walker [15,14], Szekeres [11], and Schutz [6,7]. Many statements and proofs are based on the developments by Walker and Veblen [12,13], with Hilbert [2] and Moore [4] mentioned as systems also bearing a clear resemblance. Schutz claims that his latest text differs from its predecessors significantly by making no assumption on the direction of time, causality, nor the existence of light signals. While this makes it more philosophically acceptable it makes the initial development more difficult.

There has also been recent work by Németi et al. [1] on formalising a first-order system of axioms for Minkowski space-time and they have developed an implementation of their system in Isabelle/HOL [10]. this bears little resemblance to the axiomatisation of Schutz so we have not been able to use it for help or inspiration. We cannot with authority comment on the physics-related concerns, but Schutz makes the following comment with regards to the second-order Axiom of Continuity in his system:

> "[...] *it is observed that replacement of the second-order Axiom of Continuity by an infinite schema of first-order axioms leads to some space-time models which bear a closer "physical" resemblance to Galilean space-time than to Minkowski space-time."*

[2] There is an earlier (by 24 years) text [6] also by Schutz that uses a different primitive undefined basis which we do not draw on.

Many systems also bake-in some of: congruence of line segments and angles (like Hilbert), collinearity, planes, and so on. These choices have significant consequences for the foundation of the theory, though all formalisations of the same theory will begin to look quite similar (in theory) in the statements of their theorems and their proofs at a high level once a significant foundation has been developed. Another thing that will be affected by the foundational representations is the ability to automate proofs. Whilst this can in principle be offset by more and more higher level theorems, a good foundation can significantly speed up the process.

Owing to its properties as a space-time, Schutz notes that many things which have to be assumed in other theories can be derived:

> *"As a consequence of its richer structure Minkowski space-time can be specified categorically without an axiom of uniqueness of parallelism, without axioms of congruence and without reference to polarity."*

3 Minkowski Space-time and Its Relation to Physics

Minkowski space-time is to special relativity what Euclidean geometry is to classical physics. Though we often refer to it as a space-time, the "time axis" has no special treatment (one might then prefer to call it "Minkowski space" when we are dealing only with the pure theory). In pure Minkowski space-time there is nothing stopping a particle moving along one "spatial axis" only, which means that if one of the axes is to be time, a particle can exist at several different points in time.

An event in this system is a point-like object, which exists at an instantaneous "moment in time". A path can be interpreted as a particle's "world-line", a continuous stream of such events in the direction of its travel [5]. Some of the assumptions, like an event's point-like nature, may break down or require more information when applied to the real world, and while a sufficiently sophisticated axiomatic system should be able to compensate for this[3] it is preferable for our models to not be as complicated as the real, physical world.

Schutz stresses that Minkowski space-time is not the be-all and end-all of special relativity in the same way that Euclidean geometry is not the be-all and end-all of classical physics. Even so he notes that there is little left to do in order to add relativistic mechanics once Minkowski space-time has been described.

Though Minkowski space-time and Euclidean geometry are related, and they share some similar axioms, geometries and space-times are different in important ways: space-times have a notion of "reference frames" and not all events on a path are reachable from a given event on another path.

[3] To the extent that truths about the real world can be stated and derived in a logical system.

4 Axiomatisation

The two undefined primitive sets in this theory are events ($\mathcal{E}$) and paths ($\mathcal{P}$). The word *undefined* is used because they do not have any concrete instantiation; they are just names (with types). Note that by naming these sets we have not yet assumed events or paths (i.e. particles) actually exist. Their existence and how they relate to each other are determined by the axioms.

The single undefined relation is the betweenness relation and is represented using the notation $[abc]$ where a, b, and c are events. As with the primitive sets, this is just a relation of arity three (which we may write $[\_\_\_]$), characterised further by the axioms. Minkowski space-time is therefore the triple: $\mathcal{M} = \langle \mathcal{E}, \mathcal{P}, [\_\_\_] \rangle$.

4.1 The Axioms as an Isabelle Locale

We present the axioms as an Isabelle locale here to make clear the extent of the axiomatic system up front and to highlight the differences in complexity of some of the axioms. A locale is a way to capture common assumptions and names (such as $\mathcal{E}$) so that we don't have to keep restating them as part of the assumptions to each theorem, and keeps inconsistencies from affecting anything outside the locale. We will also be using Isabelle syntax for presenting definitions and proofs.

Axioms S (symmetry/isotropy) and C (continuity) are missing here as we have not yet formalised them fully in Isabelle, which we will say more about in the conclusion (Section 6). Schutz's Axiom I1 states that the set of events is non-empty ($\mathcal{E} \neq \{\}$) and is no longer an axiom here because it appears to be unnecessary; we will discuss this shortly. Axiom I4 references a function `3-SPRAY` which we will not mention again till the conclusion because formalisation is still in progress. We have also included the additional axiom `in_path_event` capturing the relationship $\mathcal{P} \subseteq 2^{\mathcal{E}}$, which is implicit in Schutz.

O1 through O6 are the Axioms of Order and provide the primary characterisation of the betweenness relation. I1 through I7 are the Axioms of Incidence and deal with the basic relationship between events and paths as well as unreachability and dimensionality.

```
locale minkowski-spacetime =
  fixes ℰ :: 'a set
    and 𝒫 :: ('a set) set
    and [___] :: 'a ⇒ 'a ⇒ 'a ⇒ bool
  assumes in_path_event: ⟦Q ∈ 𝒫; a ∈ Q⟧ ⟹ a ∈ ℰ
      and O1: [abc] ⟹ ∃Q ∈ 𝒫. {a, b, c} ⊆ Q
      and O2: [abc] ⟹ [cba]
      and O3: [abc] ⟹ a ≠ c
      and O4: ⟦[abc]; [bcd]⟧ ⟹ [abd]
      and O5: ⟦Q ∈ 𝒫; {a, b, c} ⊆ Q; a ∈ ℰ; b ∈ ℰ; c ∈ ℰ;
               a ≠ b; a ≠ c; b ≠ c⟧ ⟹ [abc] ∨ [bca] ∨ [cab]
      and O6: ⟦Q ∈ 𝒫; R ∈ 𝒫; S ∈ 𝒫; T ∈ 𝒫; Q ≠ R; Q ≠ S; R ≠ S;
```

$$
\begin{array}{ll}
& a \in Q \cap R \wedge b \in Q \cap S \wedge c \in R \cap S;\\
& \exists d \in S.\ [bcd] \wedge (\exists e \in R.\ d \in T \wedge e \in T \wedge [cea])]\!]\\
& \Longrightarrow \exists f \in T \cap Q.\ \exists X.\ [a..f..b]X]\\
\texttt{and I2:} & [\![a \in \mathcal{E};\ \ b \in \mathcal{E};\ \ a \neq b]\!]\\
& \Longrightarrow \exists R \in \mathcal{P}.\ \exists S \in \mathcal{P}.\ a \in R \wedge b \in S \wedge R \cap S \neq \{\}\\
\texttt{and I3:} & [\![P \in \mathcal{P};\ \ Q \in \mathcal{P};\ \ a \in P;\ \ b \in P;\ \ a \in Q;\ \ b \in Q;\ \ a \neq b]\!]\\
& \Longrightarrow P = Q\\
\texttt{and I4:} & \exists x.\ \texttt{3-SPRAY}\ x\\
\texttt{and I5:} & [\![Q \in \mathcal{P};\ \ b \in \mathcal{E};\ \ b \notin Q]\!]\\
& \Longrightarrow \exists x \in \texttt{unreach}_{\subseteq}\ Q\ b.\ \exists y \in \texttt{unreach}_{\subseteq}\ Q\ b.\ x \neq y\\
\texttt{and I6:} & [\![Q \in \mathcal{P};\ \ b \in \mathcal{E};\ \ b \notin Q;\ \ Q_x \in \texttt{unreach}_{\subseteq}\ Q\ b;\\
& Q_z \in \texttt{unreach}_{\subseteq}\ Q\ b]\!]\\
& \Longrightarrow \exists X\ f.\ \texttt{ch\_by\_ordering}\ f\ X \wedge f\ 0 = Q_x\\
& \qquad\qquad \wedge f\ (\texttt{card}\ X - 1) = Q_z\\
& \qquad\qquad \wedge (\forall i \in \{1...\texttt{card}\ X - 1\}.\ f\ i \in \texttt{unreach}_{\subseteq}\ Q\ b\\
& \qquad\qquad\qquad \wedge (\forall Q_y \in Q.\ [(f\ (i-1))\ Q_y\ (f\ i)]\\
& \qquad\qquad\qquad\qquad\qquad \to Q_y \in \texttt{unreach}_{\subseteq}\ Q\ b))\\
\texttt{and I7:} & [\![Q \in \mathcal{P};\ \ b \in \mathcal{E};\ \ b \notin Q;\ \ Q_x \in Q \setminus (\texttt{unreach}_{\subseteq}\ Q, b\ );\\
& Q_y \in \texttt{unreach}_{\subseteq}\ Q\ b]\!]\\
& \Longrightarrow \exists X\ Q_n.\ [Q_x..Q_y..Q_n]X] \wedge Q_n \in Q \setminus (\texttt{unreach}_{\subseteq}\ Q, b\ )
\end{array}
$$

In Isabelle/HOL, $A \Longrightarrow B$ can be read as the sequent $A \vdash B$. The function `card` returns a natural number representing the cardinality of a set, with 0 representing both an empty and infinite set. To refer to the set of natural numbers between n and m we write $\{n...m\}$. The functions `ch_by_ordering` and $\texttt{unreach}_{\subseteq}$ are both ones we have defined and which we will describe in this section.

4.2 Axioms of Incidence (I1–I4)

The first four incidence axioms (I1–I4) are based on previous axiomatic systems for Euclidean geometry [2,4,12,13], and the last three (I5–I7) distinguish a geometry from a space-time. I5 distinguishes Minkowski from Galilean space-time.

The original presentation of I4 contains I1 ($\mathcal{E} \neq \{\}$) as an assumption. Including I1 in the system of axioms makes that assumption redundant, but then removing the assumption allows us to use I4 (as presented in the locale earlier) not only to obtain a number of paths (which enables us to obtain an event, as all paths are non-empty), but it also describes the existence of an event at which these paths all cross. It is unclear to us at this stage whether Schutz's choice of presentation has something to do with the independence of the axioms.

4.3 Axioms of Order (O1–O5)

The first five axioms of order are quite simple. O1 and O5 describe how betweenness and paths relate to each other, and O2–O4 provide us the three major properties of the betweenness relation in this system[4]: a kind of symmetry, strictness, and transitivity.

We made a change to O4 to remove the Schutz's requirement that $a \neq d$ yielding the following formalisation of the axiom in Isabelle:

$$\texttt{O4:}\ [\![[abc];\ [bcd]]\!] \Longrightarrow [abd]$$

because this leads to an impossible step in Schutz's proof of Theorem 1 where his reasoning goes

> ...leads to a contradiction since, by Axiom O4, $[abc]$ and $[bca]$ imply $[aba]$ which contradicts Axiom O3

This extra condition would preclude his application of Axiom O4. It is likely our change is what was intended in the first place seeing as the axiom is used several times as if this is how it is given.

This change allows a simplification of axiom O3. The original statement of O3 says that if a, b, and c are in a betweenness relation then they are all distinct from one another, but we need only say that a is distinct from c and then the distinctness of all three can be proven as a lemma. This simplified version is a more common way of presenting strictness in the literature. Both Tarski [3] and Veblen [12] state it this way.

We can give a minor simplification of O5 too. The original says that if the events a, b, and c all lie on a path then one of the six possible ways of putting them in the betweenness relation must be true. That is, the disjunction of these six possibilities is true. But by O2 $[xyz]$ is indistinguishable from $[zyx]$, so we do not need to include three of the six possibilities in the disjunction. This simplification makes case splits shorter, or at least allow automated techniques to go through a little quicker. A stronger O5 using exclusive-or can be proven using the other axioms.

4.4 Chains

Chains are a way of talking about back-to-back betweenness. Whereas $[abc]$ is a way of saying "a, then b, then c", chains such as $[abcd...]$ are a way of saying "a, then b, then c, then d, ..." for an arbitrary number of events. Schutz defines chains like so:

> A sequence of events $Q_0, Q_1, Q_2, ...$ (of a path Q) is called a *chain* if 1) it has two distinct events, or 2) it has more than two distinct events and for all $i \geq 2$, $[Q_{i-2}\ Q_{i-1}\ Q_i]$. A *finite chain* is denoted by writing $[Q_0 Q_1 Q_2 \cdots Q_n]$ and an *infinite chain* is denoted by writing $[Q_0 Q_1 Q_2 \cdots]$

[4] Betweenness is a common feature of axiomatised geometry, though its treatment does differ.

The major feature of this definition is that it requires us to be able to talk about both finite and infinite chains. The simplest way to closely match the above definition for chains is using an inductive definition over lists, but as lists in Isabelle cannot be infinite we have had to stray from the text and use a more involved definition.

We capture the ordering and completely characterise a chain by providing a labelling function $f : nat \Rightarrow$ `'a` which assigns each event in the set X a unique natural number which says where it appears in the ordering. X will be forced to be some subset of a path by the use of the betweenness relation in the definition.

We have adapted this idea of a labelling function for capturing the ordering underlying a chain from Scott [9]who uses it in a formalisation of Hilbert's *Grundlagen der Geometrie* [2] in HOL Light. Our slightly modified definition, in Isabelle/HOL:

```
definition ordering :: (nat ⇒ 'a) ⇒ 'a set ⇒ bool
  where
    ordering f X ≡ (∀n. (finite X → n < card X) → f n ∈ X)
                   ∧ (∀x ∈ X. (∃n. (finite X → n < card X)
                                  ∧ f n = x))
                   ∧ (∀n n' n''. (finite X → n'' < card X)
                                  ∧ n < n' ∧ n' < n''
                                  → [(f n) (f n') (f n'')])
```

The first two parts just restrict f to being a surjective function on X, with some logical case splits on finiteness. The part which quantifies over three natural numbers n, n', and n'' is the core of the definition.

Apart from the case split on finiteness it says that if n, n', and n'' are an increasing triple of natural numbers then in order for f to be a valid labelling function it should be the case that the elements returned by f using these numbers are ordered according to the betweenness relation. Various proofs about orderings are omitted here.

We can now provide two notations, one for infinite chains and one for finite, both capturing three specific events on the chain. This comes up often in Schutz. There is an inflexibility here: the moment we want to do the same thing but name four events (for example) in an order on a chain then we would need a whole new notation. As of yet we do not have an elegant way of generalising an elegant notation for this. Our notation for infinite chains capturing three events:

```
definition infinite_chain :: 'a → 'a → 'a → 'a set → bool where
    infinite_chain x y z X ≡ [xyz] ∧ {x, y, z} ⊆ X ∧ ch X
```

which we denote $[..x..y..z..]X]$ for brevity. The dots represent any number of events that are not x, y, or z. This is almost exactly as in Schutz except we have to explicitly capture the name of the set of events in order to require that it is a chain (the function `ch` does this, returning true if the set of events X is a chain; a thin wrapper around `ordering`). The finite case is written in terms of the infinite case. We do this by saying that if the requirements of the infinite case are met and additionally there are no events on X to the left of x, and no events on X to the right of z, then X must be a finite chain with x and z at either end:

```
definition finite_chain :: 'a → 'a → 'a → 'a set → bool where
      finite_chain x y z X ≡ [..x..y..z..]X] ∧ ¬(∃w ∈ X. [wxy] ∨ [yzw])
```

which we denote by $[x..y..z]X]$.

4.5 Axioms of Order (O6)

Axiom $O6$ – the axiom of collinearity – is not too complicated mathematically but its justification is not immediately clear like the others. The ancestor of axioms of this form is the Axiom of Pasch[5], though according to Schutz it differs slightly in order to maintain the system's independence. Its formal statement:

$$\begin{aligned}\texttt{O6:}\ &[\![Q \in \mathcal{P};\ R \in \mathcal{P};\ S \in \mathcal{P};\ T \in \mathcal{P};\ Q \neq R;\ Q \neq S;\ R \neq S;\\ &a \in Q \cap R \wedge b \in Q \cap S \wedge c \in R \cap S;\\ &\exists d \in S.\ [bcd] \wedge (\exists e \in R.\ d \in T \wedge e \in T \wedge [cea])]\!]\\ &\Longrightarrow \exists f \in T \cap Q.\ \exists X.\ [a..f..b]X]\end{aligned}$$

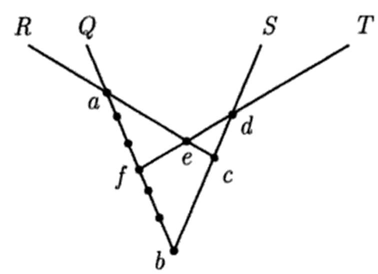


Fig. 1. Axiom O6. Figure originally produced by John Schutz [8].

See Fig. 1 for a diagram illustrating this axiom. One can understand the axioms more kinematically by imagining time moving from left to right over the diagram. Cut a vertical line in a piece of paper and move it over the diagram to simulate this.

4.6 Axioms of Incidence (I5–I7)

So far we have seen little which seems to distinguish Minkowski space-time from an ordinary geometry. These three axioms are three of the primary components of that as they characterise unreachability using the unreachable subset, which Schutz defines like so:

[5] See for example Axiom II,4 of Hilbert [2] and VIII of Veblen [12].

Given a path Q and an event $b \notin Q$, we define the *unreachable subset of Q from b* to be

$$Q(b, \emptyset) := \{x : \text{ there is no path which contains } b \text{ and } x,\ x \in Q\}$$

This says that an event cannot be reached from another if there is no path between them. The unreachable subset of a path Q with respect to some event b which is external to Q is all those events on Q which cannot be reached from b. The definition in Isabelle is very similar, only we rationalise the syntax[6] and move all of the assumptions inside the set comprehension:

$$\texttt{unreach}_{\subseteq}\ Q\ b \equiv \{x \in Q.\ Q \in \mathcal{P} \wedge b \notin Q \wedge b \in \mathcal{E} \wedge \neg(\exists R \in \mathcal{P}.\ b \in R \wedge x \in R)\}$$

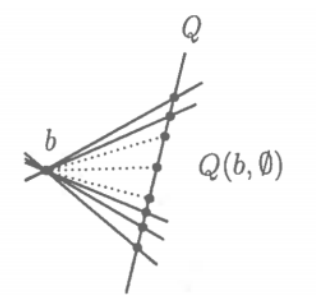


Fig. 2. Image originally from Schutz (page 15) using dotted lines to indicate unreachability.

See Fig. 2 for a diagram of this. Two extra conditions are required to account for the fact that non-events and non-paths could be passed as arguments, and one is brought into the definition ($b \notin Q$) which in the original presentation is assumed to be true whenever the notation is used. Speculation about how to physically interpret notions such as this are omitted here.

According to Schutz it is axiom I5 which excludes Galilean space-time as a possible model. Axiom I5 states that every unreachable set contains at least two events:

$$\texttt{I5:}\ [\![Q \in \mathcal{P};\ b \in \mathcal{E};\ b \notin Q]\!] \Longrightarrow \exists x \in \texttt{unreach}_{\subseteq}\ Q\ b.\ \exists y \in \texttt{unreach}_{\subseteq}\ Q\ b.\ x \neq y$$

Axiom I6 simply states that the unreachable subset of a path is connected: for any two events on it, there are a number of events between them according to the

[6] Schutz seems to use the syntax as if we are indexing into the path Q using b and uses the empty set symbol $\emptyset$ to signify the operation.

betweenness relation which are also on the unreachable set. The Isabelle/HOL version of this axiom is not easy to parse by eye because it requires the explicit use of the labelling function that we use for capturing orderings:

$$
\begin{aligned}
\texttt{I6:}\ & [\![Q \in \mathcal{P};\ b \in \mathcal{E};\ b \notin Q;\ Q_x \in \texttt{unreach}_{\subseteq}\ Q\ b;\ Q_z \in \texttt{unreach}_{\subseteq}\ Q\ b]\!] \\
& \Longrightarrow \exists X\ f.\ \texttt{ch\_by\_ordering}\ f\ X \wedge f\ 0 = Q_x \\
& \qquad\qquad \wedge f\ (\texttt{card}\ X - 1) = Q_z \\
& \qquad\qquad \wedge (\forall i \in \{1...\texttt{card}\ X - 1\}.\ f\ i \in \texttt{unreach}_{\subseteq}\ Q\ b \\
& \qquad\qquad\qquad \wedge (\forall Q_y \in Q.\ [(f\ (i-1))\ Q_y\ (f\ i)] \\
& \qquad\qquad\qquad\qquad \to Q_y \in \texttt{unreach}_{\subseteq}\ Q\ b))
\end{aligned}
$$

which in Schutz's presentation is a less imposing and involved, though refers to two additional names Q_0 and Q_n which are not needed:

> *Given any path Q, any event $b \in Q$, and distinct events $Q_x, Q_z \in Q(b, \emptyset)$, there is a finite chain $[Q_0 \cdots Q_n]$ with $Q_0 = Q_x$ and $Q_n = Q_z$ such that for all $i \in \{1, 2, \ldots, n\}$,*
> i $Q_i \in Q(b, \emptyset)$
> ii $[Q_{i-1}\ Q_y\ Q_i] \Longrightarrow Q_y \in Q(b, \emptyset)$

Axiom I7 states that there exists events on both sides of the unreachable set which bound it within some segment of the whole path. Unlike I6 this does not require referring to every intermediate event in a chain and so is a lot neater to express in Isabelle:

$$
\begin{aligned}
\texttt{I7:}\ & [\![Q \in \mathcal{P};\ b \in \mathcal{E};\ b \notin Q;\ Q_x \in Q \setminus (\texttt{unreach}_{\subseteq}\ Q, b\ );\ Q_y \in \texttt{unreach}_{\subseteq}\ Q\ b]\!] \\
& \Longrightarrow \exists X\ Q_n.\ [Q_x..Q_y..Q_n]X] \wedge Q_n \in Q \setminus (\texttt{unreach}_{\subseteq}\ Q, b\ )
\end{aligned}
$$

where $Q \setminus (\texttt{unreach}_{\subseteq}\ Q, b\ )$ should be read as "the reachable subset of Q from b".

5 Doing Proofs in the System

The theme in the second chapter (the axioms were the first chapter in the book) is the "temporal order on a path". Schutz warns that because there are events which cannot be joined by a path (due to this being a space-time), many of the proofs which have counterparts in ordinary geometries are more difficult here.

At this stage we have proven everything up to and including theorem 9, except lemma 3 and theorem 6, which we will comment on in our concluding remarks. In this section we isolate the statement and proof of theorem 5 to show what it is like to do proof in this system and highlight a case where the original prose was insufficient.

5.1 A Few Remarks

Theorem 2 says that all events on a chain are distinct and that if $0 \leq i < j < l \leq n$ then $[Q_i Q_j Q_l]$ which, though phrased differently, is exactly contained within our definition for chains and as such follows immediately. This is a typical example of how not following the text closely enough can affect the progress of formalisation.

In our proof of theorem 8 we struggled to implement an invocation of "without loss of generality" (WLOG). Simple WLOG statements such as "given three natural numbers a, b, c, assume $a < b < c$ WLOG" are not hard to emulate in Isabelle but Schutz finishes this proof with: "by cyclic interchange of the symbols a, b, c". This is problematic because cycling these symbols forces us to invoke different lemmas and proof tools. At the moment we have just copied and pasted the first part of the proof and cycled manually, changing and adding in the different lemmas and the automatic proof tools as necessary. This needs to be looked into and improved sooner rather than later before we come across more proofs like this in the text.

5.2 Theorem 5

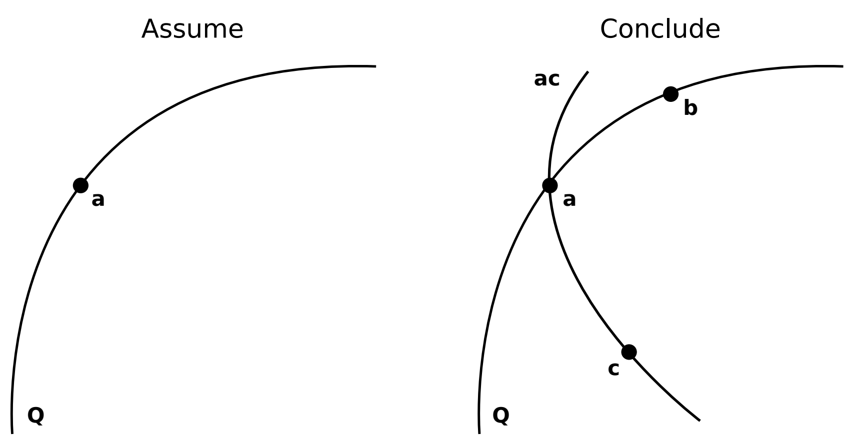


Fig. 3. A diagram representing theorem 5.

Theorem 5 states (see Fig. 3 for a diagram):

Given a path Q and an event $a \in Q$, there is
i *an event $b \in Q$ with b distinct from a, and*
ii *an event $c \notin Q$ and a path ac (distinct from Q)*

As this theorem contains two distinct conclusions connected by a conjunction, for which the proofs can be easily separated (though the second part uses the first), we first prove "$\exists b \in Q.\ b \neq a$" which we call (i) in line with Schutz, followed by "$\exists ac \in \mathcal{P}.\ ac \neq Q \wedge (\exists c \in \mathcal{E}.\ c \notin Q \wedge a \in ac \wedge c \in ac)$" which we will call (ii). The proof requires a few additional lemmas:

– There are no empty paths.
– If there is at least one path, then every event lies on some path.
– If every event is on one path, then there is only one path.

To show the last (and most involved) of these three lemmas one first needs to prove the other two. For example, in order to show that given a path Q and the fact $\forall a \in Q.\ \forall b \in Q.\ a = b$, to conclude $Q = \{a\}$ you need to know that Q cannot be empty.

As an aside, there is another lemma not proved in Schutz which is related to the second item above: the case where there are no paths. It does not seem to be used anywhere in Schutz that we can see. It is stated in Isabelle like so:

```
lemma big_bang:
       assumes no_paths: 𝒫 = {}
       shows ∃a. ℰ = {a}
```

where the name of the lemma is a tongue-in-cheek nod to the form it takes, as it looks as though it describes a point-like universe in which nothing has yet moved.

Once these lemmas have been proven we can complete part (i) (minus the assumption $a \in Q$ which we do not need) in a way that matches Schutz quite closely. To show this we present both the Isabelle and prose proof below. Note that `ge4_paths` is a lemma using axiom I4 which says that there are at least four paths, allowing us to derive a contradiction whenever we can conclude that there are less. The lemma `unreach_ge2_then_ge2` states that if the conclusion to I5 is true, which says there exists two events on the unreachable set, then there are two events on the path which the unreachable set is applied over. It is an obvious fact that the automated tools would have no problem proving on the fly but which we have added in order to make the reasoning more clear.

```
 theorem ge2_events:
       assumes path_Q: Q ∈ 𝒫
       shows ∃b ∈ Q. b ≠ a
proof -
       have d_notinQ: ∃d ∈ 𝒫. d ∉ Q
       proof (rule ccontr)
             assume ¬(∃d ∈ 𝒫. d ∉ Q)
             then have all_inQ: ∀d ∈ 𝒫. d ∈ Q by simp
             then have only_one_path: ∀P ∈ . P = Q
                by (simp add: only_one_path path_Q)
             thus False using ge4_paths by metis
       qed
       then obtain d where d ∈ 𝒫 and d ∉ Q by auto
       thus ?thesis using I5 path_Q unreach_ge2_then_ge2
       by metis
qed
```

Proof. We first show that there is an event $d \notin Q$. Suppose the contrary; namely that each event is on the path Q: then the Axiom of Unique-

*ness of Paths (Axiom I3) implies that there is only one path, namely Q, which contradicts the Axiom of Dimension (Axiom I4). Axiom I5 implies the existence of an event $b \in$ **unreach**$_\subseteq$ Q d with b distinct from a, which establishes (i).*

We do not know whether part (i) is provable using only I3 as Schutz claims here. Part (ii) also requires a number of lemmas but (aside from a lemma which states that if two events a and b are reachable from one another then there must be a path between them) they are not particularly noteworthy and are more just to make proof smoother for this theorem and in the future. After proving these intermediate lemmas the proof is again followed reasonably closely:

```
 theorem ex_crossing_at:
       assumes path_Q: Q ∈ 𝒫
           and a_inQ: a ∈ Q
       shows ∃ac ∈ 𝒫. ac ∉ Q ∧ (∃c ∈ ℰ. c ∉ Q ∧ a ∈ ac ∧ c ∈ ac)
proof -
       obtain b where b_inQ: b ∈ Q
                  and a_neq_b: a ≠ b
          using a_inQ ge2_events path_Q by blast
       have ∃R ∈ 𝒫.R ≠ Q ∧ (∃e. e ∈ R ∧ e ∈ Q)
          by (simp add: ex_crossing_path path_Q)
       then obtain R e where path_R: R ∈ 𝒫
                         and R_neq_Q: R ≠ Q
                         and e_inR: e ∈ R
                         and e_inQ: e ∈ Q by auto
       thus ?thesis
       proof cases
             assume e_eq_a: e = a
             then have ∃c. c ∈ unreach⊆ R b
                using R_neq_Q a_inQ a_neq_b b_inQ e_inR path_Q path_R
                      I5 I2 in_path_event by metis
             thus ?thesis
                using R_neq_Q e_eq_a e_inR path_Q path_R I3
                      ge2_events by metis
       next
             assume e_neq_a: e ≠ a
             then have ∃S ∈ 𝒫. S ≠ Q ∧ a ∈ S ∧ (∃c. c ∈ S ∧ c ∈ R)
                using path_past_unreach R_neq_Q a_inQ e_inQ e_inR
                      path_Q path_R by auto
             thus ?thesis
                by (metis R_neq_Q e_inR e_neq_a I3 path_Q path_R)
       qed
qed
```

Proof. The Axiom of Connectedness (Axiom I2) now implies the existence of a path R (distinct from Q) which meets Q at some event e. If this event is a then Axiom I5 implies the existence of an event c in *`unreach`$_{\subseteq}$ R b and the proof is complete; otherwise Axiom I5 and Theorem 4 imply the existence of a path (distinct from Q) though a which meets R at some event c.*

This differs from Schutz, as do many of our proofs, only in which facts are needed to prove intermediate steps. Often when writing proofs using a more informal pen and paper style one does not explicitly mention some information which is needed when it is obvious (like Q being a path), but quite often it is the case that one does not realise just how much extra information is needed in order to derive the truth of some statement and we are filling in more with intuition than we realise.

Compare, for example, the first step where Schutz simply invokes axiom I2. We instead use another lemma we have proven called `ex_crossing_path`, a 23-line proof which uses another lemma called `external_event` and so on. Our `ex_crossing_path` does use axiom I2 like Schutz, but it does not follow as easily as the prose makes it seem.

The second part of the case split appears to also differ from Schutz, but this is because we have extracted a 17-line lemma `path_past_unreach` so that this theorem can follow the prose more closely and also so that we can reuse it in the future. The unreachable subset is not part of the assumptions or conclusion of this lemma, but it is essential to proving it. If we have two distinct paths Q and R with a on Q and $b \neq a$ on the intersection, then by axiom I5 and theorem 4 (boundedness of the unreachable set), there is an unreachable set from a on one side of b on R, and on the other side of that there is an event which is reachable from a by some path. This gives us the path distinct from Q meeting at some event c, as in the last step of the prose. Once again, the Isabelle formalisation does use I5 an theorem 4 as in the prose, but we need a number of extra lemmas in order to make it work. This is extra work but does clarify exactly how I5 and theorem 4 were used in order to produce this new path, and confirms that this can indeed work.

6 Conclusion

Before beginning this work we believed that in order to provide reasonable evidence that the axiomatisation (minus axioms S, C and I4) and definitions sufficient and usable we would have to prove all of the theorems and lemmas up to and including theorem 10. While we have not been able to get as far as theorem 10, nor yet complete lemma 3 nor theorem 6, we have discovered several gaps and mistakes in the foundation of the system which should help it serve as a more solid basis moving forward. Once we have reached this stage we hope to wrap up this initial effort by completing chapter 2 of the text, formalise axioms S and C, and gain some confidence in axiom I4 and its associated definitions.

Thus far the only time we have had trouble following the prose of the original text is when we needed to add some additional lemmas to aid automation, when some lemmas not present in the text were needed, and when Schutz used the form of his inductive definition for chains in a proof.

Writing up the axioms within Isabelle posed some problems. Aside from the simplifications to the axioms we discussed, axiom O4 had an assumption which prevented it form being used to derive contradictory betweenness facts, axiom I4 relied on notions of sprays and path [in]dependence which we are not yet comfortable presenting due to the fact we have not yet formalised any proofs which rely on it (non-trivially) so that we can be confident in the correctness of our definitions, and the axiom of symmetry is not only unclear in some ways but is also quite complex and will likely be awkward to write up in Isabelle due to its use of types and different treatment of sets.

There is a great deal of future work to be done, building on what we have formalised so far. The priorities are to gain confidence in axiom I4 and to complete the proof of theorem 6, which is simply about the infinitude of paths. Tackling theorem 6 is important not only because it is next chronologically in unproven statements (after lemma 3) but because it is likely to be the first theorem aside from theorem 1 where we stray significantly from Schutz and have to develop some interesting techniques to prove it.

As a concluding remark and on a more general note, we hope that foundational work like the current one will inspire more formalisation in physics.

References

1. Andréka, H., Madarász, J., Németi, I., Székely, G.: On logical analysis of relativity theories. Magyar Filozófiai Szemle **54**(4), 204–222 (2010)
2. Hilbert, D.: Grundlagen der Geometrie: Festschrift zur Feier der Enthüllung des Gauss-Weber-Denkmals in Göttingen. Herausgegeben von dem Fest-Comitee. I. Theil. Teubner (1899)
3. Makarios, T.J.M.: A mechanical verification of the independence of Tarski's Euclidean axiom (2012)
4. Moore, R.L.: Sets of metrical hypotheses for geometry. Transactions of the American Mathematical Society **9**(4), 487–512 (1908)
5. Naber, G.L.: The geometry of Minkowski spacetime: An introduction to the mathematics of the special theory of relativity, vol. 92. Springer Science & Business Media (2012)
6. Schutz, J.W.: Foundations of special relativity: kinematic axioms for Minkowski space-time, vol. 361. Springer (1973)
7. Schutz, J.W.: An axiomatic system for Minkowski space–time. Journal of Mathematical Physics **22**(2), 293–302 (1981)
8. Schutz, J.W.: Independent axioms for Minkowski space-time, vol. 373. CRC Press (1997)
9. Scott, P.: Ordered geometry in Hilbert's Grundlagen der Geometrie. PhD thesis, University of Edinburgh (2015)
10. Stannett, M., Németi, I.: Using Isabelle/HOL to verify first-order relativity theory. Journal of Automated Reasoning **52**(4), 361–378 (2014). https://doi.org/10.1007/s10817-013-9292-7, http://dx.doi.org/10.1007/s10817-013-9292-7
11. Szekeres, G.: Kinematic geometry: an axiomatic system for Minkowski space-time. Journal of the Australian Mathematical Society **8**(2), 134–160 (1968)

12. Veblen, O.: A system of axioms for geometry. Transactions of the American Mathematical Society **5**(3), 343–384 (1904)
13. Veblen, O.: The foundations of geometry. Monographs on topics of modern mathematics, relevant to the elementary field pp. 1–51 (1914)
14. Walker, A.: Axioms for cosmology. In: Studies in Logic and the Foundations of Mathematics, vol. 27, pp. 308–321. Elsevier (1959)
15. Walker, A.G.: Foundations of relativity: Parts I and II. Proceedings of the Royal Society of Edinburgh Section A: Mathematics **62**(3), 319–335 (1948)

n-dimension Aera Method Using Exterior Products

V. Pavan

IUSTI, 5 rue Enrico Fermi, 13453 Marseille cedex 13

Abstract

In this document we present a generalization of the axioms of the Aera Method in any dimension using the interpretation of the exterior (Grassmann) product between vectors.

Keywords: Aera method, exterior product, Grassmann Algebra

1. Introduction

The Aera Method [2] was axiomatized about thirty years ago by Chou Gao and Zang. It is now known as an elegant, efficient, automative (but also human readable) way for proving non trivial theorems in geometry. Interestingly this method was used by Newton in his famous proof of gravitational interaction. To our best knowledge, this method was also proposed in dimension 3 (see the original book [2]) but not necessary in arbitrary dimension n. In this paper we use the theory of exterior algebras and exterior computations to propose an axiomatisation of the method in any dimension. We organize our paper as follows. First we present the exterior calculus and the axiomatisation of the aera method as it was proposed in [3]. Then we prove that any of the proposed axioms can be extended to arbitrary dimension.

2. Reminder on Grassmann Algebras, Exterior Computations and Aera Method

For any further questions on Grassamann algebra and exterior computations, one can refer to [**?**]. Let $\boldsymbol{E}$ be a real linear space. The exterior algebra of degree p on $\boldsymbol{E}$, noted as $\Lambda^p\boldsymbol{E}$, is a real linear space whose elements are called the p-vectors. One can define a so called exterior product between p and q- vectors (also known as "wedge" product) as follows:

$$\forall \mathbf{E} \in \Lambda^p\boldsymbol{E}, \quad \forall \mathbf{F} \in \Lambda^q\boldsymbol{E}, \quad \mathbf{E} \wedge \mathbf{F} \in \Lambda^{p+q}\boldsymbol{E} \tag{1}$$

Such a product reveals to be bi-linear, associative and satisfies the anti-commutative property as follows:

$$\forall \mathbf{E} \in \Lambda^p\boldsymbol{E}, \quad \forall \mathbf{F} \in \Lambda^q\boldsymbol{E}, \quad \mathbf{F} \wedge \mathbf{E} = (-1)^{pq}\, \mathbf{E} \wedge \mathbf{F} \tag{2}$$

By definition there holds $\Lambda^0\boldsymbol{E} = \mathbb{R}$ and the wedge product of a p-vector with a real is defined as the action of the real on the vector from the axioms of linear space. When $\boldsymbol{E}$ is of finite dimension, from any basis $\mathbf{e}_1, \cdots, \mathbf{e}_n$ of $\boldsymbol{E}$ one can construct a basis of $\Lambda^p\boldsymbol{E}$ by considering the families of exterior products:

$$\mathbf{e}_{i_1} \wedge \cdots \wedge \mathbf{e}_{i_p}, \quad 1 \le i_1 < i_2 \cdots < i_p \le n \tag{3}$$

As a consequence, any space $\Lambda^p\boldsymbol{E}$ has dimension C_n^p, with the convention that $C_n^p = 0$ si $p > n$. Then if $\boldsymbol{E}$ has dimension n, the for any $p > n$ the space $\Lambda^p\boldsymbol{E}$ reduces to $\mathbf{0}$. $\Lambda^n\boldsymbol{E}$ has dimension one and its elements are called pseudo-vetorss. A fundamental interpretation of the wedge product is as follows:

$$\mathbf{x}_1, \cdots, \mathbf{x}_p \quad \textit{are independent} \ \Leftrightarrow \mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_p \neq \mathbf{0} \in \Lambda^p\boldsymbol{E} \tag{4}$$

One can define on $\Lambda^p\boldsymbol{E}$ a dot product which can be computed thanks to the determinant and the dot product on $\boldsymbol{E}$ as follows:

$$\langle \mathbf{x}_1 \wedge \cdots \mathbf{x}_p, \mathbf{y}_1 \wedge \cdots \mathbf{y}_p \rangle = \det\left(\mathbf{x}_i \cdot \mathbf{y}_j\right) \tag{5}$$

And the definition of the dot product is extended on $\Lambda^p \boldsymbol{E}$ thanks to bi-linearity. In particular, for any independent family, one can get the Euclidean measure of any geometrical set built on the family as:

$$\mathcal{G} = \left\{ \sum_{i=1}^{i=p} u_i \mathbf{x}_i, \quad (u_1, \cdots, u_p) \in U \right\} \tag{6}$$

By using the formula:

$$\mathcal{V}(\mathcal{G}) = \left(\int_U du_1 \cdots du_p \right) \|\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_p\| \tag{7}$$

So the follwong applciations are straightforwards:

1. volume of the parallelotop built on $\mathbf{x}_1, \cdots, \mathbf{x}_p$ is $\|\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_p\|$
2. volume of the simplex built on $\mathbf{x}_1, \cdots, \mathbf{x}_p$ is $(1/p!) \|\mathbf{x}_1 \wedge \cdots \wedge \mathbf{x}_p\|$

We recall that if $\mathbf{X} \in \Lambda^{p+q} \boldsymbol{E}$ and if $\mathbf{Y} \in \Lambda^p \boldsymbol{E}$, we define the contraction product $\mathbf{X} : \mathbf{Y}$ between $\mathbf{X}, \mathbf{Y}$ by the formula:

$$\forall \mathbf{Z} \in \Lambda^q \boldsymbol{E}, \quad \langle \mathbf{X}, \mathbf{Y} \wedge \mathbf{Z} \rangle := \langle \mathbf{X} : \mathbf{Y}, \mathbf{Z} \rangle \tag{8}$$

and thanks to Riesz representation theorem the element $\mathbf{X} : \mathbf{Y}$ exists uniquely in $\Lambda^q \boldsymbol{E}$. The contraction product is an obvious bi-linear operation. Another important property of the exterior product reads as;

$$\mathbf{E} \in \Lambda^p \boldsymbol{E} = \mathbf{0} \Leftrightarrow \ \exists \mathbf{x} \in \boldsymbol{E} \ \left[\|\mathbf{x}\| \neq \mathbf{0} \Rightarrow \mathbf{E} \wedge \mathbf{x} = 0 \ \textit{and} \ \mathbf{E} : \mathbf{x} = \mathbf{0} \right] \tag{9}$$

Finally let us finish with the so called Hodge conjugation. Choose an orthonormal basis $\mathbf{e}_1, \cdots \mathbf{e}_n$ of $\boldsymbol{E}$ and consider $\mathbf{I} = \mathbf{e}_1 \wedge \cdots \wedge \mathbf{e}_n$ a unitary pseudo-vector. Then the application

$$\forall \mathbf{E} \in \Lambda^p \boldsymbol{E}, \ \mathtt{h}(\mathbf{E}) := \mathbf{I} : \mathbf{E} \in \Lambda^{n-p} \boldsymbol{E} \tag{10}$$

is called the Hodge conjugation. It is a linear involution $\mathtt{h} \circ \mathtt{h} = \mathtt{Id}$. It is straightforwards to see that there holds

$$\forall \mathbf{x} \in \boldsymbol{E}, \ \mathtt{h}(\mathbf{x}) : \mathbf{x} = \mathbf{0} \tag{11}$$

Now let us recall from [3] an (extented slighlty modified) axiomatic version of the Aera Method of Chou, Gao and Zhang's [2] in dimension two. We can define the following real valued functions:

$$(A, B) \to \overline{AB} \in \mathbb{R}, \ (A, B, C) \to \mathcal{S}_{ABC} \in \mathbb{R} \tag{12}$$

From which we derive the following shorthands

$$\mathcal{P}_{ABC} = \overline{AB}^2 + \overline{BC}^2 - \overline{AC}^2 \tag{13}$$

$$AB \| CD \Leftrightarrow \mathcal{S}_{ACD} = \mathcal{S}_{BCD} \tag{14}$$

$$AB \perp CD \Leftrightarrow \mathcal{P}_{ACD} = \mathcal{P}_{BCD} \tag{15}$$

such that the following axioms are true

1. Axiom A_1: There holds $\overline{AB} = 0$ if and only if $A = B$
2. Axiom A_2: $\mathcal{S}_{ABC} = \mathcal{S}_{CAB}$
3. Axiom A_3: $\mathcal{S}_{ABC} = -\mathcal{S}_{BAC}$
4. Axiom A_4: if $S_{ABC} = 0$ then $\overline{AB} + \overline{BC} = \overline{AC}$
5. Axiom A_5: there exists A, B, C such that $S_{ABC} \neq 0$
6. Axiom A_6: there holds $S_{ABC} = S_{DBC} + S_{ADC} + S_{ABD}$
7. Axiom A_7: for any $r \in \mathbb{R}$, there exists P such that $\mathcal{S}_{ABP} = 0$ and $\overline{AP} = r\overline{AB}$
8. Axiom A_8: if $A \neq B$, $\mathcal{S}_{ABP} = 0$, $\overline{AP} = r\overline{AB}$, $\mathcal{S}_{ABP'} = 0$, $\overline{AP'} = r\overline{AB}$ then $P = P'$

9. Axiom A_9: if $PQ \| CD$ and $\overline{PQ} = \overline{CD}$ then $DQ \| PC$
10. Axiom A_{10}: if $\mathcal{S}_{PAC} \neq 0$ and $\mathcal{S}_{ABC} = 0$ then $\overline{AB} S_{PAC} = \overline{AC} \mathcal{S}_{PAB}$
11. Axiom A_{11}: if $C \neq D$ and $AB \perp CD$ and $EF \perp CD$ then $AB \| EF$
12. Axiom A_{12}: if $A \neq B$ and $AB \perp CD$ and $AB \| EF$ then $CD \perp EF$
13. Axiom A_{13}: if $FA \perp BC$ and $\mathcal{S}_{FBC} = 0$ then $4\mathcal{S}^2_{ABC} = \overline{AF}^2 \overline{BC}^2$

3. Defining a Skew Symmetrical Product Between Points and checking Axioms *A*.1, *A*.2, *A*.3 and *A*.5

In this section we use the external product beteween vector to define a skew symmetrical produc between points. We shal note as $\mathcal{E}$ the set of points and as $\boldsymbol{E}$ the linear space of Euclidean vectors obtained by the equipolence relation on $\mathcal{E} \times \mathcal{E}$. Finally we shall note:

$$\Lambda^0 \mathcal{E} = \mathbb{R}, \quad \Lambda^1 \mathcal{E} = \mathcal{E}, \quad \forall p \geq 2, \quad \Lambda^p \mathcal{E} = \Lambda^{p-1} \boldsymbol{E} \tag{16}$$

Now we define a product between points as follows:

Definition 1. *Let $\mathcal{E}$ be an affine space of dimension n and let $\boldsymbol{E}$ be the associated real Euclidean linear space. We define the product between points $M_1, \cdots, M_p$ by the formula:*

$$\forall p \geq 2, \quad M_1 M_2 \cdots M_p = [\mathbf{M}_1\mathbf{M}_2 \wedge \mathbf{M}_2\mathbf{M}_3 \wedge \cdots \mathbf{M}_{p-1}\mathbf{M}_p] \in \Lambda^{p-1} \boldsymbol{E} \tag{17}$$

An important property of this product:

1. It can be defined for any number of points so the the computations of the functions $\bar{\cdot}$, $\mathcal{S}.$ are granted. However they do not return value in $\mathbb{R}$. However the production between p points belongs to $\Lambda^p \boldsymbol{E}$
2. we shall prove that such a product is skew-symmetrical (so Axiom A_1, A_2, A_3) are satisifed. In aprticular axiom A_1 just reads that the representative AB is the null vector is and only if $A = B$

Now let us claim the main proposition:

proposition 1. *The product between points is skew-symmetrical, that is:*

$$\forall \sigma \in S_p, \quad M_{\sigma(1)} \cdots M_{\sigma(p)} = \varepsilon(\sigma) M_1 \cdots M_p \tag{18}$$

where $\varepsilon(\sigma)$ is the so called signature of the permutation σ

Proof. It is enough to prove that for any transposition τ there holds:

$$M_{\tau(1)} \cdots M_{\tau(p)} = -M_1 \cdots M_p \tag{19}$$

Let τ be the transposition which exchanges i, j. We need to compute:

$$\mathbf{M}_1\mathbf{M}_2 \wedge \cdots \wedge \mathbf{M}_{i-1}\mathbf{M}_j \wedge \mathbf{M}_j\mathbf{M}_{i+1} \wedge \cdots \wedge \mathbf{M}_{j-1}\mathbf{M}_i \wedge \mathbf{M}_i\mathbf{M}_{j+1} \wedge \cdots \wedge \mathbf{M}_{p-1}\mathbf{M}_p \tag{20}$$

It is very clear that only the central term:

$$\mathbf{M}_{i-1}\mathbf{M}_j \wedge \mathbf{M}_j\mathbf{M}_{i+1} \wedge \cdots \wedge \mathbf{M}_{j-1}\mathbf{M}_i \wedge \mathbf{M}_i\mathbf{M}_{j+1} \tag{21}$$

is intresting for the computation. We are gpoing to use the Chasles relation as well as the linearity, associativity of the exterior product and finally the skew-symmetry:

$$\begin{array}{cccc}
\textit{number of the line} & \textit{left column} & & \textit{right column} \\
\textit{line } i-1 & \mathbf{M}_{i-1}\mathbf{M}_j & = & \mathbf{M}_{i-1}\mathbf{M}_i + \cdots + \mathbf{M}_{j-1}\mathbf{M}_j \\
\textit{line } i & \mathbf{M}_j\mathbf{M}_{i+1} & = & \mathbf{M}_j\mathbf{M}_{j-1} + \cdots + \mathbf{M}_{i+2}\mathbf{M}_{i+1} \\
\textit{line } i+1 & \mathbf{M}_{i+1}\mathbf{M}_{i+2} & = & \mathbf{M}_{i+1}\mathbf{M}_{i+2} \\
 & \vdots & = & \vdots \\
\textit{line } j-2 & \mathbf{M}_{j-2}\mathbf{M}_{j-1} & = & \mathbf{M}_{j-2}\mathbf{M}_{j-1} \\
\textit{line } j-1 & \mathbf{M}_{j-1}\mathbf{M}_i & = & \mathbf{M}_{j-1}\mathbf{M}_{j-2} + \cdots + \mathbf{M}_{i+1}\mathbf{M}_i \\
\textit{line } j & \mathbf{M}_i\mathbf{M}_{j+1} & = & \mathbf{M}_i\mathbf{M}_{i+1} + \cdots + \mathbf{M}_j\mathbf{M}_{j+1}
\end{array} \tag{22}$$

In this presentation, we need to do the exterior product between the terms of the left column and use the Chasles decomposition in the right column. We are going to show then that in the right column we just need to retai, on any line, one term. Since the exterior product is skew-symmetrical, we can retrieve in the Chasles decomposition any vector which stand in the lines from number $i+1$ to $j-2$. Then we have exactly:

$$\begin{array}{cccc} \textit{number of the line} & \textit{let column} & & \textit{right column} \\ \textit{line } i-1 & \mathbf{M}_{i-1}\mathbf{M}_j & \to & \mathbf{M}_{i-1}\mathbf{M}_i + \mathbf{M}_i\mathbf{M}_{i+1} + \mathbf{M}_{j-1}\mathbf{M}_j \\ \textit{line } i & \mathbf{M}_j\mathbf{M}_{i+1} & \to & \mathbf{M}_j\mathbf{M}_{j-1} \\ \textit{line } i+1 & \mathbf{M}_{i+1}\mathbf{M}_{i+2} & \to & \mathbf{M}_{i+1}\mathbf{M}_{i+2} \\ & \vdots & = & \vdots \\ \textit{line } j-2 & \mathbf{M}_{j-2}\mathbf{M}_{j-1} & \to & \mathbf{M}_{j-2}\mathbf{M}_{j-1} \\ \textit{line } j-1 & \mathbf{M}_{j-1}\mathbf{M}_i & \to & \mathbf{M}_{i+1}\mathbf{M}_i \\ \textit{line } j & \mathbf{M}_i\mathbf{M}_{j+1} & \to & \mathbf{M}_i\mathbf{M}_{i+1} + \mathbf{M}_{j-1}\mathbf{M}_j + \mathbf{M}_j\mathbf{M}_{j+1} \end{array} \tag{23}$$

We complete now the simplification by elimintting the vectors $\pm\mathbf{M}_i\mathbf{M}_{i+1}$ et $\pm\mathbf{M}_{j-1}\mathbf{M}_j$ from the Chasles decompostion, since thies terms appear alone on the lines i et $j-1$. There just remains one exterior product alone:

$$\begin{array}{cccc} \textit{number of line} & \textit{left column} & & \textit{right column} \\ \textit{line } i-1 & \mathbf{M}_{i-1}\mathbf{M}_j & \to & \mathbf{M}_{i-1}\mathbf{M}_i \\ \textit{line } i & \mathbf{M}_j\mathbf{M}_{i+1} & \to & \mathbf{M}_j\mathbf{M}_{j-1} \\ \textit{line } i+1 & \mathbf{M}_{i+1}\mathbf{M}_{i+2} & \to & \mathbf{M}_{i+1}\mathbf{M}_{i+2} \\ & \vdots & = & \vdots \\ \textit{line } j-2 & \mathbf{M}_{j-2}\mathbf{M}_{j-1} & \to & \mathbf{M}_{j-2}\mathbf{M}_{j-1} \\ \textit{line } j-1 & \mathbf{M}_{j-1}\mathbf{M}_i & \to & \mathbf{M}_{i+1}\mathbf{M}_i \\ \textit{line } j & \mathbf{M}_i\mathbf{M}_{j+1} & \to & \mathbf{M}_j\mathbf{M}_{j+1} \end{array} \tag{24}$$

To recover the correct order, we make the following transformations:

$$\mathbf{M}_{i+1}\mathbf{M}_i \to \mathbf{M}_i\mathbf{M}_{i+1}, \quad \mathbf{M}_j\mathbf{M}_{j-1} \to \mathbf{M}_{j-1}\mathbf{M}_j \tag{25}$$

Any of these transformations turns the vector in its opposite, so the whole exterior product is unchanged:

$$\begin{array}{cccc} \textit{number of line} & \textit{left column} & & \textit{right column} \\ \textit{line } i-1 & \mathbf{M}_{i-1}\mathbf{M}_j & \to & \mathbf{M}_{i-1}\mathbf{M}_i \\ \textit{line } i & \mathbf{M}_j\mathbf{M}_{i+1} & \to & \mathbf{M}_{j-1}\mathbf{M}_j \\ \textit{line } i+1 & \mathbf{M}_{i+1}\mathbf{M}_{i+2} & \to & \mathbf{M}_{i+1}\mathbf{M}_{i+2} \\ & \vdots & = & \vdots \\ \textit{line } j-2 & \mathbf{M}_{j-2}\mathbf{M}_{j-1} & \to & \mathbf{M}_{j-2}\mathbf{M}_{j-1} \\ \textit{line } j-1 & \mathbf{M}_{j-1}\mathbf{M}_i & \to & \mathbf{M}_i\mathbf{M}_{i+1} \\ \textit{line } j & \mathbf{M}_i\mathbf{M}_{j+1} & \to & \mathbf{M}_j\mathbf{M}_{j+1} \end{array} \tag{26}$$

Finallay, we operate the transposition which exchanges the line i and the line $j-1$. As the exterior product is skew symmetrical, this amouts to multiply by -1 the final result. So we have as expected:

$$\begin{aligned} &\mathbf{M}_{i-1}\mathbf{M}_j \wedge \mathbf{M}_j\mathbf{M}_{i+1} \wedge \cdots \wedge \mathbf{M}_{j-1}\mathbf{M}_i \wedge \mathbf{M}_i\mathbf{M}_{j+1} = \\ &\qquad - \mathbf{M}_{i-1}\mathbf{M}_i \wedge \mathbf{M}_i\mathbf{M}_{i+1} \wedge \cdots \wedge \mathbf{M}_{j-1}\mathbf{M}_j \wedge \mathbf{M}_j\mathbf{M}_{j+1} \end{aligned} \tag{27}$$

□

Before going further, let us adopt the following notation:

Definition 2. *Let $M_1 \cdots M_p$ be points of $\mathcal{E}$. We define:*

$$M_1 \cdots M_p / [M_{i_1}, \cdots, M_{i_k}] \tag{28}$$

the product between the points $M_1 \cdots M_p$ from which we have retrieved, at the corresponding places, the points $M_{i_1}, \cdots, M_{i_r}$

As for an example we have

$$M_1 \cdots M_5 / [M_1, M_3] = M_2 M_4 M_5 \tag{29}$$

4. Parallelism and Affine Independence Interpreted Thanks to Points Products

Theorem 1. *Let $\alpha_k, k \in [1, r]$ be a family of $\mathbb{R}$ such that $\sum_{k=1}^{k=r} \alpha_k \neq 0$. Then there holds:*

$$\left(\sum_{k=1}^{k=r} \alpha_k\right) \left[M_1 \cdots M_{i-1} \left(\sum_{k=1}^{k=r} \alpha_k P_{ik}\right) M_{i+1} \cdots M_p\right] = \sum_{k=1}^{k=r} \alpha_k M_1 \cdots M_{i-1} P_{ik} M_{i+1} \cdots M_p \tag{30}$$

In particular, if $\sum_{k=1}^{k=r} \alpha_k = 1$ there holds:

$$M_1 \cdots M_{i-1} \left(\sum_{k=1}^{k=r} \alpha_k P_{ik}\right) M_{i+1} \cdots M_p = \sum_{k=1}^{k=r} \alpha_k M_1 \cdots M_{i-1} P_{ik} M_{i+1} \cdots M_p \tag{31}$$

In short, the product between points is multi-affine

Proof. Let P_i be the unique point such that:

$$\forall Q \in \mathcal{E}, \quad \left(\sum_{k=1}^{k=r} \alpha_k\right) \mathbf{QP_i} = \sum_{k=1}^{k=r} \alpha_k \mathbf{QP}_{ik} \tag{32}$$

So there holds:

$$\left(\sum_{k=1}^{k=r} \alpha_k\right) M_1 \cdots M_{i-1} \left(\sum_{k=1}^{k=r} \alpha_k P_{ik}\right) M_{i+1} \cdots M_p = \left(\sum_{k=1}^{k=r} \alpha_k\right) M_1 \cdots M_{i-1} P_i M_{i+1} \cdots M_p \tag{33}$$

We exchange P_i et M_1. Then:

$$\left(\sum_{k=1}^{k=r} \alpha_k\right) M_1 \cdots M_{i-1} P_i M_{i+1} \cdots M_p = -\left(\sum_{k=1}^{k=r} \alpha_k\right) P_i M_2 \cdots M_{i-1} M_1 M_{i+1} \cdots M_p \tag{34}$$

We expand the RHS using the exterior product formula:

$$-\left(\sum_{k=1}^{k=r} \alpha_k\right) P_i M_2 \cdots M_{i-1} M_1 M_{i+1} \cdots M_p = -\left(\sum_{k=1}^{k=r} \alpha_k\right) [\mathbf{P}_i \mathbf{M}_2 \wedge \cdots] \tag{35}$$

We put then the sum of the coefficients in the first vector:

$$-\left(\sum_{k=1}^{k=r} \alpha_k\right) [\mathbf{P}_i \mathbf{M}_2 \wedge \cdots] = -\left(\left(\sum_{k=1}^{k=r} \alpha_k\right) \mathbf{P}_i \mathbf{M}_2\right) \wedge \cdots \tag{36}$$

We use the barycenter property to get:

$$\left(\sum_{k=1}^{k=r} \alpha_k\right) \mathbf{P}_i \mathbf{M}_2 = \left(\sum_{k=1}^{k=r} \alpha_k \mathbf{P}_{ik} \mathbf{M}_2\right) \tag{37}$$

and we put again the summ of the coefficients in this expression in the sequel of exterior products

$$-\left(\sum_{k=1}^{k=r} \alpha_k \mathbf{P}_{ik} \mathbf{M}_2\right) \wedge \cdots \tag{38}$$

We expand the exterior product using linearity

$$-\sum_{k=1}^{k=r} \alpha_k \left[\mathbf{P}_{i_k} \mathbf{M}_2 \wedge \cdots\right] \tag{39}$$

we re-write any of the exterior products of the sum using the definition of product between points:

$$\sum_{k=1}^{k=r} \alpha_k \left[P_{i_k} M_2 \cdots M_{i-1} M_1 M_{i+1} \cdots M_p\right] \tag{40}$$

in any term of the sum, we exchange between the first term P_{ik} and the term i. There holds:

$$\sum_{k=1}^{k=r} \alpha_k \left[M_1 \cdots M_{i-1} P_{ik} M_{i+1} \cdots M_p\right] \tag{41}$$

and we have what we wanted to. □

Now we have the following theorem of independence.

Theorem 2 (Axiom $A.5$)**.** *Let$M_1, \cdots, M_p$ a list of p points in $\mathcal{E}$. Then there are affine independent if and only if $M_1 \cdots M_p \neq \mathbf{0}$.*

Remark 1. *Then the axiom of lower dimension p is that there exists $M_1, \cdots, M_p$ such that $M_1 \cdots M_p \neq \mathbf{0}$ while the axiom of upper dimension p says that for all $M_1, \cdots, M_p$ there holds $M_1 \cdots M_p = \mathbf{0}$*

Proof. The points are affine independent if and only if the vectors $\mathbf{M}_1\mathbf{M}_2, \cdots, \mathbf{M}_{p-1}\mathbf{M}_p$ are free and so if and only if their exterior product is not zero. □

Now we can check parallelism thanks to point product as follows:

Theorem 3. *Let $M_1, \cdots, M_p$ such that $M_1 \cdots M_p \neq 0$. Then the affine space which is generated by the $P_1, \ldots, P_r$ is (weakly) parallel to the affine space which is generated by the $M_1, \cdots, M_p$ if and only if:*

$$\forall (i,j) \in [1,r], \quad P_i M_1 \cdots M_p = P_j M_1 \cdots M_p \tag{42}$$

ONe can even precise that the affine space generated by the $P_1, \cdots, P_r$ is included (so weakly parallel to) in the affine space generated by the $M_1, \cdots, M_p$ if and only if:

$$\forall i \in [1,r], \quad P_i M_1 \cdots M_p = 0 \tag{43}$$

Proof. Straightforwards: the linear space associated to the affine set generated by the points $P_k, k \in [1,r]$ eis the one generated by the vectors $\mathbf{P}_i\mathbf{P}_j, i \neq i$. But there is:

$$P_i M_1 \cdots M_p = P_j M_1 \cdots M_p \Leftrightarrow \mathbf{P}_i\mathbf{P}_j \wedge \mathbf{M}_1\mathbf{M}_2 \wedge \cdots \wedge \mathbf{M}_{p-1}\mathbf{M}_p = \mathbf{0} \tag{44}$$

As the product $M_1 \cdots M_p$ is not zero, the exterior product $\mathbf{M}_1\mathbf{M}_2 \wedge \cdots \wedge \mathbf{M}_{p-1}\mathbf{M}_p$ is not zero either. Then the equality

$$\mathbf{P}_i\mathbf{P}_j \wedge \mathbf{M}_1\mathbf{M}_2 \wedge \cdots \wedge \mathbf{M}_{p-1}\mathbf{M}_p = \mathbf{0} \tag{45}$$

implies that $\mathbf{P}_i\mathbf{P}_j$ belongs to the space generated by the $\mathbf{M}_k\mathbf{M}_{k+1}$: the affine space are parallel. the case where the exterior product is first 0 corresponds to the case where one (all the) point(s) P_i is (are) an affine combination affine of the $M_j, j \in [1,p]$ and so belong to the affine space generated by the $M_i; i \in [1,p]$ □

5. Orthogonality

Definition 3. *Let $M_1, \cdots M_{p+q}$ and $P_1, \cdots P_q$ be points of $\mathcal{E}$. We have by definition:*

$$M_1 \cdots M_{p+q} : P_1 \cdots P_q := \mathbf{M}_1\mathbf{M}_2 \wedge \cdots \wedge \mathbf{M}_{p+q-1}\mathbf{M}_{p+q} : \mathbf{P}_1\mathbf{P}_2 \wedge \cdots \wedge \mathbf{P}_{q-1}\mathbf{P}_q \tag{46}$$

Thanks toi this definition we can have straightforwards characterization of orthogonality:

Theorem 4. *Let $M_1 \cdots M_p$ et $P_1 \cdots P_q$ be two lists of points such that there holds $M_1 \cdots M_p \neq 0$ et $P_1 \cdots P_q \neq 0$ (avec $p \geq 2$ et $q \geq 2$). Then the corresponding affine space are orthogonal if and only if:*

$$\forall i, j \in [1, q]^2\,, \quad M_1 \cdots M_p : P_i P_j = 0 \tag{47}$$

which is equivalent to:

$$\forall i, j \in [1, p]^2\,, \quad P_1 \cdots P_q : M_i M_j = 0 \tag{48}$$

which is also equivalent to:

$$\forall i, j \in\in [1, p]^2\,,\ \forall k, l \in [1, q]^2\,, \quad M_i M_j : P_k P_l = 0 \tag{49}$$

Proof. This just results from the way one can expand the contraction product of a k-blade on a vector [1] □

For the sequle we use the following notation

Definition 4. *Let $M_1 \cdots M_p$ be a list of points. We put by definition:*

$$\left(M_1 \cdots M_p\right)^2 := M_1 \cdots M_p : M_1 \cdots M_p \tag{50}$$

It is very clear, thanks to the properties of the exterior product (see [1]), that $\sqrt{\left(M_1 \cdots M_p\right)^2}$ is the volume of the parallelotope with vertices $M_1 \cdots M_p$

Remark 2. *Then we have the clear interpretation that the point product $M_1 \cdots M_p$ is a generalization of the Aera Method: the product containts both the notion of orientation and the idea of volume.*

One can have a kind of generalized Pythagore theorem (in dimension 3 it is known as De Gua theorem), provided that one can define a right corner in a generalized simplex:

Definition 5. *Let $A_1, \cdots A_p$ be a simplex (or a parallelotope). We say that it has a right corner at A_k when:*

$$\forall i, j \in [1, p]\,, [i \neq j \Rightarrow \mathbf{A}_i \mathbf{A}_k : \mathbf{A}_j \mathbf{A}_k = 0] \tag{51}$$

A straightforwards application is as follows:

proposition 2. *Assume that the simplex $A_1, \cdots, A_p$ has a right corner at A_k, then there holds:*

$$\sum_{i=1, i\neq k}^{i=p} \left(A_1 \cdots A_p / A_i\right)^2 = \left(A_1 \cdots A_p / A_k\right)^2 \tag{52}$$

Proof. This is using the following property (see axiom A.4 to come):

$$\sum_{i=1}^{i=p} (-1)^{i-1} \left(A_1 \cdots A_p / \left[A_i\right]\right) = \mathbf{0} \tag{53}$$

we isolate the term where A_k is missing and we put it in the RHS:

$$\sum_{i=1, i\neq k}^{i=p} (-1)^{i-1} \left(A_1 \cdots A_p / \left[A_i\right]\right) = (-1)^k A_1 \cdots A_p / \left[A_k\right] \tag{54}$$

Now we take the square of the norm on any terms. We arrive at:

$$\left(\sum_{i=1, i\neq k}^{i=p} (-1)^{i-1} \left(A_1 \cdots A_p / \left[A_i\right]\right)\right)^2 = \left(A_1 \cdots A_p / \left[A_k\right]\right)^2 \tag{55}$$

In the LHS, the expansion will make appear the following terms:

$$\sum_{i=1,i\neq k}^{i=p} \left(A_1\cdots A_p/\left[A_i\right]\right)^2 \tag{56}$$

to which we add twice the contraction products :

$$2\left(-1\right)^{i+j}\sum_{i\neq j\neq k}\left(-1\right)^{i+j}\left(A_1\cdots A_p/\left[A_i\right]\right):\left(A_1\cdots A_p/\left[A_j\right]\right) \tag{57}$$

But in any product like:

$$M_1\cdots M_p=\mathbf{M}_1\mathbf{M}_2\wedge\mathbf{M}_2\mathbf{M}_3\wedge\cdots\wedge\mathbf{M}_j\mathbf{M}_{j+1}\wedge\cdots \tag{58}$$

we can see immediately that it can be re-arranged as:

$$M_1\cdots M_p=\mathbf{M}_1\mathbf{M}_2\wedge\mathbf{M}_1\mathbf{M}_3\wedge\cdots\wedge\mathbf{M}_1\mathbf{M}_p \tag{59}$$

So in the exterior products involving the $(A_1\cdots A_p/\left[A_i\right])$ we can re-write:

$$\left(A_1\cdots A_p/\left[A_i\right]\right)=\pm\mathbf{A}_k\mathbf{A}_1\wedge\mathbf{A}_k\mathbf{A}_2\wedge\cdots\mathbf{A}_k\mathbf{A}_p/\left[\mathbf{A}_k\mathbf{A}_i\right] \tag{60}$$

If we compute the contraction product:

$$\left(A_1\cdots A_p/A_i\right):\left(A_1\cdots A_p/\left[A_j\right]\right) \tag{61}$$

it is easy to see that in the product $(A_1\cdots A_p/A_i)$ thanks to by 60 there just misses the $\mathbf{A}_k\mathbf{A}_i$. But for $i\neq j$ this vector appears in $(A_1\cdots A_p/A_j)$ thanks to 60. By consequence, in the determinant corresponding to the matrix for the contraction product 61, there is necessary one column which is zero: the one corresponding to the dot products $\mathbf{A}_k\mathbf{A}_j:\mathbf{A}_k\mathbf{A}_i$ where the products. So we have:

$$\left[i\neq j\Rightarrow\mathbf{A}_k\mathbf{A}_j:\mathbf{A}_k\mathbf{A}_i=0\right]\Rightarrow\left[i\neq j\Rightarrow\left(A_1\cdots A_p/A_i\right):\left(A_1\cdots A_p/A_j\right)=0\right] \tag{62}$$

Finally we conclude that:

$$\left[i\neq j\Rightarrow\mathbf{A}_k\mathbf{A}_j:\mathbf{A}_k\mathbf{A}_i=0\right]\Rightarrow\sum_{i=1,i\neq k}^{i=p}\left(A_1\cdots A_p/A_i\right)^2=\left(A_1\cdots A_p/A_k\right)^2 \tag{63}$$

□

Before going further, let us indicate using the transcription of existence and uniqueness of the orthogonal product in the context of affine space:

Theorem 5. *Let$M_1,\cdots,M_p$ tel que $M_1\cdots M_p\neq 0$. Let $P\in\mathcal{E}$. Then there exists a unique point $H\in\mathcal{E}$ such that*

$$M_1\cdots M_pH=\mathbf{0}\quad et\quad M_1\cdots M_p:HP=\mathbf{0} \tag{64}$$

Proof. This is obvious because of the existence and uniqueness of the orthogonal projection: the first equation means that H belongs to the affine space generated by $M_1,\cdots,M_p$ and the second one means that HP is orthogonal to it. □

6. Proving Axioms $A.4$ and $A.6$

proposition 3 (the Axiom $A.4$)**.** *For any list $M_1\cdots M_p$ we have:*

$$\forall M_1,\cdots M_p,\quad\sum_{i=1}^{i=p}\left(-1\right)^{i-1}M_1\cdots M_p/\left[M_i\right]=\mathbf{0} \tag{65}$$

Proof. It is done by induction starting with $p = 3$. In this case, it is immediate thanks to the Chasles relation between vectors

$$M_2M_3 - M_1M_3 + M_1M_2 = M_2M_3 + M_3M_2 = \mathbf{0} \tag{66}$$

Then assume that the property is true for any $3 \leq k \leq p$. Let us prove it for $p+1$. We compute:

$$\sum_{i=1}^{i=p+1} (-1)^{i-1} M_1 \cdots M_{p+1}/\left[M_i\right] \tag{67}$$

We regroup the first two terms of the sum:

$$\mathbf{M}_2\mathbf{M}_3 \wedge \mathbf{M}_3\mathbf{M}_4 \wedge \cdots - \mathbf{M}_1\mathbf{M}_3 \wedge \mathbf{M}_3\mathbf{M}_4 \wedge \cdots + \cdots \tag{68}$$

we factorize as:

$$(\mathbf{M}_2\mathbf{M}_3 - \mathbf{M}_1\mathbf{M}_3) \wedge \mathbf{M}_3\mathbf{M}_4 \wedge \cdots + \mathbf{M}_1\mathbf{M}_2 \wedge \cdots + \cdots \tag{69}$$

So the sum redas now

$$\mathbf{M}_2\mathbf{M}_1 \wedge \mathbf{M}_3\mathbf{M}_4 \wedge \cdots + \mathbf{M}_1\mathbf{M}_2 \wedge \cdots + \cdots \tag{70}$$

which simplifies as:

$$\mathbf{M}_2\mathbf{M}_1 \wedge \mathbf{M}_3\mathbf{M}_4 \wedge \cdots + \mathbf{M}_1\mathbf{M}_2 \wedge \mathbf{M}_2\mathbf{M}_4 \wedge \cdots + \cdots \tag{71}$$

Now we can factorize by $-\mathbf{M}_1\mathbf{M}_2$ nd get:

$$-\mathbf{M}_2\mathbf{M}_1 \wedge (\mathbf{M}_3\mathbf{M}_4 \wedge \cdots - \mathbf{M}_2\mathbf{M}_4 \wedge \cdots + \cdots) \tag{72}$$

But identifying the term in parenthesis, we can write it as:

$$-\mathbf{M}_2\mathbf{M}_1 \wedge \left(\sum_{i=1}^{i=p} (-1)^{i-1} M_2 \cdots M_{p+1}/\left[M_{i+1}\right]\right) \tag{73}$$

then using the induction assumption, the term between parenthesis is null. □

proposition 4 (The axiom $A.6$)**.** *Let $M_1, \cdots, M_p$ a list of points ($p \geq 2$), with no restriction on p regarded to the dimension of $\boldsymbol{E}$. Then there holds:*

$$\forall M_1, \cdots, M_p,\ \forall P, \quad M_1 \cdots M_p = \sum_{k=1}^{k=p} M_1 \cdots M_{k-1} P M_{k+1} \cdots M_p \tag{74}$$

Proof. To prove this we begin by putting P at the first place of any terms of the RHS:

$$\sum_{k=1}^{k=p} M_1 \cdots M_{k-1} P M_{k+1} \cdots M_p = \sum_{k=1}^{k=p} (-1)^{k-1} \left(P M_1 \cdots M_p/\left[M_k\right]\right) \tag{75}$$

So we compte the RHS as:

$$[\mathbf{PM}_2 \wedge \mathbf{M}_2\mathbf{M}_3 \wedge \cdots] - [\mathbf{PM}_1 \wedge \mathbf{M}_1\mathbf{M}_3 \wedge \cdots] + \cdots \tag{76}$$

Introducing Chasles in the first term:

$$\mathbf{PM}_2 \wedge \mathbf{M}_2\mathbf{M}_3 \wedge \cdots = \mathbf{PM}_1 \wedge \mathbf{M}_2\mathbf{M}_3 \wedge \cdots + \mathbf{M}_1\mathbf{M}_2 \wedge \mathbf{M}_2\ \mathbf{M}_3 \wedge \cdots \tag{77}$$

We keep the second term of the above decomposition which is simply the point product $M_1 \cdots M_p$ and we put the first term with the others othe the sum in the RHS. Then we need to compute:

$$\mathbf{PM}_1 \wedge \mathbf{M}_2\mathbf{M}_3 \wedge \cdots - \mathbf{PM}_1 \wedge \mathbf{M}_1\mathbf{M}_3 \wedge \cdots + \mathbf{PM}_1 \wedge \mathbf{M}_1\mathbf{M}_2 \wedge \cdots + \cdots \tag{78}$$

By factorizing by $\mathbf{PM}_1$, we arrive at:

$$\mathbf{PM}_1 \wedge (M_2M_3 \cdots M_p - M_1M_3 \cdots M_p + \cdots) \tag{79}$$

That is we have

$$\mathbf{PM}_1 \wedge \left(\sum_{k=1}^{k=p} (-1)^{k-1} M_1 \cdots M_p/\left[M_k\right]\right) \tag{80}$$

But thanks to $A.4$ the term between parenthesis is zero. □

7. The Axioms *A*.7 et *A*.8

Theorem 6 (Axiom *A*.7)**.** *Let $M_1 \cdots, M_p$ a list of points in $\mathcal{E}$. Then for any $r \in \mathbb{R}$ there is a unique element $P \in \mathcal{E}$ such that:*

$$M_1 \cdots M_p P = 0 \;\; \textit{and} \; M_1 \cdots M_{p-1} P = r \; M_1 \cdots M_{p-1} M_p \tag{81}$$

Proof. If the product $M_1 \cdots M_p$ is zero, it suffices to choose $P = M_p$. Now assume that $M_1 \cdots M_p$ is not null. Then necessary $M_p \neq M_{p-1}$. In that case, there is a (unique) point P such that:

$$\mathbf{M}_{p-1}\mathbf{P} = r\mathbf{M}_{p-1}\mathbf{M}_p \Leftrightarrow \mathbf{M}_p\mathbf{P} = (r-1)\,\mathbf{M}_{p-1}\mathbf{M}_p \tag{82}$$

It is straightforwards that $M_1 \cdots M_p P = 0$. On the other hand there holds:

$$M_1 \cdots M_{p-1} P = \mathbf{M}_1\mathbf{M}_2 \wedge \cdots \wedge \mathbf{M}_{p-1}\mathbf{P} = \mathbf{M}_1\mathbf{M}_2 \wedge \cdots \wedge (r\mathbf{M}_{p-1}\mathbf{M}_p) = rM_1 \cdots M_p \tag{83}$$

□

Except when $p = 2$, there is never uniqueness of the points P satisfying the above relation. But we can characterize the set of points P that satisfy the above property:

Theorem 7 (Axiom *A*.8)**.** *For $n \geq p \geq 2$, let $M_1 \cdots M_p \neq \mathbf{0}$ be an affine set $\mathcal{V}$ of dimension $p-1$. Then for any $r \in \mathbb{R}$ there exists a unique affine set $\mathcal{W}$ (of dimension $p-2$) such that:*

$$P \in \mathcal{W} \Leftrightarrow M_1 \cdots M_p P = \mathbf{0} \;\; \textit{and} \;\; M_1 \cdots M_{p-1} P = rM_1 \cdots M_p \tag{84}$$

it is the unique affine set which is (weakly) parallel to $\mathcal{V}$ and which contains Q such that $\mathbf{M}_{p-1}\mathbf{Q} = r\mathbf{M}_{p-1}\mathbf{M}_p$

Proof. As a matter of fact, let us write

$$M_1 \cdots M_p P = \mathbf{0} \Leftrightarrow \mathbf{M}_p\mathbf{P} = \sum_{i=1}^{p-2} \alpha_i \mathbf{M}_i\mathbf{M}_{i+1} \tag{85}$$

for a given family of reals: the equivalence is because of the affine interpretation of the product betwenn points. So it is claer that the set of points P is an affine set of dimension $p-2$ which is parallel to $M_1 \cdots M_{p-1}$. On he other hand, it is clear that the point Q which satisfies:

$$\mathbf{M}_{p-1}\mathbf{Q} = r\mathbf{M}_{p-1}\mathbf{M}_p \tag{86}$$

belongs to such an affine set. But for a given point, there exists a unique affine set which is parallel to the given affine set and which contains the given point. □

8. The Axiom *A*.9

The parallelogram theorem remains true enven in dimension n.

Theorem 8. *Let $M_1, \cdots, M_p$ and $P_1, \cdots, P_p$ be two lists of points. Then thre holds the equivalence:*

$$\forall k,l \;\; P_kP_l = M_kM_l \Leftrightarrow \forall k,l \;\; P_kM_k = P_lM_l \tag{87}$$

Proof. Obvious since:

$$P_kP_l = M_kM_l \Leftrightarrow P_kM_k + M_kP_l = M_kP_l + P_lM_l \Leftrightarrow P_kM_k = P_lM_l \tag{88}$$

□

9. The Axiom $A.10$

Theorem 9 (The co-side theorem $A.10$)**.** *Let $M_1 \cdots M_p$, A, B be points which satisfy the following assumptions:*

$$\begin{array}{l} M_1 \cdots M_p \neq 0 \\ M_1 \cdots M_p A = 0 \\ M_1 \cdots M_p B = 0 \\ M_1 \cdots M_{p-1} A \neq 0, \quad M_1 \cdots M_{p-1} B \neq 0 \end{array} \tag{89}$$

Then the following formulas are equivalent:

$$\begin{array}{ll} i & M_1 \cdots M_{p-1} A = \lambda M_1 \cdots M_{p-1} B \\ ii & \exists P \quad [M_1 \cdots M_p P \neq 0 \quad \Rightarrow \quad M_1 \cdots M_{p-1} AP = \lambda M_1 \cdots M_{p-1} BP] \\ iii & \forall P \quad [\ M_1 \cdots M_p P \neq 0 \quad \Rightarrow \quad M_1 \cdots M_{p-1} AP = \lambda M_1 \cdots M_{p-1} BP] \end{array} \tag{90}$$

Proof. Let us start from ii and let us assume that there holds P satisfying $M_1 \cdots M_p P \neq 0$ such that there holds:

$$M_1 \cdots M_{p-1} AP = \lambda M_1 \cdots M_{p-1} BP \tag{91}$$

For any Q which belongs to the affine set $M_1 \cdots M_p$ (and so also to $M_1 \cdots M_{p-1} A$ and/or $M_1 \cdots M_{p-1} B$)

$$\begin{array}{l} M_1 \cdots M_{p-1} AP = (M_1 \cdots M_{p-1} A) \wedge (\mathbf{AQ} + \mathbf{QP}) = (M_1 \cdots M_{p-1} A) \wedge \mathbf{QP} \\ M_1 \cdots M_{p-1} BP = (M_1 \cdots M_{p-1} B) \wedge (\lambda \mathbf{BQ} + \lambda \mathbf{QP}) = \lambda (M_1 \cdots M_{p-1} B) \wedge \mathbf{QP} \end{array} \tag{92}$$

Because of colinearity we have $M_1 \cdots M_{p-1} AQ = \mathbf{0}$ (resp. $M_1 \cdots M_{p-1} BQ = \mathbf{0}$), and consequently:

$$(M_1 \cdots M_{p-1} A - \lambda M_1 \cdots M_{p-1} B) \wedge (\mathbf{QP}) = \mathbf{0} \tag{93}$$

Let us choose for Q the orthogonal projection of P on $M_1 \cdots M_p$, et let us note it by H. There is then:

$$(M_1 \cdots M_{p-1} A - \lambda M_1 \cdots M_{p-1} B) \wedge (\mathbf{HP}) = \mathbf{0} \tag{94}$$

Using the orthogonality of $\mathbf{HP}$ relatively to $M_1, \cdots, M_p$ there holds

$$(M_1 \cdots M_{p-1} A - \lambda M_1 \cdots M_{p-1} B) : (\mathbf{HP}) = \mathbf{0} \tag{95}$$

since $(\mathbf{HP}) \neq \mathbf{0}$ by assumption on P (since P is not in the affine set generated by $M_1 \cdots M_p$ while H is), this implies, (see the reminders on exterior product), that we have necessary:

$$M_1 \cdots M_{p-1} A - \lambda M_1 \cdots M_{p-1} B = 0 \Leftrightarrow M_1 \cdots M_{p-1} A = \lambda M_1 \cdots M_{p-1} B \tag{96}$$

So $ii \Rightarrow i$. Now let us prove that $i \Rightarrow iii$. Let us start from

$$M_1 \cdots M_{p-1} A = \lambda M_1 \cdots M_{p-1} B \tag{97}$$

take the exterior form of the point product, and multiply both terms by $\mathbf{AP}$, so there is:

$$\forall P, \quad M_1 \cdots M_{p-1} AP = \lambda M_1 \cdots M_{p-1} B \wedge (\mathbf{AP}) \tag{98}$$

Now introduce B thanks to Chasles relation in the last term $\mathbf{AP}$ of the RHS. We have:

$$M_1 \cdots M_{p-1} B \wedge (\mathbf{AP}) = M_1 \cdots M_{p-1} B \wedge (\mathbf{AB}) + M_1 \cdots M_{p-1} B \wedge (\mathbf{BP}) \tag{99}$$

But it is clear, since A, B belong to the same affine set, that the first term of the RHS is zero. So there holds:

$$M_1 \cdots M_{p-1} B \wedge (\mathbf{AP}) = M_1 \cdots M_{p-1} B \wedge (\mathbf{BP}) = M_1 \cdots M_{p-1} BP \tag{100}$$

and we have indeed

$$M_1 \cdots M_{p-1} A = \lambda M_1 \cdots M_{p-1} B \Rightarrow \forall P, \quad M_1 \cdots M_{p-1} AP = \lambda M_1 \cdots M_{p-1} BP \tag{101}$$

Since $iii \Rightarrow ii$ is trivial, the theorem is proved. □

10. The Axiom *A*.11

Theorem 10. *Let $M_1 \cdots M_p$, and $P_1 \cdots P_q$ be two set of points such that $p + q = n + 2$ where n is the dimension of the space $\boldsymbol{E}$. We assume that:*

$$M_1 \cdots M_p \neq 0 \quad and \quad P_1 \cdots P_q \neq 0, \quad and \quad \forall i, j, \quad M_1 \cdots M_p : P_i P_j = 0 \tag{102}$$

Let $Q_1 \cdots , Q_r$ another list of points. Then the following equivalence holds:

$$P_1 \cdots P_q Q_i = P_1 \cdots P_q Q_j \Leftrightarrow M_1 \cdots M_p : Q_i Q_j = 0 \tag{103}$$

Proof. First let us observe that

$$P_1 \cdots P_q Q_i = P_1 \cdots P_q Q_j \Leftrightarrow (P_1 \cdots P_q) \wedge (Q_i Qj) = \mathbf{0} \tag{104}$$

We compose by having contraction multiplication with $M_k M_l$ and we get:

$$P_1 \cdots P_q \wedge (Q_i Qj) = \mathbf{0} \Rightarrow [P_1 \cdots P_q \wedge (Q_i Qj)] : M_k M_l = \mathbf{0} \tag{105}$$

we can develop this contraction product thank to the general rules of k-blades on vectors ([1]), and because of the orthogonality we have:

$$(Q_i Q_j : M_k M_l)\, P_1 \cdots P_q = 0 \Rightarrow (Q_i Q_j : M_k M_l) = 0 \tag{106}$$

and the property f the RHS implies that $\Rightarrow M_1 \cdots M_p : Q_i Q_j = 0$. Conversely if we have:

$$(Q_i Q_j : M_k M_l) = 0 \tag{107}$$

We know that we can decompose the $Q_i Q_j$ linear combinations of the $P_l P_{l+1}$ and the $M_k M_{k+1}$: this is a consequence of the assumptions on the affine sets $M_1 \cdots M_p$ and $P_1, \cdots P_q$ which are in orthogonal, direct sum. As the projection on the $M_k M_{k+1}$ is zero (since we have $Q_i Q_j : M_k M_l = 0$), it just remains a linear combinations on the $P_l P_{l+1}$. So we have clearly:

$$P_1 \cdots P_q \wedge (Q_i Q_j) = \mathbf{0} \Leftrightarrow P_1 \cdots P_q Q_i = P_1 \cdots P_q Q_j \tag{108}$$

□

11. The Axiom *A*.12

Theorem 11. *Let $M_1 \cdots M_p$ and $P_1, \cdots P_q$ with $q \leq p$ two affine set which are weakly parallel. Then there holds:*

$$\forall i, j \quad M_1 \cdots M_p : Q_i Q_j = 0 \Rightarrow \forall i, j \quad P_1 \cdots P_q : Q_i Q_j = 0 \tag{109}$$

Proof. It suffices to expand the P_k as functions of the M_l like:

$$\mathbf{P}_k \mathbf{P}_l = \sum_{st} \pi_{kl,st} \mathbf{M}_s \mathbf{M}_t \tag{110}$$

then we use the expansion of the contraction product of a p-blade on a vector. □

12. The Axiom *A*.13

Theorem 12. *Let $M_1 \cdots M_p$ tel que $M_1 \cdots M_p \neq \mathbf{0}$. Consider H the orthogonal projection of M_p on $M_1 \cdots M_{p-1}$, that is the unique point H such that*

$$M_1 \cdots M_{p-1} H = 0 \quad et \quad M_1 \cdots M_{p-1} : H M_p = 0 \tag{111}$$

Then there holds:

$$(M_1 \cdots M_p)^2 = (M_1 \cdots M_{p-1})^2 \, (H M_p)^2 \tag{112}$$

Proof. For any H in the affine set generated by $M_1 \cdots M_{p-1}$ there holds:

$$M_1 \cdots M_p = M_1 \cdots M_{p-1} \wedge (\mathbf{M}_{p-1}\mathbf{M}_p) = M_1 \cdots M_{p-1} \wedge (\mathbf{HM}_p) \tag{113}$$

Now we compute:

$$M_1 \cdots M_{p-1} \wedge (\mathbf{HM}_p) : M_1 \cdots M_{p-1} \wedge (\mathbf{HM}_p) \tag{114}$$

Thanks to the orthogonality of **HP** relatively to $M_1 \cdots M_{p-1}$ the determinant we form to compute this contraction product has its last column and a last line which is zero except its last element which is $(HM_p)^2$. The remaining determinant to compute is finally exactly the contraction product of $M_1 \cdots M_{p-1}$ by himself and we have proved our statement.. □

References

[1] Pavan V., Exterior Algebras, ISTE-Elsevier, London, 2017, ISBN 9781785482373

[2] Chou, S.C., Gao, X.S., Zhang, J.Z.: Machine Proofs in Geometry. World Scientific, Singapore (1994)

[3] Predrag Janicic, Julien Narboux, Pedro Quaresma. The Area Method : a Recapitulation. Journal of Automated Reasoning, Springer Verlag, 2012, 48 (4), pp.489-532.

User Interface and Operations of GeometryTouch on Small and Large Touchscreens

Wei Su[1], Chuan Cai[1] and Jinzhao Wu[2]

[1] School of Information Science Engineering, Lanzhou University, Lanzhou, China 730000
[2] School of Software, Guangxi University for Nationalities, Nanning, China 530006
suwei@lzu.edu.cn

Abstract. GeometryTouch is Web-based geometry education system, which can run on Chrome or Safari browsers of touch devices. The paper introduces the user interface design principle and operation method of GeometryTouch on small and large touchscreens. Firstly, we conduct a brief analysis of the differences between mobile small devices such as 5-inch smartphones and very large touchscreen devices such as 70-inch smart whiteboards in the paper. Secondly, the paper gives some useful strategies for the user interface and operation design of GeometryTouch on both size devices.

Keywords: Multi-touch operation · Dynamic geometry system · Geometry education

1 Introduction

Multi-touch operations, which offer a rich and captivating interaction experience, have recently emerged into the IT device such as smartphone, tabl-et, touch walls and e-blackboards. Dynamic geometry systems (DGS) or interactive geometry software (IGS, also called dynamic geometry environments, DGEs) are computer programs which allow one to create and then manipulate geometric constructions, primarily in plane geometry [1-3]. In Lanzhou University, we are working on a touch-based user interface and interaction techniques for a dynamic geometry system GeometryTouch [4]. Developed by JavaScript, SVG and HTML5, GeometryTouch is Web-based geometry education system, which can run on Chrome or Safari browsers of touch devices. Users, including teachers and students, can use GeometryTouch to create geometric objects and maintain their mathematical relations under touch operations or driving data changes. The main functionalities of GeometryTouch consist:

- Drawing basic geometric shapes such as points, segments, rays, lines, circles, ellipses, polygons etc.,
- Constructing a new geometric object subject to mathematical relations with the constructed objects,
- Measuring length, slope, radius, distance, area, circumference, perimeter, angle, coordinates, parallel, perpendicular, and tangent relations,

- Constructing loci of moving points and envelopes of moving lines,
- Moving and changing objects for illustration and demonstration.

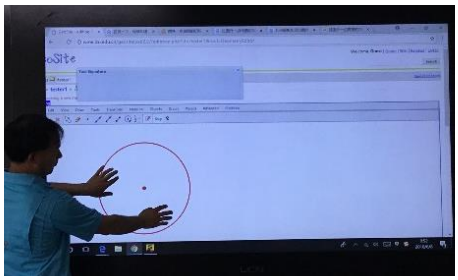
Dozens of dynamic geometry system [3], which can run on PC and be operated by mouse and keyboard, have been designed and developed in past two decades. In recent years, Isotani et al. [5] and X. hong et al. [6] respectively propose and implement an interface and interaction model that is suitable for developing a DGS for mobile touch devices. The paper introduces the design of user interface and operations of GeometryTouch on small and large touchscreens. Whether designing GeometryTouch for a 5-inch or 70-inch touchscreen, many guidelines hold true, including: (1) allowing natural gestures, (2) minimizing the interaction cost of tapping, typing, dragging and drawing geometry objects, (3) offering intuitive user feedback of each geometry manipulative.

2 GeometryTouch on Small Touchscreens

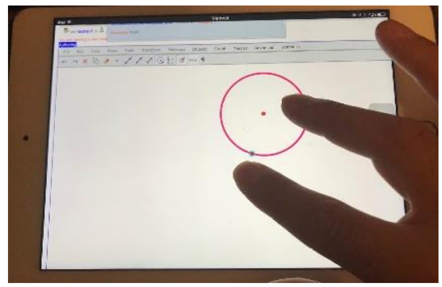
GeometryTouch starts from a traditional mouse operation dynamic geometry system, GeometryEditor [7]. As a Web-based system, GeometryEditor can run well on laptop and desktop computers. Originally, we focus on resizing and optimizing menus, interactions, content, and experiences to make GeometryTouch work well on these small screen devices such as smart phone and tablet. Interacting with small touchscreen devices touch operation can lead to fingers occluding valuable screen real estate. To improve usability of GeometryTouch on small touchscreen, wesdesign a virtual cursor for fitting precise and accurate operations. The distance between virtual cursor andtouch location will change according to device size and touch radius. Via this visual feedback, users guide the pointer into the small target by moving their finger on the screen surface and commit the target acquisition by lifting the finger. An experiment shows that by using the virtual cursor participants can select small targets with lower error rates.

3 GeometryTouch on Large Touchscreens

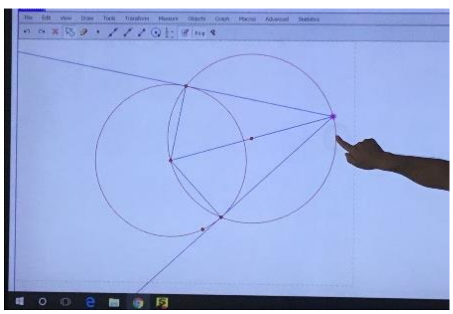
In recent years, many high and middle schools have began to use very large touchscreens (larger than 70-inch) to replace projector in their classrooms. Therefore, in LZU we begin to investigate touch operations of geometry for very large touchscreens. Only a few touch-design skills and design recommendations translate well to designing for very large touchscreens. Users' field of vision, arm motion, affordance, and fatigue are a few of the different considerations for such screens with up to 200 times the area of a smartphone [8]. Fittss Law [9] tells us that the time to acquire a target is a function of the distance to and size of the target. With small touchscreen design, all geometry objects and menus are almost at the same distance from our fingertips, so we mostly focus on target size. However, with very large touchscreens, the distance from the geometry objects becomes more relevant and we need to consider the human physical traits and capabilities including arm reach, arm motion, hand touch with palm or multiple fingertips and height of the teachers and teenage students. The figure 1 shows some comparisons of operations on iPad mini and an e-blackboard. The target size remains essential in ensuring accurate selecting and dragging-drawing geometry objects.



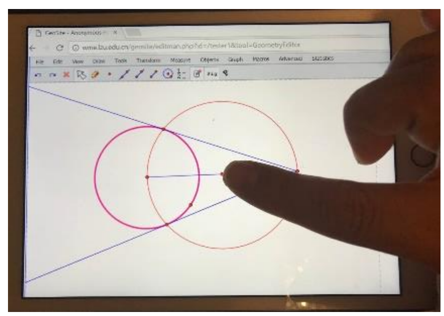
A: Two hands operations on large screen



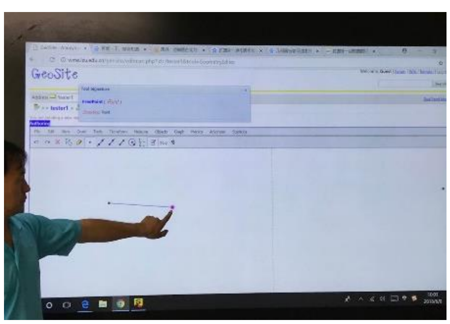
B: Two fingers operations on small screen



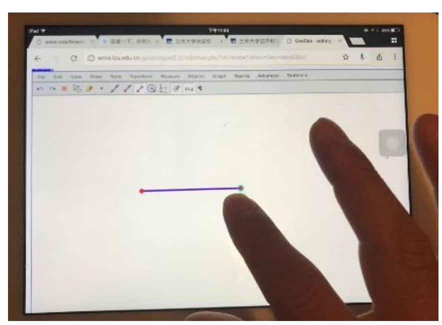
C: One finger operation on large screen



D: One finger operation on small screen



E: Select one point on large screen



F: Select one point on small screen

Fig.1 Comparisons of operations on large and small screens

4 Acknowledgements

We would like to thank Mr. Lai un for making available his GeometryEditor for further development. This work was supported by Science Foundation of Guangxi (AA17204096, AD16380076, AB17129012), Innovation Project for Lanzhou Talent (2014-RC-3) and the Fundamental Research Funds for the Central Universities (lzujbky-2017-194).

References

1. Boyko B. Bantchev. A Brief Tour to Dynamic Geometry Software. CaMSP February 2-015 Newsletter, (2010).
2. Sinclair, Nathalie, et al. Recent research on geometry education: an ICME-13 survey team report. DM 48.5 (2016): 691-719.
3. Goldenberg, E. Paul, Daniel Scher, and Nannette Feurzeig. What lies behind dynamic interactive geometry software. Research on technology and the teaching and learning of mathematics 2 (2008): 53-87.
4. Su, Wei, et al. A touch-operation-based dynamic geometry system: design and implementation. International Congress on Mathematical Software. Springer, Berlin, Heidelberg, 2014.
5. Isotani, Seiji, et al. Interactive geometry goes mobile with GeoTouch. Advanced Learning Technologies (ICALT), 2014 IEEE 14th International Conference on. IEEE, 2014.
6. X. hong, Y. Rao, J. Tang and Y. Huang. A HCI design for developing touch-operation-based DGS: What you think is what you get. The 12th International Conference on Computer Science and Education (ICCSE), Houston, 2017, pp. 51-56.
7. Xun Lai. GeometryEditor: A Web-Based System for Authoring, Sharing and Support of Plane Geometry Manipulative for Mathematics Education. AmericaKent State University. 2010.
8. ara Pernice. Very Large Touchscreens: UX Design Differs from Mobile Screens. https://www.nngroup.com/articles/very-large-touchscreen-ux-design/, accessed on 2018-2-20.
9. Soukoreff, R. William, and I. Scott MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. International journal of human-computer studies 61.6 (2004): 751-789.

A System or Automated Deduction in Engineering Mechanics

Philip Todd

Abstract. A system is presented for automated formula discovery in engineering mechanics built on a Lagrangian formulation where geometric constraints are treated as physical constraints. While the architecture of the system is described, the main focus of this paper is to highlight the interplay between architecture and user interface in generating formulas for mechanical problems. With this in mind , a number of examples are presented in the kinematics, statics and dynamics of simple mechanisms.

Keywords: Mechanics · Automated deduction

1 Introduction

The techniques of automated deduction in geometry have been applied to mechanics in a number of different settings. In [16] Wu extends his method to differential geomery and to the derivation Newton's and Kepler's Laws of planetary motion. This work is extended by Chou and Gao [2] and Wang [15]. In [3], the theorem prover ISABELLE, along with nonstandard analysis is applied to the automated derivation of the theorems in Newton's Principia. In addition to celestial mechanics, Chou and Gao apply their methods to more mundane problems in plane kinematics [1]. In [5], Groebner Bases are applied to solve nonlinear constraint problems emerging from the interplay between geometry and mechanics in the statics of trusses. The question of which beams bear zero loads in a specific problem, and always bear zero loads is addressed using the automated deduction system OTTER in [4].

In this paper, we describe a system which layers Lagrangian mechanics on top of a geometric constraint model to generate formulas in engineering mechanics. The system is designed to be general purpose within its domain of application and to give symbolic solutions for kinematic, static, kinetostatic and dynamic problems involving simple

This material is based upon work supported, in part, by the National Science Foundation under grant IIP-0750028

philt
saltire.com

Present address: Saltire Software, 12700 SW Hall Blvd., Tigard, OR 97223, USA

machines of the sort analyzed in engineering texts. An engineering mechanics model has at its root a geometry model. Our mechanics system is built on top of a constraint based geometry system *Geometry Expressions* and intertwines the constraint description with the mechanics model. In a numerical context, this approach is embedded in the product *Analytix* and described in [10]. The architecture for a symbolic system is discussed in [11]. In this paper, we present a fully realized symbolic mechanics system, discuss user interface and give usage examples. We thus explore its strengths and weaknesses, hopefully elucidating the relationship between the constraint model and the mechanics.

A fundamental characteristic of the architecture is the treatment of constraints as being "load bearing". That is, a reaction force is computed for each constraint. While this is the familiar Lagrangian formulation of mechanics, it necessitates a shift in perspective for engineers used to thinking of objects as bearing loads, rather than constraints. However it admits a very parsimonious model description, allowing the essence of an engineering problem to be expressed in a way which can generate succinct symbolic forms of the solution.

The system, *Mechanical Expressions [14]* uses a constraint based geometry description and layers on top of this kinematic elements. Velocities and accelerations may be assigned to the constraints, and the resulting velocities and accelerations of geometric elements may be measured. The underlying geometry system provides a map G from symbolic constraint values to the symbolic Cartesian coordinates of the points in the model. Differentiation of this mapping provides a kinematic analysis of the model. *Mechanical Expressions* runs on Windows or the Mac.

The Euler Lagrange equations [7] admit a simple expression in the form of the principle of virtual work. This states that at static equilibrium the work done to move the model incrementally is zero. This is an expression of the fact that the model is at an energy minimum. For each constraint, an infinitesimal change in its value will result in an infinitesimal change in the location of all the points and lines in the model. This is embodied in the Jacobian of G. The virtual work is the sum of the applied forces multiplied by the infinitesimal change in location along the axis of the applied force. We define the reaction force in a constraint to be the force which balances out the incremental virtual work done against the applied force elements.

Interpreting the constraint reaction force is part of the skill of model building in the system. For a simple truss (fig. 1) each distance constraint corresponds to the presence of a physical beam, and the reaction force in the constraint corresponds to the internal tension or compression in the beam. A model of a slider crank mechanism (fig. 2) has two distance constraints and an angle constraint. The distances correspond to physical members (the crank and the connecting rod) and would be expected to remain constant. The angle's value would change during the motion of the mechanism. The reaction force in the angle (actually a torque) would correspond to the torque required of the motor.

Inverse dynamics, or kinetostatics require the addition of mass elements and the provision of constraint values with velocities and accelerations (or the ability to describe them as time dependent functions). Given these constraint velocities and accelerations, the system can compute the velocity and acceleration of any point in the model and the angular velocity and acceleration of any line by differentiation. Inertial force elements are added to correspond to accelerating mass elements. Reaction forces may be computed as in the static case (fig. 3). Masses of bodies are represented by a point mass at the center of gravity along with a moment of inertia about the Center of grav-

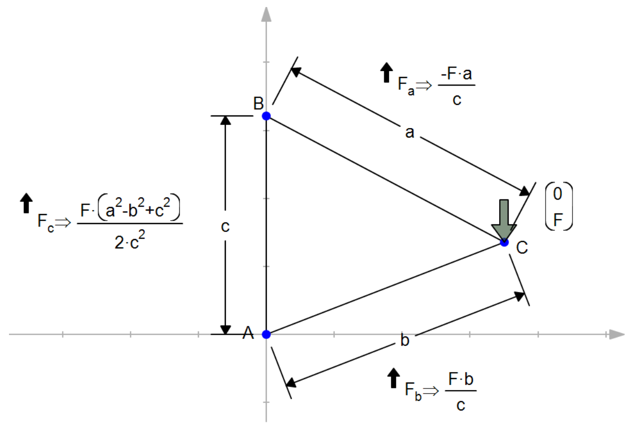


Fig. 1 A triangular truss with force applied at B. Constraint forces are displayed for the three length constraints.

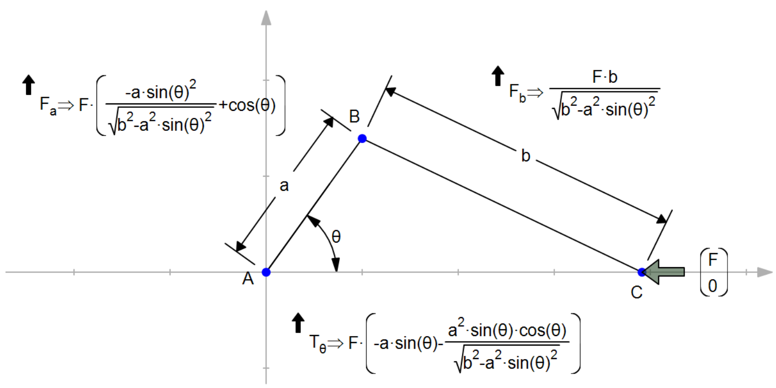


Fig. 2 A crank slider mechanism. External force is applied at C and constraint forces are displayed in the length constraints a and b and the angle constraint θ

ity attributed to any line which moves with the body. Mass elements may or may not experience a gravitational force, depending on whether the figure is specified as lying in a vertical, horizontal, or inclined plane.

In inverse dynamics, forces are derived for a given motion. By contrast a dynamic model derives motion for a given force. The specification of a dynamic model involves identifying one ore more constraints which are free to accelerate in response to unbalanced forces. An acceleration of a constraint will be mapped into accelerations of the masses in the figure, and hence change the reaction forces in the various constraints. The accelerations of the free constraints need to be such that the reaction forces in these constraints are zero. This condition can be expressed as a linear system for the free constraint accelerations.

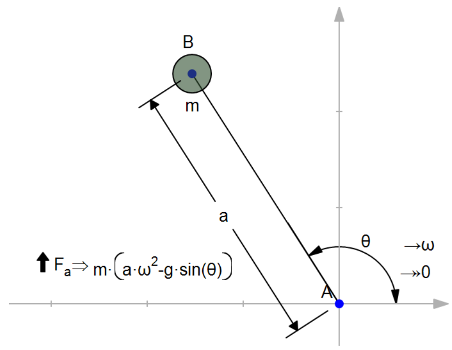


Fig. 3 AB rotates in a vertical plane with angular velocity ω. Reaction force in the length constraint is shown. One term is due to gravity acting on the mass, the other term is a centripetal pseudoforce introduced by the accelerating mass.

2 Mechanics Architecture

Geometry Expressions [12] is a dynamic geometry system with two defining characteristics. While it has a construction based geometry engine, it has a constraint based user interface along with a graph algorithm [9] which maps from a constraint based description to a construction sequence. Constraint arrangements which do not admit a construction sequence are not permitted. A second characteristic is that constructions are performed symbolically rather than numerically. Hence *Geometry Expressions* may be viewed as providing a map from constraint values to coordinate locations.

Let $q_1, \ldots, q_m$ be the (symbolic) constraint values. Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the cartesian coordinate locations of the points of the model and $\theta_1, \ldots, \theta_k$ the angles of the lines.

Let

$$\mathbf{q} = (q_1, \ldots, q_m)$$

and

$$\mathbf{x} = (x_1, y_1, \ldots, x_n, y_n, \theta_1, \ldots, \theta_k)$$

Geometry Expressions provides the function.

$$\mathbf{x}(\mathbf{q})$$

As the function is symbolic, its Jacobian:

$$J_{ij} = \frac{\partial x_i}{\partial q_j}$$

may be readily computed by differentiation. This Jacobian along with the Hessians:

$$H_{ijk} = \frac{\partial^2 x_i}{\partial q_j \partial q_k}$$

are central to the development of a symbolic mechanics capability.

Our symbolic mechanics capabilities are layered, with each layer requiring additional mechanical information on top of the pure geometry. The sequence is Kinematics, Statics, Kinetostatics, Dynamics.

2.1 Kinematics

The core kinematics problem is this: given constraint velocities and accelerations $\dot{\mathbf{q}}, \ddot{\mathbf{q}}$ to find geometry velocities and accelerations $\dot{\mathbf{x}}, \ddot{\mathbf{x}}$.

These can be evaluated directly using the Jacobian and Hessians.

$$\dot{x}_i = \sum_j J_{ij} \dot{q}_j$$

$$\ddot{x}_i = \sum_j \sum_k H_{ijk} \dot{q}_j \dot{q}_k + \sum_j J_{ij} \ddot{q}_j$$

A secondary problem is to compute the velocity and acceleration of any geometric measurement, such as a distance or an angle.

For such a problem, assume the measurement be expressed as a function $\mu(\mathbf{x})$. Then

$$\dot{\mu} = \sum_i \frac{\partial \mu}{\partial x_i} \dot{x}_i$$

$$\ddot{\mu} = \sum_j \sum_i \frac{\partial^2 \mu}{\partial x_i \partial x_j} \dot{x}_i + \sum_i \frac{\partial \mu}{\partial x_i} \ddot{x}_i$$

User Interface

Kinematics is implemented in *Mechanical Expressions* by providing velocity and acceleration input elements which may be attached to any constraint in the system. An output velocity and acceleration element may be attached to any point or line in the drawing. In the case of a point, a velocity and acceleration vector are returned. In the case of a line, an angular velocity and acceleration is returned.

An output kinematic element may also be added to any measurement, and will display the velocity and acceleration of that element. In the case of a point, these will be vector quantities. In the case of a line they will be angular velocity and acceleration. Figure 4 illustrates the solution of a kinematics problem.

2.2 Statics

Given external forces (f_i, g_i) applied to the points in the drawing, and torques T_j applied to the lines , let $\mathbf{f} = (f_1, g_1, \ldots, f_n, g_n, T_1, \ldots T_k)$. We define the generalized constraint force [8]

$$Q_j = -\sum_i f_i J_{ij}$$

By construction, we see that

$$\sum_j Q_j \delta q_j + \sum_i f_i \delta x_i = 0$$

Which can be thought of as an expression of the Principle of Virtual Work [8]

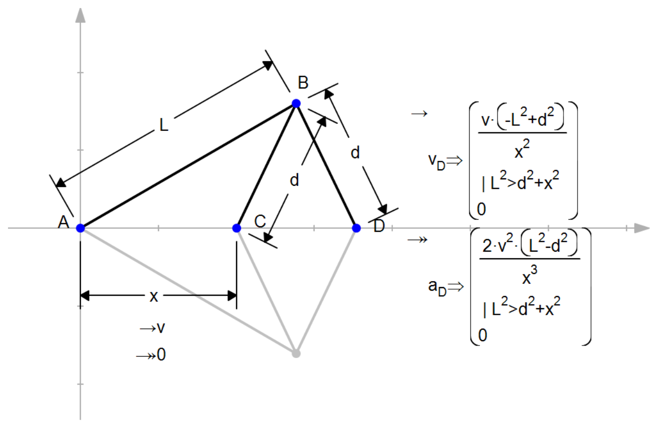


Fig. 4 Input kinematics element specifies velocity and acceleration for the length constraint x. Output kinematic element measures cartesian velocity and acceleration of point D.

In the case of a distance or length constraint, the constraint force represents the internal force in whatever element maintains the constraint. If the element is a rigid body, then this is the tension or compression force exerted to maintain the body's rigidity. If the element is a linear actuator, then this is the force expended by the actuator.

In the case of an angle constraint, the generalized constraint force represents the torque required to maintain that angle.

User Interface

Applied force elements are provided, which may be attached to points in the diagram. Examples of applied forces may be seen in figures 1and 2. In addition, spring-damper-actuator elements may be attached to pairs of points, the effect of which is to apply an equal and opposite force along the line of action of the element to each of the end points. The magnitude of the force is computed from the parameters of the element. For a spring with free length L, spring constant k and end points (x_0, y_0) and (x_1, y_1) The magnitude of the force is:

$$F = k \cdot \sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2} - L$$

For a damper with damping coefficient c, the magnitude of the force is

$$F = c \cdot \frac{(\dot{x}_1 - \dot{x}_0) \cdot (x_1 - x_0) + (\dot{y}_1 - \dot{y}_0) \cdot (y_1 - y_0)}{\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}}$$

For an actuator, the magnitude of the force is given directly by the user. Examples of such elements may be seen in fig. 6 and fig. 9. These force elements have torque analogues for lines and line pairs: applied torque attached to a line, angular spring/damper/actuators between pairs of lines. Such an element is used in fig. 13.

The principal static output is the reaction force element. This can be attached to any constraint in the model and gives the generalized constraint force. For linear constraints, this can be interpreted as an internal force. For angular constraints, it can be interpreted as a torque.

2.3 Kinetostatics

The kinetostatic, or inverse dynamic problem is essentially solving $F = m \cdot a$ for F given a. It is equivalent to the static problem with the addition of inertial forces corresponding to accelerating masses. Given masses m_i applied to the points in the drawing, and moments of inertia I_j applied to the lines, let

$$\mathbf{p} = (m_1\dot{x}_1, m_1\dot{y}_1, \ldots, m_n\dot{x}_n, m_n\dot{y}_n, I_1\dot{\theta}_1, \ldots, I_k\dot{\theta}_k)$$

Inertial force quantities $\dot{\mathbf{p}}$ are added to the derivation of the generalized constraint forces.

$$Q_j = -\sum_i (f_i - \dot{p}_i)\, J_{ij}$$

In addition to the inertial force, the mass element may experience a gravitational force of $m \cdot g \cdot sin\theta$ where θ is the angle of the model's plane to the horizontal and g is acceleration due to gravity.

Let

$$\mathbf{G} = (0, -m_1 g, \ldots, 0, -m_n g, 0, \ldots, 0)$$

then adding gravity, the generalized constraint forces are

$$Q_j = -\sum_i (f_i + G_i - \dot{p}_i)\, J_{ij}$$

User Interface

To enable the creation of kinetostatic models, *Mechanical Expressions* provides two mass elements. The point mass may be attached to a point of the model. The user is also able to specify the angle of the plane of the model to horizontal. If this angle is non-zero, gravitational forces are applied to any point mass elements. Figure 11 shows the use of both a point mass positioned at the center of gravity, and a moment of inertia applied to a line to model the mass properties of a uniform beam.

The output mechanical elements for kinetostatics are, as for statics, the generalized constraint reaction forces.

2.4 Dynamics

If kinetostatics is solving $F = m{\cdot}a$ for F given a, then dynamics is solving $F = m{\cdot}a$ for a given F. Underlying the previous forms of mechanical analysis is the assumption that the value, velocity and acceleration of the constraints are given, and hence the motion is known. If the motion is unknown and the model is fully constrained, there must be one or more constraints whose accelerations are to be derived from a consideration of the applied forces. *Mechanical Expressions* discriminates between two types of constraints: those which are held constant or in prescribed motion, and those which are free to accelerate. Constraints which are held constant typically correspond to rigid bodies, while constraints with prescribed motion typically represent motors or actuators. Constraints which are free to accelerate are geometrical but not structural.

They define the parametrization of the model but not its physics. It is this provision of parametrization which motivates the apparently oxymoronic concept of a constraint which is free to change.

The vector $\dot{p}$, the derivative of the momentum can be expressed in terms of the constraints as

$$\dot{p} = m_i \ddot{x}_i = m_i \left(\sum_j \sum_k H_{ijk} \dot{q}_j \dot{q}_k + \sum_j J_{ij} \ddot{q}_j \right)$$

If we let Q_k^*be the reaction force in free constraint k where the free constraints have 0 acceleration. Let Q_k be the reaction force with free accelerations $\alpha_1, \ldots, \alpha_r$

$$Q_k - Q_k^* = \sum_i \dot{p}_i J_{ik}$$

$$= \sum_i m_i \left(\sum_j J_{ij} \alpha_j \right) J_{ik}$$

$$= \sum_j \left(\sum_i m_i J_{ij} J_{ik} \right) \alpha_j$$

Let

$$M_{kj} = \sum_i m_i J_{ij} J_{ik}$$

Then

$$Q_k - Q_k^* = \sum_j M_{kj} \alpha_j$$

This is a linear equation in $\alpha_1, \ldots, \alpha_r$. At dynamic equilibrium, we require the reaction force in the free constraints to be 0. Hence $Q_k - Q_k^* = -Q_k^*$and we have the following linear system for the free accelerations:

$$\sum_j M_{kj} \alpha_j = -Q_k^*$$

Example

To illustrate how this dynamic solution works in practice, we use the kinetostatic features of *Mechanical Expressions* to derive reaction forces and solve the linear system explicitly. This is handled automatically by the software in the computation of resultant accelerations.

Figure 5 shows a model of a pendulum comprising equal masses at points A and B. Rather than being fixed, point A is free to slide along the x-axis. The model is constrained by the length AB, which will remain constant, by the location x of point A on the x-axis (this is the x coordinate), and by the angle θ between AB and the x-axis.

Constraint x is given velocity v and acceleration a. Angle θ is given velocity ω and acceleration α. The figure shows reaction forces computed by *Mechanical Expressions* for these two constraints:

$$F_x = m \cdot \left(-2a + \alpha \cdot sin\theta + \omega^2 cos\theta \right)$$

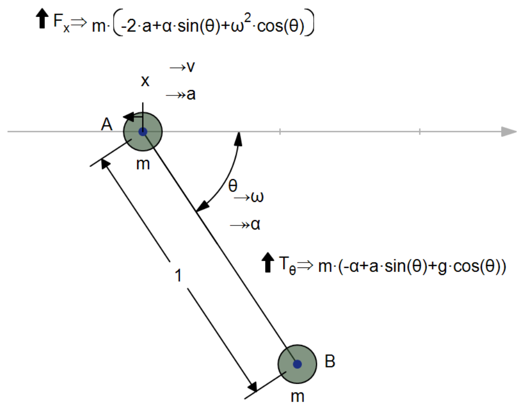


Fig. 5 Mass A is free to slide along the x-axis. AB acts as a pendulum.

$$T_\theta = m \cdot (-\alpha + a \cdot sin\theta + g \cdot cos\theta)$$

At dynamic equilibrium the accelerations a and α are such that these reaction forces are zero. This sets up the following linear system for the accelerations.

$$\begin{pmatrix} -2 & sin\theta \\ sin\theta & -1 \end{pmatrix} \begin{pmatrix} a \\ \alpha \end{pmatrix} = \begin{pmatrix} -\omega^2 cos\theta \\ -g \cdot cos\theta \end{pmatrix}$$

Whose solution is

$$a = \frac{\left(\omega^2 + g \cdot sin\theta\right) \cdot cos\theta}{2 - sin^2\theta}$$

$$\alpha = \frac{\left(2g + \omega^2 sin^2\theta\right) \cdot cos\theta}{2 - sin^2\theta}$$

User Interface

The additional input capability required for dynamic analysis is the ability to specify which of the model constraints are free to accelerate. The primary output element is the resultant acceleration, which may be measured for any of the free constraints. For example, in figure 10 angle θ is specified as free to accelerate, and its instantaneous acceleration displayed.

The free constraint accelerations, expressed in terms of constraint values and their velocities constitute a set of r second order differential equations. These can be expanded to a set of $2r$ first order differential equations by adding equations for the constraint velocities. This is a convenient form for export to mathematics systems for numerical or symbolic solution. For example, the differential equations generated in figure 5 are as follows.

$$
\begin{aligned}
&w5'(\tau) = w6(\tau)\\
&w6'(\tau) = \frac{\left(2 \cdot g + sin(w5(\tau)) \cdot w6(\tau)^2\right) \cdot cos(w5(\tau))}{1 + cos(w5(\tau))^2}\\
&w7'(\tau) = w8(\tau)\\
&w8'(\tau) = \frac{\left(g \cdot sin(w5(\tau)) + w6(\tau)^2\right) \cdot cos(w5(\tau))}{1 + cos(w5(\tau))^2}\\
&w5(0) = \theta\\
&w6(0) = \omega\\
&w7(0) = x\\
&w8(0) = v
\end{aligned}
$$

Note that new functions of time are introduced for the the constraint values and their derivatives. Displayed constraint values and velocities are used as initial conditions. The equations may be copied from *Mechanical Expressions* in a form specific to one of a number of mathematical systems. Once in the mathematics system, numerical solution of the system may readily be accomplished.

2.5 Body Mass

A body's mass properties may be modeled parsimoniously by creating a point at its center of gravity and at least one line which moves with the body. The body's mass is added to the center of gravity point. The body's moment of inertia (about the C of G) is added to any line which rotates with the model. For example, in fig 11 the mass properties of a uniform beam are modelled with a line segment constrained to have length L. A point mass m is placed at its center, while a moment of

$$\frac{mL^2}{12}$$

is added to the line.

3 Geometric Constraints and Mechanical Constraints

A characteristic of the system as described above is that the constraints which define the geometry of the model, and hence the form of all the output expressions must also be mechanical constraints: that is they should be quantities which are kept constant or whose motion is prescribed. This permits succinct statements of mechanical problems, at a cost of limiting flexibility. In addition to specifying the underlying geometry, the geometric constraints also define the coordinate system in which the mechanical model is solved. If a mass is constrained to lie at coordinates (x, y), and that constraint is made free to accelerate, then its equation of motion will be derived in cartesian coordinates. If, on the other hand, the distance to the origin is constrained to be r and the angle which a line joining the point to the origin is constrained to be θ, and both these constraints are free to accelerate, then the equation of motion will be derived in polar coordinates. In fig. 12, the mass is constrained to lie on an arbitrary curve, and the mechanics is solved in terms of parametric coordinates on that curve.

In many problems, it is convenient to parametrize the problem using geometric constraints which do indeed correspond to mechanical constraints. For some problems,

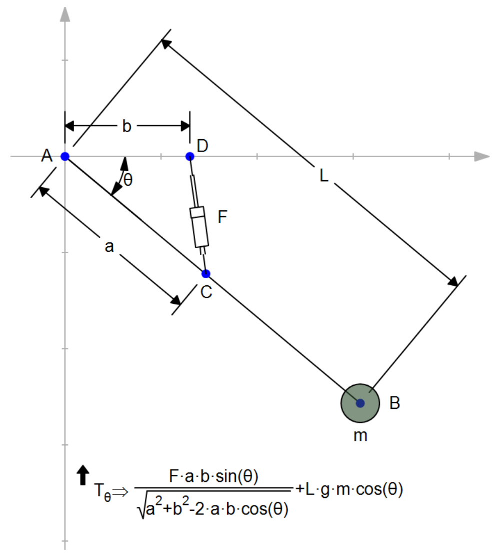


Fig. 6 Actuator CD exerts force F to hold up mass m at D. Angle θ is a geometric constraint and should bear no load. Hence the value of F should be such that $T_\theta = 0$.

however, the user may wish to parametrize the problem in terms of a geometric constraint which is not a mechanical constraint. For example, in fig. 6, the user wants an expression for the force F in the actuator as a function of the angle θ. He could have constrained the length of the actuator rather than the angle, but then his results would have been obtained in terms of this length. In fig. 6 an actuator with force F has been added and the constraint force in angle θ computed. The value of F as a function of θ may be calculated by setting the constraint force (torque) to be 0 and solving the resulting linear equation.

Let a mechanical constraint be a scalar function $\psi(\mathbf{x})$.

In general, let's assume that we have r mechanical constraints $(\psi_1, \ldots, \psi_r)$, and r geometric constraints $(q_1, \ldots, q_r)$, which are not load bearing.

Then let Ψ_i be the constraint force in mechanical constraint ψ_i and Q_j be the constraint force in geometric constraint q_j, then

$$Q_j = -\sum_i f_i J_{ij} - \sum_k \Psi_k \frac{\partial \psi_k}{\partial q_j}$$

This linear system may be solved for $\mathbf{\Psi}$ such that $\mathbf{Q}$ is zero.

From a user interface perspective, there is a need here to create three classes of constraints: those which are geometric only, those which are mechanical only, and those which are both. This distinction is not in the current system, and such problems must be handled manually.

Limitations due to Constructibility Criterion

Fundamental to the symbolic geometry system *Geometry Expressions* is the necessity of converting the constraint model into a sequential construction sequence. The con-

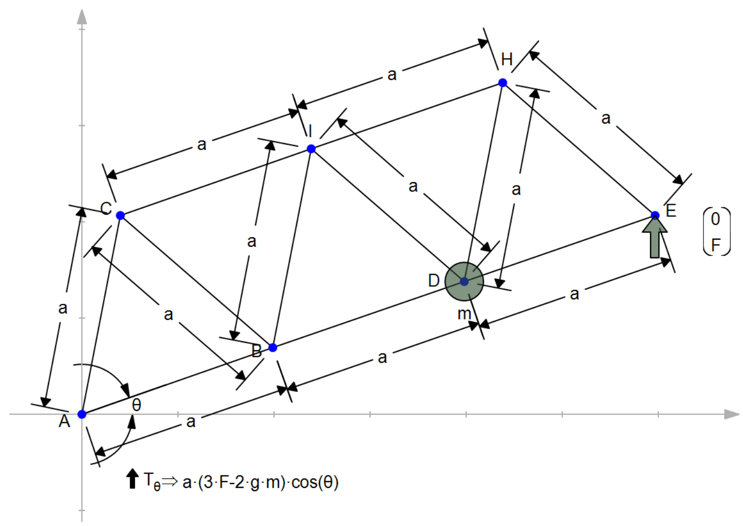


Fig. 7 The angle θ at which the bridge is cantilevered is a geometric constraint and should bear no load.

straints, in a sense are simply a User Interface device on top of a dynamic geometry construction system. The benefit of this approach is that it avoids situations which demand numerical solution. A disadvantage is that there are mechanical problems which cannot be expressed in such a constructive fashion.

For example, fig 7 shows a model of a simple truss bridge. As displayed in the model it is cantilevered at an angle of θ to the horizontal. However, this does not reflect the true mechanics of the situation, we wish the right end of the bridge, point E, to be supported by the axis. However, if we replace the angle constraint with an incidence constraint between E and the axis, the geometry becomes unconstructible from the constraints. A possible approach to this particular problem in the geometry system might be to construct the bridge at an arbitrary angle, find the location of E, then rotate by the appropriate angle such that E lies on the axis. Such an approach, implemented at the geometry level leads to unacceptable intermediate expression swell when mechanical analysis is layered on top of it.

An alternative approach is to use geometric constraints which preserve the model's constructibility (that is, we preserve the angle θ in fig 7). However, we add force elements which recapture the correct mechanical context. In the bridge example, this approach is illustrated by the addition of a vertical force F at point E. F should be set such that the constraint force in angle θ is zero. In the figure, this is done manually. However, the implementation could be automated by the addition to the system of constraints which are only geometrical, and other constraints which are only mechanical.

4 Forms of Output

Measurements of mechanical quantities are by default returned as symbolic expressions whose indeterminates are the indeterminates which are present in the constraint values and in the mechanical inputs, such as velocities and accelerations, applied forces etc. For models of even mild complexity, these symbolic outputs cna be overwhelmingly large. It is important, therefore, to provide a number of alternative forms which allow the system to give usable output as diagram complexity increases.

At the simplest level, the system has numeric values assigned to all the indeterminates, along with a user interface which allows the numeric values to be modified. Any mechanical measurement may be displayed in numeric form, where the indeterminates are replaced by their numeric values and the output expression numerically evaluated.

Intermediate to the numeric and full symbolic expressions, the system will present a result in the form of a Taylor series expanded about the current numerical values of the indeterminates. The user is able to specify the order of the Taylor Series.

For example, in figure 2 if the crank length is specified as 1 (rather than a), and the connector length 2 (rather than b), then the symbolic result for the torque in angle θ is

$$-F \cdot \left(-sin(\theta) - \frac{sin(\theta) \cdot cos(\theta)}{\sqrt{3 + cos(\theta)^2}} \right)$$

If the variable θ has numeric value 0.95, and F has numeric value 1.0, then the numeric output for the torque is

$$1.072375$$

While the 2nd order Taylor Series output is

$$-0.8613836 + 0.6530017 \cdot F + 1.813439 \cdot \theta + 0.4414461 \cdot F \cdot \theta - 0.9544417 \cdot \theta^2$$

Another useful output is in the form of source code, either code snippets, which may be embeded in the user's programs or spreadsheets, or entire JavaScript apps which allow the user to present results as interactive web object, embedding the symbolic solution in an exploratory interface (while hiding the actual symbolics.) Examples of these apps may be seen on the website [13].

5 Examples

We give a number of example problems to illustrate the different mechanical features and some techniques of using the software.

Example 1 Let ABC be a double pendulum supported at A and with equal masses at B and C. Let AB=BC. Let point C be constrained to lie on an inclined plane at angle θ to the horizontal. Find the equilibrium position of the pendulum.

To express this problem in *Mechanical Expressions*, a line is drawn passing through the origin and constrained to be at angle theta with the x axis. Point A is located at the origin, line segments AB and BC are drawn, and C is constrained to lie on the sloping line. AB and BC are constrained to have length a, and masses m are placed at B and C. Finally the angle between AB and the sloping line is constrained to be the angle ϕ (fig. 8).

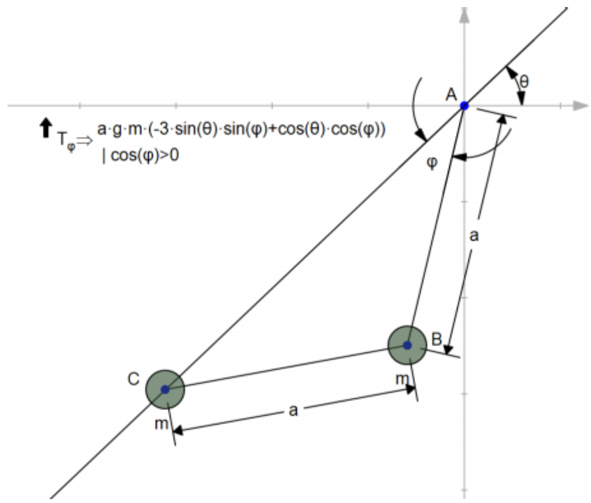


Fig. 8 Torque in the angle between AB and the inclined line is calculated. Equating this to zero finds the equilibrium position.

We can ask *Mechanical Expressions* for the torque in this angle required to maintain static equilibrium with the given geometry, yielding the expression:

$$a \cdot g \cdot m \cdot (-3 \cdot sin(\theta) \cdot sin(\phi) + cos(\theta) \cdot cos(\phi))$$

If there is nothing in place to exert the requisite torque, the system is in static equilibrium only if this reaction torque is zero. Equating the above to 0 yields the following equation for ϕ:

$$tan(\phi) = \frac{1}{3 \cdot tan(\theta)}$$

Example 2 Find a location for a zero-free-length spring to counterbalance an oven door throughout its range of motion

A zero-free-length spring is a convenient idealization of a slack spring (think of a "slinky" toy) with negligible free length. It can also be realized with a finite free-length spring by putting the spring in an enclosure, and hinging the enclosure at the natural length of the spring. Figure 2 shows a *Mechanical Expressions* model of such a door. Point B represents the center of gravity of the door, while point A, at the origin represents the door hinge. A spring with rate k, and free length 0 has one end attached to the door at distance b from the hinge, while the other end is located at coordinate location (u, v). The angle between the door and the x-axs is constrained to be θ.

In the *Mechanical Expressions* model, reaction forces in the distance constraints a, b, in the coordinate constraints (u, v), and a reaction torque in the angle θ balance the force exerted by gravity acting on the mass m and by the spring. We are particularly interested in the torque in angle θ as this does not correspond to any structural element, but reflects some externally applied force used to hold the door in a desired position. For the spring to counterbalance the door exactly, this torque should be zero.

Figure 9 displays this torque:

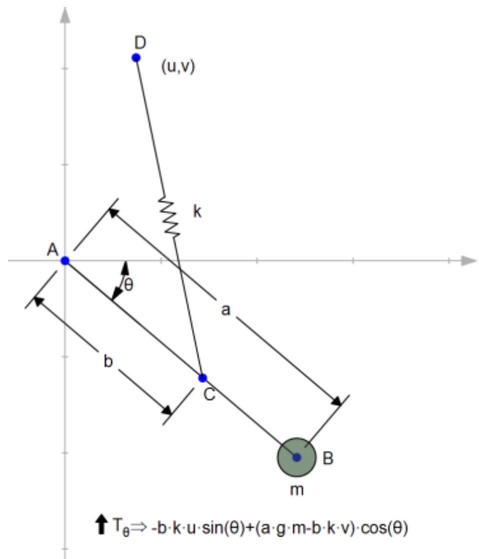


Fig. 9 Reaction torque in the angle θ is displayed. Point B is the center of gravity of the door, whose mass is m.

$$-b \cdot k \cdot u \cdot sin\theta + (a \cdot g \cdot m - b \cdot k \cdot v) \cdot cos\theta$$

This expression is zero for any θ if $u = 0$ and $b \cdot k \cdot v = a \cdot g \cdot m$. For given k and m, the designer has the freedom to choose b or v and the other is determined.

Example 3 A beam of length L has unequal masses M and m at its ends. Ignoring the mass of the beam, where would a fulcrum be located so that the angular acceleration under gravity is maximized.

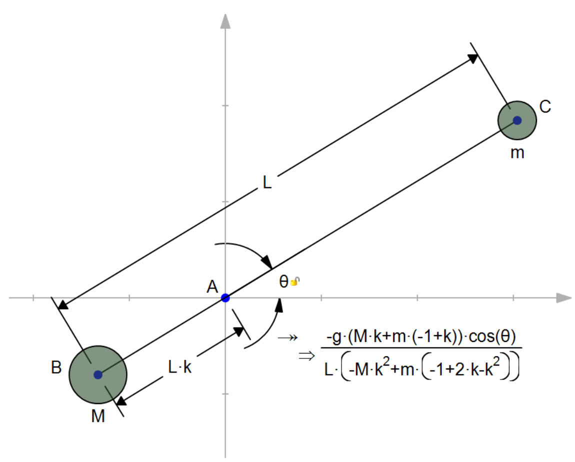


Fig. 10 Acceleration in the angle θ is displayed

In fig. 10, the fulcrum is constrained to lie proportion k along the beam, whose length and angle to the x-axis are prescribed. The acceleration of this angle may be examined.

$$\frac{-g \cdot (M \cdot k + m \cdot (-1 + k)) \cdot cos(\theta)}{L \cdot (-M \cdot k^2 + m \cdot (-1 + 2 \cdot k - k^2))}$$

To find a maximum acceleration, we need to differentiate the expression and solve. This can be done by copying from *Mechanical Expressions* into an algebra system and manipulating the result there. In this case we obtain the following solution for k.

$$k = \frac{m - \sqrt{mM}}{M + m}$$

Replacing k in *Mechanical Expressions* by this value yields the following for the acceleration:

$$\frac{-g \cdot \sqrt{M \cdot m} \cdot (M + m) \cdot cos(\theta)}{2 \cdot L \cdot M \cdot m}$$

We note that this acceleration is proportional to the ratio of the arithmetic mean and the geometric mean of m and M.

Example 4 An exhibit in a science classroom consists of a yardstick hinged at one end. At the other end are two small cups, and a ball sits in the outer cup. A wedge is placed under the yardstick so that it sits at a particular angle. when the wedge is removed, the yardstick falls, and the ball is found to have moved from the outer cup to the inner cup. Why? What restrictions are there on the angle of the wedge?

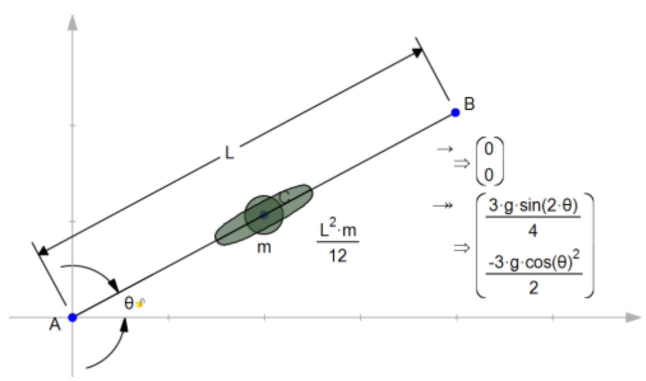


Fig. 11 Uniform beam AB is hinged at B. We measure the acceleration of point B as the beam falls under gravity.

The ruler can be modeled by a line, with one end fixed at the origin. The length of the line is constrained to be L Its center of mass is at the midpoint. A mass m is placed there. The moment of inertia about the center of mass of a uniform beam is

$$\frac{mL^2}{12}$$

This inertia is given to the line. The angle between the line and the x-axis constrained to be θ.

This angle is specified as being free to accelerate. The vertical component of the acceleration of point B is shown to be:

$$\frac{-3 \cdot g \cdot cos(\theta)^2}{2}$$

Which is greater than g so long as

$$cos(\theta) > \sqrt{\frac{2}{3}}$$

or θ less than about 35 degrees.

As the ball falls with acceleration g, under these circumstances the end of the ruler accelerates faster than the ball, allowing it to escape its cup and drop into the other, strategically positioned cup.

Example 5 A wire has the shape of the curve $(x(t), y(t))$. A bead slides on this wire under the influence of gravity. If the bead is at parametric location t, and has parametric velocity v, what is its parametric acceleration under gravity? Assume gravity operates in the negative y-direction.

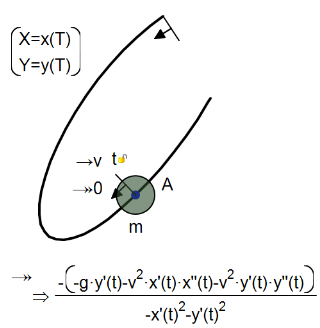


Fig. 12 A mass is located at A which is constrained to be at parametric location t on the curve $(x(t), g(t))$. It is given parametric velocity v. The resulting parametric acceleration is displayed.

Figure 12 shows a model for this problem. Point A is constrained by a parametric location constraint, whose value is specified to be t and whose velocity is specified to be v. The constraint is set to be free to accelerate, and its resultant acceleration is output.

$$a = \frac{-\left(-g \cdot y'(t) - v^2 \cdot x'(t) \cdot x''(t) - v^2 \cdot y'(t) \cdot y''(t)\right)}{-x'(t)^2 - y'(t)^2}$$

Note that v is the parametric velocity of the bead, while x',y', x'', y'' are functions describing the slope and curvature of the curve.

Example 6 Using Force to Solve Geometry Problems

In *the Mathematical Mechanic* [6], Mark Levi inverts the normal order of things, using physics to solve mathematical problems rather than the other way round. His method is to dream up a physical thought experiment whose solution coincides with the solution of the posed mathematical problem. Physical considerations can lead to slick proofs both of the thought experiment and of the corresponding mathematical problem.

In this section we illustrate how Levi's methods can be combined with *Mechanical Expressions* to solve a geometry problem.

The Drive in Movie Problem [6] (otherwise known as the rugby kick problem or the art gallery problem) is this. How far back should you park to get the best view of the movie screen at a drive-in movie?

Mathematically, we want to maximize the angle subtended by the movie screen. Levi's model for this is to take the two sight lines (AC and BC in figure 13) and add a constant torque angular actuator between them. The potential energy in such an actuator is proportional to torque times angle. As the torque is constant an angle stationary point will correspond to an energy stationary point. The physical problem of finding a static equilibrium for the model aligns with the mathematical problem of finding a maximal angle.

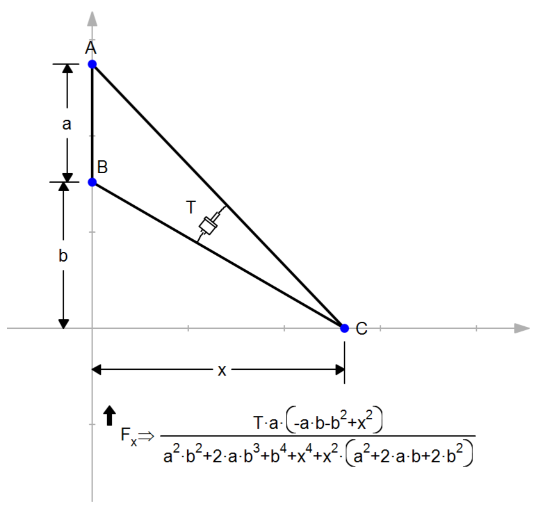


Fig. 13 A model for the drive-in movie problem

In Figure 13, the force in the distance constraint x is displayed.

$$\frac{T \cdot a \cdot \left(-a \cdot b - b^2 + x^2\right)}{a^2 \cdot b^2 + 2 \cdot a \cdot b^3 + b^4 + x^4 + x^2 \cdot (a^2 + 2 \cdot a \cdot b + 2 \cdot b^2)}$$

At equilibrium this force must be zero, leading, by inspection, to the solution:

$$x = \sqrt{b^2 + a \cdot b}$$

We admit that deploying the heavy artillery of an automated mechanics system is somewhat at odds with the spirit of Levi's approach, which is to use physics to find clever and succinct solutions. We claim this solution to be, at least, short.

6 Conclusion

Identifying the constraints of a constraint based geometry system with the load bearing constraints in a Lagrangian formulation of mechanics provides a system capable of succinct expressions of core engineering problems, such that symbolic results may be derived which are simple enough to be useful. Where sufficient simplicity cannot be achieved, symbolic results may be provided in the form of computer code for further analysis in numerical analysis environments and embedding in a program.

The system described in this paper is built on a geometry system which resolves a constraint model into a construction sequence typical of a dynamic geometry system. This places a restriction on the geometries which may be modeled in the system, but has the benefit of facilitating symbolic solution. The identification of geometric constraint with physical constraint implies that a reconfiguration of the geometric constraints to ensure constructibility may misalign them with intended structural elements of the model. For given geometry, applied forces and constraint reactions satisfy a linear system. Force elements of indeterminate strength may be added to correspond to physical constraints which are not geometrical. A linear system may be set up to find the appropriate values for these elements such that the geometric constraints which are not physical bear no load.

References

1. Chou, S. C., & Gao, X. S. (1989). Automated reasoning in mechanics using Ritt-Wu's method. University of Texas at Austin, Department of Computer Sciences. Technical Report TR-89-11.
2. Chou, S. C., & Gao, X. S. (1993). Automated Reasoning in Differential Geometry and Mechanics Using the Characteristic Set Method. Part II. Mechanical Theorem Proving. J. Autom. Reasoning, 10(2), 173-189.
3. Fleuriot, J. D., & Paulson, L. C. (1998, July). A combination of nonstandard analysis and geometry theorem proving, with application to Newton's Principia. In International Conference on Automated Deduction (pp. 3-16). Springer, Berlin, Heidelberg.
4. Ioakimidis, N. (2001). Elementary theorem-proving applications of the OTTER automated deduction system in structural mechanics. University of Patras, Greece, Department of Engineering Sciences, Division of Applied Mathematics and Mechanics. Technical Report TR-2001-07.
5. Lakmazaheri, S. (1997). Constraint-based reasoning via Grobner Bases. AI EDAM, 11(1), 5-15.
6. Levi, M. (2009). The mathematical mechanic: using physical reasoning to solve problems. Princeton University Press.
7. Levi, M. (2014). Classical mechanics with calculus of variations and optimal control: an intuitive introduction (Vol. 69). American Mathematical Soc..
8. Synge, J. L., & Griffith, B. A. (1949). Principles of mechanics. McGraw Hill
9. Todd, P. (1989). A k-tree generalization that characterizes consistency of dimensioned engineering drawings. SIAM journal on discrete mathematics, 2(2), 255-261.
10. Todd, P. (1992, June). A Constructive Variational Geometry Based Mechanism Design Software Package. In ASME DE (Vol. 46, pp. 267-273).
11. Todd, P. H., McLeod, R. J., & Harris, M. (1994, August). A system for the symbolic analysis of problems in engineering mechanics. In Proceedings of the international symposium on Symbolic and algebraic computation (pp. 84-89). ACM.
12. Todd, P. (2006, August). Geometry Expressions: a constraint based interactive symbolic geometry system. In International Workshop on Automated Deduction in Geometry (pp. 189-202). Springer, Berlin, Heidelberg.
13. Todd, P.H. (2015) "Euclid's Muse." Euclids Muse Site Wide Activity RSS, Saltire Software, www.euclidsmuse.com/.

14. Todd, P. H. (2015). Mechanical Expressions :: A New Generation of Mechanical Engineering Software. Retrieved July 29, 2018, from http://www.mechanicalexpressions.com/
15. Wang, D. (1996). A method for proving theorems in differential geometry and mechanics. In J. UCS The Journal of Universal Computer Science (pp. 658-673). Springer, Berlin, Heidelberg.
16. Wen-Tsun, W. (1991). Mechanical theorem proving of differential geometries and some of its applications in mechanics. Journal of Automated Reasoning, 7(2), 171-191.

On n-sectors of the Angles of an Arbitrary Triangle

Dongming Wang[1,2,3], Bo Huang[1], and Xiaoyu Chen[1] ✉

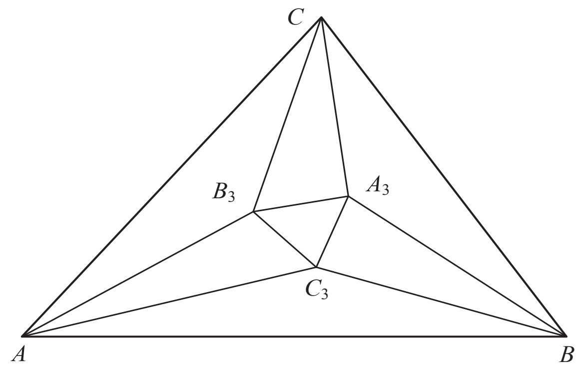
[1] LMIB – School of Mathematics and Systems Science
Beihang University, Beijing 100191, China
bohuang0407@buaa.edu.cn, chenxiaoyu@buaa.edu.cn
[2] SMS/HCIC – College of Software and Information Security
Guangxi University for Nationalities, Guangxi 530006, China
[3] Centre National de la Recherche Scientifique
75794 Paris Cedex 16, France
Dongming.Wang@lip6.fr

Abstract. Morley's theorem shows that the three points, each of which is the intersection of the two internal trisectors that are the closest to the same side of an arbitrary triangle Δ, form an equilateral triangle. This beautiful theorem was proved mechanically by Wu in 1984 in the most general form: the neighbouring trisectors of the three angles of Δ intersect to form 27 triangles in all, of which 18 are equilateral triangles, called Morley triangles. A natural question is: does there exist any equilateral triangle, other than Morley triangles, which is formed by some intersection points of the neighbouring angular n-sectors of Δ for $n \geq 3$? In this paper, we approach this question using specialized techniques with interactive, semi-automatic algebraic computations and prove that for $n = 4$ and 5 the three points, each of which is the intersection of the two internal (or two external) angular n-sectors closest to the same side of Δ, form an equilateral triangle if and only if Δ is equilateral. The computational approach we present can also be applied to other cases for specific n. How to establish the non-existence of equilateral triangles formed by the intersection points of angular n-sectors for general n remains to be an interesting question.

Keywords: Algebraic computation · Angular n-sectors · Equilateral triangle · Morley theorem · Theorem proving

1 Introduction

For an arbitrary triangle ABC let the two internal trisectors closest to the side AB intersect at point C_3, so do the two closest to the side BC at point A_3 and the two closest to the side CA at point B_3. In 1899 Frank Morley discovered that the triangle $A_3B_3C_3$ is always equilateral, no matter what the triangle ABC looks like [8, 9]. Later in 1984, Wu [15] provided a machine proof of this beautiful theorem, stated in the most general form as follows (see also [4, 13]).



Morley theorem. The neighbouring trisectors of the three angles of an arbitrary triangle intersect to form 27 triangles in all, of which 18 are equilateral.

Machine proofs of Morley's theorem using algebraic methods are often taken as examples to illustrate the surprising success of mechanized geometric theorem proving. Following Wu's pioneering work [14] in the early 1980s, extensive research has been carried out in the last three decades with many remarkable progresses (see [1, 5, 11, 12, 16, 10, 7, 17] and references therein).

There are many other proofs for Morley's theorem and there have been many attempts to generalize this theorem in different ways (see [2] for example). In this paper, we study the relationships among intersection points of the angular n-sectors of an arbitrary triangle for $n > 3$. More concretely, we are interested in the following problem.

Problem (cf. Fig. 1). Let $n \geq 3$ and $n > m > 0$ with $m/n \notin \{1/2, 1/3, 2/3\}$, and let A_n, B_n, C_n (or respectively $\bar{A}_n$, $\bar{B}_n$, $\bar{C}_n$) be the three points all inside (or all outside) the triangle ABC such that $\angle BAC_n = \angle CAB_n = \angle BAC \cdot m/n$, $\angle ABC_n = \angle CBA_n = \angle ABC \cdot m/n$ and $\angle BCA_n = \angle ACB_n = \angle BCA \cdot m/n$ (or $\angle BA\bar{C}_n = \angle CA\bar{B}_n = (\pi - \angle BAC) \cdot m/n$, $\angle AB\bar{C}_n = \angle CB\bar{A}_n = (\pi - \angle ABC) \cdot m/n$ and $\angle BC\bar{A}_n = \angle AC\bar{B}_n = (\pi - \angle BCA) \cdot m/n$). Prove or disprove that $\Delta A_nB_nC_n$ (or $\Delta \bar{A}_n\bar{B}_n\bar{C}_n$) is equilateral if and only if ΔABC is equilateral.

For the sake of convenience, we call any equilateral triangle formed by the three distinct intersection points A_n, B_n, C_n (or $\bar{A}_n$, $\bar{B}_n$, $\bar{C}_n$) for $n \geq 3$ a *Morley triangle*. Morley's theorem points out that there are 18 Morley triangles for $n = 3$. Those triangles formed by A_n, B_n, C_n (or $\bar{A}_n$, $\bar{B}_n$, $\bar{C}_n$) with $m/n \in \{1/3, 2/3\}$ are among the 18 Morley triangles. What we are interested in is the existence or non-existence of other Morley triangles. Similar problems have also been studied in [3, 18], and a strong converse of Morley's theorem was proved in [18], showing that for all ΔABC, $\Delta A_nB_nC_n$ is always equilateral if and only if $m/n \in \{1/3, 2/3\}$. However, this result does not give the necessary condition on ΔABC for $\Delta A_nB_nC_n$ to be equilateral. In this paper, we show

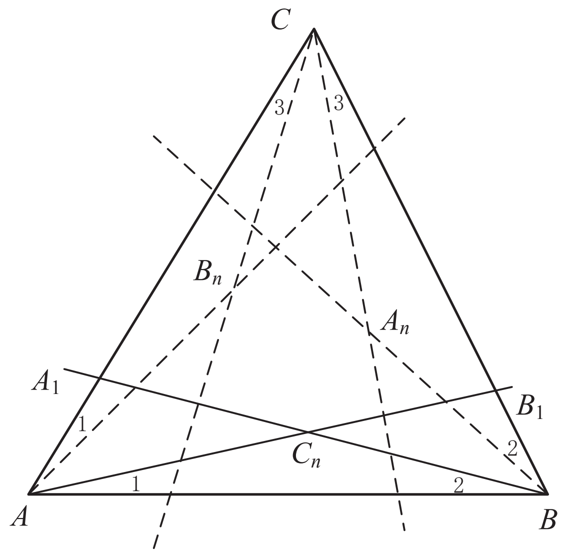


Fig. 1. n-sectors of the internal angles of ΔABC with $\angle 1 = \angle BAC \cdot m/n$, $\angle 2 = \angle ABC \cdot m/n$ and $\angle 3 = \angle BCA \cdot m/n$.

that ΔABC must be equilateral if $\Delta A_n B_n C_n$ (or $\Delta \bar{A}_n \bar{B}_n \bar{C}_n$) is equilateral for $m/n \notin \{1/2, 1/3, 2/3\}$.

The investigations in this paper, restricted to $m = 1$, are mainly based on a symbolic-computational approach with algebraic techniques used to deal with complicated polynomial and trigonometric expressions. The main results we have obtained are listed below.

1. The three lines AA_n, BB_n, CC_n are concurrent, and so are the three lines $A\bar{A}_n$, $B\bar{B}_n$, $C\bar{C}_n$.
2. $|BA^*| : |A^*C| = \frac{\angle BCA}{\angle ABC}$, $|CB^*| : |B^*A| = \frac{\angle CAB}{\angle BCA}$, $|AC^*| : |C^*B| = \frac{\angle ABC}{\angle CAB}$, where A^*, B^*, C^* are the limit points of A_n, B_n, C_n respectively as $n \to \infty$.
3. $|BA^*| : |A^*C| = \frac{\angle ABC + \angle CAB}{\angle BCA + \angle CAB}$, $|CB^*| : |B^*A| = \frac{\angle BCA + \angle ABC}{\angle CAB + \angle ABC}$, $|AC^*| : |C^*B| = \frac{\angle CAB + \angle BCA}{\angle ABC + \angle BCA}$, where A^*, B^*, C^* are the limit points of $\bar{A}_n$, $\bar{B}_n$, $\bar{C}_n$ respectively as $n \to \infty$.
4. $\Delta A_4 B_4 C_4$ is equilateral if and only if ΔABC is equilateral; $\Delta \bar{A}_4 \bar{B}_4 \bar{C}_4$ is equilateral if and only if ΔABC is equilateral.
5. $\Delta A_5 B_5 C_5$ is equilateral if and only if ΔABC is equilateral; $\Delta \bar{A}_5 \bar{B}_5 \bar{C}_5$ is equilateral if and only if ΔABC is equilateral.

These results are proved in a systematical way with interactive, semi-automatic algebraic computations. We will explain how the general approach works along with the derivation and proof of the results and report on some other experiments.

2 Properties of Intersection Points of n-sectors

This section is devoted to the proof of some properties about the intersection points of angular n-sectors of an arbitrary triangle. For brevity and without loss of generality, let the coordinates of the points be taken as $A = (0, 0)$, $B = (w, 0)$ and $C = (u^2, v^2)$ with $wuv \neq 0$. Note that $u = 0$ and $wv \neq 0$ imply $\Delta ABC \cong \Delta ABC'$ with $C' = (w, v^2)$, while $\Delta ABC'$ with $wv \neq 0$ is covered by ΔABC with $wuv \neq 0$. So we can assume that $wuv \neq 0$.

Let ΔABC be an arbitrary triangle with internal angles a, b, c at vertices A, B, C respectively, let S denote the area of the triangle ABC, and take the coordinates of the points as $A = (a_1, a_2)$, $B = (b_1, b_2)$ and $C = (c_1, c_2)$. First we recall some basic results which can be easily proved.

Lemma 21 *The area S of the triangle ABC can be expressed as*

$$S = |(c_2 - b_2)(c_1 - a_1) - (c_2 - a_2)(c_1 - b_1)|/2. \tag{2.1}$$

Proof. The lemma follows directly from the formula $2S = \pm \begin{vmatrix} a_1 & a_2 & 1 \\ b_1 & b_2 & 1 \\ c_1 & c_2 & 1 \end{vmatrix}$.

Lemma 22 *Let B_1 be the result of the point B by rotating the side AB counterclockwise α degree around A. Then the coordinates (x, y) of B_1 can be expressed as:*

$$\begin{aligned} x &= (b_1 - a_1)\cos\alpha - (b_2 - a_2)\sin\alpha + a_1, \\ y &= (b_2 - a_2)\cos\alpha + (b_1 - a_1)\sin\alpha + a_2. \end{aligned} \tag{2.2}$$

Proof. Note that

$$\overrightarrow{AB_1} \cdot \overrightarrow{AB} = |\overrightarrow{AB_1}||\overrightarrow{AB}|\cos\alpha, \tag{2.3}$$

$$|\overrightarrow{AB_1}||\overrightarrow{AB}|\sin\alpha = \pm \begin{vmatrix} a_1 & a_2 & 1 \\ b_1 & b_2 & 1 \\ x & y & 1 \end{vmatrix}, \tag{2.4}$$

and

$$|\overrightarrow{AB_1}| = |\overrightarrow{AB}| = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2}. \tag{2.5}$$

Using the equalities (2.3), (2.4) and (2.5), we have

$$\begin{aligned} x &= (b_1 - a_1)\cos\alpha - (b_2 - a_2)\sin\alpha + a_1, \\ y &= (b_2 - a_2)\cos\alpha + (b_1 - a_1)\sin\alpha + a_2. \end{aligned} \tag{2.6}$$

Now we prove the following result using the above lemmas.

Lemma 23 *For an arbitrary triangle ABC, the three lines AA_n, BB_n, CC_n are concurrent, and so are the three lines $A\bar{A}_n$, $B\bar{B}_n$, $C\bar{C}_n$.*

Proof. We only prove the case in which A_n, B_n, C_n are the intersection points of the internal angular n-sectors; the case for external angular n-sectors can be proved similarly.

Suppose that A_1 and B_1 are the results of A and B respectively by rotating counterclockwise the AB side $2\pi - \frac{b}{n}$ degree around B and the BA side $\frac{a}{n}$ degree around A. Then according to Lemma 22, we have

$$\begin{aligned} A_1 &= \left(-w\cos\left(\frac{b}{n}\right) + w, w\sin\left(\frac{b}{n}\right)\right), \\ B_1 &= \left(w\cos\left(\frac{a}{n}\right), w\sin\left(\frac{a}{n}\right)\right). \end{aligned} \tag{2.7}$$

Hence, the coordinates of the intersection point C_n of the two lines AB_1 and A_1B can be obtained by solving the linear equations of the two lines:

$$C_n = \left(\frac{wy_n}{x_n + y_n}, \frac{wx_ny_n}{x_n + y_n}\right), \tag{2.8}$$

where $x_n = \tan\left(\frac{a}{n}\right)$ and $y_n = \tan\left(\frac{b}{n}\right)$.

In a similar way, one can obtain the coordinates of the intersection points A_n and B_n as follows:

$$A_n = \left(\frac{-v^2y_nz_n + wy_n + u^2z_n}{y_n + z_n}, \frac{z_n(u^2y_n + v^2 - wy_n)}{y_n + z_n}\right), \tag{2.9}$$

$$B_n = \left(\frac{z_n(v^2x_n + u^2)}{x_n + z_n}, -\frac{z_n(u^2x_n - v^2)}{x_n + z_n}\right), \tag{2.10}$$

where $z_n = \tan\left(\frac{\pi - a - b}{n}\right)$.

To prove that AA_n, BB_n, CC_n are concurrent, let $I = (I_1, I_2)$ be the intersection point of the lines AA_n and BB_n. Using the coordinates of the points A, B, A_n, B_n, we have

$$\begin{aligned} I_1 &= \frac{w\left(-v^2y_nz_n + u^2z_n + wy_n\right)\left(u^2x_n - v^2\right)}{G}, \\ I_2 &= \frac{wz_n\left(u^2x_n - v^2\right)\left(u^2y_n + v^2 - wy_n\right)}{G}, \end{aligned} \tag{2.11}$$

where

$$G = \left(u^4 + v^4 + w^2\right)(x_n + y_n)z_n - wy_nz_n\left(v^2x_n + 2u^2\right) - v^2w(x_n + y_n + z_n).$$

Straightforward calculation according to Lemma 21 shows that

$$2S_{\Delta CC_nI} = 0,$$

which implies that AA_n, BB_n, CC_n are concurrent.

Recall that $a = \angle CAB, b = \angle ABC, c = \angle BCA$. The following result can be proved by using Lemmas 22 and 23.

Lemma 24 *When A^*, B^*, C^* are the limit points of the intersections A_n, B_n, C_n respectively as $n \to \infty$,*

$$\frac{|BA^*|}{|A^*C|} = \frac{c}{b}, \quad \frac{|CB^*|}{|B^*A|} = \frac{a}{c}, \quad \frac{|AC^*|}{|C^*B|} = \frac{b}{a}. \tag{2.12}$$

When A^*, B^*, C^* are the limit points of the intersections $\bar{A}_n$, $\bar{B}_n$, $\bar{C}_n$ respectively as $n \to \infty$,

$$\frac{|BA^*|}{|A^*C|} = \frac{b+a}{c+a}, \quad \frac{|CB^*|}{|B^*A|} = \frac{c+b}{a+b}, \quad \frac{|AC^*|}{|C^*B|} = \frac{a+c}{b+c}. \tag{2.13}$$

In both cases, the three lines AA^*, BB^*, CC^* are concurrent.

Proof. We only prove the formulas in (2.12); the formulas in (2.13) can be proved similarly, and the concurrency of AA^*, BB^*, CC^* follows directly from Lemma 23.

Taking (2.8), (2.9) and (2.10) into account and noting that

$$x_n \sim \frac{a}{n}, \quad y_n \sim \frac{b}{n}, \quad z_n \sim \frac{\pi - a - b}{n} \quad \text{as} \;\; n \longrightarrow \infty, \tag{2.14}$$

we can obtain the coordinates of the points A^*, B^*, C^* as follows:

$$\begin{aligned} A^* &= \left(\frac{u^2(\pi - a - b) + bw}{\pi - a}, \frac{v^2(\pi - a - b)}{\pi - a} \right), \\ B^* &= \left(\frac{u^2(\pi - a - b)}{\pi - b}, \frac{v^2(\pi - a - b)}{\pi - b} \right), \\ C^* &= \left(\frac{wb}{a+b}, 0 \right). \end{aligned} \tag{2.15}$$

From the above expressions of coordinates the formulas in (2.12) are obtained by computing the ratios of the lengths of segments.

3 Non-existence of Morley Triangles for Internal *n*-sectors

In this section we investigate the non-existence of Morley triangles of ΔABC for $n > 3$. For this purpose, we describe a general and computational approach that can be used to study the existence or non-existence of Morley triangles for any specific n and m. In particular, we use the approach to prove that Morley triangles do not exist for $n = 4$ and $n = 5$ with $m = 1$. Before doing so, we recall some fundamental relations about trigonometric functions for multiple angles.

Lemma 31 *For any angle θ and integer n,*

$$\begin{aligned}\cos(n\theta) &= \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k C_n^{2k} \cos^{n-2k}\theta \sin^{2k}\theta,\\ \sin(n\theta) &= \sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^k C_n^{2k+1} \cos^{n-(2k+1)}\theta \sin^{2k+1}\theta;\end{aligned} \tag{3.1}$$

$$\tan(n\theta) = \frac{\sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^k C_n^{2k+1} \tan^{2k+1}\theta}{\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k C_n^{2k} \tan^{2k}\theta}. \tag{3.2}$$

Proof. Write $z = e^{i\theta} = \cos\theta + i\sin\theta$. Then $z^n = \cos(n\theta) + i\sin(n\theta) = (\cos\theta + i\sin\theta)^n$. By using the binomial theorem and noting the parity of n, the formulas (3.1) can be easily obtained. The formula (3.2) follows directly from (3.1).

Corollary 32 *For any $n \geq 3$ and $n > m > 0$, $\cos(m\pi/n)$ is a real root of*

$$\psi_1(x) = \sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k C_n^{2k} x^{n-2k}(1-x^2)^k + (-1)^{m+1}; \tag{3.3}$$

$\sin(m\pi/n)$ *is either a real root of*

$$\bar{\psi}_1(x) = \sum_{k=0}^{\frac{n}{2}} (-1)^k C_n^{2k} (1-x^2)^{\frac{n}{2}-k} x^{2k} + (-1)^{m+1} = 0 \tag{3.4}$$

when n is even, or a real root of

$$\bar{\psi}_2(x) = \sum_{k=0}^{\frac{n-1}{2}} (-1)^k C_n^{2k+1} (1-x^2)^{\frac{n-1}{2}-k} x^{2k+1} = 0 \tag{3.5}$$

when n is odd.

3.1 General Computational Approach

It is easy to show that if ΔABC is equilateral, then so is $\Delta A_nB_nC_n$. Now suppose that $\Delta A_nB_nC_n$ is equilateral and let $\tilde{B}_n$ and $\tilde{A}_n$ be the results of B_n and A_n respectively by rotating counterclockwise the A_nB_n side $\frac{\pi}{3}$ degree around A_n and the A_nB_n side $2\pi - \frac{\pi}{3}$ degree around B_n. Then by using (2.9), (2.10) and Lemmas 21 and 22 and the relations $v^2 = \tan(a) \cdot u^2$ and $w = u^2 + \frac{\tan(a)\cdot u^2}{\tan(b)}$, the area S_1 of $\Delta\tilde{B}_nA_nC_n$ and the area S_2 of $\Delta\tilde{A}_nB_nC_n$ can be expressed as follows:

$$S_1 = |S_{\Delta\tilde{B}_nA_nC_n}| = \left| \frac{\sqrt{3}u^4 f_1}{12(t_ny_n + l_n)(t_nx_n + l_n)^2(k_ny_n + s_n)e_2^2} \right| = 0, \tag{3.6}$$

$$S_2 = |S_{\Delta \tilde{A}_n B_n C_n}| = \left| \frac{u^4 f_2}{4(t_n y_n + l_n)^2 (t_n x_n + l_n)(k_n y_n + s_n) e_2^2} \right| = 0, \tag{3.7}$$

where f_1 and f_2 are polynomials consisting of 105 and 99 terms, respectively, in $x_n, y_n, k_n, s_n, t_n, l_n, e_1, e_2$ with

$$\begin{aligned} &x_n = \tan(ma/n), \quad y_n = \tan(mb/n), \quad k_n = \cos(ma/n), \\ &s_n = \sin(ma/n), \quad t_n = \cos(m(\pi - a - b)/n), \\ &l_n = \sin(m(\pi - a - b)/n), \quad e_1 = \tan(a), \quad e_2 = \tan(b). \end{aligned} \tag{3.8}$$

Moreover, we have the following relations:

$$\begin{aligned} &k_n = \frac{1}{\sqrt{1 + x_n^2}}, \quad s_n = \frac{x_n}{\sqrt{1 + x_n^2}}, \\ &t_n = p_n \cdot \frac{1 - x_n y_n}{\sqrt{(1 + x_n^2)(1 + y_n^2)}} + q_n \cdot \frac{x_n + y_n}{\sqrt{(1 + x_n^2)(1 + y_n^2)}}, \\ &l_n = -p_n \cdot \frac{x_n + y_n}{\sqrt{(1 + x_n^2)(1 + y_n^2)}} + q_n \cdot \frac{1 - x_n y_n}{\sqrt{(1 + x_n^2)(1 + y_n^2)}}; \end{aligned} \tag{3.9}$$

$$\begin{aligned} \tan(ma) &= \frac{\sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^k C_n^{2k+1} x_n^{2k+1}}{\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k C_n^{2k} x_n^{2k}} = \frac{\sum_{k=0}^{\lfloor \frac{m-1}{2} \rfloor} (-1)^k C_m^{2k+1} e_1^{2k+1}}{\sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} (-1)^k C_m^{2k} e_1^{2k}}, \\ \tan(mb) &= \frac{\sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^k C_n^{2k+1} y_n^{2k+1}}{\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k C_n^{2k} y_n^{2k}} = \frac{\sum_{k=0}^{\lfloor \frac{m-1}{2} \rfloor} (-1)^k C_m^{2k+1} e_2^{2k+1}}{\sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} (-1)^k C_m^{2k} e_2^{2k}}, \end{aligned} \tag{3.10}$$

where $p_n = \cos(m\pi/n)$ and $q_n = \sin(m\pi/n)$ and the equalities in (3.10) follow from Lemma 31.

By substituting (3.9) into (3.6) and (3.7), we can reduce S_1 and S_2 to rational functions of $x_n, y_n, p_n, q_n, e_1, e_2$ with p_n a real root of a polynomial $\psi_1(x)$ and q_n a real root of another polynomial $\bar{\psi}_1(x)$ (or $\bar{\psi}_2(x)$) such that $p_n^2 + q_n^2 = 1$. Then for $S_1 = 0$ and $S_2 = 0$ we can construct seven polynomial equations in $x_n, y_n, p_n, q_n, e_1, e_2$:

$$\begin{cases} \bar{S}_1^{(n)}(x_n, y_n, p_n, q_n, e_1, e_2) = 0, \\ \bar{S}_2^{(n)}(x_n, y_n, p_n, q_n, e_1, e_2) = 0, \\ \psi_1(p_n) = 0, \\ \bar{\psi}(q_n) = 0, \\ H(p_n, q_n) = p_n^2 + q_n^2 - 1 = 0, \\ H_1^{(n)}(x_n, e_1) = 0, \\ H_2^{(n)}(y_n, e_2) = 0, \end{cases} \tag{3.11}$$

where $\bar{\psi}$ takes either $\bar{\psi}_1$ or $\bar{\psi}_2$ in Corollary 32, $\bar{S}_1^{(n)}$ and $\bar{S}_2^{(n)}$ can be derived from $S_1 = 0$ and $S_2 = 0$, respectively, and the expressions of $H_1^{(n)}(x_n, e_1)$ and $H_2^{(n)}(y_n, e_2)$ can be derived from the relations in (3.10).

One can solve these equations for x_n and y_n using elimination methods based on Gröbner bases and triangular sets [6]. Heuristically, we may try to compute the values of p_n and q_n, substitute them into S_1 and S_2 to obtain $\bar{S}_1(x_n, y_n, e_1, e_2)$ and $\bar{S}_2(x_n, y_n, e_1, e_2)$, and then solve the polynomial equations $\bar{S}_1 = \bar{S}_2 = H_1^{(n)} = H_2^{(n)} = 0$ for x_n, y_n. If $x_n = y_n = \tan(m\pi/3n)$ is the unique solution, then the triangle ABC is proved to be equilateral.

The approach explained above involves heavy computations with complicated expressions. It may not work effectively when n or m is big. In the following subsections, we will provide computational details for $n = 4, 5$ and $m = 1$. When $m = 1$, the relations in (3.10) have the following form

$$\begin{aligned} e_1 &= \frac{\sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^k C_n^{2k+1} x_n^{2k+1}}{\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k C_n^{2k} x_n^{2k}}, \\ e_2 &= \frac{\sum_{k=0}^{\lfloor \frac{n-1}{2} \rfloor} (-1)^k C_n^{2k+1} y_n^{2k+1}}{\sum_{k=0}^{\lfloor \frac{n}{2} \rfloor} (-1)^k C_n^{2k} y_n^{2k}}. \end{aligned} \tag{3.12}$$

By using the above relations, the set of polynomial equations in (3.11) can be reduced to another set of polynomial equations in x_n, y_n, p_n, q_n.

3.2 The Case $n = 4$

In this case, using (3.3) and (3.4), we have

$$\psi_1(p_4) = 2\left(2p_4^2 - 1\right)^2 = 0, \quad \bar{\psi}_1(q_4) = 2\left(2q_4^2 - 1\right)^2 = 0, \tag{3.13}$$

so $p_4 = q_4 = \frac{\sqrt{2}}{2}$. Then substituting (3.9) and (3.12) into (3.6) and (3.7), we obtain the following expressions:

$$S_1 = \left| \frac{(1+\sqrt{3})x_4^2(x_4^2+1)(y_4^2+1)f_1^2(x_4,y_4)\bar{H}_1^{(4)}(x_4,y_4)}{4(x_4^2+2x_4-1)^2(x_4^2-2x_4-1)^2(y_4^2-1)^2} \right| = 0, \tag{3.14}$$

$$S_2 = \left| \frac{(\sqrt{3}-1)x_4^2(x_4^2+1)(y_4^2+1)f_1^2(x_4,y_4)\bar{H}_2^{(4)}(x_4,y_4)}{4(x_4^2+2x_4-1)^2(x_4^2-2x_4-1)^2(y_4^2-1)^2} \right| = 0, \tag{3.15}$$

where

$$\begin{aligned} f_1(x_4, y_4) &= x_4y_4 + x_4 + y_4 - 1, \\ \bar{H}_1^{(4)}(x_4, y_4) &= -\sqrt{3}x_4y_4^2 + \sqrt{3}x_4^2 - \sqrt{3}x_4y_4 + \sqrt{3}y_4^2 - x_4^2y_4 + 2x_4y_4^2 + 2\sqrt{3}x_4 \\ &\quad + 3\sqrt{3}y_4 - 2x_4^2 + x_4y_4 - 2y_4^2 - 3x_4 - 6y_4 - 5 + 3\sqrt{3}, \\ \bar{H}_2^{(4)}(x_4, y_4) &= \sqrt{3}x_4y_4^2 + \sqrt{3}x_4y_4 - x_4^2y_4 + 2x_4y_4^2 - y_4\sqrt{3} + x_4^2 + x_4y_4 + y_4^2 \\ &\quad + 3x_4 + 1 - \sqrt{3}. \end{aligned} \tag{3.16}$$

Note that the equality $f_1(x_4, y_4) = 0$ implies $\frac{x_4+y_4}{1-x_4y_4} = 1 = \tan\left(\frac{a+b}{4}\right)$; thus $a + b = \pi$ and in this case ΔABC degenerates to a line.

Now compute the Gröbner basis of the polynomial set

$$\left[\bar{H}_1^{(4)}(x_4, y_4), \bar{H}_2^{(4)}(x_4, y_4)\right]$$

with respect to the lex (lex) term ordering determined by $y_4 \succ x_4$. One can find that the Gröbner basis contains the following two polynomials:

$$\begin{aligned} G_1 &= \left(x_4^2 + 1\right)(\sqrt{3} + x - 2)(x_4 + 1)^3, \\ G_2 &= -\frac{1}{13}(\sqrt{3} + 4)(x_4 + 1)\left(14\sqrt{3}x_4^3 + 13x_4^4 + 18\sqrt{3}x_4^2 - 4x_4^3 + 18\sqrt{3}x_4\right. \\ &\quad \left. +4y_4\sqrt{3} - 20x_4^2 + 18\sqrt{3} - 20x_4 - 16y_4 - 33\right). \end{aligned} \tag{3.17}$$

$G_1 = 0$ has two real roots for x_4: $x_4 = -1$ and $x_4 = 2 - \sqrt{3}$. Since $0 < a < \pi$, x_4 is required to be in $(0, 1)$. Hence the first root $x_4 = -1$ need not be considered. Substituting the second root $x_4 = 2 - \sqrt{3}$ into G_2, we have $(x_4, y_4) = (2 - \sqrt{3}, 2 - \sqrt{3})$. This means that $a = b = 4 \arctan\left(2 - \sqrt{3}\right) = \frac{\pi}{3}$. Therefore, $\Delta A_4B_4C_4$ is equilateral if and only if ΔABC is equilateral.

3.3 The Case $n = 5$

In this case, using (3.3) and (3.5), we have

$$\psi_1(p_5) = (p_5 + 1)\left(4p_5^2 - 2p_5 - 1\right)^2 = 0, \quad \bar{\psi}_2(q_5) = q_5\left(16q_5^4 - 20q_5^2 + 5\right) = 0. \tag{3.18}$$

The Gröbner basis of the polynomial set

$$\left[4p_5^2 - 2p_5 - 1, 16q_5^4 - 20q_5^2 + 5, p_5^2 + q_5^2 - 1\right]$$

with respect to the lex term ordering determined by $p_5 \succ q_5$ can be easily computed. It is

$$\left[16q_5^4 - 20q_5^2 + 5, 4q_5^2 + 2p_5 - 3\right], \tag{3.19}$$

so $p_5 = \frac{\sqrt{5}+1}{4}$, $q_5 = \frac{\sqrt{10-2\sqrt{5}}}{4}$. Substituting (3.9) and (3.12) into (3.6) and (3.7), we obtain the following expressions:

$$S_1 = \left|\bar{k} \cdot \frac{x_5^2(x_5^2 + 1)(y_5^2 + 1)f_2^2(x_5, y_5)\bar{H}_1^{(5)}(x_5, y_5)}{12800(y_5^4 - 10y_5^2 + 5)^2(5x_5^4 - 10x_5^2 + 1)^2}\right| = 0, \tag{3.20}$$

$$S_2 = \left|\bar{k} \cdot \frac{x_5^2(x_5^2 + 1)(y_5^2 + 1)f_2^2(x_5, y_5)\bar{H}_2^{(5)}(x_5, y_5)}{12800(y_5^4 - 10y_5^2 + 5)^2(5x_5^4 - 10x_5^2 + 1)^2}\right| = 0, \tag{3.21}$$

where

$$\begin{aligned}\bar{k} &= 5\sqrt{10-2\sqrt{5}}(\sqrt{5}-1)+8\sqrt{15}-20\sqrt{3},\\ f_2(x_5,y_5) &= \sqrt{5}(3+\sqrt{5})\sqrt{10-2\sqrt{5}}(x_5+y_5)+20x_5y_5-20,\end{aligned} \tag{3.22}$$

and $\bar{H}_1^{(5)}(x_5,y_5)$ and $\bar{H}_2^{(5)}(x_5,y_5)$ are polynomials, both consisting of 93 terms, in x_5, y_5.

The equality $f_2(x_5,y_5)=0$ implies $\frac{x_5+y_5}{1-x_5y_5}=\sqrt{5-2\sqrt{5}}=\frac{q_5}{p_5}=\tan\left(\frac{a+b}{5}\right)$, so $a+b=\pi$ and in this case ΔABC degenerates to a line.

Computing the Gröbner basis of the polynomial set

$$\left[\bar{H}_1^{(5)}(x_5,y_5),\bar{H}_2^{(5)}(x_5,y_5)\right]$$

with respect to the lex term ordering determined by $y_5 \succ x_5$, one can see that the Gröbner basis has five elements and the first is

$$\begin{aligned}G_1 = {} & \frac{1}{16777216}\left(x_5+\sqrt{3}\right)\left(x_5^2+1\right)^3\left(\sqrt{10-2\sqrt{5}}(\sqrt{5}-3)-2\sqrt{15}+6\sqrt{3}-4x_5\right)\\ & \left(-\sqrt{15}+\sqrt{10-2\sqrt{5}}(\sqrt{5}+2)+2x_5-3\sqrt{3}\right)\left(\sqrt{10-2\sqrt{5}}(\sqrt{5}-1)+4x_5\right)^3\\ & \left(\sqrt{10-2\sqrt{5}}(\sqrt{5}+1)+2\sqrt{15}+2\sqrt{3}-4x_5\right)\left(\sqrt{10-2\sqrt{5}}(\sqrt{5}+3)-4x_5\right)^3\\ & \left(\sqrt{10-2\sqrt{5}}(\sqrt{5}+3)+4x_5\right)^3\left(\sqrt{10-2\sqrt{5}}(\sqrt{5}+1)-2\sqrt{15}-2\sqrt{3}-4x_5\right).\end{aligned}$$

$G_1=0$ has only one real root for x_5 in $(0,1)$: $x_5=\frac{2\sqrt{3}-\sqrt{10-2\sqrt{5}}}{3+\sqrt{5}}$. Substituting this root into the second element of the Gröbner basis resulting in G_2 and solving $G_2=0$ for y_5, one may obtain $(x_5,y_5)=\left(\frac{2\sqrt{3}-\sqrt{10-2\sqrt{5}}}{3+\sqrt{5}},\frac{2\sqrt{3}-\sqrt{10-2\sqrt{5}}}{3+\sqrt{5}}\right)$. According to (3.2), we have $\tan\left(\frac{\pi}{5}\right)=\frac{3\tan\left(\frac{\pi}{15}\right)-\tan^3\left(\frac{\pi}{15}\right)}{1-3\tan^2\left(\frac{\pi}{15}\right)}=\frac{\sqrt{10-2\sqrt{5}}}{\sqrt{5}+1}$, so $\tan\left(\frac{\pi}{15}\right)$ is a real root of the polynomial

$$f=(1+\sqrt{5})x^3-3\sqrt{10-2\sqrt{5}}x^2-3(1+\sqrt{5})x+\sqrt{10-2\sqrt{5}}.$$

On the other hand, direct computation shows that $f(x_5)=0$. This means that $a=b=5\arctan\left(\frac{2\sqrt{3}-\sqrt{10-2\sqrt{5}}}{3+\sqrt{5}}\right)=\frac{\pi}{3}$. Hence $\Delta A_5B_5C_5$ is equilateral if and only if ΔABC is equilateral.

4 Non-existence of Morley Triangles for External n-sectors

4.1 General Computational Approach

As in the case for internal angular n-sectors, one can easily prove that $\Delta\bar{A}_n\bar{B}_n\bar{C}_n$ when ΔABC is equilateral. To show the necessity of the condition, we suppose that $\Delta \bar{A}_n \bar{B}_n \bar{C}_n$ is equilateral. The following notations are introduced to simplify the involved expressions:

$$\begin{aligned}
&\bar{x}_n = \tan\left(m(\pi - a)/n\right), \quad \bar{y}_n = \tan\left(m(\pi - b)/n\right), \\
&\bar{k}_n = \cos\left(m(\pi - a)/n\right), \quad \bar{s}_n = \sin\left(m(\pi - a)/n\right), \\
&\bar{t}_n = \cos\left(m(a + b)/n\right), \quad \bar{l}_n = \sin\left(m(a + b)/n\right), \\
&e_1 = \tan(a), \quad e_2 = \tan(b).
\end{aligned} \tag{4.1}$$

Similar to (3.6) and (3.7), we have the following two expressions

$$S_1 = \left| \frac{\sqrt{3} u^4 g_1}{12(\bar{t}_n \bar{y}_n + \bar{l}_n)(\bar{t}_n \bar{x}_n + \bar{m}_n)^2 (\bar{k}_n \bar{y}_n + \bar{s}_n) e_2^2} \right| = 0, \tag{4.2}$$

$$S_2 = \left| \frac{u^4 g_2}{4(\bar{t}_n \bar{y}_n + \bar{l}_n)^2 (\bar{t}_n \bar{x}_n + \bar{m}_n)(\bar{k}_n \bar{y}_n + \bar{s}_n) e_2^2} \right| = 0, \tag{4.3}$$

where g_1 and g_2 are polynomials in $\bar{x}_n, \bar{y}_n, \bar{k}_n, \bar{s}_n, \bar{t}_n, \bar{m}_n, e_1, e_2$ consisting of 105 and 99 terms, respectively. Moreover, we have the following relations:

$$\begin{aligned}
\bar{k}_n &= \frac{1}{\sqrt{1 + \bar{x}_n^2}}, \quad \bar{s}_n = \frac{\bar{x}_n}{\sqrt{1 + \bar{x}_n^2}}, \\
\bar{t}_n &= \left(2p_n^2 - 1\right) \cdot \frac{1 - \bar{x}_n \bar{y}_n}{\sqrt{(1 + \bar{x}_n^2)(1 + \bar{y}_n^2)}} + 2p_n q_n \cdot \frac{\bar{x}_n + \bar{y}_n}{\sqrt{(1 + \bar{x}_n^2)(1 + \bar{y}_n^2)}}, \\
\bar{m}_n &= -\left(2p_n^2 - 1\right) \cdot \frac{\bar{x}_n + \bar{y}_n}{\sqrt{(1 + \bar{x}_n^2)(1 + \bar{y}_n^2)}} + 2p_n q_n \cdot \frac{1 - \bar{x}_n \bar{y}_n}{\sqrt{(1 + \bar{x}_n^2)(1 + \bar{y}_n^2)}};
\end{aligned} \tag{4.4}$$

$$\begin{aligned}
\tan(m(\pi - a)) &= \frac{\sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} (-1)^k C_n^{2k+1} \bar{x}_n^{2k+1}}{\sum_{k=0}^{\left\lfloor \frac{n}{2} \right\rfloor} (-1)^k C_n^{2k} \bar{x}_n^{2k}} = \frac{\sum_{k=0}^{\left\lfloor \frac{m-1}{2} \right\rfloor} (-1)^k C_m^{2k+1} (-e_1)^{2k+1}}{\sum_{k=0}^{\left\lfloor \frac{m}{2} \right\rfloor} (-1)^k C_m^{2k} (-e_1)^{2k}}, \\
\tan(m(\pi - b)) &= \frac{\sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} (-1)^k C_n^{2k+1} \bar{y}_n^{2k+1}}{\sum_{k=0}^{\left\lfloor \frac{n}{2} \right\rfloor} (-1)^k C_n^{2k} \bar{y}_n^{2k}} = \frac{\sum_{k=0}^{\left\lfloor \frac{m-1}{2} \right\rfloor} (-1)^k C_m^{2k+1} (-e_2)^{2k+1}}{\sum_{k=0}^{\left\lfloor \frac{m}{2} \right\rfloor} (-1)^k C_m^{2k} (-e_2)^{2k}}.
\end{aligned} \tag{4.5}$$

where $p_n = \cos\left(m\pi/n\right)$ and $q_n = \sin\left(m\pi/n\right)$ and the equalities in (4.5) follow from Lemma 31. When $m = 1$, the expressions in (4.5) simplify to

$$\begin{aligned}
e_1 &= -\tan(\pi - a) = -\frac{\sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} (-1)^k C_n^{2k+1} \bar{x}_n^{2k+1}}{\sum_{k=0}^{\left\lfloor \frac{n}{2} \right\rfloor} (-1)^k C_n^{2k} \bar{x}_n^{2k}}, \\
e_2 &= -\tan(\pi - b) = -\frac{\sum_{k=0}^{\left\lfloor \frac{n-1}{2} \right\rfloor} (-1)^k C_n^{2k+1} \bar{y}_n^{2k+1}}{\sum_{k=0}^{\left\lfloor \frac{n}{2} \right\rfloor} (-1)^k C_n^{2k} \bar{y}_n^{2k}}.
\end{aligned} \tag{4.6}$$

Now substitute (4.4) and (4.6) to (4.2) and (4.3) and form the following set of polynomials:

$$\begin{cases} \bar{S}_1^{(n)}(\bar{x}_n, \bar{y}_n, p_n, q_n) = 0, \\ \bar{S}_2^{(n)}(\bar{x}_n, \bar{y}_n, p_n, q_n) = 0, \\ \psi_1(p_n) = 0, \\ \bar{\psi}(q_n) = 0, \\ H(p_n, q_n) = p_n^2 + q_n^2 - 1 = 0, \end{cases} \tag{4.7}$$

where $\bar{\psi}$ takes either $\bar{\psi}_1$ or $\bar{\psi}_2$ in Corollary 32, $\bar{S}_1^{(n)}$ and $\bar{S}_2^{(n)}$ can be derived from $S_1 = 0$ and $S_2 = 0$, respectively.

In the following subsections, we will study the special cases for $n = 4, 5$ using Gröbner bases without computing the concrete values of p_n and q_n.

4.2 The Case $n = 4$

In this case, substituting (4.4) and (4.6) into (4.2) and (4.3), we obtain the following expressions:

$$S_1 = \left| \frac{\sqrt{3}\bar{x}_4^2(\bar{x}_4^2+1)(\bar{y}_4^2+1)Q_1(\bar{x}_4, \bar{y}_4, p_4, q_4)}{2\bar{Q}_1(\bar{x}_4, \bar{y}_4)(2\bar{x}_4 p_4^2 - 2p_4 q_4 - \bar{x}_4)(2\bar{y}_4 p_4^2 - 2p_4 q_4 - \bar{y}_4)^2} \right| = 0, \tag{4.8}$$

$$S_2 = \left| \frac{-\sqrt{3}\bar{x}_4^2(\bar{x}_4^2+1)(\bar{y}_4^2+1)Q_2(\bar{x}_4, \bar{y}_4, p_4, q_4)}{2\bar{Q}_1(\bar{x}_4, \bar{y}_4)(2\bar{x}_4 p_4^2 - 2p_4 q_4 - \bar{x}_4)^2(2\bar{y}_4 p_4^2 - 2p_4 q_4 - \bar{y}_4)} \right| = 0, \tag{4.9}$$

where

$$\bar{Q}_1(\bar{x}_4, \bar{y}_4) = 6(\bar{x}_4^2 + 2\bar{x}_4 - 1)^2(\bar{x}_4^2 - 2\bar{x}_4 - 1)^2(\bar{y}_4^2 - 1)^2$$

and $Q_1(\bar{x}_4, \bar{y}_4, p_4, q_4)$, $Q_2(\bar{x}_4, \bar{y}_4, p_4, q_4)$ are polynomials, both consisting of 304 terms, in $\bar{x}_4$, $\bar{y}_4$, p_4, q_4.

Refer to (3.13) and compute the Gröbner basis of the polynomial set

$$\left[2p_4^2 - 1, 2q_4^2 - 1, Q_1(\bar{x}_4, \bar{y}_4, p_4, q_4), Q_2(\bar{x}_4, \bar{y}_4, p_4, q_4)\right]$$

with respect to the lex term ordering determined by $p_4 \succ q_4 \succ \bar{y}_4 \succ \bar{x}_4$. One can see that the Gröbner basis contains the following two polynomials:

$$\begin{aligned} G_1 &= -(\bar{x}_4 + \bar{y}_4)(-3\bar{x}_4 + \sqrt{3})(\bar{x}_4\bar{y}_4 - 1)^2, \\ G_2 &= -(\bar{x}_4 - \bar{y}_4)(\bar{x}_4 + \bar{y}_4)(\bar{x}_4\bar{y}_4 - 1)^2. \end{aligned} \tag{4.10}$$

Then it follows from (4.10) that $\bar{x}_4 = \bar{y}_4 = \frac{\sqrt{3}}{3}$. Hence $a = b = \frac{\pi}{3}$ and thus $\Delta\bar{A}_4\bar{B}_4\bar{C}_4$ is equilateral if and only if ΔABC is equilateral.

4.3 The Case $n = 5$

Substituting (4.4) and (4.6) into (4.2) and (4.3), we obtain the following expressions:

$$S_1 = \left| \frac{-\bar{x}_5^2(\bar{x}_5^2+1)(\bar{y}_5^2+1)Q_3(\bar{x}_5,\bar{y}_5,p_5,q_5)}{2\bar{Q}_2(\bar{x}_5,\bar{y}_5)(2\bar{x}_5p_5^2-2p_5q_5-\bar{x}_5)(2\bar{y}_5p_5^2-2p_5q_5-\bar{y}_5)^2} \right| = 0, \qquad (4.11)$$

$$S_2 = \left| \frac{\bar{x}_5^2(\bar{x}_5^2+1)(\bar{y}_5^2+1)Q_3(\bar{x}_5,\bar{y}_5,p_5,q_5)}{2\bar{Q}_2(\bar{x}_5,\bar{y}_5)(2\bar{x}_5p_5^2-2p_5q_5-\bar{x}_5)^2(2\bar{y}_5p_5^2-2p_5q_5-\bar{y}_5)} \right| = 0, \qquad (4.12)$$

where

$$\bar{Q}_2(\bar{x}_5,\bar{y}_5) = 2(5\bar{x}_5^4-10\bar{x}_5^2+1)^2(\bar{y}_5^4-10\bar{y}_5^2+5)^2$$

and $Q_3(\bar{x}_5,\bar{y}_5,p_5,q_5)$, $Q_4(\bar{x}_5,\bar{y}_5,p_5,q_5)$ are polynomials, both consisting of 692 terms, in $\bar{x}_5,\bar{y}_5,p_5,q_5$.

Now refer to (3.19) and compute the Gröbner basis of

$$\left[16q_5^4-20q_5^2+5, 4q_5^2+2p_5-3, Q_3(\bar{x}_5,\bar{y}_5,p_5,q_5), Q_4(\bar{x}_5,\bar{y}_5,p_5,q_5)\right]$$

with respect to the lex term ordering determined by $p_5 \succ q_5 \succ \bar{y}_5 \succ \bar{x}_5$. The Gröbner basis contains the following three polynomials:

$$\begin{aligned}
G_1 &= (-\bar{x}_5+\sqrt{3})(\bar{x}_5^4-10\bar{x}_5^2+5)^2(\bar{y}_5^4-10\bar{y}_5^2+5)\cdot g_{10}(\bar{x}_5)\cdot g_{11}(\bar{x}_5)\\
&\quad \cdot g_{12}(\bar{x}_5)\cdot g_{13}(\bar{x}_5)\cdot \bar{g}^2(\bar{x}_5,\bar{y}_5),\\
G_2 &= (\bar{x}_5-\sqrt{3})(\bar{x}_5^4-10\bar{x}_5^2+5)^2(\bar{y}_5^4-10\bar{y}_5^2+5)\cdot g_{12}(\bar{x}_5)\cdot \bar{g}^2(\bar{x}_5,\bar{y}_5)\cdot g_{20}(\bar{x}_5,\bar{y}_5),\\
G_3 &= \frac{\sqrt{3}}{3}(\bar{x}_5-\sqrt{3})(\bar{x}_5^4-10\bar{x}_5^2+5)(\bar{y}_5^4-10\bar{y}_5^2+5)\cdot \bar{g}^2(\bar{x}_5,\bar{y}_5)\cdot g_{30}(\bar{x}_5,\bar{y}_5),
\end{aligned} \qquad (4.13)$$

where

$$\begin{aligned}
g_{10}(\bar{x}_5) &= \bar{x}_5^4+2\sqrt{3}\bar{x}_5^3-8\bar{x}_5^2-6\sqrt{3}\bar{x}_5-1,\\
g_{11}(\bar{x}_5) &= -\bar{x}_5^4+2\sqrt{3}\bar{x}_5^3+8\bar{x}_5^2-6\sqrt{3}\bar{x}_5+1,\\
g_{12}(\bar{x}_5) &= \bar{x}_5^4+6\sqrt{3}\bar{x}_5^3+8\bar{x}_5^2-2\sqrt{3}\bar{x}_5-1,\\
g_{13}(\bar{x}_5) &= \bar{x}_5^8+10\sqrt{3}\bar{x}_5^7-270\bar{x}_5^6+270\sqrt{3}\bar{x}_5^5+1920\bar{x}_5^4-450\sqrt{3}\bar{x}_5^3-4050\bar{x}_5^2\\
&\quad -1350\sqrt{3}\bar{x}_5-225,
\end{aligned}$$

$$
\begin{aligned}
\bar{g}(\bar{x}_5,\bar{y}_5) &= 5\bar{x}_5^4\bar{y}_5^4 - 10\bar{x}_5^4\bar{y}_5^2 - 40\bar{x}_5^3\bar{y}_5^3 - 10\bar{x}_5^2\bar{y}_5^4 + \bar{x}_5^4 + 24\bar{x}_5^3\bar{y}_5 + 76\bar{x}_5^2\bar{y}_5^2 \\
&\quad + 24\bar{x}_5\bar{y}_5^3 + \bar{y}_5^4 - 10\bar{x}_5^2 - 40\bar{x}_5\bar{y}_5 - 10\bar{y}_5^2 + 5, \\
g_{20}(\bar{x}_5,\bar{y}_5) &= 139630603323\bar{x}_5^{15} + 1325638479773\sqrt{3}\bar{x}_5^{14} - 43618989541455\bar{x}_5^{13} \\
&\quad + 20713805047231\sqrt{3}\bar{x}_5^{12} + 1309804139334495\bar{x}_5^{11} \\
&\quad - 1593569669580023\sqrt{3}\bar{x}_5^{10} - 10683063248257395\bar{x}_5^{9} \\
&\quad + 11855658310269299\sqrt{3}\bar{x}_5^{8} + 36929053837557825\bar{x}_5^{7} \\
&\quad - 24961218229635097\sqrt{3}\bar{x}_5^{6} - 60837734425145181\bar{x}_5^{5} \\
&\quad + 10612519767932685\sqrt{3}\bar{x}_5^{4} + 30600097364253285\bar{x}_5^{3} \\
&\quad + 2197362729798195\sqrt{3}\bar{x}_5^{2} - 335010615030945\bar{x}_5 \\
&\quad - 25535661371775\sqrt{3} + 4349038510080\bar{y}_5, \\
g_{30}(\bar{x}_5,\bar{y}_5) &= 3780607290505129674\bar{x}_5^{23} + 58565295021262810097\sqrt{3}\bar{x}_5^{22} \\
&\quad - 543044613081268293018\bar{x}_5^{21} - 6830167258693173476667\sqrt{3}\bar{x}_5^{20} \\
&\quad + 4102128303756786859644 6\bar{x}_5^{19} + 239114170312647965154867\sqrt{3}\bar{x}_5^{18} \\
&\quad - 1148211248966737731853086\bar{x}_5^{17} \\
&\quad - 3624646644309581826234801\sqrt{3}\bar{x}_5^{16} \\
&\quad + 12751026613518347458501188\bar{x}_5^{15} \\
&\quad + 27683733435777510392859898\sqrt{3}\bar{x}_5^{14} \\
&\quad - 58590468745686306763205412\bar{x}_5^{13} \\
&\quad - 111521688036076998229200318\sqrt{3}\bar{x}_5^{12} \\
&\quad + 95327188785322057511669148\bar{x}_5^{11} \\
&\quad + 227081592764233751659402806\sqrt{3}\bar{x}_5^{10} \\
&\quad + 83346778311498132197833 32\bar{x}_5^{9} \\
&\quad - 201617639413877564812687890\sqrt{3}\bar{x}_5^{8} \\
&\quad - 1194048528519610368 0\sqrt{3}\bar{y}_5\bar{x}_5^{7} - 949282134153501634594701 90\bar{x}_5^{7} \\
&\quad + 72322240025105140832359205\sqrt{3}\bar{x}_5^{6} - 1970180072057357107 20\bar{y}_5\bar{x}_5^{6} \\
&\quad + 12736517637542510592 0\sqrt{3}\bar{y}_5\bar{x}_5^{5} + 495241811043203651592326 70\bar{x}_5^{5} \\
&\quad + 21114758145988443340 80\bar{y}_5\bar{x}_5^{4} - 7400684797986359317228215\sqrt{3}\bar{x}_5^{4} \\
&\quad - 6301309033677904387001370\bar{x}_5^{3} - 1870676028014056243 20\sqrt{3}\bar{y}_5\bar{x}_5^{3} \\
&\quad - 21592377557396287488 00\bar{y}_5\bar{x}_5^{2} - 2160146073089123084350 65\sqrt{3}\bar{x}_5^{2} \\
&\quad + 1830874410396735897 60\sqrt{3}\bar{y}_5\bar{x}_5 + 63727676946131438007450\bar{x}_5 \\
&\quad + 6368258818771255296 0\sqrt{3}\bar{y}_5^{2} - 995040440433008640 0\bar{y}_5 \\
&\quad + 3192191611129104585075\sqrt{3}.
\end{aligned}
$$

We want to find all the solutions of $G_1 = G_2 = G_3 = 0$ for $(\bar{x}_5, \bar{y}_5)$ with $\bar{x}_5, \bar{y}_5 \in (0, \tan\left(\frac{\pi}{5}\right)) = (0, \sqrt{5 - 2\sqrt{5}})$. One can check that if $\bar{x}_5^4 - 10\bar{x}_5^2 + 5 = 0$ (or $\bar{y}_5^4 - 10\bar{y}_5^2 + 5 = 0$), then $\bar{x}_5 = \sqrt{5 - 2\sqrt{5}}$ (or $\bar{y}_5 = \sqrt{5 - 2\sqrt{5}}$), which implies $a = 0$ (or $b = 0$); in this case ΔABC degenerates to a line. Note that $\tan(\pi - a) = \frac{\bar{x}_5^5 - 10\bar{x}_5^3 + 5\bar{x}_5}{5\bar{x}_5^4 - 10\bar{x}_5^2 + 1}$ and $\tan(\pi - b) = \frac{\bar{y}_5^5 - 10\bar{y}_5^3 + 5\bar{y}_5}{5\bar{y}_5^4 - 10\bar{y}_5^2 + 1}$. We have

$$\begin{aligned} \tan(a + b) &= -\tan(\pi - a + \pi - b) = \frac{\tan(\pi - a) + \tan(\pi - b)}{\tan(\pi - a)\tan(\pi - b) - 1} \\ &= \frac{(\bar{x}_5 + \bar{y}_5)\bar{g}(\bar{x}_5, \bar{y}_5)}{(\bar{x}_5\bar{y}_5 - 1)g^*(\bar{x}_5, \bar{y}_5)}, \end{aligned} \tag{4.14}$$

where

$$\begin{aligned} g^*(\bar{x}_5, \bar{y}_5) = {} & \bar{x}_5^4\bar{y}_5^4 - 10\bar{x}_5^4\bar{y}_5^2 - 24\bar{x}_5^3\bar{y}_5^3 - 10\bar{x}_5^2\bar{y}_5^4 + 5\bar{x}_5^4 + 40\bar{x}_5^3\bar{y}_5 + 76\bar{x}_5^2\bar{y}_5^2 \\ & + 40\bar{x}_5\bar{y}_5^3 + 5\bar{y}_5^4 - 10\bar{x}_5^2 - 24\bar{x}_5\bar{y}_5 - 10\bar{y}_5^2 + 1. \end{aligned}$$

So $\bar{g}(\bar{x}_5, \bar{y}_5) = 0$ implies $a + b = \pi$; in this case ΔABC degenerates to a line.

It is easy to check that both $g_{10}(\bar{x}_5)$ and $g_{13}(\bar{x}_5)$ have no real root in the interval $(0, 1)$. From the equalities $\tan\left(\frac{\pi}{15}\right) = \frac{2\sqrt{3} - \sqrt{10 - 2\sqrt{5}}}{3 + \sqrt{5}} = \frac{2\tan\left(\frac{\pi}{30}\right)}{1 - \tan^2\left(\frac{\pi}{30}\right)}$ and $\tan\left(\frac{2\pi}{15}\right) = \frac{2\tan\left(\frac{\pi}{15}\right)}{1 - \tan^2\left(\frac{\pi}{15}\right)}$, we obtain $\tan\left(\frac{\pi}{30}\right) = \frac{\sqrt{10 - 2\sqrt{5}}}{2} + \frac{\sqrt{3} - \sqrt{15}}{2}$ and $\tan\left(\frac{2\pi}{15}\right) = \sqrt{10 - 2\sqrt{5}}\left(1 + \frac{\sqrt{5}}{2}\right) - \frac{3\sqrt{3} + \sqrt{15}}{2}$. Note that $g_{11}(\tan\left(\frac{\pi}{30}\right)) = 0$ and $g_{12}(\tan\left(\frac{2\pi}{15}\right)) = 0$, so $G_1 = 0$ has two real roots $\bar{x}_5^{(1)} = \tan\left(\frac{\pi}{30}\right)$ and $\bar{x}_5^{(2)} = \tan\left(\frac{2\pi}{15}\right)$ in $(0, \sqrt{5 - 2\sqrt{5}})$. In what follows, we show that only the root $\bar{x}_5^{(2)} = \tan\left(\frac{2\pi}{15}\right)$ meets the requirements.

The Gröbner basis of the polynomial set $\left[g_{11}(\bar{x}_5), g_{20}(\bar{x}_5, \bar{y}_5)\right]$ with respect to the lex term ordering determined by $\bar{y}_5 \succ \bar{x}_5$ is $\left[-g_{1,1}(\bar{x}_5), \bar{y}_5 - \bar{x}_5\right]$. Therefore $(\bar{x}_5^{(1)}, \bar{y}_5^{(1)}) = (\tan\left(\frac{\pi}{30}\right), \tan\left(\frac{\pi}{30}\right))$, which implies $a = b = \frac{5\pi}{6}$. This solution does not meet the requirement that $a, b \in (0, \pi)$. Similarly, the Gröbner basis of $\left[g_{12}(\bar{x}_5), g_{30}(\bar{x}_5, \bar{y}_5)\right]$ with respect to the lex term ordering determined by $\bar{y}_5 \succ \bar{x}_5$ is $\left[g_{12}(\bar{x}_5), (-\bar{y}_5 + \bar{x}_5)(\sqrt{3} - \bar{y}_5)\right]$. Therefore $(\bar{x}_5^{(2)}, \bar{y}_5^{(2)}) = (\tan\left(\frac{2\pi}{15}\right), \tan\left(\frac{2\pi}{15}\right))$, which implies $a = b = \frac{\pi}{3}$. This proved that $\Delta\bar{A}_5\bar{B}_5\bar{C}_5$ is equilateral if and only if ΔABC is equilateral.

5 Conclusions

We have presented a general approach using algebraic and computational methods for the study of relationships among intersection points of mth angular n-sectors of an arbitrary triangle and shown that this approach can be used to prove non-existence of Morley triangles for $n = 4$ and 5 with $m = 1$. Similar results may also be established for other specific values of n and m by using the same approach with more involved computations and reasoning. Advanced techniques from complex analysis will be studied and integrated into the current approach to prove or disprove the non-existence of Morley triangles for general n and m.

References

1. B. Xia, L. Yang, *Automated Inequality Proving and Discovering*, World Scienfic, Singapore, 2016.
2. C. O. Oakley, J. C. Baker, *The Morley tridector theorem*, Am. Math. Mon., **85** (1978), 737–745.
3. D.J. Kleven, *Morley's theorem and a converse*, Am. Math. Mon., **85** (1978), 100–105.
4. D. Wang, *Elimination procedures for mechanical theorem proving in geometry*, Ann. Math. Artif. Intell., **13** (1995), 1–24.
5. D. Wang, *Geometry machines: From AI to SMC*, J. Calmet, J. A. Campbell, J. Pfalzgraf (eds.): Artificial Intelligence and Symbolic Computation, Springer, Berlin Heidelberg, 1996, pp. 213–239 (Lecture Notes in Computer Science, vol. 1138).
6. D. Wang, *Elimination Methods*, Springer Verlag, Wien New York, 2001.
7. D. Wang, *Elimination Practice: Software Tools and Applications*, Imperial College Press, London, 2004.
8. F. Morley, *On the metric geometry of the plane n-line*, Trans. Am. Math. Soc., **1** (1900), 97–115.
9. F. Morley, *On the intersections of the trisectors of the angles of a triangle*, J. Math. Assoc. Japan, **6** (1924), 260–262.
10. H. Li, *Vectorial equations solving for mechanical geometry theorem proving*, J. Automat. Reason., **25** (2000), 83–121.
11. S.-C. Chou, *Mechanical Geometry Theorem Proving*, Reidel, Dordrecht, 1988.
12. S.-C. Chou, X.-S. Gao, J.-Z. Zhang, *Machine Proofs in Geometry*, World Scienfic, Singapore, 1994.
13. T. Ida, A. Kasem, F. Ghourabi, H. Takahashi, *Morley's theorem revisited: Origami construction and automated proof*, J. Symb. Comput., **46** (2011), 571–583.
14. W.-T. Wu, *On the decision problem and the mechanization of theorem-proving in elementary geometry*, Sci. Sinica, **21** (1978), 159–172.
15. W.-T. Wu, *Basic principles of mechanical theorem proving in elementary geometries*, J. Syst. Sci. Math. Sci., **4** (1984), 207–235; J. Automat. Reason., **2** (1986) 221–252.
16. W.-T. Wu, *Mechanical Theorem Proving in Geometries: Basic Principles*, Translated from the Chinese by X. Jin and D. Wang. Springer, Wien New York, 1994.
17. W.-T. Wu, *Mathematics Mechanization*, Science Press/Kluwer Academic, Beijing, 2000.
18. Y.-C. Wong, K.-M. Tsang, *A strong converse of Morley's trisector theorem*, Am. Math. Mon., **89** (1982), 642–653.

ADG 2018

September 11-14, 2018
Nanning, China

Organized by
Scientist Mobility Station | SMS International
Guangxi University for Nationalities
Guangxi Key Laboratory of Hybrid Computation
and IC Design Analysis