Abstract

In the current Information Society the organisation of the information is key to ensure the information safekeeping and retrieval. It is of utmost importance that each and every user can find the information he/she is looking for, presented in such a way that best fit his/her needs. Geometry is no exception, the servers of geometric information should be easily and successfully searchable. By classifying the information contained in the servers of geometric information accordingly to several taxonomies, it will be possible to begin applying filters to the users’ queries, adjusting them to the perceived user’s needs. Having that in mind, the introduction of an adaptive filtering mechanisms into servers of geometric information is considered.

Different taxonomies for different goals are presented. For educational purposes, a classification like Common Core Standards should be considered, but other considerations like the complexity of the construction, the provability, by a geometry automatic theorem prover, of a given conjecture and the readability of the resulting proof, should be taken into account. For research in automated deduction purposes, other issues must be considered, e.g. efficiency and applicability of the available automated provers.

To validate the usefulness of these taxonomies it will be used, as a case study, their application to a server of geometric information. In particular, Thousands of Geometric problems for geometric Theorem Provers will be considered. TGTP is a Web-based repository of geometric problems being developed to support the testing and evaluation of geometric automated theorem proving systems. Using this system it will be analyse how the taxonomies could help to tailor the search for information adapted to each and every geometer.
1. Introduction

In the area of geometry there are now a large number of computational tools that can be used to perform many different tasks, dynamic geometry systems (DGS), computer algebra systems (CAS), geometry automatic theorem provers (GATP), among others (Quaresma, 2017). All these tools are clients of geometric information—information that can be found on repositories of geometric knowledge such as: Intergeo, TGTP, GeoGebra Materials, among others.

It can be claimed that the usefulness of such servers of geometric information, is directly related with the possibility of an easy retrieval of the information a given user is looking for. Therefore the information should be organised in such a way that it will be possible the design of filters adjusted to the user’s preferences.

The organisation of information through the taxonomy concept allows to allocate, retrieve and communicate information within a system in a logical way, that is, in classes, subclasses, sub-subclasses, and so on. Each of these levels aggregate information about the existing documents in the repository. An advantage of this form of access is the user’s guarantee of best selection of searched term, since the classes contain mutually exclusive topics.

Different taxonomies would answer to different users’ needs. The problems in the servers must be classified in such a way that, in response to a client query, only the problems in the user’s level and/or interest and/or language are returned.

If the organisation of a large field of knowledge like mathematics are to be considered, taxonomies like Mathematics Subject Classification (MSC) can be found. In this paper the much narrower scope of constructive geometry is addressed, i.e. geometric constructions made by dynamic geometry systems and geometric problems, eventually with an associated construction, manipulated by geometric automated theorem provers.

The term “geometric problem” is here used in a general way, it is used in relation to the case study, i.e. taxonomies for TGTP. There was and there is an important debate on the distinction between problems and theorems, some scholars consider that they should be distinguished, others consider that problems can be reduced to theorems, others consider that theorems can be reduced to problems. This debate is very old and continues today: see for example Morrow (1992); Heath (1908); Martin (1998); Hartshorne (2000); Sidoli (2018). However, this paper is a case study about taxonomies for TGTP, therefore theorems, their proofs, construction problems and their solutions are considered from a single and uniform taxonomic point of view.

For researchers in geometric automated theorem proving it will be interesting to look for: conjectures not yet proved by GATPs; theorems with readable proofs; theorems proved efficiently; etc. (Chou et al., 1996a,b); Janičić et al., 2012; Jiang and Zhang, 2012; Stojanović et al., 2011; Wang and Su, 2015).
If a taxonomy based on those criteria would please the researchers in automatic geometric reasoning, it might not be completely suitable for the educational community. When designing filters for educational purposes, education levels (International Standard Classification of Education [UNESCO, 2012]), levels of geometry reasoning [Usiskin, 1982] and also personal preferences must be considered. Adding to other approaches, a new approach to geometrography [Lemoine, 1902; Mackay, 1893; Pinheiro, 1974], is considered, taking into account a very interesting point of view on geometrical constructions classification. Applying geometrography’s principles to the dynamic geometry systems, it is possible to (re)define the concepts of coefficient of simplicity and a new coefficient of freedom to measure the complexity and dynamics of a DGS construction.

With reference to the paper’s case study, Thousands of Geometric problems for geometric Theorem Provers (TGTP) is a Web-based repository of geometric problems with integrated GATPs. It is being developed to support the testing and evaluation of geometric automated theorem proving systems [Quaresma, 2011]. The list of problems in TGTP can be explored with some powerful textual and geometric search mechanisms [Haralambous and Quaresma, 2018], but, if adapting the search to each user’s needs is pursued, it is necessary to introduce a classification for each TGTP problem, stating their characteristics, in face of one or more intended users’ expectations.

Originally TGTP was aimed to the geometric automatic theorem provers community, as said above to support the testing and evaluation of geometric automated theorem proving systems, so its expected audience is mostly researchers whose background is mathematics and/or computer science and whose research focus is automatic reasoning, formalisation of mathematics, artificial intelligence, among others. The interest in proofs and proving in mathematics education and the application of GATPs in mathematics education [Janičić and Quaresma, 2007; Hanna and de Villiers, 2012; Quaresma and Santos, 2016] opens a new community of potential users of TGTP [Quaresma et al., 2018a].

The classification of each TGTP problem at a given educational level, should be possible and not very difficult. The classification accordingly to a level of geometry reasoning would be more difficult. Nevertheless such classifications are useful in any educational environment, e.g. when linking with educational platforms like the Web Geometry Laboratory (WGL) [Quaresma et al., 2018b; Santos et al., 2018].

The current search mechanisms in TGTP allow its users to search for a given specific problem, or set of problems, e.g. look for Ceva’s theorem, look for all problems with the word “circumcircle” in its description, look for problems containing some given geometric configuration [Quaresma, 2011; Haralambous and Quaresma, 2018]. The introduction of taxonomies in TGTP can add a filtering step that, together with the text and geometric search mechanisms, will allow to tailor TGTP’s usefulness to each user’s needs. For example, a secondary school teacher preparing a class about circumcenter centre would choose filters: CCS classes, CO.A.1 and C.A.3; construction complexity, simple; proofs in education, verification: good, filtering the TGTP database, or any other geometry knowledge repository, in such a way that a good set of examples could be browsed and choose as teaching materials.

An interesting development of this research is the application of the taxonomies presented in this paper, to proof assistant area. We do not consider this application here, however we give some suggestions about it (see Section 2.2).

Overview of the paper. The paper is organised as follows: first, in §2 taxonomies for GATP research and taxonomies for education will be discussed. In §3 the application of taxonomies to
TGTP will be analysed providing some examples, also a new approach to Geometrography will be presented. In §4 conclusions are drawn and future work will be discussed.

2. Taxonomies

Taxonomies are a crucial component for any application of information retrieval, natural language processing and knowledge management. The design of an hierarchical faceted metadata system, allow users to browse the information accordingly to multiple categories simultaneously (Hearst, 2009). Designing such a faceted system to geometry will allow geometers to browse through the repositories of geometric knowledge in a more fulfilling way. For example an automatic deduction researcher will look for provability, for the methods used, for the readability of the proof produced and for efficiency measures. An educator will look for a given class of geometric problems, for the construction complexity and for the validation of geometric conjectures, among others.

Mathematics Subject Classification. Using the Mathematics Subject Classification (MSC) the following classes can be considered:

Geometry  generic classes

51M05 Euclidean geometries (general) and generalisations
51N10 Affine analytic geometry
51N15 Projective analytic geometry
51N20 Euclidean analytic geometry
51N25 Analytic geometry with other transformation groups
68U05 Computer graphics; computational geometry
70G55 Algebraic geometry methods

GATP geometric automated theorem proving

03B35 Mechanisation of proofs and logical operations
68T15 Theorem proving (deduction, resolution, etc.)
70G55 Algebraic geometry methods
94B27 Geometric methods (including applications of algebraic geometry)
97G70 Analytic geometry. Vector algebra
97E50 Reasoning and proving in the mathematics classroom

Education generic and deductive reasoning in education

97-XX Mathematics education
97E50 Reasoning and proving in the mathematics classroom
97Uxx Educational material and media, educational technology
68Q32 Computational learning theory
68T05 Learning and adaptive systems
2.1. Taxonomies for GATP Research

For more than half a century, mechanical theorem proving in geometry is an active research topic (Chou et al., 1994; Chou and Gao, 2001; Jiang and Zhang, 2012). This research has led to three types of GATPs: synthetic, semi-synthetic and algebraic. Synthetic GATPs have proven to be impractical for non-trivial geometry theorems (Gelernter, 1995; Coelho and Pereira, 1986). As for the semi-synthetic ones, such as the area method and the full-angle method developed by Chou, Gao and Zhang, they are capable of proving many theorems, providing readable proofs (Chou et al., 1996a,b; Janičić et al., 2012). Algebraic methods, such as the Wu’s method and the Gröbner method, have been used to prove many theorems, some of them non-trivial, within seconds, but generate very hard to read proofs (Chou et al., 1994).

It is then clear that producing fast, short and readable proofs remains a challenge to researchers in the field of automated theorem proving (Boutry et al., 2014; Chou et al., 1994, 2000; Jiang and Zhang, 2012; Pak, 2015; Stojanović et al., 2011; Wang and Su, 2015). As such, a taxonomy for GATP research must deal with the following classes: GATP provability, readability and efficiency.

**GATP Provability.** Given a conjecture and \( n \) GATPs, if the conjecture is successfully proved by \( p \) of those GATPs, the ratio \( p/n \) provides a simple method to measure this criteria.

Apart this numerical criteria one might consider a classification of the geometric conjectures accordingly to the different methods. If synthetic methods are to be considered, relying in heuristics, most of the implementations are only able to solve a certain kind of problems. Also the algebraic methods and the semi-synthetic methods have their limitations, defining classes of problems they can solve (Chou and Gao, 2001).

**Readability.** As far as the authors of this paper know, there are two proposals to measure the readability of a proof.

Chou et al. (1994, p.442) proposed a way to measure how difficult a formal proof is (using the area method). To measure this criteria the triple \((time, maxt, lems)\) is used, where:

- **time** is the time needed to complete the proof;
- **maxt** is the number of terms of the maximal polynomial occurring in the proof;
- **lems** is the number of elimination lemmas used to eliminate points from geometry quantities. In other words, **lems** is the number of deduction steps in the proof.

Using this criteria and analysing all the proofs done by the GATP they have implemented, they reach to the following thresholds of proof readability.

---

6 The polynomial of highest degree, as clear by the examples presented in the paper by Chou et al. 1994.
7 Accordingly to their data: 66.9% of the proofs has maxt ≤ 5, 42.6% has lems ≤ 10 and 73.2% has lems ≤ 20.
• According to Chou et al. (1994, p.442) a formal proof, done using the area method, is considered readable if one of the following conditions holds:
  – the maximal term in the proof is less than or equal to 5;
  – the number of deduction steps of the proof is less than or equal to 10;
  – the maximal term in the proof is less than or equal to 10 and the deduction step is less than or equal to 20.

• The de Bruijn factor (de Bruijn, 1994; Wiedijk, 2000), the quotient of the size of corresponding informal proof and the size of the formal proof, could also be used as a measure of readability. Using this quotient a proof can be considered readable if the value is less than or equal to 2 (the formal proof is at most twice as larger than a given informal proof).

In both cases it is somehow assumed, readability by experts, i.e. a given geometer, expert in the language of the prover that produced the proof. If non-experts, from the geometry not familiarised with the prover, to the complete novice on the area of the automatic deduction, are to be considered the need of synthetic proofs with natural language descriptions is felt. There are some recent approaches to this, there are even visual approaches, where the formal proof can also followed by a correspondent visual manipulation (Stojanović et al., 2011; Ye et al., 2010a,b).

The question here is how to classify the problems against this new criteria. The following schema may be considered:

1. no readable proof;
2. non-synthetic proof (i.e. a proof without a correspondent geometric description, e.g. algebraic methods);
3. semi-synthetic proof with a corresponding prover’s language rendering;
4. (semi-)synthetic proof with a corresponding natural language rendering;
5. (semi-)synthetic proof with a corresponding natural language and visual rendering;

The Chou’s conditions allows to define a threshold for area method semi-synthetic proofs. Along this lines it is possible to define coefficients of readability for other semi-synthetic methods (levels 3, 4 and 5). The de Bruijn factor can be used in all levels, but it will be more meaningful on levels above 4.

Efficiency. The amount of time needed to complete a proof should be considered, by itself, a way to measure the difficulty of a formal proof.

The space complexity is less important, because the physical constraints are nowadays less important and because the time spent waiting for the proof, or for the validation of a geometric property, is the important factor for the user.

2.2. Taxonomies in Education

Defining taxonomies in an education setting depends on many factors, the type of geometry being considered, in secondary schools most of the times will be Euclidean geometry, studied in a synthetic way (without coordinates) and analytically (with coordinates), the axiomatic system used, the type of computational tools being used, the educational level and also the development of the geometric ideas of the students depending on their behaviour, their level of geometric knowledge and personal preferences.
An interesting taxonomy to be considered in education is the USA initiative, *Common Core Standards* (CCS) (McCallum 2015; Rivera 2013; Wu 2012). According to it, the classes of geometry are: congruence (CO); similarity, right triangles, and trigonometry (SRT); circles (C); expressing geometric properties with equations (GPE); geometric measurement and dimension (GMD) and modelling with geometry (MG). Each class has its own subclasses, defined by degrees of difficulty, for example, in the class of congruence, it has the subclass, *lower degree of difficulty*, with the sub-subclasses: experiment with transformations in the plane; understand congruence in terms of rigid motions; prove geometric theorems, and the subclass, *higher degree of difficulty*, with the sub-subclass: make geometric constructions.

The major classes (abbreviations, in capital letters) are divided in subclasses with two levels, the first one designated by a capital letter and the second by a number. Below the different class/subclass/sub-subclass used to classify the examples presented in section 3.1.

**C.A.3** Circles / Understand and apply theorems about circles / Construct the inscribed and circumscribed circles of a triangle, and prove properties of angles for a quadrilateral inscribed in a circle.

**CO.A.1** Congruence / Experiment with transformations in the plane / Know precise definitions of angle, circle, perpendicular line, parallel line, and line segment, based on the undefined notions of point, line, distance along a line, and distance around a circular arc.

**CO.A.4** Congruence / Experiment with transformations in the plane / Develop definitions of rotations, reflections, and translations in terms of angles, circles, perpendicular lines, parallel lines, and line segments.

**CO.C.10** Congruence / Prove geometric theorems / Prove theorems about triangles. Theorems include: measures of interior angles of a triangle sum to 180°; base angles of isosceles triangles are congruent; the segment joining midpoints of two sides of a triangle is parallel to the third side and half the length; the medians of a triangle meet at a point.

**CO.D.12** Congruence / Make geometric constructions / Make formal geometric constructions with a variety of tools and methods (compass and straight edge, string, reflective devices, paper folding, dynamic geometric software, etc.). Copying a segment; copying an angle; bisecting a segment; bisecting an angle; constructing perpendicular lines, including the perpendicular bisector of a line segment; and constructing a line parallel to a given line through a point not on the line.

**SRT.B.4** Similarity, right triangles, and trigonometry / Prove theorems involving similarity / Prove theorems about triangles. Theorems include: a line parallel to one side of a triangle divides the other two proportionally, and conversely; the Pythagorean Theorem proved using triangle similarity.

**SRT.B.5** Similarity, right triangles, and trigonometry / Prove theorems involving similarity / Use congruence and similarity criteria for triangles to solve problems and to prove relationships in geometric figures.

---

[http://www.corestandards.org/Math/] The CCS was launched in 2009 and it is an effort of the National Governors Association and Council of Chief State School Officers.

[http://www.corestandards.org/Math/Content/HSG/]
**Geometrography.** Another approach that can be useful in this context is given by the geometrography approach (Lemoine [1902]; Mackay [1893]; Pinheiro [1974]). Geometrography, “alias the art of geometric constructions,” aims at providing a tool: to designate every geometric construction by a symbol that manifests its simplicity and exactitude; to teach the simplest way to execute an assigned construction; to discuss a known solution to a problem and eventually replacing it with a better solution; to compare different solutions for a problem, by deciding which is the most exact and the simplest solution from the point of view of geometrography (Loria [1908]). In this perspective, geometrography offers a very interesting point of view on geometrical problems and their taxonomies in education. In Lemoine’s geometrography two coefficients are defined to measure the relative difficulty to perform some geometric constructions. The approach is applied to ruler and compass geometry, i.e. geometric constructions made with the help of a ruler and a compass only. Considering the modifications proposed by Mackay [1893] the following Ruler and Compass constructions \( R \) and \( C \) and the corresponding coefficients can be considered.

1. Place the edge of the ruler in coincidence with one point .......................... \( R_1 \)
2. Place the edge of the ruler in coincidence with two points .......................... \( 2R_1 \)
3. To draw a straight line .............................................................................. \( R_2 \)
4. To put one point of the compasses on a determinate point ...................... \( C_1 \)
5. To put one point of the compasses on two determinate points ................. \( 2C_1 \)
6. To describe a circle .................................................................................. \( C_2 \)

Then a given construction is measured against the number of those elementary steps. For example, for the construction of a triangle, given its three vertices \( A, B \) and \( C \), Mackay [1893] estimate \( 4R_1 + 3R_2 \): to put the ruler in contact with \( A \) and \( B \) is \( 2R_1 \); to draw \( AB \) is \( R_2 \); with the ruler in contact with \( B \) to put it also in contact with \( C \) is \( R_1 \); to draw \( BC \) is \( R_2 \); repeat that for \( C \) and \( A \) is \( R_1 \) and finally to draw \( CA \) is \( R_2 \). In all, \( 4R_1 + 3R_2 \).

For a given construction expressed by the equation:

\[
l_1R_1 + l_2R_2 + m_1C_1 + m_2C_2
\]

where \( l_i \) and \( m_j \) are coefficients denoting the number of times any particular operation is performed. The number \( l_1 + l_2 + m_1 + m_2 \) is called the *coefficient of simplicity* of the construction, it denotes the total number of operations. The number \( l_1 + m_1 \) is called the *coefficient of exactitude* of the construction, it denotes the number of preparatory operations on which the exactitude of the construction depends (Mackay [1893]; Merikoski and Tossavainen [2010]).

Some variants of Lemoine’s geometrography can be defined, e.g. by adding rules for other idealised tools/operations (e.g. carpenter’s square, graduated rulers, etc.), or by adding a value for the change of the instrument/operation, or by considering different values for different operations (Grüttner [1912]; Loria [1908]).

Extrapolating (modernising) geometrography, considering the “tools” of dynamic geometry systems, the *coefficient of exactitude* lose its meaning, the construction will be executed by

---

1 Lemoine [1902] considers the following basic operations: L1. Place the ruler through a given point; L2. draw a line; C1. place one leg of the compass on a given point; C2. place one leg of the compass on an indeterminate point of a given line; C3. draw a circle.
the DGS, so exact (minus floating point representation considerations), but the coefficient of simplicity of the constructions can still be useful, it can be used to classify the constructions by levels of simplicity and, in this way, adding another level of meta-information. Also a new dimension can be added, the coefficient of freedom, given by the degree of freedom a given geometric object has, e.g. “a point in a line” has one degree of freedom, a point in the plane has two degrees of freedom. This new coefficient will give a value for the dynamism of the geometric construction.

The need of a classification for the complexity of construction is felt in educational settings. For example in the Euclidea11 and Geometriagon12 platforms the constructions contained in the respective repositories are classified by levels of difficulty.

Euclidea count actions with tools (L), e.g. build a line, a perpendicular line, etc. and also count movements (E), such as a construction was made using a ruler and a compass. Defining costs, (L, E), for each tool. Geometriagon defines the difficulty level, 1 ≤ n ≤ 5, awarded by the proposer and other users. Both system would benefit from the adoption of a geometrography approach.

In the following the geometrography approach to the classification of the geometric construction made using the Geometry Constructions LaTeX Converter (GCLC)13 (Janičić, 2006) is presented (GCLC provides three of the four GATPs embedded in TGTP).

Geometrography in GCLC. Apart the rendering of geometric constructions GCLC also incorporate GATPs based on the area method, Wu’s method and Gröbner Basis method, all those tools use the GCL formal language (Janičić, 2010).

Considering the operations: define a point, anywhere in the plane, D and define a given object, using other objects, C, the following values for the GCLC basic constructions are obtained:

- point – fix a point in the plane ...................................................... D
- line – uses two points ................................................................. 2C
- circle – uses two points .............................................................. 2C
- intersec – uses two lines ............................................................... 2C
- intersec – uses four points ......................................................... 4C
- intersec2 – uses a circle and a circle or line .................................. 2C
- midpoint – uses two points ........................................................... 2C
- med – uses two points ................................................................. 2C
- bis – uses three points ............................................................... 3C

---

11 https://www.euclidea.xyz/ also available as an application for smartphones.
12 http://polarprof-001-site1.htempurl.com/geometriagon/
13 http://poincare.matf.bg.ac.rs/~janicic/gclc/
The degrees of freedom are measured against the point definitions. The point definition, define a point with two degrees of freedom, the `onsegment`, `online` and `oncircle` constructions, define points with one degree of freedom.

A script that analyse any GCLC construction was implemented, giving its coefficients of simplicity (cs) and freedom (cf).

This research also open an interesting application, a geometrography approach regarding proofs and proof assistants, considering for example a combination of dynamic geometry software with proof assistants. An interesting case study could be, for example, GeoCoq (Pham and Bertot, 2012), a combination of GeoGebra and Coq. This will be object of future research.

**GATPs in Education.** There are complex problems, which a secondary school student will find difficult to prove, that can be automatically proved. For example, the Geometry Expert program (Chou et al., 1994), proved automatically about 400 problems extracted from a typical geometry secondary school book (Altshiller-Court, 2007). Chou et al. (1994, p.453) also proposed a classification of level of difficulty of geometrical theorems in terms of various types of geometric constructions involved: collinearity; parallelism; proportionality; perpendicularity; circle; angle, in ascending order of difficulty.

In Proceedings of the ICMI Study 19 Conference: Proof and Proving in Mathematics Education (Lin et al., 2009a,b) many articles exploring the use of proofs in a learning environment can be found. In (Hanna, 2000; Hanna and Sidoli, 2007; de Villiers, 1990) Michael de Villiers and Gila Hanna gave a list of the usefulness of proofs and proving in a learning environment. For the purpose of problems classification the following two points emerge:

- verification (concerned with the truth of a statement);
- explanation (providing insight into why it is true);

Considering the formal validation of properties of geometric constructions the time needed to get an answer should be used as a classification criteria.

Many of the current DGSs have already the capability of a formal validation of properties: Cinderella (Richter-Gebert and Kortenkamp, 1999) contains a randomised theorem checker; GCLC, Java Geometry Expert (JGEX) and GeoGebra (version 5) (Hohenwarter, 2002) have

---

2. [https://cinderella.de](https://cinderella.de)
4. [https://www.geogebra.org/](https://www.geogebra.org/)
a number of automated theorem provers incorporated in them, thus opening the possibility to give a formal answer to a given validation question (Botana et al., 2015; Janičić and Quaresma, 2007; Ye et al., 2011).

Extrapolating from the concept of “wait-time”—periods of silence that followed teacher questions and students’ completed responses (Rowe, 1972)—or more precisely the category of periods of silence “Post-Teacher Question Wait-Time” (Stahl, 1994), the following classes of “GATP validation time” could be defined, in terms of time, $t$, taken by the GATP to answer:

- good: $t \leq 1.5$ s;
- fair: $1.5$ s $< t \leq 3$ s;
- poor: $t > 3$ s.

The use of formal proofs in an educational setting is a more complex issue. As stated in (Hanna and de Villiers, 2012; Quaresma and Santos, 2016), formal statements have an important role in the area of education. GATPs implementing the Wu’s method, the Gröbner bases method, the area method or full-angle method, are capable of producing proofs and are very efficient, but the proofs produced are far from the level of non-expert readable proofs (Botana et al., 2015).

For a classification propose the criteria of readability (of the proof produced), efficiency, interaction between the GATP and its users, student-level, must be considered. This is an open and active research area. The following classes could be considered:

- yes: a small (in length) and fast (in processing time), readable synthetic GATP proof exist;
- maybe: a small (in length) and fast (in processing time), non-synthetic (e.g. semi-synthetic) readable GATP proof exist;
- no: no GATP proof exist or the GATP proof is neither small nor fast nor readable.

3. Taxonomies in TGTP

Applying the above defined taxonomies to the TGTP server of geometric knowledge poses a few technical problems that should not be difficult to solve. Indeed TGTP, in its current version, already provides, for each conjecture, the ratio of successful proofs to proof’s attempts, as well as other items of information needed to classify a given conjecture. TGTP also provides a proof of a conjecture if one is provided by the method’s implementation, a notable exception is the implementation of the area method using the Coq proof assistant (Janičić et al., 2012; Narboux, 2007), which only provides a proved or not proved output.

For the GCLC constructions contained in TGTP an average value of simplicity ($cs$) of 22.5 was obtained. Having that in mind, three classes of geometric constructions for an increasing level of complexity were defined: simple constructions, $1 \leq cs < 15$; average complexity constructions, $15 \leq cs < 30$; complex constructions, $cs \geq 30$.

Note that the “wait-time” concept will be used to define the threshold values for the response times of GATPs, it is not any new kind of taxonomy.

http://dpt-info.u-strasbg.fr/~narboux/area_method.html
TGTP contains 36 simple constructions; 119 average complexity constructions; 30 complex constructions.

If applied to TGTP, the educational taxonomies should also take into account the objectives of teaching, a short response time and readable demonstrations.

A first step would be to make the necessary modifications to TGTP to be able to add the missing items of information: modifying the forms used to manually manipulate the information about the geometric problems; implementing scripts capable of manipulating the information already contained in TGTP, adding, automatically, those needed items of information. A final step would be the use of all that information into filters that, by manual selection or automatic processing of users’ needs, could be used to improve the users’ queries.

3.1. TGTP Examples

In the following, some selected examples are presented. For each example, its classification, accordingly to the taxonomies defined above, is shown.

Below, the GATP provability ratio is given by the number of proofs successfully concluded over the number of attempts.

The complexity is given by the GCLC geometrography complexity coefficient.

**GEO0316—Nine Points Circle** Prove that in any triangle midpoints of each side, feet of each altitude and midpoints of the segments of each altitude from its vertex to the orthocenter lie on a circle (Chou, 1988).

MSC: 51M05, 70G55, 94B27.

GATP Provability: 1/3.

**Semi-synthetic methods:** 0/1: GCLC area method, “The conjecture is out of scope of the prover”.

**Algebraic methods:** 1/2: GCLC Wu’s method, “The conjecture successfully proved”;

GCLC Gröbner basis method, “The conjecture not proved - timeout”.

**Readability [Chou et al., 1994]:** non-synthetic proof

Wu’s Method, 16 pages long proof.

**Readability [de Bruijn, 1994]:** no readable proof.

de Bruijn factor: 16/6.

See section [Appendix A.1] for the informal proof. The formal proof is not included given its size, it can be consulted in TGTP.
Efficiency (CPU time): 0.17s.

GCLC Wu’s Method: 0.17s.

CCS: C.A.3; CO.A.1; CO.C.10; CO.D.12.

Construction Complexity: complex (cs=41).

Coefficient of simplicity: \(3 \times D + 3 \times 2C + 3 \times 2C + 3 \times 2C + 2 \times 2C + 2C + 3 \times 2C + 2 \times 2C + 2C + 2C = 41\).

Coefficient of freedom: \(3 \times 2 = 6\).

Proofs in Education:

Verification: good

GCLC Wu’s Method: 0.17s.

Explanation: no

only an algebraic, long (16 pages) GATP proof, exist.

GEO0001—Ceva’s Theorem  Let \(\triangle ABC\) be a triangle and \(P\) be any point in the plane. Let \(D = AP \cap CB\), \(E = BP \cap AC\), and \(F = CP \cap AB\). Show that:

\[
\frac{AF}{FB} \times \frac{BD}{DC} \times \frac{CE}{EA} = 1.
\]

\(P\) should not be in the lines parallels to \(AC\), \(AB\) and \(BC\) and passing through \(B\), \(C\) and \(A\) respectively (Zhang et al., 1995).

MSC: 51M05, 70G55, 94B27, 97G70, 03B35.


Readability (Chou et al., 1994): semi-synthetic proof with a corresponding prover’s language description;

GCLC area method deduction steps = 9 < 10;

GCLC area method, 4 pages long proof; GCLC Wu’s method, 7 pages long proof;

GCLC Gröbner basis method, 8 pages long proof.

Readability (de Bruijn, 1994): semi-synthetic proof with a corresponding prover’s language description;

de Bruijn factor: 1/1;

See section [Appendix A.2] for the informal proof and section [Appendix B.1] for the area method formal proof.

Efficiency (CPU time): 0.04s

GCLC area method: 0.04s; Coq area method: 3.32s; GCLC Wu’s method: 0.33s;

GCLC Gröbner basis method: 0.08s.
CCS: CO.A.4; CO.C.10; CO.D.12; SRT.B.5.

Construction Complexity: average (cs=22).

Coefficient of simplicity: $4 \times D + 6 \times 2C + 3 \times 2C = 22$.

Coefficient of freedom: $4 \times 2 = 8$.

Proofs in Education:

Verification: good

GCLC area method: 0.04s; GCLC Gröbner basis method: 0.08s.

Explanation: maybe

the area method, semi-synthetic, GATP readable proof is 1 page long (4 pages in total if title and proof information pages are to be considered) and it is produced in 0.04s.

GEO0013—Centroid Theorem The three medians of a triangle meet in a point, and each median is trisected by this point (Chou et al. [1996a]).

MSC: 51M05, 70G55, 94B27, 97G70, 03B35.


Readability (Chou et al. [1994]): semi-synthetic proof with a corresponding prover’s language description;

GCLC area method: maximal term $4 < 5$.

GCLC area method, 4 pages long proof; Wu’s method, 7 pages long proof; Gröbner basis method, 12 pages long proof.

Readability (de Bruijn [1994]): semi-synthetic proof with a corresponding prover’s language description.

de Bruijn factor: 2/1.

See section Appendix A.3 for the informal proof and section Appendix B.2 for the area method formal proof.

Efficiency (CPU time): 0.05s

GCLC area method: 0.05s; Coq area method: 3.32s; GCLC Wu’s method: 0.08s; GCLC Gröbner basis method: 0.09s.

CCS: C.A.3; CO.A.1; CO.A.4; CO.C.10; SRT.B.4.

Construction Complexity: average (cs=17).

Coefficient of simplicity: $3 \times D + 3 \times 2C + 2 \times 2C + 2 \times 2C = 17$.

Coefficient of freedom: $3 \times 2 = 6$. 

14
Proofs in Education:

Verification: good

GCLC area method: 0.05s; GCLC Wu’s method: 0.08s.

Explanation: maybe

the area method, semi-synthetic, GATP readable proof is 2 pages long (4 pages in total if title and proof information pages are to be considered) and it is produced in 0.05s.

3.2. TGTP Queries

Querying the TGTP repository is possible in three ways: a simple textual query, a more comprehensive textual search, and a geometric search (Quaresma, 2011; Haralambous and Quaresma, 2018).

The simple textual query is done using MySQL regular expressions queries over the name attribute of the Conjectures table, it will provide the list of conjectures with names containing the query as a sub-string. Another, more powerful, textual query is available, using the full-text search of MySQL. 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 (Haralambous and Quaresma, 2014) a geometric search mechanism is being developed. The queries are constructed using GeoGebra dynamic geometry system and the constructed figure is semantically compared with the figures in the repository (Haralambous and Quaresma, 2018).

The developed taxonomies will allow to apply filters to any given query, e.g. in the Intergeo repository, the “complex text search” filter, already implements an adaptive filtering, alongside the query, the user can choose some filters (e.g. type of activity).

Finally, together with some machine learning mechanisms capable of finding the level of geometric knowledge and/or preferences of TGTP users, the proposed taxonomies will open the possibility of implementing adaptive queries, where the user’s profile will be used to find the filters that best adjust to the users’ needs.

4. Conclusions & Future Work

The definition of taxonomies for geometry will enable to incorporate many useful information in the geometric objects (constructions, conjectures, proofs). All the collected information will constitute a meta-information block that would became part of the geometric objects in the repositories, in a similar way as the meta-information contained in a digital photography.

The next step is the construction of meta-information reports and filters capable of using that information to enhance the user’s queries.

The filtering mechanism will allow improved and adapted user’s queries. It will open the possibility to query for classes of geometric information, e.g. conjectures with semi-synthetic proofs; geometric construction with a coefficient of simplicity lower then 30, etc.

Considering repositories of geometric knowledge, the implementation of a taxonomies aware filtering mechanism, must be preceded by populating the repositories databases with the meta-information needed for the classification of the problems. Using TGTP as an example, adding the information can be seen as a three stages process:

**Manual Stage:** the MSC classification; the CCS classification; the de Bruijn’s factor of readability; availability of proofs for educational purposes. The procedures for inserting or updating geometric information must be changed in such a way that the meta-information can be manually added.

**Automatic Stage:** GATP provability; Chou’s factor of readability; efficiency; GCLC geometrography. The information already contained in the database is enough, it must be, automatically, parsed and transformed in meta-information whenever modifications are being made to the database.

**Machine Learning Stage:** some of the considered taxonomies are not absolute, e.g the readability is strongly dependant of the reader: GATP experts; secondary students; etc. The introduction of users’ profiles build interactively through analysis of users’ interaction with the system, will allow to adapt the filters to each users’ needs. In a first adaptive stage the users modelling can be implemented with simple quantitative or qualitative rule-based approaches, in a later stage, machine learning techniques over users’ interactions with the system should be used to have an automatic and continuous fit of the system to the user’s needs.

The presented taxonomies must be validated, improved if necessary, and other taxonomies must be taken into consideration as well. The filtering mechanism introduced by the taxonomies, the possibility of combining different filters and the combination of this filtering stage with the search mechanism must be implemented in the repositories, or in any system that remotely access the repository, e.g. WGL (Quaresma et al., 2018a).

**References**


Appendix A. Informal Proofs

Appendix A.1. Nine Points Circle Informal Proof

Taken from: [www.ma.utexas.edu/users/shirley/a333l/Handouts/euclidean-geometry-ii/9pt_circle_thm.pdf](http://www.ma.utexas.edu/users/shirley/a333l/Handouts/euclidean-geometry-ii/9pt_circle_thm.pdf), consulted 2017/12/23. Proof done with the help of lemmas to reduce the size of the proof.
The Nine-Point Circle Theorem:

For any triangle the following nine points lie on the same circle:

1) The three midpoints of the sides of the triangle,

2) The feet of the three altitudes of the triangle,

3) The midpoints of the three segments connecting vertices to the orthocenter.

This circle is called the Nine-point Circle of the triangle. Its center is the midpoint of the segment between the orthocenter and the circumcenter and its radius is $\frac{1}{2}$ the radius of the circumcircle.

In the following discussion, the following three theorems are frequently applied:

Theorem 4.2.15, The Midpoint Connection Theorem: If a line segment has as its endpoints the midpoints of two sides of a triangle then the segment is contained in a line that is parallel to the third side and the segment is one-half the length of the third side.

Theorem (NIB) 4.6, The Hypotenuse Diameter Theorem: For any right triangle, the circle which has the hypotenuse as diameter contains the vertex with the right angle.

The Perpendicularity Statement: If a line is perpendicular to one of two parallel lines, then it is perpendicular to the other.

As shown before, in Neutral Geometry the Perpendicularity Statement is equivalent to the Euclidean Parallel Postulate, so the Perpendicularity Statement is true in Euclidean Geometry.

Two Lemmas are used below and their proofs are left as exercises:

Lemma 1: A parallelogram with at least one right angle is a rectangle.

Lemma 2: A rectangle is circumscribed by a circle and each diagonal of the rectangle is a diameter of the circle.

Recall that, in Euclidean Geometry, Parallelism is Transitive (by Theorem 4.2.9):

For given lines $\ell, m$, and $n$, if $\ell \parallel m$ and $m \parallel n$ then $\ell \parallel n$.

Also, by the Perpendicularity Statement, for given lines $\ell, m$, and $t$,

if $\ell \parallel m$ and $t \perp \ell$, then $t \perp m$. 

20
Proof of the Nine-Point Circle Theorem:

Proof: Let \( \triangle ABC \) be a given triangle.

Let \( H \) denote the orthocenter and let \( S \) denote the circumcenter.

Let the following notations be defined:

1) \( M_A, M_B, \) and \( M_C \) are the midpoints of the sides of the triangle opposite vertices \( A, B, \) and \( C, \) respectively.

2) \( F_A, F_B, \) and \( F_C \) are the feet of the altitudes from vertices \( A, B, \) and \( C, \) respectively.

3) \( Q_A, Q_B, \) and \( Q_C \) are the midpoints of the segments between the orthocenter \( H \) of the triangle and the vertices \( A, B, \) and \( C, \) respectively.

\[
\begin{align*}
\overline{Q_AQ_B} & \parallel \overline{AB} \quad \text{by the Midpoint Connection Theorem applied to } \triangle HAB. \\
\overline{M_BM_B} & \parallel \overline{AB} \quad \text{by the Midpoint Connection Theorem applied to } \triangle ABC. \\
\overline{Q_AQ_B} & \parallel \overline{M_BM_B} \quad \text{by the transitivity of parallelism.} \\
\overline{Q_BM_B} & \parallel \overline{CH} \quad \text{by the Midpoint Connection Theorem applied to } \triangle HAC. \\
\overline{Q_BM_B} & \parallel \overline{CH} \quad \text{by the Midpoint Connection Theorem applied to } \triangle ABC. \\
\overline{Q_BM_B} & \parallel \overline{CH} \quad \text{by the transitivity of parallelism.} \quad \therefore \quad Q_A, Q_B, M_A, M_B \quad \text{is a parallelogram.} \\
\overline{CF_C} & \perp \overline{AB} \quad \text{by definition of “altitude”}. \quad \therefore \overline{CF_C} \perp \overline{Q_AQ_B} \quad \text{by the Perpendicularity Statement.} \\
\overline{Q_AQ_B} & \parallel \overline{CF_C} \quad \text{since } C - H - F_C. \quad \therefore \overline{Q_AQ_B} \perp \overline{Q_AQ_B} \quad \text{by the Perpendicularity Statement.} \\
\therefore \angle Q_AQ_BM_A \quad \text{is a right angle.} \quad \therefore \quad \text{By Lemma 1, } \quad Q_A, Q_B, M_A, M_B \quad \text{is a rectangle.}
\end{align*}
\]
Recall that \( \text{Q}_A \text{Q}_B \text{M}_A \text{M}_B \) is a rectangle.

By Lemma 2, there exists a circle \( c_1 \) such that

- \( c_1 \) circumscribes \( \text{Q}_A \text{Q}_B \text{M}_A \text{M}_B \) and \( c_1 = C(\text{diameter} = \text{Q}_A\text{M}_B) \) and \( c_1 = C(\text{diameter} = \text{Q}_B\text{M}_A) \).

By definition of “altitude”, \( \angle \text{Q}_A\text{F}\text{M}_B \) is a right angle.

\[ \therefore \quad \Delta \text{F}_A\text{Q}_A\text{M}_B \text{ is a right triangle with hypotenuse } \text{Q}_A\text{M}_B. \]

\[ \therefore \quad \text{Circle } c_1 \text{ contains } \text{F}_A \text{ by the Hypotenuse Diameter Theorem.} \]

By definition of “altitude”, \( \angle \text{Q}_B\text{F}\text{M}_A \) is a right angle.

\[ \therefore \quad \Delta \text{F}_B\text{Q}_B\text{M}_A \text{ is a right triangle with hypotenuse } \text{Q}_B\text{M}_A. \]

\[ \therefore \quad \text{Circle } c_1 \text{ contains } \text{F}_B \text{ by the Hypotenuse Diameter Theorem.} \]

\[ \therefore \quad \text{Circle } c_1 \text{ contains the six points: } \text{Q}_A, \text{Q}_B, \text{M}_A, \text{M}_B, \text{F}_A, \text{F}_B \]

A similar argument shows that \( \text{Q}_A \text{Q}_C \text{M}_A \text{M}_C \) is a rectangle

Then, a similar argument shows that the circle \( c_1 = C(\text{diameter} = \text{Q}_A\text{M}_C) \) circumscribes \( \text{Q}_A \text{Q}_C \text{M}_A \text{M}_C \) and that circle \( c_1 \) contains the point \( \text{F}_C \).

Note also that all three rectangle diagonals serve as diameters for circle \( c_1 \), that is, \( Q_A M_A, Q_B M_B, \) and \( Q_C M_C \), all three are diameters of circle \( c_1 \).
∴ Circle $c_1$ contains the nine points: $Q_A, Q_B, Q_C, M_A, M_B, M_C, F_A, F_B, F_C$ and it has $Q_A M_A$, $Q_B M_B$, $Q_C M_C$ as diameters.

The circle $c_1$ is called the **Nine-Point Circle**.

Let $N$ be the midpoint of $Q_A M_A$. Then, $N$ is the center of the Nine-Point Circle.

We will call this circle the **Nine-point Circle**.

It remains only to locate the center on the Euler Line and to determine the radius of the Nine-point Circle.
Recall that $S$ denotes the Circumcenter of $\triangle ABC$ and that $N$ denotes the center of the Nine-point Circle.

Let $O$ be the point where Nine-Point Circle diameter $\overline{Q_M M_S}$ intersects $\overline{HS}$.

(A technical argument, which shows that $\overline{Q_M M_S}$ and $\overline{HS}$ must intersect, is omitted here.)

We show that point $O$ and point $N$ are the same point.

It will suffice to show that $O$ is the midpoint $\overline{Q_M M_S}$ since $\overline{Q_M M_S}$ is a diameter of the Nine-Point Circle.

To do this, we show that $\triangle HQO \cong \triangle SMO$:

Now, $\overline{SM}$ lies along the perpendicular bisector of side $\overline{AC}$, so $\overline{SM} \perp \overline{AC}$.

Since $\overline{BH} \perp \overline{AC}$ also, $\overline{SM} \parallel \overline{BH}$.

Since $\overline{SM} \parallel \overline{BH}$, $\angle OQH \equiv \angle OBS$ and $\angle OM \equiv \angle OMB$ are alternate interior angles formed where transversal $\overline{Q_M M_S}$ intersects parallel lines.

∴ By the Converse of the Alternate Interior Angle Theorem, $\angle OQH \equiv \angle OBS$.

Also, $\angle HOQ = \angle SOM$ by the Vertical Angles Theorem.

Since $Q_H$ is, by definition, the midpoint of $\overline{BH}$, $H \overline{O} = \frac{1}{2} (BH)$.

By the Altitude Segment Theorem, Theorem 4.7.1, $BH = 2 (SM)$; so, $SM = \frac{1}{2} (BH)$.

Therefore, $H \overline{O} = SM$ and so, $H \overline{O} = \overline{SM}$.

∴ $\triangle OQ_H \equiv \triangle OMB$ by AAS.

Then, by CPCF, $\overline{HO} = \overline{OM}$, and so, $O$ is the midpoint of $\overline{Q_M M_S}$.

Since $\overline{Q_M M_S}$ is a diameter of the Nine-Point Circle, $O = N$, the center of the Nine-point Circle.

Also, by CPCF, $\overline{HO} = \overline{NS}$, so point $O = N$ is also the midpoint of $\overline{HS}$.

This locates the center $N$ of the Nine-Point Circle at the point midway between the Orthocenter $H$ and the Circumcenter $S$.
Thus, the center N of the Nine-point circle is the midpoint of the segment between the orthocenter H and the circumcenter S.

Therefore, the center of the Nine-Point Circle is on the Euler Line $HS$.

We show that the radius of the Nine-Point Circle is one-half the radius of the Circumcircle.

Segment $SB$ is a radius of the circumcircle (since S is the circumcenter) and segment $NQ_B$ is a radius of the Nine-point circle.

Also, N is the midpoint of $HS$, so $NQ_B$ connects the midpoints of two of the sides of $\Delta SBH$.

$\therefore$ By the Midpoint Connection Theorem applied to $\Delta SBH$, $NQ_B = \frac{1}{2} (SB)$.

Therefore, since $SB$ is the radius of the Circumcircle, the radius of the Nine-point circle is $\frac{1}{2}$ the radius of the Circumcircle.

QED
Appendix A.2. Ceva’s Theorem Informal Proof

Taken from: artofproblemsolving.com/wiki/index.php?title=Ceva’s_Theorem, consulted on 2017/12/18.

Proof

We will use the notation \([ABC]\) to denote the area of a triangle with vertices \(A, B, C\).
First, suppose \(AD, BE, CF\) meet at a point \(X\). We note that triangles \(ABD, ADC\) have the same altitude to line \(BC\), but bases \(BD\) and \(DC\). It follows that \(\frac{BD}{DC} = \frac{[ABD]}{[ADC]}\). The same is true for triangles \(XBD, XDC\), so
\[
\frac{BD}{DC} \cdot \frac{CE}{EA} \cdot \frac{AF}{FB} = \frac{[ABX]}{[AXC]} \cdot \frac{[BCX]}{[BXA]} \cdot \frac{[CAX]}{[CXB]} = 1
\]

Now, suppose \(D, E, F\) satisfy Ceva’s criterion, and suppose \(AD, BE\) intersect at \(X\). Suppose the line \(CX\) intersects line \(AB\) at \(F'\). We have proven that \(F'\) must satisfy Ceva’s criterion. This means that any two medians meet at their point two-thirds of the way from the vertex to the midpoint of the opposite side. So all three do.

Appendix A.3. Centroid Theorem Informal Proof

Taken from: new.math.uiuc.edu/public403/affine/centroid.html, consulted on 2017/12/14.

Let \(G\) be the point where medians \(BB'\) and \(CC'\) of \(\triangle ABC\) intersect. We shall show that \(G\) trisects the two medians in the sense that \(BG : GB' = 2 : 1\) and \(CG : GC' = 2 : 1\). This means that any two medians meet at their point two-thirds of the way from the vertex to the midpoint of the opposite side. So all three do.

Prior to the place in Euclid’s Elements are theorems about similar triangles and about angles made by transversals of two parallel lines. The ones we’ll use here are

**simAAA** Two triangles are similar if and only if their angles are pairwise equal.

**simSAS** Two triangles are similar if and only if two pairs of corresponding sides have the same proportion and the included angles are equal.

**altIA** Two lines are parallel if and only if two alternate interior angles they make with a transversal are equal.

**sameSA** Two lines are parallel if and only if corresponding angles on the same side of a transversal are equal.
Step 1: Apply simSAS to $\Delta AC'B'$ and $\Delta ABC$ by noting that they have $\angle A$ in common, and the adjacent sides are in the ratio of $1 : 2$. So the two triangles are similar. That, by definition of similarity, implies that $\angle B'C'A = \angle CBA$ and $C'B : BC = 1 : 2$.

Step 2: Apply sameSA to the two lines $C'B'$ and $BC$ to the first consequence: $\angle B'C'A = \angle CBA$. Therefore $C'B'||BC$. That in turn, by sameSA, implies that $\angle GC'B' = \angle GCB$ and $\angle C'B'G = \angle CBG$.

Step 3: Apply simAAA to the two triangles $\Delta GB'C'$ and $\Delta GBC$. Two pairs of their angles have already been shown to be equal. The third pair, $\angle B'GC'$ and $\angle BGC$, are equal because they are “opposite angles”. Thus the two triangles are similar.

Step 4: From the second conclusion in Step 1 we know the ratio of to be $2 : 1$. So $1 : 2 = GB : GB'$ and $1 : 2 = GC' : GC$. So $G$ does “trisect” two of the medians, as predicted. □

Appendix B. Formal Proofs

Appendix B.1. Ceva’s Theorem Formal Proof

Proof produced by GCLC area method.

\[
\frac{AF}{FB} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA} = 1 \quad \text{by the statement (0)}
\]

\[
\left(-1 \cdot \frac{AF}{FB} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA}\right) = 1 \quad \text{by geometric simplifications (1)}
\]

\[
\left(-1 \cdot \frac{S_{APC}}{S_{BPC}} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA}\right) = 1 \quad \text{by algebraic simplifications (2)}
\]

\[
\left(-1 \cdot \frac{S_{APC}}{S_{BPC}} \cdot \frac{BD}{DC} \left(-1 \cdot \frac{CE}{AE}\right)\right) = 1 \quad \text{by geometric simplifications (3)}
\]

\[
\frac{S_{APC} \cdot \frac{BD}{DC} \cdot \frac{CE}{EA}}{S_{BPC}} = 1 \quad \text{by algebraic simplifications (4)}
\]

\[
\frac{S_{APC} \cdot \frac{BD}{DC} \cdot \frac{S_{CPB}}{S_{APB}}}{S_{BPC}} = 1 \quad \text{by Lemma 8 (point $F$ eliminated (5))}
\]

\[
\frac{S_{APC} \cdot \left(-1 \cdot \frac{BD}{DC}\right) \cdot \frac{S_{CPB}}{S_{APB}}}{(-1 \cdot S_{CPB})} = 1 \quad \text{by geometric simplifications (7)}
\]
\[
\frac{(S_{APC} \cdot \overrightarrow{BD})}{S_{APB}} = 1 \quad \text{by algebraic simplifications (8)}
\]
\[
\frac{(S_{APC} \cdot \frac{S_{BPA}}{S_{CPA}})}{S_{APB}} = 1 \quad \text{by Lemma 8 (point } D \text{ eliminated) (9)}
\]
\[
\frac{(S_{APC} \cdot \frac{S_{BPA}}{S_{CPA}})}{(-1 \cdot S_{BPA})} = 1 \quad \text{by geometric simplifications (10)}
\]
\[
1 = 1 \quad \text{by algebraic simplifications (11)}
\]

NDG conditions are:
\(S_{BPA} \neq S_{CPA}\) i.e., lines BC and PA are not parallel (construction based assumption)
\(S_{APB} \neq S_{CPB}\) i.e., lines AC and PB are not parallel (construction based assumption)
\(S_{APC} \neq S_{BPC}\) i.e., lines AB and PC are not parallel (construction based assumption)
\(P_{FBE} \neq 0\) i.e., points F and B are not identical (conjecture based assumption)
\(P_{DCD} \neq 0\) i.e., points D and C are not identical (conjecture based assumption)
\(P_{EAE} \neq 0\) i.e., points E and A are not identical (conjecture based assumption)

Appendix B.2. Centroid Theorem Formal Proof
Proof produced by GCLC area method.
\[
\left\{-2 \cdot \frac{GB}{GB}\right\} = 1 \quad \text{by the statement (0)}
\]
\[
\left\{-2 \cdot \left(-1 \cdot \left(-1 \cdot \frac{DG}{BG}\right)\right)\right\} = 1 \quad \text{by geometric simplifications (1)}
\]
\[
\left\{-2 \cdot \frac{DG}{BG}\right\} = 1 \quad \text{by algebraic simplifications (2)}
\]
\[
\left\{-2 \cdot \frac{S_{DEC}}{S_{BEC}}\right\} = 1 \quad \text{by Lemma 8 (point } G \text{ eliminated) (3)}
\]
\[
\left\{-2 \cdot \frac{S_{DEC}}{S_{BEC}}\right\} = 1 \quad \text{by algebraic simplifications (4)}
\]
\[
\frac{(-2 \cdot S_{CDA})}{S_{CBE}} = 1 \quad \text{by geometric simplifications (5)}
\]
\[
\frac{(-2 \cdot S_{CDB} + \left(\frac{1}{2} \cdot (S_{CDB} + (-1 \cdot S_{CDA}))\right))}{S_{CBE}} = 1 \quad \text{by Lemma 29 (point } E \text{ eliminated) (6)}
\]
\[
\frac{((-1 \cdot S_{CDA}) + (-1 \cdot S_{CDB}))}{S_{CBE}} = 1 \quad \text{by algebraic simplifications (7)}
\]
\[
\frac{((-1 \cdot S_{CDA}) + (-1 \cdot S_{CDB}))}{(S_{CBA} + \left(\frac{1}{2} \cdot (S_{CBA} + (-1 \cdot S_{CBA}))\right))}
= 1 \quad \text{by Lemma 29 (point } E \text{ eliminated)} \quad (8)
\]

\[
\frac{((-1 \cdot S_{ACD}) + (-1 \cdot S_{BCD}))}{(S_{CBA} + \left(\frac{1}{2} \cdot (0 + (-1 \cdot S_{CBA}))\right))}
= 1 \quad \text{by geometric simplifications} \quad (9)
\]

\[
\frac{((-1 \cdot S_{ACD}) + (-1 \cdot S_{BCD}))}{\left(\frac{1}{2} \cdot S_{CBA}\right)}
= 1 \quad \text{by algebraic simplifications} \quad (10)
\]

\[
\frac{\left((-1 \cdot \left(S_{ACA} + \left(\frac{1}{2} \cdot (S_{ACC} + (-1 \cdot S_{ACA}))\right))\right) + (-1 \cdot S_{BCD})\right)}{\left(\frac{1}{2} \cdot S_{CBA}\right)}
= 1 \quad \text{by Lemma 29 (point } D \text{ eliminated)} \quad (11)
\]

\[
\frac{\left((-1 \cdot \left(0 + \left(\frac{1}{2} \cdot (0 + (-1 \cdot 0))\right)\right) + (-1 \cdot S_{BCD})\right)}{\left(\frac{1}{2} \cdot S_{CBA}\right)}
= 1 \quad \text{by geometric simplifications} \quad (12)
\]

\[
\frac{(-1 \cdot S_{BCD})}{\left(\frac{1}{2} \cdot S_{CBA}\right)}
= 1 \quad \text{by algebraic simplifications} \quad (13)
\]

\[
\frac{(-1 \cdot \left(S_{BCA} + \left(\frac{1}{2} \cdot (S_{BCC} + (-1 \cdot S_{BCA}))\right))\right)}{\left(\frac{1}{2} \cdot S_{CBA}\right)}
= 1 \quad \text{by Lemma 29 (point } D \text{ eliminated)} \quad (14)
\]

\[
\frac{(-1 \cdot \left(S_{BCA} + \left(\frac{1}{2} \cdot (0 + (-1 \cdot S_{BCA}))\right)\right))}{\left(\frac{1}{2} \cdot (-1 \cdot S_{BCA})\right)}
= 1 \quad \text{by geometric simplifications} \quad (15)
\]

\[
\frac{(-1 \cdot S_{BCA})}{\left(\frac{1}{2} \cdot (-1 \cdot S_{BCA})\right)}
= 1 \quad \text{by algebraic simplifications} \quad (16)
\]

NDG conditions are: \(S_{DEC} \neq S_{BEC}\) i.e., lines \(DB\) and \(EC\) are not parallel (construction based assumption); \(P_{GBG} \neq 0\) i.e., points \(G\) and \(B\) are not identical (conjecture based assumption).