Rating of Geometric Automated Theorem Provers

Nuno Baeta
CISUC
University of Coimbra, Portugal
nmsbaeta@gmail.com

Pedro Quaresma
CISUC / Department of Mathematics
University of Coimbra, Portugal
pedro@mat.uc.pt

The field of geometric automated theorem provers has a long and rich history, from the early synthetic provers in the 50th of last century to up to date provers.

Establishing a rating among them will be useful for the improvement of the current methods/implementations. Improvements could concern wider scope, better efficiency, proof readability and proof trustworthiness.

We need a common test bench: a common language to describe the geometric problems; a large set of problems and a set of measures to achieve such goal.

1 Introduction

When considering rating the Geometric Automated Theorem Provers (GATP) we have to consider some, somehow opposing, goals: scope; efficiency; readability; reliability of generated proofs [12, 16].

The first methods proposed, that came as early as in 1950s, adapt general-purpose reasoning approaches developed in the field of artificial intelligence, automating the traditional geometric proving processes. In order to avoid combinatorial explosion while applying postulates, many suitable heuristics, e.g. adding auxiliary elements to the geometric configuration, have been developed. Although being able to produce readable proofs, the proposed methods were very narrow-scoped and not efficient [21]. There are no recent results using this approach. It can be said that this first provers were narrow-scoped; inefficient; with readable proofs; without any attempt to testify the reliability of generated proofs.

The algebraic methods, such as the characteristic set method, the polynomial elimination method, the Gröbner basis method, and the Clifford algebra approach, reduce the complexity of logical inferences by computing relations between coordinates of geometric entities. What is gained in efficiency and wider scope is lost in the connection of the algebraic proof and the geometric reasoning. So: broad-scope; efficient; unreadable proofs, very complex algebraic proofs; without any attempt to testify the reliability of generated proofs. [3, 5, 4, 21]. This is still an active area of research [1, 10, 11, 13, 24].

In order to combine the readability of synthetic methods and efficiency of algebraic methods, some approaches, such as the area method or the full angle method, represent geometric knowledge in a form of expressions with respect to geometric invariants. These methods are broad-scoped (less than the algebraic); efficient (less than the algebraic); with readable proofs (less than the synthetics methods); some of then can have theirs proofs verified [3, 5, 4, 21, 10].

Other approaches, like the theorem prover ArgoCLP, based on coherent logic [18], or the deductive database approach [8] having been proposed with different rates of success in the different classes of geometric problems.

Submitted to:
ThEdu'18

2 Ratings

To be able to compare the different methods and implementations, alongside a standard test bench (see Section 3), we must define a set of ratings capable of assessing the GATPs in different classes: scope; efficiency; readability; reliability of generated proofs.

Scope To measure the scope of a GATP, one should consider:

- which geometries are allowed by the GATP—despite the fact that most GATPs deal only with Euclidean geometry, some exist that prove geometric problems in non-Euclidean geometries [3];
- what kind of problems are provable—as an example, the area method uses *geometric invariants* as basic quantities to prove theorems, each of which is used to deal with different geometric relations, hence allowing certain theorems to be easily proved [5].

Considering the existing methods, algebraic ones have the broadest scope. Not only have they been used to prove theorems in Euclidean and non-Euclidean geometry, but, for every geometry, the range of difficulty of the problems proved is very wide [3]. The earlier synthetic approaches were narrow scoped, e.g. the *GEOM* [9] only dealt with a limited set of geometric elements and relations [5]. In between lie semi-synthetic methods and coherent logic based methods. Semi-synthetic methods, starting with the area method which is complete for *constructive geometry* [6, 7, 10], with its *geometric invariant*, the *signed area*, allow many problems with relations like incidence and parallelism to be proved. Adding another *geometric invariant*, the *Pythagoras difference*, allows the problems with relations like perpendicularity and congruence of line segments to be easily proved. Adding other *geometric invariants* such as the *full-angle* (which gives its name to the full-angle method), the *volume*, the *vector*, allows the demonstration of an ever increasing range of theorems [5].

Synthetic and semi-synthetic methods scope may be influence by the use of deductive databases [8, 9]. Indeed, as stated in [8], unexpected results may be obtained, some of which are possibly new.

Efficiency The purpose of a GATP is, in addition to prove geometric conjectures, that these are obtained efficiently. By and large, although other resources may be involved, efficiency is related to time and memory space: we look for algorithms/implementations capable of fulfilling a proof in a reasonable amount of time and space.

Time is indeed the *natural* way to measure efficiency since it is used extensively, if not exclusively, throughout the literature [3, 4, 6, 7, 8, 10]. and, for obvious reasons, in a competition such as CADE ATP System Competition (*CASC*) [19]. Moreover, such measure is of paramount importance when considering an educational environment. Note however that when authors state something about GATP times, they do so in their settings, i.e. computer and operating system used, somehow restricting the usefulness of these results. The existence of a free and open platform where different GATPs can be tested on equal terms proves to be of utmost importance.

Regarding space, the authors are unaware of any study, presumably because these physical constraints are nowadays less important. Besides, from the users point of view time is the most important factor.

Readability Until recently, proofs in mathematics were solely made and verified by humans. With the advent of computers and automated reasoning that is no longer the case. It is therefore natural that the readability (by a human) of a computer generated proof is considered crucial. In an educational setting the proof is an object of learning by itself. In this setting the ability of a GATP to produce a synthetic proof, with the usual geometric inference rules, is of fundamental importance to its usability.

There are, at least, two different approaches to deal with the readability of proofs. In one approach only the proofs generated by the GATPs are analysed. The readability of such proofs is not only closely related to the method used, but may also be influenced by the features provided by the GATP itself. It is then possible to divide geometric proofs in the following way:

- *Algebraic proofs* are based on algebraic methods that generate proofs that may be considered illegible—the geometric conjecture is transformed in an algebraic problem with *hundreds* of terms and *dozens* of variables, loosing its tie to geometry;
- *Geometric proofs* are based on synthetic and semi-synthetic methods that generate proofs that closely resemble the human reasoning on geometry problems;
- *Geometric proofs with visual support* are also based on synthetic and semi-synthetic methods, but the GATP provides a visual and geometric step by step presentation of the proof steps [22, 23]. An example of such GATP is *JGEX*[1] [24].

The other approach, the de Bruijn factor [2, 20], involves the comparison of the *size* of an informal proof with the *size* of the corresponding machine proof, by means of a ratio. Although this approach cannot be automated, it may be useful in a classroom setting and may certainly be useful in enriching the Thousands of Geometric problems for geometric Theorem Provers[2] (*TGTP*) platform, a Web-based repository of geometric problems that was originally developed to support the testing and evaluation of geometric automated theorem proving systems [15].

Reliability By reliability is meant the confidence that we have in the proofs made by a given prover. Is the prover correctly implemented, are the proofs correct, or do we need to "prove the proofs"?

Using proof assistants like *Coq*,[3] or *Isabelle*,[4] the reliability of a given prover can be established, e.g. in [14] a implementation of the area method within *Coq* is described, where all the properties of the geometric quantities required by the area method are verified, demonstrating the correctness of the system, reducing concerns of reliability, to the trustworthiness of respective proof assistants [10, 14].

Unfortunately, within the current GATPs implementations, this is an exception, not the rule.

3 Test Bench

In order to implement a test bench, the *TGTP* platform presents itself as a solid foundation to fulfill such purpose. Due to its initial objective, as explained in [15], it already provides a centralised common repository of geometric problems[5] that may be used to test GATPs. Moreover, *TGTP* provides, as part of its infrastucture, implementations of several methods, namely: *GCLC*[6] implementations of Wu's method, Gröbner basis method and the area method; and a *Coq* implementation of the area method[7] [14]. Statistical and performance information is also supplied for all implementations, as well as a proof status for each geometric problem.

[1] `http://www.cs.wichita.edu/~ye/`
[2] `http://hilbert.mat.uc.pt/TGTP/`
[3] `https://coq.inria.fr/`
[4] `https://isabelle.in.tum.de/`
[5] In 2018–04–21 there were 236 problems.
[6] `http://poincare.matf.bg.ac.rs/~janicic/gclc/`
[7] `https://github.com/coq-contribs/area-method`

With a working test bench in operation, an interesting goal to pursue would be to operationalize a competition between GATPs, similar to *CASC*. Its ultimate goal, like in *CASC*, would be to encourage researchers to improve existing GATPs and implement new ones.

4 Conclusion and Future Work

Maybe the question "what is the best GATP of them all", can not be answered, but at least we should have some partials answers when looking for a GATP that fit a particular goal.

We need a common workbench: a common language to describe the geometric problems; a large set of problems and a set of measures to achieve such goal. Building on *TGTP* test bench [15], I2GATP common language [17] and *CASC*, CADE ATP System Competition [19], we need to improve and integrate the first two, adapting the ideas of the last, to reach the goal of a common workbench that all sort of users can use to choose the set of problems and tools that best fit their needs.

References

[1] Francisco Botana & Pedro Quaresma, editors (2015): *Automated Deduction in Geometry, 10th International Workshop, ADG 2014*. Lecture Notes in Artificial Intelligence 9201, Springer.

[2] Nicolaas Govert de Bruijn (1994): *A Survey of the Project Automath*, chapter A Survey of the Project Automath, pp. 141–161. Studies in Logic and the Foundations of Mathematics 133, North Holland, doi:10.1016/S0049-237X(08)70203-9. Available at `https://pure.tue.nl/ws/files/1892191/597622.pdf`.

[3] Shang-Ching Chou (1988): *Mechanical Geometry Theorem Proving*. Mathematics and Its Applications 41, D. Reidel Publishing Company.

[4] Shang-Ching Chou, Xiao-Shan Gao & Jing-Zhong Zhang (1994): *A Collection of 110 Geometry Theorems and Their Machine Produced Proofs Using Full-Angles*. Technical Report TR-94-4, Department of Computer Science, Wichita State University. Available at `https://www.researchgate.net/publication/239564904`.

[5] Shang-Ching Chou, Xiao-Shan Gao & Jing-Zhong Zhang (1994): *Machine Proofs in Geometry: Automated Production of Readable Proofs for Geometry Problems*. Applied Mathematics 6, World Scientific, doi:10.1142/9789812798152. Available at `https://www.researchgate.net/publication/240102887`.

[6] Shang-Ching Chou, Xiao-Shan Gao & Jing-Zhong Zhang (1996): *Automated Generation of Readable Proofs with Geometric Invariants: I. Multiple and Shortest Proof Generation*. Journal of Automated Reasoning 17(3), pp. 325–347, doi:10.1007/bf00283133.

[7] Shang-Ching Chou, Xiao-Shan Gao & Jing-Zhong Zhang (1996): *Automated Generation of Readable Proofs with Geometric Invariants: II. Theorem Proving With Full-Angles*. Journal of Automated Reasoning 17(3), pp. 349–370, doi:10.1007/BF00283134.

[8] Shang-Ching Chou, Xiao-Shan Gao & Jing-Zhong Zhang (2000): *A Deductive Database Approach to Automated Geometry Theorem Proving and Discovering*. Journal of Automated Reasoning 25(3), pp. 219–246, doi:10.1023/A:1006171315513.

[9] Helder Coelho & Luis Moniz Pereira (1986): *Automated Reasoning in Geometry Theorem Proving with Prolog*. Journal of Automated Reasoning 2(4), pp. 329–390, doi:10.1007/BF00248249.

[10] Predrag Janičić, Julien Narboux & Pedro Quaresma (2012): *The Area Method: A Recapitulation*. Journal of Automated Reasoning 48(4), pp. 489–532, doi:10.1007/s10817-010-9209-7.

[11] Predrag Janičić & Pedro Quaresma (2006): *System Description: GCLCprover + GeoThms*. In Ulrich Furbach & Natarajan Shankar, editors: *Automated Reasoning: Third International Joint Conference, IJCAR 2006, Seattle, WA, USA, August 17–20, 2006, Proceedings*, *Lecture Notes in Artificial Intelligence* 4130, Springer, pp. 145–150, doi:10.1007/11814771_13. Subseries of Lecture Notes in Computer Science.

[12] Jianguo Jiang & Jingzhong Zhang (2012): *A review and prospect of readable machine proofs for geometry theorems*. *Journal of Systems Science and Complexity* 25(4), pp. 802–820, doi:10.1007/s11424-012-2048-3. Available at `http://dx.doi.org/10.1007/s11424-012-2048-3`.

[13] Zoltán Kovács (2015): *The Relation Tool in GeoGebra 5*, pp. 53–71. *Lecture Notes in Artificial Intelligence* 9201, Springer International Publishing, doi:10.1007/978-3-319-21362-0_4.

[14] Julien Narboux (2009): *Formalization of the Area Method*. Coq user contribution. `http://dpt-info.u-strasbg.fr/~narboux/area_method.html`.

[15] Pedro Quaresma (2011): *Thousands of Geometric Problems for Geometric Theorem Provers (TGTP)*. In Pascal Schreck, Julien Narboux & Jürgen Richter-Gebert, editors: *Automated Deduction in Geometry*, *Lecture Notes in Computer Science* 6877, Springer, pp. 169–181, doi:10.1007/978-3-642-25070-5_10.

[16] Pedro Quaresma (2017): *Towards an Intelligent and Dynamic Geometry Book*. *Mathematics in Computer Science* 11(3–4), pp. 427–437, doi:10.1007/s11786-017-0302-8.

[17] Pedro Quaresma & Nuno Baeta (2015): *Current Status of the I2GATP Common Format*. In Francisco Botana & Pedro Quaresma, editors: *Automated Deduction in Geometry: 10th International Workshop, ADG 2014, Coimbra, Portugal, July 9–1, 2014, Revised Selected Papers*, *Lecture Notes in Artificial Intelligence* 9201, Springer, pp. 119–128, doi:10.1007/978-3-319-21362-0_8. Subseries of Lecture Notes in Computer Science.

[18] Sana Stojanović, Vesna Pavlović & Predrag Janičić (2011): *A Coherent Logic Based Geometry Theorem Prover Capable of Producing Formal and Readable Proofs*. In Pascal Schreck, Julien Narboux & Jürgen Richter-Gebert, editors: *Automated Deduction in Geometry: 8th International Workshop, ADG 2010, Munich, Germany, July 22-24, 2010, Revised Selected Papers*, *Lecture Notes in Artificial Intelligence* 6877, Springer, pp. 201–220, doi:10.1007/978-3-642-25070-5_12. Subseries of Lecture Notes in Computer Science.

[19] Geoff Sutcliffe & Christian Suttner (2006): *The State of CASC*. *AI Commun.* 19(1), pp. 35–48. Available at `http://dl.acm.org/citation.cfm?id=1143139.1143142`.

[20] Freek Wiedijk (2000): *The de Bruijn factor*. Poster presented at the Theorem Proving in High Order Logics, 13th International Conference, TPHOLs 2000, Portland, Oregon, USA.

[21] W.T. Wu (1984): *Automated Theorem Proving: After 25 Years*, chapter On the decision problem and the mechanization of theorem proving in elementary geometry, pp. 213–234. 29, American Mathematical Society.

[22] Zheng Ye, Shang-Ching Chou & Xiao-Shan Gao (2010): *Visually Dynamic Presentation of Proofs in Plane Geometry: Part 1. Basic Features and the Manual Input Method*. *Journal of Automated Reasoning* 45(3), pp. 213–241, doi:10.1007/s10817-009-9162-5.

[23] Zheng Ye, Shang-Ching Chou & Xiao-Shan Gao (2010): *Visually Dynamic Presentation of Proofs in Plane Geometry: Part 2. Automated Generation of Visually Dynamic Presentations with the Full-Angle Method and the Deductive Database Method*. *Journal of Automated Reasoning* 45(3), pp. 243–266, doi:10.1007/s10817-009-9163-4.

[24] Zheng Ye, Shang-Ching Chou & Xiao-Shan Gao (2011): *An Introduction to Java Geometry Expert*. In Thomas Sturm & Christoph Zengler, editors: *Automated Deduction in Geometry*, *Lecture Notes in Computer Science* 6301, Springer Berlin Heidelberg, pp. 189–195, doi:10.1007/978-3-642-21046-4_10. Available at `http://dx.doi.org/10.1007/978-3-642-21046-4_10`.