New and Interesting Theorems[†]

and Grading Proofs

Pedro Quaresma[*]

[*]CISUC / Department of Mathematics, University of Coimbra, Portugal
pedro@mat.uc.pt

ADG 2021, Hagenberg (remote), 15–17 September 2021

[†]Joint work with Pierluigi Graziani, University of Urbino, Italy

Problem 31

> *What properties can be identified to permit an automated reasoning program to find new and interesting theorems, as opposed to proving conjectured theorems?*
>
> *Automated Reasoning: 33 Basic Research Problems, Larry Wos*

Two (big!!!) problems in a single (simple) sentence:

- discover new theorems;
- select interesting theorems.

Theorem Discovery

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

Theorem Discovery

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures

- \+ stimulated by observations in the domain
- − unsound

Theorem Discovery

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures
- \+ stimulated by observations in the domain
- − unsound

generative e.g. mechanical manipulation of symbols
- \+ may generate a higher fraction of theorems than the inductive approach
- − unsound

Theorem Discovery

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures

- + stimulated by observations in the domain
- − unsound

generative e.g. mechanical manipulation of symbols

- + may generate a higher fraction of theorems than the inductive approach
- − unsound

manipulative generates conjectures from existing theorems

- + if the manipulations are satisfiability preserving, then theorems, rather than conjectures, are produced
- − the produced theorems are, in general, uninteresting

Theorem Discovery

Automated Generation of Interesting Theorems Puzis et al. [2006]:

How to do it:

inductive From facts to conjectures

- \+ stimulated by observations in the domain
- − unsound

generative e.g. mechanical manipulation of symbols

- \+ may generate a higher fraction of theorems than the inductive approach
- − unsound

manipulative generates conjectures from existing theorems

- \+ if the manipulations are satisfiability preserving, then theorems, rather than conjectures, are produced
- − the produced theorems are, in general, uninteresting

deductive application of sound inference rules to axioms

- \+ sound
- − avoid the generation of uninteresting consequences

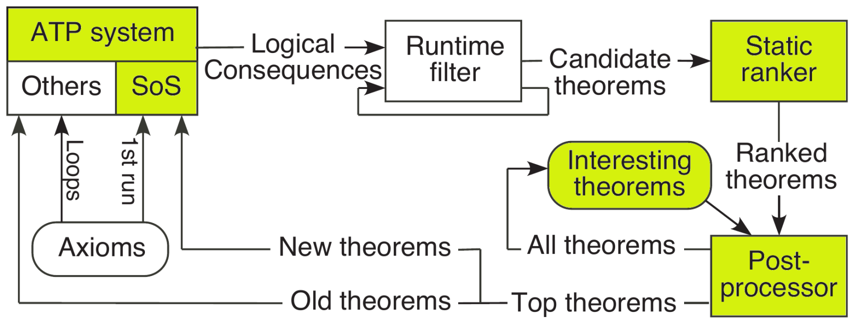
Interesting Theorems

From the set of generated conjectures to the set of interesting theorems Puzis et al. [2006]:

Pre-processor: discard obvious tautologies and apply filters:

- **Obviousness** the number of inferences in its derivation
- **Weight** the number of symbols it contains
- **Complexity** the number of distinct function and predicate symbols it contains
- **Surprisingness** measures new relationships between concepts and properties
- **Intensity** measures how much a formula summarizes information from the leaf ancestors in its derivation tree
- **Adaptivity** measures how tightly the universally quantified variables of a formula are constrained
- **Focus** measures the extent to which a formula is making a positive or negative statement about the domain of application
- **Usefulness** measures how much an interesting theorem has contributed to proofs of further interesting theorems

AGInT System



AGInT (**A**utomated **G**eneration of **In**teresting **T**heorems)

SoS – Set of Support

Puzis et al. [2006]

AGInT System — Pre-processor

Pre-processor (Runtime filter) — aggressively filter out and discard boring logical consequences.

- Discards obvious tautologies;
- Obviousness;
- Weight;
- Complexity;
- Surprisingness;
- Adaptivity;
- Focus.

AGInT — Static Ranker

Usefulness measures how much a candidate theorem has contributed to proofs of further interesting theorems: the ratio of its number of interesting descendents over its total number of descendents.

Normalization and Averaging the scores of the theorems, from each of the runtime filter and static evaluations, are normalized into the range 0.0 to 1.0.

AGInT — The Post-processor

The task of the post-processor is to remove less interesting theorems

Redundancy is tested in terms of subsumption for clauses, and deductibility for all formulae.

The second part of post-processing considers the remaining interesting theorems in pairs, in descending order of interestingness, so that each theorem is compared with every other less interesting theorem.

AGInT — Evaluation

The AGInT system has been evaluated by having it generate interesting theorems from axiom sets from the puzzles (PUZ) domain of the TPTP problem library Sutcliffe [2017], and from the axiomatization of set theory given in (McCasland et al. [2005]).

The E ATP (Schulz [2002]) was used

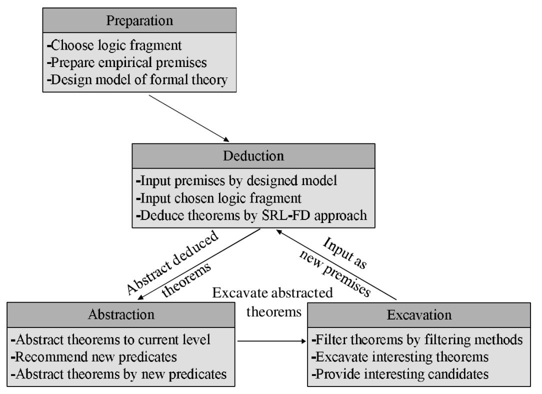
The post-processor used Otter 3.3 (McCune [2003])

It is noteworthy that in each of the PUZ tests, AGInT generates and identifies interesting theorems that are not mentioned in the original problem, i.e., interesting theorems not explicitly identified in the source domain.

Theorem Discovery

Research on Automated Theorem Finding: Current State and Future Directions Gao et al. [2014]

Strong Relevant Logic-based forward deduction approach



Relevance logic, is a kind of non-classical logic requiring the antecedent and consequent of implications to be relevantly related.

Interesting Theorems

Measuring Interestingness of Theorems in Automated Theorem Finding by Forward Reasoning: A Case Study in Tarski's Geometry Gao et al. [2018]

- **Degree of logical connectives in empirical theorems**
  the degree of logical connectives is related to the interestingness of empirical theorems, and interesting theorems always hold lower degree of logical connectives

- **Propositional schema of formula**
  The most frequent propositional schemata of known theorems is A type. A theorem is always interesting if the theorem does not contain any logical connective, because it holds clear and concise semantics. The second frequent propositional schema is $A \Rightarrow B$. We think the reason is that "if A then B" is a very frequent conditional propositional schema in any fields.

- **Abstract level of predicates and functions in one theorem**
  In the mathematical fields, mathematicians always make definition from simple to complex.
  A theorem that holds higher abstract level predicates and functions, is more interesting from the viewpoint of the meaning of the theorem.

- **Deduction distance**
  The interesting theorems are those theorems which are difficult to be reasoned out from premises.
  Therefore, if the deduction distance of an obtained theorem is long, the theorem may be interesting.

Interesting Theorems

On the notion of interestingness in automated mathematical discovery Colton et al. [2000].

A survey of five mathematical discovery programs.

Program	Year	Domains
AM	1976	set, number
GT	1987	graph
Graffiti	1988	graph, number, geometry
Bagai et al.	1993	geometry
HR	1997	finite alg. number, graph

Discovery

- **Manipulative** generate conjectures from existing theorems

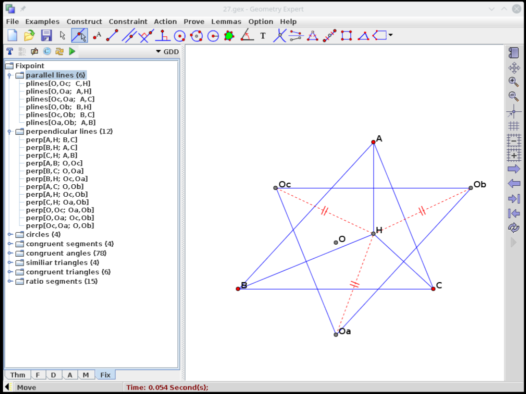
Filtering

- **Pre-processing** Discard obvious tautologies and heuristics to discard trivial conjectures
- **Surprisingness** measure new relations
- **Complexity** the number of distinct function and predicate symbols it contains
- **Usefulness** measures how much an interesting theorem has contributed to proofs of further interesting theorems

Deductive Databases

- **G**eometry **D**eductive **D**atabase **M**ethod — breadth-first forward chaining in order to reach fix-point Chou et al. [2000], Ye et al. [2011]

$$\boxed{D_0} \overset{R}{\subset} \boxed{D_1} \overset{R}{\subset} \cdots \overset{R}{\subset} \boxed{D_k} \quad \text{(fix-point)}$$



Algebraic Manipulations in Geometry

Algebraic manipulations in Geometry:

Automatic Discovery of Theorems in Elementary Geometry, Recio and Vélez [1999]

Find the missing hypotheses so that a given conclusion follows from a given incomplete set of hypotheses, by algebraic means.

A dynamic-symbolic interface for geometric theorem discovery, Botana and L. Valcarce [2002]

Towards Automated Discovery of Geometrical Theorems in GeoGebra, Kovács and Yu [2020]

Definition of point as the set of all points in the construction "equal in general" (discarding floating point differences), plus a set of properties regarding: lines; circles; parallel lines; congruent segments.

Discovery in Geometry, ADG2021

14:40 Zoltán Kovács, Tomas Recio and M. Pilar Vélez: GeoGebra Discovery in context

15:20 Philip Todd: A method for the automated discovery of angle theorems

15:40 Christopher W. Brown, Zoltán Kovács and Robert Vajda: Supporting proving and discovering geometric inequalities in GeoGebra by using Tarski

16:00 Zoltán Kovács and Róbert Vajda: Parametric Root Finding to support discovering geometric inequalities in GeoGebra

Grading GATP Proofs

¿ Interesting ATP theorems must have readable proofs ?

Grading proofs in order to establish a readability criterion.

Three proposals to measure the readability of a proof.

- The TML Criterion, by Chou et al. [1994].
- de Bruijn factor by de Bruijn [1994], Wiedijk [2000].
- Geometrography Readability Coefficient of a Proof (GRCP) by Quaresma and Graziani (in major revision).

TML Criterion

Metrics to grade proofs done by geometry automated theorem prover (GATP).

TML Criterion Chou et al. [Chou et al., 1994, p.452]

- *time* is the time needed to complete the machine proof;
- *maxt* is the number of terms of the maximal polynomial occurring in the machine proof. Thus *maxt* measures the amount of computation needed 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.

Readability Accordingly to TML Criterion

According to [Chou et al., 1994, p.452] 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

In the *de Bruijn factor* the quotient of the *size* of corresponding formal proof and the *size* of the informal (rigorous) proof is used as a measure of readability of the formal proof.

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 then a given informal proof).

Classical Geometrography

In Lemoine's Geometrography two coefficients are defined to measure the relative difficulty to perform some ruler and compass geometric constructions.

To place the edge of the ruler in coincidence with one point . . . R1
To put one point of the compasses on a determinate point C1
(. . .)

- *cs*, the coefficient of simplicity — measures the simplicity of the overall construction.
- *ce*, the coefficient of exactitude — measures the accuracy of the final construction.

Geometrography in Dynamic Geometry

Modernize version of Geometrography, taking in consideration the DGS Quaresma et al. [2020], Santos et al. [2019].
Define a point, anywhere in the plane, D, and define a given object, using other objects, C.

- cs, the coefficient of simplicity (adapted to the tools of the DGS).
- cf, the coefficient of freedom — measures the degree of movement allowed (the dynamic of the construction).

Geometrography in Automatic Theorem Proving

The same approach can be (again) extrapolated to take into consideration synthetic geometric proofs, done using the Area Method (GCLC implementation) Janičić et al. [2012].

Coefficient of Simplicity .... (CS)

(Elementary) Algebraic Simplification ....(AS)

(Elementary) Geometric Simplification ....(GS)

Application of the Area Method Lemma n ....(AMLn)

The coefficient of simplicity for a given conjecture:

$$\mathrm{CS}_{\mathrm{proof}} = n_1 + n_2 + n_3 + \sum_{j=l_1}^{l_k} \mathbf{AML_j}$$

A Geometrography Criterion

Geometrography Readability Coefficient of Proof (GRCP)

$$GRCP = ((\mathrm{CS}_{\mathrm{proof}} - \mathrm{CT}_{\mathrm{proof}}) \times (\mathrm{CD}_{\mathrm{highproof}} + \mathrm{CD}_{\mathrm{typeproof}}))$$

This coefficient relates four quantities:

- $\mathrm{CS}_{\mathrm{proof}}$, the simplicity coefficient of the proof, it gives the (geometrography) simplicity coefficient for the overall proof;
- $\mathrm{CT}_{\mathrm{proof}}$, the total number steps in the proof;
- $\mathrm{CD}_{\mathrm{highproof}}$, the number of different steps with high difficulty present in the proof;
- $\mathrm{CD}_{\mathrm{typeproof}}$, the number of different lemmas used in the proof.

To get the **high difficulty**, we have analysed the area method lemmas implemented in the GATP *GCLC* divided them into three categories: low/medium/high difficulty.

A Geometrography Criterion

Considering 71 theorems and their area method proofs, from the *TGTP* repository, a k-means clustering divides the proofs into the following classes of Geometrography readability:

- readable (high-readability) $< 80\,000$;
- $80\,000 \leq$ medium-readability $< 260\,000$;
- low-readability $\geq 260\,000$.

GRCP of TPTP's Problem GEO0021

Theorem (Circumcenter of a Triangle)

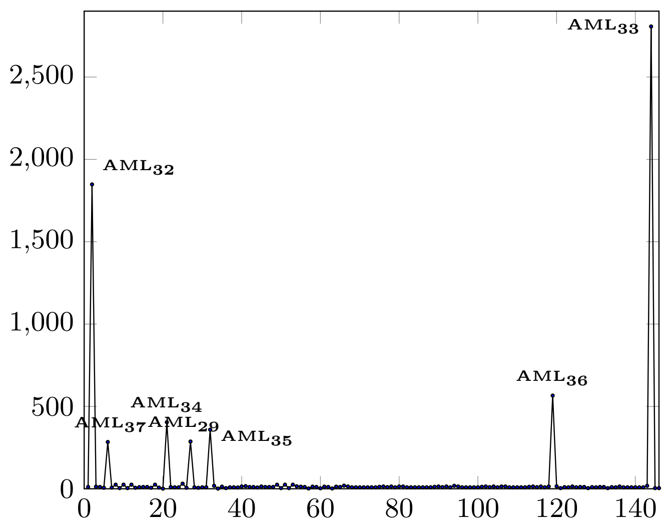
The circumcenter of a triangle can be found as the intersection of the three perpendicular bisectors

$$\texttt{GEO0021}\left\{\begin{array}{rcl} \mathrm{CS}_{\mathrm{proof}} & = & 8\,554 \\ \mathrm{CS}_{\mathrm{gcl}} & = & 11 \\ \mathrm{CT}_{\mathrm{proof}} & = & 591 \\ \mathrm{CS}_{\mathrm{proofmax}} & = & 2\,807 \\ \mathrm{CD}_{\mathrm{typeproof}} & = & 13 \\ \mathrm{CD}_{\mathrm{highproof}} & = & 3 \end{array}\right.$$

$$80\,000 \leq GRCP = 127\,408 < 260\,000$$

A **medium-readability** problem. It can be seen that it has 13 different lemmas, 3 high-difficulty step, a long proof with a significant difference between the $\mathrm{CS}_{\mathrm{proof}}$ and the number of steps of the proof

GEO0021, Geometrography Proof Trace



Comparing the Different Criteria

The GRCP criteria takes into consideration all the significant aspects of a synthetic proof, its overall difficulty, its number of steps, the number of difficulty steps and the number of different lemmas that must be applied.

The other criteria consider fewer aspects:

- de Bruijn criteria takes only in consideration the size of the proof informal proof vs formal proof.
- The TML criteria consider the number of different lemmas applied and uses the number of terms of the maximal polynomial as a way to have an approximation to the complexity of the proof.

Comparing the Different Criteria

In many cases the three criteria agree.

TGTP	TML	de Brujin	GRCP
GEO0001	$\frac{3 < 5, \text{ deduction steps}}{\text{easy}}$	$\frac{1.6 < 2}{\text{easy}}$	$\frac{564 < 80\,000}{\text{easy(high)}}$
GEO0021	$\frac{13 > 5 \text{ deduction steps \& number of terms} > 5}{\text{difficult}}$	$\frac{37.63 > 2}{\text{difficult}}$	$\frac{80\,000 \leq 127\,408 < 260\,000}{\text{difficult(medium)}}$
GEO0020	$\frac{13 > 5 \text{ deduction steps \& number of terms} > 5}{\text{difficult}}$	$\frac{47.31 > 2}{\text{difficult}}$	$\frac{80\,000 \leq 253\,920 < 260\,000}{\text{difficult(medium)}}$

Table: Comparison of the Three Criteria

New and Interesting Theorem, remains an elusive question.

. . . but we are working on it.

Obrigado

Bibliography I

Michael Beeson and Larry Wos. Finding proofs in Tarskian geometry. *Journal of Automated Reasoning*, 58(1):181–207, oct 2016. doi: 10.1007/s10817-016-9392-2.

Francisco Botana and José L. Valcarce. A dynamic-symbolic interface for geometric theorem discovery. *Computers and Education*, 38:21–35, 2002.

Shang-Ching Chou, Xiao-Shan Gao, and Jing-Zhong Zhang. *Machine Proofs in Geometry*. World Scientific, apr 1994. doi: 10.1142/2196.

Shang-Ching Chou, Xiao-Shan Gao, and Jing-Zhong Zhang. A deductive database approach to automated geometry theorem proving and discovering. *Journal of Automated Reasoning*, 25:219–246, 2000.

Simon Colton, Alan Bundy, and Toby Walsh. On the notion of interestingness in automated mathematical discovery. *International Journal of Human-Computer Studies*, 53(3):351–375, sep 2000. doi: 10.1006/ijhc.2000.0394.

N. G. de Bruijn. *Selected Papers on Automath*, volume 133 of *Studies in logic and the foundations of mathematics*, chapter A survey of the project Automath, pages 41–161. North-Holland, Amsterdam, 1994.

Bibliography II

H. Gao, J. Li, and J. Cheng. Measuring interestingness of theorems in automated theorem finding by forward reasoning based on strong relevant logic. In *2019 IEEE International Conference on Energy Internet (ICEI)*, pages 356–361. IEEE, may 2019. doi: 10.1109/ICEI.2019.00069.

Hongbiao Gao, Yuichi Goto, and Jingde Cheng. Automated theorem finding by forward deduction based on the semi-lattice model of formal theory: A case study in NBG set theory. oct 2013. doi: 10.1109/SKG.2013.36.

Hongbiao Gao, Yuichi Goto, and Jingde Cheng. Research on automated theorem finding: Current state and future directions. In *Lecture Notes in Electrical Engineering*, pages 105–110. Springer Berlin Heidelberg, 2014. doi: 10.1007/978-3-642-55038-6_16.

Hongbiao Gao, Yuichi Goto, and Jingde Cheng. A set of metrics for measuring interestingness of theorems in automated theorem finding by forward reasoning: A case study in NBG set theory. In *Intelligence Science and Big Data Engineering. Big Data and Machine Learning Techniques*, pages 508–517. Springer International Publishing, 2015. doi: 10.1007/978-3-319-23862-3_50.

Bibliography III

Hongbiao Gao, Jianbin Li, and Jingde Cheng. Measuring interestingness of theorems in automated theorem finding by forward reasoning: A case study in tarski's geometry. In *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE, oct 2018. doi: 10.1109/SmartWorld.2018.00064.

Predrag Janičić, Julien Narboux, and Pedro Quaresma. The Area Method: a recapitulation. *Journal of Automated Reasoning*, 48(4):489–532, 2012. doi: 10.1007/s10817-010-9209-7.

Zoltán Kovács and Jonathan H. Yu. Towards automated discovery of geometrical theorems in geogebra. *CoRR*, abs/2007.12447, 2020. URL https://arxiv.org/abs/2007.12447.

Bibliography IV

Roy L. McCasland, Alan Bundy, and Patrick F. Smith. Ascertaining mathematical theorems. In Jacques Carette and William M. Farmer, editors, *Proceedings of the 12th Symposium on the Integration of Symbolic Computation and Mechanized Reasoning (Calculemus 2005)*, volume 151 of *Electronic Notes in Theoretical Computer Science*, pages 21–38. Elsevier, 2005. URL `https://www.sciencedirect.com/journal/electronic-notes-in-theoretical-computer-science/vol/151/issue/1`.

William McCune. Otter3.3 reference manual. Technical Memorandum 263, Argonne National Laboratory, 2003.

Xicheng Peng, Qihang Chen, Jingzhong Zhang, and Mao Chen. Automated discovery of geometric theorems based on vector equations. *Journal of Automated Reasoning*, 65(6):711–726, apr 2021. doi: 10.1007/s10817-021-09591-2.

Yury Puzis, Yi Gao, and G. Sutcliffe. Automated generation of interesting theorems. In *FLAIRS Conference*, 2006.

Bibliography V

Pedro Quaresma and Pierluigi Graziani. Measuring the readability of a proof. (submitted to JAR).

Pedro Quaresma, Vanda Santos, Pierluigi Graziani, and Nuno Baeta. Taxonomy of geometric problems. *Journal of Symbolic Computation*, 97: 31–55, 2020. ISSN 0747-7171. doi: 10.1016/j.jsc.2018.12.004. URL http://www.sciencedirect.com/science/article/pii/S0747717118301305.

T. Recio and M. P. Vélez. Automatic discovery of theorems in elementary geometry. *J. Autom. Reason.*, 23:63–82, July 1999. ISSN 0168-7433. doi: 10.1023/A:1006135322108. URL http://dl.acm.org/citation.cfm?id=594128.594243.

Vanda Santos, Nuno Baeta, and Pedro Quaresma. Geometrography in dynamic geometry. *International Journal for Technology in Mathematics Education*, 26(2):89–96, 2019. doi: 10.1564/tme_v26.2.06.

Stephan Schulz. E - a brainiac theorem prover. *AI Commun.*, 15(2,3):111–126, August 2002. ISSN 0921-7126.

Bibliography VI

G. Sutcliffe. The TPTP Problem Library and Associated Infrastructure. From CNF to TH0, TPTP v6.4.0. *Journal of Automated Reasoning*, 59(4): 483–502, 2017.

Freek Wiedijk. The de Bruijn factor. Poster at International Conference on Theorem Proving in Higher Order Logics (TPHOL2000), 2000. Portland, Oregon, USA, 14-18 August 2000.

Larry Wos. *Automated Reasoning: 33 Basic Research Problems*. Prentice-Hall, 1988. ISBN 978-0130545527.

Larry Wos. The problem of automated theorem finding. *Journal of Automated Reasoning*, 10(1):137–138, 1993. doi: 10.1007/BF00881868.

Zheng Ye, Shang-Ching Chou, and Xiao-Shan Gao. An introduction to java geometry expert. In Thomas Sturm and Christoph Zengler, editors, *Automated Deduction in Geometry*, volume 6301 of *Lecture Notes in Computer Science*, pages 189–195. Springer Berlin Heidelberg, 2011. ISBN 978-3-642-21045-7. doi: 10.1007/978-3-642-21046-4_10. URL http://dx.doi.org/10.1007/978-3-642-21046-4_10.