

Rating of Geometric Automated Theorem Provers

Nuno Baeta¹ Pedro Quaresma²

CISUC
University of Coimbra
nmsbaeta@gmail.com

CISUC/Mathematics Department
University of Coimbra
pedro@mat.uc.pt

ThEdu'18 at FLoC 2018, 18 July 2018, Oxford, United Kingdom

Why Rate GATPs

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

Why Rate GATPs

The field of geometric automated theorem provers (GATP) 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;
- ▶ educational purposes.

Ratings

GATPs features to consider:

Ratings

GATPs features to consider:

scope;

Ratings

GATPs features to consider:

scope; efficiency;

Ratings

GATPs features to consider:

scope; efficiency; readability (of proofs);

Ratings

GATPs features to consider:

scope; efficiency; readability (of proofs); reliability.

Ratings

GATPs features to consider:

scope; efficiency; readability (of proofs); reliability.

Scope takes into account:

- ▶ which geometries are allowed by the GATP;
- ▶ what kind of problems are provable.

Ratings

GATPs features to consider:

scope; efficiency; readability (of proofs); reliability.

Efficiency is measured by:

- ▶ memory space;
- ▶ time.

Ratings

GATPs features to consider:

scope; efficiency; readability (of proofs); reliability.

Readability Two approaches:

- ▶ Only the proofs generated by the GATP are analysed. These may be:
 - ▶ algebraic proofs;
 - ▶ geometric proofs;
 - ▶ geometric proofs with visual support.
- ▶ The de Bruijn factor: Comparison of *sizes* of the informal and machine generated proofs, by means of a ratio.

Ratings

GATPs features to consider:

scope; efficiency; readability (of proofs); reliability.

Reliability gives the confidence that we have on machine generated proof.

Is the prover correctly implemented? Are the proofs correct? Do we need to “prove the proofs”?

Test Bench

With a rating criteria defined, another problems arises.

How to apply these criteria evenly?

Test Bench

With a rating criteria defined, another problems arises.

How to apply these criteria evenly?

Current situation

Statements about GATPs depend on the authors settings (computer and operating system) and theorem collection, restricting the usefulness of the obtained results.

Test Bench

With a rating criteria defined, another problems arises.

How to apply these criteria evenly?

Current situation

Statements about GATPs depend on the authors settings (computer and operating system) and theorem collection, restricting the usefulness of the obtained results.

Desirable situation

A free and open platform where different GATPs can be tested on equal terms.

Test Bench

To build such a test bench, existing formats/tools, provide a solid foundation.

- ▶ *I2GATP*, a common language to describe geometric problem, i.e., support conjectures and proofs produced by GATPs;
- ▶ *TGTP*, a centralised repository of geometric problems (which implements some GATPs as part of its current infrastructure).

Improvements are needed, to fulfil the goals.

Conclusion and Future Work

Even if we cannot answer the question “what is the best GATP of them all”, we should have partial answers for particular questions.

Conclusion and Future Work

Even if we cannot answer the question “what is the best GATP of them all”, we should have partial answers for particular questions.

A test bench provides, at least some, such answers.

Conclusion and Future Work

Even if we cannot answer the question “what is the best GATP of them all”, we should have partial answers for particular questions.

A test bench provides, at least some, such answers.

GASC, a competition between GATPs similar to CASC, should be created to encourage researchers to improve existing GATPs and implement new ones.

Conclusion and Future Work

Even if we cannot answer the question “what is the best GATP of them all”, we should have partial answers for particular questions.

A test bench provides, at least some, such answers.

GASC, a competition between GATPs similar to CASC, should be created to encourage researchers to improve existing GATPs and implement new ones.

Obrigado