Automated Generation of Formal and Readable Proofs of Mathematical Theorems — ongoing work —

Sana Stojanović Predrag Janičić
Faculty of Mathematics
University of Belgrade

SVARM 2013
Rome, Italy, January 20-21, 2013.

Overview

- Motivation
- Framework
- Case Study: Tarski's Book on Geometry
- Conclusions and further work

Readable Proofs

- Lots of research efforts have been invested into automation and formalization of theorem proving
- .. but much less efforts is invested into *readable* proofs
- By *readable proofs* we mean textbook-like proofs
- Readable proofs are typically not relevant in fields such as software verification
- ... but are very important in mathematical practice

Our Goal

- We want to build a system that will be able to:
  - efficiently prove mathematical theorems
  - generate machine verifiable proofs
  - generate readable, textbook-like proofs
- The system should be helpful to mathematicians in formalizing mathematical heritage, textbooks, etc.
- One of the key issues is finding an appropriate logical framework

What is Coherent Logic

- This work is based on coherent logic (CL)
- Coherent logic (also: *geometric logic*) is a fragment of FOL
- First used by Skolem, recently popularized by Bezem et al.
- CL has a natural proof system, based on forward reasoning
- Existential quantifiers are eliminated by introducing witnesses
- A conjecture is kept unchanged and proved directly (refutation, Skolemization and clausal form are not used)
- Generating readable and formal proofs is simple

What is Coherent Logic (2)

- CL formulae are of the form:

$$A_1(\vec{x}) \wedge \ldots \wedge A_n(\vec{x}) \Rightarrow \exists \vec{y}_1 \; B_1(\vec{x}, \vec{y}_1) \vee \ldots \vee \exists \vec{y}_m \; B_m(\vec{x}, \vec{y}_m)$$

  (A_i are literals, B_i are conjunctions of literals)
- No function symbols of arity greater than 0
- No negation (negated facts are *simulated* by new predicates)
- Intuitionistic logic
- The problem of deciding $\Gamma \vdash \Phi$ is semi-decidable

CL Realm

- A number of theories and theorems can be formulated directly and simply in CL
- Example (Euclidean geometry theorem):
  for any two points there is a point between them
- Many conjectures in geometry, abstract algebra, confluence theory, lattice theory, ... (Bezem et.al.)

CL Provers (some of)

- Euklid by Stevan Kordić and Predrag Janičić (1995)
- CL prover by Marc Bezem (2005)
- ArgoCLP by Sana Stojanović, Vesna Pavlović and Predrag Janičić (2009)
- Geo by Hans de Nivelle (2008)
- Calypso by Mladen Nikolić and Predrag Janičić (2012)

Features of CL Provers

- Sound and complete
- Ground reasoning or FOL reasoning
- Backtracking or backjumping
- Lemma learning (some)
- CDCL-based (some)
- Isabelle/Isar and natural language proofs (some)
- Still not very efficient

Example: Proof Generated by ArgoCLP

Let us prove that $p = r$ by reductio ad absurdum.

1. Assume that $p \neq r$.
   2. It holds that the point A is incident to the line q or the point A is not incident to the line q (by axiom of excluded middle).
      3. Assume that the point A is incident to the line q.
         4. From the facts that $p \neq q$, and the point A is incident to the line p, and the point A is incident to the line q, it holds that the lines p and q intersect (by axiom ax_D5).
         5. From the facts that the lines p and q intersect, and the lines p and q do not intersect we get a contradiction.

         Contradiction.
      6. Assume that the point A is not incident to the line q.
         7. From the facts that the lines p and q do not intersect, it holds that the lines q and p do not intersect (by axiom ax_nint_l_l_21).
         8. From the facts that the point A is not incident to the line q, and the point A is incident to the plane α, and the line q is incident to the plane α, and the point A is incident to the line p, and the line p is incident to the plane α, and the lines q and p do not intersect, and the point A is incident to the line r, and the line r is incident to the plane α, and the lines q and r do not intersect, it holds that $p = r$ (by axiom ax_E2).
         9. From the facts that $p = r$, and $p \neq r$ we get a contradiction.

         Contradiction.

Therefore, it holds that $p = r$.

This proves the conjecture.

Combination of Tools: Provers for Coherent Logic

- Provers for coherent logic
  - are automated
  - can export machine-verifiable proofs and readable proofs
- but...
  - are not efficient enough

Combination of Tools: Proof Assistants

- Proof assistants
  - are trusted
- but...
  - the level of automation within them is low
  - they are still not mathematician-friendly enough

Combination of Tools: Resolution Provers

- Resolution provers
  - are automated and efficient
- but...
  - they don't produce human-readable and machine verifiable proofs

Combination of Tools: Combined Power

- Therefore, we want to combine the power of:
  - Proof assistants
  - Resolution provers
  - Provers for coherent logic

Framework Description

- Sledgehammer-like:
  - using the power of external resolution provers
- Instead of trusted prover Metis, a CL prover is used and formal proofs are exported

Proving Algorithm

1. The available axioms and theorems are passed to resolution based automated theorem provers
2. If one or more resolution provers proves the conjecture, the smallest list of used axioms is used again
3. The returned list of used axioms is reversed, and the automated proving process is rerun; this is repeated until the set of used axioms is not changed
4. CL prover is invoked with the obtained list of axioms

Proving Algorithm

- For one theorem — all axioms and preceding theorems are feeded into the system
- The system works *fully automatically*, no guiding at all

Choices

- Input format for axioms and theorems: TPTP
- Resolution provers used: Vampire, E, and Spass
- CL prover used: ArgoCLP
- Output format for proofs: Isabelle/Isar and natural language

"Tarski's Book"

- Wolfram Schwabhaüser, Wanda Szmielew, and Alfred Tarski: *Metamathematische Methoden in der Geometrie* (1983)
- Culmination of a series of Tarski's axiomatization for geometry
- One of the twenty-century mathematical classics
- Self-contained: all theorems are provable from the set of starting axioms
- The set of theorems in the book makes a well-rounded set of theorems

W. Schwabhäuser
W. Szmielew A. Tarski

Metamathematische Methoden in der Geometrie

Mit 167 Abbildungen

Teil I: Ein axiomatischer Aufbau der euklidischen Geometrie
von W. Schwabhäuser, W. Szmielew und A. Tarski

Teil II: Metamathematische Betrachtungen
von W. Schwabhäuser

Springer-Verlag
Berlin Heidelberg New York Tokyo 1983

7.8 Satz. $\forall p' \exists^{=1} p\, S_a p = p'$.

7.9 Satz. $S_a p = S_a q \to p = q$.

7.10 Satz. $S_a p = p \leftrightarrow p = a$.

7.4 (bzw. 7.5) und 7.8 lassen sich zusammenfassen in

7.11 Satz. S_a *ist eine eineindeutige Abbildung des ganzen Raumes auf sich.*

Da es verschiedene Punkte gibt, ist diese Abbildung nach 7.10 nicht die identische Abbildung, aber ihre Verkettung mit sich selbst ist nach 7.7 die identische Abbildung. Das wird zusammengefaßt in

7.12 Satz. S_a *ist involutorisch.*

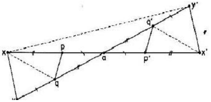
Wir wollen nun zeigen, daß S_a sogar eine Bewegung ist (4.8). Dazu benötigen wir noch, daß jede Strecke zu ihrer Bildstrecke kongruent ist, d.h.

7.13 Satz. $pq \equiv S_a(p)S_a(q)$.

Beweis: Sei $p' = S_a p$, $q' = S_a q$. Falls $p = a$ ist, ist nach 7.10 auch $p' = a$, nach Definition von q' andererseits $aq \equiv aq'$, also gilt die Behauptung. Für das folgende sei also $p \neq a$. Durch Streckenabtragung (A4) erhalten wir Punkte x, y, x', y' (Abb. 17) mit

$Bp'px \wedge px \equiv qa$, $Bq'qy \wedge qy \equiv pa$,

$Bxp'x' \wedge p'x' \equiv qa$, $Byq'y' \wedge q'y' \equiv pa$.



Nach 3.12 (mit $l=1$) und 3.9 ergeben sich dann die verallgemeinerten Zwischenbeziehungen

$B_5 xpap'x'$ und $B_5 yqaq'y'$.

Nach 2.11 (Aneinanderlegen von Strecken) ergeben sich die Streckenkongruenzen

$ax \equiv ay \equiv ay' \equiv ax'$.

Nun erhalten wir die äußere Fünf-Streckenkonfiguration

$$AFS\begin{pmatrix} x & ax' & y' \\ y' & ay & x \end{pmatrix}.$$

Wegen $p \neq a$ ist erst recht $x \neq a$, nach A5 also $x'y' \equiv xy$. Unter Benutzung vo 4.2 erhalten wir weiter

$$IFS\begin{pmatrix} y & q & ax \\ y' & q' & ax' \end{pmatrix}, \quad \text{also} \quad qx \equiv q'x',$$

$$IFS\begin{pmatrix} x & p & aq \\ x' & p' & aq' \end{pmatrix}, \quad \text{also} \quad pq \equiv p'q'. \qquad \square$$

Mit 4.8, 4.9 erhalten wir

7.14 Satz. *Jede Punktspiegelung* S_a *ist eine Bewegung und (damit) ein Automorphismus des ganzen Raumes.*

Im einzelnen wird die Eigenschaft, Automorphismus zu sein, ausgedrückt durch 7.11 und die folgenden beiden Sätze.

7.15 Satz. $Bpqr \leftrightarrow BS_a(p)S_a(q)S_a(r)$.

7.16 Satz. $pq \equiv rs \leftrightarrow S_a(p)S_a(q) \equiv S_a(r)S_a(s)$.

7.17 Satz. $Mpap' \wedge Mpbp' \to a = b$.
(*Jede Strecke hat höchstens einen Mittelpunkt.*)

Anmerkung. Die Existenz des Mittelpunktes zu jeder Strecke wird erst später bewiesen (Satz 8.22).

Beweis von 7.17. Es ist $pb \equiv p'b$, nach 7.13 (angewendet auf S_a) andererseits $p'b \equiv pS_a(b)$, also $pb \equiv pS_a(b)$. Durch Vertauschung von p mit p' ergibt sich ebenso $p'b \equiv p'S_a(b)$. Nach Voraussetzung ist auch $Bpbp'$, nach 4.19 also $b = S_a(b)$ und nach 7.10 somit $a = b$. $\square$

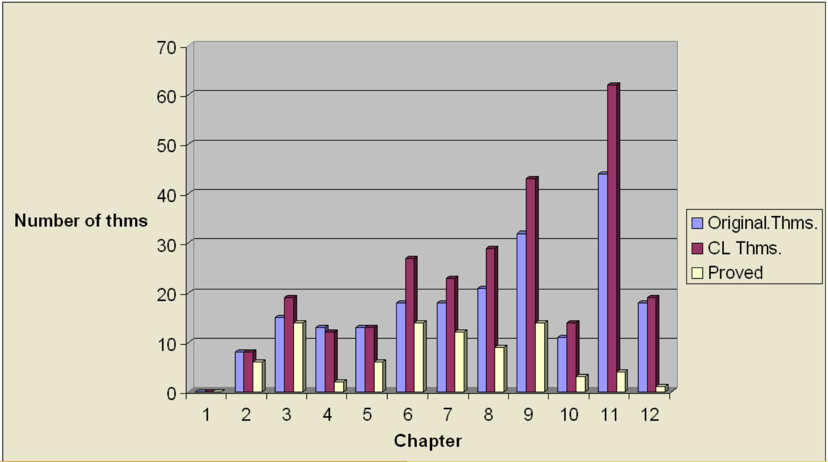
Axioms

1. $\forall A\, \forall B\; cong(A, B, B, A)$
2. $\forall A\, \forall B\, \forall P\, \forall Q\, \forall R\, \forall S\; (cong(A, B, P, Q) \wedge cong(A, B, R, S) \Rightarrow cong(P, Q, R, S))$
3. $\forall A\, \forall B\, \forall C\; (cong(A, B, C, C) \Rightarrow A = B)$
4. $\forall A\, \forall B\, \forall C\, \forall Q\, \exists X\; (bet(Q, A, X) \wedge cong(A, X, B, C))$
5. $\forall A\, \forall B\, \forall C\, \forall D\, \forall A1\, \forall B1\, \forall C1\, \forall D1\; (A \neq B \wedge bet(A, B, C) \wedge bet(A1, B1, C1) \wedge cong(A, B, A1, B1) \wedge cong(B, C, B1, C1) \wedge cong(A, D, A1, D1) \wedge cong(B, D, B1, D1) \Rightarrow cong(C, D, C1, D1))$
6. $\forall A\, \forall B\; (bet(A, B, A) \Rightarrow A = B)$
7. $\forall A\, \forall B\, \forall C\, \forall P\, \forall Q\; (bet(A, P, C) \wedge bet(B, Q, C) \Rightarrow \exists X\; (bet(P, X, B) \wedge bet(Q, X, A)))$
8. $\exists A\, \exists B\, \exists C\; (\neg bet(A, B, C) \wedge \neg bet(B, C, A) \wedge \neg bet(C, A, B))$
9. $\forall P\, \forall Q\, \forall A\, \forall B\, \forall C\; (P \neq Q \wedge cong(A, P, A, Q) \wedge cong(B, P, B, Q) \wedge cong(C, P, C, Q) \Rightarrow (bet(A, B, C) \vee bet(B, C, A) \vee bet(C, A, B)))$
10. $\forall A\, \forall B\, \forall C\, \forall D\, \forall T\; (bet(A, D, T) \wedge bet(B, D, C) \wedge A \neq D \Rightarrow \exists X\, \exists Y\; (bet(A, B, X) \wedge bet(A, C, Y) \wedge bet(X, T, Y)))$

Translation to CL – First 12 Chapters

- 211 theorems altogether in the first 12 (of 16) Chapters
  - 93 already in CL form (44%)
  - 36 can be trivially translated to CL form (17%)
  - 68 can be translated/reformulated to CL form (32%)
  - 14 involve n-tuples etc – not further considered (7%)
- 269 theorems passed to our system
- All theorems in remaining 4 chapters involve real numbers and n-tuples

Results for First 12 Chapters



Results for First 12 (of 16) Chapters (2)

- Around 1/3 of theorems proved
- Theorems proved *fully automatically*, no guiding at all
- Percentage ranges 5%-75%
- Percentage drops at final chapters

Related Work

- Quaife's work (1990) used a resolution prover
- Larry Wos and Michael Beeson (2012) used a resolution prover
- Better results, but both guided the resolution prover
- Julien Narboux (2006) used Coq

Example Again: Proof Generated by ArgoCLP

Let us prove that $p = r$ by reductio ad absurdum.

1. Assume that $p \neq r$.
   2. It holds that the point A is incident to the line q or the point A is not incident to the line q (by axiom of excluded middle).
      3. Assume that the point A is incident to the line q.
         4. From the facts that $p \neq q$, and the point A is incident to the line p, and the point A is incident to the line q, it holds that the lines p and q intersect (by axiom ax_D5).
         5. From the facts that the lines p and q intersect, and the lines p and q do not intersect we get a contradiction.

         Contradiction.
      6. Assume that the point A is not incident to the line q.
         7. From the facts that the lines p and q do not intersect, it holds that the lines q and p do not intersect (by axiom ax_nint_l_l_21).
         8. From the facts that the point A is not incident to the line q, and the point A is incident to the plane α, and the line q is incident to the plane α, and the point A is incident to the line p, and the line p is incident to the plane α, and the lines q and p do not intersect, and the point A is incident to the line r, and the line r is incident to the plane α, and the lines q and r do not intersect, it holds that $p = r$ (by axiom ax_E2).
         9. From the facts that $p = r$, and $p \neq r$ we get a contradiction.

         Contradiction.

Therefore, it holds that $p = r$.

This proves the conjecture.

Further Improvement

- Improving the quality of readable proofs may involve:
  - detecting (and omitting) trivial parts
  - avoiding a single uniform presentation scheme
  - using a wider language
  - even introducing small imperfections and typos!

Related Work

- Some methods for proving in geometry (Chou, 1990's)
- Isabelle/Isar (Wenzel, 2004) - already rather readable
- Coq (Corbineau, 2008)
- From Coq to natural language (Guilhot, Naciri, Pottier, 2003)
- "Formal proof sketches" from Mizar proofs (Wiedijk)
- "Mathematical Vernacular" - a formal language for writing readable proofs (Wiedijk)
- From tableaux-based proofs to natural language (Delahaye, Jacquel, 2012)
- Grammatical Framework (GF) – logic-based natural language processing (Ranta, 2011)

Conclusions and future work

- The presented framework can help in formalizing mathematical textbooks
- The framework can be used as an assistant to human mathematicians or in education
- There is a room for further improvements of the framework
- There is a room for improvements of "readable proofs"