MATHPROVE Dataset

Mathematical Problem-solving Dataset of Lessons and Exercises

Martin Kolář PhD et PhD

AITP 2024

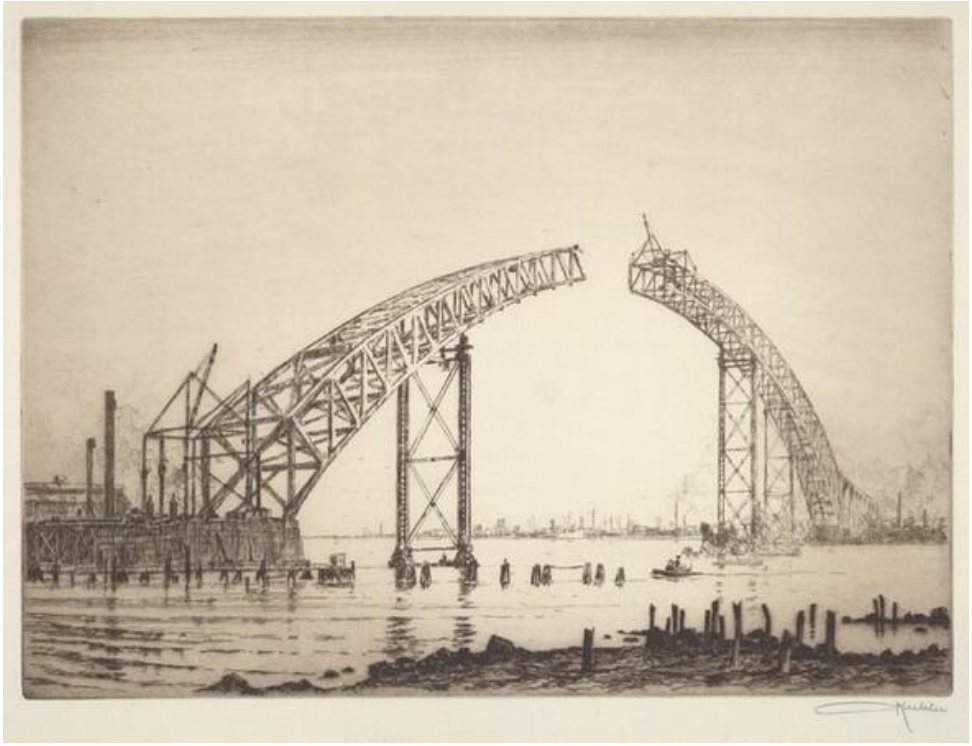
Hypothesis

Human Level Theorem Proving AI will be achieved by training models on materials humans use:

Textbooks	Already happening, in a naïve next-character way
Research papers	Not achievable through formal proof systems (Lean, Coq)
Trial and Error	Reinforcement Learning

Motivation

On one hand, we are building a huge body of proofs in formal languages, so that we can train more and more complex models to perform more and more advanced mathematics



On the other hand, we are training LLMs on huge corpora for next-token prediction, and generating natural-language proofs with RLHF

Machine-readable Theorem Proving

Math Word Problems

Methodology

Select a dozen textbooks

Select best pdf to LaTex tool

Convert, manually validate extracted data

Build data processing pipeline with regex and LLM tools to make structured data

Aggregate data

Test baseline

Implement grader and grade benchmark

Dataset

4 books

57 lessons

3091 blocks of problems

5221 exercises with worked solutions

2'508'526 chars of problems/solutions

```
14.8. Prove Theorem 14.1: Suppose $S$ is
a finite poset with $n$ elements. Then
there exists a consistent enumeration
$f: S \rightarrow\{1,2, \ldots, n\}$.
```

(a) LaTeX code of the example problem question segment

Prove Theorem 14.1: Suppose S is a finite poset with n elements. Then there exists a consistent enumeration $f : S \to \{1, 2, \ldots, n\}$.

(b) Compiled LaTeX of example problem question segment

The proof is by induction on the number n of elements in S. Suppose $n = 1$, say $S = \{s\}$. Then $f(s) = 1$ is a consistent enumeration of S. Now suppose $n > 1$ and the theorem holds for posets with fewer than n elements. Let $a \in S$ be a minimal element. (Such an element a exists since S is finite.) Let $T = S\backslash\{a\}$. Then T is a finite poset with $n - 1$ elements and hence, by induction, T admits a consistent enumeration; say $g : T \to \{1, 2, \ldots, n-1\}$. Define $f : S \to \{1, 2, \ldots, n\}$ by:

$$f(x) = \begin{cases} 1, & \text{if } x = a \\ g(x) + 1 & \text{if } x \neq a \end{cases}$$

(c) Compiled LaTeX of example problem solution segment

Dataset

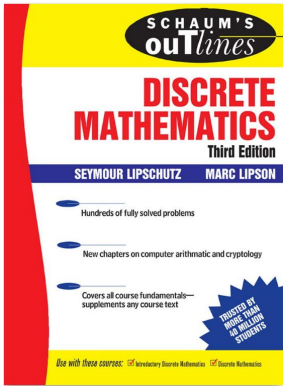
First dataset with lessons on techniques to be used in problems

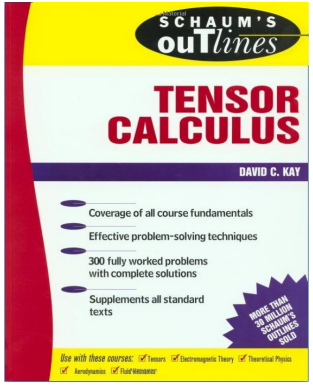
First dataset with university-level problems with worked solutions

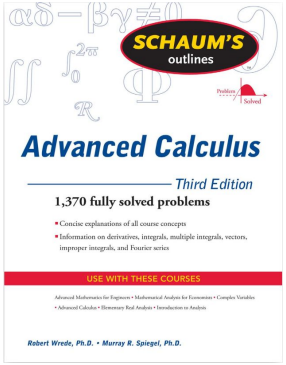
280 figures in problems and solutions

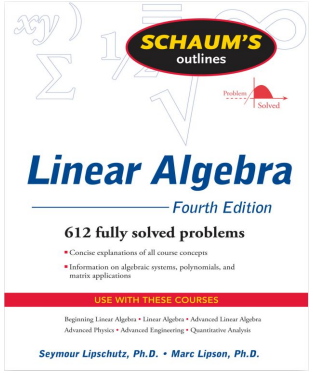
608 figures in lessons

Distributed as CC BY-NC, for Research use only, all rights remain with original authors









Usage

1) clone github.com/mrmartin/MATPROVE

2) get question and answer:

```
book_json, lesson_id, _, just_question, work_and_answer = book_qas[n]
```

3) get lesson:

```
contents, _, _ = json.load(book_json)

lesson = contents[lesson_id]['lesson']
```

Framework

pdf to LaTex	Mathpix
LaTex into chapters	custom processor (python regex)
Chapter into lesson/exercises/solution	custom processor (python regex)
Exercise into problem/solution	GPT-4

Code ready to add more textbooks with ~8 hours work per textbook

Baseline

Problem to solution	GPT-4
Proposed solution to real solution grading	GPT-3.5

Condition	**Correct Answers**	**Accuracy**
Without Lesson	3210	61.4%
With Lesson	3595	68.9%

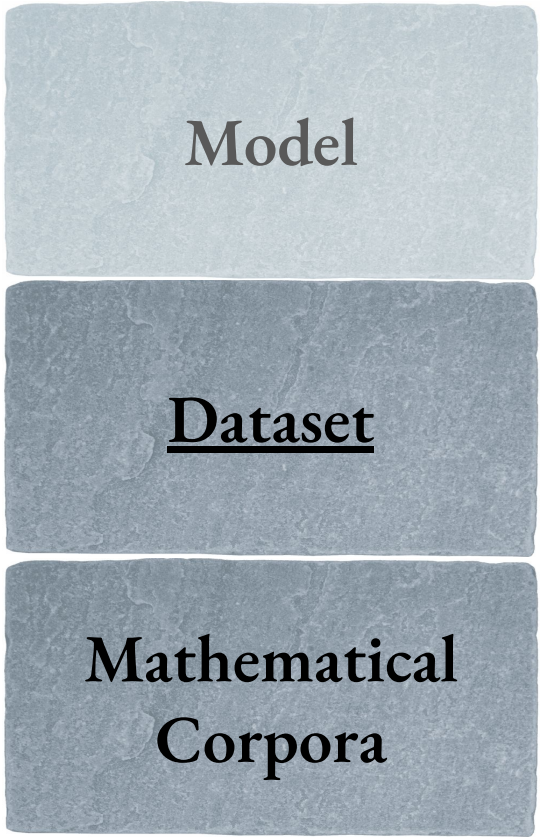
Conclusion

Hypothesis:

> Human Level Theorem Proving AI will be achieved by training models on materials humans use

Contribution:

- Dataset of lessons and worked problems in university-level mathematics
- Pipeline for converting more textbooks into this training format
- Benchmark through LLM grading



Thank You!

MATHPROVE Dataset

Mathematical Problem-solving Dataset of Lessons and Exercises

Martin Kolář PhD et PhD

AITP 2024

github.com/mrmartin/MATPROVE