Project Proposal: Automated Correction of Math Assignements

Mirek Olšák and David Stanovský

Abstract

We propose to build a system for automated correction of student assignments in pure mathematics that integrates handwritten solution recognition with evaluation of mathematical correctness and writing clarity. Using a corpus of hundreds of assignments from a single linear algebra course from approximately 300 students of mathematics, each consisting of handwritten submissions and detailed grader feedback, we propose to train and evaluate a model capable of assigning grades and generating concise, constructive comments. This project explores multimodal learning, combining visual recognition, natural language processing and formal verification, with the goal of supporting scalable and consistent assessment in mathematics education.

0.1 Introduction

Grading mathematical assignments is a time-consuming task that requires careful evaluation of both correctness and clarity of reasoning. While automated grading systems exist for structured or multiple-choice problems, open-ended handwritten solutions remain a significant challenge. Linear algebra, in particular, involves symbolic manipulation, structured reasoning, and precise notation, making it a suitable domain for studying automated assessment.

This project aims to develop a system that can process handwritten student submissions, evaluate their correctness, assess the quality of mathematical exposition, and produce both a grade and a short feedback comment. Such a system could assist instructors, provide faster feedback to students, and improve grading consistency.

0.2 Dataset

Our current dataset consists of approximately 50 homework assignments collected from around 300 students in a first year linear algebra course for students of mathematics, and about 6000 pages of graded exams on linear algebra. Each assignment includes:

- Handwritten student submissions (scanned images or digital ink),
- Corresponding grader evaluations, including numerical grades,
- Written feedback comments from graders.

The assignments focus on core topics in linear algebra, such as vector spaces, linear transformations, matrix operations, eigenvalues. The tests include straightforward numerical exercises, textbook proofs, advanced problems. The dataset is multimodal, pairing visual input (handwriting) with structured and unstructured textual annotations.

0.3 Problem Formulation

We define the task as a joint prediction problem with three outputs:

1. **Correctness score**: A numerical grade reflecting solution accuracy,

2. **Writing quality score**: An assessment of clarity and mathematical style,
3. **Feedback generation**: A short natural-language comment summarizing strengths and weaknesses.

0.4 Methodology

We propose to combine OCR to convert handwritten solutions into structured representations, followed by language models that interpret the mathematical content and extract reasoning steps. We incorporate formal verification techniques to check correctness by comparing derived results against expected properties and solutions. Finally, the system integrates these signals to predict grades and generate concise feedback comments.

Additionally, during real exam deployments, the system will incorporate extended human feedback from graders, providing additional supervisory signals that can be used to further refine the models and improve performance over time.

0.5 Evaluation

The system will be evaluated along three dimensions:

- **Grading accuracy**: Agreement between predicted and human-assigned grades,
- **Error detection**: Precision and recall of identified mistakes,
- **Feedback quality**: Human evaluation of generated comments for usefulness and clarity.

0.6 Challenges

Several challenges are anticipated:

- Variability in handwriting and notation,
- Ambiguity in partially correct solutions,
- Subjectivity in grading writing quality,
- Limited dataset size relative to model complexity.