Generating Elementary Integrable Rational Expressions

DG4D^3 – April 20, 2023

RASHID BARKET, MATTHEW ENGLAND, JUERGEN GERHARD

COVENTRY UNIVERSITY & MAPLESOFT

BARKETR@COVENTRY.AC.UK

Overview

1. Motivating the data generation method
   a) Current work
   b) Issues with current data generation

2. The Risch Algorithm
   a) Overview of the algorithm and definitions

3. Using the Risch Algorithm to generate integrals
   a) Trager-Rothstein Method
   b) Hermite Reduction

Elementary Functions

Definition: A function made by taking sums, products, roots and compositions of finitely many polynomial, rational, trigonometric, hyperbolic, and exponential functions

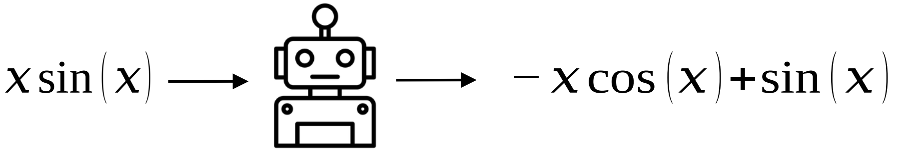
Goal: Discover elementary functions that when integrated, also result in an elementary function

(known as the logarithmic integral function)

Motivation

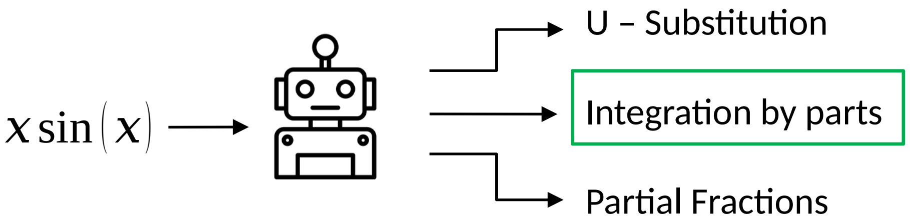
Deep Learning for Predicting an Integral

Use common NLP techniques (LSTMs, Transformers) to predict an integral given a math expression as input. This method is not as applicable to Computer Algebra Systems.



Machine Learning for Algorithm Selection

Perform feature engineering on the dataset to get relevant features for the problem. Using classical machine learning methods, choose the best algorithm to use when many are available.



Deep Learning for Symbolic Mathematics

Lample G, Charton F (Meta AI research)

Used transformer to perform integration and solve $1^{st}/2^{nd}$ order differential equations

Developed a method to generate expressions and have a labelled dataset

In testing, the model scored between 95-100% depending on the testing dataset and was able to solve some integrals that Maple, Mathematica could not

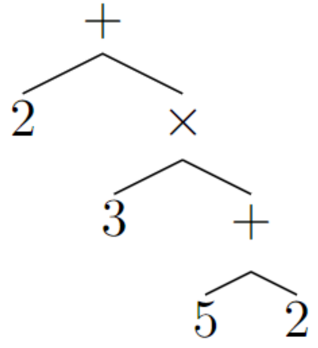
Current Methods – Random Expressions

Deep Learning for Symbolic Mathematics - Lample G, Charton F (Meta AI research)

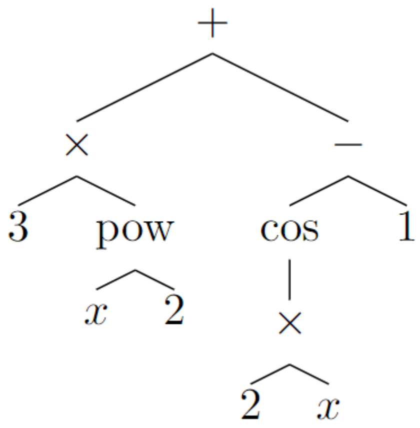
Mathematical expressions can be represented as trees:

- operators and functions as internal nodes
- numbers, constants and variables as leaves

$$2 + 3 \times (5 + 2)$$



$$3x^2 + \cos(2x) - 1$$



Creating a Labelled Dataset

Deep Learning for Symbolic Mathematics - Lample G, Charton F (Meta AI research)

- FWD: Differentiate an expression f to get f' and add the pair (f', f) to the dataset
- BWD: Integrate an expression through a Computer Algebra System (CAS) f to get F and add the pair (f, F) to the dataset
- IBP: Given two expressions f and g, calculate f' and g'. If $\int f'g$ is known then the following holds:

$$\int fg' = fg - \int f'g$$

Criticisms

1. In the BWD method, larger expressions generated are generally not elementary integrable
2. For both BWD and FWD methods, they create a pattern of expressions that are easier to predict
3. Data leakage
   a. For IBP, only 25% of the data is "unique"
   b. For FWD, only 35% of data is "unique"
   c. E.g: in the training and in the testing

The Risch Algorithm

- The algorithm takes as input an expression and outputs its’ anti-derivative
  - Anti-derivative must be elementary, otherwise it won’t output anything

- Let where is an algebraically closed field, and is an elementary extension of . Let . Then the Risch algorithm does the following:

$$\int \frac{a}{b} = \int P + \int \frac{R}{b}$$

Rational Part

Polynomial Part

n = 1

$$\int \frac{\theta^2+(2x+1)\theta+2x+1}{x(\theta+1)} = \int \frac{1}{x}\theta+3+\int \frac{1}{x(\theta+1)} = \frac{1}{2}\ln(x)^2+3x+\ln(\ln(x)+1)$$

Coventry University

The Risch Algorithm

- The algorithm takes as input an expression and outputs its' anti-derivative
  - Anti-derivative must be elementary, otherwise it won't output anything

- Let where is an algebraically closed field, and is an elementary extension of . Let . Then the Risch algorithm does the following:

$$\int \frac{a}{b} = \int P + \int \frac{R}{b}$$

$\int P$ → Polynomial Part

$\int \frac{R}{b}$ → Rational Part

n = 1

$$\int \frac{\theta^2 + (2x+1)\theta + 2x + 1}{x(\theta+1)} = \int \frac{1}{x}\theta + 3 + \int \frac{1}{x(\theta+1)} = \frac{1}{2}\ln(x)^2 + 3x + \ln(\ln(x)+1)$$

But First, Some Important Terms…

- The **resultant** of polynomials A and B, denoted as res(A, B), is a function that outputs another polynomial. This polynomial is made from looking at the roots of A and B
- A polynomial A is said to be **square-free** if it does not have as a divisor any square of a non-constant polynomial i.e. gcd(A, A')=1

E.g: is square-free whereas is *not* square-free

The Procedure

Goal: Produce a rational expression such that is elementary integrable

Idea: Given a fixed denominator , find all functions such that is elementary integrable

1. Choose a function b in for the denominator and then create a generic numerator
2. Convert   into partial fraction form, setting each numerator in the partial fraction to a generic function
3. Check to see if is square-free:
   a) If is square-free, apply the **Trager-Rothstein method**
   b) If is not square-free, apply **Hermite Reduction**

Trager-Rothstein Method

- **Key Theorem**: is elementary integrable The roots of the Trager-Rothstein resultant are constant.
- The Trager Rothstein resultant is
- Choose ourselves. Convert into partial fraction form and set a generic numerator for each fraction.
- We use the key theorem to find what the numerators must be

Trager-Rothstein Method

Example: let

1. Compute the partial fraction of

2. Calculate . The roots are

3. By the key theorem, and for constants

Demo: Let , then our expression is

Coventry University

Hermite Reduction

- Goal: Integrate a proper rational expression such that gcd and

1. Compute the Square-Free factorization for
   a) Define as
2. Solve the Diophantine Equation for
3.
4. **Key Idea**: The last integral, combined, with the results from TR-method, will give us numerators that make integrable

Hermite Reduction

Example: Let

1. Partial Fraction form:
2. Hermite Reduction:
3. TR-method: The root is . Solving for , we get:

Demo: , our expression is

Benefits

1. We can make integrand expressions as long or as short as we want. This gives a rich variety of data
   a) Consider the previous example, can be as complicated or as simple (even 0) as we wish

2. We avoid the problem of similar expressions
   a) Previous methods have a higher likelihood of producing something like and in the same dataset.

Thank you! Questions?