Proposal for a Report on the Risch Algorithm for Symbolic Integration and Implementation in the SymPy Computer Algebra System

Aaron Meurer

November 3, 2010

Contents

1 Summary

For this project, I am going to study the Risch Algorithm for symbolic integration of transcendental[1] equations. The Risch Algorithm is a complete

[1]Transcendental means that the functions handled by this sub-part of the algorithm cannot contain algebraic functions. More or less, this means that the function cannot contain radical expressions like $\sqrt{x+1}$ or $\sqrt[3]{\ln x}$. However, it can contain exponentials (e^x), logarithms ($\ln x$), or trigonometric functions ($\sin x$).

algorithm for computing elementary[2] antiderivatives, or proving that no such antiderivative exists. Over the summer of 2010, I worked under the Google Summer of Code program implementing the transcendental part of the algorithm in SymPy, a computer algebra system (CAS) written in the Python programming language.

Integration is a fundamental operation in mathematics. Most sciences that apply mathematics to themselves use calculus, which will invariably involve integration and integrals. The ability to algorithmically compute symbolic integrals is therefore of extreme practical importance. Mathematically, it is also of great interest that there exists an algorithm that not only can compute elementary symbolic integrals, but also that can prove that no such one exists when that is the case.

2 Outline

Over the summer of 2010, I worked on implementing the transcendental part of the Risch Integration Algorithm in SymPy, an open source computer algebra system written in Python. The algorithm is very complex and difficult to implement, and requires some proficiency of advanced mathematics to fully understand. This report will contain an overview of my work, and will also look at some of the other implementations of the algorithm in other open source computer algebra systems.

Integration is one of the two important operators from calculus, the other being differentiation. Unlike differentiation, however, symbolic integration (i.e., indefinite integration) of elementary functions is not a straightforward process. The methods taught in calculus, such as integration by substitution, integration by parts, trigonometric integration, and trigonometric substitution are only heuristics that can be applied to a special class of elementary functions.

For example, a calculus student would have a hard time computing

$$\int \frac{\left(1+e^{x^2}\right)\left(4x^3e^{x^2}-4x^2e^{x^2}\ln x - x + 1 - xe^{x^2}+e^{x^2}\right)}{x\left(\ln x - x\right)^2}\,dx \tag{1}$$

[2]If a function is elementary, roughly speaking, it means that it can not be represented as a combination of exponentials, logarithms, powers, and trig functions by addition, subtraction, multiplication, division, composition.

even though the solution is the relatively simple

$$\frac{\left(1+e^{x^2}\right)^2}{x-\ln x}. \tag{2}$$

The problem is further complicated by the fact that, unlike the case with differentiation, not all elementary function has an elementary antiderivative[3]. For example, the function

$$\int e^{-x^2}dx \tag{3}$$

is not elementary. Up to a constant factor, equation 3 is known as the error function, and used heavily in statistics[4].

It turns out that the Risch Algorithm not only gives a complete algorithm for symbolic integration, even for integrals as complex as the one given in equation 1, but it can also prove that no elementary antiderivative can exist for the integral, as is the case with equation 3.

The algorithm I implemented in SymPy can handle both of these case, meaning that it can produce equation 2 given equation 1, and it can prove that equation 3 is non-elementary.

3 Outcome

The report will be typset using the LaTeX typesetting system, and the bibliographies will be formatted automatically using BibTeX. The format will be a standard report format. It will have an abstract, an introduction, and sections detailing the different parts of the report.

[3]By the Fundamental Theorem of Calculus, indefinite integration is the inverse of differentiation, hence, it is sometimes also called antidifferentiation

[4]In particular, the error function represents the cumulative distribution function of the normal distribution (i.e., a bell curve), and its values are used to calculate probabilities. The fact that this function is non-elementary implies that statistical computing packages must use numerical techniques to calculate these values

4 Research Strategies

4.1 Primary Research

My main source for the algorithm was the textbook by Manuel Bronstein [3]. I have read most of this book , and have completed implementing most of the pseudocode algorithms given in it. Another important primary source will be the paper by Robert H. Risch [7], in which he originally posed a complete solution to the problem of integration in finite terms.

Because my report will also be on my implementation of the algorithm, I will focus on my own source code from the summer. I also plan on looking at implementations of the same algorithm in other open source CASs and comparing them to my own.

4.2 Secondary Research

For citation, I plan on using the default citation format in the LaTeX article document class, using BibTeX to format the citations[5].

Aside from the textbook, I have other sources. Moses [6] gives a history of the solution to the problem. Bronstein [2] is by the same author as and is referenced by [3]. The paper contains an algorithm used in integrating tangents that was not included in the textbook. Davenport [4] and Kauers [5] detail the algebraic part of the Risch Algorithm, which I may be interested in implementing in SymPy after I finish with the transcendental part. Adamchik [1] and Roach [8] detail methods for algorithmic definite integration, which would also be a natural step forward after completing the transcendental algorithm.

5 Timeline

Dates in bold are official due dates from the syllabus (revision as of October 28, 2010).

- **October 29, 2010**: Draft of proposal due in class. Peer reviews in class.
  - Finish the proposal. Make changes based on peer review feedback.

[5]See question 1 in section 6 below.

- **November 3, 2010**: Proposal (this document) due in class.
  - Start doing research. Most of the research was already done last summer, so this involve collecting the research together in a form suitable for the report.
  - Start writing the report.
- **November 8, 2010 - November 12, 2010**: Conference week.
  - At this point, I should have enough of the report done so that I will have questions to bring to the conference. Also, I should have a very rough draft ready by this point to also bring to the conference.
- **November 12, 2010**: Progress report memo 1 due in class. The bibliography must contain at least five secondary sources by this point[6].
  - Have something written for all the sections of the report. Have all secondary sources that will be used in the report.
- **November 19, 2010**: Progress report memo 2 due in class.
  - Finish most major sections of the report. Start working on the presentation.
- **November 24, 2010**: Draft of technical report due in class. Peer reviews in class.
  - Make changes from peer review feedback.
- *November 25, 2010 - November 28, 2010*: Thanksgiving break.
  - All but major parts of the report should be done by this point. Most parts of the presentation should be ready.
- **December 1, 2010**: Peer reviews in class.
  - Make changes from peer review feedback.
- **December 3, 2010**: Peer reviews in class.

[6]This won't be a problem, as I already have five sources. See the section 4.2 above and the References section below.

- Make changes from peer review feedback.
- Have the presentation and the report finalized.

- **December 6, 2010**: Presentations in class.
- **December 8, 2010**: Presentations in class.
- **December 10, 2010**: Presentations in class. Final technical report and presentation materials due in class.

6 Questions and Concerns

1. I couldn't actually figure out what the default citation format for LaTeX/BibTeX is called (like APA, MLA, etc.). Any idea what it is?

References

[1] V. S. Adamchik and O. I. Marichev. The algorithm for calculating integrals of hypergeometric type functions and its realization in reduce system. In *ISSAC '90: Proceedings of the international symposium on Symbolic and algebraic computation*, pages 212–224, New York, NY, USA, 1990. ACM.

[2] M. Bronstein. Simplification of real elementary functions. In *Proceedings of the ACM-SIGSAM 1989 international symposium on Symbolic and algebraic computation*, page 211. ACM, 1989.

[3] M. Bronstein. *Symbolic integration I: transcendental functions*. Springer Verlag, 2005.

[4] James Davenport. Integration in finite terms. *SIGSAM Bull.*, 18(2):20–21, 1984.

[5] M. Kauers. Integration of algebraic functions: a simple heuristic for finding the logarithmic part. In *Proceedings of the twenty-first international symposium on Symbolic and algebraic computation*, pages 133–140. ACM, 2008.

[6] J. Moses. Symbolic integration: the stormy decade. *Communications of the ACM*, 14(8):548–560, 1971.

[7] R.H. Risch. The Problem of Integration in Finite Terms Trans. In *Amer. Math. Soc*, volume 139, pages 167–189, 1969.

[8] Kelly Roach. Meijer g function representations. In *ISSAC '97: Proceedings of the 1997 international symposium on Symbolic and algebraic computation*, pages 205–211, New York, NY, USA, 1997. ACM.

A very thorough and well-written proposal -- 10/10