Incubating Mathematical Databases

David Roe

Department of Mathematics
MIT

July 7, 2025
LuCaNT 2

Database of databases

MathBases (mathbases.org)

- started at an AIM workshop in December 2023,
- grew out of Katja Berčič's MathDB (mathdb.mathhub.info),
- has an editorial board: Katja Berčič, Edgar Costa, Benjamin Hutz, David Lowry-Duda, David Roe, Ben Spitz, Adam Towsley,
- is a database that does not contain itself.

Mission

1. provide a searchable index of existing mathematical databases,
2. support mathematicians interested in creating new mathematical databases and improving existing ones.

Scope

MathBases currently indexes 128 databases; there are about 25 more on the to-add list.

Criteria for inclusion of a dataset

1. publicly available,
2. mathematical and of interest to research mathematicians
3. consisting of examples of mathematical objects, rather than exposition or formal statements

Database topics

51. Combinatorics
25. Number theory
22. Group theory
19. Metric geometry
12. Algebraic geometry
6. Rings and algebras
5. Geometric topology
4. Functional analysis, linear algebra
3. Dynamical systems, mathematical physics, representation theory
2. Category theory, classical analysis and ODEs, commutative algebra, differential geometry
1. Algebraic topology, analysis of PDEs, complex variables, general mathematics, general topology, logic, numerical analysis, operator algebras
0. K-theory and homology, optimization and control, probability, quantum algebra, symplectic geometry, spectral theory, statistics

Expository articles

In addition to serving as an index, MathBases aims to collect articles aimed at mathematicians who are creating databases.

- The Create page outlines the process of creating a new database, with links to more details
- At the moment, all the articles are stubs, which we hope to expand.
- We have a process for submitting articles. This could be
  - fleshing out one of the stubs;
  - writing something new, like a handbook for running computations on google cloud or a guide on accepting contributions from anonymous users;
  - an improvement to an existing article.
- Please submit things so that I don't have to write it all!

Starting a new database

In addition to expository articles, we offer several additional resources if you are interested in making a database:

- A Zulip server (part of `code4math`) where you talk to others about mathematical databases. Feel free to join, even if you don't want to make a new database!
- We plan to add templates based on existing databases (the LMFDB and MathBases to start), since modifying a template is easier than starting from scratch.
- We hope to create a community around mathematical databases; conferences like LuCaNT are part of that.

Contributing

We'd love your help! Here are some ways to contribute:

- Add an existing database to the index,
- Tell your colleagues about the site,
- Contribute an article,
- Help improve the project infrastructure (improved search interface, testing, handling dead links, displaying when a database record was last updated...),
- Fill in more metadata for databases in the index,
- Report a problem,
- Help us make a logo,
- Tell us about a database-related conference or add an idea for a new database.

The Diophantine Library

Goal

Create an online resource to help people solve, study and classify Diophantine equations.

Currently in the planning stage: I welcome suggestions and collaboration! Everything that follows is aspirational.

Equation homepages

Entry point

Enter equation(s), get a dynamically generated homepage.

- Known solutions, either stored or computed on the fly (at first just over $\mathbb{Z}$ and $\mathbb{Q}$; eventually in number fields or S-integers)
- Whether the solution set is known/expected to be finite or infinite
- Code (Sage, Magma, Pari, etc) for finding more solutions
- Methods for proving solutions complete/empty (obstructions, descent, Chabauty-Coleman...)
- The named families the equation belongs to
- Structural invariants (dimension, genus, rank, gonality, local/Brauer-Manin obstructions, L-function, subvarieties...)
- Connections to existing databases in the LMFDB: elliptic curves, modular curves, genus 2 curves....

Families of Diophantine equations

Entry point

Browse/search for families of equations.

- References to the literature and to computer algebra systems
- Transformations into a standard form
- Solutions to individual equations

Examples

Linear systems
Pythagorean triples
Pell equations
Elliptic curves
Thue: $f(X, Y) = n$

Gen. Fermat: $aX^p + bY^q = cZ^r$
Ramanujan-Nagell: $X^2 + d = k^N$
Thue-Mahler: $f(X, Y) = p_1^{N_1} \dots p_s^{N_s}$
$X^3 + Y^3 + Z^3 = n$
Erdös-Straus: $\frac{4}{n} = \frac{1}{x} + \frac{1}{y} + \frac{1}{z}$

Classifying Diophantine equations

Equation to name

For non-experts, knowing a name to Google and a transformation into a standard form is very valuable.

Given an input Diophantine equation, we want to be able to find a change of coordinates to transform it into the form of a stored family.

- Linear systems and conics are easy to classify, and there is a lot of existing software capable of finding solutions.
- For genus 1 curves not in Weierstrass form, finding an initial point can be hard. With a rational point, the rest is standard.
- In genus 2, 3 and 4, Igusa invariants and their descendants offer an identification over $\overline{\mathbb{Q}}$.
- In higher genus, want a hash function that can pick out potential curves to search for isomorphisms or birational maps. The L-function can serve as a foundation for such a hash; I'm happy to hear other ideas.

Hilbert's 10th problem

Beyond computational feasibility, by Matiyasevich's theorem we know that the solvability of some Diophantine equations is undecidable. In addition to the other information stored for each family, I plan to include examples of undecidable Diophantine equations and known results on the border between decidability and undecidability.

Contributing

I'm looking for collaborators! Here are some upcoming decisions and projects:

- What is a good name? Current contenders are the Diophantine Library and hilbert10.org.
- Create a prioritized list of families of Diophantine equations.
- Make an LMFDB-based template as a starting point (then host it on MathBases).
- Create a python package that collects classification code, processes user input, links to solvers in existing software (Sage, Magma, Pari...), and is available both as the website backend and independently.

I will be organizing workshops on this over the next several years; let me know if you're interested in attending!