* SLOW DOWN

— Big font on browser (check)

The L-functions and Modular Forms DataBase (LMFDB)

John Voight
Dartmouth College

Simons Collaboration on Arithmetic Geometry,
Number Theory and Computation
Annual Meeting
Simons Foundation, NYC
10 January 2019

Hodge Structures.

Outline of this talk

1. (Re)Introduce the LMFDB;
2. New classical modular forms functionality (private screening/world premiere!);
3. Future plans.

The LMFDB, in a nutshell

The LMFDB is a catalogue of mathematical objects arising in connection with the Langlands program, including

modular forms, L-functions, number fields, zeta zeros, elliptic curves, Maass forms, Artin representations, ...

The LMFDB was first conceived at an AIM workshop in 2007. The acknowledgements list 100 contributors. It has received substantial funding by the NSF, EPSRC, and the Simons Foundation, in addition to other smaller grants. (Thank you!)

It is now managed like a journal, with managing editors (John Cremona, John Jones, Drew Sutherland, and me) and a team of associate editors.

Nestor's computations for Serre's conj

Motivation and desiderata

- Number theory has historically been in part an experimental science: even Gauss used tables of primes. Big databases are now ubiquitous in science. [Theory leads computation in some respects.] Useful to have small tricky examples organized!
- By exhaustive computation, we can catch some issues ("sweep in all corners") and bugs in code ("compute all examples").
- Databases should be easily accessible online. (Not easy if data is scattered on homepages: learn format, install program, ... Worse still if a paper says "E-mail me if you want the data".)

"RANGE OF DATABASES IN ~~WILD~~"

WEB APPLICATION SIDE

- Data sets should be connected to one another. (Sometimes the point is to make connections across data sets.)
- Data should be reliable and have a glossary/appropriate annotations.
- Data should be searchable, so you can find particular examples of interest. (Before you prove something, it's helpful to know if there are counterexamples!)
- Data should have statistics, to notice patterns in aggregate.

[Aren't trying to replace hard computations done for other math reasons, ~~but~~ that's ~~not~~ how one]

Technical specs

no homepages
don't connect w/databases

- Size: The LMFDB includes a database of zeros of $\zeta(s)$ (1 TB), class groups of imaginary quadratic fields (2 TB), with the rest of the data totalling 600+ GB. More detailed stats are available.
- The LMFDB is currently hosted in the Google cloud rather than on a university-hosted server.
- The underlying database is Postgres, with good typing an object-relational database management system. (We moved from MongoDB, which is a document database; talk to Edgar Costa and David Roe for more details!)

$$\text{MiB} = \text{MEBIBYTE} = 2^{20} \text{ BYTES} = 1048576$$

$$\text{MB} = 10^6 \text{ BYTES.}$$

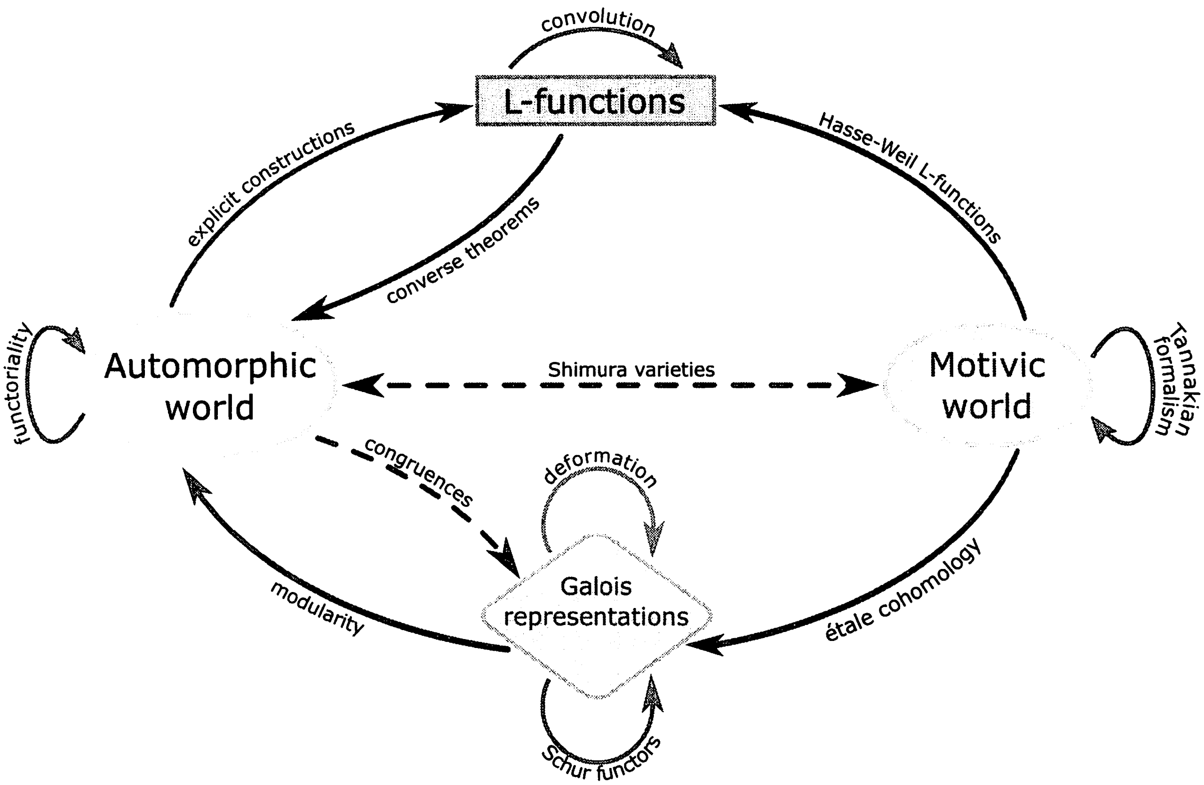
MongoDB vs Postgres

* One of Mongo's advantage was to be able to store generic objects without specified type, and this is perhaps why we chose it at the beginning, recently Postgres overcame this with the inclusion of the json and jsonb types.
* Integers limited to 64 bits, so got in the habit of storing many things as strings (which made sorting tricky among other issues)
* Indexes on lists didn't work as we wanted (e.g. list of ramified primes) - would work for "what are all the fields ramified at 5" but not "what are all the fields ramified at 2 and 3". Also inefficient (constructs query inefficiently since there are many more fields than primes)
* Mongo required more storage space since every document had to include the column names (quantify storage space change; quantify how much we're paying)
* Mongo to work efficiently required more RAM. Postgres can be faster with more RAM but works okay with less.
* Have to specify a schema in postgres in advance. This can be a hassle, but also can prevent errors (e.g. typos in column names, more complicated processing code because data isn't uniform, however also allows us to support in a standard way more complicated data structures, as we can enforce the format)
* Performance improvements, important since we don't know what kind of queries the user will request so it's hard to build all indexes. Partly because of lower storage, performance also improved even without indexes. Non-typical queries can be the most interesting mathematically.
* Postgres is a much more mature project, for example, it supports Transactions, so if you commit an error during a large data upload, things get revered to the original status

DIFFERENCE: RELATIONAL DATABASE REQUIRES A SCHEMA, WITH ~~ENTRIES~~ ROWS IN A TABLE
(DESCRIBING ATTRIBUTES/TYPES)
COLUMNS ARE SPECIFIED IN SCHEMA

IN CONTRAST, A DOCUMENT-ORIENTED DATABASE ~~[illegible]~~ W/ RECORDS THAT DESCRIBE THE DATA.

Big picture



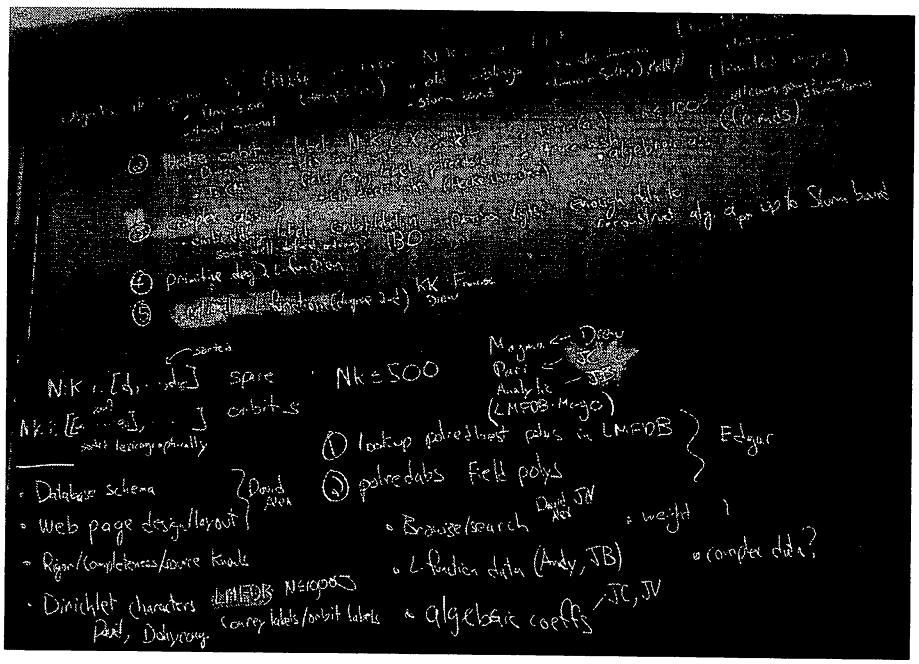
LMFDB Universe

~~X~~ SLOW DOWN

Classical modular forms

At a week-long workshop in August at MIT funded by the Simons Foundation, a significant upgrade was made to the classical modular forms. Edgar Costa and David Roe (Simons postdocs at MIT, led by Drew Sutherland) have continued the charge through this year. Thanks to Alex Best, Jonathan Bober, Andy Booker, John Cremona, and David Lowry-Duda.

lucky to have math and CS folks.



(Selective)

History

Perhaps the first systematic tabulation of modular forms was performed by Wada in 1971. The total computation time was about 300 hours on a TOSBAC-3000.

269	?
277	$(x^{11}-21x^9+152x^7-13x^6-445x^5+102x^4+450x^3-147x^2-56x+4)^2$
281	$x^{22}+50x^{20}+1081x^{18}+13285x^{16}+102611x^{14}+519821x^{12}+1748638x^{10}$ $+3865180x^8+5420534x^6+4487135x^4+1884424x^2+265883$
293	?
313	$(x^{12}-21x^{10}-5x^9+158x^8+63x^7-509x^6-241x^5+648x^4+272x^3-250x^2$ $-81x-4)^2$
317	?
337	$(x^{13}-x^{12}-22x^{11}+16x^{10}+182x^9-91x^8-697x^7+221x^6+1217x^5-215x^4$ $-808x^3+61x^2+175x)^2$
349	?
353	$x^{28}+57x^{26}+1429x^{24}+20814x^{22}+196024x^{20}+1256811x^{18}$ $+5621908x^{16}+17671372x^{14}+38764478x^{12}+58122566x^{10}$ $+57447036x^8+35475661x^6$ [illegible]

Eichler-Selberg trace formula

charpoly(T_p) on $S_2(\Gamma_0(q), \chi)$

$q \equiv 1 \pmod 4$, χ AT MOST quad.

History: Antwerp IV 1975

SOURCES AND RELIABILITY OF THE TABLES.

1. Tables 2-6.

Table 2 has been prepared for this volume by N.M. Stephens and James Davenport using the ATLAS at the Chilton laboratory. No errors are likely, but it is possible that in one or two cases a multiple of a generator has been listed rather than a generator.

Table 3 was prepared by J. Vélu; except for the eigenvalues of the W_q, the calculations have been performed independently by Stephens and Vélu, and no discrepancy has been observed; there are also strong requirements of consistency. The W_q-eigenvalues were calculated separately by D.J. Tingley, by operating on the one-dimensional homology; his results were incorporated in Vélu's tables by reading in cards. Consistency requirements make any errors in these particular calculations of Tingley hardly conceivable; copying errors are possible but unlikely. The identifying letters of the curves in this table have been added in manuscript; I hope they have never been jumbled.

Table 4 has been extracted from the unpublished 1966 Manchester thesis of F.B. Coghlan. Unless some mathematical error has remained undetected, the table should be complete.

The third column of Table 5 comes from a much larger table supplied by A.O.L. Atkin; the final column was supplied by D.J. Tingley, who has completely computed the action of the Hecke algebra on the 1-dimensional homology for N ≤ 300. His complete tables of eigenvalues will be published elsewhere. Table 6, too, is copied from a table of Atkin's; this table is new, note that the much shorter table of Deuring contains errors. Tables 5 and 6 were produced by typewriter rather than line-printer; it was very difficult to avoid slips.

11 = 11

A	0	-1	1	0	0	-	1	I1
B	0	-1	1	-10	-20	-	5	I5
C	0	-1	1	-7820	-263580	-	1	I1

14 = 2.7

A	1	0	1	-1	0	-	2, 1	I2,
B	1	0	1	-11	12	+	1, 2	I1,
C	1	0	1	4	-6	-	6, 3	I6,
D	1	0	1	-36	-70	+	3, [illegible]	I3,
E	1	0	1	-171	-874	-	18, 1	I18,
F	1	0	1	-2731	-55146	+	9, 2	I9,

15 = 3.5

A	1	1	1	0	0	-	1, 1	I1,
B	1	1	1	-5	2	+	2, 2	I2,
C	1	1	1	-10	-10	+	4, 4	I4,
D	1	1	1	-80	242	+	1, 1	I1,
E	1	1	1	-135	-660	+	8, 2	I8,
F	1	1	1	35	-28	-	2, 8	I2,
G	1	1	1	-110	-880	-	16, 1	I16,
H	1	1	1	-2160	-39540	+	4, 1	I4,

17 = 17

A	1	-1	1	-1	0	+	1	I1
B	1	-1	1	-6	-4	+	2	I2
C	1	-1	1	-1	-14	-	4	I4
D	1	-1	1	-91	-310	+	1	I1

19 = 19

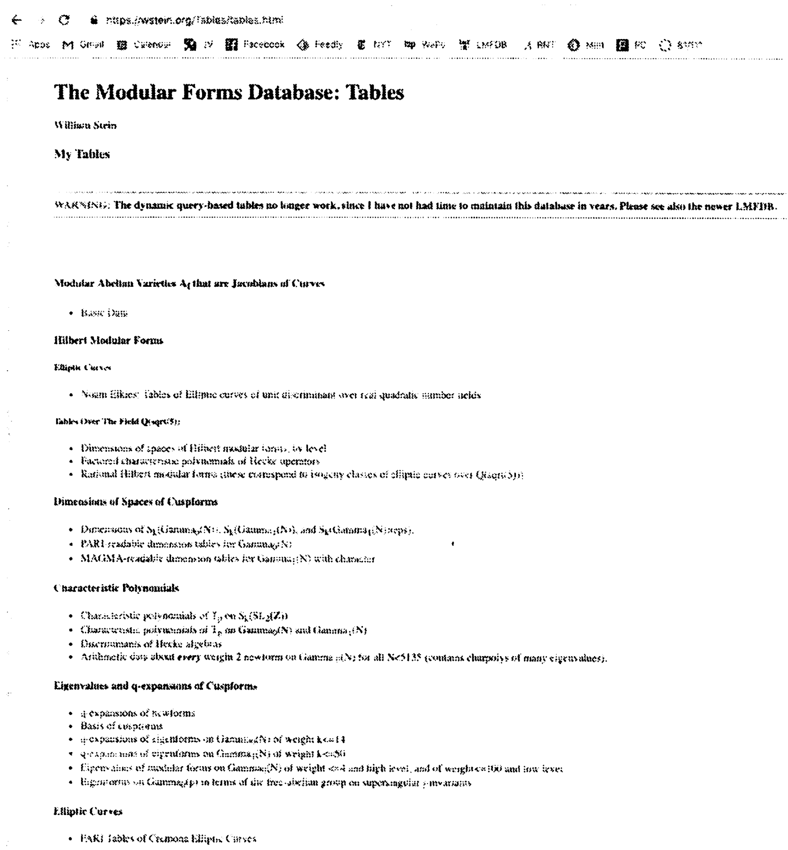
A	0	1	1	1	0	-	1	I1
B	0	1	1	-9	-15	-	3	I3
C	0	1	1	-769	-8470	-	1	I1

20 = 2.2.5

A	0	1	0	-1	0	+	4, 1	I[illegible],	[illegible]
B	0	1	0	4	4	-	8, 2	I[illegible],	[illegible]
C	0	1	0	-41	-116	+	4, 3	I[illegible],	[illegible]
D	0	1	0	-36	-140	-	8, 6	I[illegible],	[illegible]

exciting time for me as grad student.

Stein 2000s



https://wstein.org/Tables/tables.html

Apps Gmail Calendar IV Facebook Feedly NYT WaPo LMFDB BNT Mill PC

The Modular Forms Database: Tables

William Stein

My Tables

WARNING: The dynamic query-based tables no longer work, since I have not had time to maintain this database in years. Please see also the newer LMFDB.

Modular Abelian Varieties A_f that are Jacobians of Curves

- Basic Data

Hilbert Modular Forms

Elliptic Curves

- Noam Elkies' Tables of Elliptic curves of unit discriminant over real quadratic number fields

Tables Over The Field Q(sqrt(5)):

- Dimensions of spaces of Hilbert modular forms, by level
- Factored characteristic polynomials of Hecke operators
- Rational Hilbert modular forms (these correspond to isogeny classes of elliptic curves over Q(sqrt(5)))

Dimensions of Spaces of Cuspforms

- Dimensions of S_k(Gamma$_0$(N)), S_k(Gamma$_1$(N)), and S_k(Gamma$_1$(N),eps).
- PARI-readable dimension tables for Gamma$_0$(N)
- MAGMA-readable dimension tables for Gamma$_1$(N) with character

Characteristic Polynomials

- Characteristic polynomials of T_p on S_k(SL$_2$(Z))
- Characteristic polynomials of T_p on Gamma$_0$(N) and Gamma$_1$(N)
- Discriminants of Hecke algebras
- Arithmetic data about *every* weight 2 newform on Gamma$_0$(N) for all N<5135 (contains charpolys of *many* eigenvalues).

Eigenvalues and q-expansions of Cuspforms

- q-expansions of newforms
- Basis of cuspforms
- q-expansions of eigenforms on Gamma$_0$(N) of weight k<=14
- q-expansions of eigenforms on Gamma$_1$(N) of weight k<=50
- Eigenvalues of modular forms on Gamma$_0$(N) of weight <=4 and high level, and of weight<=100 and low level
- Eigenforms on Gamma$_0$(p) in terms of the free-abelian group on supersingular j-invariants

Elliptic Curves

- PARI Tables of Cremona Elliptic Curves

many cool useful things.

integrated Magma

Stein tables

Stein:		will have:
complete	418	8903
newspace data	9731	312948
newforms	~~237791~~ 84407	237791

LMFDB: Scope of data

To be systematic, our data includes two big "boxes":

- All spaces $S_k(\Gamma_1(N))$ with $N, k \in \mathbb{Z}_{\geq 1}$ and $Nk^2 \leq 4000$, and
- All spaces $S_k(\Gamma_0(N))$ with $N, k \in \mathbb{Z}_{\geq 2}$ and $Nk^2 \leq 40000$ (trivial character).

Stein:

Hecke stability.

With additional boxes to cover the spread, we easily majorize both Stein's database (and current LMFDB offerings).

Magma ( ), Pari (wt 1, 2x check) Belabas-Cohen, Booker (complex trace formula) L-function

Stein: Donnelly

- Data for 316 236 newspaces $S_k(\Gamma_0(N), \chi)$ (with 50 098 nonzero)
- 236 555 (Galois orbits of) newforms (48 324 with rational coefficients, 19 306 weight 1)
- 13 710 564 complex embedded newforms "serious data"
- Total size (including L-functions) about 300 GB

Many CPU years, CPU decade, not data you should have to recompute yourself.

order of magnitude more data

Search page and 11.2.a.a

explain labels.

20x ↑ in math size from LMFDB but only 1.5x disc space b/c improvements (a_n's past sturm bound).

check font size.

A sampling of new features (items 1–5)

sarnak suggestion

1. We bound our boxes by Nk^2, which is a good approximation to the analytic conductor $N||k||^2$ which governs the total complexity of the L-function (for more, talk to David Roberts). In particular, our data is complete in these boxes. Search page
2. Coefficients a_n are represented compactly: either sparse cyclotomic (1620.1.bp.a) or in terms of an LLL-reduced basis for the Hecke ring (6877.2.a.s). $\beta \leftrightarrow \nu$
3. We have complex eigenvalues when exact eigenvalues would be computationally infeasible (983.2.c.a), allowing for the computation of all L-functions, and can customize the display the embeddings (2412.1.b.b).
4. Self-twists (all rigorously proved), inner twists (indicated when proved or not), Atkin-Lehner eigenvalues, Hecke cutters available (2013.1.bm.c).
5. Can download the form to Magma to get more coefficients.

*SLOW DOWN
(check time)

Item 6: Trace tables (matching classical modular forms)

Dear John, please forgive a cold call.

With colleagues, I have been calculating local zeta functions for Calabi-Yau varieties in a family depending on t. In general this gives rise to a factor

$1 + aT + bpT^2 + ap^3T^3 + p^6T^4$.

For $t = 33 \pm 8\sqrt{17}$, the quartic factors in the form

$(1 - \alpha pT + p^3T^2)(1 - \beta T + p^3T^2)$

whenever 17 is a square mod p.

We have lists of data of the form $p, t, \{\alpha, \beta\}$ for when this happens, and we believe that the coefficients $\{\alpha, \beta\}$ are coefficients of a (Hilbert) modular form.

So finally I come to my question, which is: is there a practical way of extracting lists of coefficients of such modular forms from LMFDB with which to compare our lists?

We have a number of cases similar to the example above so we hope for a simple procedure that we can repeat as necessary.

Narrowing the search

- There is bad reduction at 17, and possibly at 2.
- The quartic factor is independent of the choice of $\sqrt{17}$ modulo p, so we guess that there is descent from $F = \mathbb{Q}(\sqrt{17})$ to $\mathbb{Q}$.
- We guess that the Hodge structure splits as

$$(1,1,1,1) = (0,1,1,0) + (1,0,0,1)$$

and accordingly the (likely) ℓ-adic Galois representation

$$\rho\colon \mathrm{Gal}_F \to \mathrm{GL}_4(\mathbb{Q}_\ell)$$

is the base change from $\mathbb{Q}$ to F of

$$\rho_f(-1) \oplus \rho_g$$

where f, g are classical modular forms of weights $2, 4$.
- We have a fair amount of eigenvalue data:

p	13	19	43	67	$\cdots$	
$a_p(f)$	2	-4	-4	-4	$\cdots$	34.2.b.a
$a_p(g)$	-42	60	508	-676	$\cdots$	34.4.b.a

Let's search for the form!

Item 7: Weight 1 forms

We treat weight 1 forms.

a. The data set majorizes the Buzzard–Lauder database.

b. The projective image types have been proved for all weight 1 forms, requiring a couple of new tricks. 124.1.i.a (smallest A_4), 148.1.f.a (smallest S_4), 633.1.m.b (smallest A_5); 3600.1.e.a (D_2, having RM and CM), 3997.1.cz.a (D_{285}).

c. Dimension data for subspaces by projective image type. Artin — 1161.1.i.

doubled # of A_5-exts.

look at those 0's!

give us a target 4K of 19K matched

A sampling of new features (items 7–10)

8. Can view dimension tables, with constraints. For example, rational forms.
9. Dynamic statistics: you can tabulate statistics for a specified subset of the data. For example, what are the statistics for weight and level for those forms having self-twist?
10. Provable analytic ranks (including non-self dual forms of analytic rank 1). We found a rational form (2.42.a.a) of weight 42 and a form (8.14.b.a) of weight 14 each having analytic rank 1.

Future plans

1. Annotations for interesting examples (IAS, March 2019).
2. LMFDB as like an OEIS (Online Encyclopedia of Integers Sequences) for L-functions.
   An L-function 2.2.56.1-32.1-c with 12 origins. genus 2 curve?! [Restriction of scalars?]
3. mod p Galois representations, Siegel paramodular forms, half-integer weight modular forms, hypergeometric motives, Belyi maps, small groups, ...

We hope the LMFDB will be a useful tool on your workbench!