Database tools for the large scale computation of Maximally Mutable Laurent polynomials

Giuseppe Pitton
Imperial College London

Joint work with Tom Coates (ICL) and Al Kasprzyk (Nottingham)

ICMS 2020, Brunswick
Session on Databases in Mathematics

Motivation

Context

Fanosearch Project: study Fano varieties using Mirror Symmetry.
[Coates, Corti, Galkin, Golyshev, Kasprzyk (arXiv:1212.1722)]

- Two sides of the mirror:

$$\left\{\begin{matrix}\text{some Laurent}\\ \text{polynomials}\end{matrix}\right\}\Big/\{\text{mutations}\} \Leftrightarrow \left\{\begin{matrix}\text{Fano}\\ \text{varieties}\end{matrix}\right\}\Big/\{\text{deformations}\}$$

- Laurent polynomials are simpler than Fano manifolds.

Motivation

Definition

Fix a lattice $N \equiv \mathbb{Z}^d$.

A *canonical Fano polytope* $P \subset N$ is a convex lattice polytope such that:

- the vertices $\mathcal{V}(P)$ are integral lattice points.
- The only lattice point in $\mathring{P}$ is the origin of the lattice.
- $\dim(P) = d$.

A special case

- If the polar polytope P^* is itself an integral lattice polytope, we call P *reflexive*.

Motivation

“Interesting” Laurent polynomials f:

- Given a canonical Fano polytope P,

$$\operatorname{Newt} f = P, \qquad f = \sum_{\gamma \in N \cap P} a_\gamma x^\gamma$$

Definition

Given a vector $w \in \mathrm{Hom}(N, \mathbb{Z})$, and $u \in \mathbb{C}[w^\perp]$, a *mutation* $\Phi : \mathbb{C}[N] \to \mathbb{C}[N]$ defined by:

$$\Phi : x^\gamma \mapsto x^\gamma u^{\langle w, \gamma \rangle}.$$

$$\Phi : f \mapsto \sum_{\gamma \in N \cap P} a_\gamma x^\gamma u^{\langle w, \gamma \rangle}.$$

$\Phi f \in \mathbb{C}[N]$ might constrain some a_γ.

Akhtar, Coates, Galkin, Kasprzyk [arXiv:1212.1785]

Motivation

Definition (Kasprzyk, Tveiten)

Let P be a canonical Fano polytope.
Consider

$$f = \sum_{\gamma \in N \cap P} a_\gamma x^\gamma$$

such that there exist $\Phi_1, \ldots, \Phi_r$ such that

$$\Phi_i f \in \mathbb{C}[N] \qquad \text{for } i = 1, \ldots, r$$

$$\text{fixes uniquely all values of } a_\gamma, \quad \gamma \in \partial P \setminus \mathcal{V}(P),$$

then we call f a *(rigid) Maximally Mutable* Laurent polynomial.

- if $\gamma \in \mathcal{V}(P)$, then $a_\gamma = 1$.
- If $\gamma = (0, \ldots, 0)$, then $a_\gamma = 0$.

Akhtar, Coates, Galkin, Kasprzyk [arXiv:1212.1785]

Motivation

Definition

Given a Laurent polynomial $f \in \mathbb{C}[x_1^{\pm 1}, \ldots, x_n^{\pm 1}]$, the *period* of f is:

$$\pi_f(t) = \frac{1}{(2\pi \mathrm{i})^n} \int_{|x_1|=1,\cdots,|x_n|=1} \frac{1}{1 - tf(x_1, \ldots, x_n)} \frac{\mathrm{d}x_1}{x_1} \cdots \frac{\mathrm{d}x_n}{x_n}, \qquad t \in \mathbb{C},$$

Moreover, if f is mirror to a Fano manifold,

$$\pi_f(t) = \sum_{k \geq 0} c_k t^k \qquad c_k \in \mathbb{Z}, \quad \text{for all } k.$$

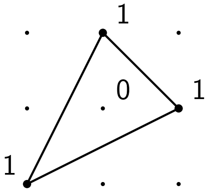
The sequence $(c_0, c_1, \ldots)$ is called the *period sequence* of f.

$$c_k = \operatorname{coeff}_1 f^k.$$

[Akhtar, Coates, Galkin, Kasprzyk (arXiv:1212.1785)]

Motivation

example

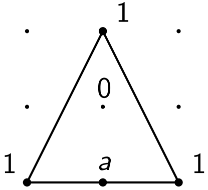


$$f = x + y + \frac{1}{xy}$$

$$\pi_f(t) = 1 + 6t^3 + 90t^6 + 1680t^9 + 34650t^{12} + 756756t^{15} + \dots$$

Motivation

example



$$f = y + \frac{1}{xy}\left(x^2 + ax + 1\right)$$

f is Maximally Mutable iff $a = 2$, and we have:

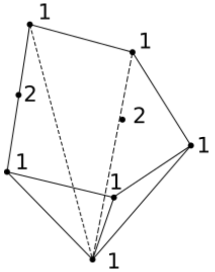
$$f = y + \frac{1}{xy}(x+1)^2 .$$

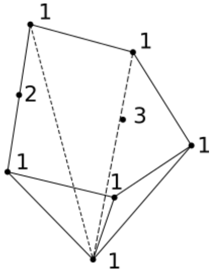
$$\pi_f(t) = 1 + 4t^2 + 36t^4 + 400t^6 + 4900t^8 + 63504t^{10} + \dots$$

Motivation

If $\dim P > 2$, P may have more than one MM Laurent polynomials.

example





$$f=\frac{1}{z}+z\left(\frac{(x+1)^2}{xy}+\frac{x^2+2x+1}{x}+y\right) \qquad g=\frac{1}{z}+z\left(\frac{(x+1)^2}{xy}+\frac{x^2+3x+1}{x}+y\right)$$

(left) $\pi_f(t) = 1 + 4t^2 + 60t^4 + 1120t^6 + 24220t^8 + \dots$

(right) $\pi_g(t) = 1 + 6t^2 + 90t^4 + 1860t^6 + 44730t^8 + \dots$

Computational aspects

Use case

Given a class of polytopes $\mathcal{U}$, for all $P \in \mathcal{U}$:

- find all Maximally-Mutable polynomials supported on P.
- Make a list of the period sequences.
- Compute other interesting quantities.

$\mathcal{U}$ can be very large

3-reflexives	3-canonicals	4-reflexives
4 319	674 688	473 800 776

Performance Requirements

- Ideally, as many cores as polytopes in $\mathcal{U}$.
- High single-core performance.
- High disk read/write rate.

Computational aspects

IO infrastructure

- Text/binary file dump.
  - As simple as possible.
  - Naturally concurrent.
  - Slow postprocessing.
- SQLite database.
  - High quality, simple to use SQL database.
  - Not designed for concurrency.
- Dedicated database server.
  - High performance, high concurrency.
  - Flexible data structures, SQL (Postgres, CockroachDB) or JSON (Mongo).
  - Non negligible set up and maintenance cost.
  - The HPC-database interface is of fundamental importance.

Our solution: in-house tooling for workers/database interaction [Coates, Kasprzyk (`https://bitbucket.org/pcas`)]

Example: canonical Fano 3-topes

Canonical Fano polytopes in 3 dimensions were classified by Kasprzyk [arXiv:0806.2604].
Up to $\mathrm{GL}(\mathbb{Z}, 3)$-equivalence, there are 674 688 canonical Fano 3-topes.
Target For each canonical Fano 3-tope P:

1. find all Maximally Mutable polynomials f supported on P.
2. For every such f, compute the period sequence.
3. Can we analise and understand this data?

Example: code metrics

The Maximally Mutables code is quite complex

The most demanding computational tasks include:

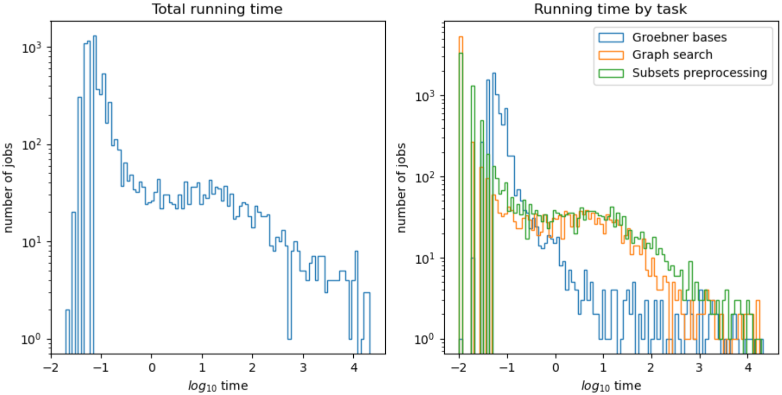
- Minkowski decompositions.
- Gröbner basis computations.
- Graph search algorithm.
- Several performance tricks.

Performance analysis and optimisation

- Code profiling (development).
- Runtime metrics (deployment).

Example: code metrics

Random sample of 10 000 polytopes.



Example: canonical Fano 3-topes

As of today, the search is 99.8% complete.
We found $\sim 170\,000$ Maximally Mutable Laurent polynomials.
These polynomials originate $\sim 8\,200$ period sequences.

- We can leverage the wealth of open source libraries for Data Science.
- Example: basic analysis of the period sequence data.
- Tools: `pandas` and `scikit-learn` libraries for python.

Example: canonical Fano 3-topes

Data Analysis

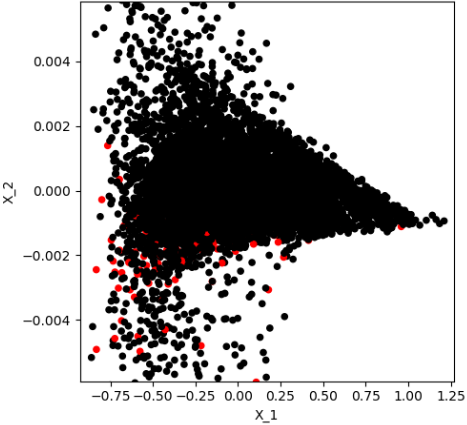
Difficulties in dealing with period sequences data:

- the coefficients of the period sequences grow quickly.
  $\hookrightarrow$ Work instead with $\log c_k$, then rescale so that
  $\log c_k \in [-1, 1]$
- The data is high-dimensional (in this computation, each datapoint is in $\mathbb{Z}^{14}$ ($[-1, 1]^{14}$ after rescaling).
  $\hookrightarrow$ Principal Component Analysis: project to linear subspaces that explain most of the variance in the data.
- Beware of low-dimensional projections of high-dimensional data.

Example: canonical Fano 3-topes

Low-dimensional projection

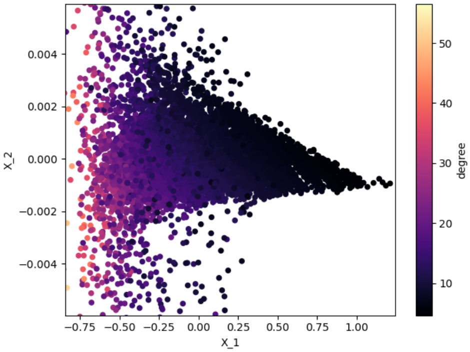
Every point represents a period sequence.
Period sequences of smooth manifolds are in red.



Example: canonical Fano 3-topes

Low-dimensional projection

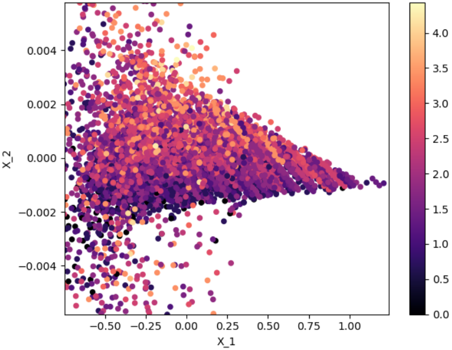
Every point represents a period sequence.
Anticanonical degree.



Example: canonical Fano 3-topes

Low-dimensional projection

Every point represents a period sequence.
Gorenstein index (log scale).



Perspectives

Opportunities and challenges ahead

- Databases are evolving quickly.
  - ↪ Automatic sharding (CockroachDB).
  - ↪ Optimised performance for SSDs.
  - ↪ Memory caches (e.g. Redis).
- Data analysis on $\sim$1 Tb.