The art of educated algorithmic guessing (with gfun)

Sergey Yurkevich

Inria Saclay, University of Vienna,
A. Bostan, H. Hauser

February, 2022

Guess the sequences

"Guessing – that's the important beginning of solving any problem."

- 1, 3, 9, 27, 81, 243,
- 0, 1, 3, 6, 10, 15,
- 1, 1, 2, 3, 5, 8,
- 1, 1, 2, 5, 14, 42,
- 1, 5, 73, 1445, 33001,

Guess the sequences

"Guessing – that's the important beginning of solving any problem."

- 1, 3, 9, 27, 81, 243, 729 3^n
- 0, 1, 3, 6, 10, 15,
- 1, 1, 2, 3, 5, 8,
- 1, 1, 2, 5, 14, 42,
- 1, 5, 73, 1445, 33001,

Guess the sequences

"Guessing – that's the important beginning of solving any problem."

- 1, 3, 9, 27, 81, 243, 729 — 3^n
- 0, 1, 3, 6, 10, 15, 21 — $(n^2 + n)/2$
- 1, 1, 2, 3, 5, 8,
- 1, 1, 2, 5, 14, 42,
- 1, 5, 73, 1445, 33001,

Guess the sequences

"Guessing – that's the important beginning of solving any problem."

- 1, 3, 9, 27, 81, 243, 729 — 3^n
- 0, 1, 3, 6, 10, 15, 21 — $(n^2 + n)/2$
- 1, 1, 2, 3, 5, 8, 13 — $F_{n+1} = F_n + F_{n-1}$
- 1, 1, 2, 5, 14, 42,
- 1, 5, 73, 1445, 33001,

Guess the sequences

"Guessing – that's the important beginning of solving any problem."

- 1, 3, 9, 27, 81, 243, 729 — 3^n
- 0, 1, 3, 6, 10, 15, 21 — $(n^2+n)/2$
- 1, 1, 2, 3, 5, 8, 13 — $F_{n+1} = F_n + F_{n-1}$
- 1, 1, 2, 5, 14, 42, 132 — $\frac{1}{n+1}\binom{2n}{n}$
- 1, 5, 73, 1445, 33001,

Guess the sequences

"Guessing – that's the important beginning of solving any problem."

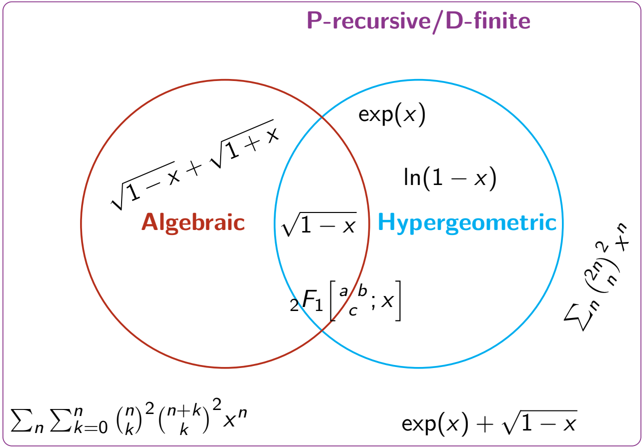
- 1, 3, 9, 27, 81, 243, 729 — 3^n
- 0, 1, 3, 6, 10, 15, 21 — $(n^2+n)/2$
- 1, 1, 2, 3, 5, 8, 13 — $F_{n+1} = F_n + F_{n-1}$
- 1, 1, 2, 5, 14, 42, 132 — $\frac{1}{n+1}\binom{2n}{n}$
- 1, 5, 73, 1445, 33001, 819005 — $\sum_k \binom{n}{k}^2\binom{n+k}{k}^2$

Importance of sequences

"Theory without practice is dead, and practice without theory is blind."

- Enumerative combinatorics (e.g. counting walks in the quarter plane)
- Algebraic number theory (e.g. irrationality proof of $\zeta(3)$)
- Analytic number theory (e.g. prime number theorem)
- Algorithmic number theory (e.g. integer factorization)
- Computer algebra (e.g. fast computation of terms in sequences)
- Algebraic geometry (e.g. counting points on curves)
- Interconnections (e.g. moonshine theory, or mirror symmetry)
- Applications in various sciences (e.g. uniqueness of the Clifford torus model)

P-recursive sequences and D-finite functions



A sequence $(u_n)_{n\geq 0}$ is **P-recursive**, if it satisfies a linear recurrence with polynomial coefficients:

$$c_r(n)u_{n+r} + \cdots + c_0(n)u_n = 0.$$

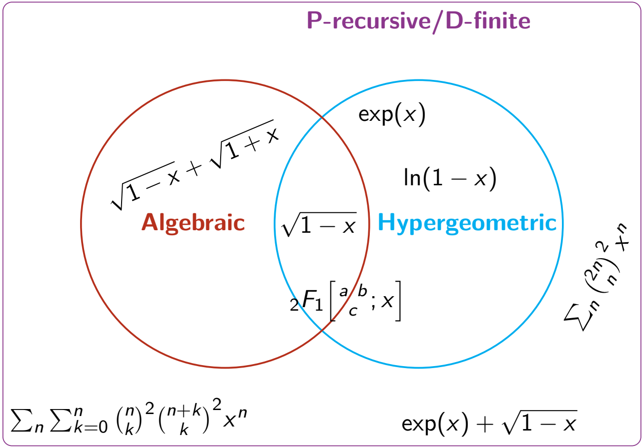
$(u_n)_{n\geq 0}$ is **hypergeometric** if $r = 1$.

Let $(\alpha)_n = \alpha\cdot(\alpha+1)\cdots(\alpha+n-1)$.

Then $u_n = \frac{(a)_n\cdot(b)_n}{(c)_n\cdot n!}$ satisfies

$$(c+n)(n+1)u_{n+1} - (a+n)(b+n)u_n = 0.$$

P-recursive sequences and D-finite functions



A power series $f(x) \in \mathbb{Q}[\![t]\!]$ is called **D-finite** if it satisfies a linear differential equation with polynomial coefficients:

$$p_n(x) f^{(n)}(x) + \cdots + p_0(x) f(x) = 0.$$

Let $(\alpha)_n = x \cdot (\alpha+1) \cdots (\alpha+n-1)$.

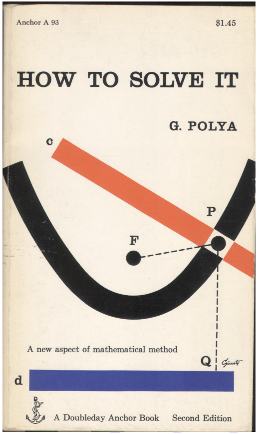
Then ${}_2F_1\left[{a\ b \atop c}; x\right] := \sum_{n \geq 0} \frac{(a)_n \cdot (b)_n}{(c)_n \cdot n!} x^n$ satisfies

$$x(1-x) f''(x) + (c-(a+b+1)x) f'(x) - abf(x) = 0.$$

The Guess-and-Prove approach

"First guess, then prove.
All great discoveries were made in this style."

- Experimental mathematics and "Guess-and-Prove" propagated by G. Pólya.
- Three step process:
  Generate data → Make conjectures → Prove them.
- Extremely fruitful when using a computer.
- P-recursive sequences/D-finite functions:
  ideal data structure for guessing.
- Find new structure.
- Find simpler formulas.
- Very efficient and easy in practice
  with computer algebra software (e.g. Maple).



The gfun package in Maple

- By Bruno Salvy and Paul Zimmermann in 1992. Constantly improved.
- Can guess a sequence from its first terms.
- Implemented closure properties and interplay between recursions and ODEs.
- Version 3.20 comes with Maple 2021

```
with(gfun);
```

[*Laplace*, *Parameters*, *algebraicsubs*, *algeqtodiffeq*, *algeqtoseries*, *algfuntoalgeq*, *borel*, *cauchyproduct*, *'diffeq*diffeq'*, *'diffeq+diffeq'*, *diffeqtohomdiffeq*, *diffeqtorec*, *guesseqn*, *guessgf*, *hadamardproduct*, *holexprtodiffeq*, *invborel*, *listtoalgeq*, *listtodiffeq*, *listtohypergeom*, *listtolist*, *listtoratpoly*, *listtorec*, *listtoseries*, *poltodiffeq*, *poltorec*, *ratpolytocoeff*, *'rec*rec'*, *'rec+rec'*, *rectodiffeq*, *rectohomrec*, *rectoproc*, *seriestoalgeq*, *seriestodiffeq*, *seriestohypergeom*, *seriestolist*, *seriestoratpoly*, *seriestorec*, *seriestoseries*]

- Newest version (3.84): perso.ens-lyon.fr/bruno.salvy/software/the-gfun-package/.

Guessing for P-recursive sequences

"Finished mathematics consists of proofs, but mathematics in the making consists of guesses."

- Given: $u_0, \dots, u_N$ terms of a sequence $(u_n)_{n\geq 0}$.
- Want: a *guess* for a linear recurrence relation for $(u_n)_{n\geq 0}$.
- Idea: Look for $r \in \mathbb{N}$ and $c_0(x), \dots, c_r(x) \in \mathbb{Q}[x]$ of some degree $d \in \mathbb{N}$ such that

$$c_r(j)u_{j+r} + \cdots + c_0(j)u_j = 0 \quad \text{holds for} \quad j = 0, \dots, N - r.$$

- Need to solve a *system of linear equations*, where the unknowns are the $(r+1)(d+1)$ coefficients of the polynomials $c_i(x)$.
- If $(r+1)(d+1) > N - r$, a non-zero solution trivially exists.
  If $(r+1)(d+1) \ll N - r$, no reason for a solution, except if $(u_n)_n$ is P-recursive.
- If no bounds on r and d, first try $r = 1$, then successively increase r, while $d \approx N/r$ such that the linear system stays over-determined.

Example

- Given

$$(u_n)_{0\leq n\leq 6} = (1, 4, 36, 400, 4900, 63504, 853776).$$

- We wish to find a linear recurrence of order $r = 1$ and degree $d \leq 2$.
- Look for a non-zero sextuple $(a, b, c, d, e, f) \in \mathbb{Q}^6$ such that

$$(fn^2 + en + d)u_{n+1} + (cn^2 + bn + a)u_n = 0, \quad n = 0, \ldots, 5.$$

- Need to solve:

$$\begin{pmatrix} u_0 & 0 & 0 & u_1 & 0 & 0 \\ u_1 & u_1 & u_1 & u_2 & u_2 & u_2 \\ u_2 & 2u_2 & 4u_2 & u_3 & 2u_3 & 4u_3 \\ u_3 & 3u_3 & 9u_3 & u_4 & 3u_4 & 9u_4 \\ u_4 & 4u_4 & 16u_4 & u_5 & 4u_5 & 16u_5 \\ u_5 & 5u_5 & 25u_5 & u_6 & 5u_6 & 25u_6 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \\ d \\ e \\ f \end{pmatrix} = 0.$$

- Solution space: span$(-4, -16, -16, 1, 2, 1)^t$. Guessed recurrence:

$$(n^2 + 2n + 1)u_{n+1} - (16n^2 + 16n + 4)u_n = 0.$$

How to do guessing efficiently?

"Educated guessing should be learned."

- Instead of solving the system over $\mathbb{Q}$ it is faster to solve over $\mathbb{F}_p$ for several primes p. Then combine the solutions using CRT.

- Exploit the structure of the obtained linear systems (generalized Hermite-Padé approximation) [Beckermann, Labahn, 1997].

Toy examples

"In mathematics often the simplest is the best."

```
> with(gfun):
> l := [1,3,9,27]:
> listtorec(l,u(n));
```

powers of 3

```
{-3*u(n) + u(n + 1), u(0) = 1}
```

```
> listtodiffeq(l,y(x));
```

```
(-3*x + 1)*y(x) - 1
```

Toy examples

"In mathematics often the simplest is the best."

```
> with(gfun):
> l := [0,1,3,6,10,15,21]:
> listtorec(l,u(n));
```

$(n^2 + n)/2$

```
{(-n - 2)*u(n) + n*u(n + 1), u(0) = 0, u(1) = 1}
```

```
> listtodiffeq(l,y(x));
```

```
(2*x + 1)*y(x) + (x^2 - x)*y'(x)
```

```
> dsolve(%);
```

```
x/(1 - x)^3
```

Toy examples

"In mathematics often the simplest is the best."

Catalan numbers

```
> with(gfun):
> l := [1,1,2,5,14,42]:
> rec := listtorec(l,u(n))[1];

        rec := {(-4*n - 2)*u(n) + (n + 2)*u(n + 1), u(0) = 1}

> rsolve(rec,u(n));

                4^n*binomial(n - 1/2, -1/2)/(n + 1)

> listtodiffeq(l,y(x));
                              FAIL

> rectodiffeq(rec,u(n),y(x));

{2*y(x) + (10*x - 2)*y'(x) + (4*x^2 - x)*y''(x), y(0) = 1, D(y)(0) = 1}

> dsolve(%);
                      2/(1 + sqrt(1 - 4*x))
```

Toy examples

"In mathematics often the simplest is the best."

```
> with(gfun):
> l := [1,3,19,147,1251,11253,104959,1004307,9793891,
96918753,970336269,9807518757,99912156111,1024622952993,10567623342519]:
> listtorec(l,u(n));
```

Apéry numbers

```
{-(n+1)^2*u(n) + (-11*n^2 - 33*n - 25)*u(n + 1) + (n+2)^2*u(n + 2)}
```

```
> listtodiffeq(l,y(x))[1];
```

```
{(x + 3)*y(x) + (3*x^2 + 22*x - 1)*y''(x) + (x^3 + 11*x^2 - x)*y'''(x)}
```

```
> dsolve(%,[hypergeometricsols]);
```

```
hypergeom([1/12, 7/12], [1], q(x))/p(x)^(1/6)
```

$$ {}_2F_1\left[\begin{matrix}1/12 & 7/12\\ & 1\end{matrix}; q(x)\right] \cdot \frac{1}{p(x)^{1/6}} $$

Application 1: The Yang-Zagier numbers

- In Arithmetic and Topology of Differential Equations, 2018 by Don Zagier:

$$c_{n-3} + 20\left(4500n^2 - 18900n + 19739\right) c_{n-2} + 80352000n(5n-1)(5n-2)(5n-4)c_n + 25\left(2592000n^4 - 16588800n^3 + 39118320n^2 - 39189168n + 14092603\right) c_{n-1} = 0,$$

with initial terms $c_0 = 1, c_1 = -161/(2^{10} \cdot 3^5)$ and $c_2 = 26605753/(2^{23} \cdot 3^{12} \cdot 5^2)$.

- Recursion comes from physics: integral over a moduli space (“topological ODE”) [Bertola, et al, 2015].
- [Yang and Zagier]: $a_n = c_n \cdot (3/5)_n \cdot (4/5)_n \cdot (2^{10} \cdot 3^5 \cdot 5^4)^n \in \mathbb{Z}$.

$$(\alpha)_n := \alpha \cdot (\alpha + 1) \cdots (\alpha + n - 1).$$

- “this is a very mysterious example [...] of numbers defined by recursions with polynomial coefficients.” – [Zagier, 2018]

Guessing for the Yang-Zagier numbers

$$a_n = c_n \cdot (3/5)_n \cdot (4/5)_n \cdot (2^{10} \cdot 3^5 \cdot 5^4)^n.$$

Three different ways to find a linear recurrence relation:

1. Use effective closure properties of P-recursive sequences.
2. First guess and then prove the recursion.
3. Guess and prove an ODE for $\sum_n a_n x^n$. Then convert it into a recurrence.

	Order of recurrence	Order of ODE
1 Closure properties	3	4
2 Guessing the recurrence	2	3
3 Guessing the ODE	3	2

The generating function of the Yang-Zagier numbers

- $f(x) = \sum_n a_n x^n$ solves

$$q_2(x)y''(x) + q_1(x)y'(x) + q_0(x)y(x) = 0, \quad \text{where} \tag{1}$$
$$q_2(x) = 5x(302400x - 31)(373248000x^2 + 216000x + 1),$$
$$q_1(x) = 1354442342400000x^3 + 64571904000x^2 - 61473600x - 31,$$
$$q_0(x) = 300(902961561600x^2 - 240974784x - 4991).$$

- Any solution of (1) is a linear combination of

$$A_1(x) := u_1(x) \cdot {}_2F_1\left[\begin{matrix}-1/60 & 11/60 \\ & 2/3\end{matrix}; q(x)\right] \text{ and } A_2(x) := u_2(x) \cdot {}_2F_1\left[\begin{matrix}19/60 & 31/60 \\ & 4/3\end{matrix}; q(x)\right].$$

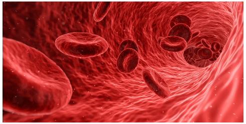
- "Guess and prove": $A_1(x)$ and $A_2(x)$ are algebraic functions.

Theorem (Bostan, Weil, Y., 2021)

The generating function of the Yang-Zagier numbers is algebraic.

Application 2: Monotonicity of Iso

- *Canham model* predicts shape of biomembranes like blood cells [Canham, 1970].
- Model asks to minimize the *Willmore energy*



$$W(S) = \int_S H^2 \mathrm{d}A,$$

over orientable closed surfaces S with prescribed genus, area and volume.

- [Yu, Chen, 2021]: The solution to the model is unique in the genus-one case, if $\mathrm{Iso}(z)$ is strictly increasing on $z \in [0, \sqrt{2}-1)$, where

$$\mathrm{Iso}(z) := 3 \cdot 2^{3/4}\pi^{3/2} \cdot \frac{\bar{V}(z^2)}{\bar{A}^{3/2}(z^2)} = \frac{2^{3/4}}{\pi}\left(\frac{3}{4} + \frac{9}{8}z^2 - \frac{243}{16}z^4 + \cdots\right),$$

$$\bar{A}(z) = \frac{1}{\sqrt{2}\pi}\int_0^{2\pi}\int_0^{2\pi}\frac{\sqrt{2}+\sin(v)}{Q(u,v,1\,;\sqrt{z})^2}\mathrm{d}u\mathrm{d}v \text{ and } \bar{V}(z) = \frac{1}{\sqrt{2}\pi}\int_0^{1}\int_0^{2\pi}\int_0^{2\pi}\frac{r\sqrt{2}+r^2\sin(v)}{Q(u,v,r\,;\sqrt{z})^3}\mathrm{d}u\mathrm{d}v\mathrm{d}r,$$

$$Q(u,v,r\,;\,z) = 1 + 2(\sqrt{2}+r\sin(v))\cos(u)z + (2 + r^2 + 2\sqrt{2}r\sin(v))z^2.$$

Investigation of $\mathrm{Iso}(z) = 3 \cdot 2^{3/4}\pi^{3/2} \cdot \bar{V}(z^2)/\bar{A}^{3/2}(z^2)$

- Guessing finds second-order differential equations for $\bar{A}(z)$ and $\bar{V}(z)$:

$$z(z-1)(z^2-6z+1)(z+1)^2\bar{A}''(z) + (z+1)(5z^4-8z^3-32z^2+28z-1)\bar{A}'(z) + (4z^4+11z^3-z^2-43z+13)\bar{A}(z) = 0$$

$$z(z-1)(z+1)(z^2-6\,z+1)^2\,\bar{V}''(z) + 3\,(3\,z^5-24\,z^4-2\,z^3+56\,z^2-25\,z+8)\,\bar{V}(z) + (z^2-6\,z+1)(7\,z^4-22\,z^3-18\,z^2+26\,z-1)\,\bar{V}'(z) = 0.$$

- Using Creative Telescoping we can prove the guesses.
- Maple solves the ODEs: $p(z) = 1 - 6z + z^2$

$$\bar{A}(z) = \frac{4(z+1)}{p(z)^{3/2}} \cdot {}_2F_1\left[\begin{matrix}-\frac{1}{2} & \frac{3}{2} \\ & 1\end{matrix}; \frac{-4z}{p(z)}\right] \text{ and } \bar{V}(z) = \frac{2}{p(z)^{3/2}} \cdot {}_2F_1\left[\begin{matrix}-\frac{3}{2} & -\frac{5}{2} \\ & 1\end{matrix}; \frac{-4z}{p(z)}\right].$$

Theorem (Melczer, Mezzarobba 2020; and Bostan, Y. 2021)

The function $\mathrm{Iso}(z)$ *is increasing on* $[0, \sqrt{2}-1)$.

Main takeaways

"We must idealize."

- P-recursive sequences are ubiquitous.
- Automated guessing allows finding structure in sequences.
- Modern computer algebra (e.g. `gfun`) makes efficient guessing easy.

Bonus: Guessing differntial equations

"Finished mathematics consists of proofs, but mathematics in the making consists of guesses."

- Given: $u_0, \ldots, u_N$ terms of a sequence $(u_n)_{n\geq 0}$.
- Want: a *guess* for a linear differential equation for $\sum_{n\geq 0} u_n x^n$.
- Idea: Look for $r \in \mathbb{N}$ and $c_0(x), \ldots, c_r(x) \in \mathbb{Q}[x]$ of some degree $d \in \mathbb{N}$ such that

$$c_r(x) \cdot \partial^r \sum_{n=0}^{N} u_n x^n + \cdots + c_0(x) \cdot \sum_{n=0}^{N} u_n x^n = 0$$

- Need to solve a *system of linear equations*, where the unknowns are the $(r+1)(d+1)$ coefficients of the polynomials $c_i(x)$.
- If $(r+1)(d+1) > N$, a non-zero solution trivially exists.
  If $(r+1)(d+1) \ll N$, no reason for a solution, except if $(u_n)_n$ is P-recursive.
- If no bounds on r and d, first try $r = 1$, then successively increase r, while $d \approx N/r$ such that the linear system stays over-determined.

Bonus: Example

- Given

$$(u_n)_{0\leq n\leq 4} = (1, 2, 6, 20, 70).$$

- We wish to find a linear differential equation of order $r = 1$ and degree $d \leq 1$.
- Look for a non-zero quadruple $(a, b, c, d) \in \mathbb{Q}^4$ such that

$$(c + dx)(1 + 2x + 6x^2 + 20x^3 + 70x^4)' + (a + bx)(1 + 2x + 6x^2 + 20x^3 + 70x^4) = 0.$$

- Need to solve:

$$\begin{pmatrix} u_0 & 0 & u_1 & 0 \\ u_1 & u_0 & 2u_2 & u_1 \\ u_2 & u_1 & 3u_3 & 2u_2 \\ u_3 & u_2 & 4u_4 & 3u_3 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} = 0.$$

- Solution space: $\text{span}(2, 0, -1, 4)^t$. Guessed differential equation:

$$(4x - 1)f'(x) + 2f(x) = 0.$$