Why You Should Remove Zeros From Data Before Guessing

Manuel Kauers, Thibaut Verron

Institute for Algebra, Johannes Kepler University, Linz, Austria

JOHANNES KEPLER UNIVERSITÄT LINZ

Problem: Determine a recurrence relation for the sequence $(a_n)_{n\in\mathbb{N}} \in \mathbb{F}_{97}^{\mathbb{N}}$ starting with

$3, 0, 0, 8, 0, 0, 18, 0, 0, 30, 0, 0, 42, 0, 0, 54, 0, 0, 68, 0, 0, 80, 0, 0, 1, 0, \dots$

0

- Generate a new sequence $\overline{a}_n = a_{3n}$ without the zeroes:

$3, 8, 18, 30, 42, 54, 68, 80, 1, \dots$

1

- Make an ansatz equation (original sequence):

$$\sum_{j=0}^{3}\sum_{i=0}^{4} c_{i,j} n^i a_{n+j} = 0$$

- Make an ansatz equation (sequence without zeros):

$$\sum_{j=0}^{1}\sum_{i=0}^{4} \overline{c}_{i,j} n^i \overline{a}_{n+j} = 0$$

Correspondence (→):

$$\sum c_{i,j/3}\left(\frac{n}{3}\right)^i a_{n+3j} = \sum c_{i,j/3}\left(\frac{n}{3}\right)^i \overline{a}_{n/3+j} = 0$$

Correspondence (←):

$$\sum \overline{c}_{i,3j}(3n)^i \overline{a}_{n+j} = \sum \overline{c}_{i,3j}(3n)^i a_{3n+3j} = 0$$

2

- Plug in the values of a_n and obtain a linear system:

19+4 linear equations, 20 unknowns:

$$\begin{pmatrix}
3 & 3 & 3 & 3 & 3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 8 & 8 & 8 & 8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8 & 16 & 32 & 64 & 31 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 8 & 24 & 72 & 22 & 66 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
8 & 32 & 31 & 27 & 11 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 18 & 72 & 94 & 85 & 49 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 18 & 90 & 62 & 19 & 95 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 18 & 11 & 66 & 8 & 48 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
18 & 29 & 9 & 63 & 53 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 30 & 16 & 15 & 8 & 56 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 30 & 46 & 77 & 34 & 78 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 30 & 76 & 5 & 45 & 17 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
30 & 9 & 90 & 27 & 76 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 42 & 32 & 29 & 96 & 87 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 42 & 74 & 38 & 30 & 39 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 42 & 19 & 34 & 20 & 46 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
42 & 61 & 17 & 27 & 60 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 54 & 23 & 8 & 7 & 91 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 54 & 77 & 11 & 57 & 22 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 54 & 34 & 25 & 84 & 96 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
54 & 88 & 50 & 24 & 93 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 68 & 21 & 45 & 41 & 74 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 68 & 89 & 58 & 16 & 78 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 68 & 60 & 13 & 40 & 41 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
68 & 31 & 7 & 36 & 5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 80 & 65 & 71 & 88 & 23 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 80 & 48 & 87 & 91 & 74 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 80 & 31 & 69 & 91 & 68 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
80 & 14 & 17 & 83 & 80 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 22 & 96 & 75 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 23 & 44 & 42 & 93 & 0 & 0 & 0 & 0 & 0
\end{pmatrix} \cdot \begin{pmatrix} c_{0,0} \\ c_{1,0} \\ c_{2,0} \\ c_{3,0} \\ c_{4,0} \\ c_{0,1} \\ c_{1,1} \\ c_{2,1} \\ c_{3,1} \\ c_{4,1} \\ c_{0,2} \\ c_{1,2} \\ c_{2,2} \\ c_{3,2} \\ c_{4,2} \\ c_{0,3} \\ c_{1,3} \\ c_{2,3} \\ c_{3,3} \\ c_{4,3} \end{pmatrix} = 0$$

Overdetermined system

- Plug in the values of $\overline{a}_n$ and obtain a linear system:

8 linear equations, 8+2 unknowns:

$$\begin{pmatrix}
3 & 3 & 3 & 3 & 3 & 8 & 8 & 8 & 8 & 8 \\
8 & 16 & 32 & 64 & 31 & 18 & 36 & 72 & 47 & 94 \\
18 & 54 & 65 & 1 & 3 & 30 & 90 & 76 & 34 & 5 \\
30 & 23 & 92 & 77 & 17 & 42 & 71 & 90 & 69 & 82 \\
42 & 16 & 80 & 12 & 60 & 54 & 76 & 89 & 57 & 91 \\
54 & 33 & 4 & 24 & 47 & 68 & 20 & 23 & 41 & 52 \\
68 & 88 & 34 & 44 & 17 & 80 & 75 & 40 & 86 & 20 \\
80 & 58 & 76 & 26 & 14 & 1 & 8 & 64 & 27 & 22
\end{pmatrix} \cdot \begin{pmatrix} \overline{c}_{0,0} \\ \overline{c}_{1,0} \\ \overline{c}_{2,0} \\ \overline{c}_{3,0} \\ \overline{c}_{4,0} \\ \overline{c}_{0,1} \\ \overline{c}_{1,1} \\ \overline{c}_{2,1} \\ \overline{c}_{3,1} \\ \overline{c}_{4,1} \end{pmatrix} = 0$$

Underdetermined system

3

- If it was generic, the system would not have any solution.
- If it was generic, the system would have a dimension 2 solution space.

4

- But it has a dimension 2 kernel:

$$(1, 0, 38, 36, 92, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 38, 77, 75, 92)^T$$
$$(0, 1, 35, 38, 16, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 49, 58, 73, 82, 68)^T$$

- It does, the kernel is generated by

$$(1, 0, 93, 80, 19, 86, 67, 54, 62, 83)^T$$
$$(0, 1, 83, 9, 41, 33, 41, 18, 4, 72)^T$$

5

- So we can deduce some relations:

$$(92n^4 + 75n^3 + 77n^2 + 38n + 7)a_{n+3} + (92n^4 + 36n^3 + 38n^2 + 1)a_n = 0$$
$$(68n^4 + 82n^3 + 73n^2 + 58n + 49)a_{n+3} + (16n^4 + 38n^3 + 35n^2 + n)a_n = 0$$

- The corresponding relations would be:

$$(83n^4 + 62n^3 + 54n^2 + 67n + 86)\overline{a}_{n+1} + (19n^4 + 80n^3 + 93n^2 + 1)\overline{a}_n = 0$$
$$(72n^4 + 4n^3 + 18n^2 + 41n + 33)\overline{a}_{n+1} + (41n^4 + 9n^3 + 83n^2 + n)\overline{a}_n = 0$$

6

- But those relations are not valid!
- We do not have enough data to obtain a recurrence!

- The same happens with other guessing techniques, such as Hermite-Padé approximation.
- The correspondence between the recurrences for (a_n) and $(\overline{a}_n)$ is always true. It ensures that removing the zeros does not lose information.
- In general, the number of values required to guess with any confidence a recurrence relation of order r and degree d is

$$(r+1)(d+2) - 1.$$

- In the example, on the left-hand side, with $r = 3$ and $d = 4$, we find that we need at least 23 points, so our 26 points should have been enough if the sequence had been generic.
- On the right-hand side, with $r = 1$ and $d = 4$, we find that we need 11 points, so our 9 points are indeed not enough.
- The actual minimal number of values, for a sequence such that $a_n = 0$ whenever $n \bmod m \neq 0$, is

$$m\left[\left(\left\lfloor\frac{r}{m}\right\rfloor + 1\right)\left(d+2\right) - 1\right]$$

- Again in the example, for the original sequence, we need 33 points, which confirms that our 26 points are not enough.

- One can also guess polynomial relations.
- There, one considers the formal power series $a(t) = \sum_{n=0}^{\infty} a_n t^n$ and looks for a relation of the form

$$P(t, a(t)) = \sum_{j=0}^{r}\sum_{i=0}^{d} c_{i,j} t^i a(t)^j = 0$$

- If $\overline{a}(t) = \sum_{n=0}^{\infty} a_n t^n$, then $a(t) = \overline{a}(t^m)$ and

$$\sum_{j=0}^{r}\sum_{i=0}^{d/m} c_{i/m,j} t^i \overline{a}(t)^j = 0.$$

- In general, the number of values required to guess with any confidence a polynomial relation of degree r in a and d in t is

$$(r+1)(d+1).$$

- For a sequence such that $a_n = 0$ whenever $n \bmod m \neq 0$, the actual value is

$$m\left(r+1\right)\left(\left\lfloor\frac{d}{m}\right\rfloor + 1\right).$$

- One can also guess differential equations of the form

$$\sum_{j=0}^{r}\sum_{i=0}^{d} c_{i,j} t^i a(t)^{(j)} = 0$$

- But $a'(t) = mt^{m-1}\overline{a}'(t^m)$ does not lead to a convenient conversion between equations for $a(t)$ and $\overline{a}(t)$
- Example: $\cos(t) = 1 + 0t - \frac{t^2}{2} + 0t^3 + \frac{t^4}{24} + \dots$ satisfies the differential equation

$$\cos''(t) + \cos(t) = 0 \qquad \text{(order 2, degree 0)}$$

- If we remove the zeros, we obtain $a(t) = \cos(\sqrt{t})$, for which the smallest differential equation is

$$4ta''(t) + 2a'(t) + a(t) = 0 \qquad \text{(order 2, degree 1)}$$

- If we add zeros, we can consider $b(t) = \cos(t^2)$, for which the smallest differential equation are

$$tb''(t) - b'(t) + 4t^3 b(t) = 0 \qquad \text{(order 2, degree 3)}$$
$$b^{(3)}(t) + 4t^2 b'(t) + 12tb(t) = 0 \qquad \text{(order 3, degree 2)}$$

References

- B. Beckermann and G. Labahn (1994). "A uniform approach for the fast computation of Matrix-type Padé approximants". In: *SIAM Journal on Matrix Analysis and Applications* 15.3, pp. 804–823
- W. Hebisch and M. Rubey (2011). "Extended Rate, more GFUN". In: *Journal of Symbolic Computation* 46.8, pp. 889–903
- M. Kauers (2009). *Guessing Handbook*. Tech. rep. 09-07. RISC-Linz
- M. Kauers (2013). "The Holonomic Toolkit". In: *Computer Algebra in Quantum Field Theory: Integration, Summation and Special Functions*. Springer, pp. 119–144
- M. Kauers, M. Jaroschek, and F. Johansson (2014). "Ore Polynomials in Sage". In: *Computer Algebra and Polynomials*. LNCS 8942. Springer, pp. 105–125
- B. Salvy and P. Zimmermann (1994). "Gfun: a Maple package for the manipulation of generating and holonomic functions in one variable". In: *ACM Transactions on Mathematical Software* 20.2, pp. 163–177