Unsupervised Discovery of Formulas for Mathematical Constants

Anonymous Author(s)
Affiliation
Address
`email`

1 Abstract

Ongoing efforts that span over decades show a rise of AI methods for scientific discovery and hypothesis creation [Fajtlowicz, 1988, Petkovsek et al., 1996, Wolfram et al., 2002, Buchberger et al., 2006, Bailey et al., 2007, Raayoni et al., 2021, Davies et al., 2021, Fawzi et al., 2022]. Despite the significant advances in the impact of AI for science, number theory in mathematics remains a persistent challenge for AI. Specifically, AI methods were not effective in creation of formulas for mathematical constants because such formulas are either true or false, with no continuous adjustments that can enhance their correctness. This entire field lacked a "distance metric" between two formulas that can guide progress. The absence of a systematic method left the realm of formula discovery elusive for automated methods. In this work, we propose a systematic methodology for categorization, characterization, and pattern identification of such formulas. The key to our methodology is introducing metrics based on the convergence dynamics of the formulas, which we utilize for the first automated clustering of mathematical formulas. We demonstrate this methodology on Polynomial Continued Fraction formulas, which are ubiquitous in their intrinsic connections to mathematical constants [Lagarias, 2013, Bowman and McLaughlin, 2002, Laughlin and Wyshinski, 2004], and generalize many mathematical functions and structures. We test our methodology on a set of 1,768,900 such formulas, identifying many known formulas for mathematical constants, and discover previously unknown formulas for π, $\ln(2)$, Gauss, and Lemniscate constants. The uncovered patterns enable a direct generalization of individual formulas to infinite families, unveiling rich mathematical structures. This success paves the way towards a generative model that creates continued fractions fulfilling specified mathematical properties, potentially accelerating by orders of magnitude the rate of discovery of useful formulas.

2 Introduction

Historically, formulas of mathematical constants were a symbol of aesthetics and beauty. Continued fraction formulas such as those for the Golden Ratio ϕ and $\tan(x)$

$$1+\frac{1}{1+\frac{1}{1+\frac{1}{1+\cdots}}}=\phi \quad \frac{x}{1-\frac{x^2}{3-\frac{x^2}{5-\cdots}}}=\tan(x) \tag{1}$$

enable calculating infinitely many digits for these constants. Discovering such formulas often leads to profound revelations regarding the properties and underlying structure of fundamental constants. For example, the continued fraction formula for $\tan(x)$, shown in Eq. 1, was used by Johann Heinrich Lambert in the first proof of the irrationality of Pi [Berggren et al., 2004]. Unfortunately, such formulas are notoriously hard to find on-demand, often relying on a mathematician's profound intuition. Part of the challenge is the lack of a well-defined 'distance' between a formula and a given constant. i.e., there is no known way to tell whether a formula is nearly accurate. The formula either works, or it does not. In other fields of science, a prediction accurate to 1000 digits is accurate enough for any practical need. However, in mathematics, if the 1001$^{\text{st}}$ digit is wrong, the formula is incorrect

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

and gives no insight regarding a correct formula. This is a substantial hurdle both for human efforts and for automated analysis, as gradient descent is generally unsuitable for binary metrics.

Recent efforts used large scale distributed computation to discover a multitude of formula hypotheses for mathematical constants [Raayoni et al., 2021, Elimelech et al., 2023]. These efforts relied mostly on exhaustive search methods. Other older applications of AI to mathematical discovery in other fields include Automated Theorem Proving [Petkovsek et al., 1996] (such as Malarea [Urban, 2007] and Flyspeck [Kaliszyk and Urban, 2012]), and Automated Conjecture Generation [Wang, 1960] (such as the Automated Mathematician [Lenat, 1982], EURISKO [Lenat and Brown, 1983, Davis and Lenat, 1982], and Graffiti [Fajtlowicz, 1988]).

This work proposes a fundamentally new methodology for automated investigation of formulas for mathematical constants. We constructed a large dataset of continued fractions, and enriched it with metrics based on their convergence dynamics, which are found to embody fundamental information about each continued fraction. This dataset enables the identification and generalization of patterns within the data. Through a process of categorization and clustering (Fig. 1), we identified subsets of continued fractions that relate to important mathematical constants. This novel method of formula discovery allowed us to identify both previously known and completely new formulas for constants such as π, $\ln(2)$, $\cot(1)$, the Golden Ratio, square roots of multiple integers, Gauss and Lemniscate constants. Often, once such a subset of formulas is identified, all its members relate to the same mathematical constant, thus exposing an internal structure that can be generalized into infinite families of such formulas.

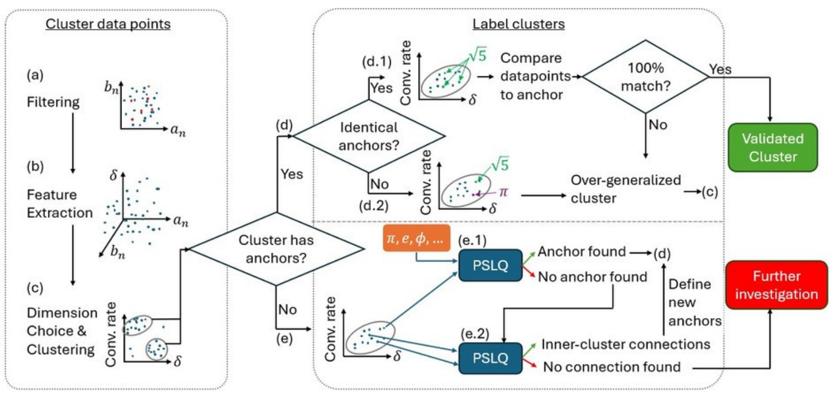

Figure 1: **Systematic clustering and labeling of formulas by dynamical metrics.** Our methodology analyzes Polynomial Continued Fractions (PCFs) in two main stages. **Clustering**: (a) Filter degenerate PCFs. (b) Evaluate PCFs and extract their dynamics-based metrics (section 3). (c) Choose the best few metrics (using the Davies-Bouldin Clustering Index [Davies and Bouldin, 1979] - see table 1) and use them to cluster the data. **Labeling**: In every cluster, look for PCFs known in the literature and use them as anchors. (d) If anchors are found in the cluster, validate that they do not contradict each other, i.e., relate to different constants, which indicates that the cluster should be split. (d.1) If all anchors are in agreement, choose a random subset of other points in the cluster and use PSLQ to validate that they also relate to the same constant. If the validation is successful, the cluster is labeled. If not, the cluster should be split. (d.2) If the anchors relate to different constants, the cluster should be split – return to step c for finer clustering of the data. When focusing on a specific cluster, the best metrics could be different than those for the full dataset. (e) If no anchor is found in a certain cluster, attempt to label by (e.1) choosing a small subset of PCFs in the cluster and running a PSLQ search for each of them against a large set of potential constants. If a connection is found, the cluster now has an anchor – return to step d. (e.2) If an anchor is still not found, attempt to connect a sample of data points within the cluster using PSLQ. If successful, conclude that the cluster is correct, but has no identified constant. Define a new label for that cluster. If PSLQ failed to connect points within the cluster, return to step c for finer clustering. If no further refinement is appropriate, flag the cluster for further analytical investigation.

As a result of our methodology, we present the most complete classification of polynomial continued fractions known to date.

Traditional clustering methods attempt to relate data points by calculating distance metrics based on the parameters of these data points. The most common approaches (like SVM) rely on linear classification, while more advanced methods rely on non-linear kernel transformations - but usually use various functions calculated directly on the data parameters. In our dataset, each point is a continued fraction formula defined by the polynomials used to construct it. We find that it is the *dynamics* of the continued fraction generated by these polynomials, rather than any direct function on their coefficients, which provides the most useful metrics for analysis. In other words, we find that the useful underlying metrics to extract from each data point are embedded within the intricate progression of the sequence created by the formula, rather than the explicit value (limit) of that formula, or the coefficients defining it. Thus, in order to assess the distance between two polynomial continued fractions, and identify relations between such formulas, it is imperative to characterize the nuanced behaviour of their sequences, analyzing trends spanning over numerous terms.

Some of the metrics we extract, such as the irrationality measure, are well-known in the mathematical community, yet were never considered for a large-scale classification effort. We develop a new algorithm - the Blind-δ algorithm - to enable the evaluation of the irrationality measure of formulas on a large scale, previously impossible.

This approach allows us to employ a novel methodology to the formula discovery challenge. We cluster formulas by their 'closeness' to other formulas according to these new metrics, which we use to identify promising formulas regardless of their numerical value (Fig. 1left). Once a candidate formula is found, we numerically validate it by calculating its value to a large precision and then identifying its relation to a mathematical constant using algorithms such as PSLQ [Ferguson and Bailey, 1992] (Fig. 1right). The "generate ⇒ validate" approach is inspired by works in AI-driven code generation [Ridnik et al., 2024] and problem solving in geometry [Trinh et al., 2024].

3 Methodology for Data-Driven Discovery

3.1 Definitions

Polynomial Continued Fractions

In this work we chose to focus on polynomial continued fraction (PCF) formulas as our test case due to the combination of their simplicity and expressive power. PCFs relate to a wide range of mathematical fields, represent a variety of constants, are equivalent to infinite sums [Euler, 1748], and cover mathematical functions such as Bessel functions, trigonometric functions, and generalized hypergeometric functions [Cuyt et al., 2008]. A PCF at depth n is defined as:

$$a_0 + \cfrac{b_1}{a_1 + \cfrac{b_2}{\ddots + \cfrac{b_n}{a_n}}} = \frac{p_n}{q_n}, \tag{2}$$

where $a_n = a(n)$ and $b_n = b(n)$ are evaluations of polynomials with integer coefficients. The PCF value is the limit $L = \lim_{n\to\infty} \frac{p_n}{q_n}$ (when it exists). The converging sequence of rational numbers $\frac{p_n}{q_n}$ provides an approximation of L, which is known as a Diophantine approximation.

The Irrationality Measure of a Number

While irrational numbers cannot be expressed using a simple quotient of integers, they can be approximated by them. Moreover, some approximations are "better" than others, and one way to evaluate their quality is by a quantity called the irrationality measure [Hardy et al., 1979].

For every $L \in \mathbb{R}$, the *irrationality measure of* L is defined as the supremum of all possible δ for which there is a sequence of distinct rational numbers $\frac{p_n}{q_n} \to L$; $\frac{p_n}{q_n} \neq L$ that satisfies

$$\left| L - \frac{p_n}{q_n} \right| < \frac{1}{q_n^{1+\delta}}. \tag{3}$$

It is known that for irrational numbers this measure is ≥ 1 (Dirichlet theorem for Diophantine approximations), and for rationals it is 0.

Given a sequence $\frac{p_n}{q_n}$ and its limit L, we define the *irrationality measure of a sequence* as

$$\delta = \frac{-\log\left|L - \dfrac{p_n}{q_n}\right|}{\log|\tilde{q}_n|} - 1, \tilde{q}_n = \frac{q_n}{\gcd(p_n, q_n)} \tag{4}$$

For a sufficiently large n. Note that the irrationality measure of L is greater or equval to the irrationality measure of any specific sequence converging to the same L.

3.2 δ-Predictor Formula

The classification of a large number of continued fraction formulas requires an efficient and accurate calculation of the irrationality measure δ for each formula. This calculation is challenging because it depends on the asymptotic behavior of the converging sequence, and because δ appears as an exponent of a large number. The δ-Predictor formula that we present here provides a way around this challenge - requiring no specific knowledge about the convergence rate and trajectory, or even about the sequence limit itself:

$$\delta_{\text{predicted}} = \lim_{n\to\infty} \frac{n \cdot \log\left|\frac{\lambda_1(n)}{\lambda_2(n)}\right|}{\log|\tilde{q}_n|} - 1 \tag{5}$$

where $\lambda_1(n)$ and $\lambda_2(n)$ are the eigenvalues of the matrix $\begin{pmatrix} 0 & b_n \\ 1 & a_n \end{pmatrix}$, $|\lambda_1(n)| > |\lambda_2(n)|$.

This formula extends a hypothesis made in a previous work [David et al., 2021], which was limited to PCFs with balanced polynomial degrees and with a $\tilde{q}_n$ that grows exponentially. As we found in this work, Eq.5 works for any converging PCF. It was validated numerically and proven for the balanced-degree case in Appendix D. This formula provides a substantial advantage in the estimate of the irrationality measure, a critical dynamical metric for our work. Specifically, the asymptotic behavior of $\tilde{q}_n$ and λ_1/λ_2 are still required for finding $\delta_{\text{predicted}}$, but they are usually easier to derive.

3.3 Discovery of Formulas by Unsupervised Learning

Each PCF formula is defined by the polynomials that generate it. This work focuses on polynomials up to 2nd degree: $a = A_2n^2 + A_1n + A_0$, $b = B_2n^2 + B_1n + B_0$, with integer coefficients in the domain $-5 \leq A_i \leq 5$, $-5 \leq B_i \leq 5$. We removed the $a = 0$ and $b = 0$ cases, as they break the PCF structure, leaving us with 1,768,900 formulas. Some of these PCFs do not converge to a single limit, rendering their measured metrics meaningless (see Appendix B for the classification method we developed to predict PCF convergence). We filtered out all formulas that do not converge, providing the final filtered dataset of 1,543,926 formulas.

Our methodology relies on dynamics-based metrics. The following metrics are calculated for each formula:

- The coefficients of the polynomials a and b, $(A_2, A_1, A_0, B_2, B_1, B_0)$. We also define the useful characteristic of which polynomial dominates the dynamics: when $2\deg(a) > \deg(b)$ the PCF is A-dominated, when $2\deg(a) < \deg(b)$ the PCF is B-dominated, and when $2\deg(a) = \deg(b)$ the PCF is balanced.
- The numerical limit of the PCF, evaluated at depth $n = 2000$.
- The irrationality measure: for each PCF, we calculate $\delta_{\text{predicted}}$ by substituting $n = 10^9$ into Eq.5, and measure δ directly using the Blind-δ algorithm (presented in section 3.4) at depth $n = 1000$ (see Fig.2a for example δ evaluations).
- The convergence rate dynamics, comprised of three parameters: we estimate the approximation error, which scales as $\epsilon(n) \sim n!^{\eta} \cdot e^{\gamma n} \cdot n^{\beta}$ for large n. We fit a curve of this form numerically (see Appendix A for more details) and store the estimate of the η, γ, β parameters (η - factorial coefficient, γ - exponential coefficient, β - polynomial coefficient).
- The growth rate of $\tilde{q}_n$. As n grows, $\tilde{q}_n \sim n!^{\eta'} \cdot e^{\gamma' n} \cdot n^{\beta'}$. We fit a curve of this form numerically and store the estimate of the η', γ', β' parameters.

Based on this set of metrics, we applied unsupervised clustering for unlabeled data (the density-based OPTICS algorithm [Ankerst et al., 1999]) and created a complete algorithm for mathematical formula discovery (Fig.1). A variety of useful formulas, formula families and data patterns were identified (see sections 4.1, 4.2 and 4.3 for selected results).

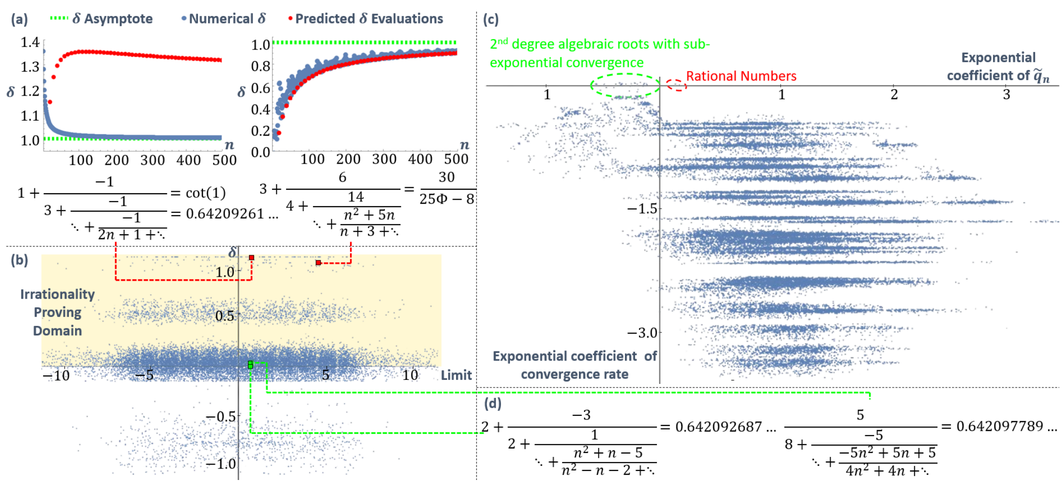

Figure 2: **Dynamics-based metrics for formulas of mathematical constants.** Analysing the convergence of polynomial continued fraction (PCF) formulas provide *dynamical metrics* that prove useful for their automated clustering and identification. (a) Irrationality measure vs. PCF depth. The simplest formula candidate identification method we used is filtering by high numeric δ. These are 2 examples of formulas for mathematical constants ($\cot(1)$ and the Silver Ratio) found this way. The irrationality measure of these constants is known to be 1 (green dashed line). The blue dots show how the numerical evaluations of δ (Eq.4) converge to the correct irrationality measure. Red dots are evaluations of the δ-Predictor (Eq.5) at finite n values. The prediction follows the numerical delta very closely in the Silver Ratio formula, while taking a completely different (and much slower) trajectory in the $\cot(1)$ formula - yet both converge to the correct value $\delta = 1$. (b) δ (at depth $n = 1000$) vs. limit value for PCFs in our set. While δ values seem to follow a pattern, the limit value distribution doesn't contain relevant information - the higher density of PCFs near the Y axis caused by our choice of a dataset with small coefficient polynomials. Note the multitude of irrationality-proving formulas, most of which are still not linked to any known constant. (c) Exponential growth coefficients of $\tilde{q}_n$ and $\epsilon(n)$ for balanced PCFs. Note the surprising "band" structure that this view reveals. A few of the clusters have been characterized, but the reason for the appearance of these "bands", as well as the properties of most clusters remain open questions for future research. (d) Examples of PCFs in the dataset that converge to a value close to the constant $\cot(1)$ ($\pm 10^{-5}$ or closer), and yet are not related to $\cot(1)$ or its formula shown in (a) - showcasing the challenge of mathematical formula discovery. Error bars not shown for visual clarity, see Appendix A for a discussion regarding measurement errors.

3.4 The Blind-δ Algorithm

The irrationality measure of a PCF is of mathematical interest, and (as we will see in section 4) is a powerful dynamical metric of a formula. Unfortunately, even if we limit ourselves to the numerically estimated δ (given a specific series and a specific depth), Eq.4 requires knowing the series limit L, making its calculation for a large set of unlabeled PCFs impossible.

The Blind-δ algorithm was created in order to circumvent this limitation. Instead of inspecting the convergence behavior of $\frac{p_n}{q_n} \to L$, we inspect the convergence behavior of $\frac{p_n}{q_n} \to \frac{p_m}{q_m}$ for some $m > n$.

This solves the prior knowledge issue, but how is it related to the actual series delta?

Given a rational approximation $\frac{p_n}{q_n} \to L$, we approximate the error rate $|\epsilon(n)|$ where $\epsilon(n) := \frac{p_n}{q_n} - L$ with

$$\frac{p_n}{q_n} - \frac{p_m}{q_m} = \epsilon(n) \cdot \left(1 - \frac{\epsilon(m)}{\epsilon(n)}\right).$$

So if $0 < s < \left|1 - \frac{\epsilon(m)}{\epsilon(n)}\right| < S$ is bounded away from zero and infinity for all n large enough, then this approximation has the same order of magnitude. This means that the error and the convergence behave similarly enough whether we use the true limit L or its approximation $\frac{p_m}{q_m}$. This condition holds whenever $|\epsilon(n)| \to 0$ fast enough, which is true for the vast majority of PCFs (see Appendix D for details).

Note that m has to grow with n. In practice the algorithm uses $m = 2n$, so in order to study δ up to $n = 1000$, we use $m = 2000$.

3.5 Choice of Metrics for Clustering

As part of the automated formula discovery flow we choose the best metrics (for each step), in terms of representation power, which is measured by applying the Davies-Bouldin Index [Davies and Bouldin, 1979] on clustering using a single metric (table 1 shows results for a randomly chosen sample of 25K converging PCFs). Note the extremely poor performance of the PCF limit L, in agreement with Fig.2b,d. This dimensionality reduction is important both for efficiency during the clustering step (especially since the size of the data set grows exponentially with PCF degree and polynomial coefficient magnitude), and for better explainability.

Table 1: Comparison of the representation power of the main dynamic metrics (lower is better).

Metric		Davies-Bouldin Index
Limit L		67.23
Irrationality measure δ		1.11
Reduced denominator $\tilde{q}_n$ growth factors $\tilde{q}_n \sim n!^{\eta'} \cdot e^{\gamma' n} \cdot P(n)$	Exponential coefficient γ'	0.51
	Factorial coefficient η'	0.13
Error rate $\lvert e(n)\rvert$ growth factors $\lvert e(n)\rvert \sim n!^{\eta} \cdot e^{\gamma n} \cdot P(n)$	Exponential coefficient γ	14.83
	Factorial coefficient η	0.77

4 Results

4.1 Discovered Formulas for Mathematical Constants

The first step in validating the dynamical metrics approach is using basic heuristics on the metric space to find PCFs related to mathematical constants. There are some PCFs in the dataset that have a known irrational limit (like the examples in Eq.1 and the PCF family

$$\cfrac{B}{A + \cfrac{B}{A + \ddots}} = \frac{2B}{A + \sqrt{A^2 + 4B}}$$

for constant A and B), so we expected to find some of them. Through this test, we also found *previously unknown* PCF formulas related to mathematical constants.

Note that known mathematical formulas are both the anchors for labeling and a test set in our method. Formulas related to the same constant or having other common properties are expected to be clustered together.

Since we are looking for irrationals, a series that converges to one of them could have an irrationality measure of 1 (or above). A natural heuristic is inspecting PCFs with $\delta \approx 1$. Another heuristic we used is focusing on PCFs with $\eta' \approx 0$, as it was a very strong indicator for mathematical constant formulas in a previous work [Elimelech et al., 2023]. Combining the two gives a subset (see Fig.3a top left) that contains PCFs such as:

$$5 + \cfrac{-10}{\ddots + \cfrac{-5n^2 - 5n}{5n + 5 + \ddots}} = 2 + \phi \qquad -3 + \cfrac{1}{\ddots + \cfrac{1}{-3 + \ddots}} = \frac{-2}{\sqrt{13} - 3} \tag{6}$$

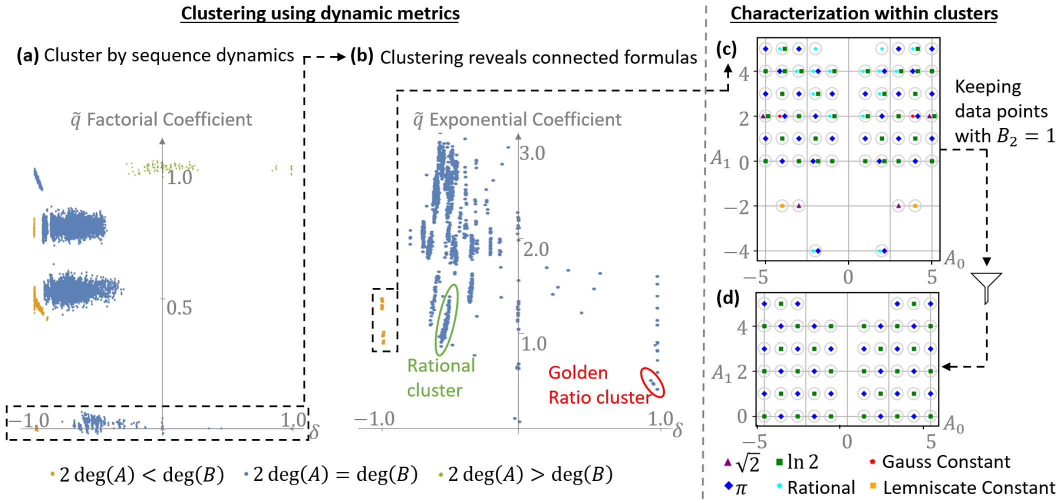

Figure 3: **Discovery of mathematical structures via analysis of dynamic metrics of formulas.** (a) Projecting the data on the δ vs. η' ($\tilde{q}_n$ factorial coefficient) plane, it's easy to see the emerging subsets. We focus on PCFs with $\eta' \approx 0$, as a previous work [Elimelech et al., 2023] indicated this is an important property. (b) Clustering in the δ vs. γ' ($\tilde{q}_n$ exponential coefficient) plane shows examples of common properties within a cluster, like rationality or convergence to a specific constant (up to a linear fractional transformation). Focusing deeper on the B-dominant cluster (as it is a clear anomaly in the $\eta' \approx 0$ subset), we used a PSLQ algorithm to identify links between these formulas and mathematical constants (which was feasible since we identified a promising subset $\sim 5,000$ times smaller than the initial dataset) and got (c) - a surprising number of novel formulas related to mathematical constants (π, $\ln(2)$, $\sqrt{2}$, Gauss and Lemniscate constants). (d) Keeping only PCFs with $B_2 = 1$ we are left with a highly symmetrical "checkerboard pattern" of formulas for π and $\ln(2)$, which was generalized into infinite formula families hypotheses (see section 4.3). Error bars not shown for visual clarity, see Appendix A for a discussion regarding measurement errors.

Removing the requirement of sub-factorial $\tilde{q}_n$ growth rate, one can find the $\cot(1)$ formula shown in Fig.2a:

$$1 + \cfrac{-1}{\ddots + \cfrac{-1}{2n+1+\ddots}} = \cot(1) \tag{7}$$

On the other hand, relaxing the limitation on δ, focusing only on $\eta' \approx 0$, a rich structure emerges (Fig.3b). Diving deeper into the B-dominated subset, we find formulas (Fig.3c) for the Gauss constant G_{GA} [Finch, 2003]:

$$4 + \cfrac{6}{\ddots + \cfrac{4n^2+2n}{4+\ddots}} = \frac{2G_{GA}}{4G_{GA}-3} \qquad 4 + \cfrac{4}{\ddots + \cfrac{4n^2+2n-2}{4+\ddots}} = \frac{4G_{GA}-1}{3G_{GA}-2} \tag{8}$$

Lemniscate constant $L_{Lemniscate}$ [Finch, 2003]:

$$4 + \cfrac{2}{\ddots + \cfrac{4n^2-2n}{4+\ddots}} = \frac{-6}{L_{Lemniscate}-4} \tag{9}$$

As well as for second order roots, π and $\ln(2)$ (see section 4.3). Note that unlike the formulas in Eq.6 and Eq.7, which are analytically proven, the formulas in Eq.8 and Eq.9 are (to the best of the authors' knowledge) *novel*. Their limits were numerically validated to a large precision, yet formal proofs for these formula hypotheses remain an open challenge.

It should be noted that usually in number theory research a bigger δ is considered "good" while a smaller (often negative) δ is considered "bad". We use δ as a metric, without "judgment". These novel formulas (Eq.8, Eq.9 and the infinite family of formulas shown in section 4.3), which have "bad" $\delta \approx -1$, are a demonstration of the advantage of our "non-judgmental" approach.

4.2 Clustering in Dynamics-Based Metric Latent Space

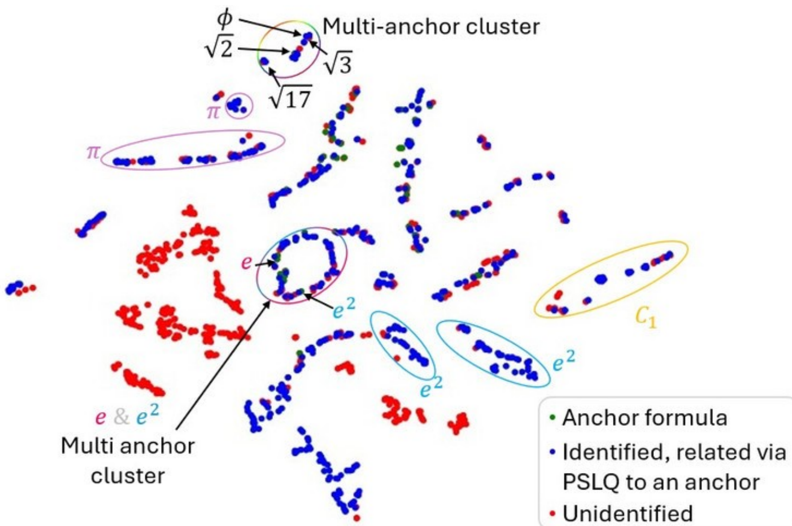

Figure 4: **Automated Formula Discovery Results:** Showcasing the automated clustering and labeling of PCFs using a set of 306 anchor formulas, connected to constants such as π, e, e^2, the continued fraction constant, the golden ratio ϕ, $\sqrt{2}$, $\sqrt{3}$, and $\sqrt{17}$. 454 PCFs were labeled. 332 are equivalent to an anchor, while *122 are novel automatically discovered mathematical formula hypotheses*. For the presentation here, the PCFs were projected to a 2D grid using tSNE (perplexity $= 10$), revealing clusters that coalesce in the full metric space. For visual clarity not all points are shown and error bars aren't shown, see Appendix A for a discussion regarding measurement errors.

This section shows that clusters in the latent space of dynamics-based metrics group together different formulas that share multiple properties including the mathematical constant to which they relate.

We'll start with 2 concrete examples for this effect. Looking at the top left cluster in Fig.3b (defined by $\tilde{q}_n$ exponential coefficient < 0.6 and $\delta > 0.9$), we recognize the canonical form of the Golden Ratio PCF (shown in Eq.1) - but also 21 additional PCFs, with different generating polynomials, some of higher degree. As it turns out, all of them are linear fractional transformations of $\sqrt{5}$ (see Appendix C), which were labeled by the formula discovery algorithm (Fig.1). Another example of property preservation within a cluster is the rational cluster marked in green on Fig.3b. The limits of the PCFs in this subset are varied, and its spread is real (not only due to numerical imperfections), and yet all its members share the rationality property - which isn't directly measured by any of the latent space dimensions.

Fig.4 showcases a collection of clusters with shared properties, visualized via tSNE. Using a set of 306 (mathematically unique) known anchor formulas, 454 PCFs were labeled. 332 are equivalent to an anchor, while *122 are novel automatically discovered mathematical formula hypotheses*.

This clustering is a result of a single iteration of the formula discovery algorithm (hence the multi-anchor clusters). Note the multi-anchor clusters of e and e^2, as well as the second order algebraic roots: these clusters failed to single out a specific constant, yet relate to constants of similar nature - suggesting meaningful clustering nevertheless.

4.3 Detecting Patterns and Underlying Structure

As mentioned in section 4.1, focusing on the B-dominant, $\eta' \approx 0$ cluster, gave rise to a multitude of formulas representing mathematical constants (see Fig.3c and d). They were discovered via a PSLQ algorithm, identifying linear fractional relations between the limit values of PCFs in the subset and notable mathematical constants (such as π or e). This is a computationally heavy operation, and it would be challenging to run it on all 1.5M formulas in the data set. Yet by first identifying the promising clusters, we reduce the search space $\sim 5,000$ times, allowing for a deeper inspection of each PCF.

Once the "checkerboard" pattern in Fig.3d was discovered, we expanded the hypothesis into 2 infinite families of PCFs with sub-exponential convergence relating to π and $\ln(2)$:

- $a_n = i + 2j + 1, b_n = n^2 + (i + k)n$, with integers $i, j \geq 0$, and $k \in \{0, 1\}$. This is expected to be related to π if $k = 1$, and to $\ln(2)$ if $k = 0$ (in fact, this pattern can be generalized even further, into a novel 3-dimensional Conservative Matrix Field, provided in Appendix C. See [Elimelech et al., 2023] for the definition of Conservative Matrix Fields).

Another formula family was discovered via clustering in the γ vs. γ' space. The algebraic roots subset (marked by a green circle in Fig.2c) was generalized into:

- $a_n = -2n + j - 1, b_n = -n^2 + jn + k$ for integer j, k such that b_n has real roots that are not positive integers. This is expected to converge to a root of b_n.

These are novel experimental results and mathematical hypotheses - awaiting proof.

5 Discussion and Outlook

This work marks an important step toward the vision of automated on-demand formula creation in mathematics. Going beyond all previous algorithms in this field, we connect the challenge of formula creation to modern approaches in AI for Science. The wide variety of novel results, from novel, automatically generated, conjectures to underlying mathematical structures and proofs, all demonstrate the power of our methodology.

The next research step directly building on our methodology could help to finally reveal the complete intricate mathematical structure of PCFs. For example, starting with the "band" structure found in Fig.2c. Further exploration of our conjectures from section 4 could have more impact on mathematics, perhaps achieving complete proofs and further generalizations.

The technique presented here can be applied to a larger scope of continued fractions and for completely different types of formulas. For more general continued fractions, dynamical metrics such as the numerical trajectories and the corresponding sequences of δ (in addition to its final value) hold valuable information even in continued fractions that do not converge at all. We expect these dynamical metrics to provide a "fingerprint" for wider families of continued fractions and perhaps even for the mathematical constants themselves. This approach will directly apply for higher polynomial degrees, larger polynomial coefficients, and for continued fractions not based on polynomials. Looking beyond continued fractions, metrics that are derived from the dynamics of a numerical calculation of certain formulas are an especially good fit for automated computer-assisted investigations. Such metrics can be measured for a variety of mathematical structures, like infinite sums, integral formulas, and partial differential equations. We believe that such dynamical metrics can unveil patterns and underlying structures in broad fields of mathematics and in other areas of science.

Our work was based on a limited-size dataset and on a small set of metrics. It would be intriguing to test the extracted conjectures on larger datasets, which can help reveal additional, more intricate, phenomena. Considering the success we had using a relatively small set of metrics, we would like to use an order-of-magnitude larger set of metrics and find what new predictions can be recovered, and whether qualitatively different types of predictions will arise.

Taking a broader perspective, the methodology presented in this work can be seen as a general prescription for tackling scientific discovery challenges, especially the ones considered as requiring intuitive leaps of extraordinary creativity, as in mathematics, theoretical physics, and a range of other fields of science and engineering.

6 References

References

Mihael Ankerst, Markus M. Breunig, Hans-Peter Kriegel, and Jörg Sander. Optics: ordering points to identify the clustering structure. *SIGMOD Rec.*, 28(2):49–60, jun 1999. ISSN 0163-5808. doi: 10.1145/304181.304187. URL `https://doi.org/10.1145/304181.304187`.

David Bailey, Jonathan Borwein, Neil Calkin, Russell Luke, Roland Girgensohn, and Victor Moll. *Experimental mathematics in action*. CRC press, 2007.

Lennart Berggren, Jonathan Borwein, Peter Borwein, and M Lambert. Mémoire sur quelques propriétés remarquables des quantités transcendentes circulaires et logarithmiques. *Pi: A Source Book*, pages 129–140, 2004.

Douglas Bowman and James McLaughlin. Polynomial continued fractions. *Acta Arith.*, 103:329–342, 2002.

Bruno Buchberger, Adrian Crăciun, Tudor Jebelean, Laura Kovács, Temur Kutsia, Koji Nakagawa, Florina Piroi, Nikolaj Popov, Judit Robu, Markus Rosenkranz, et al. Theorema: Towards computer-aided mathematical theory exploration. *Journal of applied logic*, 4:470–504, 2006.

Annie Cuyt, Vigdis Petersen, Brigitte Verdonk, Haakon Waadeland, and William B. Jones. *Handbook of continued fractions for special functions*. Springer Science & Business Media, 2008.

Nadav Ben David, Guy Nimri, Uri Mendlovic, Yahel Manor, and Ido Kaminer. On the connection between irrationality measures and polynomial continued fractions. *arXiv preprint arXiv:2111.04468*, 2021.

Alex Davies, Petar Veličković, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Peter Battaglia, Charles Blundell, András Juhász, et al. Advancing mathematics by guiding human intuition with AI. *Nature*, 600:70–74, 2021.

David L. Davies and Donald W. Bouldin. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227, 1979. doi: 10.1109/TPAMI.1979.4766909.

R. Davis and D.B. Lenat. *Knowledge-based Systems in Artificial Intelligence*. A McGraw-Hill advertising classic. McGraw-Hill International Book Company, 1982. ISBN 9780070155572. URL `https://books.google.co.il/books?id=MpVQAAAAMAAJ`.

Rotem Elimelech, Ofir David, Carlos De la Cruz Mengual, Rotem Kalisch, Wolfgang Berndt, Michael Shalyt, Mark Silberstein, Yaron Hadad, and Ido Kaminer. Algorithm-assisted discovery of an intrinsic order among mathematical constants. *arXiv preprint arXiv:2308.11829*, 2023.

Leonhard Euler. *Introductio in analysin infinitorum*, volume 1,2. Apud Marcum-Michaelem Bousquet & Socios, 1748.

Siemion Fajtlowicz. On conjectures of Graffiti. *Annals of Discrete Mathematics*, 38:113–118, 1988.

Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J R Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, et al. Discovering faster matrix multiplication algorithms with reinforcement learning. *Nature*, 610:47–53, 2022.

Helaman Ferguson and David Bailey. A Polynomial Time, Numerically Stable Integer Relation Algorithm. 1992.

Steven R. Finch. *Mathematical Constants*. Cambridge University Press, 2003.

Godfrey Harold Hardy, Edward Maitland Wright, et al. *An introduction to the theory of numbers*. Oxford university press, 1979.

Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL `https://doi.org/10.1038/s41586-020-2649-2`.

Cezary Kaliszyk and Josef Urban. Learning-assisted automated reasoning with flyspeck. *CoRR*, abs/1211.7012, 2012. URL `http://arxiv.org/abs/1211.7012`.

Jeffrey C. Lagarias. Euler's constant: Euler's work and modern developments. *Bulletin of the American Mathematical Society*, 50, 2013. ISSN 02730979. doi: 10.1090/S0273-0979-2013-01423-X.

James Mc Laughlin and Nancy J Wyshinski. Real numbers with polynomial continued fraction expansions. *arXiv preprint math/0402462*, 2004.

Douglas Lenat and John Brown. Why am and eurisko appear to work. volume 23, pages 236–240, 01 1983.

Douglas B. Lenat. The nature of heuristics. *Artificial Intelligence*, 19(2):189–249, 1982. ISSN 0004-3702. doi: https://doi.org/10.1016/0004-3702(82)90036-4. URL `https://www.sciencedirect.com/science/article/pii/0004370282900364`.

Marko Petkovsek, Herbert S. Wilf, and Doron Zeilberger. *A=B*. AK Peters Ltd, 1996.

Gal Raayoni, Shahar Gottlieb, Yahel Manor, George Pisha, Yoav Harris, Uri Mendlovic, Doron Haviv, Yaron Hadad, and Ido Kaminer. Generating conjectures on fundamental constants with the ramanujan machine. *Nature*, 590, 2021. ISSN 14764687. doi: 10.1038/s41586-021-03229-4.

Tal Ridnik, Dedy Kredo, and Itamar Friedman. Code generation with alphacodium: From prompt engineering to flow engineering. *arXiv preprint arXiv:2401.08500*, 2024.

Trieu H Trinh, Yuhuai Wu, Quoc V Le, He He, and Thang Luong. Solving olympiad geometry without human demonstrations. *Nature*, 625(7995):476–482, 2024.

Josef Urban. Malarea: a metasystem for automated reasoning in large theories. volume 257, 01 2007.

Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. doi: 10.1038/s41592-019-0686-2.

H. Wang. Toward mechanical mathematics. *IBM Journal of Research and Development*, 4(1):2–22, 1960. doi: 10.1147/rd.41.0002.

Stephen Wolfram et al. *A new kind of science*, volume 5. Wolfram media Champaign, 2002.

A Numerical Measurements and Curve Fitting

When characterizing PCFs, we use several metrics extracted from the dynamic behavior of the formula:

- The growth coefficients η, γ, β (of the form $n!^{\eta} \cdot e^{\gamma n} \cdot n^{\beta}$) of the convergence rate $\epsilon(n)$.
- $\tilde{q}_n$ (as defined in Eq.4) growth coefficients: η', γ' (of the form $n!^{\eta'} \cdot e^{\gamma' n}$).
- The δ (as defined in Eq.4) calculated using the Blind-δ Algorithm described in section 3.4.

To measure the growth coefficients of $\tilde{q}_n$ and $\epsilon(n)$, the values of $\log(\epsilon(n))$ (see section 3.4) and of $\log(\tilde{q}_n)$ were evaluated up to depth 1000.

The most resource-intensive values that are generated are p_n, q_n and $\gcd(p_n, q_n)$ - all other values are calculated from them (and require less precision). For the worst case PCF this requires 36MB of memory (without optimizations) and ~ 1.9 seconds of run time on a single core of a basic workstation, which translates to an upper cap of ~ 900 hours for the whole data set. In practice we used a high power cluster with 64 cores, which ran each iteration of the measurements in ~ 8.5 hours.

Once these values are calculated, using scipy [Virtanen et al., 2020] and numpy [Harris et al., 2020] a fit of the form $\log\left(n!^{\eta} \cdot e^{\gamma n} \cdot n^{\beta}\right)$ was calculated for $\tilde{q}_n$ and $\epsilon(n)$, producing the dynamic metrics.

A curve fit using 1000 points is a fairly heavy operation, unsuited for large scale investigations. Instead, we used an extreme down-sampling. Specifically, only 5 points were used for the fit. One may justifiably wonder if 5 data points are sufficient to fit accurately enough the desired metrics.

A test comparing between a 5 data point fit and a 1000 data point fit was done. As the test set, 50 PCFs were randomly chosen out of each of 9 categories (450 total test cases). The categories were all combinations of $\deg(a) = 0, 1, 2$ and $\deg(b) = 0, 1, 2$. Focusing on the dominant coefficients (γ and η), for each case, a full (1000 point) fit was performed (producing γ_f, η_f), and compared to the down sampled fit of 5 points (producing γ_p, η_p). We tested 2 methods of choosing the 5 points, even ($i = 6, 206, 406, 606, 806$) and logarithmic ($i = 6, 125, 250, 500, 1000$). The relative error was then calculated ($\frac{|\gamma_p - \gamma_f|}{|\gamma_f|}$ and $\frac{|\eta_p - \eta_f|}{|\eta_f|}$) for $\epsilon(n)$ and $\tilde{q}_n$. The relative errors were then averaged over the test set (results summarised in table 2) - showing the 5-point fit to be almost as good as the full 1000-point fit. In our measurements we use the logarithmic point distribution as it gives better results for most metrics.

Table 2: Comparison between 1000 point fit results and 5 point fits results (even spread and logarithmic spread).

Behavior	$\tilde{q}_n$		Convergence Rate	
Coefficient	γ'	η'	γ	η
Relative Error Average (even spread)	0.0124	0.0007	0.0078	0.0005
Relative Error Average (logarithmic spread)	0.0175	0.0004	0.0024	0.00012

Licences and versions of Python packages (used for curve fitting, clustering and large number mathematics):

Scipy (Version: 1.11.3) - BSD License (Copyright (c) 2001-2002 Enthought, Inc. 2003-2024, SciPy Developers. All rights reserved.)

gmpy2 (Version: 2.1.5) - GNU Lesser General Public License v3 or later

Numpy (Version 1.26.1)- BSD License (Copyright (c) 2005-2023, NumPy Developers. All rights reserved.)

B Classification of Continued Fractions

Not all PCFs converge. Clearly, if $\frac{p_n}{q_n}$ does not have a well defined limit, then some of our numerically measured metrics lose their meaning. Though we had algorithmic safeguards to detect such cases and remove them from the analyzed set, it was valuable to identify a pattern and formulate a rule-set that predicts the convergence of a PCF.

For that purpose we turned to the matrix representation of a continued fraction to depth n (see Appendix D.1 for details):

$$\begin{bmatrix} p_{N-1} & p_N \\ q_{N-1} & q_N \end{bmatrix} = \prod_{n=1}^{N-1} \begin{bmatrix} 0 & b_n \\ 1 & a_n \end{bmatrix} =$$
$$= \left(\prod_{n=1}^{N-1} a_{n-1} \right) \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{a_0} \end{bmatrix} \left(\prod_{n=1}^{N-1} \begin{bmatrix} 0 & \frac{b_n}{a_n a_{n-1}} \\ 1 & 1 \end{bmatrix} \right) \begin{bmatrix} 1 & 0 \\ 0 & a_{N-1} \end{bmatrix}$$

Assuming $a_n \neq 0$ for $n \geq 0$.

Analyzing the eigenvalues of the matrices within the matrix product as $n \to \infty$ allows for examining the asymptotic behavior of the continued fraction. Their characteristic polynomial is $\lambda^2 - \lambda - \frac{b_n}{a_n a_{n-1}}$, and we propose observing the discriminant of this polynomial - more specifically its dominant power of n:

$$\Delta_n = 1 + \frac{4b_n}{a_n a_{n-1}} = C_s n^s + O(n^{s-1}) \tag{10}$$

Here we assume $C_s \neq 0$ and s is some integer. Based on the data, we compiled table 3 as a summary of the conjectured behavior of any polynomial continued fraction based on s and C_s.

Table 3: Summary of PCF behavior characterized by s and C_s as defined in Eq.10.

Convergence	$C_s > 0$	$C_s < 0$
✗	$s \geq 3$	$s \geq 0$
✓	$s \leq 2$	$s \leq -1$

We can further elaborate on the converging cases by discussing the conjectured rate of convergence. Usually, a PCF is expected to converge at a sub-exponential rate, but in the case of $s = 0, C_0 > 0$ it is expected to converge faster:

- If $C_0 \neq 1$ then the PCF will converge at an exponential rate, and the exact rate of convergence increases monotonically as $C_0 \to 1$, with a vertical asymptote at $C_0 = 1$. The convergence rate is identical for C_0 and $\frac{1}{C_0}$.
- If $C_0 = 1$ then the PCF will converge at a factorial rate. More specifically, if we find the second most dominant power $\Delta_n = C_0 + \frac{C_t}{n^t} + O(\frac{1}{n^{t+1}})$ for some $C_t \neq 0$ and integer $t > 0$ then the precision will grow at a rate of $O(n!^t)$.

We used these rules (in conjunction with the measurements mentioned in section 3.3) to validate that all PCFs we analyze and cluster do converge and their measured metrics are well defined.

C Discovering equivalence of continued fractions

Polynomial continued fractions use two polynomials $a_n = a(n)$ and $b_n = b(n)$ to generate a sequence of rationals p_n/q_n. However, the same sequences with identical behaviour can be generated using more then one set of polynomials. By identifying transformations under which the dynamics of p_n/q_n remains invariant, we can formally prove equivalence between data points, validating the clustering power of the chosen metrics.

By rearranging the continued fraction definition, we can see how equivalent a_n and b_n series can arise:

$$a_0+\cfrac{b_1}{a_1+\cfrac{b_2}{a_2+\cfrac{b_3}{\ddots+\cfrac{b_n}{a_n+\ddots}}}} = a_0\left(1+\cfrac{\frac{b_1}{a_0a_1}}{1+\cfrac{\frac{b_2}{a_1a_2}}{1+\cfrac{\frac{b_3}{a_2a_3}}{\ddots+\cfrac{\frac{b_n}{a_na_{n-1}}}{1+\ddots}}}}\right) = \frac{a_0c_0}{c_0}\left(1+\cfrac{\frac{b_1c_0c_1}{a_0c_0a_1c_1}}{1+\cfrac{\frac{b_2c_1c_2}{a_1c_1a_2c_2}}{1+\cfrac{\frac{b_3c_2c_3}{a_2c_2a_3c_3}}{\ddots+\cfrac{\frac{b_nc_nc_{n-1}}{a_nc_na_{n-1}c_{n-1}}}{1+\ddots}}}}\right). \tag{11}$$

Indeed, by defining a new pair of polynomials $a'_n = a_nc_n; b'_n = b_nc_nc_{n-1}$ we get an equivalent continued fraction which converges to $\frac{c_0p_n}{q_n}$. Clearly, since the resulting sequence $\frac{p'_n}{q'_n}$ is identical to the original one, it exhibits the same dynamics. We call this process "Inflation by c_n". In particular, when $c_n = -1$, we observe that the sign of a does not affect the dynamics of the sequence - only flips the sign of the limit to $-L$. For every PCF its inflation by -1 is also contained in the data set, and clearly will have the same dynamics-based metrics. This equivalence single handedly de-facto cuts the size of the data set by half (to 771,963 converging formulas).

The metrics we are interested in are mostly not affected by a finite number of elements in the sequence. For example, both the convergence rate and δ discuss an overall trend as n grows. Consequently, we can initiate the sequence at different values of $n \neq 0$ without changing the latent parameters. When expressing these transformations as modification to the continued fraction definition, we see that the *limit* of the continued fraction might change due to this shift in sequence initiation, but only by a rational fractional transform.

$$a_0+\cfrac{b_1}{a_1+\cfrac{b_2}{a_2+\cfrac{b_3}{\ddots+\cfrac{b_n}{a_n+\ddots}}}} = \frac{p_n}{q_n} \Rightarrow a_1+\cfrac{b_2}{a_2+\cfrac{b_3}{\ddots+\cfrac{b_n}{a_n+\ddots}}} = \frac{b_1}{\frac{p_n}{q_n}-a_0}$$

For example, we consider the cluster of formulas related to the golden ratio shown in figure 3b. A large portion of these PCFs stem from transforming the known formula for the golden ratio shown in Eq.1 via the methods aforementioned. The exact transformations are depicted in Table 4.

D Analysis of the convergence rate

The growth rate for simple continued fraction or equivalently for constant linear recurrences is well understood, and usually boils down to the matrix defining the recurrence, and its eigenvalues. In our case, the coefficient in the recurrence also depend on n, so their study is more involved, however the ideas are similar, which we now describe

D.1 Approximating the error rate

To find whether or not the sequence $\frac{p_n}{q_n}$ converges and if so what is its convergence rate, we note the continued fraction formula

$$\cfrac{b(1)}{a(1)+\cfrac{b(2)}{a(2)+\cfrac{b(3)}{\ddots+\frac{b(n-1)}{a(n-1)+0}}}} = \frac{p_n}{q_n},$$

Table 4: Continued fractions converging to linear fractional transformations of the Golden Ratio ϕ, found using the top left cluster of Figure 3b. Numerous data points in this cluster exhibit identical sequence dynamics and are equivalent under the inflation and index indentation transformations. The equivalent data points create families of continued fractions in the cluster. Discrepancies between the calculated irrationality measure within the same family is ascribed to numerical inaccuracies, typically on the order 0.001. However, when comparing families, discrepancies in the irrationality measure rise to a magnitude of 0.04, suggesting potential deeper distinctions among these PCFs.

A_n	B_n	Limit	Transformation	Irrationality measure δ
1	1	ϕ	Family's canonical form	$\delta = 1.00168$
-1	1	-ϕ	Inflation by $c_n = -1$	$\delta = 1.00168$
2	4	2ϕ	Inflation by $c_n = 2$	$\delta = 1.00023$
-2	4	-2ϕ	Inflation by $c_n = -2$	$\delta = 1.00023$
$n+1$	$n(n+1)$	ϕ	Inflation by $c_n = n+1$	$\delta = 1.00168$
$-(n+1)$	$n(n+1)$	-ϕ	Inflation by $c_n = -(n+1)$	$\delta = 1.00168$
$n+2$	$(n+1)(n+2)$	2ϕ	Inflation by $c_n = n+2$	$\delta = 1.00023$
$-(n+2)$	$(n+1)(n+2)$	-2ϕ	Inflation by $c_n = -(n+2)$	$\delta = 1.00023$
$2n+1$	$(2n-1)(2n+1)$	ϕ	Inflation by $c_n = (2n+1)$	$\delta = 1.00168$
$-(2n+1)$	$(2n-1)(2n+1)$	-ϕ	Inflation by $c_n = -(2n+1)$	$\delta = 1.00168$
$2(n+1)$	$4n(n+1)$	2ϕ	Inflation by $c_n = 2(n+1)$	$\delta = 1.00023$
$-2(n+1)$	$4n(n+1)$	-2ϕ	Inflation by $c_n = -2(n+1)$	$\delta = 1.00023$
5	-5	$\phi+2$	Family's canonical form	$\delta = 1.00168$
-5	-5	$-(\phi+2)$	Inflation by $c_n = -1$	$\delta = 1.00168$
$5(n+1)$	$-5n(n+1)$	$\phi+2$	Inflation by $c_n = n+1$	$\delta = 1.00168$
$-5(n+1)$	$-5n(n+1)+0$	$-(\phi+2)$	Inflation $c_n = -(n+1)$	$\delta = 1.00168$
$n+2$	$n(n+3)$	$(30\phi+6)/19$	Family's canonical form	$\delta = 0.96967$
$-(n+2)$	$n(n+3)$	$-(30\phi+2)/19$	Inflation by $c_n = -1$	$\delta = 0.96967$
$n+3$	$(n+1)(n+4)$	$(30\phi+2)/11$	Indent $n \to n+1$	$\delta = 0.97245$
$-(n+3)$	$(n+1)(n+4)$	$-(30\phi+2)/11$	Indent $n \to n+1$ and inflation by $c_n = -1$	$\delta = 0.97245$
$n+3$	$n(n+5)$	$(750\phi+240)/361$	Family's canonical form	$\delta = 0.95243$
$-(n+3)$	$n(n+5)$	$-(750\phi+240)/361$	Inflation by $c_n = -1$	$\delta = 0.95243$

can be rewritten in matrix form as

$$\begin{pmatrix} p_{n-1} & p_n \\ q_{n-1} & q_n \end{pmatrix} = \prod_{1}^{n-1} \begin{pmatrix} 0 & b(k) \\ 1 & a(k) \end{pmatrix}.$$

In particular this implies that both p_n and q_n satisfy the same linear recurrence:

$$u_{n+1} = a(n)\, u_n + b(n)\, u_{n-1},$$

with initial conditions

$$\begin{pmatrix} p_0 & p_1 \\ q_0 & q_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Trying to determine if there is convergence, we use the Cauchy condition. For any $m \geq n$ we have that

$$\frac{p_m}{q_m} - \frac{p_n}{q_n} = \sum_{n}^{m-1} \left(\frac{p_{k+1}}{q_{k+1}} - \frac{p_k}{q_k} \right) = \sum_{n}^{m-1} \frac{p_{k+1}q_k - q_{k+1}p_k}{q_k q_{k+1}} = -\sum_{n}^{m-1} \frac{\det \begin{pmatrix} p_k & p_{k+1} \\ q_k & q_{k+1} \end{pmatrix}}{q_k q_{k+1}} = -\sum_{n}^{m-1} \frac{(-1)^k \prod_{j=1}^{k} b(j)}{q_k q_{k+1}}.$$

The sequence $\mathbb{K}_1^{\infty} \frac{b(n)}{a(n)}$ converges if and only if $\sum_1^{\infty} \frac{\prod_{j=1}^{k} b(j)}{q_k q_{k+1}}$ converges, and to the same limit L. More over, the convergence rate is

$$\epsilon(n) := \left| \frac{p_n}{q_n} - L \right| = \left| \sum_{n}^{\infty} \frac{(-1)^k \prod_{j=1}^{k} b(j)}{q_k q_{k+1}} \right|.$$

This suggests that we should understand the growth rate of both q_k and $\prod_{j=1}^{k} b(j)$. Note that the convergence and its rate might depend on the sign of $\frac{(-1)^k \prod_{j=1}^{k} b(j)}{q_k q_{k+1}}$.

1. Suppose that $\left|\frac{(-1)^k \prod_{j=1}^{k} b(j)}{q_k q_{k+1}}\right| = \frac{1}{k^d}$. If the signs do not alternate, then $\left|\sum_n^\infty \frac{(-1)^k \prod_{j=1}^{k} b(j)}{q_k q_{k+1}}\right| = \sum_n^\infty \frac{1}{k^d}$. This diverge if $d = 1$ and has order of magnitude $\frac{1}{k^{d-1}}$ for $d > 1$. However, with alternating signs we get the smaller bound

$$\sum_{2n}^{\infty} \frac{(-1)^k}{k^d} = \sum_{n}^{\infty} \left(\frac{1}{(2k)^d} - \frac{1}{(2k+1)^d}\right) = \sum_{n}^{\infty} \left(\frac{(2k+1)^d - (2k)^d}{(2k)^d (2k+1)^d}\right) \sim \sum_{n}^{\infty} \frac{d\,(2k)^{d-1}}{4k^{2d}} \sim \frac{1}{n^d}.$$

Thus, it always converges and with better rates.

2. However, for faster converging sequences we do not expect alternating sign to affect the convergence rate. For example, if $\left|\frac{\prod_1^{m-1}(-b(k))}{q_m q_{m-1}}\right| = \lambda^m$ for some $0 < \lambda < 1$, then with only positive signs the limit will be $\frac{\lambda^n}{1+\lambda}$ while for alternating signs it will be $\frac{(-\lambda)^n}{1+\lambda}$, so in any case the convergence rate is exponential.

D.2 Growth rate of $\prod_{k=1}^{m-1} |b(k)|$

Let $b(x)$ be a polynomial of degree d, with leading coefficient of absolute value B. Then there exists a constant $C > 0$ such that for any integer N we have

$$(Ne)^{-C} \le \prod_{k=1}^{N} \left|\frac{b(k)}{Bk^d}\right| \le (Ne)^{C}.$$

Proof. We may assume that the leading coefficient of b is positive. Writing $b(x) = \sum_0^d b_j x^j$ with $b_d = B \neq 0$, we want to approximate the product (of the absolute value) of

$$\tilde{b}(k) = 1 + \sum_{0}^{d-1} \frac{b_j}{B} \frac{1}{k^{d-j}}.$$

Hence, we can find an integer constant $C_0 \ge 1$ such that for all $k \ge 1$ we have

$$\left(1 - \frac{C_0}{k}\right) \le \left|\tilde{b}(k)\right| \le \left(1 + \frac{C_0}{k}\right).$$

For all k large enough, all the expression above are positive, so we get

$$\ln\left(1 - \frac{C_0}{k}\right) \le \ln\left|\tilde{b}(k)\right| \le \ln\left(1 + \frac{C_0}{k}\right).$$

With the goal of summing up these expressions from 1 to infinity, we claim that there is some constant $M > 0$ such that for any $C' \in$, and $2|C'| \le n < N$ we have that

$$\left|\sum_{n}^{N} \ln\left(1 + \frac{C'}{k}\right) - C' \ln\left(\frac{N}{n-1}\right)\right| \le M. \tag{12}$$

Given this claim we conclude that

$$-\left(C_0 \ln(N) + [M - C_0 \ln(2C_0)]\right) \le \sum_{k=2C_0+1}^{N} \ln\left|\frac{b(k)}{Bk^d}\right| \le C_0 \ln(N) + [M - C_0 \ln(2C_0)].$$

For another C large enough (independent of N), we can start the summation from $k = 1$ to get

$$-C(\ln(N) + 1) \le \sum_{k=1}^{N} \ln\left|\tilde{b}(k)\right| \le C(\ln(N) + 1).$$

Finally, exponenting it back we get the result we wanted:

$$(Ne)^{-C} \leq \prod_{k=1}^{N} \left|\tilde{b}(k)\right| \leq (Ne)^{C}.$$

We are left to prove Equation (12).

Using the Taylor expansion of $\ln(1+x)$ for $|x| \leq \frac{1}{2}$, we know that there is some large enough $0 < M_0$ such that

$$|\ln(1+x) - x| \leq M_0 x^2.$$

It follows that for $2|C'| \leq n < N$ we have

$$\left|\sum_{k=n}^{N}\left(\ln\left(1+\frac{C'}{k}\right) - \frac{C'}{k}\right)\right| \leq M_0 C'^2 \sum_{n}^{N} \frac{1}{k^2} \leq M_0 C'^2 \zeta(2).$$

In addition, we have that $\left|\sum_{n}^{N} \frac{1}{k} - \int_{n-1}^{N} \frac{1}{x}\right| \leq 1$, and

$$\int_{n-1}^{N} \frac{1}{x} = \ln\left(\frac{N}{n-1}\right).$$

Therefore

$$\left|\sum_{n}^{N} \ln\left(1+\frac{C'}{k}\right) - C' \ln\left(\frac{N}{n-1}\right)\right| \leq |C'| + M_0 C'^2 \zeta(2)$$

is uniformly bounded. □

D.3 Growth rate of q_n

The sequence q_n satisfies the linear recurrence

$$q_{n+1} = a(n)\, q_n + b(n)\, q_{n-1},$$

or in matrix form

$$(q_n, q_{n+1}) = (q_{n-1}, q_n) \overbrace{\begin{pmatrix} 0 & b(n) \\ 1 & a(n) \end{pmatrix}}^{M(n)}.$$

If both $a(x), b(x)$ are constant, and therefore $M = M(n)$ is a constant matrix, then this problem reduces to simply $(q_n, q_{n+1}) = (q_0, q_1) M^n$. Its a standard exercise to approximate q_n using the eigenvectors decomposition of M. However, in general not only $M(n)$ is non-constant, its entries have different orders of magnitude.

Thus, we would like to move to an "equivalent" system where at the very least $M(n)$ converges to some matrix M_∞, and then hope to show that the behavior of q_n can be read from the system with M_∞^n. This equivalent system will be built in two steps: first we "balance" the matrix, so its coordinates growth rate are the same, and then taking it outside as a scalar, the remaining sequence of matrices will converge.

D.3.1 Matrix balancing

This balancing is split into two cases according to the degrees of $d_a = \deg(a(x)), d_b = \deg(b(x))$.

Let $d = \max\left\{d_a, \frac{1}{2}d_b\right\}$ and denote by A, B the coefficients of x^d, x^{2d} of $a(x), b(x)$ respectively. Note that both A, B are either the corresponding leading coefficients or zero, depending on whether $d_a = d$, respectively $d_b = 2d$. If $2d_a < d_b$ and d_b is odd, then $d_a < d = \frac{d_b}{2}$, and we still consider A to be zero. Regardless of the choice of d, we see that at least one of A or B is not zero (and both if $2d_a = d_b$, which we call a "balanced" PCF).

With this choice, taking $\tilde{q}_n = \frac{q_n}{(n!)^d}$, we obtain the linear recurrence

$$\tilde{q}_{n+1} = \frac{a(n)}{(n+1)^d}\tilde{q}_n + \frac{b(n)}{(n(n+1))^d}\tilde{q}_{n-1}.$$

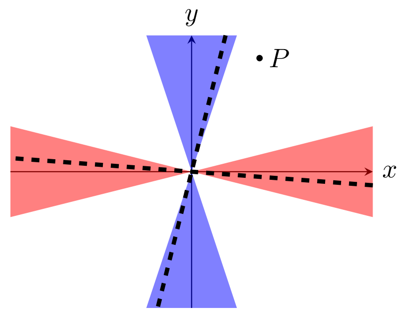

Figure 5: Convergence with variable coefficients

Letting $\tilde{a}\left(n\right)=\frac{a(n)}{(n+1)^{d}}$ and $\tilde{b}\left(n\right)=\frac{b(n)}{(n(n+1))^{d}}$, by our choice of d we see that the coefficient or the recurrence converge, and not both to zero:

$$\lim_{n\to\infty}\tilde{a}\left(n\right)=A$$
$$\lim_{n\to\infty}\tilde{b}\left(n\right)=B.$$

Here too we can also write it in a matrix form, namely

$$\left(\tilde{q}_{n},\tilde{q}_{n+1}\right)=\left(\tilde{q}_{n-1},\tilde{q}_{n}\right)\begin{pmatrix}0 & \tilde{b}\left(n\right)\\ 1 & \tilde{a}\left(n\right)\end{pmatrix}.$$

We now have a limit matrix, and the dynamics of such a matrix is well known. If both eigenvalues are real which are distinct in absolute value, then we expect exponential convergence. If both are non real, and therefore complex conjugate we expect it to behave like a rotation, and therefore will not converge. In both of these cases, since the eigenvalues are distinct in the limit, this holds for almost all n, so this behavior should hold in general.

In the discriminant zero, the situation is much more delicate, since we can converge to zero in many ways. For example, the discriminant along the way can be negative, positive or zero. In this notes we will restrict the study only to the two real eigenvalues with different absolute values.

D.3.2 Asymptotics of the continued fraction recurrence

The main goal of this section is to approximate the growth rate of a solution u_n to the recurrence

$$u_{n+1}=a_{n}u_{n}+b_{n}u_{n-1},$$

where both a_n, b_n converge (and not both to zero) or in matrix form

$$\left(\begin{smallmatrix}v_{n} & v_{n+1}\end{smallmatrix}\right)=\left(\begin{smallmatrix}v_{n-1} & v_{n}\end{smallmatrix}\right)M_{n}\quad,\quad M_{n}=\left(\begin{smallmatrix}0 & b_{n}\\ 1 & a_{n}\end{smallmatrix}\right),$$

where $M_{n}\to M:=\begin{pmatrix}0 & b\\ 1 & a\end{pmatrix}$.

The first step is the standard conjugation to a simpler matrix. Indeed, if $D=PMP^{-1}$ is simpler, e.g. diagonal, then $D_{n}:=PM_{n}P^{-1}\to D$, and $\prod_{1}^{n}M_{i}=P^{-1}\prod_{1}^{n}D_{i}P$, so we more or less need to understand $\prod_{1}^{n}D_{i}$.

In the constant diagonal case $D_{n}=\begin{pmatrix}\lambda_{1} & 0\\ 0 & \lambda_{2}\end{pmatrix}$ with $\lambda_{1}>\left|\lambda_{2}\right|$, we expect that for almost every initial condition $\left\Vert \left(\alpha_{1},\beta_{1}\right)D^{k}\right\Vert \sim\lambda_{1}^{k}$. This is true as long as the initial vector is not in $\cdot e_{2}$, and we have similar behaviour for other type of matrices. When the D_n are not constant, we need to take a little bit more care. The image you should have in mind is the following:

Instead of the two eigenvectors being on the X and Y axes, they only converge to it, so we only know that they are somewhere inside the red and blue regions. Thus, to understand this system we first need a **separation condition** saying that these regions are disjoint. Assuming the X-axis is the pulling axis (larger eigenvalue), we will need at least one point outside the error region around the Y axis, which we call the **initial condition**. Once both these conditions hold, a standard investigation of diagonalizable product will show that the point's orbit converge towards the eigenvector in the X-region. As this region shrinks to X in the limit, we see that the limit of the orbit is there as well. Suppose that $D_n \to D$ where $D = \left(\begin{smallmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{smallmatrix}\right)$ with $0 \leq \left|\frac{\lambda_-}{\lambda_+}\right| < 1$ and let $\kappa_n = \frac{1}{|\lambda_+|} \max_{k \geq n} \|D_k - D\|_\infty$.

Fix some initial z_1 and let $z_k = (z_1) \prod_1^{k-1} D_n$. Assuming that for some n we have

- **Separation condition**: $\left|\frac{\lambda_-}{\lambda_+}\right| + 4\kappa_n < 1$ and
- **Initial condition**: $|z_n| < \frac{\mu_n + \sqrt{(\mu_n - 2\kappa_n)(\mu_n + 2\kappa_n)}}{2\kappa_n}, \ \mu_n = 1 - \left|\frac{\lambda_-}{\lambda_+}\right| - 2\kappa_n.$

Then $\lim\limits_{k \to \infty} |z_k| = 0$.

Note that $\frac{\mu_n + \sqrt{(\mu_n - 2\kappa_n)(\mu_n + 2\kappa_n)}}{2\kappa_n} \sim \frac{1 - |\lambda|}{\kappa_n} \to \infty$ as $\kappa_n \to 0$, so this initial condition becomes easier to satisfy as $n \to \infty$.

Proof. First, proving the claim for $\frac{1}{\lambda_+} D_n$ instead of D_n, we may assume that the limit is $D = \begin{pmatrix} 1 & 0 \\ 0 & \lambda \end{pmatrix}$ where $\lambda = \frac{\lambda_-}{\lambda_+}$.

Next, note that whenever $\mu := 1 - |\lambda| - 2\kappa > 2\kappa > 0$, we have that $\sqrt{(\mu - 2\kappa)(\mu + 2\kappa)} = \sqrt{\mu^2 - 4\kappa^2} < \mu$. Setting

$$\nu_\pm(\kappa) = \frac{\mu_\kappa \pm \sqrt{(\mu_\kappa - 2\kappa)(\mu_\kappa + 2\kappa)}}{2\kappa},$$

we get that $0 < \nu_-(\kappa) < \nu_+(\kappa)$ are real numbers, and the condition in the assumption is $|z_n| < \nu_+(\kappa_n)$. Our main goal is to prove our process satisfies:

1. $|z_k| < \nu_+(\kappa_n)$ for all $k \geq n$ and,
2. We have $\limsup_k |z_k| \leq \nu_-(\kappa_n)$.

Assuming these two steps are true, the full proof is not too far behind. Indeed, since $D_n \to D$, the sequence $\kappa_n := \sup\limits_{k \geq n} \|D_n - D\|$ converges to zero, and note that as $\kappa_n \to 0$ we get that $\nu_+(\kappa_n) \nearrow \infty$ and $\nu_-(\kappa_n) \searrow 0$. Assuming step (1), for $k \geq m \geq n$ we have $|z_k| < \nu_+(\kappa_n) \leq \nu_+(\kappa_m)$, and by step (2) we get that $\limsup_k |z_k| \leq \nu_-(\kappa_m) \to 0$.

For the remaining of the proof, without loss of generality we may assume that $n = 1$ and just write κ, μ instead of κ_n, μ_n.

To prove these two steps, consider the change from z_k to z_{k+1}. Writing $D_k = \left(\begin{smallmatrix} 1+\varepsilon_{1,1} & \varepsilon_{1,2} \\ \varepsilon_{2,1} & \lambda+\varepsilon_{2,2} \end{smallmatrix}\right)$, since $z_{k+1} = (z_k) D_k$ and $\|D - D_k\|_\infty \leq \kappa$, we get that

$$|z_{k+1}| = \left|\frac{\varepsilon_{1,2} + z_k(\lambda + \varepsilon_{2,2})}{(1 + \varepsilon_{1,1}) + z_k \varepsilon_{2,1}}\right| \leq \frac{\kappa + |z_k|(|\lambda| + \kappa)}{1 - \kappa - \kappa|z_k|}.$$

Note that the final denominator is positive, so that the last inequality is valid. Indeed, using the conditions of the claim we get

$$1 - \kappa(1 + |z_k|) \geq 1 - \kappa\left(1 + \frac{\mu + \sqrt{(\mu - 2\kappa)(\mu + 2\kappa)}}{2\kappa}\right) > 1 - (\kappa + \mu) = |\lambda| + \kappa > 0.$$

Thus, we can rewrite the inequality as

$$|z_{k+1}| \leq M_\varepsilon(|z_k|), \quad M_\varepsilon = \begin{pmatrix} |\lambda| + \kappa & \kappa \\ -\kappa & 1 - \kappa \end{pmatrix}. \tag{13}$$

The goal now is to show that if $|z_k|$ is "large", then $|z_{k+1}|$ is much smaller, and if $|z_k|$ is small, then $|z_{k+1}|$ cannot increase too much.

A simple computations shows that the eigenvalues of this matrix are

$$\gamma_{\pm} = \frac{|\lambda| + 1 \pm \sqrt{(\mu + 2\kappa)(\mu - 2\kappa)}}{2},$$

and since $\sqrt{(\mu + 2\kappa)(\mu - 2\kappa)} \le \mu \le 1 - |\lambda|$, we get that

$$\gamma_+ > \gamma_- > 0.$$

Finally, the corresponding (right) eigenvectors are

$$v_{\pm} = \begin{pmatrix} \nu_{\mp} \\ 1 \end{pmatrix}.$$

To simplify the notations, let us conjugate by the matrix $T = \begin{pmatrix} \nu_+ & \nu_- \\ 1 & 1 \end{pmatrix}$ to obtain

$$T^{-1} M_\varepsilon T = \begin{pmatrix} \gamma_- & 0 \\ 0 & \gamma_+ \end{pmatrix}.$$

Note that the Mobius map

$$T^{-1}(z) := \frac{1}{\nu_+ - \nu_-} \begin{pmatrix} 1 & -\nu_- \\ -1 & \nu_+ \end{pmatrix} (z) = -\frac{z - \nu_-}{z - \nu_+} = -1 + \frac{\nu_- - \nu_+}{z - \nu_+}$$

sends $\nu_- \mapsto 0$, $\nu_+ \mapsto \infty$ and $0 \mapsto -\frac{\nu_-}{\nu_+} < 0$. In particular, it is monotone increasing on $[0, \nu_+)$, so that our two steps from above are equivalent to

1. $T^{-1}(|z_k|) \in [-\frac{\nu_-}{\nu_+}, \infty)$,
2. $\limsup_k T^{-1}(|z_k|) \in \left[-\frac{\nu_-}{\nu_+}, 0\right]$,

and the claim's original assumption is that $T^{-1}(|z_1|) \in [-\frac{\nu_-}{\nu_+}, \infty)$. However, now this claim is simple, since in these notations we get that

$$T^{-1}(M_\varepsilon(|z_k|)) = \left(T^{-1} M_\varepsilon T\right)\left(T^{-1}(|z_k|)\right) = \frac{\gamma_-}{\gamma_+} \cdot T^{-1}(|z_k|),$$

and $0 < \frac{\gamma_-}{\gamma_+} < 1$. Thus, if $T^{-1}(|z_k|) \in [-\frac{\nu_-}{\nu_+}, \infty)$, then so is $T^{-1}(M_\varepsilon(|z_k|)) \in [-\frac{\nu_-}{\nu_+}, \infty)$, so by Equation (13) and the monotonicity of T, we obtain that

$$T^{-1}(|z_{k+1}|) \le \frac{\gamma_-}{\gamma_+} \cdot T^{-1}(|z_k|),$$

which implies the two steps. □

Returning back to the recursion, we get the following

Suppose that we have a solution to the recurrence $v_{n+1} = a_n v_n + b_n v_{n-1}$, where $a_n \to a, b_n \to b$ and suppose that $\lambda_\pm$ are the roots of $x^2 = ax + b$ with $0 \le |\lambda_-| < \lambda_+$. Writing $\kappa'_n = \frac{1}{|\lambda_+|} \max\limits_{k \ge n} \max\{|a_k - a|, |b_k - b|\}$ and $C(\lambda_\pm) := \frac{1 + |\lambda_+|}{|\lambda_+ - \lambda_-|}$, Assume that for some n we have

- **Separation condition**: $\left|\frac{\lambda_-}{\lambda_+}\right| + 4C(\lambda_\pm)\kappa'_n < 1$ and
- **Initial condition**: $\left|\lambda_- - \frac{v_n}{v_{n-1}}\right| \ge C(\lambda_\pm)\kappa'_n \frac{|\lambda_+ - \lambda_-|}{1 - \left|\frac{\lambda_-}{\lambda_+}\right| - 4C(\lambda_\pm)\kappa'_n}$,

Then

$$\frac{v_n}{v_{n-1}} \to \lambda_+.$$

Proof. Set $M_n = \left(\begin{smallmatrix} 0 & b_n \\ 1 & a_n \end{smallmatrix}\right)$ and $M = \left(\begin{smallmatrix} 0 & b \\ 1 & a \end{smallmatrix}\right)$ as in the beginning of this section. With $P = \left(\begin{smallmatrix} 1 & \lambda_+ \\ 1 & \lambda_- \end{smallmatrix}\right)$ and $P^{-1} = \frac{1}{\lambda_- - \lambda_+}\left(\begin{smallmatrix} \lambda_- & -\lambda_+ \\ -1 & 1 \end{smallmatrix}\right)$ we have that $D = PMP^{-1} = \left(\begin{smallmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{smallmatrix}\right)$. We would like to apply Lemma D.3.2 to the matrices $D_n = PM_nP^{-1}$.

For the **separation condition** on the infinity norm, we have

$$\left\|PM_nP^{-1} - D\right\|_\infty = \left\|P\left(M_n - M\right)P^{-1}\right\|_\infty = \frac{1}{|\lambda_- - \lambda_+|}\left\|\left(\begin{smallmatrix} 1 & \lambda_+ \\ 1 & \lambda_- \end{smallmatrix}\right)\left(\begin{smallmatrix} 0 & b_n - b \\ 0 & a_n - a \end{smallmatrix}\right)\left(\begin{smallmatrix} \lambda_- & -\lambda_+ \\ -1 & 1 \end{smallmatrix}\right)\right\|_\infty$$

$$= \frac{1}{|\lambda_- - \lambda_+|}\left\|\left(\begin{smallmatrix} 1 & \lambda_+ \\ 1 & \lambda_- \end{smallmatrix}\right)\left(\begin{smallmatrix} b - b_n & b_n - b \\ a - a_n & a_n - a \end{smallmatrix}\right)\right\|_\infty \le \overbrace{\frac{1 + |\lambda_+|}{|\lambda_+ - \lambda_-|}}^{C(\lambda_\pm)}\left\|M_n - M\right\|_\infty.$$

Thus, the separation condition of this theorem implies the separation condition of Lemma D.3.2:

$$\left|\frac{\lambda_-}{\lambda_+}\right| + 4\kappa_n \le \left|\frac{\lambda_-}{\lambda_+}\right| + 4C\left(\lambda_\pm\right)\frac{1}{|\lambda_+|}\max_{k\ge n}\left\|M_n - M\right\|_\infty < 1$$

Next, for the **initial condition** , setting

$$\left(v_{k-1}\ v_k\right) := \left(v_0\ v_1\right)\left(\prod_1^{k-1} M_n\right) = \left(v_0\ v_1\right)P^{-1}\left(\prod_1^{k-1} D_n\right)P$$

we have

$$\left(\alpha_k, \beta_k\right) = \left(v_0\ v_1\right)P^{-1}\left(\prod_1^{k-1} D_n\right) = \left(v_{k-1}, v_k\right)P^{-1}.$$

Setting $z_n = \frac{\beta_n}{\alpha_n}$, we get that

$$|z_n| = \left|\frac{\beta_n}{\alpha_n}\right| = \left|\frac{-\lambda_+ v_{n-1} + v_n}{\lambda_- v_{n-1} - v_n}\right| = \left|1 + \frac{\lambda_+ - \lambda_-}{\lambda_- - \frac{v_n}{v_{n-1}}}\right| \le 1 + \left|\frac{\lambda_+ - \lambda_-}{\lambda_- - \frac{v_n}{v_{n-1}}}\right| = (*).$$

Using the assumption that $\left|\lambda_- - \frac{v_n}{v_{n-1}}\right| \ge C\left(\lambda_\pm\right)\kappa_n' \frac{|\lambda_+ - \lambda_-|}{1 - \left|\frac{\lambda_-}{\lambda_+}\right| - 4C(\lambda_\pm)\kappa_n'} \ge \kappa_n \frac{|\lambda_+ - \lambda_-|}{1 - \left|\frac{\lambda_-}{\lambda_+}\right| - 4\kappa_n}$, we see that the expression above is

$$(*) \le 1 + \frac{|\lambda_+ - \lambda_-|}{\kappa_n \frac{|\lambda_+ - \lambda_-|}{1 - \left|\frac{\lambda_-}{\lambda_+}\right| - 4\kappa_n}} = 1 + \frac{1 - \left|\frac{\lambda_-}{\lambda_+}\right| - 4\kappa_n}{\kappa_n} = \frac{2\mu_n - 2\kappa_n}{2\kappa_n} < \frac{\mu_n + \sqrt{\left(\mu_n - 2\kappa_n\right)\left(\mu_n + 2\kappa_n\right)}}{2\kappa_n}.$$

This was the second condition needed for Lemma D.3.2 , so we can now conclude that

$$\left|1 + \frac{\lambda_+ - \lambda_-}{\lambda_- - \frac{v_n}{v_{n-1}}}\right| = \left|\frac{\beta_n}{\alpha_n}\right| \to 0$$

which implies that $\frac{v_n}{v_{n-1}} \to \lambda_+$. □

D.4 Conclusion

We return now to the original problem with $\alpha = \mathbb{K}_1^\infty \frac{b(n)}{a(n)}$ and assume that $a\left(x\right), b\left(x\right)$ have degrees d_a, d_b . As mentioned before, we split our study into two cases:

The balanced case

Assume that $d_b = 2d_a = 2d$, and let A, B be the leading coefficients of $a\left(x\right), b\left(x\right)$ respectively. In this case the limit matrix is $M_\infty = \begin{pmatrix} 0 & B \\ 1 & A \end{pmatrix}$, and we assume that the roots $\lambda_\pm$ of $x^2 = Ax + B$ satisfy $0 < |\lambda_-| < \lambda_+$. Using Theorem D.3.2 once the two conditions hold, we obtain

$$\frac{q_{n+1}}{q_n}\left(n+1\right)^d = \frac{q_{n+1}/\left(n+1\right)!^d}{q_n/n!^d} \to \lambda_+,$$

implying that $q_n = n!^d \lambda_+^n \exp(o(n))$. As for the product of the $b(k)$, using Claim D.2 we have that

$$\prod_{k=1}^{N} |b(k)| = \exp(o(N)) \cdot B^N \cdot N!^{2d}.$$

Putting them together as in the error rate expression, we get :

$$\frac{\prod_{k=1}^{m-1} |b(k)|}{|q_{m-1} q_m|} = \frac{|B|^{m-1} \cdot (m-1)!^{2d}}{(m-1)!^d (m)!^d \lambda_+^{2m-1}} \exp(o(m)) = (*) .$$

Note that $|B| = |\det(M_\infty)| = |\lambda_- \lambda_+|$, so that the expression above is

$$(*) = |\lambda_- / \lambda_+|^m \cdot \exp(o(m)) = \exp(m \log |\lambda_- / \lambda_+| + o(m)) .$$

Thus, for given $\varepsilon > 0$ where $\left|\frac{\lambda_-}{\lambda_+}\right| + \varepsilon < 1$, and for any m large enough we see that $(*) \leq \left(\left|\frac{\lambda_-}{\lambda_+}\right| + \varepsilon\right)^{m-1}$. We conclude that the error rate for all n large enough is bounded from above by

$$\left|\frac{p_n}{q_n} - \alpha\right| \leq \sum_{m=n+1}^{\infty} \frac{\prod_{k=1}^{m-1} |b(k)|}{|q_{m-1} q_m|} \leq \sum_{m=n+1}^{\infty} \left(\left|\frac{\lambda_-}{\lambda_+}\right| + \varepsilon\right)^{m-1} = \left(\left|\frac{\lambda_-}{\lambda_+}\right| + \varepsilon\right)^{n} \frac{1}{1 - \left(\left|\frac{\lambda_-}{\lambda_+}\right| + \varepsilon\right)}.$$

It follows that

$$\ln\left|\frac{p_n}{q_n} - \alpha\right| \leq n \ln\left(\left(\left|\frac{\lambda_-}{\lambda_+}\right| + \varepsilon\right)\right) - \ln\left(1 - \left(\left|\frac{\lambda_-}{\lambda_+}\right| + \varepsilon\right)\right) \sim n \ln\left(\left|\frac{\lambda_-}{\lambda_+}\right|\right).$$

The unbalanced case

Suppose now that $d_b < 2d_a = 2d$, so that $B = \lim_{n\to\infty} \frac{b(n)}{(n(n+1))^{d_a}} = 0$. This time the two roots of $x^2 = Ax + 0$ are $\lambda = 0, A$. If needed, we can use a simple continued fraction inflation $\mathbb{K}_1^\infty \frac{(-1)^2 b(n)}{(-1) a(n)}$ and assume that $A > 0$. Using Theorem D.3.2, if the two conditions hold, we obtain $q_n = n!^d A^n \exp(o(n))$.

Letting $\hat{B}$ be the leading coefficient of $b(x)$ in absolute value, Claim D.2 implies that

$$\prod_{k=1}^{N} |b(k)| = \exp(o(N)) \cdot \hat{B}^N \cdot N!^{d_b}.$$

Again, together we obtain that

$$\frac{\prod_{k=1}^{m-1} |b(k)|}{|q_{m-1} q_m|} = \frac{\hat{B}^{m-1} \cdot (m-1)!^{d_b}}{(m-1)!^d m!^d A^{2m-1}} \exp(o(m)) = \frac{1}{(m-1)!^{2d_a - d_b}} \cdot \left(\frac{\hat{B}}{A^2}\right)^m \exp(o(m))$$

Similarly to the previous case, given $\varepsilon > 0$, and using the fact that $2d_a - d_b \geq 1$, for all n large enough we obtain

$$\begin{aligned}
\left|\frac{p_n}{q_n} - \alpha\right| &\leq \sum_{m=n+1}^{\infty} \frac{\prod_{k=1}^{m-1} |b(k)|}{|q_{m-1} q_m|} \leq \sum_{m=n+1}^{\infty} \frac{1}{(m-1)!^{2d_a - d_b}} \cdot \left(\frac{\hat{B}}{A^2} + \varepsilon\right)^{m-1} \\
&= \frac{1}{n!^{2d_a - d_b}} \left(\frac{\hat{B}}{A^2} + \varepsilon\right)^n \sum_{m=0}^{\infty} \left(\frac{n!}{(n+m)!}\right)^{2d_a - d_b} \cdot \left(\frac{\hat{B}}{A^2} + \varepsilon\right)^m \\
&\leq \frac{1}{n!^{2d_a - d_b}} \left(\frac{\hat{B}}{A^2} + \varepsilon\right)^n \left[\sum_{m=0}^{\infty} \left(\frac{1}{m!}\right)^{2d_a - d_b} \cdot \left(\frac{\hat{B}}{A^2} + \varepsilon\right)^m\right].
\end{aligned}$$

The infinite sum in the last exression converges to some finite limit $\tilde{C}$, so we conclude that

$$\ln\left|\frac{p_n}{q_n} - \alpha\right| \leq (d_b - 2d_a) \ln(n!) + n \ln\left|\frac{\hat{B}}{A^2} + \varepsilon\right| + \ln\left|\tilde{C}\right| \sim (d_b - 2d_a)\, n \cdot \ln|n| .$$

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes, the main results, motivations and aspirational goals are included in the abstract (section 1) and introduction (section 2).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We discuss limitations and required assumptions of the Blind-δ Algorithm in section 3.4, the dynamical metric estimation via down-sampled curve fitting in Appendix A, and of PCF convergence and the delta-predictor formula (Eq.5) in Appendix D.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory Assumptions and Proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [Yes]

Justification: Most of the results in this work are mathematical hypotheses (numerically validated) - presented without proof. The main theoretical result (Eq.5) is proven in Appendix D.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental Result Reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: The data set used was explicitly defined, the algorithmic steps presented and numerical techniques described.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code is submitted for review, and the git repository will be linked in the camera-ready version. The data set is mathematical formulas and constants - so freely available to all.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental Setting/Details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: In this work we are using unsupervised recurring clustering as the main machine learning method, and the results are mostly validated using the (test) subset of known PCF formulas for mathematical constants. This subset was included in the full data set but wasn't given any special treatment during clustering (as mentioned in section 4.1).

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment Statistical Significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: The main errors in this work originate in the dynamical metrics measurement stage. Error bars are not shown (for visual clarity), but these errors are discussed in Appendix A, and mentioned in every figure that shows measured dynamical metrics.

Guidelines:

- The answer NA means that the paper does not include experiments.

- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: There are no special or extreme resource requirements, so we did not focus on this question, but a brief discussion of required memory and runtime is in Appendix A.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code Of Ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This work does not involve human participants, special data sets, does not have implications for the broader public (only for scientists) and cannot be used to do harm.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader Impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

Justification: Although our work is applied ML, its impact is on mathematical research, not society at large.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The methods in this paper do not pose such risks.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The only existing assets that were used are open source Python packages, that were properly credited in Appendix A.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.

- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The code is released as open source with a readme and documentation.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and Research with Human Subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This research does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This research does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.