natureresearch

Peer Review File

Manuscript Title: The Ramanujan Machine – Auto-Generated Conjectures on Fundamental Constants

Editorial Notes:

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1 (Remarks to the Author):

The paper presents two algorithms for finding continued fraction representations of fundamental constants. The algorithms are numerical and therefore provide conjectures rather than formal proofs of the identities. The first algorithm (Meet-in-the-Middle algorithm) is based on an optimized enumeration of RHS and LHS of the formulas, matching these values, and subsequent validation with higher precision. The second algorithm (Descent and Repel) is based on a gradient descent optimization procedure. Using these algorithms, the paper provides a set of conjecture identities on fundamental constants, some of which are claimed to be new.

On the positive side, leveraging automation to generate new mathematical properties is a very promising and interesting area. However, the paper fails to make a strong contribution to this field.

- The problem that is being studied is elementary, and is not representative of the real complexity of most mathematical problems. This is problematic, as the paper claims to be a step in the path of using algorithms to "unveil mathematical structure", and "play the role of intuition of great mathematicians in the past, providing leads to new mathematical research". More precisely, the search space in the problem of finding PCFs is very small, which leads to exhaustive approaches working quite well. This is in stark contrast with most mathematical problems, where interesting properties have to be found in much larger search spaces (and are also hard to formalize).

- The paper could still make contributions with large impact while remaining specific to the considered problem. For example, finding more efficient ways of computing fundamental constants using PCFs, finding relations between constants, etc... The paper remains instead speculative on those problems (lots of conditionals that these will be possible in the future) but does not make any concrete contributions towards these more impactful problems.

- On the algorithmic side, the contributions are not strong enough, with the two proposed approaches being a variant of exhaustive search (with discretization of the search space), and gradient descent. In the way they are used, these are also quite specific to the problem of finding PCFs, and it is unclear how these could generalize (e.g., to problems with larger search space).

- It is unclear what the authors exactly mean by "new" conjecture. In fact, most conjectures in Table 4 are labeled as "new and proven" (e.g., 4 / (pi - 2)), while the proof in Appendix F.2 is often a mere specialization of Gauss's continued fraction. This should not be considered as a new result.

- The presentation of the algorithmic part can be significantly improved; for example:
- In Section 4, it is mentioned that "we empirically observed that all minima are global, and their errors are zero. Therefore any GD process will result in a solution with L = 0". This needs more justification.
- It is mentioned that repel mechanism is used to increase the search space, and thus "the probability of finding a match in space". In light of the above comment, do all initial conditions lead to a solution?
- The paragraph after Equation 7 is unclear, and its logic should be re-visited: Unclear why the dimensionality of the manifold is relevant to the discussion, unclear whether GD refers to vanilla gradient descent or to the version with discretization — especially since the last sentence in the

paragraph infers a property on the problem with integer constraints.
- What are x and y axes in Figure 4.
- Motivation of Descent and repel. It is mentioned that MITM-RF is not "scalable" — mention more explicitly to which parameters it is desired to scale up.

Referee #2 (Remarks to the Author):

Dear Editors,

I read the submission

"The Ramanujan Machine: Automatically Generated Conjectures on Fundamental Constants"
By: Raayoni et al
MS: 2020-04-07825

with great interest.

The paper attempts to generate new identities, specifically polynomial continued fractions, by generating large sets of potential candidates and then using a gradient-descent optimization to zoom in "correct" ones by checking to hundreds of digits of precision.
Consequently, many correct expressions are found, and impressively, new conjectures have been raised, some of which have been proven since the appearance of the paper on ArXiv and several more still open.

As stated in the appendix and the conclusions, this exercise has prompted an interactive website (in the spirit of PolyMath, GIMP, etc,) www.RamanujanMachine.com
that has inspired the mathematics community to prove some of the new conjectures.

I am very sympathetic to this experimental approach to mathematics using the best resources of today: computer algebra and machine-learning.
I therefore recommend the article for publication in Nature subject to the following revisions:

- It would be good to include a discussion on why PCFs are chosen (out of the myriad of possible mathematical structures) for the experimentation and why this is a perfectly adapted problem to their philosophy. An excellent summary of the subject is

Bowman and J. Mc Laughlin, Polynomial continued fractions, Acta Arithmetica, 103(4) (2002), 329–342

and should be cited.

- In Sec 2 for Related Works, it might be worth citing explorations in supervised and unsupervised machine-learning (very much in the spirit of this paper in attempting to find structure and generate new conjectures, and not in the style of ATP) have been applied to study of
the physical laws
"Discovering Physical Concepts with Neural Networks",
Raban Iten et al, PHYSICAL REVIEW LETTERS 124, 010508 (2020);

"Deep-Learning the Landscape", by Y.-H. He
https://arxiv.org/abs/1706.02714, Phys.Lett.B 774 (2017) 564-568

and number theory

"Machine Learning meets Number Theory: The Data Science of Birch- Swinnerton-Dyer", L. Alessandretti, A. Baronchelli, Y.-H. He
https://arxiv.org/abs/1911.02008

- Eq (5): Can the author give a brief account of the type (or full list) of the polynomials \alpha, \beta, \gamma, \delta as well as the function f_i (e.g. what does i index? how many such function are tried?) that was used in the search? How exhaustive was it? Also, please present an idea of the running time and on what machine.

- Could the authors comment on the success of generating PCFs for algebraic (e.g. Golden ratio) versus transcendental constants (e.g. Pi); do they differ? This would be substantial interest to the number theory community.

- The Descent&Repel algorithm needs to be clarified (very much line with the comment above)
* Move the comment about the variables to be optimized coming from
\alpha, \beta, \gamma, \delta, as coefficients in these polynomials to just below Eq (7), before saying there are d of these. What is the typical number of d? i.e., what are the degrees of the polynomials considered? Presumably, Fig 4, where d = 2, is only a schematic illustration.

* Why does repulsion help? i.e., why shouldn't one consider integer solutions to the coefficients which are close together in solution space?
* Write the integer-round loss function along side with {\cal L}
and explain how the 2 are used alternatingly

* in step 1 of the algorithm, what is x_t? presumably \mu is step size and x_t denote the d variables and L is script-L?

* In Fig 4, the colour-legend needs to be explained. Why is there no red? Which Log-error is used? the integer-round one or the one
in Eq 7. What are (x,y) ?

- A scope of the constants used in the search algorithm (LHS of Eq 5) should be listed explicitly, i.e., which constants (pi, e, Catalan, etc) have been searched, and which have produced good hits. This is particularly relevant to Table 2. Are there any in table which did get a hit and thus a conjecture?

- Small typographical errors such as:
* All equations and footnotes (e.g. footnote 5) need to be punctuated;
* The quotation marks are all wrongly type-set; make sure to use
` instead of ' for the beginning of the quote in LaTeX.
* Top left box of Figure 2, is that supposed to be PCF?

<u>Referee #3 (Remarks to the Author):</u>

A. This is a seminal paper describing two very efficient
algorithms to generate intriguing conjectures with
far-reaching potential applications.

B. While the LLL and PSQL were used in an ad hoc way before,
the systematic and unified approach described here is very novel
and significant.

C. Very valid and the authors share all their ample data and
output in their web-site for the benefit of the mathematical
world.

D. Since this is mathematics, the standards are much higher
than in the physical scientist, and all their conjectures
are virtually certain, even those still awaiting a formal proof.

E. Perfect

F. The references should be carefully copy-edited.
e.g. (put please check everything)

Ref. 18: the authors' names are reversed and they should include first names (or at least first initial)

Ref. 20: A=b -> A=B

Ref. 45: ditto (and also the title is wrong)

Ref. 60: ditto

G. See above, otherwise perfect

H. Very well-written, lucid, and engaging. A true tour-de-forice

Author Rebuttals to Initial Comments:

Referee #1 (Remarks to the Author):

The referee wrote-
The paper presents two algorithms for finding continued fraction representations of fundamental constants. The algorithms are numerical and therefore provide conjectures rather than formal proofs of the identities. The first algorithm (Meet-in-the-Middle algorithm) is based on an optimized enumeration of RHS and LHS of the formulas, matching these values, and subsequent validation with higher precision. The second algorithm (Descent and Repel) is based on a gradient descent optimization procedure. Using these algorithms, the paper provides a set of conjecture identities on fundamental constants, some of which are claimed to be new.

On the positive side, leveraging automation to generate new mathematical properties is a very promising and interesting area.

Reply-
We thank the referee for highlighting some of the novelty of our work and finding it promising.

The referee wrote-
However, the paper fails to make a strong contribution to this field.

Reply-
While we value the response and input, we respectfully disagree regarding the contributions of the previous version of the manuscript to the field. Our manuscript provided new formulas that were previously unknown, and whose origin (in terms of the mathematical structure they result from) is yet to be understood. These formulas include representations of the Catalan's constant and Apéry's constant presented in Table 5.

Some of the conjectures we found with the algorithms have led to new mathematical research, as shown, for example, in [arXiv:2004.00090] by Dougherty-Bliss and Zeilberger. They developed our formulas into generalized expressions with richer mathematical structure and also presented proofs for these expressions.

Having said that, we also very much understand the referee's point of view. The previous version of our manuscript was not written with the goal of emphasizing specific contributions to mathematics. It was more focused on the concept of automated conjecturing, and automatic generation of fundamental constant representations.

Therefore, we updated the manuscript with two significant additions, which both focus on demonstrating specific contributions to mathematics. The most important new result added in the manuscript was inspired by the referee's suggestion to pursue new efficient ways of computing fundamental constants. In fact, we have found a continued fraction representation of the Catalan constant that manages to beat the current record for efficient computation. This new result is explained in further detail below, alongside a comparison to the relevant literature. There is also a new section in the main text, a new section in the appendix, and a new figure - all explained below.

We believe that our manuscript is made stronger and better with the newly added results and are grateful to the referee for the comments and suggestions (especially for the direction to find more efficient ways of computing fundamental constants, which we did not consider before the review). By showing new specific contributions to the computation of the Catalan constant and its irrationality measure, we are now able to better showcase the cycle that we discuss in this work: from automatic conjecturing - providing human researchers with new leads for research - to substantially assisting in making research advancements.

We thank the referee for all the comments that helped us to improve our manuscript. A point-by-point response is provided below.

The referee wrote-

- The problem that is being studied is elementary, and is not representative of the real complexity of most mathematical problems. This is problematic, as the paper claims to be a step in the path of using algorithms to "unveil mathematical structure", and "play the role of intuition of great mathematicians in the past, providing leads to new mathematical research". More precisely, the search space in the problem of finding PCFs is very small, which leads to exhaustive approaches working quite well. This is in stark contrast with most mathematical problems, where interesting properties have to be found in much larger search spaces (and are also hard to formalize).

Reply-

The search space of PCFs is, actually, quite large in terms of the wide areas of mathematics it can cover and the many special functions it connects to. Most special functions that are ubiquitous in so many fields have PCF representations - these include all trigonometric functions, exponentials, Bessel functions, generalized hypergeometric functions, the Riemann zeta function, and many other important functions such as erf and log. Moreover, any infinite sum can be converted into a continued fraction, while the other direction is not true as many PCFs that cannot be written as infinite sums (we now provide examples of both cases in Appendix Section G, and discuss them in the new Section 5.3 of the main text).

We are not the first ones to notice the importance of PCFs in this respect. See, for example, the reference on PCFs suggested by Referee #2 [Bowman and Mc Laughlin, Polynomial continued fractions, Acta Arithmetica, 103, 329 (2002), our new reference 16]. Referee #2 also mentions that PCFs are a good fit "for the experimentation and why this is a perfectly adapted problem to their philosophy."

Thanks to the suggestions and comments of the referees, we revised the introduction to better motivate the exploration of PCFs. (we also revised Appendix Section C to point to open questions in this area related to our findings).

From a computational point of view, PCFs have a cardinality of Aleph0, and it is thus possible to enumerate over them, which makes them a good starting point for our approach. So in that sense, PCFs strike a good balance to serve as a proof-of-concept, being a countable space, yet wide enough to catch many of the interesting mathematical objects.

Having said the above, it is also worthy to note that although the results we bring are of PCFs, our algorithms are not restricted to a single structure. In fact, the algorithms we share online are designed to enable using any mathematical structure that can be represented in a parameterized and countable sequential manner. For instance, in the MITM-RF algorithm, we wrote the code implementing the LHS so that it can be chosen to be any parametric function (with a countable parameter space). Moreover, our MITM-RF algorithm now supports using any parametric function to generate the sequences of the continued fractions (this is a more complicated change to the code in comparison with the LHS, and enables searches of structures beyond PCFs). The same MITM-RF strategy also works with other countable mathematical structures that are not related to continued fractions at all.
We revised the section on the MITM-RF algorithm to explain these options in the algorithm.

The referee wrote-
- The paper could still make contributions with large impact while remaining specific to the considered problem. For example, finding more efficient ways of computing fundamental constants using PCFs, finding relations between constants, etc... The paper remains instead speculative on those problems (lots of conditionals that these will be possible in the future) but does not make any concrete contributions towards these more impactful problems.

Reply-
We thank the referee for his/her remarks and suggestions. This comment by the referee has helped us to significantly improve the outlook and prospects that our manuscript presents, and we are grateful for that.
Following this comment, we investigated how our results compare to existing methods of efficient computation of fundamental constants. As a result, we are glad to say that we found new PCF formulas of Catalan's constant that surpass the best records in two problems:

(1) The efficiency of formulas that compute the Catalan constant is defined by the number of digits found per term, normalized by the polynomial degree used in multiplications during the computation of each term (this normalization relates to the computation time of each such term). This definition is relevant for the efficiency of the state-of-the-art algorithms for the Catalan constant and other constants, as can be seen implemented in the y-cruncher project [new reference 72]. The record normalized rate was achieved in a formula by Pilehrood [new reference 70]. We found several PCFs that beat this record and present them in a new figure (Fig. 5a), and in Tables 7 and 8 in the new Appendix Section G.

(2) While the previous result is interesting from a computational point of view, a more fundamental mathematical definition of computation efficiency is related to the Diophantine approximation of a number. There, the goal is finding the rational number with the smallest denominator of any required precision. The relation between the size of the denominator and the acquired precision is given by the Liouville–Roth approximation exponent that is used as a measure of irrationality of a fundamental constant.
Specifically, for each constant X, the approximation exponent is defined as the largest constant μ for which there exists a converging sequence of rational numbers $\frac{p}{q}$ satisfying

$$|X - \frac{p}{q}| < q^{-\mu}.$$

The logic in this definition is to find the most “efficient” series of rational numbers $\frac{p}{q}$ that

converges to the constant. Here “efficient” means finding the rational number with the smallest possible denominator q for any required precision of computing the constant X. The smaller the denominator relative to the precision, the larger μ is going to be. The record value of μ for the Catalan constant was ~0.524, found in a 2003 paper [reference 11] and proven in 2016 [new reference 78]. The 2016 paper also presented a conjecture for a larger μ of ~0.554. **We found a PCF that surpasses this record, finding μ~0.567 (the comparison is presented in the new Fig. 5b).** This PCF is presented in Table 8 row 5 and sets the new record for the approximation exponent of the Catalan constant by an explicit series.

The above results are covered in greater detail in the updated manuscript in the new Section 5.3.

To summarize the changes about efficient calculations of the Catalan constant, we would like to say that this investigation, which resulted from the referee’s suggestion, clearly strengthens the impact of our manuscript. This comment led to improvements in several aspects of our manuscript, and we are grateful to the referee for it.
We find these results very exciting and hope the referee will also share our excitement.

Additional info about the underlying mathematical structure
Although it is not the focus of our work, it is interesting to explain the underlying structure that enabled finding the (complicated) PCFs that achieve the record results. Our MITM-RF algorithm found several new PCF results for the Catalan constant (Table 5) that we managed to generalize to an infinite family of PCFs. We found that the PCFs in this family are related through a lattice structure of path-invariant matrices [such path-invariant matrices were first used for other applications by Bill Gosper in unpublished work; generalizing the algebra of Wilf–Zeilberger pairs]. The lattice structure that we found from this generalization enabled us to reduce the space of enumeration in the MITM-RF algorithm and construct specialized PCFs that achieved faster convergence rates and the record result for the approximation exponent (as shown in Figs. 5a&b). Such results are exciting in the context of our work because they demonstrate how the computer-based approach can lead to complicated results that may be difficult to address without computer algorithms (see in Table 8 the polynomials of order >20 with coefficients of >30 digits). These complicated PCFs are precisely the ones that contribute to research on the computation of the Catalan constant and its Diophantine approximation.
While the underlying algebraic structure is not the focus of our work, if the referee recommends doing so, we can further elaborate on it in the manuscript.

Another specific contribution to mathematics
Apart from the new contribution that we mentioned above, we also completely revised Section 5.1 (Correspondence with the Mathematical Community) to discuss a new idea by the Zeilberger group that followed from our work and was recently put on arXiv [now our reference 20 -- the direct link is arxiv.org/abs/2004.00090 -- most recent update on arXiv from May 26]. The work by Zeilberger shows a beautiful example of the contribution our work already started having. This work also explicitly cites our manuscript as the inspiration - citing from their paper: “Inspired by the recent pioneering work, dubbed “The Ramanujan Machine” by Raayoni et al. [6], we (automatically) [rigorously] prove some of their conjectures regarding the exact values

of some specific infinite continued fractions, and generalize them to evaluate infinite families …"

To summarize the response to this comment, we copy part of the new Section 5.1 here:
"… A very recent example of this successful correspondence is the work done by Zeilberger's group [19], generalizing and proving part of the conjectures that appeared in the earlier arXiv version of our work [64] (see Appendix Section F.3). This new contribution that appeared as a consequence of our algorithms succeeded not only in proving part of our results, but also in creating infinite families of new conjectures through a new algorithm. Their method provides the proof as an inherent part of the discovery, thus the results are actually corollaries rather than conjectures. Therefore, [19] can be viewed as a successful case study of algorithms that combine ACG and ATP.

Such algorithms are important also because they complement automatic conjecturing methods, which discover formulas that still need a proof. Once proven by algorithms, the formulas can be automatically taken and used to enrich the modern "integral books" (such as Maple or Wolfram Mathematica) to assist further research efforts. Such an example from their paper is the elegant formula
...

A wide range of such identities is likely to come in handy in future approaches for different math problems, especially ones in adjacent fields (e.g., proving irrationality of zeta function values [16]). At the moment, the generalization of the results in [19] to an infinite class is done by human researchers. However, these generalizations are now being further automated and extended. This process provides an elegant example of the symbiosis between computer-generated mathematics and human-generated mathematics."

We are grateful to the referee for impelling us to sharpen the novelty of our work.

The referee wrote-
- On the algorithmic side, the contributions are not strong enough, with the two proposed approaches being a variant of exhaustive search (with discretization of the search space), and gradient descent. In the way they are used, these are also quite specific to the problem of finding PCFs, and it is unclear how these could generalize (e.g., to problems with larger search space).

Reply-
We thank the referee for this comment. While the algorithms themselves are based on well-known approaches, it is their application in a new field that is the most exciting aspect of our work. Specifically, our work is the first to use a meet-in-the-middle algorithm for an application in number theory to the best of our knowledge. It is also the first usage of any gradient descent-type algorithm in this field.
With that said, we understand the referee's point of view and made several improvements in the algorithms and their presentation in the revised manuscript:

(1) It is possible to reformulate and present the LHS and the sequences inside the continued fractions as general parametric functions (enumerating over parameters) instead of just the special case of polynomials. The code infrastructure supports it, with the aim of expanding our

search to other structures as well. Looking at the bigger picture, we do not believe it is sufficient to focus on one specific representation structure, such as PCFs, even though PCFs were initially a successful starting point for our approach. In fact, our algorithms (and code) are designed to allow using (or easily adding) any mathematical structure that can be represented in a parameterized and countable sequential manner, thus providing a flexible scheme.

(2) We improved the MITM-RF algorithm by utilizing more sophisticated data structures, which are more suited for this problem. Specifically, we now use Bloom filters instead of a regular hash-table, which drastically decreased the memory-space needed in the enumeration process. Bloom filters allow for a much lighter implementation, at the cost of some false-positives. Since false-positives are abundant in this stage anyway (due to limited precision), it does not affect the algorithm's performance. The manuscript now describes these advancements as well in the section on MITM-RF.

(3) In addition to the two points above, we have recently developed a more sophisticated conjecturing algorithm, which utilizes a variation of the Berlekamp-Massey algorithm to identify significant patterns in expressions calculated directly from fundamental constants. Our Berleklamp-Massey-based approach is used to identify an underlying mathematical structure in the representation of a target fundamental constant. This algorithm finds a generating linear-feedback-shift-register over Galois fields. Each shift register can be translated to recurrent formulas for the parameters (e.g., covering all PCFs and cases of multiple alternating polynomials). While this last development is currently not part of the manuscript, we can consider including it if the referee sees this as a crucial aspect.

The referee wrote-

- It is unclear what the authors exactly mean by "new" conjecture. In fact, most conjectures in Table 4 are labeled as "new and proven" (e.g., 4 / (pi - 2)), while the proof in Appendix F.2 is often a mere specialization of Gauss's continued fraction. This should not be considered as a new result.

Reply-

We appreciate the referee's input on this matter. We would first like to emphasize that the more significant results, and indeed the ones that drove us to submit to *Nature*, are found in Table 5. While the results in Table 4 were the first to be discovered by our algorithms, they are simpler to prove, and their mathematical context is less exciting from the point of view of pure mathematics. The results in Table 5 are more interesting than the results in Table 4, and for example, part of them led to the new findings on the Catalan constant. Unlike the results in Table 3 and 4 that received suggested proofs within several weeks to a few months (from the original appearance of our work on arXiv), the results in Table 5 have not been proven yet.

To explain why the results in Table 5 are interesting and can be considered as new results we point to a relatively famous historical precedent: The discovery of an exponentially converging series for $\zeta(3)$ led Apéry to discover a continued fraction representation that allowed him to prove its irrationality. $\zeta(3)$ is even named the Apéry constant following this important proof. This proof created a method that was later on utilized in other problems. The results we found for the Catalan constant now go in the same direction.

A similar motivation for discovering new continued fractions is also brought in other papers. e.g., *"Such identities are intrinsically fascinating, but continued fraction expansions have found wide applications in number-theoretic irrationality proofs. There is always hope that the correct continued fraction will provide a Diophantine approximation sufficiently nice to prove the irrationality of a famous constant, a la Roger Apery's proof that $\zeta(3)$ is irrational".*

[reference 19 in the revised manuscript].

Regarding the labeling of results as "known" / "new and proven" / "new and unproven", we now better explain their definitions in multiple places in the text and in Appendix Section A that collects all the PCF results and sorts them. The clarified definitions given in the revised section are also copied here:

"… each individual generated result, where it can be **known**: i.e., we have found this result (or an equivalent form of it) in the literature, and therefore it serves as a proof-of-concept for the Ramanujan Machine but is not considered new. A result could also be **new and unproven**: i.e., a new conjecture found by our algorithms that we have not found in the literature. We consider it a new conjecture until proven or until an equivalent form, unknown to us at first, is found. Finally, a result can be **proven**: i.e., a result of our algorithms that was proven after the first appearance of our work on *arXiv*. We note that some of these results are easy to prove, meaning that they can be derived from known results in a relatively straightforward manner (e.g many π results can be derived by specific specializations of Gauss' continued fraction). We provide several such examples in the later sections of the Appendix."

To summarize the response on this comment, the results in Table 4 were called "new" in the original version of the manuscript because their representations as PCFs were not shown before in the literature. However, once discovered, these PCFs were found to be relatively easy to prove as specializations of Gauss' continued fraction (each exact specialization must be found explicitly using identities of hypergeometric functions). In contrast, this situation of being "new yet easy to prove" **is not the case** with results found in Table 5 and elsewhere in our manuscript. The results brought throughout the main text concerning $\zeta(3)$, π^2, and Catalan's G are (to the best of our knowledge and research) not a specialization of any previously formulated identity. Therefore, we call these results "new and unproven".

The referee wrote-

- The presentation of the algorithmic part can be significantly improved; for example:
- In Section 4, it is mentioned that "we empirically observed that all minima are global, and their errors are zero. Therefore any GD process will result in a solution with L = 0". This needs more justification.

Reply-

We thank the referee for pointing this out. We improved the explanation in the caption and in the section on the GD process. We copy part of the improved section here:
(also see the response to Referee #2 on this)

"... Solving this optimization problem with GD appears implausible since we are only satisfied with exact zero-error integer solutions. Close to zero (yet not zero) solutions are usually meaningless as mathematical conjectures.

Nevertheless, we found a significant feature of the loss landscape of the described problem that helped us develop a slightly modified GD, which we name 'Descent&Repel' (Fig. 4). Examples of the results appear in Table 1. Without the restriction of being integers, the zero-error minima are not 0-dimensional points but rather (d-1)-dimensional manifolds with d being the number of optimization variables. Specifically, in the case plotted in Fig. 4, there are d=2 optimization variables, and therefore a 1-dimensional manifold of minima - appear as bright curves in the maps. This dimensionality of the minima is expected given the definition of the error that only poses a single constraint. We empirically validated that almost all minima have zero error (i.e., resulting in exact equality). Therefore, the GD process is expected to result in a solution with L = 0. The high dimension of the manifold of minima motivates our approach of adding the repel step to the algorithm since most minima have a neighborhood that contains additional minima."

The rest of the section provides examples and describes the algorithm in three stages (the descriptions there are also revised substantially). Moreover, we have made a similar improvement in the presentation of the MITM-GD algorithm (in Section 3).

The referee wrote-

- It is mentioned that the repel mechanism is used to increase the search space, and thus "the probability of finding a match in space". In light of the above comment, do all initial conditions lead to a solution?

Reply-

No. Most initial conditions do not lead to a solution, as they get "stuck" in an area of space where the integer parameters do not reach the global minimum of an exact zero loss (it is possible to identify when the process is stuck and when it converges to a solution). This scenario is very common in GD algorithms in the field of machine learning, where each initial condition has a high chance of reaching a local minimum instead of a global minimum. For our goal of finding a formula for a fundamental constant, a non-zero loss minimum is irrelevant, which makes our problem crucially different from other problems in machine learning that use GD-type algorithms.

The referee wrote-

- The paragraph after Equation 7 is unclear, and its logic should be re-visited: Unclear why the dimensionality of the manifold is relevant to the discussion, unclear whether GD refers to vanilla gradient descent or to the version with discretization — especially since the last sentence in the paragraph infers a property on the problem with integer constraints.

Reply-

We thank the referee for this comment. The paragraph after Eq. 7 has been revised in order to be more informative (also see the text copied above). The algorithm has discretization in its third stage, where we combine the regular gradient descent with a loss function that optimizes for integer solutions. The first stage of the algorithm performs a regular gradient descent without discretization.

Importantly, the dimensionality of the minima arises directly from the definition of the error. The error we use is the square of distance between the LHS and RHS, which only poses a single constraint on the d parameters. Therefore, the dimension of the manifold of possible solutions is typically d-1. It is the additional constraint of searching for integers that causes solutions to be extremely rare despite the high dimension of the manifold.

Following the comments of both Referees #1 and #2 on the gradient descent algorithm, the details of the algorithm are now explained better (see the revised Section 4).

The referee wrote-

- What are x and y axes in Figure 4.

Reply-

We added a sentence in the caption defining the axes. The axes are parameters of the GD and can be chosen in many different ways (examples defined in the captions of Fig. 4 and 6).

The referee wrote-

- Motivation of Descent and repel. It is mentioned that MITM-RF is not "scalable" — mention more explicitly to which parameters it is desired to scale up.

Reply-

The reason we consider MITM-RF as more limited than the GD approach is the computational complexity.
If we wish to expand to much higher order polynomials, MITM-RF is impractical due to the exponential complexity growth with the degrees of the polynomials. In contrast, GD-type searches are being used regularly in the machine learning community to optimize parameters over such large spaces. Therefore, it is possible that the GD-type search will become of greater value when the search space grows.

To conclude our response to the referee

We thank the referee for helping us improve the manuscript. We are especially grateful for the suggestion to explore the efficiency of the computation of fundamental constants using our results - this led us to a new contribution related to the Diophantine approximation of the Catalan constant.
We hope that the referee now shares our excitement about this study and its promise.

Referee #2 *(Remarks to the Author)**:***

The referee wrote-

Dear Editors,

I read the submission

"The Ramanujan Machine: Automatically Generated Conjectures on Fundamental Constants"
By: Raayoni et al
MS: 2020-04-07825

with great interest.

The paper attempts to generate new identities, specifically polynomial continued fractions, by generating large sets of potential candidates and then using a gradient-descent optimization to zoom in "correct" ones by checking to hundreds of digits of precision. Consequently, many correct expressions are found, and impressively, new conjectures have been raised, some of which have been proven since the appearance of the paper on ArXiv and several more still open.

As stated in the appendix and the conclusions, this exercise has prompted an interactive website (in the spirit of PolyMath, GIMP, etc,) www.RamanujanMachine.com that has inspired the mathematics community to prove some of the new conjectures.

I am very sympathetic to this experimental approach to mathematics using the best resources of today: computer algebra and machine-learning.
I therefore recommend the article for publication in Nature subject to the following revisions:

Reply-
We are grateful to the referee for the detailed review and constructive comments that we answer point-by-point below.

The referee wrote-
- It would be good to include a discussion on why PCFs are chosen (out of the myriad of possible mathematical structures) for the experimentation and why this is a perfectly adapted problem to their philosophy. An excellent summary of the subject is

Bowman and J. Mc Laughlin, Polynomial continued fractions, Acta Arithmetica, 103(4) (2002), 329–342

and should be cited.

Reply-
We thank the referee for this comment and now cite this reference. We also added two new paragraphs in the introduction that add motivation for the importance of PCFs, and also expand the discussion on certain applications of PCFs.

The new introduction paragraphs are also copied here

"One reason we choose to focus on PCFs in this work is their ability to capture a sweet-spot between simplicity and wide implications. Their structure is quite accessible for computer-based exploration using operations on large integers, making them a good object for a test-ground while examining an automated conjecturing approach. At the same time, PCFs turn out to naturally appear in many problems in mathematics, to connect to many special functions and generalize all infinite sums. PCFs also enable isolating unique aspects of importance to fundamental constants such as testing irrationality or normality using efficient computation methods to high precision (see Appendix sections D and G).

A possible explanation for why PCFs are so abundant in areas of mathematics is that they constitute an important special case of a general mathematical object: linear recurrences with polynomial coefficients, which for recurrences of depth 2 correspond to PCFs, and appear in this way in many mathematical problems (PCFs with alternating polynomials - as shown below - further correspond to recursion depths >2). The solutions to such recurrences are usually very complex and include special functions (i.e., hypergeometric functions, incomplete gamma function, etc.). For this reason, finding new PCF identities is valuable for different mathematical objects, especially so when incorporated as part of symbolic calculations programs (such as Maple and Wolfram Mathematica). More on PCFs in the Appendix."

The referee wrote-
- In Sec 2 for Related Works, it might be worth citing explorations in supervised and

unsupervised machine-learning (very much in the spirit of this paper in attempting to find structure and generate new conjectures, and not in the style of ATP) have been applied to study of
the physical laws
"Discovering Physical Concepts with Neural Networks",
Raban Iten et al, PHYSICAL REVIEW LETTERS 124, 010508 (2020);

"Deep-Learning the Landscape", by Y.-H. He
https://arxiv.org/abs/1706.02714, Phys.Lett.B 774 (2017) 564-568

and number theory

"Machine Learning meets Number Theory: The Data Science of Birch- Swinnerton-Dyer", L. Alessandretti, A. Baronchelli, Y.-H. He
https://arxiv.org/abs/1911.02008

Reply-
This is a very good suggestion. The revised manuscript now has an additional paragraph in the section on related works, citing these papers and a few other recent advances, especially on applying AI and modern machine learning techniques in physics.

The referee wrote-
- Eq (5): Can the author give a brief account of the type (or full list) of the polynomials \alpha, \beta, \gamma, \delta as well as the function f_i (e.g. what does i index? how many such function are tried?) that was used in the search? How exhaustive was it? Also, please present an idea of the running time and on what machine.

Reply-
We added more information on the space of parameters used in each algorithm and elaborated on the running times used for the runs we performed. These details are now provided in the respective sections in the main text and appendices.
We also added this information in the git code that we share online, so other users can estimate running time. (Specifically, there is a test code that provides many of our results, together with detailed comments that help new users to run the code for the first time).

We are still far from using the full potential of the current algorithms because more computational power can be very valuable to extend the enumeration (so far, we ran solely on simple PCs, without exploiting acceleration, distribution, nor cloud computing services). The additional information on the enumerated space and running time that we now provide in the revised manuscript can help to emphasize the potential for improvements in future work.

To give an idea of actual running times, we copy the new paragraph in Appendix Section A:

"Creation of the LHS hash table takes ~10 minutes for the typical search spaces we used of 10^8 possible terms (executed on a regular computer). The RHS enumeration step is slower, taking $\sim 10/C$ minutes for a search space of just 10^7formulas (executed on a regular computer). It is worth noting that for the PCFs we found, the required RHS search spaces were significantly larger than the required LHS search spaces, and so the running time of the RHS was the main compute bottleneck. Nevertheless, this trend may be a coincidence related to the fundamental constants or the type of LHS functions we chose. Future runs may benefit from a larger memory for larger LHS hash-tables."

The running times specified here were calculated on a computer that has only 2 cores and 8

GB of RAM - meaning no special computational resources were used. Although our resources were quite limited, we were still able to discover novel results, which shows our approach's potential.
Let us explain how these running times depend on the search space that we used: The search space size is defined as the number of possible formulas. For example, we can initiate a naive search for α as a degree 1 polynomial and β as a degree 2 polynomial, with all coefficients being between 0 and 25. This space will have ~10^7 RHS PCFs and will take ~10 minutes (this search space is good for finding pi related PCFs). Our code is quite modular, so one can facilitate understanding of the structure of expected results to more delicately fine-tune the search space to include high degree polynomials while keeping the number of terms small enough. Our longest search took ~4 days. More detailed examples now appear in our open-source github page.

The referee wrote-
- Could the authors comment on the success of generating PCFs for algebraic (e.g. Golden ratio) versus transcendental constants (e.g. Pi); do they differ? This would be of substantial interest to the number theory community.

Reply-
This is a very interesting question. We found that certain constants have many more PCFs than others and therefore our approach appears more successful for them. Examples of such constants include e and \phi.

However, we note that the absolute measure of success is not rigorously connected to the absolute number of PCFs (as the spaces are infinite). Instead, we only make such comparisons when testing different fundamental constants under the same fixed space of PCF parameters. It seems that constants like e and \phi generally have more lower-order PCFs than constants like $\zeta(3)$ and Catalan. Of course, this does not mean that there are “more” PCFs for these constants.

It is interesting to find the lowest order polynomials that are possible for each fundamental constant. We now present these questions and related questions in the revised manuscript. For example, we find that pi has lower-order PCFs than pi^2, and that this continues with larger powers, and similarly with larger arguments of the Reimann Zeta function. We updated Appendix Section C with the current values found for each constant, hoping that this will contribute to future studies or stir up new ideas in this direction.

One can see these polynomial orders as a measure of complexity that generalizes the special case of simple continued fractions. For simple continued fractions, it is known that the set of periodic sequences a_n is a subgroup of the algebraic numbers. PCFs can represent transcendental numbers as well. There may be a way to generalize this concept to define growing levels of complexity of Aleph0 numbers.

The referee wrote-
- The Descent&Repel algorithm needs to be clarified (very much line with the comment above)
* Move the comment about the variables to be optimized coming from \alpha, \beta, \gamma, \delta, as coefficients in these polynomials to just below Eq (7), before saying there are d of these. What is the typical number of d? i.e., what are the degrees of the polynomials considered? Presumably, Fig 4, where d = 2, is only a schematic illustration.

Reply-

We thank the referee for this important comment and the careful reading of our manuscript. We apologize for the typos and missing information in the description of the GD algorithm. We revised this section carefully, including the figure and its caption (and the corresponding appendix section). The missing information is now provided in the manuscript.
Our proof-of-concept runs of the Descent&Repel algorithm covered a very small subspace of the possible parameters. Most of the runs we performed were in 2 dimensions, but we also simulated larger spaces as a test of the algorithm.
The fact that already in d=2 we found results motivates trying a much larger space.

The referee wrote-

* Why does repulsion help? i.e., why shouldn't one consider integer solutions to the coefficients which are close together in solution space?

Reply-

In early attempts to use gradient descent without repulsion, we noticed that many points converge to the same minimum, often independent of the initial conditions used. Since we hoped the methodology would provide more diverse results that span a wider range of possible solutions, we experimented with different approaches to extend and improve it. After testing different methodologies, we identified the repulsion that performed well in numerical tests. Repulsion helps by causing the points to more effectively cover the space of parameters. This is, of course, a heuristic, like many algorithms in AI these days - we tune the repulsion strength so it creates a typical point-to-point distance that covers the integers in the most effective way.

Regarding the specific strength used for the repulsion, if the repulsion is too weak, then the distance between points will be too small and result in many points that converge to the same minimum. If the repulsion is too strong, then the distance between points will be much larger than the distance between adjacent integers, which will result in many potential integer solutions being missed.

We added this explanation to the section about the algorithm.

The referee wrote-

* Write the integer-round loss function along side with {\cal L}
and explain how the 2 are used alternatingly

Reply-

We thank the referee for spotting this point of confusion and of course we corrected it, writing the two types of loss functions that we use as part of the algorithm. (See the description of stages 1-3 of the algorithm on page 8, where we explain how the two loss functions are alternated in stage 3).

The referee wrote-

* in step 1 of the algorithm, what is x_t? presumably \mu is step size and x_t denote the d variables and L is script-L?

Reply-

We thank the referee for this comment. In the revised manuscript, we defined the values of x and the rest of the parameters in stage 1 of the algorithm:

“We perform a standard GD separately for each point x_i, which is a d-dimensional vector. The loss function L is defined in Eq. 7, and thus, for each point x_i we define its next iteration t+1 as

i

$x_i^{(t+1)} = x_i^{(t)} - \mu \cdot \nabla L|_{x^{(t)}}$, where μ is some small enough step size."

Additional revisions are implemented in the other parts of the algorithm (e.g., separating the definitions of the two loss functions).

The referee wrote-

* In Fig 4, the colour-legend needs to be explained.
Why is there no red?

Reply-

The two colorbars were used for two different parts of the figure. The left colorbar was for the loss landscape, while the right colorbar was for the dots that move on the landscape. We now explain it more carefully in the caption. We also modified the right colorbar to be made from colored dots, in order to make the separation clear.

The referee wrote-

Which Log-error is used? the integer-round one or the one in Eq 7. What are (x,y) ?

Reply-

There are two errors being used in the two stages of the algorithm. We apologize for the possible confusion in the previous version and clarified this issue in the improved manuscript.

The referee wrote-

- A scope of the constants used in the search algorithm (LHS of Eq 5) should be listed explicitly, i.e., which constants (pi, e, Catalan, etc) have been searched, and which have produced good hits. This is particularly relevant to Table 2. Are there any in table which did get a hit and thus a conjecture?

Reply-

We now detail the full scope of constants that were searched for in section 5.4. We also refer to that section when discussing the successful runs, which produced the results in chapter 3. The constants for which we conducted searches (in addition to the ones which produced results) are Riemann's $\zeta(5)$, Euler-Mascheroni γ, and Khinchin's K_0. Additionally, we ran some searches for the Feigenbaum's δ.

The referee wrote-

- Small typographical errors such as:
* All equations and footnotes (e.g. footnote 5) need to be punctuated;
* The quotation marks are all wrongly type-set; make sure to use
` instead of ' for the beginning of the quote in LaTeX.
* Top left box of Figure 2, is that supposed to be PCF?

Reply-

All these errors are now fixed throughout the main text and appendices.
We thank the referee for the detailed review and careful reading that helped to improve our work substantially.

To conclude our response to the referee

We are grateful for the strong positive review and for the helpful comments.

Referee #3 *(Remarks to the Author):*

The referee wrote-
A. This is a seminal paper describing two very efficient algorithms to generate intriguing conjectures with far-reaching potential applications.

B. While the LLL and PSQL were used in an ad hoc way before, the systematic and unified approach described here is very novel and significant.

C. Very valid and the authors share all their ample data and output in their web-site for the benefit of the mathematical world.

Reply-
We thank the referee for emphasizing the novelty of our work and its open-source approach that contributes to the wider community - this is especially important for us.

The referee wrote-
D. Since this is mathematics, the standards are much higher than in the physical scientist, and all their conjectures are virtually certain, even those still awaiting a formal proof.

Reply-
We are especially glad for this comment.
We think that it is a very important point of difference between our efforts with fundamental constants and the vast interest in recent years in automated conjecturing in physics. Indeed, the amount of data in a fundamental constant makes the conjectures virtually certain because of the ability to check each conjecture to arbitrary precision. This precision and the availability of data is what makes the situation very different from the attempts to use algorithms to find conjectures in physics (which has recently become a hot topic in the machine learning community). In contrast with our case, in physics, the availability of data is often the bottleneck, creating very different requirements for the algorithms.
In the revised manuscript, we now also cite key papers on AI in physics to help make the connection and show this important difference.

The referee wrote-
E. Perfect

F. The references should be carefully copy-edited.
e.g. (put please check everything)

Ref. 18: the authors' names are reversed and they should include first names (or at least first initial)

Ref. 20: A=b -> A=B

Ref. 45: ditto (and also the title is

wrong) Ref. 60: ditto

Reply-
We really appreciate the thorough reading and catching these typos. We corrected all of them.

The referee wrote-
G. See above, otherwise perfect

H. Very well-written, lucid, and engaging. A true Tour-de-forice

To conclude our response to the referee
We thank the referee for the strong positive review and for helping to improve our manuscript.

Reviewer Reports on the First Revision:

Referee #2 (Remarks to the Author):

I had a look of the revised manuscript, focusing on how the authors address my comments in particular and I am happy with the response.

Referee #3 (Remarks to the Author):

All the objections of the referees were very nicely addressed, and now the paper is yet stronger.

This is a very significant contribution to the budding field of experimental mathematics, that used be considered an oxymoron, but the future of mathematics is in that direction.

I strongly recommend acceptance.

Referee #4 (Remarks to the Author):

This manuscript describes what the authors term "The Ramanujan Machine", which is a semi-automatic algorithm and computer implementation that permits one to computationally generate conjectural identities of the general form $C = PCF(a,b)$, where C is a specific mathematical constant and $PCF(a,b)$ is a polynomial continued fraction. Here a "polynomial continued fraction" is a continued fraction whose numerators $a(n)$ and denominators $b(n)$ are given by the integer-coefficient polynomial functions $\alpha(x)$ and $\beta(x)$, respectively, for $x = 0, 1, 2, ...$.

The authors document how their algorithm and software implementation have not only recovered a number of known identities of this form, but have generated numerous previously unknown mathematical identity conjectures, some but not all of which have been subsequently proven by conventional formal methods.

The work described in the manuscript is original and mathematically significant. The manuscript itself is quite well-organized, with its main body targeted to a more general readership, and with appendices containing more technical details and background. The usage of English is very good throughout.

In general, the manuscript is well deserving of publication. But in the spirit of further improvements to the manuscript, this reviewer offers the following comments for the "Related Work" section:

1. The general approach of using algorithmic and computational tools to explore the mathematical universe and to discover conjectures worthy of further examination and formal proof is most often known in the mathematical literature as "experimental mathematics" (the authors cite several references where this terminology is used). Thus brief mention of this terminology is in order here.

2. The authors briefly mention the PSLQ algorithm and its usage to find new results on pi. This certainly is an appropriate example of "automated conjecture generation" (to use the authors' terminology). But it is not quite accurate (or at least not complete) to say that PSLQ led to a scheme that finds specific digits of pi. What should be said here is that the PSLQ algorithm was used to numerically discover a new formula for pi, which was then proved using a calculus argument, and then this new formula, after some further manipulation and analysis, gave rise to an algorithm that permits one to directly calculate strings of binary or base-16 digits of pi starting at a given position, without needing to know any of the preceding digits. By the way, it would be appropriate here to present the formula for pi discovered by PSLQ, which is:

$$\pi = \sum_{n=0}^\infty \frac{1}{16^n} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6}\right)$$

3. The authors should mention the work of Doron Zeilberger, Herbert Wilf and others, who developed a tool that proves a large class of combinatorial identities, once the identity is numerically (or otherwise) discovered. A good original reference is:

Herbert S. Wilf and Doron Zeilberger, "Rational functions certify combinatorial identities," Journal of the American Mathematical Society, vol. 3, no. 1 (Jan. 1990), 147-158.

4. The authors might also mention the work of Stephen Wolfram, who has championed experimental-computational methods to investigate the properties of cellular automata. See, for instance,

Stephen Wolfram, A New Kind of Science, Wolfram Media, Champaign, IL, 2002.