A simple information theoretical proof of the Fueter-Pólya Conjecture

Pieter W. Adriaans

ILLC, FNWI-IVI, SNE
University of Amsterdam,
Science Park 107
1098 XG Amsterdam,
The Netherlands.

Abstract

We present a simple information theoretical proof of the Fueter-Pólya Conjecture: there is no polynomial pairing function that defines a bijection between the set of natural numbers $\mathbb{N}$ and its product set $\mathbb{N}^2$ of degree higher than 2. We show that the assumption that such a function exists allows us to construct a set of natural numbers that is both compressible and dense. This contradicts a central result of complexity theory that states that the density of the set of compressible numbers is zero in the limit.

Keywords: Fueter Pólya Conjecture, Kolmogorov complexity, computational complexity, data structures, theory of computation.

1. Introduction and sketch of the proof

The set of natural numbers $\mathbb{N}$ can be mapped to its product set by the two so-called Cantor pairing functions $\pi : \mathbb{N}^2 \to \mathbb{N}$ that defines a two-way polynomial time computable bijection:

$$\pi(x, y) := 1/2(x + y)(x + y + 1) + y \tag{1}$$

The Fueter - Pólya theorem (Fueter and Pólya (1923)) states that the Cantor pairing function and its symmetric counterpart $\pi'(x, y) = \pi(y, x)$ are the only possible quadratic pairing functions. The original proof by Fueter

Email address: `P.W.Adriaans@uva.nl` (Pieter W. Adriaans)

and Pólya is complex, but a simpler version was published in Vsemirnov (2002) (cf. Nathanson (2016)). The Fueter - Pólya conjecture states that there are no other polynomial functions that define such a bijection. In this paper we present a proof of this conjecture based on the incompressibility of the set of natural numbers.

1.1. Sketch of the proof

Assume there exists a polynomial $\phi : \mathbb{N}^2 \rightarrow \mathbb{N}$ of degree $k > 2$ that describes a bijection. This allows us to construct a set $A \subset \mathbb{N}$ with the following characteristics:

1. The descriptive complexity of all elements $z = \phi(x, y)$ of A is bounded by $K(z) < \log x + \log y + O(1)$. We need $\log x$ and $\log y$ space to describe the input of an algorithm ϕ of constant size that computes $\phi(x, y)$.
2. There is a constant c such that for all elements in A we have $\phi(x, y) > cx^a y^b$ where $a + b = k$, i.e. ϕ is a function of degree k in a non-trivial way.
3. The randomness deficiency of elements of A is not bounded by a constant in the limit: $\delta(z) = \log z - K(z) = \log \phi(x, y) - K(z) \geq a \log x + b \log y - \log x - \log y - O(1)$. This is a consequence of the first two observations.
4. The density of A in the domain of ϕ is larger than 0, the density in its range is 0. At the same time ϕ is supposed to be a bijection, that conserves the densities of the underlying sets. From this contradiction various other inconsistencies can be constructed: ϕ does not exist.

2. Notation and definitions

We follow the standard reference for Kolmogorov complexity Li and Vitányi (2008). The set $\{0, 1\}^*$ contains all finite binary strings. $\mathbb{N}$ denotes the natural numbers and we identify $\mathbb{N}$ and $\{0, 1\}^*$ according to the correspondence

$$(0, \varepsilon), (1, 0), (2, 1), (3, 00), (4, 01), \ldots$$

Here ε denotes the *empty word*. The amount of information in a number is specified as $I(n) = \log_2 n$. The *length* $l(s)$ of s is the number of bits in the binary string s. Note that every natural number n corresponds to a string s such that $l(s) = \lceil \log_2(n + 1) \rceil$. If x is a string than $\overline{x}^U$ is the self delimiting code for this string in the format of the universal Turing machine U. When we select a reference prefix-free universal Turing machine U we can define the prefix-free Kolmogorov complexity $K(x)$ of an element $x \in \{0,1\}^*$ the length $l(p)$ of the smallest prefix-free program p that produces x on U:

Definition 1. $K_U(x|y) = \min_i\{l(\bar{i}) : U(\bar{i}y) = x\}$ *The actual Kolmogorov complexity of a string is defined as the one-part code:* $K(x) = K(x|\varepsilon)$

For two universal Turing machines U_i and U_j, satisfying the invariance theorem, the complexities assigned to a string x will never differ more than a constant: $|K_{U_i}(x) - K_{U_j}(x)| \leq c_{U_iU_j}$. By prefixing a print program to any string x one can prove that $\forall(x)K(x) \leq l(x) + O(1)$.

Definition 2. *The* randomness deficiency *of a string* x *is* $\delta(x) = l(x) - K(x)$. *A string* s *is* typical *if* $\delta(x) \leq \log l(x)$. *A string is* compressible *if it is not typical.*

Let A be a subset of the set of natural numbers $\mathbb{N}$. For any $n \in \mathbb{N}$ put $A(n) = \{1, 2, \ldots, n\} \cap A$. The *index function* of A is $i_A(j) = n$, where $n = a_j$ the j-th element of A. The *compression function* of A is $c_A(n) = |A(n)|$. The density of a set is defined if in the limit the distance between the index function and the compression function does not fluctuate:

Definition 3. *Let* A *be a subset of the set of natural numbers* $\mathbb{N}$ *with* $c_A(n)$ *as compression function. The lower asymptotic density* $\underline{d}(A)$ *of* $A(n)$ *in* n *is defined as:*

$$\underline{d}(A) = \liminf_{n\to\infty} \frac{c_A(n)}{n} \tag{2}$$

We call a set dense *if* $\underline{d}(A) > 0$. *The upper asymptotic density* $\overline{d}(A)$ *of* $A(n)$ *in* n *is defined as:*

$$\overline{d}(A) = \limsup_{n\to\infty} \frac{c_A(n)}{n} \tag{3}$$

The natural density $d(A)$ *of* $A(n)$ *in* n *is defined when both the upper and the lower density exist as:*

$$d(A) = \lim_{n\to\infty} \frac{c_A(n)}{n} \tag{4}$$

With these definitions we can, for any subset A of any countably infinite set $\mathcal{A}$, estimate the density based on the density of the index set of A.

Lemma 1. *Almost all strings are typical: the density of the set of compressible strings in the limit is* 0.

Proof: The set of finite binary strings is countable. The number of binary strings of length k or less is $\Sigma_{i=0}^{k} 2^i = 2^{k+1} - 1$ so the number of strings of length $k - d$, where d is a constant is at most $2^{k-d+1} - 1$. A string s is compressible if $\delta(s) \leq c \log l(s)$, i.e. $K(s) > l(s) - c \log l(s)$. The density of the number of strings that could function as a program to compress a string s in the limit is $\lim_{k\to\infty}(2^{k-c(\log k)+1} - 1)/2^k = 0$. Since the upper density is zero, the lower- and natural density are defined and both zero. □

By the correspondence between binary strings and numbers these results also hold for natural numbers. The randomness deficiency of a number is $\delta(x) = \log_2 x - K(x)$. Most numbers are typical, the density of the set of compressible numbers is 0 in the limit.

3. Proof of central theorem

The general structure of the proof is reductio ad absurdum. We assume that there is a polynomial in x and y of degree $k > 2$ that defines a bijection between $\mathbb{N}^2$ and $\mathbb{N}$. We show that the descriptive complexity of $\phi(x, y)$ has an upperbound, while the size of $\phi(x, y)$ has a lowerbound and these values diverge on dense subsets of $\mathbb{N}$. From this observation we can construct a bijection on $\mathbb{N}$ that contradicts lemma 1.

3.1. Upperbound for $\phi(x, y)$

Any function that defines a computable bijection between $\mathbb{N}$ and $\mathbb{N}^2$ also, in terms of Kolmogorov complexity, specifies a way to split any natural number in to a pair of two smaller numbers with exactly the same amount of information.

Lemma 2. *Suppose $\phi : \mathbb{N}^2 \to \mathbb{N}$ is an effectively computable bijection, then:*

$$\begin{aligned}&(\forall z \in \mathbb{N})(\exists!(x, y) \in \mathbb{N}^2\\ &\qquad \phi(x, y) = z \wedge K(z) \leq \log x + \log y + O(1)\end{aligned} \tag{5}$$

Proof: Since ϕ is a bijection the existence of a unique pair (x, y) for each z is granted. We can produce z by running the code for ϕ on a universal machine U with (x, y) as input. Let p be the prefix-free code for ϕ with constant length $O(1)$. Without loss of generality we assume that the code for x and y is provided on separate tapes, without any additional bits to separate them. The space for the code of the numbers x and y is given by $\log x$ and $\log y$ respectively. So there is a program q for U of length $l(q) = \log x + \log y + O(1)$ that produces z. This gives an upper bound for $K(z)$. $\square$

The lemmas 2 and 1 define an asymptotically rigid *information mold* for any bijection $\phi : \mathbb{N}^2 \rightarrow \mathbb{N}$. In the limit almost all numbers typical, i.e. $\log \phi(x, y) \approx \log x + \log y$ and $\log x \approx \log y$. This gives a rigid constraint for any bijection, which is the basic intuition of the proof.

3.2. Lowerbound for $\phi(x, y)$

Suppose the polynomial $\phi : \mathbb{N}^2 \rightarrow \mathbb{N}$ with degree k is a bijection. We have to prove that ϕ has a lowerbound on a dense subset of $\mathbb{N}^2$. We first prove that subsets defined by a simple linear inequality are dense provided that they are counted according to the Cantor function.

Lemma 3. *For any $h > 0 \in \mathbb{N}$ the set $A = \{(x, y) \in \mathbb{N}^2 | \ x < hy\}$ has density $d(A) = \frac{h}{1+h}$ in the set $\mathbb{N}^2$, provided that $\mathbb{N}^2$ is enumerated by $\pi(x, y)$.*

Proof: We enumerate $\mathbb{N}^2$ according to $\pi(x, y)$. The cardinality of the set $\{(x, y) \in \mathbb{N}^2 | \ x + y \leq k\}$ counted at $\pi(0, k)$ is $1/2(k)(k + 1)$. The boundary value of x on the counter diagonal is given by $k = x + y = x + hx$ which gives in the limit $\pi(0, k) = 1/2(x + hx)^2$. The cardinality of the subset $\{(x, y) \in \mathbb{N}^2 | \ (x + y \leq k) \wedge (x < hy)\}$ counted at $\pi(0, k)$ in the limit is $1/2hx(x + hx)$, wich gives for the density:

$$d(A) = \lim_{x \to \infty} \frac{1/2hx(x + hx)}{1/2(x + hx)^2} = \frac{h}{1 + h}$$

The density of the set $\{(x, y) \in \mathbb{N}^2 | \ (x + y \leq k) \wedge (x > hy)\}$ counted at $\pi(0, k)$ is $\frac{1}{1+h}$ in the limit. The density of the set $x = y$ is 0 in the limit. $\square$

We then prove that ϕ has a lower bound on such a subset. We define ϕ^- and ϕ^+ as the sets of negative and postive terms in ϕ respectively, ϕ^i is the set of terms of degree i in ϕ, with $\phi^{+i} \cup \phi^{-i} = \phi^i$. In order to prove our main result we only have to prove a weak proposition: ϕ has a lower bound $hx^a y^b$

of degree $a+b=k$ on an arbitrary small but dense infinite subset of $\mathbb{N}^2$. We say that $c_i x^a y^b$ dominates a term $c_I x^c y^d$, both of degree k in variable x if:

Definition 4. $c_i x^a y^b \succ_x c_I x^c y^d \rightarrow a+b=c+d=k \wedge a>c$

Note that ϕ^k will have two dominating terms, one in x and one in y. We have to show that a term of degree k that dominates a set of terms T with respect to a variable, in the limit dominates the sum of all variables in T within an arbitrary small neighbourhood $\epsilon>0$. We first prove the elementary case:

Lemma 4. *If $c_i x^a y^b \succ_x c_j x^c y^d$ then, there is an $\epsilon>0$ such that $\underline{d}(A)>0$, where $A=\{(x,y)\in\mathbb{N}^2| \;\; |c_i x^a y^b|-|c_j x^c y^d|>(c_i-\epsilon)x^a y^b\}$.*

Proof: Without loss of generality we assume that $c_i, c_j>0$. Dividing by $x^c y^b$, with $a-c=d-b=e$ gives: $c_i x^e - c_j y^e > (c_i-\epsilon)x^e$. This can be rewritten as: $y^e x^{-e} < \frac{c_i-(c_i-\epsilon)}{c_j} = \frac{\epsilon}{c_j}$ Now take $h=\sqrt[e]{\frac{\epsilon}{c_j}}$, which gives: $y<hx$ where $h>0$ is a constant. Consider the set $\{(x,y)\in\mathbb{N}^2| \; y<hx\}$. $\square$

Combining the previous two lemma's we can generalize this result: ϕ^{+k} will always have a positive term of degree k that dominates the sum of all terms in ϕ^{-k} on a dense subset of $\mathbb{N}^2$:

Lemma 5. *For a polynomial $\phi:\mathbb{N}^2\rightarrow\mathbb{N}$ of degree $k>2$ that defines a bijection there exists a number $h\in\mathbb{R}$ and two numbers $a,b\in\mathbb{N}$ such that $\underline{d}(A)>0$, where $A=\{(x,y)\in\mathbb{N}^2| \; \phi(x,y)>hx^a y^b\}$, provided that $\mathbb{N}^2$ is enumerated by $\pi(x,y)$.*

Proof: If ϕ^- is empty this is guaranteed as well as in the case that ϕ^{-k} is empty. This leaves the case that both ϕ^{-k} and ϕ^{+k} are not empty. The terms in ϕ^k have total ordering $\succ_x$ with a largest element $c_i x^a y^b$. We can always choose a value for $h>0$ such that this term dominates all terms in ϕ^k in A for large enough values of x and y. Consequently the dominating terms are in ϕ^{+k}. We can generalize the result of lemma 4 to the set ϕ^{-k} by observing the expression $h=\sqrt[e]{\frac{\epsilon}{c_j}}$. Here e is the difference in degree between the terms and c_j is the coefficient of the term. We only require that $h>0$, so we can always select an arbitrary small ϵ such that $h'=\sqrt[f]{\frac{\epsilon}{g}}$, where f is the maximum distance between the terms and $g=|\phi^{-k}|c_j$, where c_j the largest coefficient of terms in ϕ^{-k} and $|\phi^{-k}|$ is its cardinality. Now apply lemma 3 $\square$

3.3. Divergence of upper- and lowerbound for $\phi(x,y)$ on dense subsets of $\mathbb{N}$

Combining the results of the previous two sections we show that ϕ generates unbounded randomness deficiency on a subset with density > 0. Which is impossible because by lemma 1 the image of this set under ϕ has density 0.

Theorem 1. *There are no polynomials of degree > 2 that define a bijection between $\mathbb{N}^2$ and $\mathbb{N}$*

Proof: Suppose that such a polynomial function $\phi : \mathbb{N}^2 \to \mathbb{N}$ with degree $k > 2$ exists. We have $(\forall z \in \mathbb{N})(\exists!(x,y) \in \mathbb{N}^2)(\phi(x,y) = z)$. We make two observations:

1. By lemma 2 $K(z) \leq \log x + \log y + O(1)$
2. By lemma 5 there is a $h \in \mathbb{R}$ and two numbers $a, b \in \mathbb{N}$ such that $a + b = k$ and the set $A = \{(x,y) \in \mathbb{N}^2 \mid y < hx\}$ has density $\underline{d}(A) > 0$ with $\forall(x,y) \in A\ \phi(x,y) > hx^a y^b$.

For elements $\phi(x,y) = z$ of this set we can now estimate the randomness deficiency as $\delta(z) = \log z - K(z) \geq \log hx^a y^b - (\log x + \log y + O(1))$. This gives:

$$\delta(z) \geq a \log x + b \log y - \log x - \log y - O(1) \tag{6}$$

For $k = a + b > 2$, by lemma 1, the density of the set for which inequality 6 holds is zero in the limit. There are serveral ways to construct a contradiction on the basis of these observations. The first is that ϕ as a bijection changes the densities of the underlying sets: by lemma 5 the density of A is > 0, by equation 6 and lemma 1 it is 0. But bijections define equinumerability of sets so they cannot change densities of sets. Consequently ϕ is, contrary to our assumption, not a bijection.

A second inconsistency is constructed in the following way: All elements of $\phi(A)$ have a compressible description as solution of a function of degree $k > 2$. The density of $\phi(A)$ in $\mathbb{N}$ is 0. The elements in $\phi(A^{\complement})$, by definition, have no such compressible description of degree k. By equation 6 all elements of $\phi(A^{\complement})$ must have a description of degree 2 and by definition their density is 1. Consequently ϕ in the limit stays asymptotically close to a polynome of degree 2, except for a vanishing set of isolated points, which contradicts the fact that it has degree $k > 2$. $\square$

4. Discussion and Conclusion

The general underlying insight of this paper is that no finite function can generate more information than its input on an infinite set. Equation 6 specifies a necessary information theoretical constraint for any polynomial bijection $\phi : \mathbb{N}^2 \to \mathbb{N}$, which can only be met by functions of degree 2. By the Fueter-Pólya theorem the function $\pi : \mathbb{N}^2 \to \mathbb{N}$ is the only algebraic function on these domains that is information efficient. The essence of the proof is the observation of the fact that elements of $\mathbb{N}$ are both numbers and information bearers. As such they obey the laws of algebra as well as information theory. This dual set of constraints defines a stronger set of conditions then the ones studied in classical number theory. This observation can be developed in to a general theory about the interaction between information and computation. In this paper we have used classical Kolmogorov complexity as main tool, but a proof based solely on recursive functions and information theory is possible.

5. Acknowledgements

This research was partly supported by the Info-Metrics Institute of the American University in Washington, the Commit project of the Dutch science foundation NWO, the Netherlands eScience center and a Templeton Foundations Science and Significance of Complexity Grant supporting The Atlas of Complexity Project. I thank the editor and the anonymous referees for their insightful comments on an earlier version.

6. Bibliography

Fueter, R., Pólya, G., 1923. Rationale Abzählung der Gitterpunkte. Vierteljschr. Naturforsch. Ges. Zürich 58, 280–386.

Li, M., Vitányi, P., 2008. An Introduction to Kolmogorov Complexity and Its Applications. Springer-Verlag.

Nathanson, M. B., 2016. Cantor polynomials and the Fueter-Pólya theorem. American Mathematical Monthly 123 (10), 1001–1012.

Vsemirnov, M., 2002. Two elementary proofs of the Fueter-Pólya theorem on pairing polynomials. St Petersburg Mathematical Journal 13 (5), 705–716.