The Ulam sequence is defined as $a_1 = 1, a_2 = 2$ and a_n being the smallest integer that can be written as the sum of two distinct earlier elements in a unique way. This gives

$$1, 2, 3, 4, 6, 8, 11, 13, 16, 18, 26, 28, 36, 38, 47, \dots$$

Virtually nothing is known about the behavior of the sequence and it is described as 'quite erratic' and seems to 'not follow any recognizable pattern'. Consecutive differences do not seem to be periodic and can be large, e.g. $a_{18858} - a_{18857} = 315$. The purpose of this short note is to report the following empirical discovery: there seems to exist a real $\alpha \sim 2.57145\dots$ such that

$$\{\alpha a_n : n \in \mathbb{N}\} \qquad \text{is } \textit{not} \text{ uniformly distributed mod } 2\pi.$$

The distribution function of αa_n mod 2π seems to be supported on an interval of length ~ 3 and has a curious shape. Indeed, for the first 10.000 elements of Ulam's sequence, we have

$$\cos\left(2.57145\ a_n\right) < 0 \qquad \text{for all } a_n \notin \{2, 3, 47, 69\}\,.$$

All of this indicates that Ulam's sequence has a lot more rigidity than hithereto assumed.

A hidden signal in the Ulam sequence

Stefan Steinerberger
Research Report YALEU/DCS/TR-1508
Yale University
May 4, 2015

Approved for public release: distribution is unlimited.
Keywords: *Ulam sequence, frequency shift*

A HIDDEN SIGNAL IN THE ULAM SEQUENCE

STEFAN STEINERBERGER

ABSTRACT. The Ulam sequence is defined as $a_1 = 1, a_2 = 2$ and a_n being the smallest integer that can be written as the sum of two distinct earlier elements in a unique way. This gives

$$1, 2, 3, 4, 6, 8, 11, 13, 16, 18, 26, 28, 36, 38, 47, \dots$$

Virtually nothing is known about the behavior of the sequence and it is described as 'quite erratic' and seems to 'not follow any recognizable pattern'. Consecutive differences do not seem to be periodic and can be large, e.g. $a_{18858} - a_{18857} = 315$. The purpose of this short note is to report the following empirical discovery: there seems to exist a real $\alpha \sim 2.57145\dots$ such that

$$\{\alpha a_n : n \in \mathbb{N}\} \qquad \text{is } \textit{not} \text{ uniformly distributed mod } 2\pi.$$

The distribution function of αa_n mod 2π seems to be supported on an interval of length ~ 3 and has a curious shape. Indeed, for the first 10.000 elements of Ulam's sequence, we have

$$\cos(2.57145\ a_n) < 0 \qquad \text{for all } a_n \notin \{2, 3, 47, 69\}.$$

All of this indicates that Ulam's sequence has a lot more rigidity than hithereto assumed.

1. Introduction

Stanislaw Ulam introduced his sequence, nowadays called the Ulam sequence,

$$1, 2, 3, 4, 6, 8, 11, 13, 16, 18, 26, 28, 36, 38, 47, 48, 53, 57, 62, 69, 72, 77, 82, 87, 97 \dots\dots$$

in a 1964 survey [12] on unsolved problems. The construction is given by $a_1 = 1, a_2 = 2$ and then iteratively by choosing a_n as the smallest integer that can be written as the sum of two distinct earlier elements in a unique way. Ulam writes

> One can consider a rule for growth of patterns – in one dimension it would be merely a rule for obtaining successive integers. [...] In both cases simple questions that come to mind about the properties of a sequence of integers thus obtained are notoriously hard to answer. (Ulam, 1964)

He asks in the Preface of [13] whether it is possible to determine the asymptotic density of the sequence (this problem is sometimes incorrectly attributed to Recaman [8]). While empirically the density seems to be close to 7.4%, it is not even known whether it is bigger than 0.

Indeed, it seems that hardly anything is known at all: Section C4 of Guy's *Unsolved Problems in Number Theory* [6] contains an observation due to Burr and Zeitlin that $a_{n+1} \le a_n + a_{n-1} < 2a_n$ and that therefore every integer can be written as a sum of Ulam numbers. Different initial values a_1, a_2 can give rise to more structured sequences [2, 4, 7]: for some of them the sequence of consecutive differences $a_{n+1} - a_n$ is eventually periodic. It seems that this is not the case for Ulam's sequence: Knuth [9] remarks that $a_{4953} - a_{4952} = 262$ and $a_{18858} - a_{18857} = 315$. In general, Ulam's sequence 'does not appear to follow any recognizable pattern' [3] and is 'quite erratic' [11].

We describe the (accidental) discovery of some very surprising structure. While using Fourier series with Ulam numbers as frequencies, we noticed a persisting signal in the noise: indeed, dilating the sequence by a factor $\alpha \sim 2.57145\dots$ and considering the sequence (αa_n mod 2π) gives rise to a very regular distribution function. One surprising implication is

$$\cos(2.57145\ a_n) < 0 \qquad \text{for the first 10.000 Ulam numbers except } 2, 3, 47, 69.$$

The dilation factor α seems to be a universal constant (of which we were able to compute the first four digits); it seems likely that $\cos(\alpha a_n) < 0$ holds for all but these four Ulam numbers.

2. The Observation

2.1. **A Fourier series.** We were originally motivated by the following: suppose the Ulam sequence has a positive density, then by a classical theorem of Roth [10] it contains many 3-arithmetic progressions. We were interested in whether the existence of progressions would have any impact on the structure of the sequence (because, if $a \cdot n + b$ is in the sequence for $n = 1, 2, 3$, then either a or $2a$ is not). There is a well-known connection between randomness in sets and smallness of Fourier coefficients and this motivated us to study

$$f_N(x) = \mathrm{Re} \sum_{n=1}^{N} e^{i a_n x} = \sum_{n=1}^{N} \cos(a_n x).$$

Clearly, $f_N(0) = N$. However, since the Ulam sequence is not periodic (at least for small N) and seems to have positive density, one would expect a very large amount of cancellation outside of the origin. Much to our surprise, we discovered that this is not the case.

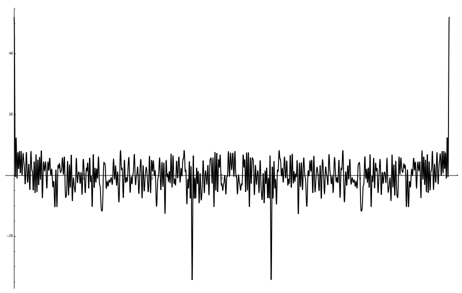

Figure 1. The function f_{50}.

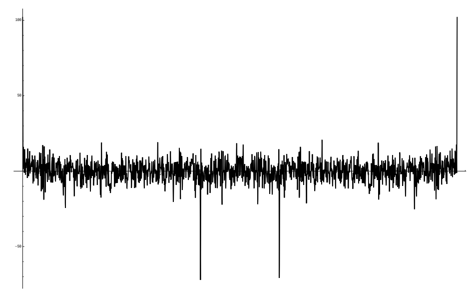

Figure 2. The function f_{100}.

We see that f_N is indeed large around the 0 and small away from the origin with the clear exception of two points: there is a clear signal in the noise; furthermore, the location of the signal is independent of N and its strength is increasing linearly in N.

2.2. **The hidden frequency.** An explicit computation with $N = 10.000$ pinpoints the location of the signal at

$$\alpha \sim 2.57145\ldots$$

and, by symmetry, at $2\pi - \alpha$. This signal acts as a hidden shift in frequency space: we can now remove that shift in frequency by considering

$$S_N = \left\{ \alpha a_n - 2\pi \left\lfloor \frac{\alpha a_n}{2\pi} \right\rfloor : n \leq N \right\} \qquad \text{instead.}$$

Usually, for a generic non-periodic sequence (or even this very sequence with a different α), one would expect the set to be asymptotically uniformly distributed mod 2π (for the simple reason that there is no good reason why any structure should be present). Here, as it turns out, the situation is very different. We clearly see two big centers of concentration; everything is strictly contained inside the interval $[1.4, 5.2]$ and after the first 69 elements of the sequence, everything is in the interval $[1.5, 4.7]$. We do not know whether the discovery of this phenomenon will allow for some additional insight into the structure of Ulam's sequence or only add to its mystery – in any case, it clearly indicates some underlying and very regular structure at work.

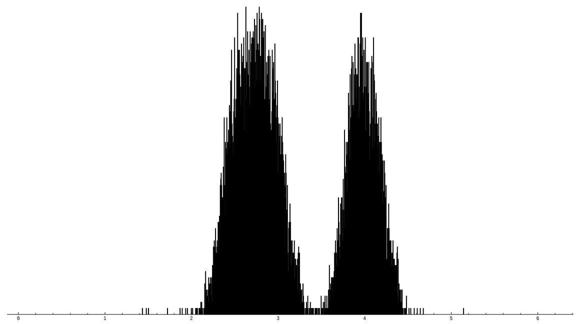

FIGURE 3. Distribution of S_N on $[0, 2\pi]$ for $N = 10.000$.

2.3. **Prior observations.** We emphasize that there have been prior observations hinting at some sort of regularity (see e.g. David Wilson's remark in the *The On-Line Encyclopedia of Integer Sequences* [9]). It is clear that even without a suitable dilation of the Ulam numbers, some form of periodicity is still visible to the human eye – the truly surprising fact is that a suitable dilation does not merely yield some form of quasi-periodicity but actually seems to give rise to a nontrivial measure on $\mathbb{T}$. The factor α is universal: if we dilate by any factor β such that α and β are linearly independent over $\mathbb{Q}$ (which is generically the case and which one would expect for $\beta = 1$), then, assuming there exists a smooth, nontrivial distribution function for ($\alpha a_n \mod 2\pi$), the ergodicity of irrationals on $\mathbb{T}$ implies that the distribution function of ($\beta a_n \mod 2\pi$) will tend towards the uniform distribution.

2.4. **Other initial values.** There has been quite some work on the behavior of Ulam-type sequences with other initial values. It has first been observed by Queneau [7] that the initial values $(a_1, a_2) = (2, 5)$ give rise to sequence where $a_{n+1} - a_n$ is eventually periodic. Finch proved that this is the case whenever only finitely many even numbers appear in the sequence and conjectured that this is the case for $(a_1, a_2) = (2, n)$ whenever $n \geq 5$ is odd. Finch's conjecture was proven by Schmerl & Spiegel [11]. Subsequently, Cassaigne & Finch [1] proved that all sequences starting from $(a_1, a_2) = (4, n)$ with $n \equiv 1 \pmod 4$ contain precisely three even integers and are thus eventually periodic. The sequence $(a_1, a_2) = (2, 3)$ as well as all sequences $(a_1, a_2) = (1, n)$ with $n \in \mathbb{N}$ do not exhibit such behavior and are being described 'erratic' in the above literature.

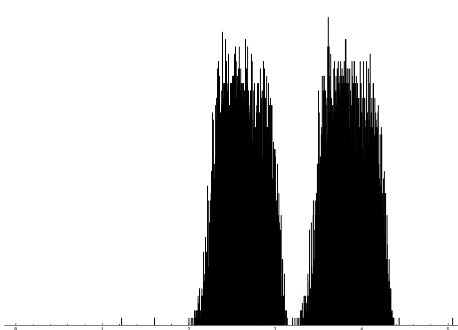

FIGURE 4. Distribution for initial value $(a_1, a_2) = (1, 3)$.

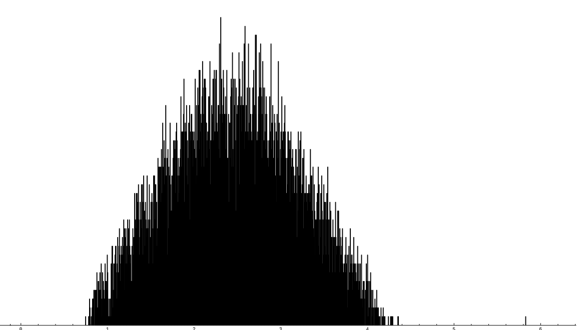

FIGURE 5. Distribution for the initial values $(a_1, a_2) = (2, 3)$.

It seems that the underlying mechanism is indeed independent of the initial values. Using again $N = 10.000$ elements, we found that the frequencies for the erratic sequences with $(a_1, a_2) = (1, 3)$ and $(a_1, a_2) = (2, 3)$ are given by

$$\alpha_{(1,3)} \sim 2.8335 \qquad \text{and} \qquad \alpha_{(2,3)} \sim 1.1650$$

and removing that hidden frequeny in the same way as above gives rise to the following two distributions, where the first one appears to be of the very same type as the distribution in the classical case while the second one is very different.

References

[1] J. Cassaigne and S. Finch, A class of 1-additive sequences and quadratic recurrences. Experiment. Math. 4 (1995), no. 1, 49-60.
[2] S. Finch, Conjectures about s-additive sequences. Fibonacci Quart. 29 (1991), no. 3, 209-214.
[3] S. Finch, Patterns in 1-additive sequences. Experiment. Math. 1 (1992), no. 1, 57-63
[4] S. Finch, On the regularity of certain 1-additive sequences. J. Combin. Theory Ser. A 60 (1992), no. 1, 123-130.
[5] V. Gardiner, R. Lazarus and N. Metropolis, On certain sequences of integers defined by sieves. Math. Mag. 29 (1956), 117-122.
[6] Richard K. Guy, Unsolved problems in number theory. Third edition. Problem Books in Mathematics. Springer-Verlag, New York, 2004.
[7] R. Queneau, Sur les suites s-additives. J. Combinatorial Theory Ser. A 12 (1972), 31-71.
[8] B. Recaman, Research Problems: Questions on a Sequence of Ulam. Amer. Math. Monthly 80 (1973), 919-920.
[9] The On-Line Encyclopedia of Integer Sequences, http://oeis.org, 30. Apr 2015, Sequence A002858
[10] K. F. Roth, On certain sets of integers. J. London Math. Soc. 28, (1953). 104-109.
[11] J. Schmerl and E. Spiegel, The regularity of some 1-additive sequences. J. Combin. Theory Ser. A 66 (1994), no. 1, 172-175.
[12] S. Ulam, Combinatorial analysis in infinite sets and some physical theories. SIAM Rev. 6 1964 343-355.
[13] S. M. Ulam, Problems in modern mathematics. Science Editions John Wiley & Sons, Inc., New York 1964

Department of Mathematics, Yale University, 10 Hillhouse Avenue, New Haven, CT 06511, USA
E-mail address: stefan.steinerberger@yale.edu