LOWER AND UPPER BOUNDS ON IRREGULARITIES OF DISTRIBUTION

Karl Levy
Department of Mathematics, BMCC CUNY, New York, New York

Received: 10/10/18, Revised: 8/15/19, Accepted: 3/23/20, Published: 4/13/20

Abstract

A sequence $(x_1, x_2, \cdots, x_{N+d})$ is an N-regular sequence with at most d irregularities if, for every $n \leq N$, each one of the intervals $[0,1),[1,2),\cdots,[n-1,n)$ contains at least one term from the sequence $(nx_1, nx_2, \cdots, nx_{n+d})$. The function $s(d)$ is equal to the largest N for which there exists an N-regular sequence with at most d irregularities. In the current paper we show that $\lfloor\sqrt{4d+895}+1\rfloor \leq s(d) < 24801d^3 + 942d^2 + 3$ for $d \geq 1$.

1. Introduction

In his long famous book, *One Hundred Problems in Elementary Mathematics* [5], Steinhaus asked a question about the regularity of sequences. Before repeating his question, we define his notion of regularity.

Definition 1. A sequence

$$(x_1, x_2, \cdots, x_N)$$

is *N-regular* if, for every $n \leq N$, each one of the intervals

$$[0,1),[1,2),\cdots,[n-1,n)$$

contains one term from the sequence

$$(nx_1, nx_2, \cdots, nx_n).$$

For example, though the sequence $(1/3, 1/2)$ is 2-regular, its extension to $(1/3, 1/2, 0)$ is not 3-regular. With the previous definition in mind, Steinhaus's question is easily given: is there a longest N-regular sequence? A few years after Steinhaus asked this, Warmus answered thus: N-regular sequences are at most seventeen elements long [6]. In the same paper Warmus gave the following example of one such maximal sequence

$$\left(\frac{4}{7},\frac{2}{7},\frac{16}{17},\frac{1}{14},\frac{8}{11},\frac{5}{11},\frac{1}{7},\frac{14}{17},\frac{3}{8},\frac{11}{17},\frac{3}{14},\frac{15}{17},\frac{1}{2},0,\frac{13}{17},\frac{5}{16},\frac{10}{17}\right).$$

Steinhaus's question was answered. But then, Berlekamp and Graham [1] asked a more general question. Again, it is helpful to first give a definition, namely, of their more relaxed notion of regularity, before repeating their question.

Definition 2. A sequence
$$(x_1, x_2, \cdots, x_{N+d})$$
is *N-regular with at most d irregularities* if, for every $n \leq N$, each one of the intervals
$$[0,1), [1,2), \cdots, [n-1, n)$$
contains at least one term from the sequence
$$(nx_1, nx_2, \cdots, nx_{n+d}).$$

For example, though the sequence $(1/3, 2/5, 1/2)$ is 2-regular with at most one irregularity, its extension to $(1/3, 2/5, 1/2, 0)$ is not 3-regular with at most one irregularity. With the previous definition in mind, Graham and Berlekamp's question can be stated as follows: for a given $d \geq 0$, what is the largest N for which there exists an N-regular sequence with at most d irregularities? Next, we introduce some notation that conveniently subsumes most of their question into a function.

Definition 3. For all $d \geq 0$, $s(d)$ is equal to the largest N for which there exists an N-regular sequence with at most d irregularities.

Graham and Berlekamp's question becomes, concisely: what is $s(d)$? To what extent has this question been answered? In terms of our function, Warmus's result from above is $s(0) = 17$. Unfortunately, for $d \geq 1$, the exact values of $s(d)$ remain unknown. There are, however, some lower and upper bounds.

For lower bounds, Oliveira's recent construction of the following 31-regular sequence with at most one irregularity [3],

$$\left(0, \frac{11}{29}, \frac{13}{16}, \frac{4}{19}, \frac{20}{29}, \frac{9}{16}, \frac{19}{20}, \frac{11}{24}, \frac{8}{29}, \frac{1}{8}, \frac{16}{21}, \frac{28}{31}, \frac{16}{25}, \frac{13}{25}, \frac{7}{22}, \frac{5}{29}, \frac{1}{12}, \frac{17}{20}, \frac{5}{12}, \frac{3}{5}, \frac{21}{29}, \right.$$
$$\left. \frac{30}{31}, \frac{7}{29}, \frac{1}{24}, \frac{10}{29}, \frac{15}{31}, \frac{24}{31}, \frac{27}{31}, \frac{19}{29}, \frac{4}{29}, \frac{17}{30}, \frac{13}{31}\right), \quad (1.1)$$

means that $s(1) \geq 31$.[1] In general, by Corollary 1, $s(d) \geq \left\lfloor \sqrt{4d + 895} + 1 \right\rfloor$ for all $d \geq 1$.

For upper bounds, in their 1970 paper [1], Berlekamp and Graham gave a proof that $s(d) < 4^{(d+2)^2}$ for all $d \geq 0$. But then in 2012, in a private email received by Graham from David and Moshe Newman, it was pointed out that the proof was

[1]Oliveira claimed to have verified that $s(1) = 31$ by an exhaustive computer search but the computer code was not provided.

incomplete. In 2013, Graham responded with a note [2] acknowledging this and pointing out that a result of the same form, namely, that $s(d) < \exp(cd^2)$, for an appropriate absolute constant c, follows directly from a fundamental inequality in Roth's paper on discrepancies [4]. Further, in the same note, Graham outlined ideas for the following improved result.

Theorem 1 (Graham). *For all $d \geq 1$, $s(d) < 16000d^3$.*

Unfortunately, it was impossible to reconstruct the details outlined in Graham's note. In Section 3, using some of the ideas outlined in Graham's note, we give a detailed proof of the slightly weaker bound, namely, that $s(d) < 24801d^3+942d^2+3$.

2. A Lower Bound for $s(d)$

First we prove something about the spacing of terms in N-regular sequences with at most d irregularities.

Lemma 1. *If*

$$(x_1, x_2, \cdots, x_{N+d})$$

is N-regular with at most d irregularities, then, for each positive integer $k \leq N$, the interval $[k-1, k+1)$ contains at least one element from the sequence

$$((N+1)x_1, (N+1)x_2, \cdots, (N+1)x_{N+d}).$$

Proof. Assume that there does exist a positive integer k such that $[k-1, k+1)$ does not contain a term from the sequence

$$((N+1)x_1, (N+1)x_2, \cdots, (N+1)x_{N+d}).$$

This is equivalent to assuming that, for each $x_i \in (x_1, x_2, \cdots, x_{N+d})$, either

$$(N+1)x_i < k-1$$

or

$$(N+1)x_i \geq k+1.$$

In the former case, this implies that $x_i < 1$. In the later case, this implies that $x_i \geq 0$. By pairing our inequalities, we have that either

$$Nx_i < k-1$$

or

$$Nx_i \geq k.$$

This is equivalent to $[k-1,k)$ not containing a term from the sequence

$$(Nx_1, Nx_2, \cdots, Nx_{N+d}).$$

This, in turn, contradicts our lemma's assumption that $(x_1, x_2, \cdots, x_{N+d})$ is N-regular with at most d regularities. $\square$

Next, we "extend the regularity" of a sequence.

Lemma 2. *If*

$$(x_1, x_2, \cdots, x_{N+d})$$

is N-regular with at most d irregularities, then there exists a sequence

$$(x'_1, x'_2, \cdots, x'_{\lceil \frac{N+1}{2} \rceil})$$

such that the sequence formed by concatenating the two,

$$(x_1, \cdots, x_{N+d}, x'_1, \cdots, x'_{\lceil \frac{N+1}{2} \rceil}),$$

is $(N+1)$-regular with at most $(d + \lceil \frac{N+1}{2} \rceil - 1)$ irregularities.

Proof. In Lemma 1, we proved that when the terms of an N-regular sequence are multiplied by $(N+1)$, they do not miss two consecutive unit-length intervals. This means that at most $\lceil \frac{N+1}{2} \rceil$ of the intervals

$$[0,1), [1,2), \cdots, [N, N+1)$$

do not contain a term from

$$((N+1)x_1, (N+1)x_2, \cdots, (N+1)x_{N+d}).$$

We pick the terms in the sequence

$$(x'_1, x'_2, \cdots, x'_{\lceil \frac{N+1}{2} \rceil})$$

so that at least one of them is in each of the at most $\lceil \frac{N+1}{2} \rceil$ intervals not containing a term from

$$((N+1)x_1, (N+1)x_2, \cdots, (N+1)x_{N+d}).$$

This guarantees that

$$(x_1, \cdots, x_{N+d}, x'_1, \cdots, x'_{\lceil \frac{N+1}{2} \rceil})$$

is an $(N+1)$-regular sequence with at most $(d + \lceil \frac{N+1}{2} \rceil - 1)$ irregularities. $\square$

Next, we put the previous two lemmas together to get our lower bound.

Theorem 2. *For all $d' \geq 1$, if*

$$s(d') \geq N,$$

then

$$s(d) \geq \left\lfloor \sqrt{4d - 4d' - 1 + (N-1)^2} + 1 \right\rfloor$$

for all $d \geq d'$.

Proof. By Definition 3, $s(d') \geq N$ if and only if an N-regular sequence with at most d' irregularities exists. By Lemma 1, if an N-regular sequence with at most d irregularities exists, then an $(N+1)$-regular seqence with at most $(d + \lceil \frac{N+1}{2} \rceil - 1)$ irregularities must also exist. Thus, again by Definition 3, an $(N+1)$-regular sequence with at most $(d + \lceil \frac{N+1}{2} \rceil - 1)$ irregularities exists, if and only if $s(d' + \lceil \frac{N+1}{2} \rceil - 1) \geq N + 1$. Recursively repeating this argument j times yields

$$s(d' + \sum_{i=1}^{j} \left(\left\lceil \frac{N+i}{2} \right\rceil - 1 \right)) \geq N + j$$

for all $j \geq 1$. We rewrite this as

$$s(d_j^*) \geq \sqrt{4d_j^* - 4d' - 1 + (N-1)^2} + 1 \tag{2.1}$$

for the increasing sequence of values

$$(d_j^*)_{j=1}^{\infty} = (d' + \sum_{i=1}^{j} \left(\left\lceil \frac{N+i}{2} \right\rceil - 1 \right))_{j=1}^{\infty}.$$

By Definition 3, $s(d)$ is an increasing, integer-valued function. By examination, we see that

$$\sqrt{4d_{j+1}^* - 4d' - 1 + (N-1)^2} + 1 = \left(\sqrt{4d_j^* - 4d' - 1 + (N-1)^2} + 1 \right) + 1$$

holds for the right-side of Inequality (2.1). Combining the previous two facts yields that

$$s(d) \geq \left\lfloor \sqrt{4d - 4d' - 1 + (N-1)^2} + 1 \right\rfloor$$

for all $d \geq d'$. □

Corollary 1. *For all $d \geq 1$,*

$$s(d) \geq \left\lfloor \sqrt{4d + 895} + 1 \right\rfloor.$$

Proof. By Definition 3, the existence of the 31-regular sequence with one irregularity, as demonstrated in [3], is equivalent to $s(1) \geq 31$. Setting $d' = 1$ and $N = 31$, in Theorem 2, yields our result. □

2.1. A Few Thoughts on Improving the Lower Bound for $s(d)$

Perhaps, by examining the geometric structure of n-regular subsequences, the maximum number of intervals left empty when multiplying the terms of an N-regular sequence by $(N+1)$ could be made fewer? Perhaps some specific N-regular sequence with at most d irregularities can be found that leaves fewer empty intervals when multiplied by $(N+1)$?

3. An Upper Bound for $s(d)$

We show that a sequence X, which we assume to be N-regular with at most d irregularities, contains a certain set of terms we call P'. We then use the set P' to show that if N were allowed to be greater than some d-dependent value, then X would be forced to have more than d irregularities. This contradicts the assumption that X has at most d irregularities. The process of establishing this contradiction yields our upper bound.

3.1. Subsequences P and P'

First, we show that there is an increasing subsequence

$$P \subset X$$

such that each of its terms, when dilated by some positive integer n_0, is contained in a unit-length interval separated from the subsequence's next dilated term by an empty unit-length interval.

Lemma 3. *Let*

$$X = (x_1, x_2, \cdots, x_{N+d})$$

be N-regular with at most d irregularities. If l, m and n_0 are all positive integer constants such that

$$l + 8md + 3 \leq n_0 \leq N,$$

then there exists a (not necessarily order preserving) subsequence taken from the first $n_0 + d$ terms of X,

$$P = (v_1 < w_1 < v_2 < w_2 < \cdots < v_{2md+1} < w_{2md+1}),$$

such that

$$n_0 v_i \in [l + 4i - 4, l + 4i - 3) \text{ and } n_0 w_i \in [l + 4i - 2, l + 4i - 1).$$

Proof. For all $n_0 \leq N$, Definition 2 guarantees that each one of the intervals

$$[0, 1), [1, 0), \cdots, [n_0 - 1, n_0)$$

contains at least one term from $(n_0x_1, n_0x_2, \cdots, n_0x_{n_0+d})$. Since $1 \leq l \leq n_0 - 8md + 3$ and $1 \leq i \leq 2md + 1$, it follows that the set of all intervals of the form

$$[l + 4i - 4, l + 4i - 3) \text{ or } [l + 4i - 2, l + 4i - 1)$$

is a subset of the intervals

$$[0, 1), [1, 0), \cdots, [n_0 - 1, n_0).$$

We construct P by picking, in ascending order, one of the n_0x_i from each of the $4md + 2$ intervals of the form

$$[l + 4i - 4, l + 4i - 3) \text{ or } [l + 4i - 2, l + 4i - 1)$$

and then dividing by the coefficient n_0. □

The previous lemma leads directly to the following bounds on both the values of and the distances between P's paired terms.

Corollary 2. *Let*

$$X = (x_1, x_2, \cdots, x_{N+d})$$

be N-regular with at most d irregularities. If l, m and n_0 are all positive integer constants such that

$$l + 8md + 3 \leq n_0 \leq N,$$

then there exists a (not necessarily order preserving) subsequence taken from the first $n_0 + d$ terms of X,

$$P = (v_1 < w_1 < v_2 < w_2 < \cdots < v_{2md+1} < w_{2md+1}),$$

such that

$$1 < n_0(w_i - v_i) < 3 \tag{3.1}$$

and

$$\frac{l}{n_0} \leq v_i < w_i < \frac{l + 8md + 3}{n_0}. \tag{3.2}$$

Proof. Since n_0v_i and n_0w_i are contained, respectively, in the intervals $[l + 4i - 4, l + 4i - 3)$ and $[l + 4i - 2, l + 4i - 1)$, two unit-length intervals that are exactly separated by a unit-length interval, it follows that

$$1 < n_0(w_i - v_i) < 3.$$

Since $n_0 v_1 < n_0 v_i < n_0 w_i < n_0 w_{2md+1}$, $n_0 v_1 \in [l, l+1)$ and $n_0 w_k \in [l + 8md + 3, l + 8md + 3)$, it follows that

$$l \leq n_0 v_i < n_0 w_i < l + 8md + 3n_0$$

and, thus, that

$$\frac{l}{n_0} \leq v_i < w_i < \frac{l + 8md + 3}{n_0}.$$

□

Next, we show that there is a certain subsequence, $P' \subset P$, made up of paired terms all separated by almost the same distance. Controlling this distance is key to forcing the contradiction that yields our upper bound for $s(d)$.

Lemma 4. *Let X be an N-regular sequence with at most d irregularities. If l, m and n_0 are all positive integer constants such that*

$$l + 8md + 3 \leq n_0 \leq N,$$

then there exists a (not necessarily order preserving) subsequence taken from the first $n_0 + d$ terms of X,

$$P' = (y_1 < z_1 < y_2 < z_2 < \cdots < y_{d+1} < z_{d+1}),$$

such that, for some positive integer $r \leq 2m$,

$$1 + \frac{r-1}{m} < n_0(z_i - y_i) \leq 1 + \frac{r}{m}$$

holds for all y_i, z_i.

Proof. By Lemma 3, there exists a (not necessarily order preserving) subsequence taken from the first $n_0 + d$ terms of X,

$$P = (v_1 < w_1 < v_2 < w_2 < \cdots < v_{2md+1} < w_{2md+1}).$$

By Corollary 2, we know that

$$1 < n_0(w_i - v_i) < 3$$

holds for the $2md + 1$ pairs w_i, v_i in P. If we partition the interval $(1, 3)$ into $2m$ subintervals of length $1/m$, then, by the pigeonhole principle, there must exist a subsequence

$$P' = (y_1 < z_1 < y_2 < z_2 < \cdots < y_{d+1} < z_{d+1})$$

of P such that, for some positive integer $r \leq 2m$,

$$1 + \frac{r-1}{m} < n_0(z_i - y_i) \leq 1 + \frac{r}{m}$$

for the $d + 1$ pairs y_i, z_i in P'. □

3.2. Increasing the Dilated Distances Between Paired Terms in P'

Next, we find an $n_1 < n_0$ such that when the paired terms of P' are dilated by this n_1, the distances between them are all slightly greater than three. This is the key to controlling the number of unit-length intervals between the dilated paired terms of P'. This, in turn, is the key to getting our upper bound on $s(d)$.

Lemma 5. *Let X be an N-regular sequence with at most d irregularities. If*

1. *l, m and n_0 are all positive integer constants such that $l + 8md + 3 \le n_0 \le N$; and*
2. *$P' \subset X$ is as given by Lemma 4,*

then there exists a positive integer constant n_1 such that both $n_0 < n_1 \le 3n_0 + (l + 8md + 3)$ and

$$3 + \frac{l + 8md + 3}{n_0} \le n_1(z_i - y_i) \le \left(1 + \frac{1}{m}\right)\left(3 + \frac{l + 8md + 3}{n_0}\right) + \frac{3}{n_0}$$

hold for all the y_i, z_i in P'.

Proof. By Lemma 4, there exists a (not necessarily order preserving) subsequence of the first $n_0 + d$ terms of X,

$$P' = (y_1 < z_1 < y_2 < z_2 < \cdots < y_{d+1} < z_{d+1}),$$

such that, for some positive integer $r \le 2m$,

$$1 + \frac{r-1}{m} < n_0(z_i - y_i) \le 1 + \frac{r}{m} \tag{3.3}$$

holds for all the y_i, z_i in P'. Multiplying Inequalities (3.3) by

$$\frac{m}{m + r - 1}\left(3 + \frac{l + 8md + 3}{n_0}\right)$$

yields

$$\begin{aligned} 3 + \frac{l + 8md + 3}{n_0} &< \frac{m}{m + r - 1}\left(3n_0 + l + 8md + 3\right)(z_i - y_i) \\ &\le \left(1 + \frac{1}{m + r - 1}\right)\left(3 + \frac{l + 8md + 3}{n_0}\right). \end{aligned}$$

The inequality to the right in (3.3) implies that $z_i - y_i \le \frac{1}{n_0}\left(1 + \frac{r}{m}\right)$. This in turn, when the ceiling function is applied, implies that

$$\begin{aligned} 3 + \frac{l + 8md + 3}{n_0} &< \left\lceil \frac{m}{m + r - 1}\left(3n_0 + l + 8md + 3\right)\right\rceil (z_i - y_i) \\ &\le \left(1 + \frac{1}{m + r - 1}\right)\left(3 + \frac{l + 8md + 3}{n_0}\right) + \frac{1}{n_0}\left(1 + \frac{r}{m}\right). \end{aligned} \tag{3.4}$$

Next, we set

$$n_1 = \left\lceil \frac{m}{m+r-1}\left(3n_0 + l + 8md + 3\right)\right\rceil. \tag{3.5}$$

Since r is between 1 and $2m$, it follows that

$$n_0 \le n_1 \le 3n_0 + (l + 8md + 3).$$

Again, since r is between 1 and $2m$, it directly follows that

$$1 + \frac{1}{m+r-1} \le 1 + \frac{1}{m}$$

and

$$\frac{1}{n_0}\left(1 + \frac{r}{m}\right) \le \frac{3}{n_0}.$$

Finally, combining these two previous inequalities and our definition of n_1 with Inequality (3.4) gives us that

$$3 + \frac{l + 8md + 3}{n_0} < n_1(z_i - y_i) \le \left(1 + \frac{1}{m}\right)\left(3 + \frac{l + 8md + 3}{n_0}\right) + \frac{3}{n_0}$$

for all the y_i, z_i in P'. □

3.3. Forcing $d + 1$ Irregularities

In this final section we show that paired terms in P' contain an extra point between them when dilated by a large fraction of the values between n_1 and a certain n_2.

Theorem 3. *Let X be an N-regular sequence with at most d irregularities. If*

1. *l, m and n_0 are all positive integer constants such that $l + 8md + 3 \le n_0 \le N$; and*
2. *$l = 351d^2$, $m = 35d$ and $n_0 = 8267d^3$; and*
3. *n_1 and P' are as given by Lemma 5; and*
4. *$n_2 = n_1 + 311d^2$,*

then, for each of the $d + 1$ pairs y_i, z_i in P', there exists some set of more than $\frac{d}{d+1}(n_2 - n_1)$ of the positive integers n between n_1 and n_2 such that at least two points from the first $n + d$ terms of the n-dilated sequence nX are contained in one of the four intervals

$$[\lfloor ny_i \rfloor, \lfloor ny_i \rfloor + 1), [\lfloor ny_i \rfloor + 1, \lfloor ny_i \rfloor + 2), [\lfloor ny_i \rfloor + 2, \lfloor nz_i \rfloor), [\lfloor nz_i \rfloor, \lfloor nz_i \rfloor + 1).$$

Proof. We begin by dilating each of the y_i from our subsequence

$$P' = (y_1 < z_1 < y_2 < z_2 < \cdots < y_{d+1} < z_{d+1})$$

by the smallest positive integer $n_i^* \geq n_1$ so that $n_i^* \cdot y_i$ is immediately to the left of a positive integer, that is, such that

$$\lfloor n_i^* \cdot y_i + y_i \rfloor = \lfloor n_i^* \cdot y_i \rfloor + 1. \tag{3.6}$$

This implies that

$$n_i^* \leq n_1 + \left\lceil \frac{1}{y_i} \right\rceil. \tag{3.7}$$

In Lemma 5, we proved that

$$n_1(z_i - y_i) \geq 3 + \frac{l + 8md + 3}{n_0}$$

for all the $y_i, z_i \in P'$. This, since $n_i^* \geq n_1$, implies that

$$n_i^*(z_i - y_i) \geq 3 + \frac{l + 8md + 3}{n_0}.$$

Rearranging the last inequality and taking the floor of both sides gives us that

$$\lfloor n_i^* \cdot z_i \rfloor - 3 \geq \left\lfloor n_i^* \cdot y_i + \frac{l + 8md + 3}{n_0} \right\rfloor. \tag{3.8}$$

Lemma 4 picked the terms of P' from Lemma 3's P, thus, by Corollary 2, we have that

$$y_i < z_i < \frac{l + 8md + 3}{n_0}. \tag{3.9}$$

This, combined with Inequality (3.8), implies the looser inequality

$$\lfloor n_i^* \cdot z_i \rfloor - 3 \geq \lfloor n_i^* \cdot y_i + y_i \rfloor.$$

Combining this looser inequality with Equation (3.6) gives us that

$$\lfloor n_i^* \cdot z_i \rfloor - \lfloor n_i^* \cdot y_i \rfloor \geq 4.$$

This means that the unit-length intervals containing $n_i^* \cdot y_i$ and $n_i^* z_i$ are separated by three unit-length intervals.

Next, for each of the y_i, we define the positive constant

$$k_i = \left\lceil \frac{1}{y_i} \left\{ 4 - \frac{n_2}{n_1} \left[\left(1 + \frac{1}{m}\right) \left(3 + \frac{l + 8md + 3}{n_0}\right) + \frac{3}{n_0} \right] \right\} \right\rceil - 1. \tag{3.10}$$

This yields the inequality

$$k_i < \frac{1}{y_i}\left\{4 - \frac{n_2}{n_1}\left[\left(1+\frac{1}{m}\right)\left(3+\frac{l+8md+3}{n_0}\right)+\frac{3}{n_0}\right]\right\}.$$

This, in turn, combined with

$$n_1(z_i - y_i) \leq \left(1+\frac{1}{m}\right)\left(3+\frac{l+8md+3}{n_0}\right)+\frac{3}{n_0}$$

from Lemma 5, implies the looser inequality

$$k_i y_i + \frac{n_2}{n_1}[n_1(z_i - y_i)] < 4.$$

This, in turn, assuming that $n + k_i \leq n_2$, implies the still looser inequality

$$k_i y_i + (n+k)(z_i - y_i) < 4, \tag{3.11}$$

This inequality, in turn, as long as n is between n_i^* and n_2 and $\lfloor ny_i \rfloor = \lfloor (n+1)y_i \rfloor - 1$, implies that the following equations hold:

$$\begin{aligned}\lfloor (n+1)z_i \rfloor - \lfloor (n+1)y_i \rfloor = \lfloor (n+2)z_i \rfloor - \lfloor (n+2)y_i \rfloor = \cdots \\ = \lfloor (n+k_i)z_i \rfloor - \lfloor (n+k_i)y_i \rfloor = 3.\end{aligned} \tag{3.12}$$

In other words, if ny_i is immediately to the left of an integer, then the next k_i dilated pairs y_i, z_i are separated by two unit-length intervals.

Observing that there are at most $\lceil \frac{1}{y_i} \rceil$ values of n for which ny_i is between two integers, we divide k_i by this amount. Also–to insure, as we assumed above, that $n + k_i \leq n_2$ for all the n for which ny_i is immediately to the left of an integer–we throw away $\lceil \frac{1}{y_i} \rceil$ of the values of n that are immediately less than n_2. Thus, we have, for at least

$$\frac{k_i}{\lceil \frac{1}{y_i} \rceil}((n_2 - \lceil \frac{1}{y_i} \rceil) - (n_i^* + 1))$$

of the positive integers n between $n_i^* + 1$ and $n_2 - \lceil \frac{1}{y_i} \rceil$, that all of the pairs ny_i, nz_i are separated by two unit-length intervals. Put concisely, we have that

$$\begin{aligned}\min_{1 \leq i \leq d+1} \frac{\#\{n : (n_i^* + 1 \leq n \leq n_2) \wedge (\lfloor nz_i \rfloor - \lfloor ny_i \rfloor = 3)\}}{n_2 - n_1} \\ \geq \min_{1 \leq i \leq d+1} \frac{k_i}{\lceil \frac{1}{y_i} \rceil} \frac{n_2 - \lceil \frac{1}{y_i} \rceil - n_i^*}{n_2 - n_1}.\end{aligned} \tag{3.13}$$

By the upper bound for n_i^*, from Inequality (3.7), and by the lower bound for y_i, from Corollary 2, we have that

$$\min_{1 \leq i \leq d+1} \frac{k_i}{\lceil \frac{1}{y_i} \rceil} \frac{n_2 - \lceil \frac{1}{y_i} \rceil - n_i^*}{n_2 - n_1} \geq \min_{1 \leq i \leq d+1} \frac{k_i}{\lceil \frac{1}{y_i} \rceil} \frac{n_2 - \lceil \frac{n_0}{l} \rceil - (n_1 + \lceil \frac{n_0}{l} \rceil)}{n_2 - n_1}.$$

Using Equation (3.10) to substitute for k_i, together with the fact that $\lceil \frac{n_0}{l} \rceil \leq \frac{n_0}{l}+1$, gives us that

$$\begin{aligned}
&\min_{1\leq i\leq d+1} \frac{k_i}{\lceil \frac{1}{y_i} \rceil} \frac{n_2-n_1-2\lceil \frac{n_0}{l} \rceil}{n_2-n_1} \\
&\geq \min_{1\leq i\leq d+1} \left(\frac{1}{y_i} \left\{ 4 - \frac{n_2}{n_1} \left[\left(1+\frac{1}{m}\right) \left(3+\frac{l+8md+3}{n_0}\right) + \frac{3}{n_0} \right] \right\} - 1 \right) \frac{1}{\lceil \frac{1}{y_i} \rceil} \cdot \\
&\qquad \left(1 - \frac{2(n_0+l)}{l(n_2-n_1)}\right).
\end{aligned}$$

Confining our focus to the above left two factors, together with the fact that $\lceil \frac{1}{y_i} \rceil \leq \frac{1}{y_i}+1$, gives us that

$$\begin{aligned}
&\min_{1\leq i\leq d+1} \left(\frac{1}{y_i} \left\{ 4 - \frac{n_2}{n_1} \left[\left(1+\frac{1}{m}\right) \left(3+\frac{l+8md+3}{n_0}\right) + \frac{3}{n_0} \right] \right\} - 1 \right) \frac{1}{\lceil \frac{1}{y_i} \rceil} \\
&\geq \min_{1\leq i\leq d+1} \left(\frac{1}{y_i} \left\{ 4 - \frac{n_2}{n_1} \left[\left(1+\frac{1}{m}\right) \left(3+\frac{l+8md+3}{n_0}\right) + \frac{3}{n_0} \right] \right\} - 1 \right) \frac{y_i}{y_i+1} \\
&= \min_{1\leq i\leq d+1} \left(\frac{1}{y_i+1} \left\{ 4 - \frac{n_2}{n_1} \left[\left(1+\frac{1}{m}\right) \left(3+\frac{l+8md+3}{n_0}\right) + \frac{3}{n_0} \right] \right\} - \frac{y_i}{y_i+1} \right).
\end{aligned}$$

Since $y_i \in P$, Corollary 2 gives us that $y_i \leq \frac{l+8md+3}{n0}$. This, in turn, gives us that

$$\begin{aligned}
&\min_{1\leq i\leq d+1} \left(\frac{1}{y_i+1} \left\{ 4 - \frac{n_2}{n_1} \left[\left(1+\frac{1}{m}\right) \left(3+\frac{l+8md+3}{n_0}\right) + \frac{3}{n_0} \right] \right\} - \frac{y_i}{y_i+1} \right) \\
&\geq \frac{n_0}{l+8md+3+n_0} \left\{ 4 - \frac{n_2}{n_1} \left[\left(1+\frac{1}{m}\right) \left(3+\frac{l+8md+3}{n_0}\right) + \frac{3}{n_0} \right] \right\} - \\
&\qquad \frac{l+8md+3}{l+8md+3+n_0}.
\end{aligned}$$

Now, $n_2 = n_1 + 190d^2$, from Premise (4), and $n_1 \geq n_0$, from Lemma 5, give us that

$$\frac{n_2}{n_1} \geq 1 + \frac{190d^2}{n_0}.$$

Combining the chain of inequalities all the way back to (3.13) gives us that

$$\min_{1\le i\le d+1} \frac{\#\{n : (n_i^*+1 \le n \le n_2) \wedge (\lfloor nz_i\rfloor - \lfloor ny_i\rfloor = 3)\}}{n_2-n_1}$$
$$\ge \left(\frac{n_0}{l+8md+3+n_0}\left\{4-\left(1+\frac{190d^2}{n_0}\right)\left[\left(1+\frac{1}{m}\right)\left(3+\frac{l+8md+3}{n_0}\right)+\right.\right.\right.$$
$$\left.\left.\left.\frac{3}{n_0}\right]\right\}-\frac{l+8md+3}{l+8md+3+n_0}\right)\left(1-\frac{n_0+l}{l(n_2-n_1)}\right).$$

By substituting in the values from Premise (2), we have that

$$\min_{1\le i\le d+1} \frac{\#\{n : (n_i^*+1 \le n \le n_2) \wedge (\lfloor nz_i\rfloor - \lfloor ny_i\rfloor = 3)\}}{n_2-n_1}$$
$$\ge \frac{\left(159357726d^5-57282192d^4-1389390d^3-264027d^2-8381d-95\right)}{1724293890(1914d^3+150d^2+1)}\cdot$$
$$\frac{\left(20710d^2-2871d-109\right)}{d^4}.$$

Since

$$\frac{\left(159357726d^5-57282192d^4-1389390d^3-264027d^2-8381d-95\right)}{1724293890(1914d^3+150d^2+1)}\cdot$$
$$\frac{\left(20710d^2-2871d-109\right)}{d^4} > \frac{d}{d+1}$$

for all $d \ge 1$, we combine these last two inequalities and have that

$$\min_{1\le i\le d+1} \frac{\#\{n : (n_i^*+1 \le n \le n_2) \wedge (\lfloor nz_i\rfloor - \lfloor ny_i\rfloor = 3)\}}{n_2-n_1} > \frac{d}{d+1}.$$

Since we have assumed X to be N-regular (with at most d irregularities), we have that each of the five intervals

$$[\lfloor n_i^*\cdot y_i\rfloor, \lfloor n_i^*\cdot y_i\rfloor+1), [\lfloor n_i^*\cdot y_i\rfloor+1, \lfloor n_i^*\cdot y_i\rfloor+2), [\lfloor n_i^*\cdot y_i\rfloor+2, \lfloor n_i^*\cdot y_i\rfloor+3),$$
$$[\lfloor n_i^*\cdot y_i\rfloor+3, \lfloor n_i^*\cdot z_i\rfloor), [\lfloor n_i^*\cdot z_i\rfloor, \lfloor n_i^*\cdot z_i\rfloor+1)$$

must contain one n_i^*-dilated term from the first n_i^*+d terms of n_i^*X, let's call them

$$n_i^*\cdot y_i < n_i^*\cdot x_i^{(a)} < n_i^*\cdot x_i^{(b)} < n_i^*\cdot x_i^{(c)} < n_i^*\cdot z_i.$$

But then Inequality (3.3) implies that, for more than $\frac{d}{d+1}(n_2-n_1)$ of the positive integers n between n_1 and n_2, the four intervals

$$[\lfloor ny_i\rfloor, \lfloor ny_i\rfloor+1), [\lfloor ny_i\rfloor+1, \lfloor ny_i\rfloor+2), [\lfloor ny_i\rfloor+2, \lfloor nz_i\rfloor), [\lfloor nz_i\rfloor, \lfloor nz_i\rfloor+1)$$

contain the five n-dilated terms

$$ny_i < nx_i^{(a)} < nx_i^{(b)} < nx_i^{(c)} < nz_i.$$

This, by the pigeonhole principle, implies that, for more than $\frac{d}{d+1}(n_2 - n_1)$ of the positive integers n between n_1 and n_2, two of the five terms (which are all from the first $n + d$ terms of the n-dilated sequence nX) are contained in one of the four intervals

$$[\lfloor ny_i \rfloor, \lfloor ny_i \rfloor + 1), [\lfloor ny_i \rfloor + 1, \lfloor ny_i \rfloor + 2), [\lfloor ny_i \rfloor + 2, \lfloor nz_i \rfloor), [\lfloor nz_i \rfloor, \lfloor nz_i \rfloor + 1).$$

□

Corollary 3. *There is an n' between n_1 and n_2 such that, for the $d + 1$ pairs y_i, z_i, one of the four intervals*

$$[\lfloor n'y_i \rfloor, \lfloor n'y_i \rfloor + 1), [\lfloor n'y_i \rfloor + 1, \lfloor n'y_i \rfloor + 2), [\lfloor n'y_i \rfloor + 2, \lfloor n'z_i \rfloor), [\lfloor n'z_i \rfloor, \lfloor n'z_i \rfloor + 1)$$

contains two of the first $n' + d$ terms from the n'-dilated sequence $n'X$.

Proof. Assume that no such n' exists. This implies that, for any single n between n_1 and n_2, there are at most d pairs y_i, z_i such that one of the four intervals

$$[\lfloor ny_i \rfloor, \lfloor ny_i \rfloor + 1), [\lfloor ny_i \rfloor + 1, \lfloor ny_i \rfloor + 2), [\lfloor ny_i \rfloor + 2, \lfloor nz_i \rfloor), [\lfloor nz_i \rfloor, \lfloor nz_i \rfloor + 1)$$

contains two of the first $n + d$ terms from the n-dilated sequence nX. But then this implies that if we sum over all n between n_1 and n_2, then the total amount of times that, for a pair y_i, z_i, one of the four intervals

$$[\lfloor ny_i \rfloor, \lfloor ny_i \rfloor + 1), [\lfloor ny_i \rfloor + 1, \lfloor ny_i \rfloor + 2), [\lfloor ny_i \rfloor + 2, \lfloor nz_i \rfloor), [\lfloor nz_i \rfloor, \lfloor nz_i \rfloor + 1)$$

contains two of the first $n + d$ terms from the n-dilated sequence nX is at most

$$d(n_2 - n_1).$$

But then this implies that, for any single pair y_i, z_i, the average amount of n between n_1 and n_2 for which one of the four intervals

$$[\lfloor ny_i \rfloor, \lfloor ny_i \rfloor + 1), [\lfloor ny_i \rfloor + 1, \lfloor ny_i \rfloor + 2), [\lfloor ny_i \rfloor + 2, \lfloor nz_i \rfloor), [\lfloor nz_i \rfloor, \lfloor nz_i \rfloor + 1)$$

contains two of the first $n + d$ terms from the n-dilated sequence nX is at most

$$\frac{d}{d+1}(n_2 - n_1).$$

This, in turn–since an average cannot be less than all of the numbers from which it is calculated–contradicts Theorem 3's result that, for *strictly more than* $\frac{d}{d+1}(n_2 - n_1)$ of the $(n_2 - n_1)$ positive integers n between n_1 and n_2, one of the four intervals

$$[\lfloor ny_i \rfloor, \lfloor ny_i \rfloor + 1), [\lfloor ny_i \rfloor + 1, \lfloor ny_i \rfloor + 2), [\lfloor ny_i \rfloor + 2, \lfloor nz_i \rfloor), [\lfloor nz_i \rfloor, \lfloor nz_i \rfloor + 1)$$

contains two of the first $n + d$ terms from the n-dilated sequence nX. □

Finally, we put the pieces together to form our upper bound on $s(d)$.

Corollary 4. *For all* $d \geq 1$,

$$s(d) < 24801d^3 + 942d^2 + 3.$$

Proof. Given a sequence X, if there is an n' such that for some $d+1$ pairs $y_i, z_i \subset X$, one of the four intervals

$$[\lfloor n'y_i \rfloor, \lfloor n'y_i \rfloor + 1), [\lfloor n'y_i \rfloor + 1, \lfloor n'y_i \rfloor + 2), [\lfloor n'y_i \rfloor + 2, \lfloor n'z_i \rfloor), [\lfloor n'z_i \rfloor, \lfloor n'z_i \rfloor + 1)$$

contains two of the first $n' + d$ terms from the n'-dilated sequence $n'X$, then, using the terminology of Definition 2, the sequence X has at least $d+1$ irregularities.

For a given d, Theorem 3 along with Corollary 3 tell us that if a sequence X has

$$24801d^3 + 942d^2 + d + 3$$

terms[2], then there must be an n' guaranteeing that the sequence X has at least $d+1$ irregularities. But then this contradicts Theorem 3's assumption that X has as most d irregularities. Thus, again using the terminology of Definition 2, if X is any N-regular sequence with at most d irregularities, then

$$N + d < 24801d^3 + 942d^2 + d + 3.$$

This, using the notation from Definition 3, is equivalent to

$$s(d) < 24801d^3 + 942d^2 + 3.$$

□

3.4. A Few Thoughts on Improving the Upper Bound for $s(d)$

Our, admittedly, technically involved demonstration of an upper bound focuses on the behavior of the terms from our sequence that are quite close to zero. Perhaps there is a technique focusing on terms from a larger portion of the unit-length interval that would yield a lower upper bound for $s(d)$?

Computer-based construction of sequences can yield exact values for $s(d)$. Unfortunately, our efforts in this direction, consisting primarily of exhaustively constructing N-regular sequences with at most d irregularities, have proven too computationally expensive.

[2] $24801d^3 + 942d^2 + d + 3$ comes from adding d to the largest possible value of Theorem 3's n_2.

References

[1] E. R. Berlekamp and R. L. Graham, Irregularities in the distributions of finite sequences, *J. Number Theory* **2** (1970), 152-161.

[2] R. L. Graham, A note on irregularities of distribution, *Integers* **13** (2013), #A53.

[3] T. Oliveira e Silva, A problem related to the 17 point problem of Steinhaus, *mathoverflow.com*, https://mathoverflow.net/q/260116 (Accessed April 20, 2016).

[4] K. F. Roth, On irregularities of distribution, *Mathematika* **1** (1954), 73-79.

[5] H. Steinhaus, *One Hundred Problems in Elementary Mathematics*, Basic Books, New York, 1964.

[6] M. Warmus, A supplementary note on the irregularities of distributions, *J. Number Theory* **8** (1976), 260-263.