GEOMETRIC PROGRESSION-FREE SEQUENCES WITH SMALL GAPS II

XIAOYU HE

Abstract. When $k$ is a constant at least 3, a sequence $S$ of positive integers is called $k$-GP-free if it contains no nontrivial $k$-term geometric progressions. Beiglböck, Bergelson, Hindman and Strauss first studied the existence of a $k$-GP-free sequence with bounded gaps. In a previous paper the author gave a partial answer to this question by constructing a 6-GP-free sequence $S$ with gaps of size $O(\exp(6 \log n / \log \log n))$. We generalize this problem to allow the gap function $k$ to grow to infinity, and ask: for which pairs of functions $(h, k)$ do there exist $k$-GP-free sequences with gaps of size $O(h)$? We show that whenever $(k(n) - 3) \log h(n) \log \log h(n) \geq 4 \log 2 \cdot \log n$ and $h, k$ satisfy mild growth conditions, such a sequence exists.

1. Introduction

Let $S$ be an increasing sequence of positive integers. We say that $S$ is $k$-GP-free if it contains no $k$-term geometric progressions with common ratio not equal to 1, where $k \geq 3$ for the problem to be nontrivial. Let $h$ be a nondecreasing function $\mathbb{N} \to \mathbb{R}^+$. We say that a sequence $S$ has gaps of size $O(h)$ if there exists a constant $C > 0$ such that for every pair $m, N \in \mathbb{N}$ with $m \leq N$, the sequence $S$ intersects the interval $[m, m + Ch(N))$.

The maximal asymptotic density of a $k$-GP-free sequence is well-studied [3, 10, 11, 15]. Beiglböck et al. [2] originally posed the related question:

Problem 1. Does there exist $k \geq 3$ and a $k$-GP-free sequence $S$ such that $S$ has gaps of size $O(1)$?

The standard example of a 3-GP-free sequence is the sequence $Q$ of positive squarefree numbers $1, 2, 3, 5, 6, 7, 10, \ldots$, which has asymptotic density $6/\pi^2$. Despite its large density, the size of its largest gaps is not known. The best unconditional result available is that of Filaseta and Trifonov [5] that $Q$ has gaps of size $O(N^{1/5} \log N)$, and Trifonov also established a generalization that the sequence of $k$-th-power-free numbers has gaps of size $O(N^{1/(2k+1)} \log N)$ [16]. Assuming the abc conjecture, Granville showed that the gaps of $Q$ are of size $O(N^\varepsilon)$ for all $\varepsilon > 0$ [7].

All of these bounds can be improved immensely if we assume the conjecture of Cramér that the gaps between consecutive primes are $O(\log^2 N)$ [4]. For a discussion of Cramér’s model and implications, see the article of Pintz [12]. The problem of bounding largest gaps between consecutive primes, both from above and below, is notoriously difficult, and the best known lower bound is

$$p_{n+1} - p_n \geq \frac{C \log p_n \log \log p_n \log \log \log \log p_n}{\log \log \log p_n}$$

for some $C > 0$ and infinitely many $n$, due to Ford, Green, Konyagin, Maynard, and Tao [6], an improvement by $\log \log \log p_n$ over the longstanding bound of Rankin [14]. The best
unconditional upper bound is \( p_{n+1} - p_n = O(N^{0.525}) \), due to Baker, Harman, and Pintz [1], with \( O(N^{1/2} \log N) \) possible assuming the Riemann hypothesis.

Instead of pursuing these notoriously difficult problems, in a previous paper the author showed that by replacing \( Q \) by a randomly constructed analogue, we can improve on Granville’s bound unconditionally.

**Theorem 2.** [8] There exists a 6-GP-free sequence \( T \) and a constant \( C > 0 \) such that the gaps of \( T \) are of size \( O(\exp(C \log N / \log \log N)) \). In fact \( C \) can be taken to be any positive real greater than \( \frac{5}{6} \log 2 \).

In this paper we generalize the Problem 1 as follows. Henceforth \( k \) is no longer a constant but a nondecreasing function \( k : \mathbb{N} \rightarrow \mathbb{R} \geq 3 \). We say that \( S \) is \( k \)-GP-free if for every \( N \in \mathbb{N} \), the finite subsequence \( S \cap \{1, 2, \ldots, N\} \) does not contain any nontrivial geometric progressions of length at least \( k(N) \).

**Problem 3.** For which pairs of functions \((h, k)\) do there exist \( k \)-GP-free sequences \( S \) such that \( S \) has gaps of size \( O(h) \)?

We call \( h \) the gap function and \( k \) the length function, and a pair \((h, k)\) feasible if such an \( S \) exists. Thus far we have only dealt with constant length function; in particular Theorem 2 shows that the pair \((\exp(C \log N / \log \log N), 6)\) is feasible. At the other end of the spectrum, it is trivial that \((1, \log N / \log 2)\) is a feasible pair, simply because the longest possible geometric progression in \( 1, \ldots, N \) has length at most \( \log N / \log 2 \). In the last section of this paper we show in fact that \((1, \varepsilon \log N)\) is feasible for any \( \varepsilon > 0 \).

To interpolate between these two situations, we prove the following theorem, extending the method used in [8] to prove Theorem 2.

For two functions \( f, g : \mathbb{N} \rightarrow \mathbb{R}^+ \) we write \( f = O(g) \) if there exists a constant \( C > 0 \) such that \( f(n) \leq Cg(n) \) for all \( n \in \mathbb{N} \) and \( f = o(g) \) if for every \( C > 0 \) the inequality \( f(n) \leq Cg(n) \) holds for all \( n \) sufficiently large. We also write \( f = \Omega(g) \) if \( g = O(f) \).

**Theorem 4.** Let \((h, k)\) be nondecreasing functions \( \mathbb{N} \rightarrow \mathbb{R}^+ \) such that \( h(n) = \Omega((\log x)^{1/(1-\log 2)}) \) and for all sufficiently large \( n \), \( k(n) > 5 \). If they satisfy

\[
(k(n) - 3) \log h(n) \log \log h(n) \geq 4 \log 2 \cdot \log n,
\]

for all sufficiently large \( n \), then there exists a \( k \)-GP-free sequence \( T \) with gaps of size \( O(h) \).

As a corollary, if \( k \) is constant we recover Theorem 2 with a weaker constant.

2. Preliminaries

In this section we generalize the GP-free process of [8] to probabilistically construct a \( k \)-GP-free sequence. First we simplify Theorem 4 by reducing the set of possible length functions \( k \). It suffices to show the following.

**Theorem 5.** If \( k \) is a nondecreasing function \( \mathbb{N} \rightarrow \{6, 8, \ldots\} \) taking on even positive integer values at least 6, and \( h : \mathbb{N} \rightarrow \mathbb{R}^+ \) is a function satisfying \( h(n) = \Omega((\log x)^{1/(1-\log 2)}) \), \( h(n) = o(\sqrt{n}) \) and

\[
(k(n) - 2) \log h(n) \log \log h(n) \geq 4 \log 2 \cdot \log n,
\]

for all \( n \) sufficiently large, then there exists a \( k \)-GP-free sequence \( T \) with gaps of size \( O(h) \).
Proof. (that Theorem 5 implies Theorem 4). Suppose Theorem 5 is true, and let k be as in Theorem 4. We can certainly round up k to the nearest integer to begin with. It is also possible to ignore the finite set of n for which k ≤ 5, since we only care about n sufficiently large. If we round k down to the nearest even integer, if it originally satisfied the inequality of Theorem 4, then it has decreased by at most 1 uniformly, so the inequality above holds. Finally, if we prove the theorem for all h(n) = o(√n), then it follows for all larger h as well, so we may as well assume h(n) = o(√n).

Let G_k be the family of all geometric progressions of positive integers such that if t is the largest term, then the length is at least k(t). Enumerate them as G_{k,i} in order lexicographically as sequences of positive integers. We assume that each G_{k,i} has common ratio r_{k,i} > 1.

Furthermore, there may be longer G_{k,i} containing shorter ones; let G_{k}^* denote the result of removing from G_k all G_{k,i} which contain some G_{k,j} with j ≠ i. Thus to find a k-GP-free sequence it suffices to construct a sequence T_k missing at least one element from each progression in G_{k}^*. Let G_{k,i}^* denote the i-th progression in G_{k}^*.

Definition 6. For a nondecreasing function k : N → {6, 8, . . .}, define the k-GP-free process as follows. Define an integer-sequence valued random variable U_k = (u_1, u_2, . . .) where u_i ∈ G_{k,i}^* such that if

\[ G_{k,i}^* = (a_i b_i^{k-1}, a_i b_i^{k-2} c_i, \ldots, a_i c_i^{k-1}), \]

then u_i is chosen from \(a_i b_i^{k/2-1} c_i^{k/2}\) and \(a_i b_i^{k/2} c_i^{k/2}\) with equal probability \(\frac{1}{2}\). Each u_i is picked independently of the others. Then T_k is the random variable whose value is the sequence of all positive integers never appearing in U_k, sorted in increasing order.

It is clear that T_k is k-GP-free by definition, as it misses at least one term out of each G_{k,i}. We now bound the probability that a given n ∈ N lies in T_k generated as above. For i, j ≥ 1, let d(n; i, j) count the number of ways to factorize n = ab^c for some a, b, c ∈ N.

Lemma 7. For a positive integer n, the sequence T_k constructed in Definition 6 contains n with probability

\[ \mathbb{P}[T_k \ni n] \geq 2^{-d(n; k(m)/2, k(m)/2 - 1)}, \]

where m is any positive integer such that any G_{k,i}^* containing n in its middle two terms has largest term at least m.

Proof. The inequality is equivalent to the statement that n is one of the middle two terms in at most d(n; k(m)/2, k(m)/2 - 1) progressions of G_{k}^*. We form an injective correspondence from progression G_{k,i}^* containing n in the middle two terms to factorizations of n as n = ab^{k(m)/2} c^{k(m)/2 - 1}. If a progression

\[ G_{k,i}^* = (a_i b_i^{k'-1}, a_i b_i^{k'-2} c_i, \ldots, a_i c_i^{k'-1}) \]

with \(b_i < c_i\) and \(k' \geq k(a_i c_i^{k'-1})\) contains n as one of the middle two terms, then certainly \(k(m) \leq k'\). Supposing \(n = a_i b_i^{k'/2-1} c_i^{k'/2}\), we map G_{k,i}^* to the factorization n = ab^{k(m)/2} c^{k(m)/2 - 1} with a = a_i b_i^{(k'-k(m))/2} c_i^{(k'-k(m))/2}, b = c_i and c = b_i. Similarly if n = a_i b_i^{k'/2} c_i^{k'/2 - 1} we take a = a_i b_i^{(k'-k(m))/2} c_i^{(k'-k(m))/2}, b = b_i and c = c_i. It is easy to see from the assumptions that \(b_i < c_i\) and that no progression in G_{k}^* strictly contains another that the correspondence above is injective, as desired.

\[ \square \]
From here we can control the total probability that \( T_k \) misses an entire interval of the form \([x, x + Ch(x)]\).

Lemma 8. For a gap function \( h(x) = o\left(x^{1-1/(k(x)-1)}\right) \) and a constant \( C > 0 \), the sequence \( T_k \) constructed in Definition 6 satisfies \( T_k \cap [x, x + Ch(x)] = \emptyset \) with probability

\[
\mathbb{P}[T_k \cap [x, x + Ch(x)] = \emptyset] \leq \exp\left(- \sum_{n \in [x, x + Ch(x)]} \exp\left(- \log 2 \cdot d\left(n; \frac{k(x)}{2}, \frac{k(x)}{2} - 1\right)\right)\right)
\]

for all \( x \) sufficiently large.

Proof. We first prove that the events \( \mathbb{P}[T_k \ni n] \) for \( n \in [x, x + Ch(x)] \) are mutually independent whenever \( x \) is sufficiently large. It suffices to show that no progression in \( G^*_k \) has both middle terms in the interval. Considering the difference between the two middle terms in a \( G^*_{k,i} \), and assuming both lie inside \([x, x + Ch(x)]\), we have

\[
\left|a_i b_i^{k/2-1} c_i^{k/2} - a_i b_i^{k/2} c_i^{k/2-1}\right| \geq \frac{x}{b_i} \geq x^{1-1/(k(m)-1)} \geq x^{1-1/(k(x)-1)}
\]

where \( k \geq k(m) \) depends on the largest term \( m = a_i c_i^{k-1} > x \). It follows that assuming \( h(x) = o\left(x^{1-1/(k(x)-1)}\right) \), for any \( C > 0 \) the middle two terms in any \( G^*_{k,i} \) with largest term at most \( x \) are further apart than \( Ch(x) \) for any \( x \) sufficiently large.

Thus the events corresponding to each \( n \) in the interval are mutually independent, and we can bound the probability involved by a product

\[
\mathbb{P}[T_k \cap [x, x + Ch(x)] = \emptyset] \leq \prod_{n \in [x, x + Ch(x)]} \left(1 - 2^{-d(n; k(m)/2, k(m)/2-1)}\right),
\]

by Lemma 7. Since the inequality \( 1 - t \leq e^{-t} \) holds for all real \( t \) we arrive at the bound

\[
\mathbb{P}[T_k \cap [x, x + Ch(x)] = \emptyset] \leq \exp\left(- \sum_{n \in [x, x + Ch(x)]} \exp\left(- \log 2 \cdot d\left(n; \frac{k(m)}{2}, \frac{k(m)}{2} - 1\right)\right)\right).
\]

Here each \( m = m(n) \) can certainly be chosen as any number at most \( n \). Thus we replace them all by \( x \), arriving at the desired bound. \( \square \)

Note that since we assumed \( h(x) = o(\sqrt{x}) \) the growth condition in Lemma 8 is automatically satisfied.

3. Proof of the Main Theorem

All that remains is to give lower bounds for the sum

\[
S(x, h, k, C) = \sum_{n \in [x, x + Ch]} \exp\left(- \log 2 \cdot d\left(n; \frac{k}{2}, \frac{k}{2} - 1\right)\right),
\]

where \( k = k(x) \) and \( h = h(x) \) are functions satisfying the conditions of Theorem 5. To this end we break down \([x, x + Ch]\) into two sets, one of which has few \((k/2 - 1)\)-power divisors, and restrict the sum to that set.
Lemma 9. There is a positive constant $B$ independent of $x$ such that for all sufficiently large $x$, 

$$S(x, h, k, C) \geq BCh(x) \exp \left( -\log 2 \exp \left( \frac{4 \log 2 \cdot \log x}{(k(x) - 2) \log h(x)} \right) \right).$$

Proof. Fix an $x > 0$ and write $k = k(x), h = h(x)$. Denote by $A$ the subset of $[x, x + Ch]$ consisting of all $n$ divisible by some $p^{k/2-1}$, where $p \leq h$. We can bound the size of $A$ by

$$|A| \leq \sum_{\text{prime } p \leq h} \left( \frac{Ch}{p^{k/2-1}} + 1 \right) \leq (\zeta(k/2 - 1) - 1)Ch + o(h),$$

where $\zeta$ is the Riemann zeta function and we used the elementary Chebyshev bound $\pi(h) = o(h)$ on the prime-counting function $\pi$. Since $k \geq 6$ and $\zeta(t) - 1 < 1$ uniformly on $t \geq 2$, there exists a constant $B$ such that for $x$, and thus $h$, sufficiently large, $|A| \leq (1 - B)Ch$.

If $n \not\in A$, we can factor $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}n'$ where $n'$ is $(k/2 - 1)$-th power free, each $\alpha_i \geq k/2 - 1$, and each $p_i \geq h$ is prime. As a result,

$$\sum_i \alpha_i \leq \frac{\log n}{\log h},$$

so by a smoothing argument we can bound $d(n; k^{k/2-1})$ subject to these assumptions,

$$d(n; k^{k/2-1}) \leq \exp \left( \log 2 \cdot \frac{\log n}{(k/2 - 1) \log h} + \log 2 \cdot \frac{\log n}{(k/2) \log h} \right),$$

where we simply bounded the number of pairs $b, c$ satisfying $b^{k/2-1}|n$ and $c^{k/2}|n$. Summing up over all terms in $[x, x + Ch]$ outside $A$, we get

$$S(x, h, k, C) \geq BCh \exp \left( -\log 2 \exp \left( \left( \frac{1}{k} + \frac{1}{k - 2} \right) \frac{(2 \log 2) \cdot \log x}{\log h} \right) \right),$$

and finally replacing $1/k \leq 1/(k - 2)$ we have the desired inequality. \hfill $\square$

Finally, we prove Theorem 5 using Lemma 9.

Proof. (of Theorem 5). By Lemma 8 it suffices to pick $h, k$ such that the sum of probabilities

$$\sum_{x \geq 1} \mathbb{P}[T_k \cap [x, x + Ch(x)) = \emptyset] \leq \sum_{x \geq 1} \exp(-S(x, h, k, C)) < 1$$

for $C$ sufficiently large, forcing the probability of finding a $T$ with gaps $O(h)$ to be nonzero. This will hold as long as the sum converges for some fixed $C$; making $C$ large enough will make the sum arbitrarily small. Now, suppose that $(k - 2) \log h \log \log h \geq 4 \log 2 \cdot \log n$ as in Theorem 5. Then, applying the inequality of Lemma 9, we have

$$S(x, h, k, C) \geq BCh \exp(-\log 2 \log h) \geq BCh^{1 - \log 2},$$

and finally since $h = \Omega((\log x)^{1/(1 - \log 2)})$, we get

$$\sum_{x \geq 1} \exp(-S(x, h, k, C)) \leq \sum_{x \geq 1} x^{-BCh D},$$

for some constant $D > 0$, so picking $C$ for which $BC > 1$ gives a convergent sum. \hfill $\square$
4. Closing Remarks

The goal of this paper was to interpolate smoothly between the two feasible pairs \((h, k) = (\exp(C \log N/ \log \log N), 6)\) and \((h, k) = (1, \log N/ \log 2)\), and we recover both pairs, up to constants, in the relation

\[(k(n) - 3) \log h(n) \log \log h(n) \geq 4 \log 2 \cdot \log n.\]

Unfortunately, when \(k\) is sufficiently close to \(\log n\), then the method of Theorem 4 fails because \(h = o((\log x)^{1/(1-\log 2)})\). Nevertheless, we expect all pairs \((h, k)\) which satisfy this inequality to be feasible. In the case that \(h = 1\) we can make an improvement on \((1, \log N/ \log 2)\).

**Proposition 10.** For any \(\varepsilon > 0\), if \(k(n) = \varepsilon \log n\) then there exists a \(k\)-GP-free sequence \(T\) with gaps of size \(O(1)\).

**Proof.** We say a positive integer \(m\) is divisible by a \(k\)-th power if \(p^{\lceil k(m) \rceil} \mid m\) for some prime \(p\), and that \(m\) is \(k\)-free otherwise. Consider the sequence \(T\) of all \(k\)-free integers; we claim that its gaps are uniformly bounded. In fact, note that if \(p^{\lceil k(m) \rceil} \mid m\) then

\[
p^{k(m)} \leq m, \\
\varepsilon \log m \cdot \log p \leq \log m, \\
\log p \leq \frac{1}{\varepsilon},
\]

and so \(p\) lies in the finite set of all primes less than \(e^{1/\varepsilon}\). In particular, for \(x\) sufficiently large, the interval \([x, x + e^{1/\varepsilon} + 1]\) will contain at least one \(k\)-free number. Indeed, it is easy to check that each \(p \leq e^{1/\varepsilon}\) contributes at most one multiple of \(p^{k(x)}\) to that interval. \(\square\)

Further improvement in the case of \(h\) small or constant along these lines is blocked by the Chinese Remainder Theorem. In particular, for \(k = o(\log n)\) and any constant \(h\) we can find infinitely many intervals \([x, x + h]\) in which each positive integer in \([x, x + h]\) is divisible by arbitrarily many \(k(x)\)-th powers of primes.

The probabilistic method in Definition 6 is by no means optimal, but is defined in such a way to guarantee the independence of events in an interval \([n, n + Ch]\). We expect that a sophisticated study of redundancies in our method can substantially improve at least the constant in Theorem 4.

5. Acknowledgements

I would like to thank Levent Alpoge and Joe Gallian for correcting many mistakes.

References

[1] R. C. Baker, G. Harman, J. Pintz, *The difference between consecutive primes, II*, Proc. London Math. Soc. 3 (2001) 83, 532–562.
[2] M. Beiglböck, V. Bergelson, N. Hindman and D. Strauss, *Multiplicative structures in additively large sets*, J. Combin. Theory Ser. A (2006) 13-7, 1219-1242.
[3] B. E. Brown and D. M. Gordon, *On sequences without geometric progressions*, Math. Comp. (1996) 65, no. 216, 1749-1754.
[4] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Mathematica (1936) 2, 23–46.
[5] M. Filaseta and O. Trifonov, *On gaps between squarefree numbers II*, J. London Math. Soc. 2 (1992) 45, 215–221.
[6] K. Ford, et al. Long gaps between primes. preprint (2014), arXiv:1412.5029.

[7] A. Granville. ABC Allows Us to Count Squarefrees. International Mathematics Research Notices (1998) 991–1009.

[8] X. He. Geometric progression-free sequences with small gaps. J. Number Theory 151 (2015), 197-210 arXiv:1501.04121.

[9] H. Maier, Primes in short intervals, Michigan Math. J. (1985) 32, 221-225.

[10] N. McNew, On sets of integers which contain no three terms in geometric progression, preprint (2014), arXiv:1310.2277.

[11] M. B. Nathanson and K. O'Bryant, Irrational numbers associated to sequences without geometric progressions, preprint (2013), arXiv:1307.8135.

[12] J. Pintz, Cramér vs. Cramér. On Cramér’s probabilistic model for primes, Funct. Approx. Comment. Math. (2007) 37-2, 361-376.

[13] R. A. Rankin, The difference between consecutive prime numbers, Proc. Edinburgh Math. Soc., (1962-1963) 13, 331-332.

[14] J. Riddell, Sets of integers containing no n terms in geometric progression, Glasgow Math. J. (1969) 10, 137-146.

[15] O. Trifonov, On gaps between k-free numbers, J. Number Theory (1995) 55, 46-59.