

# Bounded gaps and perfect power gaps in sequences of consecutive primes

Katalin Gyarmati

## Abstract

We study whether several consecutive prime gaps can all be relatively large at the same time, or is it possible that all are squares or perfect powers, or perhaps none of them are squares? A few related results and problems are also presented.

## 1 Introduction

It has long been conjectured that there are infinitely many twin primes. Building on classical sieving techniques, significant recent progress has been made toward proving this conjecture by Goldston, Pintz, and Yıldırım [13, 14, 15], Zhang [37], and Maynard [23]. The most notable result was proved within the framework of the Polymath Project [29], [30] about ten years ago. Namely, they proved that

$$p_{n+1} - p_n \leq 246 \tag{1}$$

holds infinitely many times, where  $p_n$  denotes the  $n$ -th prime number.

---

2020 Mathematics Subject Classification: Primary: 11N05.

Keywords and phrases: prime gaps, squares and powers.

Research supported by the Hungarian National Research Development and Innovation Funds KKP133819.

It is also interesting to study the maximum value of  $p_{n+1} - p_n$  for primes less than  $x$ . This problem was studied, for example, by Westzynthius [38], Erdős [8], and Rankin [31]. In 1997, Erdős [9] offered a reward of \$10,000 to anyone who could improve on the previous best result, but this was only achieved in 2016 by Ford, Green, Konyagin, Tao [10], and Maynard [24], [25] (also see the joint paper [11]). Here, they proved that

$$p_{n+1} - p_n > \frac{c \log n \log \log n \log \log \log n}{\log \log \log n}$$

holds for infinitely many integers  $n$ , where  $c > 0$  is an absolute constant. Consider the following problem: is it possible to find many large consecutive prime gaps? Shiu [35], Cramér [6], Maier [22], and Pintz [27] also worked on this problem. Finally, Ford, Maynard, and Tao [12] proved the sharpest theorem:

**Theorem A.** [Ford, Maynard, Tao]. *Let  $k \geq 1$  be an integer and*

$$G_k(x) \stackrel{\text{def}}{=} \max_{p_{n+k} \leq x} \min\{p_{n+1} - p_n, p_{n+2} - p_{n+1}, \dots, p_{n+k} - p_{n+k-1}\}.$$

*Then*

$$G_k(x) \gg \frac{1}{k^2} \frac{\log x \log \log x \log \log \log x}{\log \log x}, \quad (2)$$

*where the implied constant is absolute and effective.*

The following result, Theorem 1, provides a lower bound for  $G_k(x)$  that, while staying within the average spacing of  $\log x$ , offers important insights for larger values of  $k$ . For the case  $k = 1$ , the result is essentially a consequence of the Prime Number Theorem and could be considered trivial. However, for  $k \geq 2$ , the estimate is derived using specific bounds from sieve theory. This result sharpens Theorem A only for large  $k$ .

**Theorem 1.** *There exists a positive constant  $x_0$  such that if  $x_0 < x \in \mathbb{R}$  and  $2 \leq k \in \mathbb{N}$ , then*

$$G_k(x) > 0.1504 \frac{\log x}{k}. \quad (3)$$

If  $k$  is large in terms of  $x$ , i.e.,

$$k \gg \frac{\log \log x \log \log \log x}{\log \log \log x},$$

then the estimate (3) is sharper than the estimate (2).

Today, according to (1), we know that there are infinitely many consecutive primes whose difference is  $\leq 246$ . Related to this, I ask the following problem:

**Problem 1.** *Is it possible to prove the existence of an absolute constant  $C$  for which there are infinitely many integers  $n$  with*

$$p_{n+1} - p_n < C$$

and

$$p_{n+1} - p_n$$

*is a square? Is a similar statement true for squarefree numbers in place of squares? Is such a statement true if we require the difference  $p_{n+1} - p_n$  to be twice a prime (or a prime itself) instead of a square?*

I conjecture that the above problem is true for  $C = 246^2$ , but I have not been able to prove this. However, a slightly weaker statement follows immediately from Pintz, Steiger, and Szemerédi theorem [28] or Bloom and Maynard's theorem [3].

**Theorem 2.** *There are infinitely many prime pairs  $p, q$  for which  $p - q$  is a square.*

Problem 1 is related to the famous Dickson conjecture (see Conjecture 1 later). Surprisingly, if we replace the squares in Problem 1 with perfect powers, then the statement can be easily proved, following the Banks, Freiberg, and Turnage-Butterbaugh's theorem [2].

**Theorem 3.** *There are infinitely many integers  $n$  for which*

$$p_{n+1} - p_n < e^{1.834 \cdot 10^{76}}$$

*and  $p_{n+1} - p_n$  is a perfect power.*

The method of the proof of Theorem 3 can also be used to prove the following more general statement:

**Theorem 4.** *For every integer  $m \geq 1$ , there are infinitely many integers  $n$  for which*

$$p_{n+m} - p_{n+1} < e^{e^{10.6m}}$$

*and the difference between any two of the consecutive primes  $p_{n+1}, p_{n+2}, \dots, p_{n+m}$  is a perfect power.*

**Problem 2.** *The upper bounds in Theorems 3 and 4 are clearly not optimal. It might be interesting to sharpen the constants in these theorems significantly.*

In addition to the cases concerning squares and powers, it is also interesting to study sets of primes for which the difference between two primes is never a square. Then, using the results of Ruzsa [34] and Lewko [21], I will prove the following:

**Theorem 5.** *Let  $\gamma = \frac{1}{2} + \frac{\log 12}{2 \log 205} = 0.7334117\dots$ . If  $x \geq 2$ , then in the set  $\{1, 2, \dots, x\}$  there exist at least  $\frac{1}{205}x^\gamma / \log x$  primes such that the difference between any two primes is never a square.*

We may also study the case when we require that the primes in Theorem 5 should be consecutive primes. In this case, I have been able to prove the existence of a much smaller set of primes with the desired property.

**Theorem 6.** *There exists a positive constant  $x_0$  such that if  $x \geq x_0$ , then in the set  $\{1, 2, \dots, x\}$  there exist  $k = \lceil 0.24(\log x)^{1/4} \rceil$  consecutive primes such that the difference between any two of these primes is never a square.*

**Problem 3.** *Is Theorem 6 true for  $k = \lfloor x^c \rfloor$  consecutive primes, where  $c > 0$  is an absolute constant?*

The structure of this paper is as follows. In Section 2, we provide the proof of Theorem 1 and discuss the underlying sieve methods. Section 3 is devoted to the proof of Theorems 3 and 4, while the proofs of the remaining two theorems are detailed in Sections 4 and 5. To achieve these results, we utilize a combination of prime number theory and combinatorial methods. All numerical computations throughout the proofs were performed using the SAGE programming language [36]; the corresponding source codes are provided in the Appendix.

## 2 Proof of Theorem 1.

Throughout the proof, we may assume that  $k \leq 0.0752 \log x$ ; otherwise, the theorem is trivial. Let

$$r = 0.1504 \frac{\log x}{k} \geq 2.$$

Then,

$$rk = 0.1504 \log x.$$

Assume that  $x$  is sufficiently large ( $x > x_0$ ). In the interval  $[1, x]$ , we indicate the primes with two colors, red and green. The color of a prime  $1 < p \leq x$  is red if there exists another prime  $q > p$  for which  $q - p \leq r$ . The color of a prime  $1 < p \leq x$  is green if there does not exist such a prime  $q$ , i.e., for all  $q > p$  primes, we have

$$q - p > r.$$

Let  $R$  denote the number of red primes and  $G$  denote the number of green primes in the interval  $[1, x]$ . The main idea of the proof is to find many consecutive green primes in the interval  $[1, x]$ . To do this, we will prove the following:

**Lemma 1.** *There exists a constant  $x_0$  such that for all  $x > x_0$ , the number of red primes in  $[1, x]$  satisfies*

$$R \leq 6.646 \frac{rx}{(\log x)^2}.$$

Before proving the lemma, let us see how the theorem follows from it. In the interval  $[1, x)$ , let us call a subinterval  $[a, b)$  red if all the primes it contains are red. Similarly, let us call a subinterval  $[a, b)$  green if all the primes it contains are green. The interval  $[1, x)$  can be partitioned into the union of disjoint red and green intervals. Since these intervals alternate in colour (and we may assume that  $[1, x)$  starts with a red interval), it follows that a red interval is followed by a green one, and a green interval is followed by a red one. In this partition, the number of green intervals is less than or equal to the number of red intervals, which in turn is less than or equal to the total number of red primes. That is,

$$\#(\text{red intervals}) \leq R \leq 6.646 \frac{rx}{(\log x)^2}.$$

By Rosser and Schoenfeld's result [33, Theorem 9]

$$\pi(x) > \frac{x}{\log x}, \quad \text{if } x \geq 17. \tag{4}$$

Thus, for the number of green primes, we have

$$G \geq \frac{x}{\log x} - R \geq \frac{x}{\log x} - 6.646 \frac{rx}{(\log x)^2}.$$

That is, in the above division of  $[1, x]$  into disjoint intervals, we have at most  $6.646 \frac{rx}{(\log x)^2}$  green intervals, which contain more than  $\frac{x}{\log x} - 6.646 \frac{rx}{(\log x)^2}$  green primes. Then, by the pigeon-hole principle, there exists a green interval that contains at least

$$\frac{\frac{x}{\log x} - 6.646 \frac{rx}{(\log x)^2}}{6.646 \frac{rx}{(\log x)^2}}$$

green primes, which is

$$\begin{aligned}
&> 0.1504 \frac{\log x}{r} - 1 \\
&= 0.1504 \frac{\log x}{0.1504 \log x/k} - 1 \\
&= k - 1.
\end{aligned}$$

Denote the first  $k$  of these green primes by  $p_n, p_{n+1}, \dots, p_{n+k-1}$ . Since these primes are consecutive primes in a green interval, the distance between two adjacent ones is  $> r = 0.1504 \frac{\log x}{k}$ , from which Theorem 1 follows. All that remains is to prove Lemma 1.

**The proof of Lemma 1.** A classical upper bound for the number of prime pairs with a given difference  $h$  can be found in the work of Halberstam and Richert [18, Theorem 3.11]. While more recent developments (see e.g. [19, 20]) provide sharper constants or fully explicit versions of this inequality, and explicit bounds holding for all  $x > 1$  are also available in the literature (see e.g. Riesel and Vaughan [32, Lemma 5]), the following asymptotic form is sufficient for our current purposes.

**Lemma 2** ([18, Theorem 3.11]). *If  $N \rightarrow \infty$  and  $2 \mid h$ , then*

$$\begin{aligned}
&|\{p : p \leq N, p + h = p' \text{ and } p, p' \text{ are primes}\}| \\
&\leq 8 \prod_{2 < p} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{2 < p|h} \frac{p-1}{p-2} \cdot \frac{N}{(\log N)^2} \left(1 + O\left(\frac{\log \log N}{\log N}\right)\right),
\end{aligned}$$

where the implied constant factor used in the  $O$ -term is absolute and does not depend on  $h$ .

In the following, we denote the integer part of  $r$  by  $[r]$ . If  $h$  is odd, then  $|\{p : p \leq N, p + h = p' \text{ and } p, p' \text{ are primes}\}| \leq 1$ , since by parity, we know that one of the primes  $p$  and  $p'$  is even, so it is 2. Using this and Lemma 2,

we find that the number of red primes is

$$\begin{aligned}
R &\leq \sum_{\substack{h=2 \\ 2|h}}^{[r]} 8 \prod_{2 < p} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{2 < p|h} \frac{p-1}{p-2} \cdot \frac{x}{(\log x)^2} \left(1 + O\left(\frac{\log \log x}{\log x}\right)\right) \\
&+ \sum_{\substack{h=1 \\ h \text{ is odd}}}^{[r]} 1.
\end{aligned} \tag{5}$$

In the above estimate,  $O$  is independent of  $h$  and  $\sum_{\substack{h=1 \\ h \text{ is odd}}}^{[r]} 1 \leq [r] < \log x$ , thus, if  $x$  is sufficiently large, then

$$R < 8.0001 \sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{2 < p} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{2 < p|h} \frac{p-1}{p-2} \cdot \frac{x}{(\log x)^2}.$$

A numerical computation using SAGE programming language (Appendix, Program 1) shows that

$$\prod_{2 < p} \left(1 - \frac{1}{(p-1)^2}\right) < \prod_{2 < p < 5000} \left(1 - \frac{1}{(p-1)^2}\right) < 0.6602.$$

These Euler products can be computed to high precision; for an extensive survey and numerical values, see e.g. Moree [26]. Thus,

$$R < 5.2817 \sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{2 < p|h} \frac{p-1}{p-2} \cdot \frac{x}{(\log x)^2}. \tag{6}$$

So, we need to give an upper bound for

$$\sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{2 < p|h} \frac{p-1}{p-2}.$$

Here,

$$\begin{aligned}
\frac{p-1}{p-2} &= \left(1 + \frac{1}{p}\right) \cdot \frac{(p-1)p}{(p-2)(p+1)} \\
&= \left(1 + \frac{1}{p}\right) \left(1 + \frac{2}{(p-2)(p+1)}\right) \\
&< \left(1 + \frac{1}{p}\right) e^{2/((p-2)(p+1))}.
\end{aligned}$$

Thus,

$$\begin{aligned} \sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{2 < p|h} \frac{p-1}{p-2} &< \sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{2 < p|h} \left(1 + \frac{1}{p}\right) e^{2/((p-2)(p+1))} \\ &= \sum_{\substack{h=2 \\ 2|h}}^{[r]} e^{\sum_{2 < p|h} 2/((p-2)(p+1))} \prod_{2 < p|h} \left(1 + \frac{1}{p}\right). \end{aligned} \quad (7)$$

Let  $T = 20000$ , then

$$\begin{aligned} \sum_{2 < p|h} 2/((p-2)(p+1)) &< \sum_{2 < p} 2/((p-2)(p+1)) \\ &= \sum_{2 < p \leq T+1} 2/((p-2)(p+1)) + \sum_{T+2 \leq p} 2/((p-2)(p+1)). \end{aligned}$$

Using the SAGE programming language (Appendix, Program 2), we get

$$\sum_{2 < p \leq T+1} 2/((p-2)(p+1)) < 0.7271.$$

Furthermore,

$$\begin{aligned} \sum_{T+2 \leq p} 2/((p-2)(p+1)) &< \sum_{n=T+2}^{\infty} 2/((n-2)(n+1)) \\ &< \sum_{n=T+2}^{\infty} 2/((n-2)(n-1)) \\ &< \sum_{n=T+2}^{\infty} \left(\frac{2}{n-2} - \frac{2}{n-1}\right) \\ &= \frac{2}{T} = 0.0001. \end{aligned}$$

So,

$$\sum_{2 < p|h} 2/((p-2)(p+1)) < 0.7272.$$

Thus,

$$e^{\sum_{2 < p|h} 2/((p-2)(p+1))} < e^{0.7272} < 2.0693.$$

Writing this in inequality (7), we get

$$\begin{aligned}
\sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{2 < p|h} \frac{p-1}{p-2} &< 2.0693 \sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{2 < p|h} \left(1 + \frac{1}{p}\right) \\
&= 2.0693 \sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{p|h} \left(1 + \frac{1}{p}\right) / \left(1 + \frac{1}{2}\right) \\
&< 1.3796 \sum_{\substack{h=2 \\ 2|h}}^{[r]} \prod_{p|h} \left(1 + \frac{1}{p}\right) \\
&= 1.3796 \sum_{\substack{h=2 \\ 2|h}}^{[r]} \sum_{\substack{d|h \\ |\mu(d)|=1}} \frac{1}{d} \\
&= 1.3796 \sum_{\substack{d=1 \\ |\mu(d)|=1}}^{[r]} \frac{1}{d} \sum_{\substack{h=1 \\ 2|h \\ d|h}}^{[r]} 1 \\
&= 1.3796 \sum_{\substack{d=1 \\ |\mu(d)|=1 \\ d \text{ even}}}^{[r]} \frac{1}{d} \sum_{\substack{h=1 \\ d|h}}^{[r]} 1 + 1.3796 \sum_{\substack{d=1 \\ |\mu(d)|=1 \\ d \text{ odd}}}^{[r]} \frac{1}{d} \sum_{\substack{h=1 \\ 2d|h}}^{[r]} 1 \\
&\leq 1.3796 \sum_{\substack{d=1 \\ |\mu(d)|=1 \\ d \text{ even}}}^{[r]} \frac{r}{d^2} + 1.3796 \sum_{\substack{d=1 \\ |\mu(d)|=1 \\ d \text{ odd}}}^{[r]} \frac{r}{2d^2} \\
&< 1.3796r \sum_{\substack{d=1 \\ |\mu(d)|=1 \\ d \text{ even}}}^{\infty} \frac{1}{d^2} + 1.3796r \sum_{\substack{d=1 \\ |\mu(d)|=1 \\ d \text{ odd}}}^{\infty} \frac{1}{2d^2} \tag{8}
\end{aligned}$$

Let  $T = 5000$ . Using the SAGE programming language (Appendix Programs 3 and 4) we get

$$\sum_{\substack{d=2 \\ |\mu(d)|=1 \\ d \text{ even}}}^T \frac{1}{d^2} = \sum_{\substack{d=1 \\ d \text{ even}}}^T \frac{\mu^2(d)}{d^2} < 0.304$$

$$\sum_{\substack{d=1 \\ |\mu(d)|=1 \\ d \text{ odd}}}^{T+1} \frac{1}{2d^2} = \sum_{\substack{d=1 \\ d \text{ odd}}}^{T+1} \frac{\mu^2(d)}{d^2} < 0.6079.$$

Moreover,

$$\begin{aligned} \sum_{\substack{d=T+2 \\ |\mu(d)|=1 \\ d \text{ even}}}^{\infty} \frac{1}{d^2} &< \sum_{\substack{d=T+2 \\ d \text{ even}}}^{\infty} \frac{1}{d(d-2)} < \sum_{\substack{d=T+2 \\ d \text{ even}}}^{\infty} \frac{1}{2} \left( \frac{1}{d-2} - \frac{1}{d} \right) \\ &= \frac{1}{2T} = 0.0001 \end{aligned}$$

and

$$\begin{aligned} \sum_{\substack{d=T+3 \\ |\mu(d)|=1 \\ d \text{ odd}}}^{\infty} \frac{1}{2d^2} &< \sum_{\substack{d=T+3 \\ d \text{ odd}}}^{\infty} \frac{1}{2d(d-2)} < \sum_{\substack{d=T+3 \\ d \text{ odd}}}^{\infty} \frac{1}{4} \left( \frac{1}{d-2} - \frac{1}{d} \right) \\ &< \frac{1}{4T} = 0.00005. \end{aligned}$$

Thus, by (8) we get

$$\sum_{\substack{h=2 \\ 2|h}}^{\lfloor r \rfloor} \prod_{2 < p|h} \frac{p-1}{p-2} < 1.3796r (0.3041 + 0.60795) < 1.2583r.$$

Writing this in (6), we have

$$R < 6.646 \frac{rx}{(\log x)^2},$$

which proves Lemma 1. This completes the proof of Theorem 1.  $\square$

**Remark.** In Theorem 1, the constant factor could be further sharpened by employing more sophisticated sieve estimates. For instance, one could use the classical result of Chen [5] or the more recent refinements by Johnston [19] and Lichtman [20]. However, even with these high-precision improvements, the constant 0.1504 in Theorem 1 would improve only by a few thousandths, which does not qualitatively change the nature of the result.

### 3 Proof of Theorem 2.

We can prove that for every integer  $M$ , there exist two distinct primes greater than  $M$  ( $p$  and  $q$ ), such that  $p - q$  is a square number. Throughout the proof, we fix  $M$ . Next, we will use the theorem of Bloom and Maynard [3, Theorem 1], who proved the following:

**Lemma 3** (Bloom-Maynard). *There exists positive constants  $N_0, c_0$  and  $c_1$  such that if  $N \geq N_0$ ,  $\mathcal{A} \subseteq \{1, 2, \dots, N\}$  and*

$$|\mathcal{A}| \geq c_0 \frac{N}{(\log N)^{c_1 \log \log \log N}},$$

*then  $\mathcal{A}$  contains two different elements whose difference is a square.*

It is important to emphasize that Lemma 3 could be replaced by the density theorem of Pintz, Steiger, and Szemerédi [28]. Their result was the first to provide a sufficiently strong upper bound for the size of a set lacking square differences to be applicable in the context of prime gaps. In the following proof, we use the Bloom–Maynard formulation for its convenient explicit constants, but the underlying arithmetic conclusion remains consistent with the Pintz–Steiger–Szemerédi approach.

Next, we choose constants  $N_0, c_0, c_1$  which satisfy Lemma 3. Moreover, let  $N \geq N_0$  and  $\mathcal{P}$  be the set of primes between  $M + 1$  and  $N$ . Then, if  $N$  is sufficiently large depending on  $M$  and  $c_0, c_1$ , we can apply the prime number theorem:

$$|\mathcal{P}| \geq 0.9 \frac{N}{\log N} \geq c_0 \frac{N}{(\log N)^{c_1 \log \log \log N}}.$$

So, using Lemma 3, we proved that there exist two distinct elements  $p \in \mathcal{P}$  and  $q \in \mathcal{P}$ , for which  $p - q$  is a square number. Then, by the definition of  $\mathcal{P}$ ,  $p, q$  are primes and  $M < q < p$ , which completes the proof of the statement.  $\square$

## 4 Proofs of Theorems 3 and 4.

The definition of the admissible set was proposed during the development of the well-known Dickson conjecture, which is as follows.

**Definition 1.** *We say that the set  $\{g_1x + h_1, g_2x + h_2, \dots, g_kx + h_k\} \subseteq \mathbb{Z}[x]$  is admissible if*

$$\prod_{i=1}^k (g_i x + h_i)$$

*has no fixed prime divisor  $p$ , i.e.,*

$$\left| \left\{ n \pmod{p} : \prod_{i=1}^k (g_i n + h_i) \equiv 0 \pmod{p} \right\} \right| < p$$

*holds for all primes  $p$ .*

If we consider the case  $g = 1$ , this definition states that the set  $\{x + h_1, x + h_2, \dots, x + h_k\} \subset \mathbb{Z}[x]$  is admissible if and only if for every prime  $p$ , the set  $\{h_1, h_2, \dots, h_k\}$  does not include a complete residue system modulo  $p$ .

**Conjecture 1** (Dickson's conjecture). *If the set  $\{g_1x + h_1, g_2x + h_2, \dots, g_kx + h_k\} \subseteq \mathbb{Z}[x]$  is admissible, then there are infinitely many positive integers  $n$  for which every element of the set  $\{g_1n + h_1, g_2n + h_2, \dots, g_kn + h_k\}$  is prime.*

Dickson's conjecture remains unsolved; however, one of the most significant breakthroughs towards this open problem is attributed to Maynard [23] and Tao. Their work established the existence of bounded gaps between primes for any  $k$ -tuple, significantly advancing the earlier pioneering results in this direction. The following formulation of their result was given by Granville [16, Theorem 6.4].

**Lemma 4** (Maynard-Tao). *For every  $m \in \mathbb{N}$ , there exists a number  $k_m$  depending only on  $m$  such that if  $k \geq k_m$  and the set  $\{g_1x + h_1, \dots, g_kx + h_k\}$  is admissible, then there exist infinitely many positive integers  $n$  for which the set  $\{g_1n + h_1, \dots, g_kn + h_k\}$  contains at least  $m$  primes.*

**Definition 2.** Denote the smallest constant  $k_m$  for which Lemma 4 holds by  $K_m$ .

Maynard [23] and Tao also gave an upper bound for  $K_m$ ; namely, they proved that

$$K_m \log K_m < e^{8m+4}. \quad (9)$$

This inequality follows from the methods established by Maynard and Tao [23]; however, for a version that provides this explicit form, we refer to the exposition by Granville [16, Theorem 6.4].

Later, Maynard [23, Propositions 4.2. and 4.3.] proved that  $K_2 \leq 105$  and  $K_m \leq cm^2 e^{4m}$ . The sharpest upper bounds were proved in the framework of the Polymath Project [30, Theorem 3.2]:

$$K_2 \leq 50, \quad (10)$$

$$K_m \leq ce^{(4-28/157)m}. \quad (11)$$

In order to prove Theorems 3 and 4, I will use estimates (9) and (10). The main tool in the proofs is the theorem of Banks, Freiberg, and Turnage-Butterbaugh [2].

**Lemma 5** (Banks, Freiberg, Turnage-Butterbaugh). *Let  $m, k \in \mathbb{N}$ ,  $m \geq 2$ ,  $k \geq K_m$ . Furthermore, let  $\{x+b_1, \dots, x+b_k\} \subset \mathbb{Z}[x]$  be an admissible set and  $g \in \mathbb{N}$  such that  $\gcd(g, b_1 \cdots b_k) = 1$ . Then, there is a subset  $\{h_1, \dots, h_m\} \subset \{b_1, \dots, b_k\}$  such that there exist infinitely many  $r \in \mathbb{N}$  for which the elements of  $\{gr + h_1, \dots, gr + h_m\}$  are consecutive primes.*

We will use this lemma for a properly constructed set  $\{b_1, \dots, b_{K_m}\}$ . This set will be of the form  $\{Wa, 2Wa, \dots, K_m Wa\}$ , where  $a$  and  $W$  are defined in the following lemma.

**Lemma 6.** *Let  $m \geq 2$  be an integer. Define  $W$  by*

$$W = \prod_{\substack{p \leq K_m \\ p \text{ prime}}} p. \quad (12)$$

Then, there exist an  $0 < a \in \mathbb{N}$ , for which all elements of

$$\{Wa, 2Wa, \dots, (K_m - 1)Wa\}$$

are perfect power. Furthermore, we also have

$$a < e^{e^{10.5m}}, \quad (13)$$

and, in the special case of  $m = 2$  we also have

$$a < 10^{1.8339 \cdot 10^{76}}. \quad (14)$$

**Proof of Lemma 6.** First we prove (13). Let  $s$  be the largest integer such that  $p_s \leq K_m$ , where  $p_1, p_2, \dots, p_s$  denote the sequence of primes in increasing order. In other words,  $p_1, \dots, p_s$  lists all the primes in the interval  $[2, K_m]$ . Then,  $W = p_1 p_2 \cdots p_s$ . We also need the first  $K_m$  primes, which are  $p_1 = 2, p_2 = 3, \dots, p_{K_m}$ . Next, we will find an integer  $a$ , which is of the form

$$a = p_1^{\alpha_1} \cdots p_s^{\alpha_s},$$

$a$  is as small as possible, and for  $1 \leq i \leq K_m$ ,  $iWa$  is a perfect  $p_i$ -th power (i.e., it is of the form  $iWa = x^{p_i}$ , where  $0 < x \in \mathbb{N}$ ). This determines the remainder of the numbers  $\alpha_1, \alpha_2, \dots, \alpha_s$  modulo  $p_i$  for each  $1 \leq i \leq K_m$ . That is, for a fixed  $j$ , the remainder of an  $\alpha_j$  is given for each of the moduli  $p_1, p_2, \dots, p_{K_m}$ . Thus,  $\alpha_j$  satisfies a simultaneous congruence system, where the moduli  $p_1, p_2, \dots, p_{K_m}$  are pairwise relative primes. According to the Chinese remainder theorem, such an  $\alpha_j$  exists, and for the smallest  $\alpha_j$  with this property, we have

$$0 \leq \alpha_j \leq p_1 p_2 \cdots p_{K_m}.$$

Thus,

$$a = p_1^{\alpha_1} \cdots p_s^{\alpha_s} \leq (p_1 \cdots p_s)^{p_1 p_2 \cdots p_{K_m}}.$$

According to Rosser and Schoenfeld's result [33, Theorem 9],  $\prod_{p \leq x} p < e^{1.01624x}$ , so

$$W = \prod_{p \leq K_m} p = p_1 p_2 \cdots p_s < e^{1.01624K_m} \leq e^{1.01624e^{8m+4}}. \quad (15)$$

We note that while this bound has been significantly improved in recent years (see e.g., [4]), the classical bound is sufficient for our purposes. On the other hand, according to Dusart's result [7, Lemma 1],  $p_k \leq k(\log k + \log \log k)$ , if  $k \geq 6$ , we can also estimate  $a$ . Let  $K = \max\{K_m, 6\}$ . Then,

$$\begin{aligned} a &\leq \left( e^{1.01624K_m} \right)^{p_1 \cdots p_{K_m}} \\ &= e^{1.01624K_m p_1 \cdots p_{K_m}} \\ &< e^{1.01624K \cdot e^{1.01624K(\log K + \log \log K)}}. \end{aligned}$$

Here, by (9) we have  $K \log K < e^{8m+4}$ , so  $K < e^{8m+4}$  and  $\log K + \log \log K < 8m + 4$ . Thus,

$$\begin{aligned} a &< e^{1.01624e^{8m+4} e^{1.01624e^{8m+4}(8m+4)}} \\ &< e^{e^{10.5m}}, \quad \text{if } m \geq 3. \end{aligned} \quad (16)$$

Note that for  $m = 2$  the upper estimate in (16) is approximately

$$10^{2.0407 \cdot 10^{4282528195}},$$

which is an order of magnitude worse than the upper estimate given in (14).

Next, we prove (14). This requires slightly more precise calculations, for which I used the SAGE programming language again. In the proof, we use the fact that  $K_2 \leq 50$  (see (10)). Let us list the primes between 1 and 50:  $p_1 = 2, p_2 = 3, \dots, p_{15} = 47$ . Then,  $W \leq p_1 \cdots p_{15}$ . Now, we will find an integer  $a$  of the form

$$a = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_{15}^{\alpha_{15}}, \quad (17)$$

such that for every  $1 \leq i \leq 49$ ,  $iWa$  is a perfect power. To do this, we divide the set  $\{1, 2, \dots, 49\}$  into three parts:

$$\begin{aligned}\mathcal{H}_1 &= \{1, 4, 9, 16, 25, 49\}, \\ \mathcal{H}_2 &= \{3, 24\}, \\ \mathcal{H}_3 &= \{1, 2, \dots, 49\} \setminus (\mathcal{H}_1 \cup \mathcal{H}_2).\end{aligned}$$

Then,  $|\mathcal{H}_3| = 41$ . Let us define a bijection  $f$  between the elements of  $\mathcal{H}_3$  and the prime numbers  $\{p_3 = 5, p_4 = 7, p_5 = 11, \dots, p_{43}\}$ . We will construct a positive integer  $a$  for which:

$Wa$  is a square,

$3Wa$  is a cube,

and for  $i \in \mathcal{H}_3$ ,  $iWa$  is an  $f(i)$ -th power, (i.e., it is of the form  $iWa = x^{f(i)}$ , where  $0 < x \in \mathbb{N}$  and  $f(i)$  is a prime in the set  $\{p_3, p_4, \dots, p_{41}\}$ ).

Then, for every  $i \in \{1, 2, \dots, 49\}$ ,  $iWa$  is a perfect power since for  $i \in \mathcal{H}_1$ ,  $iWa$  is a square, for  $i \in \mathcal{H}_2$ ,  $iWa$  is a cube, and for  $i \in \mathcal{H}_3$ ,  $iWa$  is an  $f(i)$ -th power. As before, now, each exponent  $\alpha_j$  in the prime factorization of the integer  $a$  given in (17) must satisfy a simultaneous congruence system, where the moduli are the first 43 primes. Since these moduli are also pairwise relative primes, we can use the Chinese remainder theorem, and get

$$\alpha_j < \prod_{j=1}^{43} p_j \stackrel{\text{def}}{=} E.$$

Then,

$$a \leq (p_1 \cdots p_{15})^E.$$

I calculated this value using the SAGE programming language (Appendix, Program 5), and got the following result:

$$\begin{aligned}W &\leq p_1 p_2 \cdots p_{15} < 10^{17,789} \\ a &< 10^{1.8339 \cdot 10^{76}}.\end{aligned}\tag{18}$$

This completes the proof of Lemma 6.  $\square$

Now, we can prove Theorems 3 and 4. In the case of the proof of Theorem 3, let  $m = 2$ . In Theorems 3 and 4, we look for an infinite number of consecutive primes  $p_{n+1}, p_{n+2}, \dots, p_{n+m}$  such that for  $1 \leq i < j \leq m$ ,  $p_{n+j} - p_{n+i}$  is a perfect power, and

$$p_{n+m} - p_{n+1} < e^{e^{10.6m}}, \quad \text{if } m \geq 3,$$

respectively,

$$p_{n+2} - p_{n+1} < 10^{1.834 \cdot 10^{76}}, \quad \text{if } m = 2.$$

For this, consider the  $a$  and  $W$  given in Lemma 6. Then, every element of the set  $\{Wa, 2Wa, \dots, (K_m - 1)Wa\}$  is a perfect power. Next, we would like to apply Lemma 5 with  $k = K_m$ ,  $\{b_1, b_2, \dots, b_k\} = \{Wa, 2Wa, \dots, K_m Wa\}$  and  $g = 1$ . Then, the set  $\{x + b_1, \dots, x + b_{K_m}\}$  is admissible since  $\{Wa, 2Wa, \dots, K_m Wa\}$  does not contain a complete residue system for any prime  $p$ . This is because if  $p \leq K_m$ , then by  $p \mid W$ , every element of the set is congruent to 0 modulo  $p$ , and if  $p > K_m$ , then the set has fewer elements than  $p$ . By Lemma 5, there is a subset  $\{h_1, h_2, \dots, h_m\} \subset \{Wa, 2Wa, \dots, K_m Wa\}$  for which there exists infinitely many  $r$  such that  $r + h_1, r + h_2, \dots, r + h_m$  are consecutive primes. Let

$$p_{n+1} = r + h_1, \dots, p_{n+m} = r + h_m.$$

Then, for  $1 \leq i < j \leq m$ ,

$$p_{n+j} - p_{n+i} = h_j - h_i \in \{Wa, 2Wa, \dots, (K_m - 1)Wa\},$$

such that  $p_{n+j} - p_{n+i}$  is a perfect power. On the other hand, for the  $a$  given in Lemma 5, we have (13) or (14), and for  $W$ , we have (15) or (18); thus,

$$\begin{aligned} p_{n+m} - p_{n+1} &\leq (K_m - 1)Wa < K_m W e^{e^{10.5m}} \\ &< e^{8m+4} e^{1.01624e^{8m+4}} e^{e^{10.5m}} < e^{e^{10.6m}}, \quad \text{if } m \geq 3, \end{aligned}$$

and for  $m = 2$ , we have

$$\begin{aligned} p_{n+2} - p_{n+1} &\leq (K_2 - 1)Wa < K_2W10^{1.8339 \cdot 10^{76}} \\ &< 50 \cdot 10^{17.781} 10^{1.8339 \cdot 10^{76}} < 10^{1.834 \cdot 10^{76}}. \end{aligned}$$

This completes the proofs of Theorems 3 and 4.  $\square$

**Remark.** If we could find 50 distinct squares such that the difference between any two is also a square, then Problem 1 could be proved using the method described here. However, I strongly doubt that such a set of 50 elements exists. In fact, I do not think there is even a set of five elements with the above property (see Conjecture 4 in [17]). It is also worth mentioning that based on the Bombieri-Lang conjecture, an absolute constant upper bound can be given for the size of sets of integers where the difference between any two numbers is a square (for example, see the Note after Theorem 4.6 in [1]). Unfortunately, we do not know whether this constant is greater than or less than 50.

## 5 Proof of Theorem 5.

We can assume that  $x \geq 17$  since the theorem is trivial for  $2 \leq x \leq 16$ . Let  $D(x)$  denote the number of elements of the largest set in  $[1, x]$  that do not have two elements whose difference is a square, and let  $D_P(x)$  denote the number of elements of the largest set of primes in  $[1, x]$  that do not have two elements whose difference is a square. Ruzsa [34], in disproving a conjecture of Erdős, proved the following:

**Theorem B.**

$$D(x) > \frac{1}{65}x^\gamma,$$

where

$$\gamma = \frac{1}{2} \left( 1 + \frac{\log 7}{\log 65} \right) = 0.733077 \dots$$

In fact, Ruzsa proved a slightly more general theorem. Namely, if  $r(m)$  denotes the maximum number of residue classes modulo  $m$  such that the difference between any two residue classes is not a square modulo  $m$ , then the following holds.

**Theorem C.** *For all squarefree  $m$  we have*

$$D(x) \geq \frac{1}{m} x^{\gamma(m)},$$

where

$$\gamma(m) = \frac{1}{2} + \frac{\log r(m)}{2 \log m}.$$

Ruzsa noticed that

$$r(65) \geq 7. \tag{19}$$

Namely, if we consider the pairs

$$(0, 0), (0, 2), (1, 8), (2, 1), (2, 3), (3, 9), (4, 7)$$

where we assign to the pairs of numbers the residue class modulo 65, whose residue modulo 5 is the first member of the pair, while the residue 13 is the second member of the pair, then for these residue classes mod 65, there are no two elements in the set whose difference is a square modulo 65, since the difference between any two elements mod 5 or mod 13 is a quadratic non-residue. This proves (19). Later, Lewko [21] gave a similar construction with modulus 205, proving that

$$r(205) \geq 12.$$

The only question is whether the proof of Theorem C can be transferred to the case where the set can only contain prime numbers. We will prove the following.

**Lemma 7.** *For all  $x \geq 17$  and squarefree  $m$  we have*

$$D_P(x) \geq \frac{x^{\gamma(m)}}{m \log x},$$

where

$$\gamma(m) = \frac{1}{2} + \frac{\log r(m)}{2 \log m}.$$

Using Lemma 7 with  $m = 205$  and  $r(205) \geq 12$ , we immediately obtain the statement of Theorem 5. Next, the proof of Lemma 7 follows.

**Proof of Lemma 7.** Let  $R \subset \mathbb{Z}_m$  be a set such that the difference between any two elements is not a square modulo  $m$  and  $|R| = r(m)$ . If  $r \in \mathbb{Z}_m$ , then let  $\bar{r}$  denote the smallest nonnegative integer that represents  $r$  among integers, i.e.,

$$r \equiv \bar{r} \pmod{m}, \quad 0 \leq \bar{r} < m.$$

If  $0 \leq a < m$ , define  $R_a$  by the following set:

$$R_a \stackrel{\text{def}}{=} \{\overline{r+a} : r \in R\} \subset \{0, 1, 2, \dots, m-1\}.$$

Then, the difference between two elements in  $R_a$  is never a square modulo  $m$ , since the difference between elements in  $R_a$  modulo  $m$  is the same as the difference between the corresponding elements in  $R$ .

Let  $n = \lfloor \log_m x \rfloor$ , so that

$$m^n \leq x < m^{n+1}. \tag{20}$$

We consider the  $k$ -tuples  $(i_0, i_2, \dots, i_t)$ , where  $t$  is the largest even integer less than  $n$ , and  $k = \lfloor n/2 \rfloor + 1$  denotes the number of coordinates. For each such tuple, where  $0 \leq i_{2j} < m$  are integers, we define the set  $S(i_0, i_2, \dots, i_t)$  as the collection of integers that can be written in the form

$$s = \sum_{j=0}^{n-1} r_j m^j + 1,$$

where  $r_j \in R_{i_j}$  if  $j$  is even, and  $0 \leq r_j < m$  is an arbitrary integer if  $j$  is odd. We will prove that the difference between any two elements in  $S(i_0, i_2, \dots, i_t)$

cannot be a square. Suppose that  $s - s' = u^2$ , where  $s, s' \in S(i_0, i_2, \dots, i_t)$ .

Let

$$s = \sum_{j=0}^{n-1} r_j m^j + 1, \quad s' = \sum_{j=0}^{n-1} r'_j m^j + 1.$$

Let  $k$  denote the smallest index  $k$  for which  $r_k \neq r'_k$ . Then:

$$u^2 = s - s' = (r_k - r'_k)m^k + zm^{k+1}.$$

If  $k$  is odd, then  $m^k \mid u^2$ , but  $m^{k+1} \nmid u^2$ , which is impossible for squarefree  $m$ . If  $k$  is even, let  $k = 2\ell$ , and then

$$(u/m^\ell)^2 \equiv r_k - r'_k \pmod{m}, \quad \text{where } r_k, r'_k \in R_{i_k},$$

which contradicts the fact that the difference between any two elements in  $R_{i_k}$  is not a square modulo  $m$ .

For every integer  $1 \leq s \leq m^n$ ,  $s - 1$  can be written in the  $m$  number system:

$$s = (s - 1) + 1 = \sum_{j=0}^{n-1} r_j m^j + 1.$$

For a fixed even  $j$ , there exist exactly  $r(m)$  values of  $i$  such that  $0 \leq i < m$  and  $r_j \in R_i$ . That is, for a given integer  $1 \leq s \leq m^n$ , there exist a total of  $r(m)^{\lfloor (n+1)/2 \rfloor}$  pieces of  $\lfloor (n+1)/2 \rfloor$ -tuples  $(i_0, i_2, i_4, \dots, i_t)$  such that

$$s \in S(i_0, i_2, \dots, i_t).$$

This is true for all integers  $1 \leq s \leq m^n$ , including the primes. Thus,

$$\begin{aligned} & |\{(p, (i_0, i_2, \dots, i_t)) : 1 \leq p \leq m^n \text{ prime}, p \in S(i_0, \dots, i_t), 0 \leq i_j < m, \}| \\ & \hspace{15em} (21) \\ & = r(m)^{\lfloor (n+1)/2 \rfloor} \pi(m^n). \end{aligned}$$

Since the number of  $k$ -tuples  $(i_0, i_2, \dots, i_t)$  with  $0 \leq i_j < m$  is exactly  $m^k$ , according to the pigeon-hole principle, there exists a  $k$ -tuple  $(i_0, i_2, \dots, i_t)$  such that  $S(i_0, i_2, \dots, i_t)$  contains at least

$$\frac{r(m)^k \pi(m^n)}{m^k}$$

primes. If we prove that

$$\frac{r(m)^{\lfloor (n+1)/2 \rfloor} \pi(m^n)}{m^{\lfloor (n+1)/2 \rfloor}} \geq \frac{x^{\gamma(m)}}{m \log x}, \quad (22)$$

then the number of primes in this set  $S(i_0, i_2, \dots, i_t)$  is  $\geq \frac{x^{\gamma(m)}}{m \log x}$ , and the difference between any two primes is not a square number, which proves the statement of Lemma 7. So, we only need to prove (22). In [33, Corollary 1], Rosser and Schoenfeld proved that for  $x \geq 17$ ,  $\pi(x) > \frac{x}{\log x}$ . So,

$$\frac{r(m)^{\lfloor (n+1)/2 \rfloor} \pi(m^n)}{m^{\lfloor (n+1)/2 \rfloor}} \geq \frac{r(m)^{\lfloor (n+1)/2 \rfloor} m^n}{m^{\lfloor (n+1)/2 \rfloor} \log(m^n)}.$$

By (20),

$$\begin{aligned} \frac{r(m)^{\lfloor (n+1)/2 \rfloor} \pi(m^n)}{m^{\lfloor (n+1)/2 \rfloor}} &\geq \frac{r(m)^{\lfloor (n+1)/2 \rfloor} m^n}{m^{\lfloor (n+1)/2 \rfloor} \log x} \\ &= \frac{1}{\log x} m^{[(n+1)/2](\log r(m)/\log m - 1) + n}. \end{aligned}$$

Since by  $\log r(m)/\log m - 1$  is negative,

$$\begin{aligned} \frac{r(m)^{\lfloor (n+1)/2 \rfloor} \pi(m^n)}{m^{\lfloor (n+1)/2 \rfloor}} &\geq \frac{1}{\log x} m^{(n+1)(\log r(m)/\log m - 1)/2 + n} \\ &= \frac{1}{m \log x} m^{(n+1) \log r(m)/(2 \log m) + (n+1)/2}. \end{aligned}$$

Thus, by (20),

$$\begin{aligned} \frac{r(m)^{\lfloor (n+1)/2 \rfloor} \pi(m^n)}{m^{\lfloor (n+1)/2 \rfloor}} &\geq \frac{1}{m \log x} x^{\log r(m)/(2 \log m) + 1/2} \\ &= \frac{x^{\gamma(m)}}{m \log x}, \end{aligned}$$

which proves (22). This completes the proof of the lemma.  $\square$

## 6 Proof of Theorem 6.

Let us list the primes between 2 and  $x$  in increasing order:

$$2 = p_1 < p_2 < p_3 < \dots < p_s \leq x.$$

Let

$$t = [0.24(\log x)^{1/4}], \quad (23)$$

$$r = \frac{t^{1/3}(\log x)^{2/3}}{3.323^{1/3}}. \quad (24)$$

We color the primes in the interval  $[1, x]$  with three colors. The color of the prime  $2 \leq p_i \leq x$  is red if there exists an integer  $1 \leq h \leq r$  such that  $p_i + h^2$  is also prime. The color of the prime  $2 \leq p_i \leq x$  is yellow if

$$p_{i+t} - p_i > r^2.$$

The primes that are not colored red or yellow are colored green. In the proof, we look for  $t$  consecutive green primes. If  $p_i, p_{i+1}, \dots, p_{i+t-1}$  are  $t$  consecutive green primes, then for  $0 \leq m < n \leq t-1$ ,

$$p_{i+m} - p_{i+n}$$

is not a square. This is because if

$$p_{i+m} - p_{i+n} = h^2$$

holds, then for  $h \leq r$ , the color of  $p_{i+n}$  is red, and if  $h > r$ , then by  $p_{i+t} - p_i > p_{i+m} - p_{i+n} = h^2 > r^2$ , the color of  $p_i$  is yellow. Next, we will use the following lemma:

**Lemma 8.** *For sufficiently large  $x$ , denote the number of red primes in the interval  $[1, x]$  by  $R$ , and the number of yellow primes by  $Y$ . Then*

$$R \leq 6.646 \frac{rx}{(\log x)^2}, \quad (25)$$

$$Y \leq \frac{tx}{r^2}. \quad (26)$$

Before proving the lemma, let us see how the theorem follows from it. We divide the interval  $[1, x]$  into disjoint subintervals of the form  $[a, b)$ . We call such a subinterval *red* if all primes within it are either yellow or red;

otherwise, if a subinterval contains only green primes, it is called a *green interval*.

The interval  $[1, x)$  can be represented as a union of alternating disjoint red and green intervals. By our construction, the first interval starting from 1 is red (since the initial small primes do not satisfy the conditions for being green). Because red and green intervals alternate, the number of green intervals is at most the number of red intervals. Note that each red interval contains at least one red or yellow prime by definition. Thus, the total number of red and yellow primes,  $R + Y$ , provides an upper bound for the number of such intervals:

$$R + Y \leq 6.646 \frac{rx}{(\log x)^2} + \frac{tx}{r^2}.$$

Using the prime number theorem in the form  $\pi(x) > x/\log x$ , the number of green primes  $G$  satisfies:

$$G \geq \pi(x) - (R + Y) \geq \frac{x}{\log x} - 6.646 \frac{rx}{(\log x)^2} - \frac{tx}{r^2}.$$

Thus, we have at most  $6.646 \frac{rx}{(\log x)^2} + \frac{tx}{r^2}$  green intervals, which contain more than  $\frac{x}{\log x} - 6.646 \frac{rx}{(\log x)^2} - \frac{tx}{r^2}$  green primes. So, by the pigeon-hole principle, there exists a green interval containing at least

$$H \stackrel{\text{def}}{=} \frac{\frac{x}{\log x} - 6.646 \frac{rx}{(\log x)^2} - \frac{tx}{r^2}}{6.646 \frac{rx}{(\log x)^2} + \frac{tx}{r^2}}$$

If this quantity is greater than  $0.24(\log x)^{1/4}$ , the statement of the theorem is proved. Indeed,

$$H = \frac{\frac{x}{\log x}}{6.646 \cdot \frac{rx}{(\log x)^2} + \frac{tx}{r^2}} - 1.$$

Then, according to (24), given as the definition of  $r$

$$H = \frac{(\log x)^{1/3}}{3 \cdot 3.323^{2/3} t^{1/3}} - 1,$$

where by (23), we have

$$H > 0.2408 \cdot (\log x)^{1/4},$$

and this was to be proved. All that remains is to prove Lemma 8.

**Proof of Lemma 8.** By Lemma 2. for the number of red primes we have

$$\begin{aligned} R &\leq \sum_{\substack{h=2 \\ 2|h}}^{[r]} 8 \prod_{2 < p} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{2 < p|h^2} \frac{p-1}{p-2} \cdot \frac{x}{(\log x)^2} \left(1 + O\left(\frac{\log \log x}{\log x}\right)\right) \\ &\quad + \sum_{\substack{h=1 \\ h \text{ is odd}}}^{[r]} 1. \end{aligned}$$

But  $p \mid h^2$  holds if and only if  $p \mid h$ , thus

$$\begin{aligned} R &\leq \sum_{\substack{h=2 \\ 2|h}}^{[r]} 8 \prod_{2 < p} \left(1 - \frac{1}{(p-1)^2}\right) \prod_{2 < p|h} \frac{p-1}{p-2} \cdot \frac{x}{(\log x)^2} \left(1 + O\left(\frac{\log \log x}{\log x}\right)\right) \\ &\quad + \sum_{\substack{h=1 \\ h \text{ is odd}}}^{[r]} 1. \end{aligned}$$

From here, we can perform the same calculations as in the proof of Lemma 2, and we get

$$R \leq 6.646 \frac{rx}{(\log x)^2},$$

which proves (25). In order to prove (26), consider the primes in  $[1, x]$  in the

following sequences

$$\begin{aligned}
 & p_1, p_{t+1}, p_{2t+1}, \dots \\
 & p_2, p_{t+2}, p_{2t+2}, \dots \\
 & \vdots \\
 & p_t, p_{2t}, p_{3t}, \dots
 \end{aligned}$$

Here, in the  $j$ -th sequence

$$2 \leq p_j, p_{t+j}, p_{2t+j}, \dots, p_{\ell t+j} \leq x,$$

where  $\ell = \ell(j)$  is the largest integer such that  $p_{\ell t+j} \leq x$ , there can be at most  $\frac{x}{r^2}$  yellow primes. This follows from the fact that

$$x > (p_{t+j} - p_j) + (p_{2t+j} - p_{t+j}) + (p_{3t+j} - p_{2t+j}) + \dots + (p_{\ell t+j} - p_{(\ell-1)t+j}),$$

and in the above sum, at most  $\frac{x}{r^2}$  differences can be greater than  $r^2$  (otherwise their sum would exceed  $x$ ). Since we have  $t$  such sequences, the total number of yellow primes  $Y$  is at most  $\frac{tx}{r^2}$ . This completes the proof of the lemma.  $\square$

## 7 Appendix

Here are the SAGE programs required in the proofs of Theorems 1, 3, and 6. I ran these codes on the following online platform:

<https://sagecell.sagemath.org/>

\*\*\*

### Program 1

n=5000

prod( 1-1/((p-1)\*\*2) for p in prime\_range(3,n)).n()

**Result:** 0.660175738989977

\*\*\*

### Program 2

T=20000

```
sum(2/((p-2)*(p+1)) for p in prime_range(3,T+2)).n()
```

**Result:** 0.727089417741948

\*\*\*

### Program 3

T=5000

```
sum(moebius(d)**2/(d**2) for d in range(2, T+1,2)).n()
```

**Result:** 0.303923082993008

\*\*\*

### Program 4

T=5000

```
sum(moebius(d)**2/(2*d**2) for d in range(1, T+2,2)).n()
```

**Result:** 0.607886600147474

\*\*\*

### Program 5

```
W=prod(p for p in prime_range(2,50))
```

```
P=Primes()
```

```
E=prod(P.unrank(j) for j in range(0,43))
```

```
exponent=(E*log(W,10)).n()
```

```
print(log(W,10).n())
```

```
print(exponent)
```

**Result:** 17.7887972765829, 1.83383491155388e76

## References

- [1] N. Alon, O. Angel, I. Benjamini, and E. Lubetzky, *Sums and products along sparse graphs*, Israel J. Math. 188(1) (2012), 353–384.
- [2] W. D. Banks, T. Freiberg, and C. L. Turnage-Butterbaugh, *Consecutive primes in tuples*, Acta Arith. 167(3) (2015), 261-266.
- [3] T. F. Bloom and J. Maynard, *A new upper bound for sets with no square differences*, Compos. Math. 158(8) (2022), 1777-1798.
- [4] M. Bordignon, D. R. Johnston, and V. S. Starichkova, *An explicit version of Chen’s theorem and the linear sieve*, Int. J. Number Theory (2025).
- [5] J.-R. Chen, *On the Goldbach’s problem and the sieve methods*, Sci. Sin. 21 (1978), 701-739.
- [6] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Prace Mat.-Fiz. 45 (1937), 51-74.
- [7] P. Dusart, *The  $k^{\text{th}}$  prime is greater than for  $k \geq 2$* , Math. Comput. 68(225) (1999), 411-415.
- [8] P. Erdős, *On the difference of consecutive primes*, Q. J. Math., Oxf. Ser. 6 (1935), 124-128.
- [9] P. Erdős, *Some unsolved problems*, in: B. Bollobás, et al. (ed.), Combinatorics, geometry and probability. A tribute to Paul Erdős. Proceedings of the conference dedicated to Paul Erdős on the occasion of his 80th birthday, Cambridge, UK, 26 March 1993. Cambridge: Cambridge University Press, 1-10 (1997).
- [10] K. Ford, B. Green, S. Konyagin, and T. Tao, *Large gaps between consecutive prime numbers*, Ann. Math. (2) 183(3) (2016), 935-974.

- [11] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao, *Long gaps between primes*, J. Am. Math. Soc. 31(1) (2018), 65-105.
- [12] K. Ford, J. Maynard, and T. Tao, *Chains of large gaps between primes*, in: Pintz, János (ed.) et al., Irregularities in the distribution of prime numbers, From the era of Helmut Maier's matrix method and beyond. Cham: Springer. 1-21 (2018).
- [13] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, *Primes in tuples. I*, Ann. Math. 170(2) (2009), 819–862.
- [14] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, *Primes in tuples. II*, Acta Math. 204(1) (2010), 1-47.
- [15] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, *Primes in tuples. III*, Funct. Approximatio, Comment. Math. 35 (2006), 79-89.
- [16] A. Granville, *Primes in intervals of bounded length*, Bull. Am. Math. Soc., New Ser. 52(2) (2015), 171-222.
- [17] K. Gyarmati, *On Diophantine square tuples*, Fibonacci Q. 63(3) (2025), 554-569.
- [18] H. Halberstam and H. E. Richert, *Sieve Methods*, London Mathematical Society Monographs. 4. London-New York San Francisco: Academic Press, a subsidiary of Harcourt Brace Jovanovich, Publishers. xiii, 1974, p. 117.
- [19] D. R. Johnston, *An effective Bombieri–Vinogradov error term for sifting problems*, arXiv:2510.10853 (2025).
- [20] J. D. Lichtman, *Primes in arithmetic progressions to large moduli, and Goldbach beyond the square-root barrier*. arXiv:2309.08522 (2023).

- [21] M. Lewko, *An improved lower bound related to the Furstenberg-Sárközy theorem*, Electron. J. Comb. 22(1) (2015), Research Paper P1.32, 6 p.
- [22] H. Maier, *Chains of large gaps between consecutive primes*, Adv. Math. 39 (1981), 257-269.
- [23] J. Maynard, *Small gaps between primes*, Ann. Math. (2) 181(1) (2015), 383-413.
- [24] J. Maynard, *Large gaps between primes*, Ann. Math. (2) 183(3) (2016), 915-933.
- [25] J. Maynard, *Large gaps between primes* Notices Am. Math. Soc. 64(5) (2017), 438-440.
- [26] P. Moree, *Some number-theoretical constants*, .
- [27] J. Pintz, *On the distribution of gaps between consecutive primes*, arXiv:1407.2213 (2014).
- [28] J. Pintz, W. L. Steiger, and E. Szemerédi, *On sets of natural numbers whose difference set contains no squares*, J. Lond. Math. Soc., II. Ser. 37(2) (1989), 219-231.
- [29] Polymath, *New equidistribution estimates of Zhang type*, Algebra Number Theory 8(9) (2014), 2067–2199, arXiv:1402.0811.
- [30] Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes*, Res. Math. Sci. 1, Paper No. 12 (2014), 83p., arXiv:1407.4897.
- [31] R. A. Rankin, *The difference between consecutive prime numbers*, J. Lond. Math. Soc. 13 (1938), 242-247.
- [32] H. Riesel and R. C. Vaughan, *On sums of primes*, Ark. Mat. 21.1 (1983), pp. 45–74.

- [33] J. B. Rosser, and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Ill. J. Math. 6 (1962), 64-94.
- [34] I. Z. Ruzsa, *Difference sets without squares*, Period. Math. Hung. 15 (1984), 205-209.
- [35] D. K. L. Shiu, *Strings of congruent primes*, J. Lond. Math. Soc., II. Ser. 61(2) (2000), 359-373.
- [36] W. A. Stein et al. *Sage Mathematics Software (Version 10.5)*, The Sage Development Team, 2024, <http://www.sagemath.org>.
- [37] Y. Zhang, *Bounded gaps between primes*, Ann. Math. (2) 179(3) (2014), 1121-1174.
- [38] E. Westzynthius, *Über die Verteilung der Zahlen, die zu den  $n$  ersten Primzahlen teilerfremd sind.*, Comment. Phys.-Math., Soc. Sci. Fennica 5(25) (1931), 1-37.