

Polignac Numbers, Conjectures of Erdős on Gaps between Primes, Arithmetic Progressions in Primes, and the Bounded Gap Conjecture

by

János PINTZ

Rényi Mathematical Institute of the Hungarian Academy of Sciences

Budapest, Reáltanoda u. 13–15

H-1053 Hungary

e-mail: pintz@renyi.hu

Abstract

In the present work we prove a number of surprising results about gaps between consecutive primes and arithmetic progressions in the sequence of generalized twin primes which could not have been proven without the recent fantastic achievement of Yitang Zhang about the existence of bounded gaps between consecutive primes. Most of these results would have belonged to the category of science fiction a decade ago. However, the presented results are far from being immediate consequences of Zhang's famous theorem: they require various new ideas, other important properties of the applied sieve function and a closer analysis of the methods of Goldston–Pintz–Yıldırım, Green–Tao, and Zhang, respectively.

1 Introduction

1.1. Very recently Yitang Zhang, in a fantastic breakthrough solved the Bounded Gap Conjecture, a term formulated in recent works of mine, often in collaboration with D. Goldston and C. Yıldırım. Let p_n denote the n^{th} prime, \mathcal{P} the set of all primes and

$$d_n = p_{n+1} - p_n \tag{1.1}$$

the n^{th} difference between consecutive primes.

Theorem A (Y. Zhang [28, 2013]). $\liminf_{n \rightarrow \infty} d_n \leq 7 \cdot 10^7$.

Earlier, in a joint work with D. Goldston and C. Yıldırım we showed [11, 2009] this under the deep unproved condition that primes have a distribution level greater than $1/2$, in other words, under the assumption that the exponent $1/2$ in the Bombieri–Vinogradov Theorem can be improved.

Theorem B ([11, 2009]). *If the primes have a distribution level $\vartheta > 1/2$, then with a suitable explicit constant $C(\vartheta)$ we have*

$$\liminf_{n \rightarrow \infty} d_n \leq C(\vartheta). \tag{1.2}$$

If the Elliott–Halberstam Conjecture (EH) is true, i.e. $\vartheta = 1$ or, at least, $\vartheta > 0.971$ then

$$\liminf_{n \rightarrow \infty} d_n \leq 16. \tag{1.3}$$

⁰Supported by OTKA grants K 100291, NK 104183 and ERC-AdG. 228005.

We say that ϑ is a level of distribution of primes if

$$\sum_{q \leq X^{\vartheta - \varepsilon}} \max_{(a, q) = 1} \left| \sum_{\substack{p \equiv a \pmod{q} \\ p \leq X}} \log p - \frac{X}{\varphi(q)} \right| \ll_{A, \varepsilon} \frac{X}{(\log X)} \quad (1.4)$$

for any $A, \varepsilon > 0$. The Bombieri–Vinogradov Theorem asserts that $\vartheta = 1/2$ is a level of distribution and the Elliott–Halberstam Conjecture [4, 1970] asserts that $\vartheta = 1$ is also an admissible level. Zhang could not prove (1.4) with a level $> 1/2$ but showed that

(i) it is possible to neglect in (1.4) all moduli q with a prime divisor $> X^b$, where b is any constant, and

(ii) managed to show for these smooth moduli a result which is similar but weaker than the analogue of (1.4) with $\vartheta = 1/2 + 1/584$, which, however suits to apply our method of Theorem B.

I have to remark that the present author mentioned the phenomenon (i) in his lecture at the conference of the American Institute of Mathematics in November 2005, Palo Alto, and worked it out in a joint paper with Y. Motohashi [18, 2008], however, without being able to show (ii).

The present author worked out recently various consequences of a hypothetical distribution level $\vartheta > 1/2$ ([19, 2010] and [22, 2013]) which partly appeared with proofs [19, 2010], partly exist at a level of announcements [22, 2013]. The purpose of the present work is to show that most of these conditional results can be shown unconditionally using Zhang’s fantastic result [28, 2013] or in some sense, his method, coupled with other ideas.

2 History of the problems. Formulation of the results

2.1 Approximations to the Twin Prime Conjecture

The problem of finding small gaps between primes originates from the

Twin Prime Conjecture. $\liminf_{n \rightarrow \infty} d_n = 2$.

Since by the Prime Number Theorem the average gap size is $\log n$, Hardy and Littlewood considered first already in 1926 in an unpublished manuscript (see, however [24, 1940]) the upper estimation of

$$\Delta_1 = \liminf_{n \rightarrow \infty} \frac{d_n}{\log n} \quad (2.1)$$

and showed $\Delta_1 \leq 2/3$ under the assumption of the Generalized Riemann Hypothesis (GRH).

Erdős [5, 1940] was the first to show unconditionally

$$\Delta_1 < 1 - c_0, \quad c_0 > 0 \quad (2.2)$$

with an unspecified but explicitly calculable positive constant c_0 . The full history with about 12–15 improvements concerning the value of Δ_1 is contained in [11, 2009], so we list here just the most important steps:

$$\Delta_1 < 0.4666 \quad (\text{Bombieri, Davenport [1, 1966]}), \quad (2.3)$$

$$\Delta_1 < 0.2485 \quad (\text{H. Maier [17, 1988]}). \quad (2.4)$$

This was the best result until 2005 when we showed in a joint work with D. Goldston and C. Yıldırım what we called the Small Gap Conjecture:

Theorem C ([11, 2009]). $\Delta_1 = 0$.

Soon after it we improved this to

Theorem D ([12, 2010]). *We have*

$$\liminf_{n \rightarrow \infty} \frac{d_n}{(\log n)^{1/2} (\log \log n)^2} < \infty. \quad (2.5)$$

Finally a few years ago I improved the exponent to $3/7$ and announced the result at the Journées Arithmétiques, Vilnius, 2011, and Turán Memorial Conference, Budapest, 2011.

Theorem E ([21, 2013]).

$$\liminf_{n \rightarrow \infty} \frac{d_n}{(\log n)^{3/7} (\log \log n)^{4/7}} < \infty. \quad (2.6)$$

In a joint work with B. Farkas and Sz. Gy. Révész [9, 2013] we also showed that essential new ideas are necessary to improve (2.6).

2.2 Polignac numbers

The Twin Prime Conjecture appeared already in a more general form in 1849 in a work of de Polignac.

First we give two definitions.

Definition 1. A positive even number $2k$ is a strong Polignac number, or briefly a Polignac number if $d_n = 2k$ for infinitely many values of n .

Definition 2. A positive even number $2k$ is a weak Polignac number if it can be written as the difference of two primes in an infinitude of ways.

The set of (strong) Polignac numbers will be denoted by \mathcal{D}_s , the set of weak Polignac numbers by \mathcal{D}_w . (We have trivially $\mathcal{D}_s \subseteq \mathcal{D}_w$.)

Polignac's Conjecture ([23, 1849]). *Every positive even integer is a (strong) Polignac number.*

Since the smallest weak Polignac number has to be a strong Polignac number, an easy consideration gives that using $|A|$ (or sometimes $\#A$) to denote the number of elements of a set A , the following proposition is true.

Proposition. *The following three statements are equivalent:*

- (i) *the Bounded Gap Conjecture is true;*
- (ii) *there is at least one (strong) Polignac number, i.e. $|\mathcal{D}_s| \geq 1$;*
- (iii) *there is at least one weak Polignac number, i.e. $|\mathcal{D}_w| \geq 1$.*

The above very simple proposition shows that the Bounded Gap Conjecture itself leaves still many problems open about Polignac numbers or weak Polignac numbers, but solves the crucial problem that their set is nonempty. We will prove several unconditional results about the density and distribution of (strong) Polignac numbers.

Theorem 1. *There exists an explicitly calculable constant c such that for $N > N_0$ we have at least cN Polignac numbers below N , i.e. Polignac numbers have a positive lower asymptotic density.*

Theorem 2. *There exists an ineffective constant C' such that every interval of type $[M, M + C']$ contains at least one Polignac number.*

Remark. As mentioned earlier, the term Polignac number means always strong Polignac numbers.

2.3 The normalized value distribution of d_n

The Prime Number Theorem implies

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum \frac{d_n}{\log n} = 1, \quad (2.7)$$

so it is natural to investigate the series $d_n/\log n$. Denoting by J the set of limit points of $d_n/\log n$, Erdős [7, 1955] conjectured

$$J = \left\{ \frac{d_n}{\log n} \right\}' = [0, \infty]. \quad (2.8)$$

While Westzynthius [27, 1931] proved more than 80 years ago that $\infty \in J$, no finite limit point was known until 2005 when we showed the Small Gap Conjecture, i.e. Theorem C [11, 2009], which is equivalent to $0 \in J$.

Interestingly enough Erdős [7, 1955] and Ricci [25, 1954] proved independently about 60 years ago that J has a positive Lebesgue measure. What I can show is a weaker form of Erdős's conjecture (2.8).

Theorem 3. *There is an ineffective constant $c > 0$ such that*

$$[0, c] \subset J. \quad (2.9)$$

Kálmán Győry asked me at the Turán Memorial Conference in Budapest, 2011 whether it is possible to find a form of the above result which gives answers about the more subtle value-distribution of d_n if we use a test-function $f(n) \leq \log n$, $f(n) \rightarrow \infty$ as $n \rightarrow \infty$.

Definition 3. Let \mathcal{F} denote the class of functions $f : \mathbb{Z}^+ \rightarrow \mathbb{R}^+$ with a slow oscillation, when for every $\varepsilon > 0$ we have an $N(\varepsilon) > 0$ such that

$$(1 - \varepsilon)f(N) \leq f(n) \leq (1 + \varepsilon)f(N) \text{ for } N \leq n \leq 2N, \quad N > N(\varepsilon). \quad (2.10)$$

Theorem 4. *For every function $f \in \mathcal{F}$, $f(n) \leq \log n$, $\lim_{n \rightarrow \infty} f(n) = \infty$ we have an ineffective constant $c_f > 0$ such that*

$$[0, c_f] \subset J_f := \left\{ \frac{d_n}{f(n)} \right\}'. \quad (2.11)$$

2.4 Comparison of two consecutive values of d_n

Erdős and Turán proved 65 years ago [8, 1948] that $d_{n+1} - d_n$ changes sign infinitely often. This was soon improved by Erdős [6, 1948] to

$$\liminf_{n \rightarrow \infty} \frac{d_{n+1}}{d_n} < 1 < \limsup_{n \rightarrow \infty} \frac{d_{n+1}}{d_n}. \quad (2.12)$$

Erdős [7, 1955] wrote seven years later: "One would of course conjecture that

$$\liminf_{n \rightarrow \infty} \frac{d_{n+1}}{d_n} = 0, \quad \limsup_{n \rightarrow \infty} \frac{d_{n+1}}{d_n} = \infty, \quad (2.13)$$

but these conjectures seem very difficult to prove."

In Section 6 I will show this in a much stronger form:

Theorem 5. *We have*

$$\liminf_{n \rightarrow \infty} \frac{d_{n+1}/d_n}{(\log n)^{-1}} < \infty \quad (2.14)$$

and

$$\limsup_{n \rightarrow \infty} \frac{d_{n+1}/d_n}{\log n} > 0. \quad (2.15)$$

2.5 Arithmetic progressions in the sequence of generalized twin primes

Based on the method of I. M. Vinogradov [26, 1937] van der Corput [2, 1939] showed the existence of infinitely many 3-term arithmetic progressions in the sequence of primes. The problem of the existence of infinitely many k -term arithmetic progressions was open for all $k \geq 4$ until 2004, when B. Green and T. Tao [14, 2008] found their wonderful result that primes contain k -term arithmetic progressions for every k . I found recently [19, 2010] a common generalization of Green–Tao’s result and our Theorem B [12, 2010] under the deep assumption that there is a level $\vartheta > 1/2$ of the distribution of primes. Combining Zhang’s method with that of [19, 2010] I can prove now the following

Theorem 6. *There is a $d \leq 7 \times 10^7$ such that there are arbitrarily long arithmetic progressions of primes with the property that $p' = p + d$ is the prime following p for each element of the progression.*

3 Preparation for the proofs

All the proofs use some modified form of the conjecture of Dickson [3, 1904] about k -tuples of primes. (His original conjecture referred for linear forms with integer coefficients.) Let $\mathcal{H} = \{h_i\}_{i=1}^k$ be a k -tuple of different non-negative integers. We call \mathcal{H} admissible, if

$$P_{\mathcal{H}}(n) = \prod_{i=1}^k (n - h_i) \quad (3.1)$$

has no fixed prime divisor, that is, if the number $\nu_p(\mathcal{H})$ of residue classes covered by $\mathcal{H} \bmod p$ satisfies

$$\nu_p(\mathcal{H}) < p \text{ for } p \in \mathcal{P}. \quad (3.2)$$

This is equivalent to the fact that the singular series

$$\mathfrak{S}(\mathcal{H}) = \prod_p \left(1 - \frac{\nu_p(\mathcal{H})}{p}\right) \left(1 - \frac{1}{p}\right)^{-k} > 0. \quad (3.3)$$

Dickson conjectured that if \mathcal{H} is admissible, then all $n + h_i$ will be primes simultaneously for infinitely many values of n .

Hardy and Littlewood [16, 1923], probably unaware of Dickson’s conjecture, formulated a quantitative version of it, according to which

$$\pi_{\mathcal{H}}(x) = \sum_{\substack{n \leq x \\ n+h_i \in \mathcal{P}(1 \leq i \leq k)}} 1 = (\mathfrak{S}(\mathcal{H}) + o(1)) \frac{x}{\log^k x}. \quad (3.4)$$

In the work [11, 2009] we attacked (but missed by a hair’s breadth) the following weaker form of Dickson’s conjecture which I called

Conjecture DHL ($k, 2$). *If \mathcal{H} is an admissible k -tuple, then $n + \mathcal{H}$ contains at least two primes for infinitely many values of n .*

It is clear that if DHL($k, 2$) is proved for any k (or even for any single k -tuple \mathcal{H}_k), then the Bounded Gap Conjecture is true. DHL($k, 2$) was shown very recently by Y. T. Zhang [28, 2013] for $k \geq k_0 = 3.5 \times 10^6$ and this implied his Theorem A, the infinitude of gaps of size $\leq 7 \cdot 10^7$.

However, results of this type cannot exclude the existence of other primes and therefore give information on numbers expressible as difference of two primes, in the optimal case of Zhang’s very strong Theorem A prove the existence of many weak Polignac numbers. However, they do not

provide more information about (strong) Polignac numbers than the very deep fact that $\mathcal{D}_s \neq \emptyset$ and they do not help in showing any of Theorems 1–6. For example, in case of Theorem 6 they do not yield, let say, 4-term arithmetic progressions of primes and a bounded number d such that $p + d$ should also be prime for all four elements of the progression (even if we do not require that p and $p + d$ should be *consecutive* primes).

By a combination of the ideas of D. Goldston, C. Yıldırım, Y. Zhang and mine, I am able to show a much stronger form of Conjecture DHL($k, 2$) which may be applied towards the proof of Theorems 1–6. In case of Theorem 6 the ground-breaking ideas of Green and Tao [14, 2008] have to be used too, of course.

In view of Zhang’s recent result the stronger form of Conjecture DHL($k, 2$) will obtain already the name Theorem DHL*($k, 2$) for large enough k and can be formulated as follows, first still as a conjecture for $k < 3.5 \cdot 10^6$.

Let $P^-(m)$ denote the smallest prime factor of m .

Conjecture DHL* ($k, 2$). *Let $k \geq 2$ and $\mathcal{H} = \{h_i\}_{i=1}^k$ be any admissible k -tuple, $N \in \mathbb{Z}^+$, $\varepsilon > 0$ sufficiently small ($\varepsilon < \varepsilon_0$)*

$$\mathcal{H} \subset [0, H], \quad H \leq \varepsilon \log N, \quad P_{\mathcal{H}}(n) = \prod_{i=1}^k (n + h_i). \quad (3.5)$$

We have then positive constants $c_1(k)$ and $c_2(k)$ such that the number of integers $n \in [N, 2N)$ such that $n + \mathcal{H}$ contains at least two consecutive primes and almost primes in each components (i.e. $P^-(P_{\mathcal{H}}(n)) > n^{c_1(k)}$) is at least

$$c_2(k) \mathfrak{S}(\mathcal{H}) \frac{N}{\log^k N} \quad \text{for } N > N_0(\mathcal{H}). \quad (3.6)$$

One can see that we have a looser condition than in DHL($k, 2$) as far as the elements of \mathcal{H} are allowed to tend to infinity as fast as $\varepsilon \log N$. On the other hand we get stronger consequences as

- (i) we can prescribe that the two primes $n + h_i$ and $n + h_j$ in our k -tuple should be *consecutive*;
- (ii) we have almost primes in each component $n + h_i$;
- (iii) we get the lower estimate (3.6) for the number of the required n ’s with the above property.

The condition $n \in [N, 2N)$ makes usually no problem but in case of the existence of Siegel zeros some extra care is needed if we would like to have effective results (see Section 8 for this).

After this it is easy to formulate (but not to prove) our

MAIN THEOREM. *Conjecture DHL*($k, 2$) is true for $k \geq 3.5 \times 10^6$.*

Proof. Since this result contains Zhang’s Theorem and even more, it is easy to guess that a self contained proof would be hopelessly long (and difficult). We therefore try to describe only the changes compared to different earlier works.

The first pillar of Zhang’s work is the method of proof of our Theorem B. Although he supposes \mathcal{H} as a constant the method of proof of Theorem B (see Propositions 1 and 2 in [11, 2009]) allow beyond $H \ll \log N$ (required by (3.5) above) the much looser condition $H \ll N^{1/4-\varepsilon}$ for any $\varepsilon > 0$.

The second pillar of Zhang’s work is to show that distribution of primes according to non-smooth moduli, i.e. without any prime divisor $> N^b$ for any fixed small constant b , can be neglected. As mentioned in the Introduction we showed this already much earlier in a joint work with Y. Motohashi [18, 2008]. This work also supposed \mathcal{H} to be a constant but the only place where actually more care is needed in the proof is (3.11) of [18, 2008]. On the other hand, allowing here the condition

$$H \ll \log N, \quad (3.7)$$

the same simple argument as in Section 6 of [11, 2009] adds an additional error term

$$k^2 \log \log \log R \ll \log \log \log R \quad (3.8)$$

to the right-hand side of (3.11) of [18, 2008] which is far less than the present error term $\log R_0 = \log R / (\log \log R)^5$. Otherwise the proof works without any change, everything remains uniform under our condition (3.7) above.

Remark. The crucial Lemmas 3 and 4 of our work [18, 2008] contain an additional factor $\gamma(n, \mathcal{H})$. However, by the definition (4.17) of [18, 2008] we have $\gamma(n, \mathcal{H}) = 1$ if $P^-(P_{\mathcal{H}}(n)) > R^\eta$ for any fixed $\eta > 0$. In such a way the extra factor $\gamma(n, \mathcal{H})$ does not affect the validity of our Lemma 2 below since the asymptotic provided by Lemma 3 for the right-hand side of (3.11) (and similarly the analogue of it for primes, Lemma 4) is the same as if we used the constant weight 1 instead of $\gamma(n, \mathcal{H})$.

No change is required in the third pillar of Zhang's work where he proves some sort of extension of the Bombieri–Vinogradov theorem for smooth moduli and the residue classes appearing by the method of Theorem B.

However, the proof of Theorem DHL^{*}($k, 2$) (for k large enough, $k \geq 3.5 \cdot 10^6$) requires a further important idea, namely Lemmas 3–4 of the author's work [19, 2010]. This we formulate now as

Lemma 1. *Let $N^{c_0} < R \leq \sqrt{N/p}(\log N)^{-C}$, $p \in \mathcal{P}$, $p < R^{C_0}$ with a sufficiently small positive c_0 and sufficiently large C . Then we have with the notation*

$$\Lambda_R(n; \mathcal{H}, k + \ell) = \frac{1}{(k + \ell)!} \sum_{d \leq R, d | P_{\mathcal{H}}(n)} \mu(d) \left(\log \frac{R}{d} \right)^{k + \ell} \quad (\ell \leq k) \quad (3.9)$$

the relation

$$\sum_{\substack{n \in [N, 2N] \\ p | P_{\mathcal{H}}(n)}} \Lambda_R(n; \mathcal{H}, k + \ell)^2 \ll_k \frac{\log p}{p \log R} \sum_{n \in [N, 2N]} \Lambda_R(n; \mathcal{H}, k + \ell)^2. \quad (3.10)$$

Lemma 1 immediately implies

Lemma 2. *Let $N^{c_0} < R \leq N^{1/(2+\eta)}(\log N)^{-C}$, $\eta > 0$. We have then*

$$\sum_{\substack{n \in [N, 2N] \\ P^-(P_{\mathcal{H}}(n)) < R^\eta}} \Lambda_R(n; \mathcal{H}, k + \ell)^2 \ll_k \eta \sum_{n \in [N, 2N]} \Lambda_R(n; \mathcal{H}, k + \ell)^2. \quad (3.11)$$

Remark. Lemmas 1 and 2 were already proved in [19, 2010] under the loose condition $H \ll \log N$.

Lemma 2 asserts that numbers n where $P_{\mathcal{H}}(n)$ has a prime factor $< R^\eta$ (equivalently $< N^b$) with a small enough value of η (or b) might be neglected, since the weight used in all proofs is actually of type (3.9). The value of η (or b) depends on k .

These results play a crucial role in the common generalization of the Green–Tao theorem and of our Theorem B (cf. [19, 2010]) and also in the proof that prime gaps $< \varepsilon \log p$ form a positive proportion of all gaps for any $\varepsilon > 0$ (proved in a joint work with Goldston and Yıldırım).

These four pillars lead finally to the stronger form of Theorem DHL^{*}($k, 2$) if we combine it with a standard assertion following from Selberg's sieve, which we can formulate in this special case as

Lemma 3. *Let $0 < \alpha < 1/2$ be any constant. Then*

$$\sum_{\substack{n \in [N, 2N] \\ P^-(P_{\mathcal{H}}(n)) > N^\alpha}} 1 \ll_k \frac{N \alpha^{-k}}{\log^k N} \mathfrak{S}(\mathcal{H}). \quad (3.12)$$

Proof. This is Theorem 5.1 of [15, 1974] or Theorem 2 in §2.2.2 of [13, 2001]. This is also valid if we assume only $H \ll \log N$. \square

We further need a generalization of Gallagher's theorem proved by the author which we formulate as

Lemma 4. *Let \mathcal{H}_k be an arbitrary admissible k -tuple with*

$$\mathcal{H} = \mathcal{H}_k \subseteq [0, H]. \quad (3.13)$$

Then we have for any $\eta > 0$

$$S_{\mathcal{H}}(H) := \frac{1}{H} \sum_{h=1}^H \frac{\mathfrak{S}(\mathcal{H} \cup h)}{\mathfrak{S}(\mathcal{H})} = 1 + O(\eta) \quad (3.14)$$

if

$$H \geq \exp(k^{1/\eta}). \quad (3.15)$$

Proof. This is Theorem 1 of [20, 2010]. \square

Remark. As it is easy to see Lemma 4 implies Gallagher's classical theorem [10, 1976] on the singular series.

Combining the proofs of Theorems A and B (in the modified forms mentioned above) with the assertion of Lemma 2 we obtain under the weaker condition (3.6), i.e. for all admissible k -tuples $\mathcal{H}_k = \{h_i\}_{i=1}^k$, $h_i < h_{i+1}$,

$$\mathcal{H} = \mathcal{H}_k \subseteq [0, \varepsilon \log N], \quad (3.16)$$

at least

$$c_2(k) \mathfrak{S}(\mathcal{H}) \frac{N}{\log^k N} \quad (3.17)$$

numbers $n \in [N, 2N)$ such that $n + \mathcal{H}$ contains at least two primes and almost primes in each component, i.e.

$$P^-(P_{\mathcal{H}}(n)) > n^{c_1(k)}. \quad (3.18)$$

However, we must show the same with *consecutive* primes as well.

We can define for any subset V of $\{1, 2, \dots, k\}$ the set

$$V(N) = \{n \in [N, 2N) : n + h_i \in \mathcal{P} \Leftrightarrow i \in V\}. \quad (3.19)$$

Since k is bounded the number of possible subsets V is also bounded, therefore we can choose a V_0 such that

$$V_0 \subset \{1, 2, \dots, k\}, \quad |V_0| \geq 2, \quad |V_0(N)| \geq c_3(k) \mathfrak{S}(\mathcal{H}) \frac{N}{\log^k N}. \quad (3.20)$$

Choosing two arbitrary consecutive elements $i, j \in V_0(N)$ with $i < j$ we have at least

$$c_3(k) \mathfrak{S}(\mathcal{H}) \frac{N}{\log^k N} \quad (3.21)$$

numbers $n \in [N, 2N)$ such that for any $\mu \in (i, j)$

$$n + h_{\mu} \notin \mathcal{P}. \quad (3.22)$$

We have to assure, however, additionally that for a positive proportion of these numbers n we have also

$$n + h \notin \mathcal{P} \quad \text{for } h \notin \mathcal{H}, \quad h_i < h < h_j. \quad (3.23)$$

Applying Lemma 3 for all these values h and summing up we arrive at the conclusion that

$$\begin{aligned} \sum_{\substack{n \in [N, 2N] \\ P^-(P_{\mathcal{H}}(n)) > N^{c_1(k)}}} \sum_{\substack{h_i < h < h_j \\ n+h \in \mathcal{P}}} 1 &\leq C_4(k) \frac{N}{\log^{k+1} N} \sum_{h_i < h < h_j} \mathfrak{S}(\mathcal{H} \cup h) \\ &\leq 2C_4(k) \frac{N \varepsilon \mathfrak{S}(\mathcal{H})}{\log^k N} \end{aligned} \quad (3.24)$$

by Lemma 4. This shows that if ε was chosen sufficiently small depending on k , i.e.

$$\varepsilon < \varepsilon_0(k) \quad (3.25)$$

then our original primes $n + h_i$ and $n + h_j$ are *consecutive* for at least

$$c_5(k) \mathfrak{S}(\mathcal{H}) \frac{N}{\log^k N} \quad (3.26)$$

elements $n \in [N, 2N)$, thereby showing our Main Theorem. \square

4 Polignac numbers

1. Proof of Theorem 1

The fact that $\text{DHL}^*(k, 2)$ implies the positivity of the asymptotic lower density of Polignac numbers is expressed as Corollary 1 of [19, 2010] with the value

$$\frac{1}{k(k-1)} \prod_{p \leq k} \left(1 - \frac{1}{p}\right) \quad (4.1)$$

and proved in Section 11. The above value is about $e^{-\gamma}/(k^2 \log k)$ for large values of k . (If $k \rightarrow \infty$, they are asymptotically equal.)

1. Proof of Theorem 2

We again suppose $\text{DHL}^*(k, 2)$ for $k \geq k_0$.

The reasoning is more intricate in this case and the resulting value C is ineffective. The trivial relation

$$h_k - h_1 \geq (k-1) \min_{i>j} (h_i - h_j) \quad (h_i < h_{i+1}) \quad (4.2)$$

even gives the impression that the best we can prove about localization of Polignac numbers is an interval of type

$$[M, (k-1)M]. \quad (4.3)$$

Let us suppose that Theorem 2 is false. Then we have for any $C_0 > 0$ an infinite series of intervals

$$I_\nu := [M_\nu, M_\nu + C_\nu], \quad M_\nu > C_\nu > 4M_{\nu-1}, \quad M_1 > C_0 \quad (4.4)$$

such that

$$\mathcal{D}_s \cap \left(\bigcup_{\nu=1}^{\infty} I_\nu \right) = \emptyset. \quad (4.5)$$

For $p > k$ we have clearly

$$\nu_p(\mathcal{H}_k) < p \quad \text{for } p \in \mathcal{P}, \quad (4.6)$$

so we have no problem of choosing an admissible system \mathcal{H}_k in a sufficiently long interval (e.g. if C_k is large enough). Let

$$\mathcal{H}_k := \{h_\nu\}_{\nu=1}^k, \quad h_\nu \in I'_\nu := [M_\nu + C_\nu/2, M_\nu + C_\nu]. \quad (4.7)$$

For $h_\nu \in I_\nu$, $h_\mu \in I_\mu$, $\nu < \mu$ we have then

$$h_\mu - h_\nu \in [M_\mu + C_\mu/2 - 2M_{\mu-1}, M_\mu + C_\mu] \subset I_\mu. \quad (4.8)$$

Since in case of $k \geq k_0$ at least one of the numbers

$$h_\mu - h_\nu \quad (1 \leq \nu < \mu \leq k) \quad (4.9)$$

can be written as a difference of two consecutive primes, (4.8) contradicts to (4.5) and thus proves our theorem.

5 The normalized value distribution of d_n

1. Proof of Theorems 3–4

Since Theorem 4 implies Theorem 3 it is sufficient to prove the latter. The structure of the proof will follow that of Theorem 2 proved in the previous section. Suppose that Theorem 4 is false. In this case we have for a sufficiently small $c^* > 0$ an infinite series of intervals

$$J_\nu = [c_\nu, c_\nu + \delta_\nu], \quad c_\nu > 4\delta_\nu > 20c_{\nu+1}, \quad c_1 < c^* \quad (5.1)$$

such that for K large enough

$$\left\{ \frac{d_n}{f(n)} \right\}_{n=N}^\infty \cap \left(\bigcup_{\nu=1}^K J_\nu \right) = \emptyset, \quad \text{where } N = N(K) > 0. \quad (5.2)$$

Let

$$I_\nu(n) := [c_\nu f(n), (c_\nu + \delta_\nu)f(n)], \quad \nu = 1, 2, \dots, K. \quad (5.3)$$

Then we have

$$d_n \notin \bigcup_{\nu=1}^K I_\nu(n) \quad \text{for } \nu = 1, 2, \dots, K, \quad n \in [N, 2N), \quad N > N(K). \quad (5.4)$$

Using our Main Theorem, similarly to (4.7) we can construct an admissible k -tuple $\mathcal{H}_k = \{h_\nu\}_{\nu=1}^k$ with $3.5 \cdot 10^6 \leq k \leq K$ such that $h_1 > h_2 > \dots > h_k$ and with a sufficiently small $\varepsilon > 0$

$$h_\nu \in I'_\nu(N) := \left[\left(c_\nu + \frac{\delta_\nu}{2} \right) (1 + \varepsilon)f(N), (c_\nu + \delta_\nu)(1 - \varepsilon)f(N) \right]. \quad (5.5)$$

For $h_\mu \in I'_\mu$, $h_\nu \in I'_\nu$, $\mu < \nu$ we have for $N > \max(N(K), N_0(\varepsilon))$

$$h_\mu - h_\nu \in \left[\left(c_\mu + \frac{\delta_\mu}{2} - 2c_{\mu+1} \right) (1 + \varepsilon)f(N), (c_\mu + \delta_\mu)(1 - \varepsilon)f(N) \right] := I_\mu^*(N). \quad (5.6)$$

Since we have for any $1 \leq \mu < \nu \leq k_0$ and any $n \in [N, 2N)$

$$I_\mu^*(N) \subset I_\mu(n), \quad (5.7)$$

the fact that by the Main Theorem $h_\mu - h_\nu = d_n$ for some $n \in [N, 2N)$, $1 \leq \mu < \nu \leq k_0$ contradicts to (5.4) and thus proves Theorems 3 and 4.

6 Comparison of two consecutive values of d_n

1. Proof of Theorem 5

Since the proof of the two inequalities are completely analogous, we will only prove the second one. The basis of it is the Main Theorem proved in Section 3. We can start with an arbitrary admissible k -tuple $\mathcal{H} = \mathcal{H}_k = \{h_i\}_{i=1}^k$, $h_1 < h_2 < \dots < h_k$ with $k \geq 3.5 \cdot 10^6$. Let with a fixed sufficiently small $c_1(k)$ define

$$B(i, j, N) = \left\{ n \leq N, n + h_i \in \mathcal{P}, n + h_j \in \mathcal{P}, P^-(\mathcal{P}_{\mathcal{H}}(n)) > n^{c_1(k)} \right\}, \quad (6.1)$$

$$\mathcal{T} = \left\{ (i, j); j > i, \limsup_{N \rightarrow \infty} \frac{|B(i, j, N)| \log^k N}{N} > 0 \right\} \quad (6.2)$$

and let us choose the pair (i, j) with maximal value of j , afterwards that with maximal value of $i < j$. Then for any $h_\mu \in (h_i, h_j)$ (i.e. $i < \mu < j$) we have clearly

$$\limsup_{N \rightarrow \infty} \frac{|B(\mu, j, N)| \log^k N}{N} = 0 \quad (6.3)$$

so all components $n + h_\mu$ between $n + h_i$ and $n + h_j$ are almost always composite if $n \in B(i, j, N)$ and $N = N_\nu \rightarrow \infty$ through a suitable sequence N_ν .

On the other hand if we have an arbitrary $h \in (h_i, h_j)$, $h \notin \mathcal{H}$, then the assumption $n + h \in \mathcal{P}$ implies for $\mathcal{H}^+ = \mathcal{H} \cup h$

$$P^-(\mathcal{P}_{\mathcal{H}^+}(n)) > n^{c_1(k)}. \quad (6.4)$$

However, by Lemma 3 the number of such $n \leq N$ is for all N

$$\ll_{k, c_1} \frac{\mathfrak{S}(\mathcal{H} \cup \{h\})N}{\log^{k+1} N} \ll_{k, c_1} \frac{\mathfrak{S}(\mathcal{H})N \log h_k}{\log^{k+1} N} \ll_{k, c_1, \mathcal{H}} \frac{N}{\log^{k+1} N}. \quad (6.5)$$

This, together with (6.3) shows that we have at least

$$(c_1(k, \mathcal{H}) + o(1)) \frac{N}{\log^k N} \quad (6.6)$$

values $n \leq N$ with $n + h_i, n + h_j$ being consecutive primes for some sequence $N = N_\nu \rightarrow \infty$.

Let us consider now these differences. Let

$$n + h_i = p_\nu \in \mathcal{P}, \quad n + h_j = p_{\nu+1} \in \mathcal{P}, \quad d_\nu = h_j - h_i \ll 1 \quad (6.7)$$

where

$$\log n \sim \log \nu \sim \log N. \quad (6.8)$$

Suppose now that the second inequality of Theorem 5 is false. Then we have for all those values of ν with an arbitrary $\varepsilon > 0$ for $N > N(\varepsilon)$

$$d_{\nu+1} \leq d_\nu \varepsilon \log N \leq C \varepsilon \log N, \quad C = h_k - h_1. \quad (6.9)$$

The already quoted sieve of Selberg (Lemma 3) gives an upper estimate how often this might happen for any particular value

$$d_{\nu+1} = d \leq C \varepsilon \log N. \quad (6.10)$$

Adding it up until $C \varepsilon \log N$ we obtain at most

$$\ll_{k, c_1} \frac{N}{\log^{k+1} N} \sum_{h=1}^{C \varepsilon \log N} \mathfrak{S}(\mathcal{H} \cup h) \ll_{k, c_1} \mathfrak{S}(\mathcal{H}) \frac{CN}{\log^k N} \varepsilon \quad (6.11)$$

by Lemma 4. This means that in view of (6.6) this cannot hold for all $N = N_\nu \rightarrow \infty$. This contradiction proves Theorem 5.

7 Arbitrarily long arithmetic progressions of generalized twin primes

Proof of Theorem 6

In this case we have to use again our crucial Main Theorem and the rest of the machinery executed in Section 7 of [19, 2010]. This yields the combination of Zhang's theorem with that of Green and Tao [14, 2008].

8 How to make Zhang's theorem effective?

Our last point is that in its original form Zhang's theorem is ineffective since it uses Siegel's theorem. His result and similarly to it all results of the present work can be made effective in the following way.

According to the famous theorem of Landau–Page there is at most one real primitive character χ with a modulus q (and the characters induced by it) which might cause ineffectivity in the Bombieri–Vinogradov theorem, and this modulus has to satisfy

$$(\log X)^2 \leq q \leq (\log X)^{\omega(X)} \quad (8.1)$$

for any $\omega(X) \rightarrow \infty$ as $X \rightarrow \infty$. This modulus can cause any problem only in the case

$$q \mid P_{\mathcal{H}}(n). \quad (8.2)$$

However, our Lemma 3 says that we can neglect all numbers n with

$$P^-(P_{\mathcal{H}}(n)) < n^{c_1(k)}, \quad (8.3)$$

so (prescribing additionally $q \nmid d$ in the definition (3.9) of $\Lambda_R(n; \mathcal{H}, k, \ell)$) both Zhang's theorem and all our present results become effective.

Remark. It is an interesting phenomenon that in the sieving process yielding bounded gaps between primes and in all our present results we can choose

$$\lambda_d = 0 \quad (8.4)$$

if d has either a prime divisor

$$> N^{c'(k)} \quad (8.5)$$

or if it has a prime divisor

$$< N^{c''(k)}. \quad (8.6)$$

References

- [1] E. BOMBIERI, H. DAVENPORT, Small differences between prime numbers, *Proc. Roy. Soc. Ser. A* **293** (1966), 1–18.
- [2] J. G. VAN DER CORPUT, Über Summen von Primzahlen und Primzahlquadraten, *Math. Ann.* **116** (1939), 1–50.
- [3] L. E. DICKSON, A new extension of Dirichlet's theorem on prime numbers, *Messenger of Math.* (2), **33** (1904), 155–161.
- [4] P. D. T. A. ELLIOTT, H. HALBERSTAM, A conjecture in prime number theory, *Symposia Mathematica* Vol. 4 (1970) (INDAM, Rome, 1968/69), 59–72, Academic Press, London.

- [5] P. ERDŐS, The difference of consecutive primes, *Duke Math. J.* **6** (1940), 438–441.
- [6] P. ERDŐS, On the difference of consecutive primes, *Bull. Amer. Math. Soc.* (1948), 885–889.
- [7] P. ERDŐS, Some problems on the distribution of prime numbers, *Teoria dei Numeri, Math. Congr. Varenna, 1954*, 8 pp., 1955.
- [8] P. ERDŐS, P. TURÁN, On some new questions on the distribution of prime numbers, *Bull. Amer. Math. Soc.* **54** (1948), 371–378.
- [9] B. FARKAS, J. PINTZ AND SZ. GY. RÉVÉSZ, On the optimal weight function in the Goldston–Pintz–Yıldırım method for finding small gaps between consecutive primes, submitted.
- [10] P. X. GALLAGHER, On the distribution of primes in short intervals, *Mathematika* **23** (1976), 4–9.
- [11] D. A. GOLDSTON, J. PINTZ, C. YILDIRIM, Primes in tuples. I, *Ann. of Math. (2)* **170** (2009), no. 2, 819–862.
- [12] D. A. GOLDSTON, J. PINTZ, C. YILDIRIM, Primes in tuples. II, *Acta Math.* **204** (2010), no. 1, 1–47.
- [13] G. GREAVES, *Sieves in Number Theory*, Springer, Berlin, 2001.
- [14] B. GREEN, T. TAO, The primes contain arbitrarily long arithmetic progressions, *Ann. of Math. (2)* **167** (2008), no. 2, 481–547.
- [15] H. HALBERSTAM, H.-E. RICHERT, *Sieve methods*, Academic Press, New York, 1974.
- [16] G. H. HARDY, J. E. LITTLEWOOD, Some problems of ‘Partitio Numerorum’, III: On the expression of a number as a sum of primes, *Acta Math.* **44** (1923), 1–70.
- [17] H. MAIER, Small differences between prime numbers, *Michigan Math. J.* **35** (1988), 323–344.
- [18] Y. MOTOHASHI, J. PINTZ, A smoothed GPY sieve, *Bull. London Math. Soc.* **40** (2008), no. 2, 298–310.
- [19] J. PINTZ, Are there arbitrarily long arithmetic progressions in the sequence of twin primes? *An irregular mind. Szemerédi is 70*. Bolyai Soc. Math. Stud. 21, Springer, 2010, pp. 525–559.
- [20] J. PINTZ, On the singular series in the prime k -tuple conjecture, arXiv: 1004.1084v1 [math.NT]
- [21] J. PINTZ, Some new results on gaps between consecutive primes, *Turán Memorial, Number Theory, Combinatorics and Analysis*, de Gruyter, Berlin, to appear.
- [22] J. PINTZ, *Paul Erdős and the Difference of Primes. Erdős Centennial*, Bolyai Soc. Math. Stud., Springer, 2013, to appear.
- [23] A. DE POLIGNAC, Six propositions arithmologiques déduites de crible d’Ératosthène, *Nouv. Ann. Math.* **8** (1849), 423–429.
- [24] R. A. RANKIN, The difference between consecutive prime numbers. II, *Proc. Cambridge Philos. Soc.* **36** (1940), 255–266.
- [25] G. RICCI, Sull’andamento della differenza di numeri primi consecutivi, *Riv. Mat. Univ. Parma* **5** (1954), 3–54.

- [26] I. M. VINOGRADOV, Representation of an odd number as a sum of three prime numbers, *Doklady Akad. Nauk. SSSR* **15** (1937), 291–294 (Russian).
- [27] E. WESTZYNTHIUS, Über die Verteilung der Zahlen, die zu der n ersten Primzahlen teilerfremd sind, *Comm. Phys. Math. Helsingfors (5)* **25** (1931), 1–37.
- [28] YTANG ZHANG, Bounded gaps between primes, *Ann. of Math.*, to appear.