

On the difference of primes

by
János Pintz

1 Introduction

In the present work we investigate some approximations to generalizations of the twin prime conjecture. The twin prime conjecture appeared in print already the first time in a more general form, due to de Polignac [Pol] in 1849:

Conjecture C1. *Every even number can be written in infinitely many ways as the difference of two consecutive primes.*

Kronecker [Kro] mentioned in 1901 the same conjecture in a weaker form as

Conjecture C2. *Every even number can be expressed in infinitely many ways as the difference of two primes.*

Finally, Maillet [Mai] formulated it in 1905 as

Conjecture C3. *Every even number is the difference of two primes.*

We remark that it is easy to see that the existence of at least one even number satisfying either C1 or C2 is equivalent to

Conjecture C4 (Bounded Gap Conjecture). *If p_n denotes the n^{th} prime then*

$$(1.1) \quad \liminf_{n \rightarrow \infty} (p_{n+1} - p_n) < \infty.$$

Supported by OTKA (Grants No. K72731, K67676) and ERC/AdG. 228005.

We will therefore concentrate on Conjecture C3 which shows the closest analogy to Goldbach's conjecture among Conjectures C1–C3.

Let us call n a Goldbach number if it is the sum of two primes and a Maillet number if it is the difference of two primes. As an approximation to the conjectures of Goldbach and Maillet (C3) one can ask how long an interval can be if it contains no Goldbach numbers (i.e. no sum of two primes) or, no Maillet numbers (i.e. no difference of two primes), respectively. Montgomery–Vaughan [MV] and Ramachandra [Ram] observed that if for some $0 \leq \vartheta_1, \vartheta_2 \leq 1$

$$(1.2) \quad \pi(x + x^{\vartheta_1}) - \pi(x) \geq \frac{cx^{\vartheta_1}}{\log x}$$

and

$$(1.3) \quad \pi(n + n^{\vartheta_2}) - \pi(n) > 0 \text{ for almost all } n \in [x, 2x) \text{ where } x \rightarrow \infty$$

then for $x > x_0$ the interval

$$(1.4) \quad I = [x, x + C'x^{\vartheta}], \quad \vartheta = \vartheta_1\vartheta_2$$

contains at least one Goldbach number. The sharpest known results, $\vartheta_1 = 21/40$ by R. C. Baker, G. Harman and the author [BHP] and $\vartheta_2 = 1/20$ of Ch. Jia [Jia] imply that I contains Goldbach numbers if $\vartheta = 21/800$. The same method applies to Maillet numbers without any change. Thus the interval

$$(1.5) \quad I^* = [x, x + Cx^{21/800}], \quad x > x_0$$

contains an even integer which can be written as the difference of two primes.

Under supposition of the Riemann Hypothesis (RH) it was proved by Linnik [Lin], later by Kátai [Kat] that the interval

$$(1.6) \quad I(C_2) = [x, x + C_1(\log x)^{C_2}]$$

contains Goldbach numbers for $C_2 > 3$ [Lin], respectively for $C_2 = 2$ [Kat].

We announce the unconditional

Theorem 1. *A positive proportion of even numbers in an interval of type $[x, x + (\log x)^C]$ can be written as the difference of two primes if $C > C_0$ and $x > x_0$.*

In order to illustrate the method we will prove here a result which shows that the best known exponent $21/800$ (see (1.5)) can be replaced by an arbitrary positive number.

Theorem 2. *Let $\varepsilon > 0$ be arbitrary. The interval $[x, x + x^\varepsilon]$ contains even numbers which can be written as the difference of two primes if $x > x_0(\varepsilon)$.*

It was shown in [GPY] that the Bounded Gap Conjecture C4 is true, equivalently there is at least one de Polignac number if primes have an admissible level $\vartheta > 1/2$ of distribution. This means that

$$(1.7) \quad \sum_{q \leq x^{\vartheta-\varepsilon}} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{p \equiv a \pmod{q} \\ p \leq x}} \log p - \frac{x}{\varphi(q)} \right| \ll_{\varepsilon, A} \frac{x}{(\log x)^A}$$

for any $\varepsilon > 0$ and $A > 0$. In [Pin1] it was proved that if $\vartheta > 1/2$ then de Polignac numbers have a positive (lower) density. We announce here the stronger but still conditional

Theorem 3. *If primes have an admissible level $\vartheta > 1/2$ of distribution then there exists a constant $C(\vartheta)$ such that for $x > C_0(\vartheta)$ the interval $[x, x + C(\vartheta)]$ contains at least one number which can be written in infinitely many ways as the difference of two consecutive primes.*

About 60 years ago Erdős [Erd] and Ricci [Ric] independently proved that the set J of limit points of the sequence

$$(1.8) \quad \frac{p_{n+1} - p_n}{\log p_n}, \text{ or equivalently that of } \frac{p_{n+1} - p_n}{\log n}$$

has positive Lebesgue measure but no finite point of J was known until [GPY], which implied $0 \in J$.

Supposing $\vartheta > 1/2$ we can show the much stronger

Theorem 4. *Let $g(n) < \log n$ be any monotonically increasing positive function with $\lim_{n \rightarrow \infty} g(n) = \infty$. Let us suppose that primes have an admissible level $\vartheta > 1/2$ of distribution. Then we have a constant $c(g, \vartheta)$ such that*

$$(1.9) \quad [0, c(g, \vartheta)] \subset J.$$

The details of proofs for Theorems 1, 3 and 4 will appear elsewhere. We remark here that their proofs (similar to that of Theorem 2) will be completely ineffective (independently of some eliminable ineffectivity originating from the use of Bombieri–Vinogradov theorem, which uses the ineffective theorem of Siegel for \mathcal{L} -zeros).

2 Notation and lemmata

Let \mathcal{P} denote the set of primes. Let $\varepsilon > 0$ and let C_0 be a sufficiently large constant depending on ε . Let us suppose the existence of an infinite sequence $(H_\nu \in \mathbb{Z})$

$$(2.1) \quad I_\nu = [H_\nu, H_\nu + H_\nu^\varepsilon], \quad H_\nu^\varepsilon > 2H_{\nu-1}, H_1 > C_0$$

such that

$$(2.2) \quad (\mathcal{P} - \mathcal{P}) \cap \left(\bigcup_{\nu=1}^{\infty} I_\nu \right) = \emptyset.$$

Let

$$(2.3) \quad k = \lceil (6/\varepsilon)^2 \rceil.$$

We call a k -tuple $\mathcal{H} = \{h_i\}_{i=1}^k$, $0 \leq h_1 < h_2 < \dots < h_k$ ($h_i \in \mathbb{Z}$) admissible if the number $\nu_p(\mathcal{H})$ of residue classes covered by $\mathcal{H} \pmod p$ satisfies

$$(2.4) \quad \nu_p(\mathcal{H}) < p \text{ for } p \in \mathcal{P}.$$

In this case the corresponding singular series is positive, i.e.

$$(2.5) \quad \mathfrak{S}(\mathcal{H}) := \prod_p \left(1 - \frac{1}{p}\right)^{-k} \left(1 - \frac{\nu_p(\mathcal{H})}{p}\right) > 0.$$

Since (2.4) is trivially true for $p > k$ it is easy to choose an admissible system

$$(2.6) \quad \mathcal{H}_k = \{h_\nu\}_{\nu=1}^k, \quad h_\nu \in I'_\nu := [H_\nu + H_\nu^\varepsilon/2, H_\nu + H_\nu^\varepsilon]$$

if C_0 was chosen large enough (depending on ε). This system $\mathcal{H}_k = \mathcal{H}$ will be fixed for the rest of the work.

We now choose any sufficiently large $t > \nu_0(\varepsilon, C_0)$ and denote

$$(2.7) \quad T := H_t^\varepsilon/2, \quad N := H_t + T, \quad I'_t := [N, N + T] \subset I_t.$$

It will be very important that our condition (2.1) guarantees that for $h_\nu \in I_\nu$, $h_\mu \in I_\mu$, $\nu < \mu$ we have

$$(2.8) \quad h_\mu - h_\nu \geq H_\mu + H_\mu^\varepsilon/2 - H_{\mu-1} \geq H_\mu, \quad \text{so } h_\mu - h_\nu \in I_\mu.$$

We will use this relation for $\nu, \mu \in [1, k] \cup \{t\}$.

We will choose our weights a_n depending on $\mathcal{H} = \mathcal{H}_k$ exactly as in [GPY]:

$$(2.9) \quad a_n := \Lambda_R(n; \mathcal{H}, \ell)^2 := \left(\frac{1}{(k + \ell)!} \sum_{\substack{d|P_{\mathcal{H}}(n) \\ d \leq R}} \mu(d) \left(\log \frac{R}{d} \right)^{k+\ell} \right)^2,$$

where we define

$$(2.10) \quad P_{\mathcal{H}}(n) := \prod_{i=1}^k (n + h_i), \quad \ell = \left\lceil \frac{\sqrt{k}}{2} \right\rceil, \quad \mathcal{L} = \log N, \quad R = N \exp \left(-\sqrt{\log N} \right).$$

Let further $\chi_{\mathcal{P}}(n)$ denote the characteristic function of the primes. The notation $n \sim N$ will abbreviate $n \in [N, 2N]$.

We remark that for any admissible k -tuple \mathcal{H} we have

$$(2.11) \quad \mathfrak{S}(\mathcal{H}) \geq \prod_{p \leq 2k} \frac{1}{p} \prod_{p > 2k} \left(1 - \frac{k}{p} \right) \left(1 - \frac{1}{p} \right)^{-k} = c_1(k).$$

In the following formulae the o symbol will refer to the case $N \rightarrow \infty$ which is equivalent to $t \rightarrow \infty$. We will use the notation

$$(2.12) \quad A := \sum_{n \sim N} a_n, \quad B := B_R(N, k, \ell) := \binom{2\ell}{\ell} \frac{\log^{2k+\ell} R}{(k + 2\ell)!}.$$

The following lemmata are special cases of Propositions 1 and 2 of [GPY]. The only change is that although Proposition 2 has originally a condition $h_i \leq R$, this can in fact be replaced without any change in the proof by $h_i \leq 4N$, for example. Lemma 2.4 is a well-known sieve estimate (see Theorem 4.4 of [Mon], for example).

Lemma 1. $A = (\mathfrak{S}(\mathcal{H}) + o(1))B$.

Lemma 2. *If $h_i \in \mathcal{H}$ then*

$$(2.13) \quad S_i := \sum_{n \sim N} a_n \chi_{\mathcal{P}}(n + h_i) = \frac{2\ell + 1}{2\ell + 2} \frac{(\mathfrak{S}(\mathcal{H}) + o(1))B}{k + 2\ell + 1}.$$

Lemma 3. *If $h_0 \notin \mathcal{H}$, $h_0 \leq 4N$ then*

$$(2.14) \quad S_0 := S(h_0) := \sum_{n \sim N} a_n \chi_{\mathcal{P}}(n + h_0) = \frac{(\mathfrak{S}(\mathcal{H} \cup \{h_0\}) + o(1))B}{\mathcal{L}}.$$

Lemma 4. *$\pi(x + y) - \pi(x) \leq \frac{2y}{\log y}$ for any $x, y \geq 1$.*

Our last lemma is Lemma 3.4 of [Pin3] (and a slight variant of Theorem 2 of [Pin2]).

Lemma 5. *Let $\mathcal{H}_k \subset [0, H]$ be a fixed k -element admissible set, $k \geq k_0$, C a sufficiently large constant, $m \in \mathbb{Z}$, $\delta > 0$. Then, for $H > \exp\left(\frac{Ck}{\delta \log k}\right)$ we have*

$$(2.15) \quad S_{\mathcal{H}}(M, H) = \frac{1}{H} \sum_{m \in [M, M+H]} \frac{\mathfrak{S}(\mathcal{H} \cup \{m\})}{\mathfrak{S}(\mathcal{H})} \geq 1 - \delta.$$

3 Proof of Theorem 2

We begin with the very important remark that (cf. (2.2) and (2.8)) due to our definition of I and to $h_\mu - h_\nu \in I_\mu$ (for $\mu > \nu$), the set $\{n + h_i\}_{i=1}^k$ contains at most one prime for any given n . Let

$$(3.1) \quad \mathcal{D}_0 = \{n \sim N; n + h_i \notin \mathcal{P}, i = 1, 2, \dots, k\}, \quad \mathcal{D}_1 = [N, 2N] \setminus \mathcal{D}_0.$$

Then we have by (2.10)–(2.12) and Lemmas 1 and 2, for $k > k_0$

$$(3.2) \quad \begin{aligned} A_0 &:= \sum_{n \in \mathcal{D}_0} a_n = A - \sum_{i=1}^k S_i \\ &= (\mathfrak{S}(\mathcal{H}) + o(1))B \left(1 - \left(1 - \frac{1}{2\ell + 2}\right) \left(1 - \frac{2\ell + 1}{k + (2\ell + 1)}\right)\right) \\ &\leq \frac{7\mathfrak{S}(\mathcal{H})B}{3\sqrt{k}}. \end{aligned}$$

The second important remark is that if $n \in \mathcal{D}_1$ and $m \in I'_t$ then $n + m \notin \mathcal{P}$. Namely, $n \in \mathcal{D}_1$ and $n + m \in \mathcal{P}$ would imply $n + h_i \in \mathcal{P}$ for some i ; consequently

$$(3.3) \quad m - h_i \in \mathcal{P} - \mathcal{P},$$

which is a contradiction to (2.2) in view of $m \in I_t$. Hence,

$$(3.4) \quad S := \sum_{n \in \mathcal{D}_0} a_n \sum_{m \in I'_t} \chi_{\mathcal{P}}(n+m) = \sum_{m \in I_t} \sum_{n \sim N} a_n \chi_{\mathcal{P}}(n+m).$$

Considering the LHS, from (3.2), (2.3) and Lemma 4 we obtain

$$(3.5) \quad S \leq A_0 \cdot \frac{2T}{\log T} < \frac{5\mathfrak{S}(\mathcal{H})BT}{\varepsilon \mathcal{L} \sqrt{k}} \leq \frac{5\mathfrak{S}(\mathcal{H})BT}{6\mathcal{L}}.$$

On the other hand, considering the RHS, Lemmas 3 and 5 imply

$$(3.6) \quad S \geq \sum_{m \in I_t} \frac{(\mathfrak{S}(\mathcal{H} \cup \{m\}) + o(1))B}{\mathcal{L}} > \frac{5\mathfrak{S}(\mathcal{H})BT}{6\mathcal{L}},$$

which contradicts (3.5). This proves Theorem 2.

References

- [BHP] Baker, R. C., Harman, G., Pintz, J., The difference between consecutive primes, II., *Proc. London Math. Soc.* (3) 83 (2001), no. 3, 532–562.
- [Erd] Erdős, P., Some problems on the distribution of prime numbers, *C.I.M.E. Teoria dei numeri, Math. Congr. Varenna, 1955*, 8 pp., Roma: Istituto Matematico dell'Università 1955. [Zbl 0067.2704].
- [GPY] Goldston, D. A., Pintz, J., Yıldırım, C. Y., Primes in tuples. I, *Ann. of Math.* 170 (2009), 819–862.
- [Jia] Jia, Chaohua, Almost all short intervals containing prime numbers, *Acta Arith.* 76 (1996), 21–84.
- [Kat] Kátai, I., A comment on a paper of Ju. V. Linnik. (Hungarian), *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl.* 17 (1967), 99–100.
- [Kro] Kronecker, L., *Vorlesungen über Zahlentheorie, I.*, p. 68, Teubner, Leipzig, 1901.
- [Lin] Linnik, Yu. V., Some conditional theorems concerning the binary Goldbach problem. (Russian), *Izv. Akad. Nauk SSSR* 16 (1952), 503–520.

- [Mai] Maillet, E., L'intermédiaire des math. 12 (1905), p. 108.
- [Mon] Montgomery, H. L., *Topics in Multiplicative Number Theory*, Lecture Notes in Math. 227, Springer, Berlin, Heidelberg, New York, 1971.
- [MV] Montgomery, H. L., Vaughan, R. C., The exceptional set in Goldbach's problem, *Acta Arith.* 27 (1975), 353–370.
- [Pin1] Pintz, J., Are there arbitrarily long arithmetic progressions in the sequence of twin primes?, in: *An irregular mind. Szemerédi is 70*, Bolyai Soc. Math. Studies, Vol. 21. Eds.: I. Bárány, J. Solymosi, Springer, 2010. pp. 525–559.
- [Pin2] Pintz, J., On the singular series in the prime k-tuple conjecture, preprint, arXiv: 1004.10841 [mathNT]
- [Pin3] Pintz, J., The bounded gap conjecture and bounds between consecutive Goldbach numbers, *Acta Arith.*, to appear.
- [Pol] Polignac, A. de, Recherches nouvelles sur les nombres premiers, *Comptes Rendus Acad. Sci. Paris* 29 (1849), 397–401, Rectification: *ibid.* pp.738–739.
- [Ram] K. Ramachandra, On the number of Goldbach numbers in small intervals, *J. Indian Math. Soc. (N.S.)* 37 (1973), 157–170.
- [Ric] Ricci, Giovanni, Recherches sur l'allure de la suite $\{(p_{n+1}-p_n)/\log p_n\}$. (French), Colloque sur la Théorie des Nombres, Bruxelles, 1955, pp. 93–106. Georges Thone, Liège; Masson and Cie, Paris, 1956.

János Pintz

Rényi Mathematical Institute of the Hungarian Academy of Sciences

Budapest, Reáltanoda u. 13–15

H-1053 Hungary

E-mail: pintz@renyi.hu