

SMALL GAPS BETWEEN PRIMES

Nasr-allah Hitar

Abstract

Astract. in this paper , We will prove that

$$\liminf_n (\sqrt{p_{n+1}} - \sqrt{p_n}) < 1$$

We introduce a refinement of the (james maynard method) GPY sieve method for studying prime k -tuples and small gaps between primes. This refinement avoids previous limitations of the method, and allows us to show that for each k , the prime k -tuples conjecture holds for a positive proportion of admissible k -tuples. In particular, $\liminf_n (p_{n+m} - p_n) < \infty$ for every integer m . We also show that $\liminf (p_{n+1} - p_n) \leq 600$, and, if we assume the Elliott-Halberstam conjecture, that $\liminf (p_{n+1} - p_n) \leq 12$ and $\liminf_n (p_{n+2} - p_n) \leq 600$.

1 Introduction (James maynard method to prove that $\liminf_n (p_{n+2} - p_n) \leq 600$.)

We say that a set $\mathcal{H} = \{h_1, \dots, h_k\}$ of distinct non-negative integers is 'admissible' if, for every prime p , there is an integer a_p such that $a_p \not\equiv h \pmod{p}$ for all $h \in \mathcal{H}$. We are interested in the following conjecture.

Conjecture (Prime k -tuples conjecture). Let $\mathcal{H} = \{h_1, \dots, h_k\}$ be admissible. Then there are infinitely many integers n such that all of $n + h_1, \dots, n + h_k$ are prime.

When $k > 1$ no case of the prime k -tuples conjecture is currently known. Work on approximations to the prime k -tuples conjecture has been very successful in showing the existence of small gaps between primes, however. In their celebrated paper [5], Goldston, Pintz and Yıldırım introduced a new method for counting tuples of primes, and this allowed them to show that

$$\liminf_n \frac{p_{n+1} - p_n}{\log p_n} = 0.$$

The recent breakthrough of Zhang [9] managed to extend this work to prove

$$\liminf_n (p_{n+1} - p_n) \leq 70000000,$$

thereby establishing for the first time the existence of infinitely many bounded gaps between primes. Moreover, it follows from Zhang's theorem that the number of admissible sets of size 2 contained in $[1, x]^2$ which satisfy the prime 2-tuples conjecture is $\gg x^2$ for large x . Thus, in this sense, a positive proportion of admissible sets of size 2 satisfy the prime 2-tuples conjecture. The recent polymath project [7] has succeeded in reducing the bound (1.2) to 4680, by optimizing Zhang's arguments and introducing several new refinements.

The above results have used the 'GPY method' to study prime tuples and small gaps between primes, and this method relies heavily on the distribution of primes in arithmetic progressions. Given $\theta > 0$, we say the primes have 'level of distribution θ ' if, for every $A > 0$, we have

$$\sum_{q \leq x^\dagger} \max_{(a, q)=1} \left| \pi(x; q, a) - \frac{\pi(x)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^A}.$$

¹ We note that different authors have given slightly different names or definitions to this concept. For the purposes of this paper, (1.3) will be our definition of the primes having level of distribution θ . Theorem 1.3. We have

$$\liminf_n (p_{n+1} - p_n) \leq 600.$$

We emphasize that the above result does not incorporate any of the technology used by Zhang to establish the existence of bounded gaps between primes. The proof is essentially elementary, relying only on the Bombieri-Vinogradov theorem. Naturally, if we assume that the primes have a higher level of distribution, then we can obtain stronger results.

Theorem 1.4. Assume that the primes have level of distribution θ for every $\theta < 1$. Then

$$\liminf_n (p_{n+1} - p_n) \leq 12,$$

$$\liminf_n (p_{n+2} - p_n) \leq 600.$$

Although the constant 12 of Theorem 1.4 appears to be optimal with our method in its current form, the constant 600 appearing in Theorem 1.3 and Theorem 1.4 is certainly not optimal. By performing further numerical calculations our method could produce a better bound, and also most of the ideas of Zhang's work (and the refinements produced by the polymath project) should be able to be combined with this method to reduce the constant further. We comment that the assumption of the Elliott-Halberstam conjecture allows us to improve the bound on Theorem 1.1 to $O(m^3 e^{2m})$.

2 An IMPROVED GPY SIEVE METHOD

We first give an explanation of the key idea behind our new approach. The basic idea of the GPY method is, for a fixed admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$, to consider the sum

$$S(N, \rho) = \sum_{N \leq n < 2N} \left(\sum_{i=1}^k \chi_{\mathbb{P}}(n + h_i) - \rho \right) w_n.$$

Here $\chi_{\mathbb{P}}$ is the characteristic function of the primes, $\rho > 0$ and w_n are non-negative weights. If we can show that $S(N, \rho) > 0$ then at least one term in the sum over n must have a positive contribution. By the non-negativity of w_n , this means that there must be some integer $n \in [N, 2N]$ such that at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime. (Here $\lfloor x \rfloor$ denotes the largest integer less than or equal to x .) Thus if $S(N, \rho) > 0$ for all large N , there are infinitely many integers n for which at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime (and so there are infinitely many bounded length intervals containing $\lfloor \rho + 1 \rfloor$ primes).

The weights w_n are typically chosen to mimic Selberg sieve weights. Estimating (2.1) can be interpreted as a ' k -dimensional' sieve problem. The standard Selberg k -dimensional weights (which can be shown to be essentially optimal in other contexts) are

$$w_n = \left(\sum_{\substack{d | \prod_{i=1}^k (n + h_i) \\ d < R}} \lambda_d \right)^2, \quad \lambda_d = \mu(d) (\log R/d)^k.$$

With this choice we find that we just fail to prove the existence of bounded gaps between primes if we assume the Elliott-Halberstam conjecture. The key new idea in the paper of Goldston, Pintz and Yıldırım [5] was to consider more general sieve weights of the form

$$\lambda_d = \mu(d) F(\log R/d),$$

for a suitable smooth function F . Goldston, Pintz and Yıldırım chose $F(x) = x^{k+l}$ for suitable $l \in \mathbb{N}$, which has been shown to be essentially optimal when k is large. This allows us to gain a factor of approximately 2 for large k over the previous choice of sieve weights. As a result we just fail to prove bounded gaps using the fact that the primes have exponent of distribution θ for any $\theta < 1/2$, but succeed in doing so if we assume they have level of distribution $\theta > 1/2$.

The new ingredient in our method is to consider a more general form of the sieve weights

$$w_n = \left(\sum_{d_i | n + h_i \downarrow i} \lambda_{d_1, \dots, d_k} \right)^2.$$

Using such weights with $\lambda_{d_1, \dots, d_k}$ is the key feature of our method. It allows us to improve on the previous choice of sieve weights by an arbitrarily large factor, provided that k is sufficiently large. It is the extra flexibility gained by allowing the weights to depend on the divisors of each factor individually which gives this improvement.

The idea to use such weights is not entirely new. Selberg [8, Page 245] suggested the possible use of similar weights in his work on approximations to the twin prime problem, and Goldston and Yıldırım [6] considered similar weights in earlier work on the GPY method, but with the support restricted to $d_i < R^{1/k}$ for all i .

We comment that our choice of $\lambda_{d_1, \dots, d_k}$ will look like

$$\lambda_{d_1, \dots, d_k} \approx \left(\prod_{i=1}^k \mu(d_i) \right) f(d_1, \dots, d_k),$$

for a suitable smooth function f . For our precise choice of $\lambda_{d_1, \dots, d_k}$ (given in Proposition 4.1) we find it convenient to give a slightly different form of $\lambda_{d_1, \dots, d_k}$, but weights of the form (2.5) should produce essentially the same results.

3 Notation

We shall view k as a fixed integer, and $\mathcal{H} = \{h_1, \dots, h_k\}$ as a fixed admissible set. In particular, any constants implied by the asymptotic notation o, O or \ll may depend on k and \mathcal{H} . We will let N denote a large integer, and all asymptotic notation should be interpreted as referring to the limit $N \rightarrow \infty$.

All sums, products and suprema will be assumed to be taken over variables lying in the natural numbers $\mathbb{N} = \{1, 2, \dots\}$ unless specified otherwise. The exception to this is when sums or products are over a variable p , which instead will be assumed to lie in the prime numbers $\mathbb{P} = \{2, 3, \dots\}$.

Throughout the paper, φ will denote the Euler totient function, $\tau_r(n)$ the number of ways of writing n as a product of r natural numbers and μ the Moebius function. We will let ϵ be a fixed positive real number, and we may assume without further comment that ϵ is sufficiently small at various stages of our argument. We let p_n denote the n^{th} prime, and $\#\mathcal{A}$ denote the number of elements of a finite set \mathcal{A} . We use $\lfloor x \rfloor$ to denote the largest integer $n \leq x$, and $\lceil x \rceil$ the smallest integer $n \geq x$. We let (a, b) be the greatest common divisor of integers a and b . Finally, $[a, b]$ will denote the closed interval on the real line with endpoints a and b , except for in Section 5 where it will denote the least common multiple of integers a and b instead.

4 OUTLINE OF THE PROOF

We will find it convenient to choose our weights w_n to be zero unless n lies in a fixed residue class $v_0(\text{mod } W)$, where $W = \prod_{p \leq D_0} p$. This is a technical modification which removes some minor complications in dealing with the effect of small prime factors. The precise choice of D_0 is unimportant, but it will suffice to choose

$$D_0 = \log \log \log N,$$

so certainly $W \ll (\log \log N)^2$ by the prime number theorem. By the Chinese remainder theorem, we can choose v_0 such that $v_0 + h_i$ is coprime to W for each i since \mathcal{H} is admissible. When $n \equiv v_0 \pmod{W}$, we choose our weights w_n of the form (2.4). We now wish to estimate the sums

$$S_1 = \sum_{\substack{N \leq n < 2N \\ n \equiv v_0 \\ (\text{mod } W)}} \left(\sum_{d_i | n + h_i \forall i} \lambda_{d_1, \dots, d_k} \right)^2,$$

$$S_2 = \sum_{\substack{N \leq n < 2N \\ n \equiv v_0 \\ (\text{mod } W)}} \left(\sum_{(\text{mod } W)}^k \chi_{\mathbb{P}}(n + h_i) \right) \left(\sum_{d_i | n + h_i \forall i} \lambda_{d_1, \dots, d_k} \right)^2.$$

We evaluate these sums using the following proposition.

Proposition 4.1. Let the primes have exponent of distribution $\theta > 0$, and let $R = N^{\theta/2-\delta}$ for some small fixed $\delta > 0$. Let $\lambda_{d_1, \dots, d_k}$ be defined in terms of a fixed smooth function F by

$$\lambda_{d_1, \dots, d_k} = \left(\prod_{i=1}^k \mu(d_i) d_i \right) \sum_{\substack{r_1, \dots, r_k \\ d_i | r_i \forall i \\ (r_i, W) = 1 \forall i}} \frac{\mu \left(\prod_{i=1}^k r_i \right)^2}{\prod_{i=1}^k \varphi(r_i)} F \left(\frac{\log r_1}{\log R}, \dots, \frac{\log r_k}{\log R} \right),$$

whenever $\left(\prod_{i=1}^k d_i, W \right) = 1$, and let $\lambda_{d_1, \dots, d_k} = 0$ otherwise. Moreover, let F be supported on $\mathcal{R}_k = \left\{ (x_1, \dots, x_k) \in [0, 1]^k : \sum_{i=1}^k x_i \leq 1 \right\}$. Then we have

$$S_1 = \frac{(1 + o(1)) \varphi(W)^k N (\log R)^k}{W^{k+1}} I_k(F)$$

$$S_2 = \frac{(1 + o(1)) \varphi(W)^k N (\log R)^{k+1}}{W^{k+1} \log N} \sum_{m=1}^k J_k^{(m)}(F),$$

provided $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$ for each m , where

$$I_k(F) = \int_0^1 \cdots \int_0^1 F(t_1, \dots, t_k)^2 dt_1 \cdots dt_k,$$

$$J_k^{(m)}(F) = \int_0^1 \cdots \int_0^1 \left(\int_0^1 F(t_1, \dots, t_k) dt_m \right)^2 dt_1 \cdots dt_{m-1} dt_{m+1} \cdots dt_k.$$

We recall that if S_2 is large compared to S_1 , then using the GPY method we can show that there are infinitely many integers n such that several of the $n + h_i$ are prime. The following proposition makes this precise.

Proposition 4.2. Let the primes have level of distribution $\theta > 0$. Let $\delta > 0$ and $\mathcal{H} = \{h_1, \dots, h_k\}$ be an admissible set. Let $I_k(F)$ and $J_k^{(m)}(F)$ be given

as in Proposition 4.1 and let \mathcal{S}_k denote the set of Riemann-integrable functions $F : [0, 1]^k \rightarrow \mathbb{R}$ supported on $\mathcal{R}_k = \left\{ (x_1, \dots, x_k) \in [0, 1]^k : \sum_{i=1}^k x_i \leq 1 \right\}$ with $I_k(F) \neq 0$ and $J_k^{(m)}(F) \neq 0$ for each m . Let

$$M_k = \sup_{F \in \mathcal{S}_k} \frac{\sum_{m=1}^k J_k^{(m)}(F)}{I_k(F)}, \quad r_k = \left\lceil \frac{\theta M_k}{2} \right\rceil.$$

Then there are infinitely many integers n such that at least r_k of the $n + h_i$ ($1 \leq i \leq k$) are prime. In particular, $\liminf_n (p_{n+r_k-1} - p_n) \leq \max_{1 \leq i, j \leq k} (h_i - h_j)$.
Proof of Proposition 4.2 We let $S = S_2 - \rho S_1$, and recall that from Section 2 that if we can show $S > 0$ for all large N , then there are infinitely many integers n such that at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime.

We put $R = N^{\theta/2-\delta}$ for a small $\delta > 0$. By the definition of M_k , we can choose $F_0 \in \mathcal{S}_k$ such that $\sum_{m=1}^k J_k^{(m)}(F_0) > (M_k - \delta) I_k(F_0) > 0$. Since F_0 is Riemann-integrable, there is a smooth function F_1 such that $\sum_{m=1}^k J_k^{(m)}(F_1) > (M_k - 2\delta) I_k(F_1) > 0$. Using Proposition 4.1, we can then choose $\lambda_{d_1, \dots, d_k}$ such that

$$\begin{aligned} S &= \frac{\varphi(W)^k N (\log R)^k}{W^{k+1}} \left(\frac{\log R}{\log N} \sum_{j=1}^k J_k^{(m)}(F_1) - \rho I_k(F_1) + o(1) \right) \\ &\geq \frac{\varphi(W)^k N (\log R)^k I_k(F_1)}{W^{k+1}} \left(\left(\frac{\theta}{2} - \delta \right) (M_k - 2\delta) - \rho + o(1) \right). \end{aligned}$$

If $\rho = \theta M_k/2 - \epsilon$ then, by choosing δ suitably small (depending on ϵ), we see that $S > 0$ for all large N . Thus there are infinitely many integers n for which at least $\lfloor \rho + 1 \rfloor$ of the $n + h_i$ are prime. Since $\lfloor \rho + 1 \rfloor = \lceil \theta M_k/2 \rceil$ if ϵ is suitably small, we obtain Proposition 4.2.

Thus, if the primes have a fixed level of distribution θ , to show the existence of many of the $n + h_i$ being prime for infinitely many $n \in \mathbb{N}$ we only require a suitable lower bound for M_k . The following proposition establishes such a bound for different values of k .

Proposition 4.3. Let $k \in \mathbb{N}$, and M_k be as given by Proposition 4.2 Then

- (1) We have $M_5 > 2$.
- (2) We have $M_{105} > 4$.
- (3) If k is sufficiently large, we have $M_k > \log k - 2 \log \log k - 2$.

We now prove Theorems 1.1, 1.2, 1.3 and 1.4 from Propositions 4.2 and 4.3,

First we consider Theorem 1.3, We take $k = 105$. By Proposition 4.3, we have $M_{105} > 4$. By the Bombieri-Vinogradov theorem, the primes have level of distribution $\theta = 1/2 - \epsilon$ for every $\epsilon > 0$. Thus, if we take ϵ sufficiently small, we have $\theta M_{105}/2 > 1$. Therefore, by Proposition 4.2, we have $\liminf (p_{n+1} - p_n) \leq \max_{1 \leq i, j \leq 105} (h_i - h_j)$ for any admissible set $\mathcal{H} = \{h_1, \dots, h_{105}\}$. By computations performed by Thomas Engelsma (unpublished), we can choose \mathcal{H} such that $0 \leq h_1 < \dots < h_{105}$ and $h_{105} - h_1 = 600$. This gives Theorem 1.3.

If we assume the Elliott-Halberstam conjecture then the primes have level of distribution $\theta = 1 - \epsilon$. First we take $k = 105$, and see that $\theta M_{105}/2 > 2$ for ϵ sufficiently small (since $M_{105} > 4$). Therefore, by Proposition 4.2, $\liminf_n (p_{n+2} - p_n) \leq \max_{1 \leq i, j \leq 105} (h_i - h_j)$. Thus, choosing the same admissible set \mathcal{H} as above, we see $\liminf_n (p_{n+2} - p_n) \leq 600$ under the Elliott-Halberstam conjecture.

Next we take $k = 5$ and $\mathcal{H} = \{0, 2, 6, 8, 12\}$, with $\theta = 1 - \epsilon$ again. By Proposition 4.3 we have $M_5 > 2$, and so $\theta M_5/2 > 1$ for ϵ sufficiently small. Thus, by Proposition 4.2, $\liminf_n (p_{n+1} - p_n) \leq 12$ under the Elliott-Halberstam conjecture. This completes the proof of Theorem 1.4

² Explicitly, we can take $\mathcal{H} = \{0, 10, 12, 24, 28, 30, 34, 42, 48, 52, 54, 64, 70, 72, 78, 82, 90, 94, 100, 112, 114, 118, 120, 124, 132, 138, 148, 154, 168, 174, 178, 180, 184, 190, 192, 202, 204, 208, 220, 222, 232, 234, 250, 252, 258, 262, 264, 268, 280, 288, 294, 300, 310, 322, 324, 328, 330, 334, 342, 352, 358, 360, 364, 372, 378, 384, 390, 394, 400, 402, 408, 412, 418, 420, 430, 432, 442, 444, 450, 454, 462, 468, 472, 478, 484, 490, 492, 498, 504, 510, 528, 532, 534, 538, 544, 558, 562, 570, 574, 580, 582, 588, 594, 598, 600\}$.

This set was obtained from the website <http://math.mit.edu/~primegaps/> maintained by Andrew Sutherland. Finally, we consider the case when k is large. For the rest of this section, any constants implied by asymptotic notation will be independent of k . By the Bombieri-Vinogradov theorem, we can take $\theta = 1/2 - \epsilon$. Thus, by Proposition 4.3, we have for k sufficiently large

$$\frac{\theta M_k}{2} \geq \left(\frac{1}{4} - \frac{\epsilon}{2}\right) (\log k - 2 \log \log k - 2).$$

We choose $\epsilon = 1/k$, and see that $\theta M_k/2 > m$ if $k \geq Cm^2 e^{4m}$ for some absolute constant C (independent of m and k). Thus, for any admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ with $k \geq Cm^2 e^{4m}$, at least $m + 1$ of the $n + h_i$ must be prime for infinitely many integers n . We can choose our set \mathcal{H} to be the set $\{p_{\pi(k)+1}, \dots, p_{\pi(k)+k}\}$ of the first k primes which are greater than k . This is admissible, since no element is a multiple of a prime less than k (and there are k elements, so it cannot cover all residue classes modulo any prime greater than k .) This set has diameter $p_{\pi(k)+k} - p_{\pi(k)+1} \ll k \log k$. Thus $\liminf_n (p_{n+m} - p_n) \ll k \log k \ll m^3 e^{4m}$ if we take $k = \lceil Cm^2 e^{4m} \rceil$. This gives Theorem 1.1.

We can now establish Theorem 1.2 by a simple counting argument. Given m , we let $k = \lceil Cm^2 e^{4m} \rceil$ as above. Therefore if $\{h_1, \dots, h_k\}$ is admissible, then there exists a subset $\{h'_1, \dots, h'_m\} \subseteq \{h_1, \dots, h_k\}$ with the property that there are infinitely many integers n for which all of the $n + h'_i$ are prime ($1 \leq i \leq m$).

We let \mathcal{A}_2 denote the set formed by starting with the given set $\mathcal{A} = \{a_1, \dots, a_r\}$, and for each prime $p \leq k$ in turn removing all elements of the residue class modulo p which contains

the fewest integers. We see that $\#\mathcal{A}_2 \geq r \prod_{p \leq k} (1 - 1/p) \gg_m r$. Moreover, any subset of \mathcal{A}_2

of size k must be admissible, since it cannot cover all residue classes modulo p for any prime $p \leq k$. We let $s = \#\mathcal{A}_2$, and since r is taken sufficiently large in terms of m , we may assume that $s > k$.

We see there are $\binom{s}{k}$ sets $\mathcal{H} \subseteq \mathcal{A}_2$ of size k . Each of these is ad-

missible, and so contains at least one subset $\{h'_1, \dots, h'_m\} \subseteq \mathcal{A}_2$ which satisfies the prime m -tuples conjecture. Any admissible set $\mathcal{B} \subseteq \mathcal{A}_2$ of size m is contained in $\binom{s-m}{k-m}$ sets $\mathcal{H} \subseteq \mathcal{A}_2$ of size k . Thus there are at least $\binom{s}{k} \binom{s-m}{k-m}^{-1} \gg_m s^m \gg_m r^m$ admissible sets $\mathcal{B} \subseteq \mathcal{A}_2$ of size m which satisfy the prime m -tuples conjecture. Since there are $\binom{r}{m} \leq r^m$ sets $\{h_1, \dots, h_m\} \subseteq \mathcal{A}$, Theorem 1.2 holds.

We are left to establish Propositions 4.1 and 4.3.

Corollary 1 $\lim_n \inf(\sqrt{p_{n+1}} - \sqrt{p_n}) < 1$ We get easily the following equivalent proposition : $\lim_n \inf(\sqrt{p_{n+1}} - \sqrt{p_n}) < 1 \Leftrightarrow \lim_n \inf(p_{n+1} - p_n < 1 + 2\sqrt{p_n})$ (james maynard proved that :)

$$\lim_n \inf(p_{n+1} - p_n) < 600$$

and while we have $\lim_n \inf(\sqrt{p_n}) > 600 \therefore$
 $\lim_n \inf(\sqrt{p_{n+1}} - \sqrt{p_n}) < 1$ QED.

5 References

- [1] P. D. T. A. Elliott and H. Halberstam. A conjecture in prime number theory. In Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69), pages 59–72. Academic Press, London, 1970.
- [2] J. Friedlander and A. Granville. Limitations to the equi-distribution of primes. I. Ann. of Math. (2), 129(2):363–382, 1989.
- [3] D. A. Goldston, S. W. Graham, J. Pintz, and C. Y. Yıldırım. Small gaps between products of two primes. Proc. Lond. Math. Soc. (3), 98(3):741–774, 2009.
- [4] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in tuples. III. On the difference $p_n + p_n$. Funct. Approx. Comment. Math., 35:79–89, 2006.
- [5] D. A. Goldston, J. Pintz, and C. Y. Yıldırım. Primes in tuples. I. Ann. of Math. (2), 170(2):819–862, 2009.
- [6] D. A. Goldston and C. Y. Yıldırım. Higher correlations of divisor sums related to primes. III. Small gaps between primes. Proc. Lond. Math. Soc. (3), 95(3):653–686, 2007.
- [7] D. H. J. Polymath. A new bound for gaps between primes. Preprint.
- [8] A. Selberg. Collected papers. Vol. II. Springer-Verlag, Berlin, 1991. With a foreword by K. Chandrasekharan.
- [9] Y. Zhang. Bounded gaps between primes. Ann. of Math.(2), to appear.
- [10] J. MAYNARD. SMALL GAPS BETWEEN PRIMES. Centre de recherches mathématiques, Université de Montréal, Pavillon André-Aisenstadt, 2920 Chemin de la tour, Room 5357, Montréal (Québec) H3T 1J4

(Special thanks to my colleague ayman chaafi for his intellectual assistance)