Bounded prime gaps in short intervals

Johan Andersson

Abstract

We generalise Zhang’s and Pintz recent results on bounded prime gaps to give a lower bound for the number of prime pairs bounded by $6 \cdot 10^7$ in the short interval $[x, x + x(\log x)^{-A}]$. Our result follows only by analysing Zhang’s proof of Theorem 1, but we also explain how a sharper variant of Zhang’s Theorem 2 would imply the same result for shorter intervals.

Yitang Zhang [11] in his recent landmark paper proved the there are infinitely many weak prime pairs. More precisely he proved that

$$\lim \inf_{n \to \infty} (p_{n+1} - p_n) < 7 \cdot 10^7.$$  \hfill (1)

His method is a variant of the method of Goldston-Pintz-Yıldırım [1, 2]. For a nice introduction to this method see Soundararajan [9]. Zhang not only realised that it is sufficient to consider divisors $d$ with only small prime factors in the definition of the real arithmetical function $\lambda(n)$ that is fundamental in the method [1], but he also managed to utilise this insight. As in previous work on related results, Zhang used the notation of admissible set. A set

$$\mathcal{H} = \{h_1, \ldots, h_k\}$$  \hfill (2)

is called admissible if there is no obvious arithmetical obstruction that makes it impossible for $\mathcal{H} + n$ to be prime for all elements (such as $\mathcal{H} = \{1, 2\}$) except for a possible exceptional case. The Hardy-Littlewood conjecture asserts that the number of $n$ less than $x$ such that $n + \mathcal{H}$ consists of only primes should be asymptotic to $\mathcal{S}(\mathcal{H})x/(\log x)^k$, where $\mathcal{S}(\mathcal{H})$ is a certain positive constant. While far from being proven, it gives good intuition what the right answer should be for these questions.

Zhang’s result (1) follows from his result that for any admissible $\mathcal{H}$ with at least $3.5 \cdot 10^6$ elements there are infinitely many $n$ such that the set $n + \mathcal{H}$ contains at least two primes. By being more careful with the estimates and choice of the admissible set, Trudgian [10] already improved Zhang’s constant $7 \cdot 10^7$ to $6 \cdot 10^7$. There is also a team effort led by Terence Tao that as of this moment has managed to improve the bound to a number slightly less than...
4 \cdot 10^5$; for the latest results see [8]. Since this is work in progress and depends on more substantial changes to the proof of Zhang’s Theorem 1, we have chosen not to use this result in this version of the paper.

Zhang’s bounded prime gap theorem ultimately follows from the inequality

$$\sum_{x \leq n \leq 2x} \left( \sum_{i=1}^{k_0} \theta(n + h_i) - \log(3x) \right) \lambda(n))^2 \geq (\omega + o(1))x(\log x)^{k_0 + l_0 + 1}, \quad (3)$$

for some constant $\omega > 0$ where

$$\theta(n) = \begin{cases} \log n, & \text{if } n \text{ is prime}, \\ 0, & \text{otherwise}, \end{cases}$$

and $\lambda$ is defined as follows (following Zhang [11 (2.12)]):

$$\lambda(n) = \sum_{d \mid (P(n), P)} \mu(d)g(d), \quad g(y) = \begin{cases} \frac{1}{(k_0 + l_0)^2} \left( \log \frac{D}{y} \right)^{k_0 + l_0}, & y < D, \\ 0, & y \geq D, \end{cases} \quad (4)$$

where

$$k_0 = 3.5 \cdot 10^6, \quad l_0 = 180, \quad \varpi = 1/1168, \quad D = x^{\varpi + 1/4}, \quad D_1 = x^\varpi, \quad P = \prod_{p \leq D_1} p, \quad P(n) = \prod_{i=1}^{k_0} (n + h_i).$$

By estimating $\lambda(n)$ trivially by its definition (4) and absolute values we see that

$$\lambda(n) \ll \tau \left( \prod_{i=1}^{k_0} (n + h_i) \right) (\log D)^{k_0 + l_0},$$

where $\tau(n)$ denotes the divisor function, and from the well known estimate $\tau(n) \ll n^{\varepsilon}$ one sees that

$$\lambda(n) \ll_{\varepsilon} x^\varepsilon, \quad (n \ll x). \quad (5)$$

The equations (3) and (5) immediately yield that the number of weak prime pairs less than $x$ is at least of the order $x^{1-\varepsilon}$ for any $\varepsilon > 0$, and thus Zhang’s proof method gives somewhat stronger results than he states in his paper. The point here is that we do not only have that the left hand side of the inequality in (3) is positive, but also that it is bounded from below by $x^{1-\varepsilon}$. Pintz [7] was the first to come out with a paper including such a result. He furthermore managed to prove this result where $x^{1-\varepsilon}$ is replaced by a power of $\log x$.

In this short paper we will consider the question of existence of pairs of primes with bounded gaps in short intervals, such as $[x, x + \Delta(x)]$ for some function $\Delta(x) = o(x)$. It turns out that in a similar way as in Pintz [7] it is sufficient to analyse the proof of Theorem 1 in Zhang [11] and use his version of Theorem 2 (which seems to be the deepest part of his paper) as in, in order to prove the following result.

\footnote{His results are in fact much more general and much deeper than this, see his paper.}
**Theorem.** Suppose that $A \geq 0$ and $\varepsilon > 0$. Then the interval $[x, x + x/(\log x)^A]$ contains at least $(1 + o(1))x^{1-\varepsilon}$ pairs of primes $p_1, p_2$ such that $1 < |p_1 - p_2| < 6 \cdot 10^7$.

**Proof.** By using Trudgian’s construction [10] of an admissible set it follows immediately from Lemma 1.

**Remark 1.** By some more work, such as using estimates for the divisor functions (Lemma 8 in Zhang [11]) or similar methods as in Pintz [7] it is possible to improve the lower bound by replacing $x^{-\varepsilon}$ with a power of $\log x$.

**Lemma 1.** Assume that $H$ is an admissible set with at least $k_0 = 3.5 \cdot 10^6$ elements. Then, given $\varepsilon, A > 0$, the interval $[x, x + x/(\log x)^A]$ contains at least $(1 + o(1))x^{1-\varepsilon}$ integers $n$ such that the set $n + H$ contains at least two primes.

**Proof.** Lemma 1 follows from using Lemma 2 to estimate the error term in Lemma 3 and Eq. (5).

As the second Lemma we will state a variant of [11, Theorem 2] which is a strong version of the Bombieri-Vinogradov theorem.

**Lemma 2.** Let $\Delta(x) = x(\log x)^{-A}$, and assume that

$$\Delta(\gamma; d, c) = \sum_{\substack{n \equiv c \pmod{d} \\{x \leq n \leq x + \Delta(x)\}}} \gamma(n) - \frac{1}{\phi(d)} \sum_{x \leq n \leq x + \Delta(x)} \gamma(n) \quad \text{for} \quad (d, c) = 1.$$  \hspace{1cm} (6)

Then for any $B > 0$ and $1 \leq i \leq k_0$

$$\sum_{d \leq D^2} \sum_{c \in \mathcal{C}_i(d)} |\Delta(\theta; d, c)| \ll_{A, B} \Delta(x)(\log x)^{-B},$$

where

$$\mathcal{C}_i(d) = \{c : 1 \leq c \leq d, (c, d) = 1, P(c - h_i) \equiv 0 \pmod{d}\} \quad \text{for} \quad 1 \leq i \leq k_0.$$

**Proof.** Although stated for short intervals $[x, x + \Delta(x)]$ instead of $[x, 2x]$ this follows directly from Zhang’s theorem 2. This is because $\Delta(x)$ is just $x$ multiplied by a power of $\log x$ and this power can be accounted for in the error term.

**Remark 2.** If we can prove Lemma 2 for $\Delta(x) = x^\theta$ and some $7/12 < \theta < 1$ it would follow that the intervals $[x, x + x^\theta]$ contains at least $(1 + o(1))x^{\theta-\varepsilon}$ weak prime pairs. This would be of great interest. We have chosen to formulate Lemma 3 such that a proof of Lemma 2 in a short interval will immediately give such a consequence. Although we have not yet checked all the details of Zhang’s proof of Theorem 2, his approach seems promising also for short intervals. In particular he at many places gets somewhat sharper results than he needs that might be more useful in the short interval case. We also remark that the corresponding short interval version of the classical Bombieri-Vinogradov theorem has been proved by Perelli-Pintz-Salerno [5, 6].
Lemma 3. Assume that $\mathcal{H}$ is admissible and $\mathcal{H}$ contains at least $k_0 = 3.5 \cdot 10^6$ elements. Then there exists some $B > 0$ such that with $\Delta(\theta; d, c)$ defined by (6) and $0 < \Delta(x) \ll x$ we have that

$$\sum_{x \leq n \leq x + \Delta(x)} \left( \sum_{i=1}^{k_0} \theta(n + h_i) - \log(3x) \right) (\lambda(n))^2 \geq (\omega + o(1)) \Delta(x)(\log x)^{k_0 + l_0 + 1} + O \left( (\log x)^{B} \sum_{d \leq D^2} \sum_{c \in C_i(d)} |\Delta(\theta; d, c)| \right) + O \left( x^{7/12} \right),$$

where we can choose $\omega = \exp(-5 \cdot 10^7)$.

Proof. We will follow Zhang [11] and just indicate where changes are needed. The essence in the changes is that we would like $n \sim x$ to mean $x \leq n \leq x + \Delta(x)$ instead of $x \leq n \leq 2x$. We now proceed as in Zhang [11]. We write the left hand side of the lemma as

$$S_2 - \log(3x)S_1,$$

where

$$S_1 = \sum_{x \leq n \leq x + \Delta(x)} (\lambda(n))^2,$$

and

$$S_2 = \sum_{x \leq n \leq x + \Delta(x)} (\lambda(n))^2 \sum_{i=1}^{k_0} \theta(n + h_i).$$

We will treat $S_1$ and $S_2$ separately.

Upper bound for $S_1$

We follow Zhang [11], section 4. We treat the inner sum that corresponds to the first displayed formula on p. 16 in the same way. Proceeding in a similar manner, instead of eq. (4.1) in Zhang we get

$$S_1 = T_1 \Delta(x) + O(D^{2+\varepsilon}).$$

The quantity $T_1$ is the same in Zhang, and from his (4.19) we obtain similarly to (4.20) in his paper that

$$S_1 \leq \frac{(1 + \kappa_1 + o(1))}{(k_0 + 2l_0)!} \left( \frac{2l_0}{l_0} \right) \mathcal{S}(\mathcal{H}) x(\log D)^{k_0 + 2l_0} + O(D^{2+\varepsilon}),$$

for some $\kappa_1 < \exp(-1200)$. 

\[ \text{4} \]
Lower bound for $S_2$

Again we will change the summation order as in Zhang [11, p. 22] and we get

$$S_2 = k_0 T_2 \sum_{x \leq n \leq x + \Delta(x)} \theta(n) + k_0 \sum_{i=1}^{k_0} O(\mathcal{E}_i),$$

where $T_2$ is defined as in his paper and

$$\mathcal{E}_i = \sum_{d < D^2 \atop d \mid \mathcal{P}} \tau_3(d) \rho_2(d) \sum_{c \in C_i(d)} |\Delta(\theta; d, c)|,$$  \hspace{1cm} (8)

where we now keep in mind that $\Delta(\theta; d, c)$ is defined by (6), i.e. summing over $n$ in shorter intervals than in Zhang [11]. We may now assume that $\Delta(x) \gg x^{7/12} \log x/\phi(d)$ in the first sum, and estimates for sum of divisor functions in short intervals ([11, Lemma 8]), this gives us the estimate

$$\mathcal{E}_i \ll (\log x)^B \sqrt{\Delta(x)} \left( \sum_{d < D^2 \atop d \mid \mathcal{P}} \sum_{c \in C_i(d)} |\Delta(\theta; d, c)| \right),$$ \hspace{1cm} (9)

for some positive constant $B > 0$. Since $T_2$ is the same as in Zhang’s paper, it can be estimated by [11, Eq (5.5)]. Corresponding to Eq (5.6) in Zhang we get the inequality

$$S_2 \geq \frac{k_0 (1 - \kappa_2)}{(k_0 + 2l_0 + 1)!} \left( \frac{2l_0 + 2}{l_0 + 1} \right) \mathcal{G}(\mathcal{H}) \Delta(x) (\log D)^{k_0 + 2l_0 + 1} (1 + o(1)).$$ \hspace{1cm} (10)

---

3We are grateful to GH and Denis Chaperon de Lauzi at mathoverflow for explaining precisely how Zhang used Cauchy’s inequality. See [http://mathoverflow.net/questions/132452](http://mathoverflow.net/questions/132452).

4This corresponds to each integer being prime in the residue class for the short interval.
By the inequalities (7) and (10) we obtain
\[ S_2 - \log(3x)S_1 \geq \omega \mathbb{E}(\mathcal{H}) \Delta(x)(\log D)^{k_0+2l_0+1}(1 + o(1)) + O(D^{2+\epsilon}) + O \left( \sum_i \mathcal{E}_i \right), \] (11)
where
\[ \omega = \frac{1}{(k_0 + 2l_0)!} \left( \frac{2l_0}{l_0} \right) \left( \frac{2(2l_0 + 1)k_0(1 - \kappa_2)}{(l_0 + 1)(k_0 + 2l_0 + 1)} - \frac{4(1 + \kappa_1)}{1 + 4\varpi} \right). \]

While Zhang never calculates \( \omega \) and just say it is positive, by the estimates \( \kappa_1 < \exp(-1200) \) and \( \kappa_2 < 10^8 \exp(-1200) \) from Zhang [11], and the numerical values of \( l_0, k_0, \varpi \) it is easy to use Mathematica/Sage to see that \( \omega = 3.647 \cdot 10^{-21385285} > \exp(-5 \cdot 10^7) \). The results follows by combining (9) with (11) with the fact that with our choice or \( D \) we have \( O(D^{2+\epsilon}) = O(x^{7/12}) \).

References

[1] D. A. Goldston, J. Pintz, C. Yıldırım, Primes in tuples. I, Ann. of Math. (2) 170 (2009), no. 2, 819–862. [arXiv:math/0508185]
[2] D. A. Goldston, J. Pintz, C. Yıldırım, Primes in tuples. II, Acta Math. 204 (2010), no. 1, 1–47. [arXiv:0710.2728]
[3] D. R. Heath-Brown The number of primes in a short interval. J. Reine Angew. Math. 389 (1988), 2263.
[4] Y. Motohashi, J. Pintz, A smoothed GPY sieve, Bull. London Math. Soc. 40 (2008), no. 2, 298–310. [arXiv:math/0602599]
[5] Perelli, A.; Pintz, J.; Salerno, S. Bombieri’s theorem in short intervals. Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) 11 (1984), no. 4, 529-539.
[6] Perelli, A.; Pintz, J.; Salerno, S. Bombieri’s theorem in short intervals. II. Invent. Math. 79 (1985), no. 1, 1-9.
[7] J. Pintz, Polignac Numbers, Conjectures of Erdős on Gaps between Primes, Arithmetic Progressions in Primes, and the Bounded Gap Conjecture [arXiv:1305.6289v1]
[8] Polymath, http://michaelnielsen.org/polymath1/index.php?title=Bounded_gaps_between_primes. Work in progress.
[9] K. Soundararajan Small gaps between prime numbers: The work of Goldston-Pintz-Yildirim [arXiv:math/0605696]
[10] T. S. Trudgian A poor man’s improvement on Zhang’s result: there are infinitely many prime gaps less than 60 million [arXiv:1305.6369v1]
[11] Yitang Zhang, Bounded gaps between primes, Ann. of Math., to appear.