ACTA ARITHMETICA
160.1 (2013)

Primes in tuples IV: Density of small gaps between consecutive primes

by

Daniel Alan Goldston (San Jose, CA), János Pintz (Budapest) and Cem Yalçın Yildirim (İstanbul)

1. Introduction. Denote, as usual, by $\pi(x)$ the number of primes not exceeding x, and by p_n the nth prime. The prime number theorem says $\pi(x) \sim x/\log x$ as $x \to \infty$, so that on average $p_{n+1} - p_n$ is about $\log p_n$. In [GPY1] we proved

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log p_n} = 0, \tag{1.1}$$

that is, 'small gaps' (smaller than any positive constant times the average gap) between primes exist ad infinitum. Concerning the distribution of gaps between primes, it is conjectured (see [S1], [S2]) that

$$\#\{p_n \le x;\, p_{n+1} \in (p_n + \alpha \log p_n, p_n + \beta \log p_n)\} \sim \pi(x) \int_\alpha^\beta e^{-t}\, dt \tag{1.2}$$

as $x \to \infty$, for any two fixed real numbers $\beta > \alpha \ge 0$. Gallagher's calculation [Ga] shows that this conjecture can be deduced from the Hardy–Littlewood prime k-tuples conjecture (see [S2]). Hence an immediate query to be conducted was whether the small gaps between primes attested to by the proof of (1.1) constitute a positive proportion of the set of all gaps between consecutive primes.

The main result of this article is

Theorem 1. *For sufficiently small but fixed $\eta > 0$,*

$$\sum_{\substack{N < p_j \le 2N \\ p_{j+1} - p_j \le \eta \log N}} 1 \gtrsim e^{-c\eta^{-6}} \frac{N}{\log N} \quad (N \to \infty), \tag{1.3}$$

with $c = \lceil 65 \cdot 4^6 \cdot \log 2 \rceil = 184544$. Thus, for any fixed $\eta > 0$,

$$\#\{p_n \le x;\, p_{n+1} - p_n \le \eta \log p_n\} \gg_\eta \pi(x). \tag{1.4}$$

2010 *Mathematics Subject Classification*: Primary 11N05; Secondary 11N36.
Key words and phrases: consecutive primes, gaps, density.

DOI: 10.4064/aa160-1-3

While dwelling upon how (1.1) was obtained in the next section, we will introduce some concepts and notation. The proof of Theorem 1 is given in the third and fourth sections. Theorem 2 in the fifth section expresses that 'very small gaps' are sparse. In the last section we conditionally obtain stronger results, by assuming more than what the Bombieri–Vinogradov theorem supplies.

A rather qualitative concise version of the results of this paper was presented in [GPY3]. Here we carry out the calculations in full detail so that in the results the dependence on η is explicit. This explicitness provides, within the framework of our method, a new quantitative manifestation of the effect of the extent of assumed information about how well the primes are distributed in arithmetic progressions in addition to that provided by the results in [GPY1]. To obtain the explicit estimates it was necessary to develop the propositions given in the third section which may be useful in various other problems.

2. Preliminaries. A bibliography of former results for the limit in (1.1) was given in [GPY1]. Before [GPY1] the best known result for this limit was

$$\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{\log p_n} \le 0.2484\ldots,$$

due to Maier [M]. In his proof Maier constructed thin sets in which the density of primes is larger by a factor of e^γ than on average, and paralleling the Bombieri–Davenport–Huxley method ([BD], [Hu]) with necessary modifications, he obtained this upper bound which is $e^{-\gamma}$ times the result of the older method. Since Maier's method involved working within a thin set of integers, the gaps indicated by his result could not constitute a positive proportion of all gaps between consecutive primes. The second best result was the upper bound 1/4 by Goldston and Yıldırım [GY2]. The origin of its method also being the Bombieri–Davenport proof, the gaps of [GY2] were shown to make up a positive proportion. Another positive proportion result was obtained by Bazzanella, Languasco and Zaccagnini [BLZ] for larger gaps. They showed that, as $x \to \infty$, the number of primes $p \le x$ for which the interval $(p, p+\eta\log x]$, where η is any constant $> 1/2$, contains at least one prime is $\gtrsim \Delta(\eta)x/\log x$ with the distinction that an explicit expression for $\Delta(\eta)$ was provided.

We now outline how the result (1.1) was attained. For a k-tuple

$$\mathcal{H} = \{h_1, \ldots, h_k\} \quad \text{with distinct integers } h_1, \ldots, h_k \in [1, h], \tag{2.1}$$

and for a prime p, denote by $\nu_p(\mathcal{H})$ the number of distinct residue classes modulo p occupied by the entries of $\mathcal{H}$. The singular series associated with

$\mathcal{H}$ is defined as

$$\mathfrak{S}(\mathcal{H}) := \prod_p \Big(1-\frac{1}{p}\Big)^{-k}\Big(1-\frac{\nu_p(\mathcal{H})}{p}\Big), \tag{2.2}$$

the product being convergent because $\nu_p(\mathcal{H}) = k$ for $p > h$. We say that $\mathcal{H}$ is *admissible* if

$$P_{\mathcal{H}}(n) := (n+h_1)\cdots(n+h_k) \tag{2.3}$$

is not divisible by a fixed prime number for every n, which is equivalent to $\nu_p(\mathcal{H}) \neq p$ for all p and therefore also to $\mathfrak{S}(\mathcal{H}) \neq 0$. That $\{n+h_1,\dots,n+h_k\}$ is a prime tuple, i.e. each entry is prime, is equivalent to $P_{\mathcal{H}}(n)$ being a product of k primes. Since the generalized von Mangoldt function

$$\Lambda_k(m) := \sum_{d|m} \mu(d)\Big(\log\frac{m}{d}\Big)^k$$

vanishes when m has more than k distinct prime factors, the quantity

$$\frac{1}{k!}\sum_{\substack{d|P_{\mathcal{H}}(n)\\ d\le R}} \mu(d)\Big(\log\frac{R}{d}\Big)^k$$

with the truncation $d \le R$ may be employed in detecting prime tuples, albeit roughly (the contribution from proper prime power factors is negligible; $1/k!$ is a normalization factor). However, it turns out that an additional crucial idea is needed, and that is to give up trying to count tuples consisting of primes exclusively in favour of including tuples with primes in many entries. This brings about the use of

$$\Lambda_R(n;\mathcal{H},\ell) := \frac{1}{(k+\ell)!}\sum_{\substack{d|P_{\mathcal{H}}(n)\\ d\le R}} \mu(d)\Big(\log\frac{R}{d}\Big)^{k+\ell} \qquad (0\le\ell\le k-2). \tag{2.4}$$

Let

$$\theta(n) := \begin{cases} \log n & \text{if } n \text{ is prime,}\\ 0 & \text{otherwise,}\end{cases} \tag{2.5}$$

and

$$\Theta(n,h) := \sum_{1\le h_0\le h} \theta(n+h_0). \tag{2.6}$$

The proof of (1.1) is achieved by showing the positivity of the quantity

$$S_R(N,k,\ell,h) := \sum_{N<n\le 2N} (\Theta(n,h)-\log 3N)\Big(\sum_{\substack{\mathcal{H}\subset[1,h]\\ |\mathcal{H}|=k}} \Lambda_R(n;\mathcal{H},\ell)\Big)^2. \tag{2.7}$$

Here, as $N\to\infty$, for a result of the type (1.1) we need $\epsilon\log N \ll h \ll \log N$ with an arbitrarily small but fixed $\epsilon > 0$, and the larger the truncation level

R is relative to N, the better detection will be provided by (2.4). The tuple size k is taken to be arbitrarily large but fixed. In fact, for the proof of (1.1) it suffices to consider the simpler expression where the inner sum consists only of the diagonal terms $\Lambda_R(n;\mathcal{H},\ell)^2$, and a modified version of this will be used in Section 4. The expression in (2.7) is needed to achieve a better result in the case of the gaps $p_{n+r}-p_n$ with $r\ge 2$ in [GPY1] and for a stronger quantitative version of (1.1) in [GPY2].

The essential information on primes we need beyond the prime number theorem concerns the level of distribution of primes in arithmetic progressions. We say that the primes have a *level of distribution* ϑ if

$$\sum_{q\le Q}\max_{\substack{a\\(a,q)=1}}\Bigg|\sum_{\substack{p\text{ prime}\\p\le N\\p\equiv a\,(\mathrm{mod}\,q)}}\log p-\frac{N}{\phi(q)}\Bigg|\ll_{\epsilon,A}\frac{N}{(\log N)^A} \tag{2.8}$$

for any $A>0$ and any $\epsilon>0$ with

$$Q=N^{\vartheta-\epsilon}. \tag{2.9}$$

According to the Bombieri–Vinogradov theorem, for any $A>0$ there is a $B=B(A)$ such that (2.8) holds with $Q=N^{1/2}(\log N)^{-B}$, so that the primes are known to have level of distribution $1/2$. The Elliott–Halberstam conjecture is that the primes have level of distribution 1.

The following are relevant special cases of the propositions from [GPY1]. For an admissible k-tuple $\mathcal{H}$, we have

$$\sum_{n\le N}\Lambda_R(n;\mathcal{H},\ell)^2\sim\binom{2\ell}{\ell}\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}\mathfrak{S}(\mathcal{H})N \tag{2.10}$$

as $R,N\to\infty$, for $R\ll N^{1/2}(\log N)^{-8M}$ where $M=k+\ell$, and $h\le R^C$ for any given constant $C>0$. In the situation of weighting with the primes, for $1\le h_0\le h$ writing $m=1$ when $h_0\in\mathcal{H}$ and $m=0$ when $h_0\notin\mathcal{H}$, if $\mathcal{H}\cup\{h_0\}$ is admissible we have

$$\sum_{N\le n}\theta(n+h_0)\Lambda_R(n;\mathcal{H},\ell)^2\sim\binom{2(\ell+m)}{\ell+m}\frac{\mathfrak{S}(\mathcal{H}\cup\{h_0\})}{(k+2\ell+m)!}N(\log R)^{k+2\ell+m} \tag{2.11}$$

as $R,N\to\infty$, provided that $R\ll_M N^{1/4}(\log N)^{-B(M)}$ for a sufficiently large positive constant $B(M)$, and $h\le R$. The upper bound for R is forced by the dependence of the proof of (2.11) on the Bombieri–Vinogradov theorem, and for the unconditional results in [GPY1], taking $R=N^{1/4-\epsilon}$ suffices. More generally, assuming that the primes have level of distribution ϑ with a fixed $\vartheta\in[1/2,1]$, (2.11) holds with $R\ll N^{\vartheta/2-\epsilon}$ and $h\le R^\epsilon$ for any $\epsilon>0$.

The proof of (2.10) and (2.11) may be summarized as follows. Upon writing the left-hand sides explicitly by substituting (2.4), the sum over n is carried to the innermost position and easily evaluated. Then a Mellin transform converts the expressions into integrals over vertical lines in the complex plane. The integrands contain Dirichlet series which encode the arithmetic information from the tuples. The integrals are evaluated by shifting the lines of integration appropriately and by calculating the residues and the bounds for the integrals over the new contours.

For the calculation of $S_R(N,k,\ell,h)$, the general versions of (2.10) and (2.11) are employed in the expression on the right-hand side of (2.7), and then Gallagher's [Ga] result

$$\sum_{\substack{\mathcal{H}\subset[1,h]\\|\mathcal{H}|=k}} \mathfrak{S}(\mathcal{H}) \sim h^k \quad \text{for fixed } k \text{ as } h\to\infty \tag{2.12}$$

(where each set is counted $k!$ times due to all of its permutations) is needed to complete the calculation. The parameters which appear in this process are chosen judiciously, in particular k has to be arbitrarily large but fixed and the optimal order of magnitude of the integer ℓ turns out to be $\sqrt{k}$.

For a proof of (1.4) the weights $\Lambda_R(n;\mathcal{H},\ell)^2$ have to be removed, which is customarily achieved by an application of the Cauchy–Schwarz inequality. Straightforward adaptation of the argument in [GY2] brings in the fourth moment of prime tuple approximants. Specifically, for a proof of (1.4), one needs to show that

$$\sum_{N<n\le 2N} \Big(\sum_{\substack{\mathcal{H}\subset[1,h]\\|\mathcal{H}|=k}} \Lambda_R(n;\mathcal{H},\ell)\Big)^4 \ll N(\log N)^{4k+4\ell}. \tag{2.13}$$

However, some calculations indicate that the truth of

$$\sum_{N<n\le 2N} \sum_{\substack{\mathcal{H}\subset[1,h]\\|\mathcal{H}|=k}} \Lambda_R(n;\mathcal{H},\ell)^4 \ll N(\log N)^{4k+4\ell} \tag{2.14}$$

is questionable. It seems that there exist n having a lot of divisors for which $|\Lambda_R(n;\mathcal{H},\ell)|$ becomes exceptionally large, thereby preventing an estimate such as (2.14).

3. The modified propositions. The lack of success from a direct use of results from [GPY1] for the proof of a positive proportion result notwithstanding, a version of (2.10) and (2.11) in which the n with $P_{\mathcal{H}}(n)$ having small prime factors are discounted vouchsafes the solution. We define

$$\mathcal{P}(x) := \prod_{p_n\le x} p_n. \tag{3.1}$$

We shall use the following results which are consequences of (2.10), (2.11) and Lemmas 4 and 5 of Pintz's work [P].

PROPOSITION 1. *For $N^{c_1} \le R \le N^{1/(2+\delta)}(\log N)^{-c_2}$ where c_1 and c_2 are suitably chosen constants depending on k and $\ell \asymp \sqrt{k}$ (c_1 can be taken to be $1/5$ and c_2 is sufficiently large), $\delta > 0$ small compared to $k^{-3/2}$, and $\mathcal{H}$ admissible with $h \ll \log R$ and $h \to \infty$ with N, we have*

$$\sum_{\substack{N<n\le 2N\\ (P_{\mathcal{H}}(n),\mathcal{P}(R^\delta))=1}} \Lambda_R(n;\mathcal{H},\ell)^2 \sim (1+O(k^3\delta^2))\binom{2\ell}{\ell}\frac{\mathfrak{S}(\mathcal{H})}{(k+2\ell)!}N(\log R)^{k+2\ell}. \tag{3.2}$$

PROPOSITION 2. *Upon the conditions of Proposition 1 and the notation introduced in connection with (2.11), if the level of distribution of primes is $\vartheta \ge 1/2$, then for $N^{c_1} \le R \le N^{(\vartheta-\epsilon)/(2+\delta)}(\log N)^{-c_2}$ ($\epsilon > 0$ arbitrarily small but fixed) and $\mathcal{H} \cup \{h_0\}$ admissible, we have*

$$\sum_{\substack{N<n\le 2N\\ (P_{\mathcal{H}}(n),\mathcal{P}(R^\delta))=1}} \theta(n+h_0)\Lambda_R(n;\mathcal{H},\ell)^2 \sim (1+O(k^3\delta^2))\binom{2(\ell+m)}{\ell+m}\frac{\mathfrak{S}(\mathcal{H}\cup\{h_0\})}{(k+2\ell+m)!}N(\log R)^{k+2\ell+m}; \tag{3.3}$$

in case $\mathcal{H}\cup\{h_0\}$ is not admissible, instead of the right-hand side of (3.3) we have $o(N(\log R)^{k+2\ell+m})$.

Proof. From [P], along with (2.10) and (2.11), these results are obvious except that the present error term $O(k^3\delta^2)$ meant with an absolute constant comes out as $O(\delta)$ with the constant implied depending on k and ℓ. Since we shall use the dependence on k and ℓ of the error term, we give its proof. An examination of the proof of Pintz's Lemma 3 reveals that we need to have more precise versions of formulas (6.17), (6.18), (6.25) and (6.26) of [P]. First we evaluate

$$\mathcal{T}_{q,1}(1+\alpha) := \frac{1}{\ell!}\left[\left(\frac{d}{d\xi}\right)^{\ell}\left(\frac{(1+\alpha+\xi)^{k+2\ell}}{(1+\xi)^k}\right)\right]_{\xi=0} = (1+\alpha)^{k+\ell}\sum_{m=0}^{\ell}\binom{2\ell-m}{\ell}\binom{k+m-1}{m}(-\alpha)^m. \tag{3.4}$$

This indicates that it would be opportune to restrict $|\alpha|$ to values small compared to $1/k$. Assuming this, and recalling that $\ell \asymp \sqrt{k}$, from (3.4) we see

$$\mathcal{T}_{q,1}(1+\alpha) = \binom{2\ell}{\ell}\left[1+\left(\frac{k}{2}+\ell\right)\alpha+\left(\frac{k^2}{8}+\frac{k\ell}{2}-\frac{3k}{8}+\frac{\ell^2}{2}-\frac{\ell}{2}-\frac{k(k+1)}{8(2\ell-1)}\right)\alpha^2+O\big((k\alpha)^3\big)\right]. \tag{3.5}$$

We will denote by K the coefficient of α^2 in the last line. This is used in (6.8) of [P]. We recall that the prime number q has the value R^β in the statement of Lemma 3. In the last factor of the integrand of (6.8) there are four terms. With the notation introduced in (6.11) of [P] we have the following. The first term has $R_1 = R_2 = R$, so that $\alpha = 0$, and we get the contribution

$$\binom{2\ell}{\ell}\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}G_q(0,0).$$

The second term has $R_1 = R/q$, $R_2 = R$, so that $\alpha = -\beta$, and we get the contribution

$$\binom{2\ell}{\ell}\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}G_q(0,0)\Big[1-\Big(\frac{k}{2}+\ell\Big)\beta+K\beta^2+O((k\beta)^3)\Big].$$

The third term has $R_1 = R$, $R_2 = R/q$, so that $\alpha = \frac{\beta}{1-\beta}$, and we get the contribution

$$\binom{2\ell}{\ell}\frac{(\log R^{1-\beta})^{k+2\ell}}{(k+2\ell)!}G_q(0,0)\Big[1+\Big(\frac{k}{2}+\ell\Big)\frac{\beta}{1-\beta}+K\Big(\frac{\beta}{1-\beta}\Big)^2+O((k\beta)^3)\Big].$$

The fourth term has $R_1 = R_2 = R/q$, so that $\alpha = 0$, and we get the contribution

$$\binom{2\ell}{\ell}\frac{(\log R^{1-\beta})^{k+2\ell}}{(k+2\ell)!}G_q(0,0).$$

Combining these as in (6.8) of [P] we obtain

(3.6)
$$\binom{2\ell}{\ell}\frac{(\log R)^{k+2\ell}}{(k+2\ell)!}G_q(0,0)\Big[\Big(\frac{k^2}{4}+k\ell+\ell^2+\frac{k}{4}+\frac{k(k+1)}{4(2\ell-1)}\Big)\beta^2+O((k\beta)^3)\Big]$$

in place of the main term of (6.25) of [P]. Hence in the new version for (6.1) of [P], instead of β/q we have

$$\frac{\nu_q(\mathcal{H})}{q}\,\frac{G_q(0,0)}{G(0,0)}\Big[\Big(\frac{k^2}{4}+k\ell+\ell^2+\frac{k}{4}+\frac{k(k+1)}{4(2\ell-1)}\Big)\beta^2+O((k\beta)^3)\Big]$$
$$=\frac{\nu_q(\mathcal{H})}{q}\Big(1-\frac{\nu_q(\mathcal{H})}{q}\Big)^{-1}\Big(\frac{k^2\beta^2}{4}+\text{smaller terms}\Big),$$

i.e. we can express the new version of Lemma 3 of [P] as

$$\sum_{\substack{N<n\le 2N\\ q\mid P_{\mathcal{H}}(n)}}\Lambda_R(n;\mathcal{H},\ell)^2 \le \frac{\nu_q(\mathcal{H})}{q-\nu_q(\mathcal{H})}\,\frac{k^2\beta^2}{3}\sum_{N<n\le 2N}\Lambda_R(n;\mathcal{H},\ell)^2. \tag{3.7}$$

Here $N^{c_1} \le R \le N^{1/(2+\delta)}(\log N)^{-c_2}$, q is a prime number for which we write $q = R^\beta$, with $0 < \beta \le \delta$ where δ is small compared to $k^{-3/2}$ say, k is sufficiently large, and $\ell \asymp \sqrt{k}$.

We know that $\nu_q(\mathcal{H}) \le \min(q-1,k)$ since $\mathcal{H}$ is admissible. For $q \le k$, we have $\nu_q(\mathcal{H}) \le q-1$, so that $\frac{\nu_q(\mathcal{H})}{q-\nu_q(\mathcal{H})} \le q-1$. For $q > k$, we take $\nu_q(\mathcal{H}) \le k$, so that $\frac{\nu_q(\mathcal{H})}{q-\nu_q(\mathcal{H})} \le \frac{k}{q-k}$. Summing over all primes $q \le R^\delta$ we obtain a new version of Lemma 4 of [P] as

$$\sum_{\substack{N<n\le 2N\\(P_{\mathcal{H}}(n),\mathcal{P}(R^\delta))>1}} \Lambda_R(n;\mathcal{H},\ell)^2 \le \frac{k^3\delta^2}{4} \sum_{N<n\le 2N} \Lambda_R(n;\mathcal{H},\ell)^2 \tag{3.8}$$

if k is large enough. We see that we have to choose δ small enough so that $k^3\delta^2$ will be small. Now by (2.10) and (2.11) we immediately obtain (3.2). When there is the twisting with primes, the proof runs similarly and Proposition 2 also follows. ∎

4. Proof of Theorem 1. Our aim is to obtain an inequality of the form

$$\sum_{\substack{N<p_j\le 2N\\ p_{j+1}-p_j\le h}} 1 \gg_\eta \pi(N) \sim \frac{N}{\log N} \quad (N\to\infty) \tag{4.1}$$

for

$$h = \eta \log N, \quad \eta > 0 \text{ arbitrarily small but fixed.} \tag{4.2}$$

Let

$$Q(N,h) := \sum_{\substack{N<n\le 2N\\ \pi(n+h)-\pi(n)>1}} 1. \tag{4.3}$$

If n is an integer for which $\pi(n+h)-\pi(n) > 1$, then there must be a j such that $n < p_j$ and $p_{j+1} \le n+h$. Thus $p_{j+1}-p_j < h$ and $p_{j+1}-h \le n < p_j$, so that there are less than $\lfloor h \rfloor$ such integers n corresponding to each such gap. Therefore

$$Q(N,h) \le h \sum_{\substack{N<p_j\le 2N\\ p_{j+1}-p_j\le h}} 1 + O(Ne^{-C\sqrt{\log N}}), \tag{4.4}$$

where we have used the prime number theorem with error term to remove the prime gaps which overlap the endpoints (this is explicitly shown in [GY1]).

Instead of S_R which was defined in (2.7), we will work with

$$\tilde{S}_R := \frac{1}{N(h\log R)^k} \sum_{N<n\le 2N} (\Theta(n,h) - \log 3N)\Big(\sum_{\mathcal{H}}{}^{*} \Lambda_R(n;\mathcal{H},\ell)^2\Big), \tag{4.5}$$

where

$$\sum_{\mathcal{H}}{}^{*} := \sum_{\substack{\mathcal{H}\subset[1,h],|\mathcal{H}|=k\\ \mathcal{H}\text{ admissible}\\ (P_{\mathcal{H}}(n),\mathcal{P}(R^\delta))=1}} . \tag{4.6}$$

We note that, as a function of η, k and ℓ will be chosen sufficiently large but fixed, and $\delta > 0$ will be chosen sufficiently small but fixed (see (4.22) below).

From (4.5) we have, when N is sufficiently large,

$$\tilde{S}_R \le \frac{1}{N(h\log R)^k} \sum_{\substack{N<n\le 2N\\ \Theta(n,h)\ge \frac{3}{2}\log N}} \Theta(n,h) \sideset{}{^*}\sum_{\mathcal{H}} \Lambda_R(n;\mathcal{H},\ell)^2 \tag{4.7}$$

$$\le \frac{1}{N(h\log R)^k}\Big\{ \sum_{\substack{N<n\le 2N\\ \Theta(n,h)\ge \frac{3}{2}\log N}} 1\Big\}^{1/2} \Big\{ \sum_{N<n\le 2N} \Theta(n,h)^2 \Big(\sideset{}{^*}\sum_{\mathcal{H}} \Lambda_R(n;\mathcal{H},\ell)^2\Big)^2\Big\}^{1/2}$$

$$= \frac{Q(N,h)^{1/2}}{N(h\log R)^k} I^{1/2},$$

where

$$I = \sum_{1\le h',h''\le h} \sum_{\substack{\mathcal{H}_i\subset[1,h],|\mathcal{H}_i|=k\\ \mathcal{H}_i \text{ admissible}\\ i=1,2}} \sum_{\substack{N<n\le 2N\\ (P_{\mathcal{H}_1\cup\mathcal{H}_2}(n),\mathcal{P}(R^\delta))=1}} \theta(n+h')\theta(n+h'') \times \Lambda_R(n;\mathcal{H}_1,\ell)^2\Lambda_R(n;\mathcal{H}_2,\ell)^2. \tag{4.8}$$

Here for a number n to make a nonzero contribution, both of $n+h'$ and $n+h''$ must be prime, so that $((n+h')(n+h''),\mathcal{P}(R^\delta)) = 1$ and writing $\mathcal{H}_0 = \{h'\}\cup\{h''\}\cup\mathcal{H}_1\cup\mathcal{H}_2$ we can re-express the condition on n as $(P_{\mathcal{H}_0}(n),\mathcal{P}(R^\delta)) = 1$. We also observe that, since all prime factors of $P_{\mathcal{H}}(n)$ in $\sum^*_{\mathcal{H}}$ are greater than R^δ, the number of squarefree divisors of $P_{\mathcal{H}}(n)$ is at most $2^{\frac{k\log 3N}{\delta\log R}}$. Hence for any term in $\sum^*_{\mathcal{H}}$ we have

$$|\Lambda_R(n;\mathcal{H},\ell)| \le \frac{2^{\frac{k\log 3N}{\delta\log R}}}{(k+\ell)!}(\log R)^{k+\ell}. \tag{4.9}$$

Such an estimate could not be written if we had $\delta = 0$, which is the case with (2.10) and (2.11). Thus this is the crucial point which makes the current argument work. Substituting (4.9) in (4.8), we have

$$I \le \frac{2^{\frac{4k\log 3N}{\delta\log R}}(\log R)^{4(k+\ell)}(\log 3N)^2}{(k+\ell)!^4} \sum_{1\le h',h''\le h} \sum_{\substack{\mathcal{H}_i\subset[1,h],|\mathcal{H}_i|=k\\ \mathcal{H}_i \text{ admissible}\\ i=1,2}} \sum_{\substack{N<n\le 2N\\ (P_{\mathcal{H}_0}(n),\mathcal{P}(R^\delta))=1}} 1. \tag{4.10}$$

For a given $\mathcal{H}_0\subset\{1,\dots,h\}$ with $|\mathcal{H}_0| = k+r$, $0\le r\le k+2$, denoting by $D(k,r)$ the number of quadruples $h',h'',\mathcal{H}_1,\mathcal{H}_2$ corresponding to $\mathcal{H}_0$, we

re-express (4.10) as

(4.11)
$$I \le \frac{2^{\frac{4k\log 3N}{\delta\log R}}(\log R)^{4(k+\ell)}(\log 3N)^2}{(k+\ell)!^4}\sum_{r=0}^{k+2}D(k,r)\sum_{|\mathcal{H}_0|=k+r}\sum_{\substack{N<n\le 2N\\(P_{\mathcal{H}_0}(n),\mathcal{P}(R^\delta))=1}}1.$$

We now invoke the main theorem of Selberg's upper bound sieve ([HR, Theorem 5.1] or [Gr, §2.2.2, Theorem 2]) that for any set $\mathcal{H}$ and $\delta < 1/2$,

$$\sum_{\substack{N<n\le 2N\\(P_{\mathcal{H}}(n),\mathcal{P}(R^\delta))=1}}1 \le \frac{N|\mathcal{H}|!\mathfrak{S}(\mathcal{H})}{(\log R^\delta)^{|\mathcal{H}|}}(1+o(1)) \quad (N\to\infty), \tag{4.12}$$

which gives upon using (2.11) that

(4.13)
$$\begin{aligned} I &\lesssim N\frac{2^{\frac{4k\log 3N}{\delta\log R}}(\log R)^{4(k+\ell)}(\log N)^2}{(k+\ell)!^4}\sum_{r=0}^{k+2}\frac{(k+r)!D(k,r)}{(\delta\log R)^{k+r}}\sum_{\mathcal{H}_0;\,|\mathcal{H}_0|=k+r}\mathfrak{S}(\mathcal{H}_0)\\ &\lesssim N\frac{2^{\frac{4k\log 3N}{\delta\log R}}(\log R)^{4(k+\ell)}(\log N)^2}{(k+\ell)!^4}\sum_{r=0}^{k+2}(k+r)!D(k,r)\left(\frac{h}{\delta\log R}\right)^{k+r}. \end{aligned}$$

To deal with the inner sum here, first note that

$$D(k,r) = \frac{k!^2(k+r)!(k^4+3k^3+(3r+2)k^2+4rk+r^2)}{r!^2(k+2-r)!} \tag{4.14}$$

(here the factor $k!^2$ comes from the ordering of the elements within the k-tuples $\mathcal{H}_1$ and $\mathcal{H}_2$). We skip the proof of (4.14) since it follows from an elementary combinatorial calculation, and after our choice of parameters the order of magnitude is much smaller than that of the heftiest factor $2^{\frac{4k\log 3N}{\delta\log R}}$. Now for $u>0$, we have

$$\begin{aligned} \sum_{r=0}^{k+2}(k+r)!D(k,r)u^{k+r} &\le u^k(k+1)^2(k+2)^2\sum_{r=0}^{k+2}\frac{k!^2(k+r)!^2}{r!^2(k+2-r)!}u^r\\ &= u^k(k+2)!\sum_{r=0}^{k+2}\frac{(k+r)!^2}{r!}\binom{k+2}{r}u^r \le (2k+2)!^2u^k(1+u)^{k+2}, \end{aligned} \tag{4.15}$$

so that

(4.16)
$$I \lesssim N(\log R)^{4(k+\ell)}(\log N)^2\frac{(2k+2)!^2}{(k+\ell)!^4}2^{\frac{4k\log 3N}{\delta\log R}}\left(\frac{h}{\delta\log R}\right)^k\left(1+\frac{h}{\delta\log R}\right)^{k+2}.$$

Using (4.16) and (4.4) in (4.7), we obtain

$$\tilde{S}_R \lesssim \Big(h \sum_{\substack{N<p_j\le 2N\\ p_{j+1}-p_j\le h}} 1 + O(Ne^{-C\sqrt{\log N}})\Big)^{1/2} \tag{4.17}$$
$$\times \frac{(\log R)^{(k+2\ell)}\log N}{N^{1/2}h^k}\,\frac{(2k+2)!}{(k+\ell)!^2}\,2^{\frac{2k\log 3N}{\delta\log R}}\left(\frac{h}{\delta\log R}\right)^{k/2}\left(1+\frac{h}{\delta\log R}\right)^{(k+2)/2}.$$

Now we calculate $\tilde{S}_R$ using Propositions 1 and 2. From Proposition 1 and (2.12) we see that

$$\sum_{N<n\le 2N}(\log 3N)\sum_{\mathcal{H}}{}^{*}\Lambda_R(n;\mathcal{H},\ell)^2 \tag{4.18}$$
$$\sim \big(1+O(k^3\delta^2)\big)\binom{2\ell}{\ell}\frac{h^k}{(k+2\ell)!}N(\log R)^{k+2\ell}\log N.$$

Similarly, Proposition 2 and (2.12) imply

$$\sum_{\substack{\mathcal{H}\subset[1,h],\,|\mathcal{H}|=k\\ \mathcal{H}\text{ admissible}}}\sum_{h_i\in\mathcal{H}}\sum_{\substack{N<n\le 2N\\ (P_{\mathcal{H}}(n),\mathcal{P}(R^\delta))=1}}\theta(n+h_i)\Lambda_R(n;\mathcal{H},\ell)^2 \tag{4.19}$$
$$\sim \big(1+O(k^3\delta^2)\big)\binom{2\ell+2}{\ell+1}\frac{kh^k}{(k+2\ell+1)!}N(\log R)^{k+2\ell+1},$$

and

$$\sum_{\substack{\mathcal{H}\subset[1,h],\,|\mathcal{H}|=k\\ \mathcal{H}\text{ admissible}}}\sum_{\substack{1\le h_0\le h\\ h_0\notin\mathcal{H}}}\sum_{\substack{N<n\le 2N\\ (P_{\mathcal{H}}(n),\mathcal{P}(R^\delta))=1}}\theta(n+h_0)\Lambda_R(n;\mathcal{H},\ell)^2 \tag{4.20}$$
$$\gtrsim \big(1+O(k^3\delta^2)\big)\binom{2\ell}{\ell}\frac{h^{k+1}}{(k+2\ell)!}N(\log R)^{k+2\ell}.$$

Putting (4.18)–(4.20) in (4.5) we obtain

$$\tilde{S}_R \gtrsim \tag{4.21}$$
$$\frac{\binom{2\ell}{\ell}}{(k+2\ell)!}(\log N)(\log R)^{2\ell}\left\{\frac{k}{k+2\ell+1}\,\frac{2(2\ell+1)}{\ell+1}\,\frac{\log R}{\log N}+\eta-1+O(k^3\delta^2)\right\}.$$

Now given a small fixed $\eta>0$ if we take

$$\ell=\lfloor 4/\eta\rfloor,\quad k=2(\ell+1)(2\ell+1),\quad \delta=1/\ell^4,\quad R=N^{1/4(1+\delta)}, \tag{4.22}$$

then for the factor in brackets in (4.21) we see that

$$\frac{k}{k+2\ell+1}\,\frac{2(2\ell+1)}{\ell+1}\,\frac{\log R}{\log N}+\eta-1+O(k^3\delta^2)>\frac{\eta}{2} \tag{4.23}$$

for sufficiently small η, so that $\tilde{S}_R>0$.

Noting that (4.2) and (4.22) imply $\frac{h}{\delta \log R} = 4\eta(\ell^4+1) > 16\frac{\ell^4+1}{\ell+1} \ge 16$, we will use $1+\frac{h}{\delta \log R} < \frac{2h}{\delta \log R}$. Then, from (4.17), (4.21) and (4.23), we have

$$\sum_{\substack{N<p_j\le 2N\\ p_{j+1}-p_j\le h}} 1 \gtrsim \frac{N}{\log N}\,\frac{1}{2^{\frac{4k\log 3N}{\delta\log R}}}\,\frac{\binom{2\ell}{\ell}^2 (k+\ell)!^4\delta^{2k+2}}{\eta(k+2\ell)!^2(2k+2)!^2 2^{k+10}}. \tag{4.24}$$

With the values specified in (4.22), the dominating factor in the coefficient on the right-hand side of (4.16) is

$$2^{-\frac{4k\log 3N}{\delta\log R}} > e^{-65(4/\eta)^6\log 2}. \tag{4.25}$$

The other factors in (4.24) give rise to exponents which are $O\big(\frac{1}{\eta^2}\log\frac{1}{\eta}\big)$. Thus the proof of Theorem 1 is finished. (This is not the strongest estimate the present method yields. Taking $\delta = 1/\ell^{3+c_3}$ with any fixed $c_3 > 1/2$ leads to an estimate of the type (1.3) with $\eta^{-(5+c_3)}$ instead of η^{-6}.)

Note that we could have kept $Q(N,h)/h$ as the left-hand side of (4.24), which yields $Q(N,h) \gg N$, meaning that the proportion of integers $n \in [N,2N]$ for which one can find at least two primes within a distance of h from n is positive.

5. Sparsity of very small gaps between primes. The following result expresses that very small gaps between consecutive primes occur rarely, in the sense that such gaps do not constitute a positive proportion of all gaps between consecutive primes.

THEOREM 2. *For any $h > 2$, as $x \to \infty$, we have*

$$\#\{p_n \le x;\, p_{n+1}-p_n \le h\} \ll \min(h/\log x, 1)\,\pi(x). \tag{5.1}$$

In particular, if $h = o(\log x)$, then

$$\#\{p_n \le x;\, p_{n+1}-p_n \le h\} = o(\pi(x)). \tag{5.2}$$

We remark that for $h = \eta\log x$, $0 < \eta < 1$, the upper estimate for the density of small gaps given by (5.1) corresponds to the conjectured density $1-e^{-\eta}$ from (1.2) apart from the constant implied by the $\ll$ symbol; i.e. the simple upper estimate argument in the proof given below is optimal except for this constant.

Proof. Given two prime numbers p, p' satisfying $0 < p'-p \le h$, let us write $u+h_1 = p$, $u+h_2 = p'$. There are h ordered pairs (u,h_1) with $1 \le h_1 \le h$ such that $u+h_1 = p$, and for any ordered pair (u,h_1) the value

of h_2 with $h_1 < h_2 \le 2h$ is fixed. Hence we see that

$$h \sum_{\substack{N<p,p'\le 2N\\ 0<p'-p\le h}} 1 < \sum_{\substack{1\le h_1,h_2\le 2h\\ h_1\ne h_2}} \sum_{\substack{N/2<u<3N\\ u+h_1,\,u+h_2 \text{ prime}}} 1 \ll \sum_{\substack{1<h_1,h_2\le 2h\\ h_1\ne h_2}} \mathfrak{S}(\{h_1,h_2\})\frac{N}{(\log N)^2} \ll \frac{h^2N}{(\log N)^2}, \tag{5.3}$$

where we have used the well-known (see [HR, Theorem 5.7] or [Gr, §2.3.3, Theorem 4]) sieve bound for prime tuples

$$\sum_{N<n\le 2N} \theta(n+h_1)\cdots\theta(n+h_k) \lesssim 2^k k!\mathfrak{S}(\mathcal{H})N \tag{5.4}$$

with $k = 2$, and Gallagher's result (2.12). Thus we have obtained

$$\sum_{\substack{N<p,p'\le 2N\\ 0<p'-p\le h}} 1 \ll \frac{hN}{(\log N)^2} \ll \frac{h}{\log N}\pi(N). \ \blacksquare \tag{5.5}$$

Note that (5.4) used with $k = 3$ shows that of the p, p' in (5.5), the number of those which are not consecutive is $\ll (h/\log N)^2\pi(N)$.

6. Conditional results. For the circumstance specified by (4.2) we shall now consider the consequence of assuming that the level of distribution of primes ϑ is greater than $1/2$. The conditions of Propositions 1 and 2 allow us to take

$$R = N^{\frac{\vartheta-\epsilon}{2(1+\delta)}} \tag{6.1}$$

with ϵ and δ arbitrarily small fixed positive numbers. We let

$$\ell = \lfloor \sqrt{k}/2 \rfloor. \tag{6.2}$$

For a given $\vartheta > 1/2$, we determine $k = k(\vartheta)$ sufficiently large and ϵ and δ small enough so as to ensure that the quantity $\frac{k}{k+2\ell+1}\frac{2(2\ell+1)}{\ell+1}\frac{\log R}{\log N} - 1$ occurring in (4.21) is positive. Now k is not necessarily large enough to satisfy (3.8) and the corresponding inequality when there is the twisting with primes, so instead of the error term $O(k^3\delta^2)$ in Propositions 1 and 2, and in (4.21) we will have the cruder $O_k(\delta)$ (or else we can re-do the calculation as of (3.4) up until (3.8) without having error terms in what will correspond to (3.5) and (3.6), but this will not be necessary for our purpose). By choosing a smaller δ if necessary, we will have the factor in brackets in (4.21) (with 0 in place of η) greater than a positive quantity which ultimately depends only on ϑ. Hence, comparing (4.17) and (4.21) we immediately obtain

THEOREM 3. *Assume that the primes have a level of distribution $\vartheta > 1/2$. Let η be a fixed positive small number. Then there exists an integer $k(\vartheta)$ and*

a constant $c_4(\vartheta)$ such that

$$\sum_{\substack{N<p_j\le 2N\\ p_{j+1}-p_j\le \eta\log N}} 1 \gtrsim c_4(\vartheta)\eta^{k(\vartheta)-1}\frac{N}{\log N} \quad (N\to\infty). \tag{6.3}$$

Notice that the unconditional estimate (1.3) in which η takes place exponentially, gets improved to estimates involving just powers of η when it is assumed that the primes have a level of distribution greater than $1/2$. By comparing the factor in brackets in (4.21) with the corresponding factor in the argument in [GPY1], we see that the smallest possible $k(\vartheta)$ we can assert is either the smallest $r = r(\vartheta)$ such that every admissible r-tuple is guaranteed by the proof of Theorem 1 of [GPY1] to contain at least two primes infinitely often, or it is $r+1$ (depending on the value of ϑ). The greater the level of distribution, the smaller power of η will be needed in (6.3). A table of values of $r(\vartheta)$ was provided between (3.4) and (3.5) of [GPY1] (to avoid confusion, we have renamed the k of that table as r here). Thus, if $\vartheta > \frac{20}{21}$, then we can take $k = 7$, $\ell = 1$, so that the η-dependent factor on the right-hand side of (6.3) is η^6. However, we recall that assuming $\vartheta \ge 0.971$ and considering a linear combination of the $\Lambda_R(n;\mathcal{H},\ell)$ with $k = 6$ and $\ell = 0, 1$, the argument for proving Theorem 1 of [GPY1] still works, so that under this assumption we can get a lower bound in (6.3) which has η^5. We also see from (1.2) that the true order of magnitude of the η-dependent factor on the right-hand side of (6.3) is believed to be η. Thus for this argument to lead to the true order of magnitude we need to be able to work with admissible pairs (2-tuples). But this seems to require improving the results of [GPY1] to the extent of proving the twin prime conjecture assuming the Elliott–Halberstam conjecture.

When ϑ is slightly greater than $1/2$, from the condition in Proposition 2 we write

$$R = N^{\frac{1/2+\xi}{2(1+\delta)}}, \tag{6.4}$$

where we assume that $\xi > 0$ is small. We take

$$k = 2(\ell+1)(2\ell+1), \quad \delta = 1/\ell^4, \tag{6.5}$$

so that

$$\frac{k}{k+2\ell+1}\,\frac{2(2\ell+1)}{\ell+1}\,\frac{\log R}{\log N} + \eta - 1 + O(k^3\delta^2) = \eta + 2\xi - \frac{1}{\ell} - \frac{2\xi}{\ell} + O\Big(\frac{1}{\ell^2}\Big). \tag{6.6}$$

For a given ξ, we determine ℓ by

$$\ell = \lceil 1/\xi \rceil, \tag{6.7}$$

and then the quantity in (6.6) is $> \eta + \xi/2$ if ξ is sufficiently small. Hence from (4.21) we now have

$$\tilde{S}_R > \frac{\binom{2\ell}{\ell}}{(k+2\ell)!}(\log N)(\log R)^{2\ell}\bigg(\eta + \frac{\xi}{2}\bigg). \tag{6.8}$$

As before, we derive an upper bound for $\tilde{S}_R$ starting from (4.7), together with (4.8), (4.13) and (4.15). In our case $u = h/(\delta \log R)$, and upon using the relations in (4.2), (6.4), (6.5) and (6.7), we have

$$u = \frac{4(1+\lceil 1/\xi\rceil^4)}{1+2\xi}\eta. \tag{6.9}$$

This is a small quantity if for a given small ξ we take η small enough, say $\eta \le \xi^4/5$, so that we can say $(1+u)^{k+2} < 2^{k+2}$. Using this in (4.15) and (4.13) gives

$$I < N(\log R)^{4(k+\ell)}(\log N)^2\bigg(\frac{h}{\delta\log R}\bigg)^k \frac{2^{\frac{4k\log 3N}{\delta\log R}+k+2}(2k+2)!^2}{(k+\ell)!^4}. \tag{6.10}$$

Plugging this in (4.7), and using that together with (4.4) and (6.8) we obtain

$$\sum_{\substack{N<p_j\le 2N\\ p_{j+1}-p_j\le h}} 1 \gtrsim \frac{N}{\log N}\eta^{k-1}\bigg(\frac{\xi}{3}\bigg)^2\bigg(\frac{4\delta(1+\delta)}{1+2\xi}\bigg)^k \frac{\binom{2\ell}{\ell}^2(k+\ell)!^4k!^2}{(k+2\ell)!^2(2k+2)!^2 2^{\frac{4k\log 3N}{\delta\log R}+k+2}}. \tag{6.11}$$

By (6.4), (6.5) and (6.7), all of the factors after η^{k-1} can be expressed in terms of ξ. What interests us most is the power of η, so we re-express (6.11) as

$$\sum_{\substack{N<p_j\le 2N\\ p_{j+1}-p_j\le \eta\log N}} 1 \gtrsim c_5(\xi)\eta^{4\xi^{-2}+14\xi^{-1}+11}\frac{N}{\log N} \quad (N\to\infty), \tag{6.12}$$

valid when the level of distribution of primes is assumed to allow us to take R as in (6.4), which can be re-written as

$$R = N^{\frac{1+2\xi}{4(1+\lceil 1/\xi\rceil^{-4})}}, \tag{6.13}$$

for fixed $\eta \in (0, \xi^4/5]$. Here ξ has to be sufficiently small, which ensures that ℓ and k are sufficiently large so as to permit the inequality (6.8).

In [GPY1, §3] it was shown that under the Elliott–Halberstam conjecture we have

$$\liminf_{n\to\infty}\frac{p_{n+2}-p_n}{\log p_n} = 0. \tag{6.14}$$

Our method also shows that such gaps occur in positive proportion. To see

this we consider

$$\tilde{S}_{R,2} := \frac{1}{N(h\log R)^k}\sum_{N<n\le 2N}(\Theta(n,h)-2\log 3N)\Big(\sum_{\mathcal{H}}{}^{*}\,\Lambda_R(n;\mathcal{H},\ell)^2\Big). \tag{6.15}$$

As was done in [GPY1] along with the modification provided by (3.8), we find

$$\tilde{S}_{R,2} \gtrsim \frac{\binom{2\ell}{\ell}}{(k+2\ell)!}(\log N)(\log R)^{2\ell}\left\{\frac{k}{k+2\ell+1}\,\frac{2(2\ell+1)}{\ell+1}\,\frac{\log R}{\log N}+\eta-2-k^3\delta^2\right\}. \tag{6.16}$$

Here we are assuming that $\vartheta = 1$, and so we can take $R = N^{1/(2(1+\delta))}$. In the proof of (6.14), k is taken to be sufficiently large, and $\ell = \lfloor\sqrt{k}/2\rfloor$. If δ is taken to be accordingly small, say $\delta = 1/k^2$, then the quantity in brackets in (6.16) is

$$\begin{aligned}
&> 2\left(1-\frac{2\ell+1}{k}-\frac{1}{2\ell}\right)(1-\delta)+\eta-2-k^3\delta^2 \\
&> \eta-\frac{2(2\ell+1)}{k}-\frac{1}{\ell}-2\delta-k^3\delta^2 \\
&> \eta-\frac{2(\sqrt{k}+1)}{k}-\frac{2}{\sqrt{k}-2}-\frac{2}{k^2}-\frac{1}{k} \\
&> \eta-\frac{5}{\sqrt{k}}-\frac{3}{k}-\frac{2}{k^2} \quad (\text{for } k>36) \\
&> \eta-\frac{6}{\sqrt{k}} \\
&> \eta/2 \quad (\text{for } k>144/\eta^2).
\end{aligned} \tag{6.17}$$

The rest of the argument is almost identical to what was done as of (4.7), the only changes are that we now have the summation condition $\Theta(n,h)\ge \frac{5}{2}\log N$, and $\frac{h}{\delta\log R} = 2\eta\big(1+\frac{1}{\delta}\big)$ being not small we should use some bound like $(1+u)^{k+2}\le(2u)^{k+2}$ (cf. between (6.9) and (6.10)). The following is the result of this calculation.

THEOREM 4. *Assuming the Elliott–Halberstam conjecture we have*

$$\sum_{\substack{N<p_j\le 2N\\ p_{j+2}-p_j\le\eta\log N}} 1 \gtrsim e^{c_6\eta^{-2}\log\eta}\frac{N}{\log N} \qquad (N\to\infty) \tag{6.18}$$

($c_6 = 5$ gives a valid result if η is small enough).

Acknowledgements. The first author was supported by NSF; this study was done while the second and third authors were members at the Institute for Advanced Study, Princeton during Fall 2009 and were supported by the Oswald Veblen Fund; the second author also acknowledges the partial support of ERC-AdG. No. 228005.

References

[BLZ] D. Bazzanella, A. Languasco and A. Zaccagnini, *Prime numbers in logarithmic intervals*, Trans. Amer. Math. Soc. 362 (2010), 2667–2684.

[BD] E. Bombieri and H. Davenport, *Small differences between prime numbers*, Proc. Roy. Soc. Ser. A 293 (1966), 1–18.

[Ga] P. X. Gallagher, *On the distribution of primes in short intervals*, Mathematika 23 (1976), 4–9; Corrigendum, ibid. 28 (1981), 86.

[GPY1] D. A. Goldston, J. Pintz and C. Y. Yıldırım, *Primes in tuples I*, Ann. of Math. 170 (2009), 819–862.

[GPY2] D. A. Goldston, J. Pintz and C. Y. Yıldırım, *Primes in tuples II*, Acta Math. 204 (2010), 1–47.

[GPY3] D. A. Goldston, J. Pintz and C. Y. Yıldırım, *Positive proportion of small gaps between consecutive primes*, Publ. Math. Debrecen 79 (2011), 433–444.

[GY1] D. A. Goldston and C. Y. Yıldırım, *Higher correlations of divisor sums related to primes I: Triple correlations*, Integers 3 (2003), A5, 66 pp.

[GY2] D. A. Goldston and C. Y. Yıldırım, *Higher correlations of divisor sums related to primes III: Small gaps between primes*, Proc. London Math. Soc. (3) 95 (2007), 653–686.

[Gr] G. Greaves, *Sieves in Number Theory*, Springer, Berlin, 2001.

[HR] H. Halberstam and H.-E. Richert, *Sieve Methods*, Dover, New York, 2011.

[Hu] M. N. Huxley, *Small differences between consecutive primes II*, Mathematika 24 (1977), 142–152.

[M] H. Maier, *Small differences between prime numbers*, Michigan Math. J. 35 (1988), 323–344.

[P] J. Pintz, *Are there arbitrarily long arithmetic progressions of twin primes?*, in: An Irregular Mind, I. Barany and J. Solymosi (eds.), Bolyai Soc. Math. Stud. 21, János Bolyai Math. Soc., Budapest, 2010, 525–559.

[S1] K. Soundararajan, *Small gaps between prime numbers: The work of Goldston–Pintz–Yıldırım*, Bull. Amer. Math. Soc. (N.S.) 44 (2007), 1–18.

[S2] K. Soundararajan, *The distribution of prime numbers*, in: Equidistribution in Number Theory, an Introduction, NATO Sci. Ser. II Math. Phys. Chem. 237, Springer, Dordrecht, 2007, 59–83.

Daniel Alan Goldston
Department of Mathematics
San Jose State University
San Jose, CA 95192, U.S.A.
E-mail: goldston@math.sjsu.edu

János Pintz
Rényi Mathematical Institute
of the Hungarian Academy of Sciences
H-1364 Budapest, Pf 127, Hungary
E-mail: pintz@renyi.hu

Cem Yalçın Yıldırım
Department of Mathematics
Boğaziçi University
Bebek, İstanbul 34342, Turkey
E-mail: yalciny@boun.edu.tr

Received on 24.10.2012
and in revised form on 18.2.2013 (7241)