On small gaps between primes and almost prime powers

János Pintz*

1. In two subsequent works, joint with D. Goldston and C. Y. Yıldırım [GPY1, GPY2] we showed that for the sequence p_n of primes

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

and even

$$\liminf_{n\to\infty} \frac{p_{n+1}-p_n}{(\log p_n)^{1/2}(\log\log p_n)^2} < \infty. \tag{1.2}$$

A crucial ingredient of the proof was the celebrated Bombieri–Vinogradov theorem, which asserts that $\vartheta = 1/2$ is an admissible level of distribution of primes, that is,

$$\sum_{q\le N^\vartheta/\log^C N} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{p\equiv a \ (\mathrm{mod}\ q) \\ p\le N}} 1 - \frac{li\, N}{\varphi(q)} \right| \ll_A \frac{N}{\log^A N} \tag{1.3}$$

holds with $\vartheta = 1/2$ for any $A > 0$, $C > C(A)$. The method also yielded [GPY1] that if $\vartheta > 1/2$ is an admissible level of distribution of primes then for any *admissible* k-element set $\mathcal{H} = \{h_i\}_{i=1}^k$ (that is, if $\mathcal{H}$ does not occupy all residue classes $\bmod p$ for any prime p) the set $n + \mathcal{H} := \{n + h_i\}_{i=1}^k$ contains at least two primes for infinitely many values of n if $k \ge k_0(\vartheta)$. Consequently we have infinitely many bounded gaps between primes, more precisely

$$\liminf_{n\to\infty}(p_{n+1}-p_n) \le C(\vartheta). \tag{1.4}$$

The strongest possible hypothesis on the uniform distribution of primes in arithmetic progressions, the Elliott–Halberstam [EH] conjecture stating the

*Supported by OTKA Grants K72731, K67676 and ERC-AdG.228005.

admissibility of the level $\vartheta = 1$ (with $N/\log^C N$ replaced by $N^{1-\varepsilon}$ for any $\varepsilon > 0$), or slightly weaker, even the assumption $\vartheta \geq 0.971$ implies gaps of size at most 16 infinitely often, in fact,

$$k_0(0.971) = 6, \qquad C(0.971) = 16. \tag{1.5}$$

If $\vartheta = 1/2 + \delta$ is near to $1/2$, that is, δ is a small positive number, one can take for $\delta \to 0^+$

$$k_0\left(\frac{1}{2} + \delta\right) = \left(2\left\lceil \frac{1}{2\delta} \right\rceil + 1\right)^2, \qquad C\left(\frac{1}{2} + \delta\right) \sim 2\delta^{-2} \log \frac{1}{\delta}. \tag{1.6}$$

This situation suggests that one might take some prime-like set $\mathcal{P}'$ just slightly more dense than the set $\mathcal{P}$ of primes, that is, for any $\varepsilon > 0$ a set $\mathcal{P}_\varepsilon$ such that

$$\mathcal{P} \subset \mathcal{P}'_\varepsilon := \{b_n\}_{n=1}^\infty, \quad \pi'_\varepsilon(N) := \#\{n \leq N,\ n \in \mathcal{P}'\} < \pi(n)(1+\varepsilon) \tag{1.7}$$

which has bounded gaps infinitely often, that is,

$$\liminf_{n\to\infty} (b_{n+1} - b_n) < \infty. \tag{1.8}$$

Of course adding $p+1$ to the set $\mathcal{P}$ for infinitely many primes would trivially satisfy the requirements but we are looking for some arithmetically interesting set $\mathcal{P}'_\varepsilon$ with some similarity to primes or prime powers. (Adding just prime powers to $\mathcal{P}$ raises the number of elements just with a quantity $\sim 2N^{1/2}/\log N$ which is negligible compared to $\pi(N)$.) One possibility is to add some numbers which are similar to prime powers. To avoid confusion with almost primes we will introduce the following

Definition. For any $\varepsilon \geq 0$ a natural number n is called ε-balanced if for any prime divisors p, q of n we have

$$\min(p,q) \geq \big(\max(p,q)\big)^{1-\varepsilon}. \tag{1.9}$$

Remark. With this definition 0-balanced numbers larger than 1 are exactly the primes and prime powers.

Let us denote the set of ε-balanced numbers by $\mathcal{P}_\varepsilon$, the total number of prime divisors of n by $\Omega(n)$ and let

$$\mathcal{P}_{\varepsilon,r} := \big\{n \in \mathcal{P}_\varepsilon,\ \Omega(n) = r\big\}, \quad \mathcal{P}_\varepsilon := \bigcup_{r=1}^\infty \mathcal{P}_{\varepsilon,r}. \tag{1.10}$$

(In this way we can talk about almost prime-squares ($r=2$), almost prime-cubes ($r=3$) etc.)

To have an idea about the quantity

$$\pi_{\varepsilon,r}(N) := \#\{N \le n < 2N;\ n \in \pi_{\varepsilon,r}(N)\} \tag{1.11}$$

we remark that denoting by $P^-(n)$ and $P^+(n)$ the least, resp., the greatest prime factor of n we have obviously

$$n \in \pi_{\varepsilon,r}(N) \Longrightarrow N^{(1-\varepsilon)/r} \le P^-(n) \le P^+(n) \le (2N)^{1/(r(1-\varepsilon))}. \tag{1.12}$$

Reversed, we have also clearly for $n \in [N, 2N)$, $\Omega(n) = r$ by $(1+\varepsilon/2)(1-\varepsilon) \le 1-\varepsilon/2$

$$N^{(1-\varepsilon/2)/r} \le P^-(n) \le P^+(n) \le N^{(1+\varepsilon/2)/r} \Longrightarrow n \in \pi_{\varepsilon,r}(N). \tag{1.13}$$

In order to simplify the calculation of the density of the ε-balanced numbers we will work with the smaller subsets of $\mathcal{P}_{\varepsilon,r}$, defined by

$$\mathcal{P}^*_{r,\varepsilon}(N) := \Big\{N \le n < 2N,\ \Omega(n) = r,\\ N^{(1-\varepsilon/2)/r} \le P^-(n) \le P^+(n) \le N^{(1+\varepsilon/2)/r}\Big\}. \tag{1.14}$$

The prime number theorem implies with easy calculations that by

$$a_1 := (1-\varepsilon/2)/r, \qquad a_2 := (1+\varepsilon/2)/r,$$
$$I := \big[N^{a_1}, N^{a_2}\big], \qquad J(\mathbf{u}) := (N/u_1 \dots u_{r-1}, 2N/u_1 \dots u_{r-1}]$$

(1.15)

$$\pi^*_{r,\varepsilon}(N) : = \#\big\{n \in \mathcal{P}^*_{r,\varepsilon}(N)\big\} = \sum_{\substack{N \le p_1 \dots p_r < 2N \\ p_i \in I}} 1 \sim$$
$$\sim \int_I \dots \int_I \prod_{i=1}^{r=1} \frac{1}{\log u_i} \int_{I \cap J(\mathbf{u})} \frac{1}{\log t} du_1 \dots du_{r-1}\, dt \sim$$
$$\sim \frac{N}{\log N} \int_{a_1}^{a_2} \dots \int_{a_1}^{a_2} \frac{d\alpha_1 \dots d\alpha_{r-1}}{\alpha_1 \dots \alpha_{r-1}(1-\alpha_1-\dots-\alpha_{r-1})} =: \frac{C_0(r,\varepsilon)N}{\log N}.$$

Here we have obviously for $\varepsilon \to 0$

$$C_0(r,\varepsilon) \le \left(\frac{\varepsilon}{r}\right)^{r-1} \frac{r^r}{(1-\varepsilon/2)^r} = \frac{r\varepsilon^{r-1}}{(1-\varepsilon/2)^r}. \tag{1.16}$$

Since for $\varepsilon < \varepsilon_0$ we have $\mathcal{P}_{r,\varepsilon}(N) \subset \mathcal{P}^*_{r,3\varepsilon}(N)$ the above assertion shows that the number of ε-balanced composite numbers (the counting function of $\mathcal{P}'_\varepsilon \setminus \mathcal{P}$) is negligible compared to that of the primes, since even in total

$$\sum_{r=2}^{\infty} C_0(r,\varepsilon) < 3\varepsilon \quad \text{if} \quad \varepsilon < c_0. \tag{1.17}$$

After this preparation we can formulate our result.

Theorem 1. *Let $r = 2$ or 3, $\varepsilon > 0$. Then the set of ε-balanced numbers with either one or r prime factors contains infinitely many bounded gaps, but has $(1 + O(\varepsilon))\pi(N)$ elements below N.*

2. We will actually prove a stronger result.

Theorem 2. *Let $r = 2$ or 3, $\varepsilon > 0$ and let $\mathcal{H}$ be an arbitrary k-element admissible set of non-negative integers, $k > k_0(\varepsilon)$. Then the k-tuple $n + \mathcal{H}$ contains at least two ε-balanced numbers with either one or r prime factors for infinitely many values of n.*

Proof. Similarly to the role of the Bombieri–Vinogradov theorem (1.3) in the proof of (1.1)–(1.2) we need the analogous assertion for the ε-balanced numbers in $\mathcal{P}^*_{r,\varepsilon}(N)$ defined in (1.14). □

Theorem 3. *We have for any $A > 0$ with $C > C(A)$*

$$\sum_{q \le \sqrt{N}/\log^C N} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{\substack{n \equiv a(q) \\ n \in \mathcal{P}^*_{r,\varepsilon}(N)}} 1 - \frac{C_0(r,\varepsilon) li\, N}{\varphi(q)} \right| \ll_{A,r} \frac{N}{\log^A N}. \tag{2.1}$$

The proof runs analogously to the proof of Vaughan [Vau] of the Bombieri–Vinogradov theorem or one may apply some form of generalized Bombieri–Vinogradov type theorems, as that of Y. Motohashi [Mot] or Pan Cheng Dong [Pan]. The latter asserts that for any $\alpha > 0$, $\varepsilon > 0$ and any $f(m) \ll 1$ we have

(2.2)
$$\sum_{q \le \sqrt{N}/\log^C N} \max_{\substack{a \\ (a,q)=1}} \left| \sum_{m \le N^{1-\alpha}} f(m) \Bigg(\sum_{\substack{mp \le N \\ mp \equiv a (\bmod q)}} 1 - \frac{li \frac{N}{m}}{\varphi(q)} \Bigg) \right| \ll_{\alpha,A} \frac{N}{\log^A N}.$$

The work [GPY1] was based on two main lemmas describing properties of the crucial weight function $\left(\mathcal{H} = \{h_i\}_{i=1}^k\right)$

(2.3)
$$\Lambda_R(n;\mathcal{H},l) = \frac{1}{(k+l)!} \sum_{d \mid P_{\mathcal{H}}(n), d \le R} \mu(d) \log^{k+l} \frac{R}{d}, \quad P_{\mathcal{H}}(n) := \prod_{i=1}^k (n+h_i).$$

The formulation of the main lemmas need the singular series

$$\mathfrak{S}(\mathcal{H}) = \prod \left(1 - \frac{\nu_p(\mathcal{H})}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}, \tag{2.4}$$

where $\nu_p(\mathcal{H})$ denotes the number of residue classes occupied by $\mathcal{H} \bmod p$, for any prime p. The admissible property of $\mathcal{H}$ means $\nu_p(\mathcal{H}) < p$ for any p, or equivalently $\mathfrak{S}(\mathcal{H}) \ne 0$. The two main lemmas below are special cases of Propositions 1 and 2 of [GPY1].

In the following let $\eta > 0$, k, l bounded, but arbitrarily large integers, $n \sim N$ substitutes $n \in [N, 2N)$

$$\max_{h_i \in \mathcal{H}} h_i \ll \log N, \quad R > N^{c_0}, \quad \chi_{\mathcal{P}}(n) = \begin{cases} 1 & \text{if } n \in \mathcal{P}, \\ 0 & \text{if } n \notin \mathcal{P}. \end{cases} \tag{2.5}$$

Lemma 1. *For $R \le \sqrt{N}/(\log N)^C$, $N \to \infty$, we have*

$$\sum_{n \sim N} \Lambda_R(n;\mathcal{H},k+l)^2 = \binom{2l}{l} \frac{N(\log R)^{k+2l}(\mathfrak{S}(\mathcal{H}) + o(1))}{(k+2l)!}. \tag{2.6}$$

Lemma 2. *For $h \in \mathcal{H}$, $R \le N^{1/4}/(\log N)^C$, $C > C(A)$, $N \to \infty$, we have*

(2.7)
$$\sum_{n \sim N} \Lambda_R(n;\mathcal{H},k+l)^2 \chi_{\mathcal{P}}(n+h) = \binom{2l+2}{l+1} \frac{N(\log R)^{k+2l+1}(\mathfrak{S}(\mathcal{H}) + o(1))}{(k+2l+1)! \log N}.$$

In the proof of Lemma 2 actually just two properties of the primes are used:

(i) their distribution in residue classes is on average regular as described by the Bombieri–Vinogradov theorem;

(ii) if $n + h_0 \in \mathcal{P}$, $n \sim N$, then $\mathcal{P}_{\mathcal{H}}(n)$ and $\mathcal{P}_{\mathcal{H}\setminus\{h\}}(n)$ have the same divisors below R, that is, $n + h_0$ has no prime divisor below R.

The first property is shared by the elements of $\mathcal{P}^*_{r,\varepsilon}(N)$ as shown by (2.1), the only change being the factor $C_0(r,\varepsilon)$. In the cases $r = 2$ and $r = 3$ they obviously share property (ii) as well.

In such a way with the notation

$$\mathcal{P}(N) = [N, 2N) \cap \mathcal{P}, \quad \widetilde{\mathcal{P}}_{r,\varepsilon}(N) = \mathcal{P}(N) \cup \mathcal{P}^*_{r,\varepsilon}(N) \tag{2.8}$$

we obtain in exactly the same way as Lemma 2, for the characteristic function $\chi_{\widetilde{\mathcal{P}}}$ of the set $\widetilde{\mathcal{P}}$ the following

Lemma 3. *For* $R \le N^{1/4}/(\log N)^C$, $C > C(A, r, \varepsilon)$, $r = 2$ *or* 3, $N \to \infty$, *we have*

$$\sum_{n \sim N} \Lambda_R(n; \mathcal{H}, l)^2 \chi_{\widetilde{\mathcal{P}}}(n + h_0) = \tag{2.9}$$

$$= \binom{2l+2}{l+1} \frac{N(\log R)^{k+2l+1} \mathfrak{S}(\mathcal{H})(1 + C_0(r, \varepsilon) + o(1))}{(k + 2l + 1)! \log N}.$$

In this case we have, similarly to (3.3) of [GPY1],

(2.10)

$$S := \sum_{n \sim N} \left(\sum_{i=1}^{k} \chi_{\widetilde{\mathcal{P}}}(n + h_i) - 1 \right) \Lambda_R(n; \mathcal{H}, l)^2 \sim$$

$$\sim \binom{2l}{l} \frac{N(\log R)^{k+2l} \mathfrak{S}(\mathcal{H})}{(k+2l)!} \left(\frac{k}{k + 2l + 1} \cdot \frac{2l+1}{2l+2} (1 + C_0(r, \varepsilon)) - 1 \right) > 0$$

if we choose $l = \left\lfloor \sqrt{k}/2 \right\rfloor$, $k > k_0(r, \varepsilon)$, which proves Theorem 2, consequently also Theorem 1 for $r = 2, 3$.

References

[EH] P. D. T. A. Elliott, H. Halberstam, *A conjecture in prime number theory*, Symposia Mathematica **4** INDAM, Rome, 59–72, Academic Press, London, 1968/69.

[GPY1] D. A. Goldston, J. Pintz, C. Y. Yıldırım, Primes in tuples I, *Ann. of Math. (2)* **170** (2009), no. 2, 819–862.

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

[Mot] Y. Motohashi, An induction principle for the generalization of Bombieri's prime number theorem, *Proc. Japan Acad.* **52** (1976), 273–275.

[Pan] Cheng-Dong Pan, A New Mean Value Theorem and Its Applications, *Recent progress in analytic number theory*, Vol. 1 (Durham, 1979), 275–287, Academic Press, London–New York, 1981.

[Pin] J. Pintz, Are there arbitrarily long arithmetic progressions in the sequence of twin primes?, preprint, arXiv:1002.2899

[Vau] R. C. Vaughan, An Elementary Method in Prime Number Theory, *Recent progress in analytic number theory*, Vol. 1 (Durham, 1979), 341–348, Academic Press, London–New York, 1981.

János Pintz
Rényi Mathematical Institute of the Hungarian Academy of Sciences
Budapest
Reáltanoda u. 13–15
H-1053 Hungary
E-mail: pintz@renyi.hu