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

\begin{equation}
\liminf_{n \to \infty} \frac{p_{n+1} - p_n}{\log p_n} = 0,
\end{equation}

and even

\begin{equation}
\liminf_{n \to \infty} \frac{p_{n+1} - p_n}{(\log p_n)^{1/2}(\log \log p_n)^{1/2}} < \infty.
\end{equation}

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,

\begin{equation}
\sum_{q \leq N^{\vartheta}/\log^{C_N}(a,q)} \max_{a \equiv 0 \pmod{q}} \left| \sum_{p \leq N} 1 - \frac{\text{li} N}{\varphi(q)} \right| \ll_A \frac{N}{\log^A N}
\end{equation}

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 mod $p$ 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 $k \geq k_0(\vartheta)$. Consequently we have infinitely many bounded gaps between primes, more precisely

\begin{equation}
\liminf_{n \to \infty} (p_{n+1} - p_n) \leq C(\vartheta).
\end{equation}

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 \( N / \log C \) 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,

\[
(1.5) \quad k_0(0.971) = 6, \quad C(0.971) = 16.
\]

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

\[
(1.6) \quad k_0 \left( \frac{1}{2} + \delta \right) = \left( 2 \left\lfloor \frac{1}{2\delta} \right\rfloor + 1 \right)^2, \quad C \left( \frac{1}{2} + \delta \right) \sim 2\delta^{-2} \log \frac{1}{\delta}.
\]

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

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

which has bounded gaps infinitely often, that is,

\[
(1.8) \quad \limsup_{n \to \infty}(b_{n+1} - b_n) < \infty.
\]

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 \( \varepsilon \)-balanced if for any prime divisors \( p, q \) of \( n \) we have

\[
(1.9) \quad \min(p, q) \geq (\max(p, q))^{1-\varepsilon}.
\]

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

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

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

\[
\pi_{\varepsilon,r}(N) := \# \{ N \leq n < 2N; \ n \in \pi_{\varepsilon,r}(N) \}
\]

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

\[
(1.12) \quad n \in \pi_{\varepsilon,r}(N) \implies N^{(1-\varepsilon)/r} \leq P^-(n) \leq P^+(n) \leq (2N)^{1/(r(1-\varepsilon))}.
\]

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

\[
(1.13) \quad N^{(1-\varepsilon/2)/r} \leq P^-(n) \leq P^+(n) \leq N^{(1+\varepsilon/2)/r} \implies n \in \pi_{\varepsilon,r}(N).
\]

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

\[
\mathcal{P}_{r,\varepsilon}^*(N) := \left\{ N \leq n < 2N, \ \Omega(n) = r, \quad N^{(1-\varepsilon/2)/r} \leq P^-(n) \leq P^+(n) \leq N^{(1+\varepsilon/2)/r} \right\}.
\]

The prime number theorem implies with easy calculations that by

\[
\begin{align*}
&\quad a_1 := (1 - \varepsilon/2)/r, \quad a_2 := (1 + \varepsilon/2)/r, \\
&\quad I := [N^{a_1}, N^{a_2}], \quad J(u) := (N/u_1 \ldots u_{r-1}, 2N/u_1 \ldots u_{r-1}]
\end{align*}
\]

\[
(1.15) \quad \pi_{r,\varepsilon}^*(N) := \# \{ n \in \mathcal{P}_{r,\varepsilon}^*(N) \} = \sum_{N \leq p_1 \ldots p_r < 2N} 1 \sim \\
\quad \sim \int_I \ldots \int_I \prod_{i=1}^{r-1} \frac{1}{\log u_i} \int_{I \cap J(u)} \frac{1}{\log t} du_1 \ldots du_{r-1} dt \sim \\
\quad \sim \frac{N}{\log N} \int_{a_1}^{a_2} \ldots \int_{a_1}^{a_2} \frac{d\alpha_1 \ldots d\alpha_{r-1}}{\alpha_1 \ldots \alpha_{r-1}(1 - \alpha_1 - \cdots - \alpha_{r-1})} =: C_0(r, \varepsilon)N/\log N.
\]

Here we have obviously for \(\varepsilon \to 0\)

\[
(1.16) \quad C_0(r, \varepsilon) \leq \left(\frac{\varepsilon}{r}\right)^{r-1} \frac{r^r}{(1-\varepsilon/2)^r} = \frac{r^{r-1}}{(1-\varepsilon/2)^r}.
\]
Since for $\epsilon < \epsilon_0$ we have $\mathcal{P}_{r,\epsilon}(N) \subset \mathcal{P}^*_{r,3\epsilon}(N)$ the above assertion shows that the number of $\epsilon$-balanced composite numbers (the counting function of $\mathcal{P}^*_\epsilon \setminus \mathcal{P}$) is negligible compared to that of the primes, since even in total

\begin{equation}
\sum_{r=2}^{\infty} C_0(r,\epsilon) < 3\epsilon \quad \text{if} \quad \epsilon < \epsilon_0.
\end{equation}

After this preparation we can formulate our result.

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

2. We will actually prove a stronger result.

**Theorem 2.** Let $r = 2$ or $3$, $\epsilon > 0$ and let $\mathcal{H}$ be an arbitrary $k$-element admissible set of non-negative integers, $k > k_0(\epsilon)$. Then the $k$-tuple $n + \mathcal{H}$ contains at least two $\epsilon$-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 in the proof of (1.1)–(1.2) we need the analogous assertion for the $\epsilon$-balanced numbers in $\mathcal{P}^*_{r,\epsilon}(N)$ defined in (1.14).

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

\begin{equation}
\sum_{q \leq \sqrt{N}} \max_{d \mid q} \frac{\max_{a \mid q} \sum_{n \equiv a(q)}^{\phi(q)} 1 - C_0(r,\epsilon)\text{li}\ N}{\phi(q)} \ll A,r \frac{N}{\log^4 N}.
\end{equation}

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 $\alpha > 0$, $\epsilon > 0$ and any $f(m) \ll 1$ we have

\begin{equation}
\sum_{q \leq \sqrt{N}} \max_{d \mid q} \frac{\sum_{m \leq N^{1-\alpha}} f(m) \left( \sum_{m p \leq N \mid m \equiv a(q) \mod q} 1 - \frac{\text{li} N}{\phi(q)} \right)}{\phi(q)} \ll A,r \frac{N}{\log^4 N}.
\end{equation}
The work [GPY1] was based on two main lemmas describing properties of the crucial weight function \( \mathcal{H} = \{ h_i \}_{i=1}^k \) of the form
\[
\Lambda_R(n; \mathcal{H}, l) = \frac{1}{(k+l)!} \sum_{d \mid P_H(n), d \leq R} \mu(d) \log^{k+l} \frac{R}{d}, \quad P_H(n) := \prod_{i=1}^k (n + h_i).
\]
The formulation of the main lemmas need the singular series
\[
S(\mathcal{H}) = \prod \left( 1 - \frac{\nu_p(\mathcal{H})}{p} \right) \left( 1 - \frac{1}{p} \right)^{-k},
\]
where \( \nu_p(\mathcal{H}) \) denotes the number of residue classes occupied by \( \mathcal{H} \mod p \), for any prime \( p \). The admissible property of \( \mathcal{H} \) means \( \nu_p(\mathcal{H}) < p \) for any \( p \), or equivalently \( S(\mathcal{H}) \neq 0 \). 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 \( n \in [N, 2N) \)
\[
\max_{h_i \in \mathcal{H}} h_i \ll \log N, \quad R > N^c_0, \quad \chi_P(n) = \begin{cases} 1 & \text{if } n \in \mathcal{P}, \\ 0 & \text{if } n \notin \mathcal{P}. \end{cases}
\]

**Lemma 1.** For \( R \leq \sqrt{N}/(\log N)^C \), \( N \to \infty \), we have
\[
\sum_{n \sim N} \Lambda_R(n; \mathcal{H}, k + l)^2 = \left( \frac{2l}{l} \right) N(\log R)^{k+2l} \left( S(\mathcal{H}) + o(1) \right) \frac{(k+l)!}{(k+2l)!}.
\]

**Lemma 2.** For \( h \in \mathcal{H}, R \leq N^{1/4}/(\log N)^C, C > C(A), N \to \infty \), we have
\[
\sum_{n \sim N} \Lambda_R(n; \mathcal{H}, k + l)^2 \chi_P(n + h) = \left( \frac{2l + 2}{l + 1} \right) N(\log R)^{k+2l+1} \left( S(\mathcal{H}) + o(1) \right) \frac{(k+2l+1)! \log N}{(k+2l+1)!}.
\]

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 \( \mathcal{P}_H(n) \) and \( \mathcal{P}_{\mathcal{H}\backslash\{h\}}(n) \) 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
\[(2.8) \quad \mathcal{P}(N) = [N, 2N) \cap \mathcal{P}, \quad \mathcal{P}_{r, \varepsilon}(N) = \mathcal{P}(N) \cup \mathcal{P}^*_r(N)\]
we obtain in exactly the same way as Lemma 2 for the characteristic function \(\chi_{\mathcal{P}}\) of the set \(\mathcal{P}\) the following

**Lemma 3.** For \(R \leq N^{1/4}/(\log N)^C, \quad C > C(A, r, \varepsilon), \quad r = 2 \text{ or } 3, \quad N \to \infty\), we have
\[(2.9) \quad \sum_{n \sim N} \Lambda_R(n; \mathcal{H}, l)^2 \chi_{\mathcal{P}}(n + h_0) = \]
\[= \binom{2l + 2}{l + 1} N(\log R)^{k + 2l + 1} \mathcal{G}(\mathcal{H})(1 + C_0(r, \varepsilon) + o(1)) (k + 2l + 1)! \log N.\]

In this case we have, similarly to (3.3) of [GPY1],
\[(2.10) \quad S = \sum_{n \sim N} \left(\sum_{i=1}^{k} \chi_{\mathcal{P}}(n + h_i) - 1\right) \Lambda_R(n; \mathcal{H}, l)^2 \sim \]
\[\sim \binom{2l}{l} N(\log R)^{k + 2l} \mathcal{G}(\mathcal{H}) \left(\frac{k}{k + 2l + 1} \cdot \frac{2l + 1}{2l + 2} (1 + C_0(r, \varepsilon)) - 1\right) > 0\]
if we choose \(l = \left\lfloor \frac{\sqrt{k}}{2} \right\rfloor, \quad k > k_0(r, \varepsilon)\), 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