Large gaps between consecutive prime numbers containing perfect powers

Kevin Ford, D. R. Heath-Brown and Sergei Konyagin

Abstract For any positive integer k, we show that infinitely often, perfect k-th powers appear inside very long gaps between consecutive prime numbers, that is, gaps of size

$$c_k \frac{\log p \log_2 p \log_4 p}{(\log_3 p)^2},$$

where p is the smaller of the two primes.

1 Introduction

In 1938, Rankin [11] proved that the maximal gap, $G(x)$, between primes $\leqslant x$, satisfies[1]

$$G(x) \geqslant (c+o(1)) \frac{\log x \log_2 x \log_4 x}{(\log_3 x)^2}, \tag{1}$$

with $c = \frac{1}{3}$. The following six decades witnessed several improvements of the constant c; we highlight out only a few of these. First, Rankin's own improvement [12] $c = e^\gamma$ in 1963 represented the limit of what could be achieved by inserting into Rankin's original 1938 argument best possible bounds on counts of "smooth" numbers. This record stood for a long time until Maier and Pomerance [8] introduced new ideas to improve the constant to $c = 1.31256e^\gamma$ in 1989; these were refined by Pintz [10], who obtained $c = 2e^\gamma$ in 1997. Very recently, the first and third authors together with Green and Tao [3] have shown that c can be taken arbitrarily large. Independently, this was also proven by Maynard [9].

Ford
Department of Mathematics, 1409 West Green Street, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA, e-mail: ford@math.uiuc.edu

Heath-Brown
Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG, UK e-mail: rhb@maths.ox.ac.uk

Konyagin
Steklov Mathematical Institute, 8 Gubkin Street, Moscow, 119991, Russia e-mail: konyagin@mi.ras.ru

[1] As usual in the subject, $\log_2 x = \log\log x$, $\log_3 x = \log\log\log x$, and so on.

Rankin's lower bound (1) is probably very far from the truth. Based on a probabilistic model of primes, Cramér [1] conjectured that

$$\limsup_{X\to\infty} \frac{G(X)}{\log^2 X} = 1,$$

and Granville [2], using a refinement of Cramér's model, has conjectured that the lim sup above is in fact at least $2e^{-\gamma} = 1.1229\ldots$. Cramér's model also predicts that the normalized prime gaps $\frac{p_{n+1}-p_n}{\log p_n}$ should have exponential distribution, that is, $p_{n+1} - p_n \geqslant C \log p_n$ for about $e^{-C}\pi(X)$ primes $\leqslant X$.

Our aim in this paper is to study whether or not long prime gaps, say of the size of the right hand size of the inequality in (1), occur when we impose that an integer of a specified type lies inside the interval. To be precise, we say that a number m is "prime avoiding with constant c" if $m+u$ is composite for all integers u satisfying

$$|u| \leqslant c\frac{\log m \log_2 m \log_4 m}{(\log_3 m)^2}.$$

Here we will be concerned with prime avoiding perfect powers.

Theorem 1. *For any positive integer k, there are a constant $c = c(k) > 0$ and infinitely many perfect k-th powers which are prime-avoiding with constant c.*

It seems possible that the methods of Green, Ford, Konyagin and Tao, or of Maynard, might be adapted to handle slightly longer intervals containing a k-th power but no primes. However we leave this possibility aside for the time being.

2 Sieve estimates

Throughout, constants implied by the Landau O-symbol and Vinogradov $\ll$-symbol are absolute unless otherwise indicated, e.g. by a subscript such as $\ll_u$. The symbols p and q will always denote prime numbers. Denote by $P^+(n)$ the largest prime factor of a positive integer n, and by $P^-(n)$ the smallest prime factor of n.

We need several standard lemmas from sieve theory, the distribution of "smooth" numbers, and the distribution of primes in arithmetic progressions.

Lemma 2.1. *For large x and $z \leqslant x^{\log_3 x/(10\log_2 x)}$, we have*

$$\#\{n \leqslant x : P^+(n) \leqslant z\} \ll \frac{x}{\log^5 x}.$$

Proof. This follows from standard counts of smooth numbers. Lemma 1 of Rankin [11] also suffices. □

Lemma 2.2. *Let $\mathscr{R}$ denote any set of primes and let $a \in \{-1,1\}$. Then*

$$\#\{p \leqslant x : p \not\equiv a \pmod r\ (\forall r \in \mathscr{R})\} \ll \frac{x}{\log x} \prod_{\substack{p\in\mathscr{R}\\ p\leqslant x}} \left(1 - \frac{1}{p}\right).$$

Proof. Standard sieve methods [4]. □

Finally, we require a bound of "large sieve" type for averages of quadratic character sums.

Lemma 2.3. *For any set $\mathscr{P}$ of primes in $[2,x]$, and for any $\varepsilon > 0$,*

$$\sum_{\substack{m\leqslant x\\ m\ odd}}\mu^2(m)\Big|\sum_{p\in\mathscr{P}}\left(\frac{p}{m}\right)\Big|^2\ll_\varepsilon x^{2+\varepsilon}.$$

Proof. This follows immediately from Theorem 1 of [5]. □

3 k-th power residues and prime ideals

One of our principle tools is the following estimate for an average of counts of solutions of a certain k-th power congruence.

Lemma 3.1. *Let k be a positive integer. For any non-zero integer u and any prime p write*

$$\rho_{u,k}(p)=\rho_u(p)=\#\{n\pmod p:\ n^k+u\equiv 0\pmod p\}.$$

Then for any fixed $\varepsilon>0$ and $x\geqslant 2$ we have

$$\prod_{x<p\leqslant y}\left(1-\frac{\rho_u(p)}{p}\right)\ll_{k,\varepsilon}|u|^\varepsilon\frac{\log x}{\log y}.$$

The proof of this is based on the Prime Ideal Theorem, and we begin by giving a formal statement of an appropriate form of the latter.

Lemma 3.2. *There is an effectively computable absolute constant $c>0$ with the following property. Let K be an algebraic number field of degree n_K, and write d_K for the absolute value of the discriminant of K. Let β_0 be the largest simple real zero of $\zeta_K(s)$ in the interval $[\frac{1}{2},1]$ if any such exists. Then*

$$|\pi_K(x)-\mathrm{Li}(x)|\leqslant\mathrm{Li}(x^{\beta_0})+O\left(x\exp\{-cn_K^{-1/2}\log^{1/2}x\}\right)$$

for $x\geqslant\exp\{10n_K\log^2 d_K\}$, where as usual, $\pi_K(x)$ denotes the number of prime ideals of K with norm at most x, and $\mathrm{Li}(x)=\int_2^x dt/\log t$. We omit the first summand on the right hand side if β_0 does not exist.

This follows from Theorem 1.3 of Lagarias and Odlyzko [7], on choosing $L=K$ in their notation. The reader should note that the counting function $\pi_C(x,L/K)$ of [7] excludes ramified primes, but the number of these is $O(n_K\log d_K)$, which is majorized by $x\exp\{-cn_K^{-1/2}\log^{1/2}x\}$.

In order to handle the term involving the possible simple real zero β_0 we use the following result of Heilbronn [6, Theorem 1].

Lemma 3.3. *A simple real zero of $\zeta_K(s)$ must be a zero of $\zeta_k(s)$ for some quadratic subfield k of K.*

It follows that β_0 is a zero for some quadratic Dirichlet L-function $L(s,\chi)$, with a character χ of conductor dividing d_K. Thus Siegel's Theorem shows that $1-\beta_0\geqslant c(\varepsilon)d_K^{-\varepsilon}$, for any fixed $\varepsilon>0$, with an ineffective constant $c(\varepsilon)>0$. We then deduce that

$$\mathrm{Li}(x^{\beta_0})\ll x^{\beta_0}\leqslant x\exp\{-c(\varepsilon)d_K^{-\varepsilon}\log x\}\leqslant x\exp\{-c(\varepsilon)n_K^{-1/2}\log^{1/2}x\}$$

if $n_K\log x\geqslant d_K^{2\varepsilon}$. We therefore obtain the following version of the Prime Ideal Theorem.

Lemma 3.4. *For any $\eta>0$ there is an ineffective constant $C(\eta)>0$ with the following property. Let K be an algebraic number field of degree n_K, and write d_K for the absolute value of the discriminant of K. Then*

$$\pi_K(x) = \mathrm{Li}(x) + O(x\exp\{-C(\eta)n_K^{-1/2}\log^{1/2}x\})$$

for

$$x \geqslant \exp\left\{\max\left(10n_K\log^2 d_K\,,\, n_K^{-1}d_K^{\eta}\right)\right\}.$$

Proof (Proof of Lemma 3.1.). Since it may happen that $-u$ is a perfect power we begin by taking a to be the largest divisor of k for which $-u$ is a perfect a-th power. Then if $-u = v^a$ and $k = ab$ we see firstly that the polynomial $X^b - v$ is irreducible over the rationals, and secondly that $n^b \equiv v \pmod p$ implies $n^k + u \equiv 0 \pmod p$, whence

$$\rho_{u,k}(p) \geqslant \rho_{-v,b}(p). \tag{1}$$

We will apply Lemma 3.4 to the field $K = \mathbb{Q}(\theta)$, where θ is a root of $X^b - v$. Thus K has degree $b \leqslant k$. Moreover its discriminant will be a divisor of

$$D := \mathrm{Disc}(1,\theta,\theta^2,\dots,\theta^{b-1}) = (-1)^{b-1}b^b v^{b-1}.$$

We now set

$$x_0 = C(k,\eta)\exp(|D|^{\eta}).$$

If we choose the constant $C(k,\eta)$ sufficiently large then whenever $x \geqslant x_0$ we will have

$$x\exp\{-C(\eta)b^{-1/2}\log^{1/2}x\} \leqslant x\log^{-2}x$$

and

$$x \geqslant \exp\left\{\max\left(10b(\log|D|)^2\,,\, b^{-1}|D|^{\eta}\right)\right\}.$$

It therefore follows from Lemma 3.4 that

$$\pi_K(x) = \mathrm{Li}(x) + O_{k,\eta}\left(x\log^{-2}x\right)$$

for $x \geqslant x_0$.

We now write $\nu_K(p)$ for the number of first degree prime ideals of K lying above p. Then

$$\pi_K(x) = \sum_{p\leqslant x}\nu_K(p) + O_k(\sum_{p^e\leqslant x,\, e\geqslant 2} 1).$$

Moreover, by Dedekind's Theorem we will have $\rho_{-v,b}(p) = \nu_K(p)$ whenever $p \nmid D$. In the remaining case in which $p \mid D$ we have $\rho_{-v,b}(p) \leqslant b \leqslant k$ and $\nu_K(p) \leqslant b \leqslant k$. It therefore follows that

$$\pi_K(x) = \sum_{p\leqslant x}\rho_{-v,b}(p) + O_k(x^{1/2}) + O_k(\log|D|),$$

so that

$$\sum_{p\leqslant x}\rho_{-v,b}(p) = \mathrm{Li}(x) + O_{k,\eta}(x\log^{-2}x)$$

when $x \geqslant x_0$.

We now observe that

$$\prod_{x<p\leqslant y}\left(1-\frac{\rho_u(p)}{p}\right)\leqslant \exp\left\{-\sum_{x<p\leqslant y}\frac{\rho_u(p)}{p}\right\}$$
$$\leqslant \exp\left\{-\sum_{x<p\leqslant y}\frac{\rho_{-v,b}(p)}{p}\right\}$$

by (1). Assuming that $y\geqslant x_0$ we may then use summation by parts to calculate that

$$\begin{aligned}\sum_{x<p\leqslant y}\frac{\rho_{-v,b}(p)}{p} &\geqslant \sum_{\max(x,x_0)<p\leqslant y}\frac{\rho_{-v,b}(p)}{p}\\ &= \log\log y-\log\log\big(\max(x,x_0)\big)+O_{k,\eta}(1)\\ &\geqslant \log\log y-\log\log x-\log\log x_0+O_{k,\eta}(1)\\ &= \log\log y-\log\log x-\eta\log|D|+O_{k,\eta}(1)\\ &\geqslant \log\log y-\log\log x-\eta k\log|u|+O_{k,\eta}(1).\end{aligned}$$

We therefore have

$$\prod_{x<p\leqslant y}\left(1-\frac{\rho_u(p)}{p}\right)\ll_{k,\eta}|u|^{k\eta}\frac{\log x}{\log y}$$

when $y\geqslant x_0$. Of course this estimate is trivial when $y\leqslant x_0$ since one then has $\log y\ll_{k,\eta}|D|^{\eta}\ll_{k,\eta}|u|^{k\eta}$. The lemma then follows.

4 Main argument

Fix a positive integer k. Let x be a large number, sufficiently large depending on k, let c_1 and c_2 be two positive constants depending on k to be chosen later, and put

$$N=\prod_{p\leqslant x}p,\qquad z=x^{c_1\log_3 x/\log_2 x},\qquad y=\frac{c_2x\log x\log_3 x}{(\log_2 x)^2}.$$

In the rest of the paper we will prove the following lemma.

Lemma 4.1. *There is a number $m\leqslant 2N$ such that m^k+u is composite for $|u|\leqslant y$.*

Theorem 1 will follow upon observing that $m^k\leqslant e^{kx+o(x)}$ as $x\to\infty$ and consequently that

$$y\gg_k\frac{\log(m^k)\log_2(m^k)\log_4(m^k)}{(\log_3(m^k))^2}.$$

We will select m be choosing residue classes for m modulo p for primes $p\leqslant x$. Let

$$\mathscr{P}_1=\{p:p\leqslant\log x\text{ or }z<p\leqslant x/4\},\qquad \mathscr{P}_2=\{p:\log x<p\leqslant z\}.$$

We first choose

$$\begin{aligned}m&\equiv 0\pmod p\quad(p\in\mathscr{P}_1),\\ m&\equiv 1\pmod p\quad(p\in\mathscr{P}_2).\end{aligned}\tag{1}$$

Observe that $p|(m^k+u)$ if $p|u$ for some $p\in\mathscr{P}_1$. Because $y<(x/4)\log x$, any remaining value of u is thus either composed only of primes in $\mathscr{P}_2$ (in particular, u is z-smooth), including $|u|=1$, or $|u|$ is a prime larger than $x/4$. For any u in the latter category such that $p|(u+1)$ for some $p\in\mathscr{P}_2$, $p|(m^k+u)$. Let U denote the set of exceptional values of u, that is, the set of $u\in[-y,y]$ not divisible by any prime in $\mathscr{P}_1$, and such that if $|u|$ is prime then $p\nmid(u+1)$ for all $p\in\mathscr{P}_2$. By Lemmas 2.1 and 2.2, if c_1 is sufficiently small, then

$$|U|\ll\frac{y}{\log^5 x}+\frac{y}{\log x}\prod_{p\in\mathscr{P}_2}\left(1-\frac{1}{p}\right)\ll\frac{y\log_2 x}{\log x\log z}=\frac{c_2}{c_1}\frac{x}{\log x}.$$

Choosing c_2 appropriately, we can ensure that $|U|\leqslant\delta x/\log x$, where $\delta>0$ depends on k (δ will be chosen later).

The remaining steps depend on whether k is odd or even. If k is odd, the construction is very easy. For each $u\in U$, associate with u a different prime $p_u\in(x/4,x]$ such that $(p_u-1,k)=1$ (e.g., one can take $p_u\equiv 2 \pmod k$ if $k\geqslant 3$). Then every residue modulo p_u is a k-th power residue, and we take m in the residue class modulo p_u such that

$$m^k\equiv -u \pmod{p_u}\quad (u\in U). \tag{2}$$

By the prime number theorem for arithmetic progressions, the number of available primes is at least

$$x/(2\phi(k)\log x)\geqslant|U|$$

if δ is small enough. With this construction, $p_u|(m^k+u)$ for every $u\in U$. Therefore, m^k+u is divisible by a prime $\leqslant x$ for every $|u|\leqslant y$. Furthermore, (1) and (2) together imply that m is defined modulo a number N', where $N'|N$. Therefore, there is an admissible value of m satisfying $N<m\leqslant 2N$. The prime number theorem implies that $N=e^{x+o(x)}$, thus $m^k-y>x$. Consequently, m^k+u is composite for $|u|\leqslant y$.

Now suppose that k is even. There do not exist primes for which every residue modulo p is a k-th power residue. However, we maximize the density of k-th power residues by choosing primes p such that $(p-1,k)=2$, e.g. taking $p\equiv 3 \pmod{2k}$. For such primes p, every quadratic residue is a k-th power residue. Let

$$\mathscr{P}_3=\{x/4<p\leqslant x/2 : p\equiv 3 \pmod{2k}\}.$$

By the prime number theorem for arithmetic progressions, $|\mathscr{P}_3|\geqslant x/(5\phi(2k)\log x)$. We aim to associate numbers $u\in U$ with distinct primes $p_u\in\mathscr{P}_3$ such that $\left(\frac{-u}{p_u}\right)=1$. This ensures that the congruence $m^k+u\equiv 0 \pmod{p_u}$ has a solution. We, however, may not be able to find such p for every $u\in U$, but can find appropriate primes for most u. Let

$$U'=\left\{u\in U:\left(\frac{-u}{p}\right)=1 \text{ for at most } \frac{\delta x}{\log x} \text{ primes } p\in\mathscr{P}_3\right\}.$$

The numbers $u\in U\backslash U'$ may be paired with different primes $p_u\in\mathscr{P}_3$ such that $\left(\frac{-u}{p_u}\right)=1$. We then may take m such that

$$m^k\equiv -u \pmod{p_u}\quad (u\in U\backslash U'). \tag{3}$$

Next we will show that $|U'|$ is small. Write

$$S=\sum_{u\in U}\Big|\sum_{p\in\mathscr{P}_3}\left(\frac{-u}{p}\right)\Big|^2.$$

Each u may be written uniquely in the form $u=su_1^2u_2$, where $s=\pm1$, $u_2>0$ and u_2 is squarefree. By quadratic reciprocity,

$$\left(\frac{-u}{p}\right) = (-s)\left(\frac{u_2}{p}\right) = (-s)(-1)^{\frac{u_2-1}{2}}\left(\frac{p}{u_2}\right),$$

since $p \equiv 3 \pmod 4$. Given u_2, there are at most $\sqrt{y/u_2} \leqslant \sqrt{y}$ choices for u_1. Hence, using Lemma 2.3,

$$\begin{aligned} S &= \sum_{u\in U}\Big|\sum_{p\in\mathscr{P}_3}\left(\frac{p}{u_2}\right)\Big|^2 \\ &\leqslant \sum_{u_2\leqslant y} 2y^{1/2}\Big|\sum_{p\in\mathscr{P}_3}\left(\frac{p}{u_2}\right)\Big|^2 \\ &\ll_\varepsilon x^{5/2+\varepsilon}. \end{aligned}$$

Now let $\delta = \frac{1}{15\phi(2k)}$, so that $\delta x/\log x \leqslant \frac{1}{3}|\mathscr{P}_3|$. If $u \in U'$, then clearly

$$\Big|\sum_{p\in\mathscr{P}_3}\left(\frac{-u}{p}\right)\Big| \geqslant \frac{1}{3}|\mathscr{P}_3| \geqslant \delta\frac{x}{\log x}.$$

It follows that $|S| \gg |U'|(x/\log x)^2$, and consequently that

$$|U'| \ll_\varepsilon x^{1/2+2\varepsilon}. \tag{4}$$

Let $A \mod M$ denote the set of numbers m satisfying the congruence conditions (1) and (3), where $0 \leqslant A < M$. Thus, if $m \equiv A \pmod M$ and $u \notin U'$, then $m^k + u$ is divisible by a prime $\leqslant x/2$. Let

$$K = \prod_{x/2<p\leqslant x} p.$$

We'll take $m = Mj + A$, where $1 \leqslant j \leqslant K$, and aim to show that there exists a value of j so that $(mj+A)^k + u$ is composite for every $u \in U'$. By sieve methods (see [4]),

$$\begin{aligned} \sum_{j=1}^{K} \#\{u \in U' : (Mj+A)^k + u \text{ prime}\} &= \sum_{u\in U'} \#\{1 \leqslant j \leqslant K : (Mj+A)^k + u \text{ prime}\} \\ &\ll \sum_{u\in U'} K \prod_{y<q\leqslant\sqrt{K}}\left(1 - \frac{\rho_u(q)}{q}\right). \end{aligned}$$

By Lemma 3.1, the above product is

$$\ll_{k,\varepsilon} u^{\varepsilon/2}\frac{\log y}{\log K} \ll_{k,\varepsilon} u^{\varepsilon/2}\frac{\log x}{x}.$$

Combined with our estimate (4) for the size of $|U'|$, we find that

$$\sum_{1\leqslant j\leqslant K} \#\{u \in U' : (Mj+A)^k + u \text{ prime}\} \ll_{k,\varepsilon} \frac{K}{x^{1/2-4\varepsilon}}.$$

It follows that the left hand side above is zero for some j. That is, $(Mj+A)^k + u$ is composite for every $u \in U'$. Therefore, $(Mj+A)^k + u$ is composite for every u satisfying $|u| \leqslant y$. Finally, we note that $Mj + A \leqslant 2N$, and the proof of Lemma 4.1 is complete.

Remark 1. For odd k the constant $c(k)$ in Theorem 1 is effective. For even k it is ineffective due to the use of Siegel's theorem in the proof of Lemma 3.1.

Acknowledgements Research of the third author was partially performed while he was visiting the University of Illinois at Urbana–Champaign. Research of the first author was partially performed while visiting the University of Oxford. Research of the first and third authors was also carried out in part at the University of Chicago, and they are thankful to Prof. Wilhelm Schlag for hosting these visits.

The first author was supported by NSF grant DMS-1201442. The research of the second author was supported by EPSRC grant EP/K021132X/1. The research of the the third author was partially supported by Russian Foundation for Basic Research, Grant 14-01-00332, and by Program Supporting Leading Scientific Schools, Grant Nsh-3082.2014.1.

References

1. H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers,* Acta Arith. **2** (1936), 396–403.
2. A. Granville, *Harald Cramér and the distribution of prime numbers*, Scandanavian Actuarial J. **1** (1995), 12–28.
3. K. Ford, B. Green, S. Konyagin and T. Tao, *Large gaps between consecutive prime numbers*, `arXiv:1408.4505`.
4. H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, London, 1974.
5. D. R. Heath-Brown, *A mean value estimate for real character sums*, Acta Arith. **72** (1995), 235–275.
6. H. Heilbronn, *On real zeros of Dedekind ζ-functions,* Canad. J. Math. **25** (1973), 870–873.
7. J.C. Lagarias and A.M. Odlyzko, *Effective versions of the Chebotarev density theorem,* Algebraic number fields, (Proc. Sympos., Univ. Durham, Durham, 1975), 409–464. (Academic Press, London, 1977).
8. H. Maier and C. Pomerance, *Unusually large gaps between consecutive primes*. Trans. Amer. Math. Soc. **322** (1990), no. 1, 201–237.
9. J. Maynard, *Large gaps between primes*, `arXiv:1408.5110`.
10. J. Pintz, *Very large gaps between consecutive primes.* J. Number Theory **63** (1997), no. 2, 286–301.
11. R. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.
12. R. A. Rankin, *The difference between consecutive prime numbers. V,* Proc. Edinburgh Math. Soc. (2) 13 (1962/63), 331–332.