Prime avoiding numbers

Mikhail Gabdullin (joint work with Artyom Radomskii)

Steklov Mathematical Institute, Moscow

Moscow, 11th of November 2022

Let p_n be the n^{th} prime and

$$G(X) = \max_{p_{n+1} \leqslant X} (p_{n+1} - p_n)$$

denote the largest gap between consecutive primes up to X.

The Prime Number Theorem together with a simple averaging argument implies that $G(X) \geqslant (1 + o(1)) \log X$.

The expected size of $G(X)$ is of order $(\log X)^2$: Cramér made this conjecture based on a probabilistic model of primes.

The best known upper bound for $G(X)$ is only

$$G(X) \ll X^{0.525}$$

due to Baker, Harman, and Pintz (2001).

Let p_n be the n^{th} prime and

$$G(X) = \max_{p_{n+1} \leqslant X} (p_{n+1} - p_n)$$

denote the largest gap between consecutive primes up to X.

The Prime Number Theorem together with a simple averaging argument implies that $G(X) \geqslant (1 + o(1)) \log X$.

The expected size of $G(X)$ is of order $(\log X)^2$: Cramér made this conjecture based on a probabilistic model of primes.

The best known upper bound for $G(X)$ is only

$$G(X) \ll X^{0.525}$$

due to Baker, Harman, and Pintz (2001).

Let p_n be the n^{th} prime and

$$G(X) = \max_{p_{n+1} \leqslant X} (p_{n+1} - p_n)$$

denote the largest gap between consecutive primes up to X.

The Prime Number Theorem together with a simple averaging argument implies that $G(X) \geqslant (1 + o(1)) \log X$.

The expected size of $G(X)$ is of order $(\log X)^2$: Cramér made this conjecture based on a probabilistic model of primes.

The best known upper bound for $G(X)$ is only

$$G(X) \ll X^{0.525}$$

due to Baker, Harman, and Pintz (2001).

Let p_n be the n^{th} prime and

$$G(X) = \max_{p_{n+1} \leqslant X} (p_{n+1} - p_n)$$

denote the largest gap between consecutive primes up to X.

The Prime Number Theorem together with a simple averaging argument implies that $G(X) \geqslant (1 + o(1)) \log X$.

The expected size of $G(X)$ is of order $(\log X)^2$: Cramér made this conjecture based on a probabilistic model of primes.

The best known upper bound for $G(X)$ is only

$$G(X) \ll X^{0.525}$$

due to Baker, Harman, and Pintz (2001).

Prime gaps: lower bounds

Rankin in 1938 was the first to prove the bound of the type

$$G(X) \geqslant (c+o(1))\frac{\log X \log\log X \log\log\log\log X}{(\log\log\log X)^2},$$

improving the previous results of Westzynthius and Erdős. Rankin proved the mentioned bound with $c = 1/3$, and for about next 80 years this constant was increased many times, the last being $c = 2e^{\gamma}$ due to Pintz (1997).

In 2016, Ford, Green, Konyagin, Tao and independently Maynard showed by different approaches that c can be taken arbitrarily large, giving the affirmative answer for a long-standing conjecture of Erdős. In 2018 all these five authors together, combining their ideas, made a further breakthrough establishing that

$$G(X) \gg \frac{\log X \log\log X \log\log\log\log X}{\log\log\log X}.$$

Rankin in 1938 was the first to prove the bound of the type

$$G(X) \geqslant (c+o(1)) \frac{\log X \log\log X \log\log\log\log X}{(\log\log\log X)^2},$$

improving the previous results of Westzynthius and Erdős. Rankin proved the mentioned bound with $c = 1/3$, and for about next 80 years this constant was increased many times, the last being $c = 2e^{\gamma}$ due to Pintz (1997).

In 2016, Ford, Green, Konyagin, Tao and independently Maynard showed by different approaches that c can be taken arbitrarily large, giving the affirmative answer for a long-standing conjecture of Erdős. In 2018 all these five authors together, combining their ideas, made a further breakthrough establishing that

$$G(X) \gg \frac{\log X \log\log X \log\log\log\log X}{\log\log\log X}.$$

In 2015, Ford, Heath-Brown, and Konyagin introduced the notion of prime avoidance. For a positive integer n, let $F(n)$ denote the distance from n to the nearest prime number (clearly, the maximum value of $F(n)$ taken over all $n \leqslant X$ has the same order as $G(X)$). They called n a "prime avoiding number with constant c", if

$$F(n) \geqslant c \frac{\log n \log\log n \log\log\log\log n}{(\log\log\log n)^2},$$

and proved that 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.

Using the "new" method from the mentioned work of five authors, Maier and Rassias recently extended this result in the following way: there exists $c(k)$ and infinitely many numbers n of the form $n = p^k$, p is a prime, with

$$F(n) \geqslant c(k) \frac{\log n \log\log n \log\log\log\log n}{\log\log\log n}.$$

In 2015, Ford, Heath-Brown, and Konyagin introduced the notion of prime avoidance. For a positive integer n, let $F(n)$ denote the distance from n to the nearest prime number (clearly, the maximum value of $F(n)$ taken over all $n \leqslant X$ has the same order as $G(X)$). They called n a "prime avoiding number with constant c", if

$$F(n) \geqslant c\frac{\log n \log\log n \log\log\log\log n}{(\log\log\log n)^2},$$

and proved that 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.

Using the "new" method from the mentioned work of five authors, Maier and Rassias recently extended this result in the following way: there exists $c(k)$ and infinitely many numbers n of the form $n = p^k$, p is a prime, with

$$F(n) \geqslant c(k)\frac{\log n \log\log n \log\log\log\log n}{\log\log\log n}.$$

We consider the following additive problem related to prime avoidance: can we prove that any large positive integer N can be represented as

$$N = n_1 + n_2, \qquad \text{where both } F(n_1) \text{ and } F(n_2) \text{ are large?}$$

To do so, one may try to use the "usual" approach for getting a long string of composite numbers, which originated from the work of Westzynthius. In that approach, one defines a "smooth" number $m \leqslant X$,

$$m \equiv 0 \pmod{p} \quad \text{for all } p \leqslant \log\log X \text{ and } (\log X)^{o(1)} < p \leqslant 0.5 \log X, \tag{0.1}$$

and chooses m modulo the remaining primes $p < \log X$ so that the $G(X)$ numbers starting from $m+2$ are composite. This argument gives about $(\log X)^{1+o(1)}$ numbers which are prime avoiding with distance $(\log X)^{1+o(1)}$.

But if we take two prime avoiding numbers n_1 and n_2 constructed in this way, then their sum (which we want to be an arbitrary N) is also close to a smooth number; the distance from $N \leqslant 2X$ to a number m' obeying (0.1) is at most $(\log X)^{1+o(1)}$, whereas there are at most

$$X \exp(-(0.5 + o(1)) \log X) = X^{1/2 - o(1)}$$

such numbers m'. So, N is not arbitrary at all.

We consider the following additive problem related to prime avoidance: can we prove that any large positive integer N can be represented as

$$N = n_1 + n_2, \qquad \text{where both } F(n_1) \text{ and } F(n_2) \text{ are large?}$$

To do so, one may try to use the "usual" approach for getting a long string of composite numbers, which originated from the work of Westzynthius. In that approach, one defines a "smooth" number $m \leqslant X$,

$$m \equiv 0 \pmod{p} \quad \text{for all } p \leqslant \log\log X \text{ and } (\log X)^{o(1)} < p \leqslant 0.5 \log X, \tag{0.1}$$

and chooses m modulo the remaining primes $p < \log X$ so that the $G(X)$ numbers starting from $m + 2$ are composite. This argument gives about $(\log X)^{1+o(1)}$ numbers which are prime avoiding with distance $(\log X)^{1+o(1)}$.

But if we take two prime avoiding numbers n_1 and n_2 constructed in this way, then their sum (which we want to be an arbitrary N) is also close to a smooth number; the distance from $N \leqslant 2X$ to a number m' obeying (0.1) is at most $(\log X)^{1+o(1)}$, whereas there are at most

$$X \exp(-(0.5 + o(1)) \log X) = X^{1/2 - o(1)}$$

such numbers m'. So, N is not arbitrary at all.

We consider the following additive problem related to prime avoidance: can we prove that any large positive integer N can be represented as

$$N = n_1 + n_2, \qquad \text{where both } F(n_1) \text{ and } F(n_2) \text{ are large?}$$

To do so, one may try to use the "usual" approach for getting a long string of composite numbers, which originated from the work of Westzynthius. In that approach, one defines a "smooth" number $m \leqslant X$,

$$m \equiv 0 \pmod{p} \quad \text{for all } p \leqslant \log\log X \text{ and } (\log X)^{o(1)} < p \leqslant 0.5 \log X, \tag{0.1}$$

and chooses m modulo the remaining primes $p < \log X$ so that the $G(X)$ numbers starting from $m+2$ are composite. This argument gives about $(\log X)^{1+o(1)}$ numbers which are prime avoiding with distance $(\log X)^{1+o(1)}$.

But if we take two prime avoiding numbers n_1 and n_2 constructed in this way, then their sum (which we want to be an arbitrary N) is also close to a smooth number; the distance from $N \leqslant 2X$ to a number m' obeying (0.1) is at most $(\log X)^{1+o(1)}$, whereas there are at most

$$X \exp(-(0.5 + o(1)) \log X) = X^{1/2 - o(1)}$$

such numbers m'. So, N is not arbitrary at all.

Simple probabilistic argument (which is to be discussed later) gives the following.

Proposition

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where $F(n_i) \gg \log N$, $i = 1, 2$.

Our goal is to improve the lower bound from this proposition by obtaining a result where $\log N$ is multiplied by some growing function. For a number $\rho \in (0,1)$, we define

$$C(\rho) = \sup\left\{\delta \in (0, 1/2) : \frac{6 \cdot 10^{2\delta}}{\log(1/(2\delta))} < \rho\right\}. \quad (0.2)$$

Our main result is the following.

Theorem

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where

$$F(n_i) \geqslant (\log N)(\log\log N)^{C(1/2)-o(1)},$$

for $i = 1, 2$.

Note that $C(1/2) > 1/325565$.

Simple probabilistic argument (which is to be discussed later) gives the following.

Proposition

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where $F(n_i) \gg \log N$, $i = 1, 2$.

Our goal is to improve the lower bound from this proposition by obtaining a result where $\log N$ is multiplied by some growing function. For a number $\rho \in (0, 1)$, we define

$$C(\rho) = \sup \left\{ \delta \in (0, 1/2) : \frac{6 \cdot 10^{2\delta}}{\log(1/(2\delta))} < \rho \right\}. \quad (0.2)$$

Our main result is the following.

Theorem

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where

$$F(n_i) \geqslant (\log N)(\log\log N)^{C(1/2)-o(1)},$$

for $i = 1, 2$.

Note that $C(1/2) > 1/325565$.

Prime avoiding numbers is a basis of order 2

Simple probabilistic argument (which is to be discussed later) gives the following.

Proposition

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where $F(n_i) \gg \log N$, $i = 1, 2$.

Our goal is to improve the lower bound from this proposition by obtaining a result where $\log N$ is multiplied by some growing function. For a number $\rho \in (0, 1)$, we define

$$C(\rho) = \sup\left\{\delta \in (0, 1/2) : \frac{6 \cdot 10^{2\delta}}{\log(1/(2\delta))} < \rho\right\}. \quad (0.2)$$

Our main result is the following.

Theorem

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where

$$F(n_i) \geqslant (\log N)(\log\log N)^{C(1/2)-o(1)},$$

for $i = 1, 2$.

Note that $C(1/2) > 1/325565$.

Theorem

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where

$$F(n_i) \geqslant (\log N)(\log\log N)^{C(1/2)-o(1)},$$

for $i = 1, 2$.

Note that $C(1/2) > 1/325565$.

This admits the following interpretation. Let us consider the set

$$\left\{n \geqslant 3 : F(n) \geqslant (\log n)(\log\log n)^{\delta}\right\}$$

(which is also kind of a set of prime avoiding numbers). Our Theorem then implies that this set is a basis of order 2 for any $\delta < C(1/2)$.

Theorem

Every sufficiently large positive integer N can be represented as the sum $N = n_1 + n_2$, where

$$F(n_i) \geqslant (\log N)(\log\log N)^{C(1/2)-o(1)},$$

for $i = 1, 2$.

Note that $C(1/2) > 1/325565$.

This admits the following interpretation. Let us consider the set

$$\left\{n \geqslant 3 : F(n) \geqslant (\log n)(\log\log n)^{\delta}\right\}$$

(which is also kind of a set of prime avoiding numbers). Our Theorem then implies that this set is a basis of order 2 for any $\delta < C(1/2)$.

To prove our Theorem, we apply the technique from the recent paper of Ford, Konyagin, Maynard, Pomerance, and Tao, where the authors used so-called hypergraph covering lemma to detect long gaps in general sieved sets.

Definition (Sieving System)

A sieving system is a collection $\mathcal{I}$ of sets $I_p \subset \mathbb{Z}/p\mathbb{Z}$ of residue classes modulo p for each prime p. Moreover, we have the following definitions.

- (Non-degeneracy) We say that the sieving system is non-degenerate if $|I_p| \leqslant p-1$ for all p.
- (B-Boundedness) Given $B > 0$, we say that the sieving system is B-bounded if $|I_p| \leqslant B$ for all primes p.
- (One-dimensionality) We say that the sieving system is one-dimensional if

$$\prod_{p \leqslant x} \left(1 - \frac{|I_p|}{p}\right) \sim \frac{C_1}{\log x}, \quad (x \to \infty),$$

  for some constant $C_1 > 0$.
- (ρ-supportedness) Given $\rho > 0$, we say that the sieving system system is ρ-supported if

$$\lim_{x \to \infty} \frac{|p \leqslant x : |I_p| \geqslant 1|}{x/\log x} = \rho.$$

Theorem (FKMPT, 2021)

For a sieving system defined above, the sieved set

$$S_x = S_x(\mathcal{I}) = \mathbb{Z} \backslash \bigcup_{p \leqslant x} I_p$$

(the set of integers which do not belong to any I_p for all $p \leqslant x$) contains a gap of size $x(\log x)^{C(\rho)-o(1)}$, where $C(\rho)$ is defined in (0.2) and the rate of decay in $o(1)$ depends on $\mathcal{I}$. Moreover, $C\rho > e^{-1-6/\rho}$.

Despite the fact that this general bound applied to the Eratosthenes sieve (that is, the sieving system with $I_p = \{0\}$ for all p) yields only a bound

$$G(X) \gg (\log X)(\log\log X)^{C(1)-o(1)} \gg (\log X)(\log\log X)^{1/835},$$

which is weaker than the aforementioned estimates, it has the advantage of not dealing with "smooth" numbers from the above discussion, and this is crucial for us.

Theorem (FKMPT, 2021)

For a sieving system defined above, the sieved set

$$S_x = S_x(\mathcal{I}) = \mathbb{Z} \setminus \bigcup_{p \leqslant x} I_p$$

(the set of integers which do not belong to any I_p for all $p \leqslant x$) contains a gap of size $x(\log x)^{C(\rho)-o(1)}$, where $C(\rho)$ is defined in (0.2) and the rate of decay in $o(1)$ depends on $\mathcal{I}$. Moreover, $C\rho > e^{-1-6/\rho}$.

Despite the fact that this general bound applied to the Eratosthenes sieve (that is, the sieving system with $I_p = \{0\}$ for all p) yields only a bound

$$G(X) \gg (\log X)(\log\log X)^{C(1)-o(1)} \gg (\log X)(\log\log X)^{1/835},$$

which is weaker than the aforementioned estimates, it has the advantage of not dealing with "smooth" numbers from the above discussion, and this is crucial for us.

Corollary

Let $f : \mathbb{Z} \mapsto \mathbb{Z}$ *be a polynomial of a degree* $d \geqslant 1$ *with positive leading term. Then for sufficiently large* X, *there is a string of consecutive natural numbers* $n \in [1, X]$ *of length* $\geqslant (\log X)(\log\log X)^{C(1/d)-o(1)}$ *for which* $f(n)$ *is composite, where* $C(1/d) > e^{-6d+1}$.

Corollary

Let $f : \mathbb{Z} \mapsto \mathbb{Z}$ *be a polynomial of a degree* $d \geqslant 2$ *with positive leading term, irreducible over* $\mathbb{Q}$, *and with full Galois group* S_d. *Then for sufficiently large* X, *there is a string of consecutive natural numbers* $n \in [1, X]$ *of length* $\geqslant (\log X)(\log\log X)^{1/325565}$ *for which* $f(n)$ *is composite.*

Corollaries for polynomials

Corollary

Let $f : \mathbb{Z} \mapsto \mathbb{Z}$ *be a polynomial of a degree* $d \geqslant 1$ *with positive leading term. Then for sufficiently large* X*, there is a string of consecutive natural numbers* $n \in [1, X]$ *of length* $\geqslant (\log X)(\log\log X)^{C(1/d)-o(1)}$ *for which* $f(n)$ *is composite, where* $C(1/d) > e^{-6d+1}$.

Corollary

Let $f : \mathbb{Z} \mapsto \mathbb{Z}$ *be a polynomial of a degree* $d \geqslant 2$ *with positive leading term, irreducible over* $\mathbb{Q}$*, and with full Galois group* S_d*. Then for sufficiently large* X*, there is a string of consecutive natural numbers* $n \in [1, X]$ *of length* $\geqslant (\log X)(\log\log X)^{1/325565}$ *for which* $f(n)$ *is composite.*

For a number $x \geqslant 2$, let

$$S_x = \{n \in \mathbb{Z} : n \not\equiv 0 \pmod p \text{ for each } p \leqslant x\}.$$

Let $z = x/2$ and let a number $b' \in \mathbb{Z}/P(z)\mathbb{Z}$ be chosen uniformly at random. We consider the random sets

$$A_{b'} := (S_z - b') \cap [-y, y]$$

and

$$A_{N-b'} := (S_z - N + b') \cap [-y, y],$$

where $y = \lfloor 0.08x \rfloor$.

We have

$$\mathbb{E}|A_{b'}| = \mathbb{E} \sum_{|n| \leqslant y} 1_{n \in S_z - b'} = \sum_{|n| \leqslant y} \prod_{p \leqslant z} \mathbb{P}(b' \not\equiv -n \pmod p)$$

$$= \sum_{|n| \leqslant y} \prod_{p \leqslant z} (1 - 1/p) = \frac{(2e^{-\gamma} + o(1))y}{\log z},$$

where γ is Euler's constant, and similarly

$$\mathbb{E}|A_{N-b'}| = \frac{(2e^{-\gamma} + o(1))y}{\log z}.$$

For a number $x \geqslant 2$, let

$$S_x = \{n \in \mathbb{Z} : n \not\equiv 0 \pmod p \text{ for each } p \leqslant x\}.$$

Let $z = x/2$ and let a number $b' \in \mathbb{Z}/P(z)\mathbb{Z}$ be chosen uniformly at random. We consider the random sets

$$A_{b'} := (S_z - b') \cap [-y, y]$$

and

$$A_{N-b'} := (S_z - N + b') \cap [-y, y],$$

where $y = \lfloor 0.08x \rfloor$.

We have

$$\mathbb{E}|A_{b'}| = \mathbb{E} \sum_{|n| \leqslant y} 1_{n \in S_z - b'} = \sum_{|n| \leqslant y} \prod_{p \leqslant z} \mathbb{P}(b' \not\equiv -n \pmod p))$$

$$= \sum_{|n| \leqslant y} \prod_{p \leqslant z} (1 - 1/p) = \frac{(2e^{-\gamma} + o(1))y}{\log z},$$

where γ is Euler's constant, and similarly

$$\mathbb{E}|A_{N-b'}| = \frac{(2e^{-\gamma} + o(1))y}{\log z}.$$

Therefore, if x is large enough,

$$\mathbb{E}(|A_{b'}| + |A_{N-b'}|) \leqslant \frac{5y}{\log x}.$$

Thus, there is a choice b' modulo $P(z)$ such that

$$|A_{b'}| + |A_{N-b'}| \leqslant \frac{0.4x}{\log x}.$$

Let $P_{z,x} = \prod_{z<p\leqslant x} p$ and

$$S_{z,x} = \{n \in \mathbb{Z} : n \not\equiv 0 \pmod p \quad \forall p \in (z,x]\}.$$

We set $b \equiv b' \pmod{P(z)}$ and claim that there is a choice $b \pmod{P_{z,x}}$ (let us denote it b'') such that

$$(S_x - b) \cap [-y,y] = (S_x - N + b) \cap [-y,y] = \varnothing. \tag{0.3}$$

To see that this is possible, note that

$$S_x - b = \{n \in \mathbb{Z} : n \not\equiv -b \pmod p \quad \forall p \leqslant x\} = (S_z - b') \cap (S_{z,x} - b'');$$

further, for each element $m \in A_{b'}$ we take a prime $q \in (z,x]$ and define $b \equiv b_q \pmod q$ such that $m \equiv -b_q \pmod q$; so, $m \notin S_{z,x} - b''$ and thus $m \notin S_x - b$. We do similarly for each $m \in A_{N-b'}$.

Now let $f(n) = \min\{|n-l| : l \in S_x\}$. Then

$$F(b) \geqslant f(b) \geqslant y, \quad F(N-b) \geqslant f(N-b) \geqslant y$$

for our choice of $b \pmod{P(x)}$. Now we choose $x \approx \log(N/2)$ maximally so that $P(x) \leqslant N/2$. We thus see that it is possible to take that b with $b \in [N/4, 3N/4]$; then $N - b \in [N/4, 3N/4]$ as well. This completes the proof.

To prove the Theorem, it is enough to show that for any fixed $\delta < C(1/2)$ and $y = [x(\log x)^{\delta}]$ there exists a choice of b modulo $P(x/2)$ such that

$$\left|\Big((S_{x/2} - b) \cup (S_{x/2} - N + b)\Big) \cap [-y, y]\right| \leqslant \left(\frac{1}{2} - \varepsilon\right) \frac{x}{\log x}$$

for some $\varepsilon > 0$.

Now let $f(n) = \min\{|n-l| : l \in S_x\}$. Then

$$F(b) \geqslant f(b) \geqslant y, \quad F(N-b) \geqslant f(N-b) \geqslant y$$

for our choice of $b \pmod{P(x)}$. Now we choose $x \approx \log(N/2)$ maximally so that $P(x) \leqslant N/2$. We thus see that it is possible to take that b with $b \in [N/4, 3N/4]$; then $N - b \in [N/4, 3N/4]$ as well. This completes the proof.

To prove the Theorem, it is enough to show that for any fixed $\delta < C(1/2)$ and $y = \lceil x(\log x)^{\delta} \rceil$ there exists a choice of b modulo $P(x/2)$ such that

$$\left| \Big((S_{x/2} - b) \cup (S_{x/2} - N + b) \Big) \cap [-y, y] \right| \leqslant \left(\frac{1}{2} - \varepsilon \right) \frac{x}{\log x}$$

for some $\varepsilon > 0$.

THANK YOU FOR YOUR ATTENTION !