Bull. Aust. Math. Soc. **111** (2025), 469–477
doi:10.1017/S0004972724001096

DIGITALLY RESTRICTED SETS AND THE GOLDBACH CONJECTURE

JAMES CUMBERBATCH

(Received 21 May 2024; accepted 28 September 2024; first published online 4 February 2025)

Abstract

We show that for any set D of at least two digits in a given base b, almost all even integers taking digits only in D when written in base b satisfy the Goldbach conjecture. More formally, if $\mathcal{A}$ is the set of numbers whose digits base b are exclusively from D, almost all elements of $\mathcal{A}$ satisfy the Goldbach conjecture. Moreover, the number of even integers in $\mathcal{A}$ which are less than X and not representable as the sum of two primes is less than $|\mathcal{A} \cap \{1, \dots, X\}|^{1-\delta}$.

2020 *Mathematics subject classification*: primary 11P32; secondary 11A63.

Keywords and phrases: Goldbach, digits, exceptional sets, digitally restricted sets.

1. Introduction

The Goldbach conjecture states that for all sufficiently large even n there exist primes p_1, p_2 such that

$$p_1 + p_2 = n. \tag{1.1}$$

In this paper, we study how many exceptions to the Goldbach conjecture can have restricted digits. Given a base b and a set of digits $D \subseteq \{0, 1, 2, \dots, b-1\}$ with $|D| > 2$ we define $\mathcal{A}_k$ to be the set of k-digit numbers whose base-b representations only contain elements of D. More formally, $\mathcal{A}_k$ is the set of numbers of the form

$$\sum_{i=0}^{k} a_i b^i$$

with $a_i \in D$. We call sets $\mathcal{A}_k$ of this form, with general k and D, *digitally restricted sets*, and similarly we call their elements digitally restricted numbers. Let $\mathcal{A}$ be the union over k of $\mathcal{A}_k$, that is, the set of digitally restricted integers with any number of digits. Let E be the set of even n which are not the sum of two primes and let $E(X) = E \cap \{1, \dots, X\}$.

The author's work is supported by NSF grant DMS-2001549.

© The Author(s), 2025. Published by Cambridge University Press on behalf of Australian Mathematical Publishing Association Inc.

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

THEOREM 1.1. *There exists some* $\delta = \delta(D, b)$ *such that* $|E(X) \cap \mathcal{A}| \ll |\mathcal{A}(X)|^{1-\delta}$.

To see the relevance of this result, we look at historical results on the Goldbach conjecture. First conjectured by Goldbach in 1742, little progress occurred until the work of Vinogradov [13] in 1937, who showed that all sufficiently large odd numbers are the sum of three primes. Heilbronn [4] noted in his review that one could use the same method to prove that almost all even numbers are the sum of two primes, and within a year Estermann [3], van der Corput [11] and Chudakov [2] all independently did so. Although van der Corput's paper appears in the 1936 edition of *Acta Arithmetica*, it was submitted in 1937, after Vinogradov's work. The initial bound of $|E(X)| \ll X(\log X)^{-c}$ has since been improved, most notably by Montgomery and Vaughan in [8] to $|E(X)| \ll X^{1-\delta}$ for some $\delta > 0$, and the current best δ which can be found in the literature is $\delta = 0.28$ by Pintz in [10].

Another motivation for our result is previous study of how many exceptions lie in other specific thin sets. This was first studied by Perelli in [9], who showed that for any polynomial ϕ of degree k with integer coefficients the values of $2\phi(n)$ obey the Goldbach conjecture for all but $N(\log N)^{-A}$ many $n < N$. This estimate was improved by Brüdern *et al.* in [1] who lowered the bound on the exceptions to $N^{1-c/k}$ for some positive absolute constant c. When bounding the number of exceptions in a thin set, it is important to check that the thin set is thinner than the potential set of exceptions, or the result will trivially follow without looking at any specifics of the thin set. In the case of values of a polynomial we can see that provided the polynomial is not linear it may have at most $cX^{1/2}$ many values less than X, which is considerably fewer than $X^{0.72}$ many values in the thin set. When considering digitally restricted sets, we first note that by counting strings of length k in an alphabet with $|D|$ letters,

$$|\mathcal{A}_k| = |D|^k = X^{(\log|D|)/\log b},$$

and hence

$$|\mathcal{A}(X)| \ll X^{(\log|D|)/\log b}.$$

When $|D| < b^{0.72}$, there are thus significantly fewer than $X^{0.72}$ numbers with restricted digits less than X, and Theorem 1.1 is nontrivial.

Also motivating our result is work studying the set $\mathcal{A}$. The analytic structure of $\mathcal{A}$ has led to many results. Particularly notable results include those of Maynard [7] who proved that $\mathcal{A}$ has infinitely many primes provided that the digit set D omits sufficiently few digits, and Mauduit and Rivat [6] who proved that sums of digits of primes are well distributed in residue classes.

2. Preliminaries

We first introduce notation. As is common in analytic number theory, let $e(\alpha) := e^{2\pi i\alpha}$ and let the symbol $f(x) \ll g(x)$ mean that there exists a positive constant c such that $|f(x)| \le cg(x)$ after restricting to an appropriate domain. Whenever we use the letter ε, we mean that the statement is true for every $\varepsilon > 0$. Throughout this paper,

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

the letter p always denotes primes. Take k, the number of digits, to be sufficiently large, and let $X = b^k$, so that $\mathcal{A}_k$ has only values less than X.

When discussing digitally restricted sets we need to deal with the technicality of leading zeros. We want to study all digitally restricted numbers less than X, of any number of digits, but when $0 \notin D$ we see that $\mathcal{A}_k$ only contains numbers with precisely k digits and when $j < k$ we have $\mathcal{A}_j \cap \mathcal{A}_k = \emptyset$, in contrast to the case where $0 \in D$, in which a j-digit number is also a k-digit number with $k - j$ leading zeros and $\mathcal{A}_j \subset \mathcal{A}_k$. When $0 \in D$ we may show that $|\mathcal{A}_k \cap E(X)|$ is small to obtain the desired result, but when $0 \notin D$ we note that $|\mathcal{A} \cap E(X)| = \sum_{j \le k} |\mathcal{A}_k \cap E(X)|$, so proving the result for $\mathcal{A}_k$ will prove it for $\mathcal{A}(X)$ in either case, and we will henceforth deal only with $\mathcal{A}_k$.

Now we define the generating function for $\mathcal{A}_k$ to be

$$f(\alpha) := \sum_{n \in \mathcal{A}_k} e(n\alpha).$$

The generating function for primes on the other hand will require a parameter. Let δ_0 be a sufficiently small constant, which may depend on b, to be chosen later. We define

$$S(\alpha) := \sum_{X^{6\delta_0} < p < X} e(p\alpha) \log p.$$

For any $\mathfrak{B} \subset [0, 1]$ and any even $n < X$, we define

$$r(n, \mathfrak{B}) := \int_{\mathfrak{B}} S(\alpha)^2 e(-n\alpha)\, d\alpha.$$

By orthogonality, $r(n, [0, 1])$ counts solutions to $p_1 + p_2 = n$ with weight $\log p_1 \log p_2$. To evaluate this integral we apply the Hardy–Littlewood circle method. Define the major arcs to be

$$\mathfrak{M}_{\delta_0} := \bigcup_{0 < q < X} \{\alpha \in [0, 1] : \|q\alpha\| < X^{6\delta_0 - 1}\} \tag{2.1}$$

and the minor arcs to be

$$\mathfrak{m}_{\delta_0} := [0, 1] \setminus \mathfrak{M}_{\delta_0}. \tag{2.2}$$

In Section 3 we establish Lemmas 3.1 and 3.2, the facts about $f(\alpha)$ and $\mathcal{A}_k$ which we will use in the proof. In Section 4 we combine Section 3 with commonly known facts about S to prove Theorem 1.1.

3. Properties of digital restrictions

In this section we prove the properties of digitally restricted sets which we will need to show Theorem 1.1. The two results we need are Lemma 3.1, a mean value estimate, and Lemma 3.2, a bound on how many elements of $\mathcal{A}_k$ can be multiples of a given large value.

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

LEMMA 3.1. *For all $\delta > 0$, there exists a $t_0 = t_0(b, \delta)$ such that whenever $t \geq t_0$,*

$$\int_0^1 |f(\alpha)|^t \, d\alpha \ll |\mathcal{A}_k|^t X^{\delta - 1}.$$

PROOF. We prove the bound by exploiting the independence of digits of values in $\mathcal{A}_k$. By treating each digit as its own variable, the number of variables goes to infinity while the number of values any variable can take remains constant.

Note that by orthogonality, if $t = 2s$ is an even integer, the integral counts solutions to the equation

$$\sum_{i=1}^{s} (x_i - y_i) = 0$$

with $x_i, y_i \in \mathcal{A}_k$. We now interpret each x_i and y_i as a linear combination of k variables, each corresponding to a digital place. Let

$$x_i = \sum_{j=0}^{k-1} x_{i,j} b^j$$

and the same for y_i. We are now counting solutions to

$$\sum_{j=0}^{k} b^j \sum_{i=1}^{s} (x_{i,j} - y_{i,j}) = 0 \tag{3.1}$$

with $x_{i,j}, y_{i,j} \in D$.

First, if every allowed digit shares a common factor then we know that every $x_{i,j}$ and $y_{i,j}$ will share that common factor, permitting us to divide everything by that common factor. Thus, we may assume that the allowed digits do not share a common factor. To count solutions to (3.1) we investigate what addition looks like in base b, and in particular we look at each digital place. Let

$$\sigma_j := \sum_{i=1}^{s} (x_{i,j} - y_{i,j})$$

and

$$\tau_j := \sum_{l<j} b^l \sigma_l.$$

Now, we note that for (3.1) to hold, we require that it holds modulo b, and thus $\sigma_0 \equiv 0 \mod b$, or to be precise, the 1s digits must add up to 0 mod b. Further, we know that (3.1) holds mod b^j for any j, meaning that $\tau_j \equiv -b^j \sigma_j \mod b^{j+1}$. After we have chosen the digits up to j we will thus require that $\sigma_j \equiv n \mod b$ for some fixed n which is determined by the digits up to j.

No matter what the digits in the other places are, we will thus have at most

$$\max_{n<b} \#\{(x_{0,j}, x_{1,j}, \ldots, x_{s,j}, y_{0,j}, \ldots, y_{s,j}) : \sigma_j \equiv n \bmod b\}$$

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

choices for how to assign the jth digits. By orthogonality, we can count these using the function

$$u(n,b,D) := \frac{1}{b}\sum_{0\le a<b}\left(\left|\sum_{d\in D} e\Big(\frac{da}{b}\Big)\right|^{2s} e\Big(-\frac{na}{b}\Big)\right).$$

Noting that this does not depend on j, we thus have an upper bound on the number of solutions to (3.1) which is the product of the upper bounds on the sums of each digit, which is bounded by the maximum of $u(n,b,D)^k$ over all possible n. Finally, noting that by the lack of a common factor of acceptable digits, for all nonzero a,

$$\left|\sum_{d\in D} e\Big(\frac{da}{b}\Big)\right| < |D|.$$

Dividing both sides by $|D|$ and noting that there are finitely many nonzero a then yields

$$\max_{0<a<b}\frac{1}{D}\left|\sum_{d\in D} e\Big(\frac{da}{b}\Big)\right| < 1. \tag{3.2}$$

We now choose s to be at least

$$\frac{\log(\delta/a)}{\log\left(\max_{0<a<b}\frac{1}{D}\left|\sum_{d\in D} e\left(\frac{da}{b}\right)\right|\right)}.$$

By (3.2), this must be a finite constant depending only on b, D and δ, which makes it a valid choice of s. Summing over all values of a then yields

$$\sum_{0\le a<b}\left(\frac{1}{D}\left|\sum_{d\in D} e\Big(\frac{da}{b}\Big)\right|^{2s}\right) < 1+\delta.$$

Rearranging this and noting that since $|D|\ge 2$, we have $|D|^{\delta} > 1+\delta$, we can see that

$$u(b;D) = \frac{1}{b}\sum_{0\le a<b}\left|\sum_{d\in D} e\Big(\frac{da}{b}\Big)\right|^{2s} \le \frac{(1+\delta_s)}{b}|D|^{2s} \le \frac{1}{b}|D|^{2s+\delta}.$$

Taking the product across all digital places now yields the desired result. □

LEMMA 3.2. *For any m, we have*

$$\#\{n\in\mathcal{A}_k : m \mid n\} \le \frac{b|\mathcal{A}_k|}{b^{-1}+(m/2)^{\log|D|/\log b}} \ll \frac{|\mathcal{A}_k|}{m^{\log|D|/\log b}}.$$

PROOF. Note that for every integer $n\in\mathcal{A}_k$ which is a multiple of m, we may change the last $\lfloor \log(m/2)/\log b\rfloor$ many digits and obtain $|D|^{\lfloor\log(m/2)/\log b\rfloor} \ge b^{-1}(m/2)^{\log|D|/\log b}$ many other values in $\mathcal{A}_k$ which are less than $m/2$ away from n. Since any integer may be less than $m/2$ from at most one multiple of m, this provides a map from values in $\mathcal{A}_k$ which are multiples of m to disjoint sets of at least $b^{-1}(m/2)^{\log|D|/\log b}$ many integers in $\mathcal{A}_k$ which are not multiples of m. Taking the disjoint union of all these sets, along

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

with the associated multiples of m, returns a subset of $\mathcal{A}_k$, and hence the sum of the cardinalities must be at most $|\mathcal{A}_k|$. More formally, we see that

$$\sum_{\substack{n\in\mathcal{A}_k\\ m|n}} 1 + b^{-1}(m/2)^{\log|D|/\log b} \le |\mathcal{A}_k|.$$

and hence, dividing by the term being summed, we see that

$$\sum_{\substack{n\in\mathcal{A}_k\\ m|n}} 1 \le \frac{b|\mathcal{A}_k|}{b^{-1} + (m/2)^{\log|D|/\log b}},$$

which is the desired result. □

REMARK 3.3. The above lemma also holds, with the same proof, if we count any particular residue class modulo m. In particular, it is not unique to 0.

Since the expected number of multiples of m lying in $\mathcal{A}_k$ would be $|\mathcal{A}_k|/m$, it is natural to ask whether we can obtain this or possibly $|\mathcal{A}_k|/m^c$ for some absolute c which does not depend on b and $|D|$. Unfortunately, this is ruled out by those m which are multiples of powers of b. Taking $m = b^j$, we see that $n \in \mathcal{A}_k$ being a multiple of m is equivalent to the last j digits being 0, and that when 0 is an allowed digit there will be $N/N^{j/k} = N/m^{\log|D|/\log b}$ of these. For values m which are coprime to b, a significant amount of research has been done (see, for example, [5]), but this only applies to small values of m and not the large values we will need, about which few positive results are known. Konyagin [5] showed that we cannot hope for too much, as there exist relatively small coprime moduli, of the order of $\exp(c\log X/\log\log X)$, in which $\mathcal{A}$ is not evenly distributed between residue classes. Despite not being evenly distributed, it is still expected that $\mathcal{A}$ is not too badly distributed. In particular, Konyagin conjectured that $\mathcal{A}(X)$ meets every residue class modulo m whenever $(m, b) = 1$ and $m < X^c$ for some constant $c = c(b)$.

4. Exceptional sets in Goldbach's conjecture

In this section, following [1], we use what we have proven about numbers with restricted digits to prove Theorem 1.1. We now study $r(n, \mathfrak{B})$ as we defined it in Section 2. In particular, we study $r(n, \mathfrak{M})$ and $r(n, \mathfrak{m})$ as defined in (2.1) and (2.2), respectively. Our goal will be to show, for almost all n in $\mathcal{A}_k$, that $r(n, \mathfrak{m}) \ll X^{1-\delta_1}$ and $r(n, \mathfrak{M}) \gg X^{1-\delta_2}$ with $\delta_1 < \delta_2$. First, we handle the minor arcs.

LEMMA 4.1. *There exists $\delta_1 = \delta_1(b, D) > 0$ such that for all sufficiently large X, there exists a set $B \subset \mathcal{A}_k$ with $|B| \ll |\mathcal{A}_k|X^{-\delta_1}$, and for all even n in $\mathcal{A}_k \setminus B$ we have*

$$|r(n, \mathfrak{m})| \ll |\mathcal{A}_k|X^{-\delta_1}$$

PROOF. By symmetry of $\mathfrak{m}$, we see that $r(n, \mathfrak{m})$ is real. Set

$$\eta(n) := \operatorname{sign}(r(n, \mathfrak{m})).$$

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

Now, we let

$$K(\alpha) := \sum_{\substack{0<n<X \\ n\in\mathcal{A}}} \eta(n)e(-n\alpha)$$

and rewrite the expression we are trying to bound as

$$\sum_{n\in\mathcal{A}_k} |r(n,\mathfrak{m})| = \int_{\mathfrak{m}} S(\alpha)^2 K(-\alpha)\,d\alpha.$$

From Hölder's inequality, this is at most

$$\Big(\sup_{\alpha\in\mathfrak{m}}|S(\alpha)|\Big)^{2/t}\Big(\int_0^1 |S(\alpha)|^2\,d\alpha\Big)^{1-1/t}\Big(\int_0^1 |K(-\alpha)|^t\,d\alpha\Big)^{1/t},$$

and we can now estimate each term individually.

First, by [12],

$$\sup_{\alpha\in\mathfrak{m}}|S(\alpha)| \ll X^{1-3\delta_0}(\log X)^4.$$

By applying orthogonality, we trivially have

$$\int_0^1 |S(\alpha)|^2\,d\alpha \ll \sum_{P<p<X}(\log p)^2 \ll X\log X.$$

Now we estimate the contribution of K. Observe that for integer values of t, the mean value $\int_0^1 |K(\alpha)|^{2t}\,d\alpha$ counts solutions with $x_i, y_i \in \mathcal{A}_k$ to the additive equation

$$\sum_{i\le t}(x_i - y_i) = 0,$$

with each solution being weighted with ± 1. Thus, it is clearly less than or equal to the unweighted version, namely $\int_0^1 |f(\alpha)|^{2t}\,d\alpha$, which when we take t sufficiently large is itself $O(|\mathcal{A}_k|^{2t}X^{\delta_0-1})$ by Lemma 3.1.

Thus, in total,

$$\begin{aligned}\int_0^1 S(\alpha)^2K(-\alpha)\,d\alpha &\ll |\mathcal{A}_k|X^{1-1/t+(2-6\delta_0)/t+(-1+\delta)/t}(\log X)^{8/t+1-1/t}\\ &= |\mathcal{A}_k|X^{1-5\delta_0/t}(\log X)^{1+7/t}\end{aligned}$$

implying that

$$\sum_{\substack{n<X \\ n\in\mathcal{A}}} |r(n,\mathfrak{m})| \ll NX^{1-2\delta_2}$$

for some $\delta_2 > 0$. Now using the trivial bound, that

$$|\{n\in\mathcal{A}_k : |r(n,\mathfrak{m})| > X^{1-\delta_1}\}| \le X^{\delta_1-1}\sum_{n\in|\mathcal{A}_k|}|r(n,\mathfrak{m})|,$$

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

we obtain the desired result. We note that we just bounded the number of values in $\mathcal{A}$ with a large minor arc, and the number of even values in $\mathcal{A}$ with a large minor arc is clearly even less. □

Next we deal with the major arcs.

LEMMA 4.2. *Given Y such that $1 \le Y \le |\mathcal{A}_k|^{\delta_0}$, we have*

$$r(n, \mathfrak{M}) \gg XY^{-1/2}(\log X)^{-1}$$

for all even $n \in \mathcal{A}_k(X)$ except for at most $O(|\mathcal{A}_k|^{1+\varepsilon} Y^{-(\log|D|)/\log b})$ many n.

PROOF. This follows immediately from the proof of Lemma 2 in [1]. Note that the only step in the proof which depends on the set being counted occurs in the display following (18) of the latter paper, where numbers n such that $(n, \tilde{r}) > Y$ were discarded. We replace that with

$$\sum_{\substack{d|\tilde{r} \\ d>Y}} \sum_{\substack{n\in\mathcal{A}_k \\ D|n}} 1 \ll \sum_{\substack{d|\tilde{r} \\ d>Y}} (|\mathcal{A}_k|/d^{\log|D|/\log b}) \ll |\mathcal{A}_k|^{1+\varepsilon} Y^{-\log|D|/\log b}.$$

If we discard $O(|\mathcal{A}_k|^{1+\varepsilon} Y^{-\log|D|/\log b})$ many numbers, the rest of the proof follows. □

Now that we have both the major arcs and the minor arcs, we are equipped to prove Theorem 1.1. Taking Y to be less than $X^{\delta_1/3}$, we see that there is some set B such that for all $n \in B$, we have $r(n, \mathfrak{m}) = o(r(n, \mathfrak{M}))$ and

$$|\mathcal{A} \setminus B| \ll |\mathcal{A}_k|^{1+\varepsilon} X^{-(\delta_1 \log|D|)/(3\log b)} + |\mathcal{A}_k| X^{-\delta_1}.$$

Recalling that the number of solutions to (1.1) is

$$r(n, [0,1]) = r(n, \mathfrak{m}) + r(n, \mathfrak{M}) \gg r(n, \mathfrak{M}) > 0,$$

for all n other than those exceptions, shows that the Goldbach conjecture holds for all other values in $\mathcal{A}_k$.

Acknowledgements

The author would like to thank Trevor Wooley for considerable assistance on this paper. The author would also like to thank the anonymous referee for many helpful suggestions.

References

[1] J. Brüdern, K. Kawada and T. D. Wooley, 'Additive representation in thin sequences, II: The binary Goldbach problem', *Mathematika* **47**(1–2) (2000), 117–125.
[2] N. G. Chudakov, 'Sur le problème de Goldbach', *C. R. (Dokl.) Acad. Sci. URSS, n. Ser.* **17** (1937), 335–338.
[3] T. Estermann, 'On Goldbach's problem: proof that almost all even positive integers are sums of two primes', *Proc. Lond. Math. Soc. (2)* **44**(1) (1938), 307–314.
[4] H. Heilbronn, 'Review of I. M. Vinogradov, "Representation of an odd number as a sum of three primes"', *Z. Math.* **16** (1937), 291–292.

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press

[5] S. Konyagin, 'Arithmetic properties of integers with missing digits: distribution in residue classes', *Period. Math. Hungar.* **42**(1–2) (2001), 145–162.
[6] C. Mauduit and J. Rivat, 'Sur un problème de Gelfond: la somme des chiffres des nombres premiers', *Ann. of Math. (2)* **171** (2010), 1591–1646.
[7] J. Maynard, 'Primes with restricted digits', *Invent. Math.* **217**(1) (2019), 127–218.
[8] H. L. Montgomery and R. C. Vaughan, 'The exceptional set of Goldbach's problem', *Acta Arith.* **27** (1975), 353–370.
[9] A. Perelli, 'Goldbach numbers represented by polynomials', *Rev. Mat. Iberoam.* **12**(2) (1996), 477–490.
[10] J. Pintz, 'A new explicit formula in the additive theory of primes with applications II. The exceptional set in Goldbach's problem', Preprint, 2018, arXiv:1804.09084.
[11] J. van der Corput, 'Sur l'hypothèse de Goldbach pour presque tous les nombres pairs', *Acta Arith.* **2**(2) (1936), 266–290.
[12] R. C. Vaughan, 'Sommes trigonométriques sur les nombres premiers', *C. R. Acad. Sci. Paris Sér. A-B* **285**(16) (1977), 981–983.
[13] I. M. Vinogradov, 'Representation of an odd number as a sum of three primes', *C. R. (Dokl.) Acad. Sci. URSS, n. Ser.* **15** (1937), 169–172.

JAMES CUMBERBATCH, Department of Mathematics,
Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, USA
e-mail: jcumberb@purdue.edu

https://doi.org/10.1017/S0004972724001096 Published online by Cambridge University Press