

ARTICLE

An explicit economical additive basis

Vishesh Jain¹, Huy Tuan Pham², Mehtaab Sawhney³ and Dmitrii Zakharov⁴

¹Department of Mathematics, Statistics, and Computer Science, University of Illinois Chicago, Chicago, IL, USA,

²Department of Mathematics, Stanford University, Stanford, CA, USA, ³Department of Mathematics, Columbia University, New York, NY, USA, and ⁴Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Corresponding author: Vishesh Jain; Email: visheshj@uic.edu

(Received 22 August 2024; accepted 19 May 2025)

Abstract

We present an explicit subset $A \subseteq \mathbb{N} = \{0, 1, \dots\}$ such that $A + A = \mathbb{N}$ and for all $\varepsilon > 0$,

$$\lim_{N \rightarrow \infty} \frac{\left| \{(n_1, n_2) : n_1 + n_2 = N, (n_1, n_2) \in A^2\} \right|}{N^\varepsilon} = 0.$$

This answers a question of Erdős.

Keywords: Additive number theory; additive bases

2020 MSC Codes: 11B13

1. Introduction

Sidon asked (see [7, 20]) whether there exists a set $A \subseteq \mathbb{N}$ such that $A + A = \mathbb{N}$ (i.e. A is an additive basis of order 2) and for all $\varepsilon > 0$,

$$\lim_{N \rightarrow \infty} \frac{\left| \{(n_1, n_2) : n_1 + n_2 = N, (n_1, n_2) \in A^2\} \right|}{N^\varepsilon} = 0.$$

Erdős [7] answered Sidon's question by showing that there exists an additive basis of order 2 which, in fact, satisfies the stronger bound

$$\limsup_{N \rightarrow \infty} \frac{\left| \{(n_1, n_2) : n_1 + n_2 = N, (n_1, n_2) \in A^2\} \right|}{\log N} < \infty.$$

It is a major open problem whether there exists an additive basis of order 2 for which the factor of $\log N$ in the denominator can be replaced by an absolute constant; Erdős and Turán [6] famously conjectured that this is impossible.

Erdős's proof of the existence of A is randomized; in modern notation, one simply includes the number n in the set A with probability proportional to $C(\log n)^{1/2}n^{-1/2}$. Kolountzakis [12] derandomized (a variation of) Erdős's proof in the sense that one can deterministically generate the elements of $A \cap \{0, \dots, N\}$ in time $N^{O(1)}$. We remark that a number of variants of the original result of Erdős have been developed including results of Erdős and Tetali [11] which prove the analogous result for higher order additive bases and results of Vu [21] regarding economical versions of Waring's theorem.

The focus of this work is on 'explicit' constructions. Erdős several times [7–9] asked for an explicit set A which affirmatively answers Sidon's question and, in fact, offered \$100 for a

solution [9]. We note that if one takes A to be the set of squares, then $A + A$ contains all primes which are 1 mod 4 and by the divisor bound $A + A$ has multiplicities bounded by $N^{o(1)}$. Therefore, if one is willing to assume strong number-theoretic conjectures, one can take A to be the set of numbers n which are within $O((\log n)^{O(1)})$ a square. The purpose of this note is to present an explicit construction unconditionally.

Given a set A which is either a subset of $\mathbb{N}$ or $\mathbb{Z}/q\mathbb{Z}$ for some q , we denote by $\sigma_A(n)$ the number of representations $n = a + a'$ or $n \equiv a + a' \pmod{q}$, where $a, a' \in A$.

Theorem 1.1. *There is an explicit set $A \subset \mathbb{N}$ and absolute constants $C, c > 0$ such that for every $n \in \mathbb{N}$, we have $1 \leq \sigma_A(n) \leq Cn^{c/\log \log n}$.*

Our starting point is an explicit construction of a set $A_p \subseteq \mathbb{Z}/(p^2\mathbb{Z})$ due to Ruzsa [18] (for a prime $p \equiv 3, 5 \pmod{8}$) such that $A_p + A_p = \mathbb{Z}/(p^2\mathbb{Z})$ and $\sigma_{A_p}(r) = O(1)$ for all $r \in \mathbb{Z}/(p^2\mathbb{Z})$. Given this set and a sequence of squares of primes $\mathbf{b} = (p_1^2, p_2^2, \dots)$, A is given by

$$A = \bigcup_{k \geq 1} \{\overline{a_k a_{k-1} \dots a_1}^{\mathbf{b}}, a_j \in A_{p_j} \text{ for } j = 1, \dots, k-1, \text{ and } a_k \in \{0, \dots, p_k^2 - 1\}\}.$$

Here the overline denotes a generalized base expansion in base $\mathbf{b}$, that is

$$\overline{a_k a_{k-1} \dots a_1}^{\mathbf{b}} = \sum_{i=1}^n a_i \prod_{j < i} b_j;$$

here $b_j = p_j^2$. In words, the set is given by specifying that all but the leading digit of the generalized base expansion lies in the specially constructed sets A_p . By working upward from the smallest digit, one can use the property that $A_p + A_p$ covers all residues in $\mathbb{Z}/(p^2\mathbb{Z})$ to see that all natural numbers are represented. The fact that no number is represented too many times is similarly derived using that $A_p + A_p$ is ‘flat’, in particular that the multiplicities are bounded by $p^{o(1)}$. We remark that generalized bases have been utilized in a variety of questions related to Sidon sets, including works of Ruzsa [19], Cilleruelo, Kiss, Ruzsa and Vinuesa [4], and Pillate [17].

The construction above is clearly ‘explicit’ in the traditional mathematical sense (the sets A_p are given by the union of 3 parabolas over $\mathbb{F}_p \times \mathbb{F}_p$ and then projecting; see Lemma 2.2). Furthermore, we can also examine the word ‘explicit’ from the computational complexity perspective. In analogy with the long and established line of work on explicit Ramsey graphs [2, 5, 15, 16], we say that a set $A \subset \mathbb{N}$ is explicit if one may test membership $n \in A$ in $(\log n)^{O(1)}$ -time, that is, polynomial in the number of digits. We note that for the purpose of simply obtaining an upper bound of $N^{o(1)}$ in Theorem 1.1, one can actually choose the primes p sufficiently small (e.g. of size $\log \log \log N$ say) and find a suitable set A_p by brute force enumeration. However Ruzsa’s [18] construction is ‘strongly explicit’ (i.e. membership can be tested in time $O((\log p)^{O(1)})$).

The current bottleneck in Theorem 1.1 (given Ruzsa’s construction) is finding the smallest prime in an interval $[N, 2N]$. Under strong number-theoretic conjectures (e.g. Cramer’s conjecture) finding such a prime would take time $O((\log N)^{O(1)})$ due to the AKS primality testing algorithm [1]. Assuming this to be the case (or under a more traditional ‘mathematical’ definition of explicit), we can choose the primes more carefully to obtain an improved upper bound of $\exp(O((\log N)^{1/2}))$ (see Section 2.2). The limiting feature of our construction now is that in the top block, one is forced to allow ‘all possibilities’. We believe obtaining an explicit construction achieving $\sigma_A(N) \leq \exp((\log N)^\epsilon)$ or better would be of substantial interest.

Notation

Throughout this paper we let $[N] = \{0, \dots, N-1\}$ and $\mathbb{N} = \{0, 1, 2, \dots\}$. We let $\lfloor x \rfloor$ denote the largest integer less than or equal to x . We use standard asymptotic notation, for example, $f \lesssim g$ if

$|f(n)| \leq C|g(n)|$ for some constant C and all large enough n . We usually denote by c, C absolute constants which may change from line to line.

2. Proof of Theorem 1.1

We formally define the notion of a generalized base.

Definition 2.1. Let $\mathbf{b} = (b_1, b_2, \dots)$ be an infinite set of integers such that $b_i \geq 2$. Given any integer $x \in \mathbb{N}$ there exists a representation $x = \overline{a_n \dots a_1}^{\mathbf{b}}$ with $0 \leq a_i \leq b_i - 1$ such that

$$x = \sum_{i=1}^n a_i \prod_{j < i} b_j.$$

Here an empty product (when $i = 1$) is treated as 1.

Remark. If one requires $a_n \neq 0$ (e.g. does not have leading zeros) the representation is unique. When $b_j = g$ for all j , we recover precisely the base- g expansion.

A crucial piece in our construction is an ‘economical’ modular additive basis of order 2 over $\mathbb{Z}/(p^2\mathbb{Z})$ due to Ruzsa [18, Theorem 1]; the precise constant M in the result below has been studied in [3].

Lemma 2.2. *There exists an absolute constant $M \geq 1$ such that the following holds. Consider a prime p such that $p \equiv 3, 5 \pmod{8}$. There exists a set $A_p \subseteq \mathbb{Z}/(p^2\mathbb{Z})$ such that for all $r \in \mathbb{Z}/(p^2\mathbb{Z})$, we have $1 \leq \sigma_{A_p}(r) \leq M$. Furthermore, given p and $x \in \mathbb{Z}/(p^2\mathbb{Z})$, one can check whether $x \in A_p$ in time $O((\log p)^{O(1)})$.*

For completeness, and especially in order to discuss the second part of the statement, we present the proof of Lemma 2.2 in Section 2.1.

We next need the following basic fact about deterministically finding primes, which is immediate via (say) the Sieve of Eratosthenes. Using a more sophisticated algorithm of Lagarias and Odlyzko [14], one may obtain a run time of $O(N^{1/2+o(1)})$ in the statement below; runtimes of the form $O(N^{o(1)})$ remain a major open problem.

Lemma 2.3. *Let $N \geq C_{2.3}$. Then one may produce the smallest prime $p \in [N, 2N]$ such that $p \equiv 3 \pmod{8}$ in time $O(N^{1+o(1)})$.*

We now are in position to give the proof of Theorem 1.1.

Proof of Theorem 1.1. Let $f: \mathbb{N} \rightarrow \mathbb{N}$ be an arbitrary monotone increasing function such that $f(k) \geq C_0$ for some large constant C_0 and all k . Let $p_1 < p_2 < \dots$ be a sequence of primes such that $p_k \equiv 3 \pmod{8}$ and $p_k \in [f(k), 2f(k))$ is the least such prime.¹ Define $\mathbf{b} = (b_1, b_2, \dots)$ by setting $b_k = p_k^2$ for all $k \geq 1$. We are going to define our set A in terms of its expansion in the generalized base $\mathbf{b}$. Namely, for each $k \geq 1$ let $A_k \subset \{0, 1, \dots, p_k^2 - 1\}$ be the set from Lemma 2.2 (where we lift elements mod p_k^2 to their integer representatives) and consider the set

$$A = \bigcup_{k \geq 1} \{\overline{a_k a_{k-1} \dots a_1}^{\mathbf{b}}, a_j \in A_j \text{ for } j = 1, \dots, k-1, \text{ and } a_k \in \{0, \dots, b_k - 1\}\}. \quad (2.1)$$

We begin by showing that $A + A = \mathbb{N}$. For any $n \in \mathbb{N}$, we (uniquely) write $n = \overline{n_k \dots n_1}^{\mathbf{b}}$ for some $k \geq 1$; we will construct the representation $n = a + a'$ for $a, a' \in A$ digit by digit. First, since A_1 is an order 2 additive basis mod b_1 , there exist $a_1, a'_1 \in A_1$ such that $n_1 \equiv a_1 + a'_1 \pmod{b_1}$.

¹That such a prime exists for $f(k)$ larger than an absolute constant follows via the Siegel-Walfisz theorem (see e.g. [13, Theorem 12.1]).

Let $c_1 = \lfloor \frac{a_1 + a'_1}{b_1} \rfloor \in \{0, 1\}$ be the carry bit. Next, there exist $a_2, a'_2 \in A_2$ such that $n_2 - c_1 \equiv a_2 + a'_2 \pmod{b_2}$. As before, define the carry bit c_2 and continue in the same fashion to produce sequences of digits $a_1, \dots, a_{k-1}$ and $a'_1, \dots, a'_{k-1}$ and a carry bit $c_{k-1} \in \{0, 1\}$. Finally, let $a_k = n_k - c_k \in \{0, \dots, b_k - 1\}$ and consider the elements

$$a = \overline{a_k \dots a_1}^{\mathbf{b}}, \quad a' = \overline{a'_k \dots a'_1}^{\mathbf{b}}.$$

By construction, we have $n = a + a'$ and $a, a' \in A$.

Next, we bound the number of possible representations $n = a + a'$ with $a, a' \in A$. Write $n = \overline{n_k \dots n_1}^{\mathbf{b}}$, $a = \overline{a_\ell \dots a_1}^{\mathbf{b}}$, and $a' = \overline{a'_{\ell'} \dots a'_1}^{\mathbf{b}}$, where $\ell, \ell' \leq k$ are the digit lengths of a and a' . We may assume that $\ell \leq \ell'$ (this costs us a factor of 2 in the number of representations). Since $a, a' \in A$, we have $a_i \in A_i$ for $i \leq \ell - 1$ and $a'_i \in A_i$ for $i \leq \ell' - 1$ but the top digits a_ℓ , (respectively, $a'_{\ell'}$) can be arbitrary elements of $\{0, \dots, b_\ell - 1\}$, (respectively, $\{0, \dots, b_{\ell'} - 1\}$). By Lemma 2.2, we can choose a_1, a'_1 such that $n_1 \equiv a_1 + a'_1 \pmod{b_1}$ in at most M ways. Given a choice of a_1, a'_1 , there are at most M pairs a_2, a'_2 with $n_2 - c_1 \equiv a_2 + a'_2 \pmod{b_2}$, where $c_1 = \lfloor \frac{a_1 + a'_1}{b_1} \rfloor \in \{0, 1\}$ is the carry. Continuing in this fashion for $j = 1, \dots, \ell - 1$, we get that there are at most $M^{\ell-1}$ ways to fix the first $\ell - 1$ digits $a_1, \dots, a_{\ell-1}$ and $a'_1, \dots, a'_{\ell-1}$. We can fix a_ℓ and a'_ℓ in at most b_ℓ ways. Given this choice, the digits $a'_{\ell+1}, \dots, a'_{\ell'}$ are uniquely determined. Putting this together, we obtain the following upper bound on the number of representations $n = a + a'$:

$$\sigma_A(n) \leq 2 \sum_{\ell=1}^k b_\ell M^{\ell-1} \leq 2b_k M^k \leq 8f(k)^2 M^k, \quad (2.2)$$

where we used $b_k = p_k^2 \leq (2f(k))^2$ and $M \geq 2$. On the other hand, $b_j = p_j^2 \geq f(j)^2$ and so

$$n \geq b_1 \dots b_{k-1} \geq b_{\lfloor k/2 \rfloor} \dots b_{k-1} \geq f(\lfloor k/2 \rfloor)^{k/2}. \quad (2.3)$$

Hence, $k \leq \frac{2 \log n}{\log f(\lfloor k/2 \rfloor)}$ and substituting this in Eq. (2.2), we obtain the bound

$$\sigma_A(n) \leq 4f(k)^2 n^{c/\log f(\lfloor k/2 \rfloor)}.$$

Note that the right hand side is $n^{o(1)}$ for any sufficiently slowly growing function f . Owing to the computational considerations in the next paragraph, we take $f(k) = k$ which leads to $k \lesssim \frac{\log n}{\log \log n}$ and $\sigma_A(n) \lesssim n^{c/\log \log n}$.

Finally, we quickly verify that testing membership $a \in A$ can be done in time $O((\log a)^{O(1)})$. Indeed, given $a \in \mathbb{N}$, we can compute all primes p_k for $k \leq c \log a$ in time $(\log a)^{O(1)}$ (Lemma 2.3), compute the base $\mathbf{b}$ expansion $a = \overline{a_k \dots a_1}$ in time $(\log a)^{O(1)}$, and check that $a_j \in A_j$ for $j = 1, \dots, k - 1$ in time $O(k(\log f(k))^{O(1)})$ using Lemma 2.2. $\square$

2.1 Modular construction and computational details

We record the proof of Lemma 2.2, following Ruzsa [18]. For $n \in \mathbb{Z}, p \in \mathbb{N}$ we write $(n \bmod p)$ for the unique $n' \in \{0, 1, \dots, p - 1\}$ congruent to n modulo p . The following is exactly [18, Lemma 3.1].

Lemma 2.4. *Let $p \equiv 3, 5 \pmod{8}$. Define $B_p \subseteq \{0, \dots, 2p^2\}$ by*

$$\begin{aligned} B_p = & \{x + 2p(3x^2 \bmod p) : x \in \{0, \dots, (p-1)\}\} \\ & \cup \{x + 2p(4x^2 \bmod p) : x \in \{0, \dots, (p-1)\}\} \\ & \cup \{x + 2p(6x^2 \bmod p) : x \in \{0, \dots, (p-1)\}\}. \end{aligned}$$

We have $\sup_{n \in \mathbb{Z}} \sigma_{B_p}(n) \leq 18$ and furthermore, for all $0 \leq n < p^2$, at least one of the six numbers

$$n - p, n, n + p, n + p^2 - p, n + p^2, n + p^2 + p$$

appears in the set $B_p + B_p$.

Given Lemma 2.4, we prove Lemma 2.2.

Proof of Lemma 2.2. Let $B'_p = B_p + \{-p, 0, p\}$ (viewed as a subset of $\mathbb{Z}$) and set

$$A_p = \{x \bmod p^2 : x \in B'_p\} \subseteq \mathbb{Z}/(p^2\mathbb{Z}).$$

Applying Lemma 2.4, we immediately have:

- $B'_p + B'_p \subseteq [-2p, 5p^2]$
- For all $0 \leq n < p^2$, one of n or $n + p^2$ appears in $B'_p + B'_p$
- We have that $\sup_{n \in \mathbb{Z}} \sigma_{B'_p}(n) \leq 9 \cdot 18 = 162$.

Noting that $n \equiv n + p^2 \bmod p^2$, it follows that $A_p + A_p = \mathbb{Z}/(p^2\mathbb{Z})$. Furthermore we have that $\sup_{n \in \mathbb{Z}} \sigma_{A_p}(n) \leq 6 \cdot 9 \cdot 18 = 594$; this is immediate as $B'_p + B'_p \subseteq [-2p, 5p^2]$ and there are at most 6 representatives in this interval for a given residue modulo p^2 .

We now discuss the time complexity of testing membership in A_p . Given p , and $x \in \mathbb{Z}/(p^2\mathbb{Z})$, we consider the unique representative $x' \in \{0, \dots, p^2 - 1\}$. Noting that $B'_p \subseteq [-p, 2p^2 + p]$, it suffices by construction to test whether at least one of $x' - p^2, x', x' + p^2, x' + 2p^2$ is in B'_p . This is equivalent to checking whether at least one of $x' + \{-p^2, 0, p^2, 2p^2\} + \{-p, 0, p\}$ is in B_p ; in particular, one of at most 12 distinct given elements is in B_p .

To test whether $y \in \{0, \dots, 2p^2\}$ is in B_p amounts to testing whether $y = z + 2p(3z^2 \bmod p)$, $y = z + 2p(4z^2 \bmod p)$, or $y = z + 2p(6z^2 \bmod p)$ for an integer $z \in \{0, \dots, p - 1\}$. Given y , the 'candidate' z is forced to be the unique number in $\{0, \dots, p - 1\}$ equivalent to $y \bmod p$ and we can then simply compute $(3z^2 \bmod p)$, $(4z^2 \bmod p)$, and $(6z^2 \bmod p)$. This procedure clearly takes time $O((\log p)^{O(1)})$. $\square$

2.2 Assuming deterministic polynomial time algorithms for locating primes

For the remainder of this section we will operate under the following assumption.

Assumption 2.5. There exists a deterministic algorithm which outputs the least prime which is $3 \bmod 8$ in the interval $[N, 2N]$ in time $O((\log N)^{O(1)})$.

To obtain a better upper bound on $\sigma_A(n)$, we take the function f in the proof of Theorem 1.1 to be $f(k) = \exp(ck)$. It follows from (2.2) and (2.3) that $\sigma_A(n) \lesssim \exp(Ck)$ and $n \gtrsim \exp(ck^2)$, thus giving the bound $\sigma_A(n) \lesssim \exp(C\sqrt{\log n})$. To test membership, we need to construct primes p of order at most $\exp(ck) \approx \exp(c\sqrt{\log n})$ which can be done in $(\log n)^{O(1)}$ -time under Assumption 2.5.

Acknowledgements

V.J. is supported by NSF CAREER award DMS-2237646. H.P. is supported by a Clay Research Fellowship and a Stanford Science Fellowship. M.S. is supported by NSF Graduate Research Fellowship Programme DGE-2141064. D.Z. is supported by the Jane Street Graduate Fellowship. We thank Zach Hunter and Sándor Kiss for carefully reading the manuscript and suggesting improvements and references.

References

- [1] Agrawal, M., Kayal, N. and Saxena, N. (2004) PRIMES is in P. *Ann. Math. (2)* **160** 781–793.
- [2] Barak, B., Rao, A., Shaltiel, R. and Wigderson, A. (2012) 2-source dispersers for $n^{o(1)}$ entropy, and Ramsey graphs beating the Frankl–Wilson construction. *Ann. Math. (2)* **176** 1483–1543.
- [3] Chen, Y.-G. (2008) *The analogue of Erdős–Turán conjecture in $\mathbb{Z}_m$* . *J. Numb. Theory*, **128** 2573–2581.
- [4] Cilleruelo, J., Kiss, S. Z., Ruzsa, I. Z. and Vinuesa, C. (2010) Generalization of a theorem of Erdős and Rényi on Sidon sequences. *Rand. Struct. Algorithms* **37** 455–464.
- [5] Cohen, G. (2021) Two-source dispersers for polylogarithmic entropy and improved Ramsey graphs. *SIAM J. Comput.* **50** STOC16–30–STOC16–67.
- [6] Erdős, P. and Turán, P. (1941) On a problem of Sidon in additive number theory, and on some related problems. *J. Lond. Math. Soc.* **16** 212–215.
- [7] Erdős, P. (1954) On a problem of Sidon in additive number theory. *Acta Sci. Math. (Szeged)* **15** 255–259.
- [8] Erdős, P. (1995) Some of my favourite problems in number theory, combinatorics, and geometry. *Combin. Week (Portuguese) (São Paulo, 1994)* **2** 165–186.
- [9] Erdős, P. (1997) Some of my favorite problems and results, The mathematics of Paul Erdős, I, *Algorithms Combin.*, vol. **13**. Berlin: Springer, pp. 47–67.
- [10] Erdős, P. and Graham, R. L. (1980) Old and new problems and results in combinatorial number theory, *Monographies de L'Enseignement Mathématique [Monographs of L'Enseignement Mathématique]*, **28**. Geneva: L'Enseignement Mathématique, Université de Genève.
- [11] Erdős, P. and Tetali, P. (1990) Representations of integers as the sum of k terms. *Rand. Struct. Algorithms* **1** 245–261.
- [12] Kolountzakis, M. N. (1995) An effective additive basis for the integers. *Discrete Math* **145** 307–313.
- [13] Koukoulopoulos, D. (2019) The distribution of prime numbers, *Graduate Studies in Mathematics*, vol. **203**. Providence: American Mathematical Society.
- [14] Lagarias, J. C. and Odlyzko, A. M. (1987) Computing $\pi(x)$: an analytic method. *J. Algorithms* **8** 173–191.
- [15] Li, X. (2016) Improved two-source extractors, and affine extractors for polylogarithmic entropy. In *57th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2016*. Los Alamitos: IEEE Computer Society, pp. 168–177.
- [16] Li, X. (2023) Two source extractors for asymptotically optimal entropy, and (many) more. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science—FOCS. 2023*. Los Alamitos: IEEE Computer Society, pp. 1271–1281.
- [17] Pilatte, C. (2024) A solution to the Erdős–Sárközy–Sós problem on asymptotic Sidon bases of order 3. *Compositio Mathematica* **160** 1418–1432.
- [18] Ruzsa, I. Z. (1990) A just basis. *Monatsh. Math.* **109** 145–151.
- [19] Ruzsa, I. Z. (1998) An infinite Sidon sequence. *J. Numb. Theory* **68** 63–71.
- [20] Sidon, S. (1932) Ein satz über trigonometrische polynome und seine anwendung in der theorie der Fourier-Reihen. *Math. Ann.* **106** 536–539.
- [21] Van, H. V. U. (2000) On a refinement of Waring's problem. *Duke Math. J.* **105** 107–134.