RECIPROCAL SUMS OVER SEPARATED PAIRS AND TRIPLES: ERDŐS PROBLEM #289

PINGYOU LTD

ABSTRACT. Every positive rational number is a sum of reciprocals over exactly k intervals of positive integers of length two or three, for every sufficiently large integer k; the intervals may be placed beyond any prescribed bound and with any prescribed minimum separation. For the number 1 this answers a question of Erdős and Graham. For fixed rational, bound and separation there are at least $2^{k/21}$ such representations at every sufficiently large k, with all denominators $O(k^{2-\delta})$ for some $\delta > 0$, and a representation can be chosen with $o(k)$ intervals of length three. The proof cancels the leading prime-power residue of the reciprocal sum by pairs of consecutive integers supplied by primitive points on modular hyperbolas and restricted sumsets, completes the remaining rational by an elementary construction on a fixed family of reserved pairs, and changes the number of intervals by independent sum-preserving splittings of cells with powersmooth denominators. One correction and one completion serve every choice of splittings, and a comparison of consecutive scales makes the attainable counts consecutive.

1. INTRODUCTION

For integers $1 \le a \le b$ write

$$H[a,b] = \sum_{n=a}^{b} \frac{1}{n}.$$

Erdős and Graham [9, p. 34] ask whether, for every sufficiently large integer k, there are finite intervals $I_1, \dots, I_k$ of positive integers with

$$1 = \sum_{i=1}^{k} \sum_{n \in I_i} \frac{1}{n}, \tag{1}$$

and they reproduce the representation

$$2 = H[2,7] + H[9,10] + H[17,18] + H[34,35] + H[84,85] \tag{2}$$

found by Montgomery and by Hickerson in answer to a problem of L.-S. Hahn in the *American Mathematical Monthly* (Problem E2689). The intervals in (2) are pairwise disjoint and no two of them are adjacent, where two disjoint intervals $[a,b]$ and $[c,d]$ of integers with $b < c$ are *adjacent* if $c = b + 1$. This is the reading of the question adopted in Bloom's database [2], where it is Problem 289: the intervals are to be distinct, pairwise disjoint and pairwise nonadjacent, each of length at least two. As Kovač observed in the discussion recorded there, the question is easy if intervals may be repeated or may overlap, because repeated denominators absorb all congruence conditions; distinct denominators must satisfy them exactly (Lemma 3.2). Related results prescribe other properties of the denominators of an Egyptian fraction: Croot's theorem on denominators in short intervals [5], his proof of the colouring conjecture of [9] [6], and Bloom's density theorem [3]; see [4] for a survey and [12] for recent work on several further questions of Erdős on unit fractions.

Two intervals $I < I'$ of integers are *separated by at least g integers* if $\min I' - \max I \ge g + 1$; for $g = 1$ this says that I and I' are disjoint and nonadjacent.

Theorem A. Let r be a positive rational number and let $B \ge 0$ and $g \ge 1$ be integers. There are constants $A > 0$ and $0 < \delta < 1/5$, and an integer $k_0 = k_0(r, B, g)$, with the following properties for every integer $k \ge k_0$.

(i) There are at least $2^{k/21}$ distinct families of k intervals $I_1, \dots, I_k$ of integers greater than B, each of length 2 or 3, any two of which are separated by at least g integers, such that

$$\sum_{i=1}^{k} \sum_{n \in I_i} \frac{1}{n} = r \qquad \text{and} \qquad \max \bigcup_{i=1}^{k} I_i \le Ak^{2-\delta}. \tag{3}$$

(ii) There is a family with the same sum, separation and bound in which at most $Ak^{1-\delta/4}$ of the intervals have length 3.

Families are counted without an ordering of their intervals.

Corollary B. Erdős Problem 289 has an affirmative answer: for every sufficiently large integer k there are pairwise disjoint and pairwise nonadjacent intervals $I_1, \dots, I_k$ of positive integers, each of length two or three, satisfying (1).

The inputs to the proof are two estimates for complete exponential sums, the Weil–Estermann bound for Kloosterman sums and a bound of Shparlinski for sums with inverse squares, the restricted-sumset theorem of Dias da Silva and Hamidoune, and elementary prime-number estimates; the argument for the count of primitive points on modular hyperbolas follows Shparlinski [15] and is reproduced in full. The proof is constructive once the thresholds in these inputs are made explicit, but it gives no usable value of $k_0(r, B, g)$, and the exponents $1/21$ and δ are not optimised (Remark 9.4). Assertion (ii) is obtained from a different choice than the families of assertion (i); the theorem does not assert exponential multiplicity among representations with few intervals of length 3.

1.1. **Legal sets and the three operations.** A finite set S of positive integers is *legal* if each maximal run of consecutive elements of S has at least two elements. The runs are the *components* of S, $\kappa(S)$ is their number, and $\rho(S) = \sum_{n \in S} 1/n$ is the *reciprocal sum* of S. Legal sets with k components correspond bijectively to families of k pairwise disjoint, pairwise nonadjacent intervals of length at least two, and Theorem A states that for every $k \ge k_0$ there is a legal set S with

$$\rho(S) = r, \qquad \kappa(S) = k, \qquad \min S > B,$$

all of whose components have length 2 or 3, any two of them separated by at least g integers. For a positive rational x let $\lambda(x)$ be the largest exact prime-power divisor of the denominator of x. The proof combines three exact operations on legal sets.

(i) *Descent* (Theorem 5.1). Fix a level L and put $K = \operatorname{lcm}(1, \dots, L)$. Every positive rational x exceeding a quantity $E(L)$, with $E(L) \to 0$ as $L \to \infty$, can be written as $x = x' + \sum_{\mathcal{P} \in \mathcal{C}} \rho(\mathcal{P})$, where $Kx' \in \mathbb{Z}$, $0 \le x - x' < E(L)$, and $\mathcal{C}$ is a set of pairwise disjoint, pairwise nonadjacent pairs of consecutive integers with $|\mathcal{C}| \le H(\lambda(x))$, where $H(Y) \ll_\nu Y^{1+\nu}$ for every $\nu > 0$. The pairs have the form $\{Qm, Qm+1\}$ with $Q = p^e$ the largest prime-power divisor of Qm, all prime-power divisors of $Qm+1$ below Q, and $Q\big(1/(Qm) + 1/(Qm+1)\big) \equiv m^{-1} \pmod p$; subtracting pairs of prescribed residue lowers λ (Lemma 3.3), and pairs of every residue exist by Shparlinski's asymptotic formula for primitive points on modular hyperbolas together with the theorems of Dias da Silva–Hamidoune and Cauchy–Davenport (Proposition 4.7).

(ii) *Completion* (Lemma 6.2). Reserved pairs $\{Kd+1, Kd+2\}$, $d \in U$, are fixed once and for all, where U is the union of J disjoint finite sets $\mathcal{D}_j$ with $\sum_{d \in \mathcal{D}_j} 1/d = 1$; such sets exist beyond any bound by an elementary greedy argument (Lemma 6.1). Adjoining Kd to a reserved pair creates a triple and no new component, and $\sum_{d \in \mathcal{D}_j} 1/(Kd) = 1/K$; so every x' with $Kx' \in \mathbb{Z}$ and $0 < x' \le J/K$ is absorbed exactly.

(iii) *Splitting* (Corollary 7.4). For $n = Kd + 8t$, where t is a fixed integer depending on g, the cell consisting of the five triples $[a, a+2]$ with $a \in \{10n, 15n, 12n+2, 18n+3, 36n+6\}$ has five

components; deleting the first element of each triple and adjoining the pair $\{6n, 6n+1\}$ leaves the reciprocal sum unchanged, since

$$\frac{1}{10n}+\frac{1}{15n}=\frac{1}{6n}, \qquad \frac{1}{12n+2}+\frac{1}{18n+3}+\frac{1}{36n+6}=\frac{1}{6n+1},$$

and produces six pairs. All elements of all cells in both states, of all reserved pairs and triples, and of all correction pairs lie in nine residue bands modulo $R = 210t$, each band a run of two or three consecutive residues occupied by one kind of component, any two bands separated by at least g residues (Fig. 2); so any subset of the cells may be split independently of everything else, and any two components of the final set are separated by at least g integers.

At a scale T, among the cells with $T \le d \le CT$ those whose denominators are $T^{1-\varepsilon}$-powersmooth number $m_T \asymp T$, and a prefix of them has reciprocal sum within $O(1/T)$ of a fixed target. The descent of the resulting rational uses $j_T = O(T^{1-\varepsilon/2})$ pairs, and one descent and one completion serve all 2^{m_T} choices of split cells: this gives an injection Ψ_T of $\{0,1\}^{m_T}$ into the legal sets with reciprocal sum r, the set $\Psi_T(\sigma)$ having $C_0 + 5m_T + j_T + |\sigma|$ components (Theorem 8.4). The admissible families at scales T and $T+1$ differ in $O(T^{\varepsilon})$ cells, so $m_{T+1} - m_T = o(T)$ (Lemma 9.2), and an elementary counting argument (Proposition 2.1) then yields every sufficiently large count with exponential multiplicity.

1.2. **Organisation and notation.** Section 2 contains the counting argument, Section 3 introduces leading residues, Sections 4 and 5 contain the descent, Sections 6 and 7 the completion and the splitting, Section 8 the construction at a fixed scale, and Section 9 the comparison of consecutive scales and the proof of Theorem A. Appendix A records exact computations for small numbers of intervals.

For a positive integer n let $P(n)$ denote the largest prime power p^a with $p^a \mid n$, and put $P(1) = 1$; thus $P(n)$ is the largest *exact* prime-power divisor of n, and n is Y*-powersmooth* if $P(n) \le Y$. For a prime p, v_p is the p-adic valuation on $\mathbb{Q}$ and $\mathbb{Z}_{(p)}$ is the ring of rationals with denominator prime to p, so that $\mathbb{Z}_{(p)}/p\mathbb{Z}_{(p)} = \mathbb{F}_p$. For an integer $M \ge 1$, $\varphi(M)$ is Euler's function, $\tau(M)$ the number of divisors of M, and $e(t) = \exp(2\pi i t)$. The letter p always denotes a prime and $Q = p^e$ a prime power with $e \ge 1$; sums over Q are sums over prime powers. Implied constants in $O(\cdot)$, $\ll$ and $o(\cdot)$ are absolute unless a subscript indicates the parameters they depend on, and $o(1)$ denotes a quantity tending to zero as the indicated variable tends to infinity.

2. Independent choices and component counts

The passage from independent local choices to every sufficiently large count is an elementary counting fact. It is isolated here, so that the construction of Sections 4 to 8 only has to deliver its hypotheses. For $\sigma \in \{0,1\}^m$ write $|\sigma|$ for the number of coordinates of σ equal to 1.

Proposition 2.1. *Let $q \ge 1$ be an integer. Suppose that for every sufficiently large integer T there are a set $\mathcal{X}_T$, an integer-valued function κ on $\mathcal{X}_T$, a positive integer m_T, a nonnegative integer b_T, and an injective map*

$$\Psi_T\colon \{0,1\}^{m_T} \to \mathcal{X}_T \qquad \text{with} \qquad \kappa(\Psi_T(\sigma)) = qm_T + b_T + |\sigma|,$$

and that

$$m_T \asymp T, \qquad m_{T+1} - m_T = o(T), \qquad b_T = o(T) \qquad (T \to \infty).$$

Then there is an integer k_0 such that for every integer $k \ge k_0$ there is a T with $k \ge qm_T$ for which at least $2^{k/(3(q+2))}$ distinct $\sigma \in \{0,1\}^{m_T}$ satisfy $\kappa(\Psi_T(\sigma)) = k$.

Proof. Put $a_T = qm_T + b_T$ and

$$u_T = a_T + \lceil m_T/3 \rceil, \qquad v_T = a_T + \lfloor 2m_T/3 \rfloor.$$

Then $v_T - u_T \geq m_T/3 - 2$, which is $\asymp T$, and

$$u_{T+1} - u_T = q(m_{T+1} - m_T) + (b_{T+1} - b_T) + \big(\lceil m_{T+1}/3\rceil - \lceil m_T/3\rceil\big) = o(T),$$

since $|b_{T+1} - b_T| \leq b_T + b_{T+1}$. Hence $u_{T+1} \leq v_T$ and $u_T \leq v_{T+1}$ for all sufficiently large T; fix T_* from which both inequalities hold, and put $k_0 = u_{T_*}$. Let $k \geq k_0$. Since $v_T \to \infty$ there is a least $T \geq T_*$ with $v_T \geq k$; if $T = T_*$ then $u_T \leq k \leq v_T$, and otherwise $v_{T-1} < k \leq v_T$ and $u_T \leq v_{T-1} < k$. In both cases $u_T \leq k \leq v_T$.

For this T put $s = k - a_T$, so that $\lceil m_T/3\rceil \leq s \leq \lfloor 2m_T/3\rfloor$, and put $\ell = \min\{s, m_T - s\}$, so that $m_T/3 \leq \ell \leq m_T/2$. The $\binom{m_T}{s}$ vectors σ with $|\sigma| = s$ satisfy $\kappa(\Psi_T(\sigma)) = k$, and

$$\binom{m_T}{s} = \binom{m_T}{\ell} = \prod_{i=0}^{\ell-1} \frac{m_T - i}{\ell - i} \geq \Big(\frac{m_T}{\ell}\Big)^{\ell} \geq 2^{\ell} \geq 2^{m_T/3}.$$

Finally $k \geq u_T \geq a_T \geq qm_T$, and $k \leq v_T \leq (q+1)m_T + b_T \leq (q+2)m_T$ once $b_T \leq m_T$, which holds for all large T; so $2^{m_T/3} \geq 2^{k/(3(q+2))}$. □

3. Leading residues

Definition 3.1. Let x be a positive rational number with reduced denominator D. The *level* of x is $\lambda(x) = P(D)$, the largest exact prime-power divisor of D; if x is an integer then $\lambda(x) = 1$. If $\lambda(x) = Q = p^e > 1$ and y is a rational number with $v_p(y) \geq -e$, the *residue of y at level Q* is

$$\mathrm{res}_Q(y) = Qy \bmod p \in \mathbb{F}_p,$$

the image of $Qy \in \mathbb{Z}_{(p)}$ in $\mathbb{Z}_{(p)}/p\mathbb{Z}_{(p)}$. The *leading residue* of x is $\mathrm{res}_Q(x)$ for $Q = \lambda(x)$.

Since Qx is a unit in $\mathbb{Z}_{(p)}$ when $\lambda(x) = Q$, the leading residue of x is a nonzero element of $\mathbb{F}_p$. The map $y \mapsto \mathrm{res}_Q(y)$ is additive on the rationals with $v_p(y) \geq -e$.

The first lemma is the necessary condition that any exact representation with distinct denominators must satisfy at every prime.

Lemma 3.2. *Let S be a finite set of positive integers, let p be a prime dividing some element of S, and let $q = p^e$ be the largest power of p dividing an element of S. If $v_p(\rho(S)) > -e$, and in particular if $\rho(S)$ is an integer, then*

$$\sum_{\substack{n\in S\\ q\mid n}} \Big(\frac{n}{q}\Big)^{-1} \equiv 0 \pmod{p}.$$

Proof. If $n \in S$ and $q \mid n$, then n/q is prime to p because q is the largest power of p dividing an element of S, and $q/n = (n/q)^{-1}$ in $\mathbb{Z}_{(p)}$. If $n \in S$ and $q \nmid n$, then $v_p(n) \leq e - 1$ and $v_p(q/n) \geq 1$. Hence

$$q\rho(S) \equiv \sum_{\substack{n\in S\\ q\mid n}} \Big(\frac{n}{q}\Big)^{-1} \pmod{p\mathbb{Z}_{(p)}},$$

and $v_p(q\rho(S)) \geq 1$ by hypothesis. □

The leading residue of x is cancelled by subtracting a rational of the following kind.

Lemma 3.3. *Let x be a positive rational with $\lambda(x) = Q = p^e > 1$, and let y be a rational with $v_p(y) \geq -e$ and $\mathrm{res}_Q(y) = \mathrm{res}_Q(x)$ such that every exact prime-power divisor of the reduced denominator of y, other than its p-part, is smaller than Q. Then every exact prime-power divisor of the reduced denominator of $x - y$ is smaller than Q. In particular $\lambda(x - y) < Q$ if $x \neq y$.*

Proof. Since $\operatorname{res}_Q(x-y)=0$, the number $Q(x-y)$ lies in $p\mathbb{Z}_{(p)}$ and $v_p(x-y)\ge 1-e$; so the p-part of the reduced denominator of $x-y$ divides $p^{e-1}<Q$. For a prime $\ell\neq p$, $v_\ell(x-y)\ge\min\{v_\ell(x),v_\ell(y)\}$, and the ℓ-parts of the reduced denominators of x and of y are smaller than Q, the former because $\lambda(x)=Q$ and the latter by hypothesis. □

4. Correction pairs

Throughout the rest of the paper $g\ge 1$ is the integer of Theorem A, and t is a fixed integer with

$$t\equiv 1 \pmod{35}, \qquad 2t-9\ge g; \tag{4}$$

in particular $t\ge 36$. Put $R=210t$. All correction pairs will be placed in two residue classes modulo R, and all other components of the final sets in seven further classes (Section 7); the parameter t spreads these classes far enough apart to leave g integers between any two components. Constants are allowed to depend on t, hence on g, without further mention.

4.1. **Two counts of points on modular hyperbolas.** For an integer $M\ge 2$, a unit u modulo M and real numbers $U,V\ge 0$, put

$$W_{M,u}(U,V)=\#\{(a,b)\in\mathbb{Z}^2:\ 1\le a\le U,\ 1\le b\le V,\ ab\equiv u \pmod M\},$$
$$N_{M,u}(U,V)=\#\{(a,b) \text{ counted by } W_{M,u}(U,V):\ \gcd(a,b)=1\},$$

and

$$c(M)=\frac{6}{\pi^2}\prod_{\ell\mid M}\Big(1+\frac{1}{\ell}\Big)^{-1}, \tag{5}$$

the product being over the primes dividing M. For d prime to M let $\bar d$ denote the inverse of d modulo M. The two estimates for complete exponential sums used below are

$$\max_{b\in\mathbb{Z}}\Big|\sum_{\substack{1\le d\le M\\ \gcd(d,M)=1}} e\Big(\frac{a\bar d^{\,j}+bd}{M}\Big)\Big|\le\big(M\gcd(a,M)\big)^{1/2+o(1)}, \qquad j\in\{1,2\}, \tag{6}$$

for every integer a, where $o(1)\to 0$ as $M\to\infty$ uniformly in a. For $j=1$ this is the Weil–Estermann bound for Kloosterman sums [11, Corollary 11.12]; for $j=2$ it is the special case of the bound of [14] recorded in [15, Lemma 1]. By the completion method [11, §12.2], the same bounds hold, with an additional factor $O(\log M)$, for the corresponding sums over d in any interval of at most M consecutive integers.

Lemma 4.1. *Let $M\ge 2$ and let u be a unit modulo M.*

(a) For all real U,V with $0\le U,V\le M$,

$$W_{M,u}(U,V)=\frac{\varphi(M)UV}{M^2}+O\big(M^{1/2+o(1)}\big).$$

(b) For every interval J of at most M consecutive integers and every real X with $1\le X\le M$,

$$\#\big\{d\in J:\ \gcd(d,M)=1,\ (u\bar d^{\,2} \bmod M)\in[1,X]\big\}=\frac{X}{M}\,\#\{d\in J:\gcd(d,M)=1\}+O\big(M^{1/2+o(1)}\big).$$

The implied constants and the $o(1)$ are uniform in u, U, V, J and X.

Proof. (a) If $U<1$ or $V<1$ the count is zero and the main term is $O(1)$. Otherwise, for $1\le a\le U$ with $\gcd(a,M)=1$ the condition $ab\equiv u \pmod M$ reads $b\equiv u\bar a \pmod M$, and there are no solutions when $\gcd(a,M)>1$. Detecting the congruence with additive characters modulo M,

$$W_{M,u}(U,V)=\frac{1}{M}\sum_{h=0}^{M-1}\Big(\sum_{1\le b\le V}e\Big(\frac{hb}{M}\Big)\Big)\Big(\sum_{\substack{1\le a\le U\\ \gcd(a,M)=1}}e\Big(-\frac{hu\bar a}{M}\Big)\Big).$$

The term $h=0$ equals $\lfloor V\rfloor M^{-1}\#\{a\le U:\gcd(a,M)=1\}$, and $\#\{a\le U:\gcd(a,M)=1\}=\sum_{d\mid M}\mu(d)\lfloor U/d\rfloor=\varphi(M)U/M+O(\tau(M))$; since $U,V\le M$ and $\tau(M)=M^{o(1)}$, the term $h=0$ is $\varphi(M)UV/M^2+O(M^{o(1)})$. For $0<h<M$ the inner sum over b is at most $M/\min\{h,M-h\}$ in absolute value, and the inner sum over a is bounded by $(M\gcd(hu,M))^{1/2+o(1)}=(M\gcd(h,M))^{1/2+o(1)}$, by (6) with $j=1$ and completion. Hence the terms with $h\neq 0$ contribute at most

$$M^{-1/2+o(1)}\sum_{0<h<M}\frac{M\gcd(h,M)^{1/2}}{\min\{h,M-h\}}\ll M^{1/2+o(1)}\sum_{g\mid M}g^{1/2}\sum_{\substack{1\le h\le M\\ g\mid h}}\frac{1}{h}\ll M^{1/2+o(1)}\tau(M)\log M,$$

which is $M^{1/2+o(1)}$.

(b) Detect the condition $(u\bar d^{\,2}\bmod M)\in[1,X]$ with additive characters in the same way: the term $h=0$ is $\lfloor X\rfloor M^{-1}\#\{d\in J:\gcd(d,M)=1\}$, which differs from the stated main term by at most 1, and for $0<h<M$ the sum over $x\in[1,X]$ of $e(hx/M)$ is at most $M/\min\{h,M-h\}$, while the sum of $e(-hu\bar d^{\,2}/M)$ over $d\in J$ prime to M is bounded by $(M\gcd(h,M))^{1/2+o(1)}$, by (6) with $j=2$ and completion. The same summation over h gives the error term. □

Lemma 4.2. *For every integer $M\ge 2$, every unit u modulo M and all real U,V with $1\le U,V<M$,*

$$N_{M,u}(U,V)=c(M)\frac{UV}{M}+O\big(M^{3/4+o(1)}\big),$$

where the implied constant and the $o(1)$ are uniform in u, U and V.

Proof. For $u=1$ this is [15, Theorem 3]; the argument is reproduced with a unit right-hand side and fixed truncation parameters. For $d\ge 1$ let M_d be the number of pairs (a,b) counted by $W_{M,u}(U,V)$ with $d\mid a$ and $d\mid b$. If $\gcd(d,M)>1$ then $M_d=0$, because d^2 divides a number congruent to the unit u modulo M. If $\gcd(d,M)=1$, writing $a=ds$ and $b=dt$ shows that M_d is the number of $(s,t)\in[1,U/d]\times[1,V/d]$ with $st\equiv u\bar d^{\,2}\pmod M$, and $u\bar d^{\,2}$ is a unit. By Möbius inversion, $N_{M,u}(U,V)=\sum_{d\ge1}\mu(d)M_d$.

Three bounds for M_d are used. First, by Lemma 4.1(a),

$$M_d=\frac{\varphi(M)UV}{d^2M^2}+O\big(M^{1/2+o(1)}\big).\tag{7}$$

Second, the product $st\le UV/d^2$ lies in one residue class modulo M, so it takes at most $UV/(d^2M)+1$ values, and each value has $M^{o(1)}$ ordered factorisations; hence

$$M_d\le\Big(\frac{UV}{d^2M}+1\Big)M^{o(1)}.\tag{8}$$

Third, let $\Delta\ge M^{1/2}$. For $d\in[\Delta,2\Delta)$ prime to M with $M_d>0$, the residue $u\bar d^{\,2}$ has a representative st in $[1,UV/\Delta^2]$, and $UV/\Delta^2\le M$ because $UV<M^2$; there are no such d if $UV/\Delta^2<1$, and otherwise, by Lemma 4.1(b) applied to $J=[\Delta,2\Delta)$ and $X=UV/\Delta^2$, their number is at most $UV/(\Delta M)+O(M^{1/2+o(1)})$. Since $UV/d^2\le M$, each such d has a unique admissible product st, with $M^{o(1)}$ factorisations. Summing over the dyadic ranges with $\Delta\ge D_2$, of which there are $O(\log M)$ below M, gives, for $D_2\ge M^{1/2}$,

$$\sum_{d>D_2}M_d\ll\Big(\frac{UV}{D_2M}+M^{1/2}\Big)M^{o(1)}.\tag{9}$$

Put $D_1=\lceil M^{1/4}\rceil$ and $D_2=\lceil M^{1/2}\rceil$, and use (7) for $d\le D_1$, (8) for $D_1<d\le D_2$ and (9) beyond. Since $UV<M^2$, the total error is

$$O\Big(D_1M^{1/2+o(1)}+\Big(\frac{UV}{D_1M}+D_2+\frac{UV}{D_2M}+M^{1/2}\Big)M^{o(1)}\Big)=O\big(M^{3/4+o(1)}\big),$$

and extending the main term

$$\sum_{\substack{d\le D_1\\ \gcd(d,M)=1}} \mu(d)\frac{\varphi(M)UV}{d^2M^2}$$

to all d costs $O(UV/(MD_1)) = O(M^{3/4})$. Finally

$$\frac{\varphi(M)}{M}\sum_{\substack{d\ge 1\\ \gcd(d,M)=1}}\frac{\mu(d)}{d^2} = \prod_{\ell\mid M}\Big(1-\frac{1}{\ell}\Big)\cdot\frac{6}{\pi^2}\prod_{\ell\mid M}\Big(1-\frac{1}{\ell^2}\Big)^{-1} = c(M).$$

The uniformity in u holds because every exponential sum used carries the unit u in its coefficient, and $\gcd(hu, M) = \gcd(h, M)$. □

Both lemmas transfer to boxes $I \times I$ with $I = (A, A'] \cap \mathbb{Z}$ and $0 \le A < A' < M$ by inclusion and exclusion, with the same error terms.

4.2. **Admissible parameters.**

Proposition 4.3. *There is a constant $Q_0 \ge R$ such that for every prime power $Q = p^e > Q_0$ there is a set $\mathcal{M}_Q$ of integers with $|\mathcal{M}_Q| \ge Q^{1-o(1)}$, the $o(1)$ tending to zero as $Q \to \infty$ through prime powers, such that every $m \in \mathcal{M}_Q$ satisfies*

$$\frac{Q}{5} < m < Q, \qquad p \nmid m, \qquad Qm+1 = ab \ \text{ for some integers } a, b < Q \text{ with } \gcd(a,b) = 1, \tag{10}$$

and

$$Qm \equiv \begin{cases} 42t \pmod R, & p \ne 5,\\ 60t \pmod R, & p = 5. \end{cases} \tag{11}$$

Proof. Let $I_Q = \{a \in \mathbb{Z} :\ Q/2 < a < Q\}$, so that $\ell_Q := |I_Q| = Q/2 + O(1)$, and consider the pairs $(a, b) \in I_Q^2$ with $\gcd(a, b) = 1$ and

$$ab \equiv 1 \pmod Q, \qquad ab \equiv \begin{cases} 42t+1 \pmod R, & p \ne 5,\\ 60t+1 \pmod R, & p = 5. \end{cases} \tag{12}$$

The right-hand sides are units modulo $R = 2\cdot 3\cdot 5\cdot 7\cdot t$: each is 1 modulo every prime dividing t; $42t+1$ is 1 modulo 2, 3 and 7 and is $2t+1 \equiv 3$ modulo 5; and $60t+1$ is 1 modulo 2, 3 and 5 and is $4t+1 \equiv 5$ modulo 7, by (4). The two congruences in (12) are compatible whenever $Q > R$: if $p \mid R$ then $e > v_p(R)$, and $p^{v_p(R)}$ divides $42t$ when $p \ne 5$ (for $p \in \{2, 3, 7\}$ because $v_p(42t) = v_p(210t)$, and for the remaining primes $p \mid t$ because then $v_p(42t) = v_p(t) = v_p(210t)$) and divides $60t$ when $p = 5$ (because $5 \nmid t$), so both congruences require $ab \equiv 1$ modulo $p^{v_p(R)}$. Hence, for $Q > R$, (12) is a single congruence $ab \equiv u \pmod M$ with

$$M = \operatorname{lcm}(Q, R), \qquad Q < M \le RQ,$$

and a unit u modulo M. The primes dividing M divide Rp, so $c(M) \ge c_0$ for a constant $c_0 > 0$ depending only on t. By Lemma 4.2, transferred to the box $I_Q \times I_Q$ (which is legitimate because $Q - 1 < M$), the number of coprime pairs $(a, b) \in I_Q^2$ satisfying (12) is

$$c(M)\frac{\ell_Q^2}{M} + O\big(Q^{3/4+o(1)}\big), \qquad \text{and} \qquad c(M)\frac{\ell_Q^2}{M} \ge \frac{c_0}{5R}\,Q \tag{13}$$

for all sufficiently large Q, since $\ell_Q \ge Q/2 - 1$ and $M \le RQ$; the error term $O(M^{3/4+o(1)})$ of Lemma 4.2 is $O(Q^{3/4+o(1)})$ because $M \le RQ$.

Among these pairs we discard those with $v_p(ab-1) > e$. If $e = 1$ there are none, because $0 < ab - 1 < Q^2 = p^2$ and $p \mid ab - 1$. If $e \ge 2$ then $p \le Q^{1/2}$, and a discarded pair satisfies $ab \equiv 1 \pmod{pQ}$ together with the second congruence in (12). Since $e \ge v_p(R)$, the modulus $\operatorname{lcm}(pQ, R)$

equals pM, and the right-hand side is again a unit, by the same compatibility argument. By Lemma 4.1(a), applied to the modulus pM and the box $I_Q \times I_Q$ without the coprimality condition, the number of discarded pairs is at most

$$\frac{\varphi(pM)}{(pM)^2}\ell_Q^2 + O\big((pM)^{1/2+o(1)}\big) = \frac{\varphi(M)}{pM^2}\ell_Q^2 + O\big(Q^{3/4+o(1)}\big), \tag{14}$$

using $\varphi(pM) = p\varphi(M)$ and $pM \le RQ^{3/2}$. The ratio of the main term in (14) to the main term in (13) is

$$\frac{1}{p}\cdot\frac{\varphi(M)/M}{c(M)} = \frac{\pi^2}{6p}\prod_{\ell\mid M}\Big(1-\frac{1}{\ell^2}\Big) \le \frac{\pi^2}{12} < 1.$$

Consequently, for all sufficiently large Q, at least $c_1 Q$ coprime pairs $(a,b) \in I_Q^2$ satisfy (12) and $v_p(ab-1) = e$, where $c_1 > 0$ depends only on t.

For each such pair put $m = (ab-1)/Q$. Then $Qm+1 = ab$ with $a, b < Q$ coprime; $p \nmid m$ because $v_p(ab-1) = e$; $m < Q$ because $ab < Q^2$; and $m > Q/4 - 1/Q \ge Q/5$ for $Q \ge 5$. Reducing $ab-1$ modulo R gives (11). A given m arises from at most $\tau(Qm+1)$ ordered pairs (a,b), and $\tau(n) = n^{o(1)}$ with $Qm+1 < Q^2$; so the set $\mathcal{M}_Q$ of values of m has at least $c_1 Q \cdot Q^{-o(1)}$ elements. □

Remark 4.4. The estimate of Lemma 4.2 is nontrivial only when UV exceeds $M^{3/2+\nu}$ for some $\nu > 0$, as noted in [15]; in the proof above $UV \asymp Q^2 \ge M^2/R^2$. The coprimality of a and b is what keeps every prime-power divisor of $Qm+1$ below Q; the discarded pairs were counted without it, which is why Lemma 4.1(a) suffices there. For hyperbolas $ab \equiv u \pmod M$ with a general unit u see also [15, §4] and [16].

4.3. **Correction pairs.**

Definition 4.5. Let $Q = p^e > Q_0$ and $m \in \mathcal{M}_Q$. The *correction pair at level Q with parameter m* is

$$\mathcal{P}_{Q,m} = \{Qm,\ Qm+1\}.$$

Its *centre* is the even integer Qm.

Lemma 4.6. *Let $Q = p^e > Q_0$ and $m \in \mathcal{M}_Q$.*

(a) $P(Qm) = Q$ and $P(Qm+1) < Q$.
(b) $v_p(\rho(\mathcal{P}_{Q,m})) = -e$ and $\operatorname{res}_Q(\rho(\mathcal{P}_{Q,m})) = m^{-1} \bmod p$.
(c) $\rho(\mathcal{P}_{Q,m}) < 10/Q^2$.
(d) Every element of $\mathcal{P}_{Q,m}$ is congruent to $42t$, $42t+1$, $60t$ or $60t+1$ modulo R, and exceeds $Q^2/5$.
(e) If $\mathcal{P}_{Q,m} \ne \mathcal{P}_{Q',m'}$ (with $Q' > Q_0$, $m' \in \mathcal{M}_{Q'}$), then the two pairs are disjoint and their centres differ by at least $18t$; in particular the pairs are separated by at least $18t - 2 \ge g$ integers.

Proof. (a) Since $p \nmid m$, the exact p-part of Qm is Q, and every other exact prime-power divisor of Qm divides $m < Q$. Write $Qm+1 = ab$ with $\gcd(a,b) = 1$ and $a, b < Q$; every prime power dividing ab divides exactly one of a and b and is therefore smaller than Q.

(b) $v_p(1/(Qm)) = -e$ and $v_p(1/(Qm+1)) = 0$, whence $v_p(\rho(\mathcal{P}_{Q,m})) = -e$ and

$$Q\rho(\mathcal{P}_{Q,m}) = \frac{1}{m} + \frac{Q}{Qm+1} \equiv m^{-1} \pmod{p\mathbb{Z}_{(p)}},$$

because $Qm+1 \equiv 1 \pmod p$ is a unit and $p \mid Q$.

(c) $\rho(\mathcal{P}_{Q,m}) < 2/(Qm) < 10/Q^2$ since $m > Q/5$.

(d) The centre is $42t$ or $60t$ modulo R by (11), and $Qm > Q^2/5$.

(e) The centre Qm is even, so the pairs are disjoint unless they have the same centre; by (a) the centre determines $Q = P(Qm)$ and then m. Two distinct integers each congruent to $42t$ or $60t$

modulo R differ by a number congruent to 0 or $\pm 18t$ modulo $R = 210t$, hence by at least $18t$; and $18t - 2 \geq 2t - 9 \geq g$ by (4). $\square$

4.4. **Every residue at small cost.** For $A \subseteq \mathbb{F}_p$ and an integer $h \geq 1$ let $h^{\wedge}A$ denote the set of sums of h distinct elements of A. The two facts needed are the theorem of Dias da Silva and Hamidoune [8] (see also [1]),

$$|h^{\wedge}A| \geq \min\{p,\ h(|A| - h) + 1\}, \tag{15}$$

and the Cauchy–Davenport theorem [7], $|A + B| \geq \min\{p, |A| + |B| - 1\}$ for nonempty $A, B \subseteq \mathbb{F}_p$; both are proved in [13, Chapters 2 and 3].

For a prime power $Q = p^e > Q_0$ put

$$N(Q) = \begin{cases} \lceil 2Q/|\mathcal{M}_Q| \rceil, & e = 1, \\ p - 1, & e \geq 2. \end{cases} \tag{16}$$

By Proposition 4.3, $N(Q) \leq Q^{o(1)}$ when Q is prime, and $N(Q) \leq Q^{1/2}$ when Q is a proper prime power.

Proposition 4.7. *There is a constant $Q_1 \geq Q_0$ such that for every prime power $Q = p^e > Q_1$ and every $c \in \mathbb{F}_p$ there is a subset $\mathcal{S} \subseteq \mathcal{M}_Q$ with*

$$\sum_{m \in \mathcal{S}} m^{-1} \equiv c \pmod{p}, \qquad |\mathcal{S}| \leq N(Q) \leq Q^{1/2}, \qquad \sum_{m \in \mathcal{S}} \rho(\mathcal{P}_{Q,m}) < 10\, Q^{-3/2}.$$

Proof. Suppose first that $e = 1$, so $Q = p$. Distinct $m \in \mathcal{M}_p \subset (0, p)$ have distinct inverses modulo p, so $A = \{m^{-1} \bmod p : m \in \mathcal{M}_p\}$ has $s := |A| = |\mathcal{M}_p| \geq p^{1-o(1)}$ elements, and $h := N(p) = \lceil 2p/s \rceil$ satisfies $h \leq s/2$ once p is large. Then $h(s - h) \geq hs/2 \geq p$, so $h^{\wedge}A = \mathbb{F}_p$ by (15): every $c \in \mathbb{F}_p$ is the sum of exactly h distinct elements of A, that is, of the inverses of h distinct elements of $\mathcal{M}_p$.

Suppose next that $e \geq 2$. Then $|\mathcal{M}_Q| \geq Q^{1-o(1)} \geq p - 1$ for Q large, and the inverses $c_1, \dots, c_{p-1}$ of $p - 1$ distinct elements of $\mathcal{M}_Q$ are nonzero elements of $\mathbb{F}_p$, possibly with repetitions. The set of subset sums of $\{c_1, \dots, c_i\}$ is $\{0, c_1\} + \cdots + \{0, c_i\}$, and by the Cauchy–Davenport theorem it has at least $\min\{p, i + 1\}$ elements; for $i = p - 1$ it is all of $\mathbb{F}_p$. So every $c \in \mathbb{F}_p$ is a sum of at most $p - 1$ of the c_i, with distinct indices.

In both cases $|\mathcal{S}| \leq N(Q)$, and $N(Q) \leq Q^{1/2}$ for Q large, since $N(Q) \leq Q^{o(1)}$ for primes and $p - 1 < Q^{1/2}$ for proper powers. The mass bound follows from Lemma 4.6(c). $\square$

Corollary 4.8. *For $L \geq Q_1$ and $Y \geq L$ put*

$$E(L) = 10 \sum_{Q > L} Q^{-3/2}, \qquad H(Y) = \sum_{Q_1 < Q \leq Y} N(Q),$$

both sums over prime powers. Then $E(L) \leq 20 L^{-1/2}$, and for every $\nu > 0$

$$H(Y) \ll_{\nu} Y^{1+\nu}.$$

Proof. $E(L) \leq 10 \sum_{n > L} n^{-3/2} \leq 10 \int_L^{\infty} t^{-3/2}\, dt = 20 L^{-1/2}$. For $H(Y)$, the primes $p \leq Y$ contribute at most $\sum_{p \leq Y} N(p)$, and $N(p) \leq p^{\nu}$ for all $p > p_0(\nu)$ by Proposition 4.3, so their contribution is $O_{\nu}(1) + Y \cdot Y^{\nu}$. The proper prime powers contribute

$$\sum_{e \geq 2} \sum_{p^e \leq Y} (p - 1) \leq \sum_{n \leq Y^{1/2}} n + \sum_{3 \leq e \leq \log_2 Y} \sum_{n \leq Y^{1/e}} n = O(Y) + O\big(Y^{2/3} \log Y\big). \qquad \square$$

5. Descent

Correction pairs cancel the leading residue of a rational at any level above a fixed cutoff L, and iterating in descending order of level brings the denominator down to a divisor of $\operatorname{lcm}(1,\dots,L)$.

Theorem 5.1. *Let $L \ge Q_1$, let $K = \operatorname{lcm}(1,\dots,L)$, let $Y \ge L$, and let x be a positive rational with $\lambda(x) \le Y$ and $x > E(L)$. There is a finite set $\mathcal{C}$ of correction pairs with the following properties.*

(a) $\mathcal{C}$ is the union of batches at pairwise distinct levels Q with $L < Q \le Y$, the batch at level Q consisting of at most $N(Q)$ correction pairs at that level; in particular $|\mathcal{C}| \le H(Y)$.

(b) Any two pairs in $\mathcal{C}$ are separated by at least g integers; their elements are congruent to $42t$, $42t+1$, $60t$ or $60t+1$ modulo R and exceed $L^2/5$.

(c) The rational $x' = x - \sum_{\mathcal{P}\in\mathcal{C}} \rho(\mathcal{P})$ satisfies $x - E(L) < x' \le x$ and $Kx' \in \mathbb{Z}$.

Proof. Put $x_0 = x$ and $\mathcal{C}_0 = \varnothing$, and suppose that $x_i > x - E(L)$ has been defined, with $\lambda(x_i) \le Y$, together with a set $\mathcal{C}_i$ of correction pairs, the union of batches at distinct levels above L. If $\lambda(x_i) \le L$ the process stops. Otherwise put $\Lambda_i = \lambda(x_i) = p_i^{e_i}$, so that $L < \Lambda_i \le Y$ and $\Lambda_i > Q_1$, and let $c_i = \operatorname{res}_{\Lambda_i}(x_i)$ be the leading residue of x_i. By Proposition 4.7 there is $\mathcal{S}_i \subseteq \mathcal{M}_{\Lambda_i}$ with $\sum_{m\in\mathcal{S}_i} m^{-1} \equiv c_i \pmod{p_i}$ and $|\mathcal{S}_i| \le N(\Lambda_i)$. Put

$$y_i = \sum_{m\in\mathcal{S}_i} \rho(\mathcal{P}_{\Lambda_i,m}), \qquad x_{i+1} = x_i - y_i, \qquad \mathcal{C}_{i+1} = \mathcal{C}_i \cup \{\mathcal{P}_{\Lambda_i,m} : m \in \mathcal{S}_i\}.$$

By Lemma 4.6(b), each $\rho(\mathcal{P}_{\Lambda_i,m})$ has $v_{p_i} \ge -e_i$ and residue m^{-1} at level Λ_i, so $v_{p_i}(y_i) \ge -e_i$ and, by additivity, $\operatorname{res}_{\Lambda_i}(y_i) = c_i = \operatorname{res}_{\Lambda_i}(x_i)$. By Lemma 4.6(a), every exact prime-power divisor of the reduced denominator of y_i other than its p_i-part divides one of the integers $\Lambda_i m$ or $\Lambda_i m + 1$ with $m \in \mathcal{S}_i$ and is smaller than Λ_i. Lemma 3.3 therefore gives $\lambda(x_{i+1}) < \Lambda_i$, provided $x_{i+1} \ne 0$, which is shown below. In particular the levels $\Lambda_0 > \Lambda_1 > \dots$ are strictly decreasing, no level is used twice, and every level used lies in $(L, Y]$; the process stops after finitely many steps.

By Proposition 4.7, $y_i < 10\Lambda_i^{-3/2}$. Since the levels are distinct prime powers exceeding L,

$$\sum_i y_i < 10 \sum_{Q>L} Q^{-3/2} = E(L),$$

so $x_{i+1} > x - E(L) > 0$ at every step, which justifies the induction, and the final rational x' satisfies $x - E(L) < x' \le x$. At termination $\lambda(x') \le L$: every exact prime-power divisor of the reduced denominator of x' is at most L and hence divides K, so $Kx' \in \mathbb{Z}$. This proves (c).

Let $\mathcal{C}$ be the final set of pairs. Its batches are the sets $\{\mathcal{P}_{\Lambda_i,m} : m \in \mathcal{S}_i\}$, at the distinct levels $\Lambda_i \in (L, Y]$, with $|\mathcal{S}_i| \le N(\Lambda_i)$; so $|\mathcal{C}| \le \sum_i N(\Lambda_i) \le H(Y)$, which proves (a). Distinct pairs in $\mathcal{C}$ are separated by at least g integers by Lemma 4.6(e), and the congruence and size assertions in (b) follow from Lemma 4.6(d) and $\Lambda_i > L$. □

6. Reserved pairs

The descent of Section 5 leaves a positive rational whose denominator divides $K = \operatorname{lcm}(1,\dots,L)$. It is absorbed exactly by completing some members of a fixed family of reserved pairs $\{Kd+1, Kd+2\}$ to triples $\{Kd, Kd+1, Kd+2\}$. The indices d are grouped by the following elementary fact.

Lemma 6.1. *For every integer $M \ge 1$ there is a finite set $\mathcal{D}$ of integers greater than M with $\sum_{d\in\mathcal{D}} 1/d = 1$.*

Proof. Let $N > M$ be the least integer with $\sum_{d=M+1}^{N} 1/d \ge 1$; it exists because the harmonic series diverges. If equality holds, $\mathcal{D} = \{M+1,\dots,N\}$ will do. Otherwise

$$x = 1 - \sum_{d=M+1}^{N-1} \frac{1}{d} \qquad \text{satisfies} \qquad 0 < x < \frac{1}{N}.$$

Expand x by the greedy algorithm: for a rational $a/b \in (0,1)$ in lowest terms put $q = \lceil b/a \rceil$ and pass to $a/b - 1/q$. Since $q - 1 < b/a$, the new numerator $aq - b$ satisfies $0 \le aq - b < a$, so the numerators in lowest terms strictly decrease and the expansion terminates; and $a/b - 1/q = (aq-b)/(bq) < a/(bq) < 1/(q(q-1))$, so the next denominator exceeds $q(q-1) \ge q$; the denominators therefore strictly increase. The first denominator is $\lceil 1/x \rceil > N$. Adjoining the denominators of the expansion to $\{M+1, \dots, N-1\}$ gives $\mathcal{D}$. □

6.1. **The fixed constants.** Fix the positive rational r and the integer $B \ge 0$ of Theorem A (the integer g and t, R were fixed in Section 4), and put

$$\eta = \frac{r}{1000}. \tag{17}$$

Choose an integer L so large that

$$L \ge Q_1, \qquad E(L) < \frac{\eta}{4}, \qquad L \ge R, \qquad L \ge \frac{10}{\eta}, \qquad L^2 > 5B, \tag{18}$$

which is possible by Corollary 4.8, and put

$$K = \operatorname{lcm}(1, \dots, L), \qquad J = \lceil 4K\eta \rceil. \tag{19}$$

Then $R \mid K$ and $K\eta \ge L\eta \ge 10$. By Lemma 6.1, applied J times with M equal to the larger of B and the largest integer already chosen, there are pairwise disjoint finite sets $\mathcal{D}_0, \dots, \mathcal{D}_{J-1}$ of integers greater than B with

$$\sum_{d \in \mathcal{D}_j} \frac{1}{d} = 1 \qquad (0 \le j < J). \tag{20}$$

Put

$$U = \bigcup_{j=0}^{J-1} \mathcal{D}_j, \qquad \mathcal{A} = \bigcup_{d \in U} \{Kd+1, Kd+2\}, \qquad C_0 = |U|, \qquad A_0 = \rho(\mathcal{A}). \tag{21}$$

All of r, B, η, L, K, J, the sets $\mathcal{D}_j$, U, $\mathcal{A}$, C_0 and A_0 remain fixed for the rest of the paper. For $d \in U$ the pair $R_d = \{Kd+1, Kd+2\}$ is the *reserved pair* with index d.

Lemma 6.2. *(a) $\mathcal{A}$ is a legal set with $\kappa(\mathcal{A}) = C_0$ components, namely the pairs R_d, any two of which are separated by at least g integers; $A_0 < 10\eta$; and every element of $\mathcal{A}$ exceeds B.*

(b) Let x be a rational number with $0 < x \le 2\eta$ and $Kx \in \mathbb{Z}$. There is a set $\mathcal{D} \subseteq U$ with

$$\sum_{d \in \mathcal{D}} \frac{1}{Kd} = x.$$

For any $\mathcal{D} \subseteq U$ the set $\mathcal{A}' = \mathcal{A} \cup \{Kd : d \in \mathcal{D}\}$ is legal, $\rho(\mathcal{A}') = A_0 + \sum_{d \in \mathcal{D}} 1/(Kd)$, $\kappa(\mathcal{A}') = C_0$, its components are the pairs R_d for $d \in U \setminus \mathcal{D}$ and the triples $\{Kd, Kd+1, Kd+2\}$ for $d \in \mathcal{D}$, any two of them are separated by at least g integers, and all its elements are congruent to 0, 1 or 2 modulo R.

Proof. (a) For distinct $d, d' \in U$ the pairs R_d and $R_{d'}$ are separated by at least $K-3$ integers, and $K - 3 \ge R - 3 > g$ by (4). By (20),

$$A_0 \le \frac{2}{K} \sum_{d \in U} \frac{1}{d} = \frac{2J}{K} \le \frac{8K\eta + 2}{K} = 8\eta + \frac{2}{K} \le \Big(8 + \frac{1}{5}\Big)\eta < 10\eta,$$

using $J \le 4K\eta + 1$ and $K \ge 10/\eta$. Every element of $\mathcal{A}$ is at least $K \min U + 1 > B$.

(b) The number $w = Kx$ is an integer with $1 \le w \le 2K\eta < J$. Put $\mathcal{D} = \mathcal{D}_0 \cup \dots \cup \mathcal{D}_{w-1}$; by (20),

$$\sum_{d \in \mathcal{D}} \frac{1}{Kd} = \frac{1}{K} \sum_{j < w} \sum_{d \in \mathcal{D}_j} \frac{1}{d} = \frac{w}{K} = x,$$

and $\mathcal{D} \subseteq U$. For any $\mathcal{D} \subseteq U$, adjoining Kd to $\mathcal{A}$ extends the reserved pair R_d to the triple $\{Kd, Kd+1, Kd+2\}$; the integer $Kd-1$ is not in $\mathcal{A}'$, because it is congruent to -1 modulo K while the elements of $\mathcal{A}'$ are congruent to 0, 1 or 2 modulo K, and $K > 3$. So the components of $\mathcal{A}'$ are as stated, their number is C_0, consecutive ones are separated by at least $K-3 \geq g$ integers, and $\rho(\mathcal{A}') = A_0 + \sum_{d\in\mathcal{D}} 1/(Kd)$. The congruences modulo R follow from $R \mid K$. □

7. Cells

Definition 7.1. For an integer $d \geq 1$ put $n = Kd + 8t$ and

$$\mathcal{A}_n = \{10n,\ 15n,\ 12n+2,\ 18n+3,\ 36n+6\}.$$

The *cell* with index d is the union of five triples,

$$Z_d = \bigcup_{a\in\mathcal{A}_n} [a, a+2],$$

and the *split cell* with index d is the union of six pairs,

$$Z_d^* = [6n, 6n+1] \cup \bigcup_{a\in\mathcal{A}_n} [a+1, a+2].$$

The common reciprocal sum of the two is denoted $h(d)$.

Thus Z_d^* is obtained from Z_d by deleting the first element of each of the five triples and adjoining the pair $\{6n, 6n+1\}$. In terms of d the fifteen elements of Z_d are the integers $\alpha d + \beta$ with (α, β) in the list

$$\begin{gathered}(10K,\ 80t+i), \qquad (15K,\ 120t+i), \qquad (12K,\ 96t+2+i),\\ (18K,\ 144t+3+i), \qquad (36K,\ 288t+6+i), \qquad i \in \{0,1,2\},\end{gathered} \tag{22}$$

and Z_d^* contains in addition $6Kd + 48t$ and $6Kd + 48t + 1$.

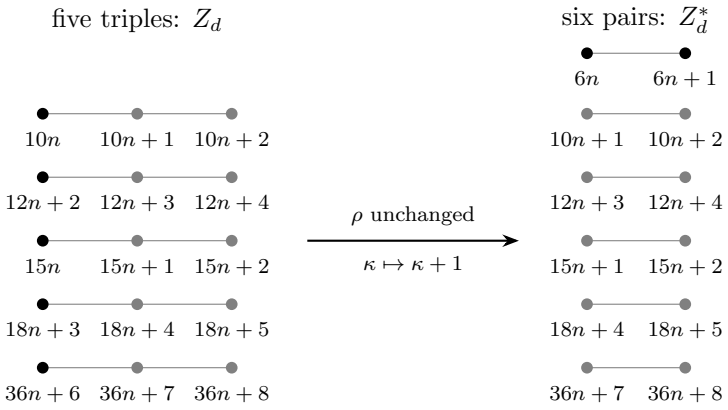

FIGURE 1. The two states of one cell. The five black elements on the left and the two black elements on the right have the same reciprocal sum; the ten gray elements are retained. Splitting a cell changes nothing outside it.

Lemma 7.2. *For every $d \geq 1$:*

(a) $\rho(Z_d^*) = \rho(Z_d) = h(d)$;

(b) Z_d is legal with $\kappa(Z_d) = 5$ and all components of length 3; Z_d^* is legal with $\kappa(Z_d^*) = 6$ and all components of length 2;

(c)

$$\frac{1}{K(d+1)} \leq h(d) \leq \frac{1}{Kd}, \qquad h(d) = \frac{1}{Kd} + O_{K,t}\big(d^{-2}\big). \tag{23}$$

Proof. (a) Since $12n+2 = 2(6n+1)$, $18n+3 = 3(6n+1)$ and $36n+6 = 6(6n+1)$,

$$\frac{1}{10n} + \frac{1}{15n} = \frac{1}{6n}, \qquad \frac{1}{12n+2} + \frac{1}{18n+3} + \frac{1}{36n+6} = \frac{1}{6n+1}\Big(\frac{1}{2} + \frac{1}{3} + \frac{1}{6}\Big) = \frac{1}{6n+1},$$

so the five deleted elements and the two adjoined ones have the same reciprocal sum.

(b) In increasing order the five triples of Z_d begin at $10n$, $12n+2$, $15n$, $18n+3$, $36n+6$, and the six pairs of Z_d^* begin at $6n$, $10n+1$, $12n+3$, $15n+1$, $18n+4$, $36n+7$. In each list the difference between consecutive starting points is at least $2n$, so the displayed intervals are pairwise disjoint and pairwise nonadjacent.

(c) Each pair (α, β) in (22) has $0 < \beta < \alpha$, because $\beta \le 288t+8$ and $\alpha \ge 10K \ge 2100t$. Hence $1/(\alpha(d+1)) \le 1/(\alpha d+\beta) \le 1/(\alpha d)$, and summing over the fifteen elements, with $3\big(\frac{1}{10} + \frac{1}{15} + \frac{1}{12} + \frac{1}{18} + \frac{1}{36}\big) = 1$, gives the two bounds; $1/(\alpha d+\beta) = 1/(\alpha d) + O_{K,t}(d^{-2})$ gives the asymptotic. □

Modulo $R = 210t$ the elements of $Z_d \cup Z_d^*$ lie in the six bands

$$[48t, 48t{+}1],\ [78t{+}6, 78t{+}8],\ [80t, 80t{+}2],\ [96t{+}2, 96t{+}4],\ [120t, 120t{+}2],\ [144t{+}3, 144t{+}5], \tag{24}$$

independently of d: the first is the class of the pair $\{6n, 6n+1\}$, the second that of the triple at $36n+6 = 36Kd + 288t + 6$, and the others are the classes of the triples at $10n$, $12n+2$, $15n$ and $18n+3$. The reserved pairs and their completions occupy $[0, 2]$, and the correction pairs occupy $[42t, 42t+1]$ and $[60t, 60t+1]$, by Lemma 6.2(b) and Lemma 4.6(d). In cyclic order the nine bands are

$$\begin{aligned} &[0, 2], \quad [42t, 42t+1], \quad [48t, 48t+1], \quad [60t, 60t+1], \quad [78t+6, 78t+8], \\ &[80t, 80t+2], \quad [96t+2, 96t+4], \quad [120t, 120t+2], \quad [144t+3, 144t+5], \end{aligned} \tag{25}$$

and the numbers of residues strictly between consecutive bands, the last one counted cyclically up to $210t$, are

$$42t-3, \quad 6t-2, \quad 12t-2, \quad 18t+4, \quad 2t-9, \quad 16t-1, \quad 24t-5, \quad 24t, \quad 66t-6,$$

all at least $2t-9 \ge g$ by (4). Figure 2 shows the nine bands.

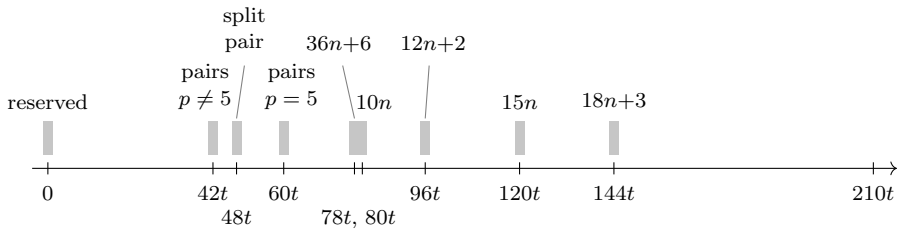

FIGURE 2. The nine bands of residues modulo $R = 210t$ occupied by components: the reserved pairs and triples at $[0, 2]$, the correction pairs at $[42t, 42t+1]$ and $[60t, 60t+1]$, the pair $\{6n, 6n+1\}$ of a split cell at $[48t, 48t+1]$, and the five triples of a cell at $[78t+6, 78t+8]$, $[80t, 80t+2]$, $[96t+2, 96t+4]$, $[120t, 120t+2]$ and $[144t+3, 144t+5]$. Band widths are not to scale. The narrowest gap, between $78t+8$ and $80t$, contains $2t-9$ residues.

Proposition 7.3. *Let $\mathcal{A}'$ be a set as in Lemma 6.2(b), let $\mathcal{C}$ be a finite set of correction pairs, let E be a finite set of positive integers, and for each $d \in E$ let Z'_d be either Z_d or Z_d^*. Then*

$$S = \mathcal{A}' \cup \bigcup_{\mathcal{P} \in \mathcal{C}} \mathcal{P} \cup \bigcup_{d \in E} Z'_d$$

is a legal set whose components are exactly the components of $\mathcal{A}'$, the pairs in $\mathcal{C}$, and the components of the sets Z'_d; any two components of S are separated by at least g integers; every component of S has length 2 or 3; and

$$\rho(S) = \rho(\mathcal{A}') + \sum_{\mathcal{P}\in\mathcal{C}} \rho(\mathcal{P}) + \sum_{d\in E} h(d), \qquad \kappa(S) = C_0 + |\mathcal{C}| + \sum_{d\in E} \kappa(Z'_d).$$

Proof. It suffices to show that any two of the listed components are separated by at least $g \geq 1$ integers: they are then pairwise disjoint and pairwise nonadjacent runs of length 2 or 3, so their union is a legal set having exactly these runs as components, and the formulas for $\rho(S)$ and $\kappa(S)$ follow.

Let X and Y be two of the listed components with $x = \max X < y = \min Y$. Every listed component lies, modulo R, inside one of the nine bands (25). Suppose first that X and Y lie in different bands. Then $y - x$ is congruent modulo R to the difference of a residue in the band of Y and a residue in the band of X, and every positive integer in that residue class is at least 1 plus the number of residues strictly between the band of X and the band following it in cyclic order, hence at least $2t - 8$. So $y - x \geq 2t - 8$, and at least $2t - 9 \geq g$ integers lie strictly between X and Y.

Suppose next that X and Y lie in the same band. If both come from $\mathcal{A}'$, then $y - x \geq K - 2$, because the components of $\mathcal{A}'$ start at integers of the form Kd or $Kd + 1$ with distinct d and have length at most 3. If both come from cells, they belong to the same linear family, with coefficient $\alpha \geq 6K$ and different indices, so $y - x \geq 6K - 2$. If both are correction pairs, their centres differ by at least $18t$ by Lemma 4.6(e), so $y - x \geq 18t - 1$. In each case $y - x - 1 \geq g$, because $K \geq R = 210t > 2t - 9 \geq g$.

The component lengths are 2 or 3 for $\mathcal{A}'$ (Lemma 6.2), 2 for $\mathcal{C}$, and 3 or 2 for the cells (Lemma 7.2(b)). □

Corollary 7.4. *In the situation of Proposition 7.3, let $E' \subseteq E$ be a set of indices with $Z'_d = Z_d$ for every $d \in E'$. Replacing Z_d by Z^*_d for the indices $d \in E'$ produces a legal set with the same reciprocal sum, with any two components separated by at least g integers, and with exactly $|E'|$ more components.*

Proof. Apply Proposition 7.3 to the new choice of the sets Z'_d and use Lemma 7.2(a),(b). □

8. Powersmooth cells at a fixed scale

Put

$$G = r - A_0 - 2\eta, \tag{26}$$

which is positive since $A_0 < 10\eta$ by Lemma 6.2(a). Choose an integer $C > 1$ with

$$\frac{1}{K}\log C > 2r, \tag{27}$$

then a real number ε with

$$0 < \varepsilon < \min\Big\{\frac{1}{10},\ \frac{rK}{100(C-1)}\Big\}, \tag{28}$$

and put $\Gamma = 36KC + 300t$. The constants C, ε and Γ are fixed from now on. For an integer $T \geq 1$ put

$$Y_T = T^{1-\varepsilon}. \tag{29}$$

Definition 8.1. A cell Z_d is *admissible at scale T* if all fifteen elements of Z_d are Y_T-powersmooth. Write $\mathcal{L}_T$ for the set of $d \in [T, CT]$ such that Z_d is admissible at scale T.

Admissibility refers to the cell Z_d only; the elements $6Kd + 48t$ and $6Kd + 48t + 1$ of Z^*_d are not required to be powersmooth.

8.1. **Few cells are lost.**

Lemma 8.2. *As $T \to \infty$,*

$$\#\big([T, CT] \cap \mathbb{Z} \setminus \mathcal{L}_T\big) \le 30(C-1)\varepsilon T + o(T),$$

where the $o(T)$ depends on K, t, C and ε.

Proof. If Z_d is not admissible, some element $\alpha d + \beta$ in (22) is divisible by a prime power $q > Y_T$, and $q \le \alpha d + \beta \le 36KCT + 288t + 8 < \Gamma T$. The fifteen pairs (α, β) are fixed, with $\alpha \in \{10K, 12K, 15K, 18K, 36K\}$, and the prime divisors of α are at most L. For each of the fifteen forms we count the $d \in [T, CT]$ with $q \mid \alpha d + \beta$ for some prime power q with $Y_T < q < \Gamma T$.

Consider first $q = p$ prime. For T so large that $Y_T > L$, every such p is prime to α, so the d with $p \mid \alpha d + \beta$ form a single residue class modulo p, and there are at most $(C-1)T/p + 1$ of them in $[T, CT]$. By Mertens' theorem [10, Theorem 427],

$$\sum_{Y_T < p < \Gamma T} \frac{1}{p} = \log \frac{\log(\Gamma T)}{\log Y_T} + o(1) = \log \frac{\log \Gamma + \log T}{(1-\varepsilon)\log T} + o(1) = -\log(1-\varepsilon) + o(1),$$

and $-\log(1-\varepsilon) \le \varepsilon/(1-\varepsilon) \le 2\varepsilon$ for $\varepsilon \le 1/10$. Since Γ is fixed, $\pi(\Gamma T) = o(T)$ by Chebyshev's bound [10, Theorem 7]. The primes therefore account for at most $(C-1)T(2\varepsilon + o(1)) + \pi(\Gamma T) = 2(C-1)\varepsilon T + o(T)$ values of d for each form.

Consider next $q = p^a$ with $a \ge 2$. Put $\gamma = \gcd(\alpha, q) \le \alpha \le 36K$. If the congruence $\alpha d \equiv -\beta \pmod q$ is solvable, its solutions form one residue class modulo q/γ, so at most $\gamma\big((C-1)T/q + 1\big)$ values of $d \in [T, CT]$ satisfy it. Summing over the proper prime powers q with $Y_T < q < \Gamma T$,

$$\sum_{\substack{Y_T < q < \Gamma T \\ q \text{ proper}}} \gamma\Big(\frac{(C-1)T}{q} + 1\Big) \le 36K\Big((C-1)T \sum_{\substack{q > Y_T \\ q \text{ proper}}} \frac{1}{q} + \#\{q < \Gamma T : \ q \text{ proper}\}\Big).$$

The series $\sum_q 1/q$ over proper prime powers converges, so its tail beyond Y_T is $o(1)$, and the number of proper prime powers below x is at most $\sum_{2 \le a \le \log_2 x} x^{1/a} = O(x^{1/2})$. So the proper prime powers account for $o(T)$ values of d for each form.

Summing over the fifteen forms gives the lemma. □

8.2. **Selection of a prefix.**

Lemma 8.3. *For all sufficiently large T the following hold.*

(a) $\sum_{d \in \mathcal{L}_T} h(d) > r$.

(b) Let D_T be the least integer in $[T, CT]$ such that $\sum_{d \in \mathcal{L}_T,\ d \le D_T} h(d) \ge G$, let $\mathcal{E}_T = \{d \in \mathcal{L}_T : d \le D_T\}$ and $m_T = |\mathcal{E}_T|$. Then

$$G \le \sum_{d \in \mathcal{E}_T} h(d) < G + \frac{1}{KT} \tag{30}$$

and

$$KG\,T \le m_T \le (C-1)T + 1. \tag{31}$$

Proof. By (23),

$$\sum_{T \le d \le CT} h(d) = \frac{1}{K} \sum_{T \le d \le CT} \frac{1}{d} + O_{K,t}\Big(\frac{1}{T}\Big) = \frac{1}{K}\log C + o(1),$$

which exceeds $2r$ for large T by (27). Every $d \in [T, CT]$ has $h(d) \le 1/(KT)$, so by Lemma 8.2 the cells with $d \in [T, CT]$ that are not admissible have total reciprocal sum at most

$$\big(30(C-1)\varepsilon T + o(T)\big)\frac{1}{KT} = \frac{30(C-1)\varepsilon}{K} + o(1) < \frac{3r}{10} + o(1)$$

by (28), which is less than $r/2$ for large T. Hence $\sum_{d\in\mathcal{L}_T} h(d) > 2r - r/2 - o(1) > r$ for large T, proving (a).

Since $r > G$, (a) shows that D_T exists and $D_T \le CT$. The sum in (30) is at least G by definition, and it was smaller than G before the last term $h(D_T) \le 1/(KT)$ was added; this gives (30). The lower bound in (31) follows from $G \le m_T/(KT)$, and the upper bound from $\mathcal{E}_T \subseteq [T, CT]$. □

8.3. **One correction for all cell states.** For $\sigma \in \{0,1\}^{\mathcal{E}_T}$ write $Z_d^{(0)} = Z_d$ and $Z_d^{(1)} = Z_d^*$, and $|\sigma|$ for the number of indices d with $\sigma_d = 1$.

Theorem 8.4. *There is a constant A_1, depending on r, B and g, such that for all sufficiently large T there are a set $\mathcal{A}'_T$ as in Lemma 6.2(b), a set $\mathcal{C}_T$ of correction pairs, and an injective map*

$$\Psi_T\colon \{0,1\}^{\mathcal{E}_T} \to \{\textit{finite legal sets}\}$$

with the following properties, where $j_T = |\mathcal{C}_T|$ and $a_T = C_0 + 5m_T + j_T$.

(a) $0 \le j_T \le H(Y_T) \ll T^{1-\varepsilon/2}$.

(b) For every σ, the set $\Psi_T(\sigma)$ has reciprocal sum r and exactly $a_T + |\sigma|$ components; every component has length 2 or 3; any two components are separated by at least g integers; and

$$B < \min \Psi_T(\sigma), \qquad \max \Psi_T(\sigma) \le A_1 T^{2-2\varepsilon}.$$

(c) If exactly u coordinates of σ are 0, then $\Psi_T(\sigma)$ has $5u + \tau_T$ components of length 3, where $0 \le \tau_T \le C_0$ does not depend on σ.

Proof. Let T be large enough for Lemma 8.3, and put

$$x_T = r - A_0 - \sum_{d\in\mathcal{E}_T} h(d) = G + 2\eta - \sum_{d\in\mathcal{E}_T} h(d). \tag{32}$$

By (30), $2\eta - 1/(KT) < x_T \le 2\eta$, so

$$\frac{7\eta}{4} < x_T \le 2\eta \tag{33}$$

once $1/(KT) < \eta/4$. The reduced denominator of x_T divides the least common multiple of the denominator of r, the elements of $\mathcal{A}$ and the elements of the cells Z_d, $d \in \mathcal{E}_T$. The last are Y_T-powersmooth by admissibility, and the finitely many fixed integers among the former are Y_T-powersmooth as soon as Y_T exceeds their largest exact prime-power divisor. Hence $\lambda(x_T) \le Y_T$ for large T. Also $x_T > 7\eta/4 > E(L)$ by (18).

Apply Theorem 5.1 to $x = x_T$ with $Y = Y_T$ (and L, K as in (18), (19)). It gives a set $\mathcal{C}_T$ of correction pairs at levels in $(L, Y_T]$, with $j_T := |\mathcal{C}_T| \le H(Y_T)$, and the rational

$$x_f = x_T - \sum_{\mathcal{P}\in\mathcal{C}_T} \rho(\mathcal{P})$$

satisfies $x_T - E(L) < x_f \le x_T$ and $Kx_f \in \mathbb{Z}$. By (33) and $E(L) < \eta/4$,

$$\frac{3\eta}{2} < x_f \le 2\eta.$$

By Lemma 6.2(b) there is $\mathcal{D} \subseteq U$ with $\sum_{d\in\mathcal{D}} 1/(Kd) = x_f$; let $\mathcal{A}'_T = \mathcal{A} \cup \{Kd : d \in \mathcal{D}\}$, so that $\rho(\mathcal{A}'_T) = A_0 + x_f$ and $\kappa(\mathcal{A}'_T) = C_0$, and put $\tau_T = |\mathcal{D}|$, the number of triples of $\mathcal{A}'_T$. Define

$$\Psi_T(\sigma) = \mathcal{A}'_T \cup \bigcup_{\mathcal{P}\in\mathcal{C}_T} \mathcal{P} \cup \bigcup_{d\in\mathcal{E}_T} Z_d^{(\sigma_d)}. \tag{34}$$

By Proposition 7.3, $\Psi_T(\sigma)$ is legal, its components have length 2 or 3 and are pairwise separated by at least g integers, and

$$\kappa(\Psi_T(\sigma)) = C_0 + j_T + 5(m_T - |\sigma|) + 6|\sigma| = a_T + |\sigma|$$

and

$$\rho(\Psi_T(\sigma)) = A_0 + x_f + \sum_{\mathcal{P}\in\mathcal{C}_T} \rho(\mathcal{P}) + \sum_{d\in\mathcal{E}_T} h(d) = r$$

by (32). The components of length 3 are the τ_T triples of $\mathcal{A}'_T$ and the five triples of each unsplit cell, which proves (c). The map is injective: the pair $\{6Kd+48t, 6Kd+48t+1\}$ is contained in $\Psi_T(\sigma)$ if and only if $\sigma_d = 1$, because its residue band $[48t, 48t+1]$ is used by no other kind of component and distinct indices give distinct pairs; so $\Psi_T(\sigma)$ determines σ.

The elements of $\mathcal{A}'_T$ exceed B by Lemma 6.2; those of the correction pairs exceed $L^2/5 > B$ by Theorem 5.1(b) and (18); and those of the cells, in either state, are at least $6KT > B$ once $T \ge B$. For the upper bound, every element of $\mathcal{A}'_T$ is at most $M_0 := K\max U + 2$, a fixed integer; every element of a cell, in either state, is less than ΓT; and every element of a correction pair at level $Q \le Y_T$ is less than $Q^2 \le T^{2-2\varepsilon}$, by (10). Hence $\max\Psi_T(\sigma) \le \max\{M_0, \Gamma T, T^{2-2\varepsilon}\} \le A_1 T^{2-2\varepsilon}$ for a suitable A_1.

Finally, Corollary 4.8 with $\nu = \varepsilon/(2(1-\varepsilon))$ gives $H(Y_T) \ll T^{(1-\varepsilon)(1+\nu)} = T^{1-\varepsilon/2}$, which proves (a). □

Corollary 8.5. *For all sufficiently large T and every integer s with $0 \le s \le m_T$, at least $\binom{m_T}{s}$ distinct legal sets have reciprocal sum r, exactly $a_T + s$ components, and the properties listed in Theorem 8.4(b). In particular every integer in $[a_T, a_T + m_T]$ is the number of components of such a set.*

Proof. The sets $\Psi_T(\sigma)$ with $|\sigma| = s$ are distinct by injectivity, and Theorem 8.4(b) applies to each. □

9. Consecutive scales and the proof of Theorem A

Proposition 2.1 will be applied to the maps Ψ_T of Theorem 8.4, with $q = 5$ and $b_T = C_0 + j_T$. The hypothesis $b_T = o(T)$ is Theorem 8.4(a), and $m_T \asymp T$ is (31); what remains is the change of m_T from one scale to the next. The number j_T of correction pairs need not vary regularly with T, and no such regularity is used.

Lemma 9.1. $|\mathcal{L}_T \triangle \mathcal{L}_{T+1}| = O_{K,C,\varepsilon}(T^{\varepsilon})$.

Proof. The two sets of indices differ outside $[T+1, CT]$ in at most $1+C$ elements. On $[T+1, CT]$ every cell admissible at scale T is admissible at scale $T+1$, because $Y_{T+1} > Y_T$. Conversely, a cell Z_d with $d \in [T+1, CT]$ admissible at scale $T+1$ but not at scale T has an element $\alpha d + \beta$ in (22) divisible by a prime power q with $Y_T < q \le Y_{T+1}$. Since

$$Y_{T+1} - Y_T = (T+1)^{1-\varepsilon} - T^{1-\varepsilon} \le (1-\varepsilon)T^{-\varepsilon} < 1,$$

there is at most one integer in $(Y_T, Y_{T+1}]$. If it is a prime power q, then, as in the proof of Lemma 8.2, for each of the fifteen forms at most $\gcd(\alpha, q)\big((C-1)T/q + 1\big) \le 36K\big((C-1)T^{\varepsilon} + 1\big)$ indices $d \in [T, CT]$ satisfy $q \mid \alpha d + \beta$, using $q > Y_T = T^{1-\varepsilon}$. □

Lemma 9.2. $|m_{T+1} - m_T| = O_{K,C,\varepsilon}(T^{\varepsilon})$ *for all sufficiently large T.*

Proof. Put $\Lambda = \mathcal{L}_T \cap \mathcal{L}_{T+1}$, $P = \mathcal{E}_T$ and $P' = \mathcal{E}_{T+1}$, and let $s = |\mathcal{L}_T \triangle \mathcal{L}_{T+1}|$. All indices under consideration lie in $[T, C(T+1)]$, so by (23) every $h(d)$ involved satisfies

$$h_{\min} := \frac{1}{K(C(T+1)+1)} \le h(d) \le \frac{1}{KT} =: h_{\max}, \qquad \frac{h_{\max}}{h_{\min}} \le 3C.$$

The set $P \cap \Lambda = \{d \in \Lambda : d \le D_T\}$ is an initial segment of Λ in the natural order, and so is $P' \cap \Lambda = \{d \in \Lambda : d \le D_{T+1}\}$; hence one of them contains the other. By symmetry assume $P \cap \Lambda \subseteq P' \cap \Lambda$ and put $\Delta = (P' \cap \Lambda) \setminus (P \cap \Lambda)$. Then

$$\sum_{d\in P'} h(d) - \sum_{d\in P} h(d) = \sum_{d\in\Delta} h(d) + \sum_{d\in P'\setminus\Lambda} h(d) - \sum_{d\in P\setminus\Lambda} h(d).$$

The left-hand side is less than $h_{\max}$ by (30) at scales T and $T+1$, the middle sum on the right is nonnegative, and $|P \setminus \Lambda| \le s$ because $P \setminus \Lambda \subseteq \mathcal{L}_T \setminus \mathcal{L}_{T+1}$. Hence

$$|\Delta|\, h_{\min} \le \sum_{d \in \Delta} h(d) < (1+s) h_{\max}, \qquad \text{so} \qquad |\Delta| \le 3C(1+s).$$

Since $P \triangle P' \subseteq (P \setminus \Lambda) \cup (P' \setminus \Lambda) \cup \Delta$,

$$|m_{T+1} - m_T| \le |P \triangle P'| \le 2s + 3C(1+s),$$

and $s = O_{K,C,\varepsilon}(T^{\varepsilon})$ by Lemma 9.1. □

The argument uses no information about the gaps between consecutive elements of $\mathcal{L}_T$: it counts admissible indices between the two stopping points, not the distance between them.

Proof of Theorem A. The constants were fixed at the start of Section 4, in Section 6.1 and at the start of Section 8, in the order g, t, R; r, B, η; L, K, J; $\mathcal{D}_0, \dots, \mathcal{D}_{J-1}, U, C_0, A_0$; C; ε; Γ; they do not depend on T. For T large, Theorem 8.4 provides Ψ_T, and (31), Theorem 8.4(a) and Lemma 9.2 give

$$m_T \asymp T, \qquad m_{T+1} - m_T = o(T), \qquad b_T := C_0 + j_T = o(T).$$

Put $\delta = 2\varepsilon$; then $0 < \delta < 1/5$ by (28).

(i) By Proposition 2.1 with $q = 5$, there is k_0 such that every $k \ge k_0$ is the number of components of $\Psi_T(\sigma)$ for some T with $k \ge 5m_T$ and at least $2^{k/21}$ distinct σ. These sets $\Psi_T(\sigma)$ are distinct legal sets, and by Theorem 8.4(b) each is a family of k intervals of integers greater than B, of lengths 2 or 3, any two separated by at least g integers, with reciprocal sum r and largest element at most $A_1 T^{2-2\varepsilon}$. Since $k \ge 5m_T \ge 5KGT$ by (31), the largest element is at most $A_1 (5KG)^{-(2-\delta)} k^{2-\delta}$.

(ii) Put $v_T = a_T + m_T = C_0 + 6m_T + j_T$. By Lemma 9.2 and Theorem 8.4(a),

$$|v_T - v_{T-1}| \le 6|m_T - m_{T-1}| + j_T + j_{T-1} = O\big(T^{1-\varepsilon/2}\big).$$

Let T_* be such that Theorem 8.4 and this bound hold for all $T \ge T_*$, and let $k > v_{T_*}$. Since $v_T \ge m_T \to \infty$, there is a least $T \ge T_*$ with $v_T \ge k$, and then $T > T_*$ and $v_{T-1} < k \le v_T$. Put $u = v_T - k$; then $0 \le u < v_T - v_{T-1} = O(T^{1-\varepsilon/2})$, so $u \le m_T$ for large k. Choose σ with exactly u coordinates equal to 0. Then $\Psi_T(\sigma)$ has $a_T + m_T - u = k$ components, and by Theorem 8.4(c) at most $5u + C_0 = O(T^{1-\varepsilon/2})$ of them have length 3. As $k \ge a_T \ge 5m_T \ge 5KGT$, this is $O(k^{1-\varepsilon/2}) = O(k^{1-\delta/4})$, and the largest element is $O(k^{2-\delta})$ as in (i). Enlarging A and k_0 to cover both assertions completes the proof. □

Proof of Corollary B. Take $r = 1$, $B = 0$ and $g = 1$ in Theorem A(i). □

Corollary 9.3. *Let $r > 0$ be rational, let $g \ge 1$ be an integer, and let S_0 be a finite set of positive integers with $\rho(S_0) < r$ whose components have length 2 or 3 and are pairwise separated by at least g integers. For every sufficiently large integer k there is a set $S \supseteq S_0$ with $\rho(S) = r$ and exactly k components, all of length 2 or 3 and pairwise separated by at least g integers, and all elements of $S \setminus S_0$ can be required to exceed any prescribed bound.*

Proof. Apply Theorem A to the rational $r - \rho(S_0) > 0$, to a bound $B \ge \max S_0 + g$ at least as large as the prescribed one, and to the count $k - \kappa(S_0)$; adjoin S_0. Every component of S_0 is separated from every new component by at least g integers, so the components of the union are those of the two parts. □

Remark 9.4. Every threshold in the proof is effective in principle: the exponential-sum bounds (6), the divisor bound and the prime-number estimates used in Lemma 8.2 admit effective constants, and Lemma 6.1 is a terminating algorithm. None of the thresholds is small. The reserved family alone has $C_0 = |U|$ components with U built from $J = \lceil 4K\eta \rceil$ unit-fraction expansions and $K = \operatorname{lcm}(1, \dots, L)$, the condition $\log C > 2rK$ forces a small ε in (28), and T_* must be large enough

for the sieve of Lemma 8.2. The argument therefore gives no practical value of $k_0(r, B, g)$ and no algorithm for a prescribed modest k; the exponents $1/21$ and δ are not optimised. What is known for small k when $r = 1$ is collected in Appendix A.

Remark 9.5. Split cells and correction pairs consist of pairs, so the only components of length 3 in $\Psi_T(\sigma)$ come from unsplit cells and from the completed reserved pairs; this is what gives Theorem A(ii). Whether every sufficiently large k admits a representation of r by k intervals of length 2 only is not known to the author; the cell of Definition 7.1 cannot decide it, since its unsplit state consists of triples, and a construction with pairs alone would need a different local identity.

Appendix A. Finite representations of 1

The statements in this appendix concern representations of 1 by pairwise disjoint, pairwise non-adjacent intervals of length at least two, with no restriction on the lengths. They were established by exhaustive enumeration in exact rational arithmetic, with the pruning of Lemma 3.2, and are not used in the proof of Theorem A.

Proposition A.1. *(i) No such representation has all denominators at most* 84*. Exactly four have all denominators at most* 85*; they have* 8*,* 9*,* 10 *and* 11 *intervals.*
(ii) Every such representation with all denominators at most 99 *has at least eight intervals, and there is one with seven intervals and largest denominator* 100*.*
(iii) Such representations exist for every number of intervals from 7 *to* 21*.*

Exact verification. A representation cannot contain 1, since it would then contain 2 and its reciprocal sum would exceed 1. Start with the set $\{2, \dots, N\}$ of candidate denominators, and repeat the following two reductions until nothing changes: remove every candidate neither of whose neighbours is a candidate; and, for each prime p, with q the largest power of p dividing a candidate, compute in $\mathbb{F}_p$ all subset sums of the residues $(n/q)^{-1}$ over the candidates n divisible by q, and remove every such n that lies in no subset with sum 0. Both reductions remove only integers that belong to no representation, the second by Lemma 3.2. For each maximal run of the remaining candidates enumerate every subset without isolated elements, including the empty one; since no subset can join two different runs, these local choices enumerate exactly the legal subsets of the candidate set. With D the least common multiple of the candidates and weight D/n attached to n, a meet-in-the-middle enumeration retains the choices of total weight D, and each retained set is rechecked in rational arithmetic. For $N = 84$ nothing is retained; for $N = 85$ exactly the four sets displayed below are retained; for $N = 99$ seven sets are retained, each with at least eight intervals. This proves (i) and the first assertion of (ii). The displayed seven-interval identity proves the second assertion of (ii), and the examples recorded in Table 1, each rechecked in rational arithmetic for disjointness, nonadjacency, lengths and sum, prove (iii). □

The four representations with all denominators at most 85 are

$$1 = H[7,10] + H[14,15] + H[17,19] + H[34,36] + H[44,45] + H[56,57] + H[76,77] + H[84,85],$$
$$1 = H[5,6] + H[14,15] + H[17,18] + H[20,22] + H[27,28] + H[33,34] + H[44,45] + H[54,55] + H[84,85],$$
$$1 = H[6,8] + H[14,15] + H[17,18] + H[26,27] + H[34,35] + H[44,45] + H[54,56] + H[65,66] + H[77,78] + H[84,85],$$
$$1 = H[7,8] + H[11,12] + H[17,18] + H[21,22] + H[26,28] + H[33,34] + H[44,45] + H[54,56] + H[65,66] + H[77,78] + H[84,85],$$

and a representation with seven intervals is

$$1 = H[4,5] + H[9,11] + H[21,22] + H[25,26] + H[65,66] + H[77,78] + H[99,100].$$

Thus 85 is the smallest possible largest denominator, and 100 the smallest for seven intervals. Whether fewer than seven intervals can occur is not decided by these computations. Of the examples in Table 1, those with 7, 9, 10, 11, 12 and 20 intervals use intervals of length 2 or 3 only; the others contain longer intervals, so they do not exhibit the conclusion of Theorem A for small k.

TABLE 1. Examples for Proposition A.1(iii): number of intervals k, largest denominator, and largest interval length.

k	largest denominator	largest length	k	largest denominator	largest length
7	100	3	15	190	4
8	85	4	16	190	4
9	85	3	17	190	5
10	85	3	18	190	4
11	85	3	19	190	5
12	176	3	20	196	2
13	145	4	21	2080	5
14	196	4			

REFERENCES

[1] N. Alon, M. B. Nathanson and I. Ruzsa, *The polynomial method and restricted sums of congruence classes*, J. Number Theory **56** (1996), no. 2, 404–417.

[2] T. F. Bloom, *Erdős Problem #289*, `https://www.erdosproblems.com/289`, accessed 25 September 2026.

[3] T. F. Bloom, *On a density conjecture about unit fractions* (with an appendix by T. F. Bloom and B. Mehta), J. Eur. Math. Soc. **27** (2025), no. 11, 4563–4589.

[4] T. F. Bloom and C. Elsholtz, *Egyptian fractions*, Nieuw Arch. Wiskd. (5) **23** (2022), no. 4, 237–245.

[5] E. S. Croot III, *On unit fractions with denominators in short intervals*, Acta Arith. **99** (2001), no. 2, 99–114.

[6] E. S. Croot III, *On a coloring conjecture about unit fractions*, Ann. of Math. (2) **157** (2003), no. 2, 545–556.

[7] H. Davenport, *On the addition of residue classes*, J. London Math. Soc. **10** (1935), 30–32.

[8] J. A. Dias da Silva and Y. O. Hamidoune, *Cyclic spaces for Grassmann derivatives and additive theory*, Bull. London Math. Soc. **26** (1994), no. 2, 140–146.

[9] P. Erdős and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathématique **28**, Université de Genève, Geneva, 1980.

[10] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 6th ed., Oxford University Press, Oxford, 2008.

[11] H. Iwaniec and E. Kowalski, *Analytic number theory*, Amer. Math. Soc. Colloq. Publ. **53**, American Mathematical Society, Providence, RI, 2004.

[12] V. Kovač and T. Tao, *On several irrationality problems for Ahmes series*, Acta Math. Hungar. **175** (2025), 572–608.

[13] M. B. Nathanson, *Additive number theory: inverse problems and the geometry of sumsets*, Graduate Texts in Mathematics **165**, Springer-Verlag, New York, 1996.

[14] I. E. Shparlinski, *On exponential sums with sparse polynomials and rational functions*, J. Number Theory **60** (1996), no. 2, 233–244.

[15] I. E. Shparlinski, *Primitive points on a modular hyperbola*, Bull. Polish Acad. Sci. Math. **54** (2006), no. 3–4, 193–200.

[16] I. E. Shparlinski, *Modular hyperbolas*, Jpn. J. Math. **7** (2012), no. 2, 235–294.

PINGYOU LTD, LONDON, UNITED KINGDOM
Email address: `budden@pingyou.com`