hexagon:2611.01294v1

Second-Order Extremal Asymptotics for Unit Fractions and Dickman Phase Transitions

Pengcheng Chen

Abstract

Let $E(N)$ denote the largest cardinality of a subset of $[1, N]$ with no reciprocal subsum equal to one. Refining the first-order theorem of Liu–Sawhney, we obtain the second-order formula

$$E(N) = (1 - e^{-1})N + \big(\mathcal{C}(2/e) + o(1)\big)\frac{N \log\log N}{\log N},$$

Here ρ is the Dickman function and $\mathcal{C}(a)$ is the integral of the positive part of $a - 1 + \rho(v)$ over $[0, \infty)$. Numerically, $\mathcal{C}(2/e) \approx 1.0871215914020186$. The lower bound is an explicit construction. The converse combines a uniform deletion-cost ledger for prime-power classes with sparse modular repairs and scalar exact completion on the actual rational lattice. At the local scale $x = \lfloor (\log p)^v \rfloor$ we establish the limiting deletion fraction $1 - \rho(v)$ for modular inverse zero sums, together with stability and a sparse common pool representing every residue after subcritical deletions. We include the fine- and coarse-scale completion proofs. The global asymptotic does not use the companion entropy and mixing results; their implications are recorded separately.

MSC 2020: 11D68, 11N25, 11B75.
Keywords: modular reciprocals, Dickman function, smooth numbers, zero-sum-free sets, Egyptian fractions, extremal stability.

1 Main results

Throughout, intervals of integers are intersected with $\mathbb{N}$, and logarithms are natural. For a finite set A of positive integers write $R(A) = \sum_{n \in A} 1/n$, with $R(\varnothing) = 0$. For $1 \le x < p$, let

$$\sigma(p, x) = \min\{|F| : F \subseteq [1, x],\ \nexists \varnothing \neq S \subseteq [1, x] \setminus F \text{ with } \sum_{n \in S} n^{-1} = 0 \text{ in } \mathbb{F}_p\}.$$

The largest prime factor is $P^+(n)$, with $P^+(1) = 1$. The Dickman function ρ is one on $[0, 1]$, is continuous on $[0, \infty)$, and obeys $v\rho'(v) = -\rho(v - 1)$ for $v > 1$. It is positive, strictly decreasing for $v > 1$, and tends to zero. In particular $1 - \rho(v) = \log v$ for $1 \le v \le 2$. Write

$$f(v) = 1 - \rho(v), \qquad \mathcal{C}(a) = \int_0^\infty (a - f(v))_+ \, dv \quad (0 < a < 1),$$

where $t_+ = \max(t, 0)$. The integral is finite.

Theorem 1.1 (Local blocking threshold)**.** *For fixed $v > 0$ and $x = \lfloor (\log p)^v \rfloor$,*

$$\sigma(p, x) = (1 - \rho(v) + o(1))x. \tag{1.1}$$

For $v > 1$ and fixed $0 < \epsilon < 1 - \rho(v)$, suppose $F \subseteq [1, x]$ has $|F| \le (1 - \rho(v) - \epsilon)x$. Then there is a common pool $C \subseteq [c_\epsilon x, x] \setminus F$ of size $O_{v,\epsilon}(x/(\log x)^{1/20})$ whose reciprocal subset sums represent every residue modulo p. The zero residue has a nonempty representative. For fixed $v < 1$, one has $\sigma(p, x) = 0$ eventually.

Corollary 1.2 (Local stability)**.** *If $U \subseteq [1,x]$ is zero-sum-free for modular inverse sums and $|U| \geq (\rho(v) - o(1))x$, then*

$$|U \triangle \{n \leq x : P^+(n) \leq \log p\}| = o(x).$$

A recent preprint of Howe, Huffman, Kohlmeyer and Kruer [4, Lemma 3.3] states a uniform sparsity bound for nonsmooth elements in inverse-zero-free sets, for general moduli. In the prime case that bound, summed over dyadic intervals and combined with the elementary smooth-number construction below, implies the limiting deletion fraction and local stability above. We do not use that preprint as an input: our argument also supplies a sparse common pool for all residues, which is needed in the global converse.

This stability conclusion is at the scale x, not a characterization of exact finite maximizers. For $x \asymp \log p$, the relative deletion proportion tends to zero; the first nonempty zero sum at that scale is a different question, discussed in Section 7.

Let $E(N)$ be the maximum size of a subset of $[1,N]$ containing no subset with reciprocal sum one, and let $h(N) = N - E(N)$ be the corresponding minimum deletion number. Erdős and Graham [2] posed the extremal question (Problem 300 in the current index). Croot [3] disproved their conjecture that $E(N) \sim N$, and Liu and Sawhney [1, Theorem 1.3 in arXiv v1] proved $E(N) = (1 - e^{-1} + o(1))N$. Our aim is to determine the matching second-order term. Set

$$L = \log N, \qquad \lambda = \log L, \qquad \delta_N = \lambda/L, \qquad a = 2/e.$$

Theorem 1.3 (Second-order unit-sum-free asymptotics)**.** *The matching second-order formula is*

$$E(N) = (1 - e^{-1})N + \mathcal{C}(2/e)\, N\delta_N + o(N\delta_N). \tag{1.2}$$

In particular the coefficient is determined by the unique $v_* > 1$ with $\rho(v_*) = 1 - 2/e$:

$$\mathcal{C}(2/e) = \int_0^{v_*} (2/e - 1 + \rho(v))\, dv.$$

For reference, $v_* = 2.091155468101753\ldots$ and $\mathcal{C}(2/e) \approx 1.0871215914020186$. Any sequence of unit-sum-free sets of cardinality $(1 - e^{-1} + o(1))N$ differs by $o(N)$ points from $(N/e, N] \cap \mathbb{N}$.

The lower bound on $E(N)$ in (1.2) is elementary once the smooth-number asymptotic is known; it is proved separately in Section 4. The matching upper bound is new relative to the earlier one-sided note on this problem and is the purpose of Sections 5–6. In particular the coefficient $2/e$ is strictly greater than $\log 2$, so the cost profile cannot be truncated to $\log v$ for $v \leq 2$ and a constant thereafter. We must use the full Dickman function. The proof is based on a scalar version of the arithmetic phase-transition method, with all required completion estimates reproduced here. For the numerical evaluation, $\rho(2) = 1 - \log 2 > 1 - 2/e$ and $\rho(v) = 1 - \log v + \int_2^v \log(t-1)\, dt/t$ on $[2,3]$. Using $\log(t-1) \leq t - 2$ and $t \geq 2$ gives $\rho(3) \leq 1 - \log 3 + 1/4 < 1 - 2/e$. Thus $2 < v_* < 3$. Integration by parts using $v\rho'(v) = -\rho(v-1)$ yields

$$\mathcal{C}(2/e) = \int_0^{v_*-1} \rho(t)\, dt = 2(v_* - 1) - (v_* - 1)\log(v_* - 1) - 1. \tag{1.3}$$

A rational interval check, supplied in the accompanying calculation, uses the series for $\log t$ and e, and the expansion of $\log(1+s)/(2+s)$ for $0 \leq s < 1$. This gives the displayed numerical values without affecting the proof of the asymptotic formula. No claim of priority over all other unpublished work is made.

2 Arithmetic estimates and an exact completion lemma

We use the following standard prime estimates (with $t \to \infty$):

$$\log \operatorname{lcm}(1, \dots, \lfloor t \rfloor) \ll t, \qquad \sum_{p \le t} p^{-1} = \log\log t + c_M + O(1/\log t), \qquad \pi(t) \ll t/\log t, \tag{2.1}$$

$$\sum_{n \le t} (3/2)^{\Omega(n)} \ll t(\log(2t))^{1/2}, \qquad \sum_{p^j > y,\ j \ge 2} p^{-j} \ll y^{-1/2}. \tag{2.2}$$

Here Ω counts prime factors with multiplicity. The moment bound follows by writing $(3/2)^{\Omega}$ as $1 * g$ with nonnegative multiplicative $g(p^j) = \frac{1}{2}(3/2)^{j-1}$, then applying Mertens' estimate to $\prod_{p \le t}(1 + \sum_{j \ge 1} g(p^j)/p^j)$. For the last bound split at $p = \sqrt{y}$: below it the geometric tail for each prime is $O(1/y)$; above it sum $O(p^{-2})$ over all integers. The fixed-parameter smooth-number estimate, in the uniform form of [5, Theorem 1.1], is

$$\Psi(t, y) := \#\{n \le t : P^+(n) \le y\} = (\rho(\log t/\log y) + o(1))t \tag{2.3}$$

for $y, t \to \infty$ with the ratio in a fixed compact subset of $(0, \infty)$; the case $y \ge t$ is immediate. The bound is uniform for bounded $\log t/\log y \ge 1$. We also use the standard upper-bound sieve (e.g. [1, Lemma 2.4, arXiv v1]): the count of integers in $[1, t]$ avoiding all primes in a set $\mathcal{P}$ below $\exp(\log t/\sqrt{\log\log t})$ is $\ll t \prod_{p \in \mathcal{P}}(1 - 1/p)$, with an absolute constant. Only the two widely separated scales below require this input.

For a finite pool P let $Q = \operatorname{lcm}(P)$, and say a power s appears if it is a prime power dividing some element of P. We use two regimes with $X \to \infty$, $l = \log X$:

Fine regime. $m = X/l^3$, $Y = X/l^B$ for a sufficiently large fixed B. The pool lies in $[m, X]$, has $|P| \ge c_0 X$, and $|P \cap s\mathbb{N}| \ge \eta X/s$ for every appearing power, where $c_0, \eta > 0$ are fixed.

Coarse regime. $m = X/l^{1/100}$, $Y = Xe^{-12l^{9/10}}$, $P \subseteq [m, X]$, $|P| \ge Xl^{-3/50}$, and $|P \cap s\mathbb{N}| \ge l^{-7/100}X/s$ for every appearing power.

In both regimes all prime-power divisors are at most Y and $\Omega(n) \le 20 \log l$ for each $n \in P$. Thus $\log Q \ll Y$. The following lemma is a scalar specialization of the synchronized-carrier argument; the proof is included to avoid assuming any precise-cardinality or finer-scale external grid theorem.

Lemma 2.1 (Synchronized carriers)**.** *For any appearing power s, after removing $o(X/s)$ of its multiples in the fine regime or $o(l^{-7/100}X/s)$ in the coarse regime, the remaining s-multiples contain a fiber $d_s U_s$ with $s \mid d_s$, $U_s \subseteq [1, z_s]$, $z_s = X/d_s$. There are fixed constants $\nu, J, d, \kappa > 0$ in the fine regime such that*

$$l^{d\nu} \le z_s \le l^{3+dJ}, \qquad |U_s| \ge \kappa z_s.$$

In the coarse regime one can ensure

$$e^{l^{7/10}} \le z_s \le l^{1/100} e^{l^{9/10}}, \qquad |U_s| \gg l^{-27/100} z_s.$$

In each regime any two such fibers have common primes dividing some member of each of their cofactor sets U_s, with product exceeding X. All statements are uniform for the indicated small deletions; the fine constant B is chosen sufficiently large after d, J, ν.

Proof. Let $t = X/s$, and consider the surviving cofactor set $T \subseteq [1, t]$. For any y let $k_y = \prod_{p \le y} p^{v_p(k)}$. The sum of $\log k_y$ over $k \le t$ is $\ll t \log y$, by summing valuations and using the prime estimates.

In the fine regime take $y = l^d$ with $d \geq 1$. For $t \geq y^3$, if $\mathcal{P}$ consists of any subcollection of primes up to y with reciprocal sum w, elementary expansion and counting multiples give

$$\sum_{k \leq t} \left(\sum_{p \in \mathcal{P}} \mathbf{1}_{p|k} - w \right)^2 \leq tw + 2w|\mathcal{P}| \leq 2tw \tag{2.4}$$

for large t (the last inequality also holds whenever $2|\mathcal{P}| \leq t$). Indeed the off-diagonal count is at most $t/(pp')$, the diagonal count at most t/p, and the first-moment floor errors cost at most $2w|\mathcal{P}|$. Choose $\nu > 0$ small enough that the reciprocal sum on $[y^\nu, y]$ is larger than $16/\eta$; its limit is $\log(1/\nu)$. By (2.4), fewer than $(\eta/8)t$ integers can avoid all these primes. Choose a fixed J large enough, using the logarithmic moment, that at most $(\eta/8)t$ have $k_y > y^J$. A fixed positive fraction of t remains in T with $y^\nu \leq k_y \leq y^J$, since $|T| \geq (\eta - o(1))t$.

Partition these cofactors by their unique y-rough part $j = k/k_y$. A nonempty fiber has $z = t/j$ between y^ν and $l^3 y^J$, since $k \geq (m/X)t$. The number of y-smooth integers up to z is at least $c_J z$ uniformly in this range, independently of $d \geq 1$, by (2.3) since $\log z / \log y \leq J + 3$ (and the case $z \leq y$ is trivial). The full smooth fibers for distinct rough parts are disjoint in $[1, t]$, hence $\sum_j z_j \ll_J t$. At least one surviving fiber has density $\kappa > 0$ depending only on η, J, ν. Take $d_s = sj$. This order of choosing constants makes κ independent of the later choice of d.

In the coarse regime $t \geq e^{12l^{9/10}}$. Take $y = e^{l^{4/5}}$. The upper-bound sieve and Mertens show that $O(tl^{-1/10})$ integers lack every prime in $[e^{l^{7/10}}, y]$; this is within the sieve range because $\log t \geq 12l^{9/10}$. The logarithmic moment removes another $O(tl^{-1/10})$ integers with $k_y > e^{l^{9/10}}$. These losses are $o(tl^{-7/100})$. The same rough-part partition yields the asserted interval of z values. The sum of $1/j$ over y-rough parts $j \leq t$ is bounded above by $\prod_{y<p\leq t}(1 - 1/p)^{-1} \ll l^{1/5}$. Thus $\sum_j z_j \ll tl^{1/5}$ and a fiber has density $\gg l^{-7/100-1/5}$.

For fine synchronization use primes in $[l^2, l^{d\nu/4}]$. Applied to a cofactor set of density at least κ, (2.4) implies that the reciprocal sum of the primes it misses is at most $2/\kappa$: indeed at least κz_s terms in the square sum equal w^2, and these primes are at most $z_s^{1/4}$. Choose d large enough that the total reciprocal sum, tending to $\log(d\nu/8)$, exceeds $4/\kappa + 2$. Then the common primes for two fibers have reciprocal sum at least one for large X, and each is at least l^2. Their product exceeds X. Now choose $B > 3d$ to ensure $t \geq y^3$ throughout.

For coarse synchronization use primes in $[e^{l^{1/20}}, e^{l^{3/5}}]$. The upper-bound sieve applies since $z_s \geq e^{l^{7/10}}$: its logarithmic cutoff is at least $l^{7/10}/\sqrt{\log\log z_s}$, which exceeds $l^{3/5}$ for large X. If a set of density $\gg l^{-27/100}$ misses primes of reciprocal sum w, the sieve gives $l^{-27/100} \ll e^{-w}$; thus $w \leq (27/100)\log l + O(1)$. The total reciprocal sum is $(11/20)\log l + o(1)$, so two fibers have common prime reciprocal sum at least $(1/100)\log l - O(1)$. Since the primes are all at least $e^{l^{1/20}}$, their product exceeds X. This proves the lemma in both regimes. □

Lemma 2.2 (Scalar exact completion)**.** *In either regime, if $u \in Q^{-1}\mathbb{Z}$ and*

$$l^{-4} \leq u/R(P) \leq 1 - l^{-4},$$

then there is a subset of P of reciprocal sum u. More precisely, for independent Bernoulli variables X_n of the common parameter $u/R(P)$, with $Z = \sum_{n \in P} X_n/n$, one has

$$\mathbb{P}(Z = u) \sim \frac{1}{Q\sqrt{2\pi \operatorname{Var} Z}}. \tag{2.5}$$

In the fine regime B is fixed sufficiently large in terms of c_0, η only. The bounds are uniform in P, u satisfying the conditions.

Proof. Write $e(t) = e^{2\pi it}$ and $\|t\| = \operatorname{dist}(t, \mathbb{Z})$. The centered characteristic function is $\phi(\xi) = \mathbb{E}e(\xi(Z - u))$. The Bernoulli product estimate gives

$$|\phi(\xi)| \le \exp(-cl^{-4} \sum_{n \in P} \|\xi/n\|^2). \tag{2.6}$$

In the fine regime choose $w > 3 + dJ + 2$, then put $W = X/l^w$ and $T_* = l^{2w+10}$; increase B so $B > 2w + 12$ as well as $B > 3d$. In the coarse regime put $W = Xe^{-2l^{9/10}}$, $T_* = e^{5l^{9/10}}$. In either case every carrier in Lemma 2.1 has $d_s > W$, and T_* is negligible relative to the minimum allowed size of a power fiber.

At a frequency $\xi \in [-Q/2, Q/2]$, call an appearing power bad if at least T_* of its multiples in P have distance from ξ to their own multiples at least $W/2$. If there are j bad powers, at least $jT_*/(20 \log l)$ distinct integers witness badness, since each is divisible by at most $20 \log l$ prime powers. Each witness has $\|\xi/n\| \ge W/(2X)$. By (2.6) the modulus is then bounded by

$$\exp(-cjT_* l^{-4} W^2/(X^2 \log l)) \le X^{-20j}$$

for large X. The exponent is at least of order $jl^6/\log l$ in the fine regime and at least $cje^{l^{9/10}}/(l^4 \log l)$ in the coarse regime. Both dominate $20j \log X$ uniformly.

For a good power remove its fewer than T_* witnesses. Lemma 2.1 gives a fiber $d_s U_s$. For each member of the fiber there is a multiple within $W/2$ of ξ. Since all such multiples are themselves multiples of $d_s > W$, they must equal one integer h_s. Each common prime of two cofactor fibers divides the difference of their corresponding integers. Their product exceeds $X > W$, hence all the h_s agree. Thus ξ lies within $W/2$ of a multiple of the lcm Q' of the good powers (take $Q' = 1$ if there are none). In a period of length Q the measure of these frequencies is at most $(W + 1)Q/Q'$, and $Q/Q' \le X^j$, since all the missing prime-power contributions can be bounded by the product of the bad powers. There are at most X^j choices of a set of j bad powers. Consequently the integral over $j \ge 1$ is $O(\sum_{j \ge 1} X^{1-18j})$.

When $j = 0$, the frequency lies within $W/2$ of zero modulo Q. On the chosen period this is $|\xi| \le W/2$. Since $W/m = o(1)$ the phases are unwrapped here. Write $V = \operatorname{Var} Z$. We have $V \gg |P|/(l^4 X^2)$ and $V \le X/m^2$. On $|\xi| \le m^{3/5}$ the centered Bernoulli expansion gives $\phi(\xi) = e^{-2\pi^2 V \xi^2}(1 + o(1))$ uniformly: the total cubic error is $O(Xm^{-6/5}) = o(1)$. This interval lies inside $[-Q/2, Q/2]$ since $Q \ge \max P \ge |P|$, and inside the central $W/2$ interval for large X. Even if some such frequencies are classified as bad, the product bound above controls their contribution; in fact each is good because its distance from zero is less than $W/2$. On $m^{3/5} \le |\xi| \le W/2$ the unwrapped bound gives $|\phi(\xi)| \le \exp(-cl^{-4}|P|\xi^2/X^2)$. At the lower endpoint its exponent in magnitude is $\gg X^{1/5} l^{-38/5}$ in the fine regime and $\gg X^{1/5} l^{-4-3/50-3/250}$ in the coarse regime. Since $W \le X$, its integral is superpolynomially small. The same estimate controls the Gaussian tails. The full Gaussian integral is $1/\sqrt{2\pi V} \gg m/\sqrt{X}$, hence the preceding absolute errors are negligible compared with it. Continuous inversion on the actual lattice $Q^{-1}\mathbb{Z}$, using that $Qu \in \mathbb{Z}$, yields (2.5) by integrating ϕ over one period and dividing by Q. In particular the probability is positive. □

3 Pruned grids and local Dickman laws

Lemma 3.1 (Pruned rational grid)**.** *Fix $0 < c \le 1$ and $c' > 0$. Let $A_0 \subseteq [cX, X]$ have size at least $c'X(\log X)^{-1/20}$. One can remove $o(X(\log X)^{-1/20})$ elements to obtain A, such that, with $Q = \operatorname{lcm}(A)$, every $u \in Q^{-1}\mathbb{Z}$ in the interval*

$$R(A)/\log X \le u \le R(A)/(1 + 1/\log X) \tag{3.1}$$

is a reciprocal subset sum of A. In particular its central third of grid points is filled.

Proof. Write $l = \log X$. Remove integers having a prime-power divisor above $Xe^{-12l^{9/10}}$ or $\Omega(n) > 20 \log l$. By (2.1)–(2.2) this costs $O(Xl^{-1/10})$. Repeatedly remove all current multiples of an appearing power s when fewer than $l^{-7/100}X/s$ survive. Each power triggers at most once. The total cost is $O(Xl^{-7/100} \log l) = o(Xl^{-1/20})$, by summing reciprocals of all prime powers. The final set has size at least $(c'/2)Xl^{-1/20} \geq Xl^{-3/50}$ for large X. It lies above $X/l^{1/100}$ eventually, and satisfies all the coarse hypotheses. For a target in (3.1), its normalized mass lies in $[l^{-4}, 1 - l^{-4}]$. Lemma 2.2 gives the representation. □

The pruning and scalar inversion are an alternative to applying the more general Proposition 5.2 of Liu–Sawhney [1]. We include them so that the later fine completion step has a common proof, rather than assuming a finer error estimate from their fixed-scale statement.

Lemma 3.2 (Sparse pool for modular repair)**.** *Fix $\epsilon > 0$. If $U \subseteq [1, X]$ has size at least ϵX, and each member has a prime divisor exceeding $z \geq 2$, there exists $C \subseteq U \cap [c_\epsilon X, X]$ with*

$$|C| \asymp X(\log X)^{-1/20}, \quad R(C) \asymp_\epsilon (\log X)^{-1/20}, \quad \log \operatorname{lcm}(C) \gg z \log z \, (\log X)^{-1/20}.$$

All grid points of the central third of $[0, R(C)]$ are filled. If $p > X$ is prime and $R(C) \operatorname{lcm}(C)/p \to \infty$, then the inverse subset sums of C cover all residues modulo p, with a nonempty representative for zero.

Proof. Discard denominators below $\epsilon X/4$ and choose $\left\lfloor X(\log X)^{-1/20} \right\rfloor$ of the remaining integers. Apply Lemma 3.1, retaining the asserted order of size and mass. Each prime above z divides at most X/z of the selected integers. Choosing one such prime factor from each integer proves that the lcm contains at least $|C|z/X$ distinct primes above z, yielding the logarithmic bound without any assumption about multiple large factors. The consecutive grid numerators in the positive central third contain a complete residue system modulo p under the ratio hypothesis. The denominator $\operatorname{lcm}(C)$ is coprime to p, so the represented sums span every residue. Positivity ensures nonemptiness even for zero. □

Proof of the upper bound in Theorem 1.1. Put $L_p = \log p$ and $y = cL_p$ for an absolute constant $c > 0$ sufficiently small. Keep only integers whose prime-power divisors are all at most y. Their lcm divides $Q_0 = \operatorname{lcm}(1, \ldots, \lfloor y \rfloor)$. By (2.1), $Q_0 H_x < p$ for large p, where $H_x = \sum_{n \leq x} 1/n$. Every nonempty sum has a positive numerator at most $Q_0 H_x$, so it cannot vanish modulo p. The retained set differs from the y-smooth integers by at most $O(x/\sqrt{y}) = o(x)$, by the prime-power tail in (2.2). Equation (2.3) gives its size as $(\rho(v) + o(1))x$. For fixed $v < 1$ all denominators are already below y eventually. This gives the upper bound and the exact assertion below exponent one. □

Proof of the lower bound and coverage. The lower bound for $v \leq 1$ is trivial. For $v > 1$ put $L_p = \log p$ and $z = L_p$. If $|F| \leq (f(v) - \epsilon)x$, the smooth-number estimate implies that at least $(\epsilon/2)x$ survivors have a prime factor above z for all sufficiently large p. Apply Lemma 3.2 with $X = x$ to these survivors. Its sparse pool satisfies

$$\log \operatorname{lcm}(C) \gg_\epsilon \frac{L_p \log L_p}{(\log x)^{1/20}} \gg_{v,\epsilon} L_p (\log L_p)^{19/20}, \qquad |\log R(C)| = O_\epsilon(\log \log x).$$

Thus $R(C) \operatorname{lcm}(C)/p \to \infty$ and its positive grid spans all residues, with a nonempty representative for zero. This proves the pool property and the lower bound since ϵ is arbitrary. □

Proof of Corollary 1.2. A zero-sum-free set U has only $o(x)$ elements with a prime factor above $\log p$. Otherwise a positive-density subsequence of those elements alone supplies a sparse modular pool by Lemma 3.2, with the preceding logarithmic lcm estimate. For $v \le 1$ there are no such elements. Its cardinality hypothesis and the asymptotic cardinality of the $(\log p)$-smooth set imply that only $o(x)$ smooth integers are missing. This gives the stated symmetric difference. □

Remark 3.3 (Uniformity on bounded exponent intervals). The estimates for the Dickman profile are uniform when v ranges in any fixed compact interval of positive exponents. To see this, the upper-bound construction uses uniform (2.3); for the lower bound it is enough to consider exponents with $f(v) \ge \epsilon$. The argument with $z = \log p$ works uniformly there: the smooth-number estimate is uniform and $\log x \ll \log_2 p$, so the lcm margin $\gg_\epsilon (\log p)(\log_2 p)^{19/20}$ holds uniformly for v in the given compact interval. Near exponent one the desired error follows directly from the upper bound and nonnegativity. This uniformity will be used for the large-prime blocks below.

4 The elementary second-order lower bound

Retain L, λ, δ_N, a from Section 1 and write $\delta = \delta_N$ in this and the following sections. For a fixed $B > 0$ set $v_p = \log(N/p)/\lambda$ for primes $p > N/L^B$. We shall use the scale integration rule

$$\sum_{N/L^B < p \le N} \frac{g(v_p)}{p} = \delta \int_0^B g(v)\, dv + o(\delta) \tag{4.1}$$

for each fixed continuous g on $[0, B]$, or an interval indicator. For an interval $[b_1, b_2]$ this follows from Mertens' theorem by taking the difference of $\log\log(N/L^{b_i})$, which equals $(b_2 - b_1)\lambda/L + O_B(\lambda^2/L^2)$ with the appropriate order of the endpoints. The prime harmonic error is $O(1/L) = o(\delta)$ and endpoint primes cost $o(\delta)$. Finite partitions and uniform approximation prove the continuous case. In sums of the form $\sum_{p > N/L^B} \lfloor N/p \rfloor\, g(v_p)$, the floor error is $O_g(\pi(N)) = o(N\delta)$.

Proposition 4.1 (Explicit deletion construction)**.** *For every fixed $V > 1$ there is a unit-sum-free set of size at least*

$$(1 - e^{-1})N + N\delta \int_0^V (2/e - f(v))\, dv + o(N\delta).$$

In particular the lower bound in Theorem 1.3 holds with coefficient $\mathcal{C}(2/e)$, without using any of the completion lemmas.

Proof. For every prime $p > N/L^V$ delete all $pm \le N$ for which the cofactor m has a prime-power divisor exceeding cL, with a sufficiently small fixed $c > 0$. Call this actual deletion set F_V, and let V_p^{all} denote the union of all full multiple classes of the attacked primes. These classes are disjoint for large N, since $p > \sqrt{N}$. Any subset of the surviving integers of reciprocal sum one would have to avoid the entire union V_p^{all}: multiplying its equation by a given p and reducing modulo p forces the sum of the inverses of its used cofactors to be zero. All retained cofactors divide $Q_0 = \mathrm{lcm}(1, \dots, \lfloor cL \rfloor)$, and a nonempty cofactor sum has positive numerator at most $Q_0 H_{\lfloor L^V \rfloor} < p$ by (2.1), so zero modulo p is impossible.

For $p > N/L^V$, the excluded fraction among its $\lfloor N/p \rfloor$ cofactors is $f(v_p) + o(1)$ uniformly when v_p stays in any fixed compact interval of positive exponents. This follows from (2.3), since $\log(cL) = \lambda + O(1)$ and the higher prime-power tail costs $o(1)$ in relative count. For $v_p \le 1/2$ the

cofactor bound is less than cL for large N, so nothing is excluded. Uniform continuity and the floor estimate therefore give, by (4.1),

$$|F_V| = N\delta \int_0^V f(v)\,dv + o(N\delta). \tag{4.2}$$

For each prefix $[1,t] \subseteq [1,N]$, summing the prime multiple counts gives uniformly $|V_p^{\mathrm{all}} \cap [1,t]| = V\delta t + o(N\delta)$, with the error from floors $O(\pi(N))$. Write $d = 1 - V\delta$. Thus the effective set outside the entire attacked union has counting function $dt + o(N\delta)$, uniformly in t. By partial summation the reciprocal mass of effective denominators in $(tN, N]$ for t in a fixed neighborhood of $1/e$ is $d\log(1/t) + o(\delta)$, uniformly. Delete its initial segment up to $N\exp(-1/d) + o(N\delta)$, taking a positive displacement $o(N\delta)$ sufficient to ensure that the mass of the remaining effective denominators is strictly less than one. This is possible since the mass derivative in t is bounded away from zero in magnitude, and the uniform error is $o(\delta)$. These additional actual deletions lie outside the attacked union and cost

$$dNe^{-1/d} + o(N\delta) = N/e - (2/e)VN\delta + o(N\delta).$$

Any exact unit sum would have to lie entirely in the remaining effective set, whose total reciprocal mass is less than one. Combining its cost with (4.2) proves the claimed construction. The integrand $a - f(v)$ is strictly positive up to v_* and negative afterwards, so the best fixed V is v_*, giving $\mathcal{C}(a)$. This is strictly beyond two since $a = 2/e > \log 2 = f(2)$. □

5 First-order stability of minimum transversals

Call $F \subseteq [1,N]$ a transversal if its complement contains no unit sum. For the converse we need control of the location of a near-minimal transversal, not merely its cardinality. We first give a scalar coarse argument recovering first-order stability. It uses no fine approximation and also supplies the first-order cardinality threshold.

Lemma 5.1 (Initial-segment stability)**.** *If F is a transversal with $|F| = (e^{-1} + o(1))N$, then*

$$|F \triangle [1, \lfloor N/e \rfloor]| = o(N).$$

Moreover any transversal has $|F| \geq (e^{-1} - o(1))N$.

Proof. For the second claim, if $|F| \leq (e^{-1} - \epsilon)N$ for fixed $\epsilon > 0$ on an infinite subsequence, discard all remaining integers below cN with $0 < c < \epsilon/2$. The total reciprocal mass of the surviving set is at least the tail harmonic mass obtained by removing the smallest $|F| + \lceil cN \rceil$ integers; hence it exceeds one by a fixed margin. It has positive density and lies in $[cN, N]$. Lemma 3.1 removes $o(N)$ additional points, with reciprocal mass loss $o(1)$. On the resulting pool A the target one is on the actual rational grid (it is an integer), and $R(A) > 1 + \epsilon'$ for some fixed $\epsilon' > 0$, while $R(A) = O_c(1)$. Thus one satisfies (3.1), a contradiction. The opposite first-order bound follows from the elementary tail attack (delete $\lfloor N/e \rfloor + o(N)$ smallest integers so that the remaining mass is strictly less than one); it is also the classical Liu–Sawhney theorem.

For stability put $h = |F|$. If $|F \triangle [1,h]|/2 \geq \kappa N$ for a fixed $\kappa > 0$ on a subsequence, let $F' = F \cup [1, \lfloor cN \rfloor]$, $h' = |F'|$. By the triangle inequality, the displacement $d' = |F' \triangle [1,h']|/2$ is at least $(\kappa - c)N$. Relative to the upper tail $[h'+1, N]$, the survivor set $[1,N] \setminus F'$ restores d' smaller denominators and removes d' larger denominators. By monotonicity of reciprocals, the minimum possible increase of mass is

$$\sum_{i=1}^{d'} \left(\frac{1}{h' - i + 1} - \frac{1}{h' + i} \right) \geq \frac{(d')^2}{N^2}.$$

Thus its mass is at least $\log(N/h') + (\kappa - c)^2 + o(1)$. Since $h'/N \le e^{-1} + c + o(1)$, taking $c > 0$ sufficiently small in terms of κ makes this exceed one by a fixed positive margin. The same coarse grid pruning supplies an exact unit sum, a contradiction. Therefore $|F \triangle [1, h]| = o(N)$, and $h = N/e + o(N)$ gives the lemma. □

6 A matching global converse

We prove that every minimum transversal has

$$h(N) \ge N/e - \mathcal{C}(a)N\delta - o(N\delta). \tag{6.1}$$

We give the proof in detail because a local Dickman law alone does not imply this estimate. The essential step is to make the total cost of arithmetic repairs smaller than the second-order scale while keeping a healthy global rational grid.

Let F be a minimum transversal, $h = |F|$, and $F_0 = [1, \lfloor N/e \rfloor]$. By Lemma 5.1 and the elementary tail attack, $h = N/e + o(N)$ and $|F \triangle F_0| = o(N)$. Fix $\eta > 0$ small, with $\eta < (1 - a)/4$. Choose the fine-regime constant B in Lemma 2.2 for density bound $c_0 = (1 - e^{-1})/2$ and local bound $\eta/3$, also taking $B > v_* + 1$ and $B > 4$. All these choices precede the later choices of the small repair parameters. Put $Y = N/L^B$ and call primes $p > Y$ large; their full multiple classes are pairwise disjoint. Denote their union in $[1, N]$ by V_{all}. By (4.1), $|V_{\text{all}}| = O_B(N\delta)$.

6.1 Marking damaged powers at the small scale

Define E_0 to be the integers below N/L^3, those divisible by a prime power $p^j > Y$ with $j \ge 2$, and those with $\Omega(n) > 20\lambda$. Equations (2.1)–(2.2) imply

$$|E_0| = o(N\delta). \tag{6.2}$$

In particular the higher-power contribution is $O(NY^{-1/2})$ and the multiplicity contribution is $O(NL^{1/2-20\log(3/2)})$. Write

$$G = [1, N] \setminus (F \cup E_0 \cup V_{\text{all}}).$$

Mark each prime power $s \le Y$ for which $|G \cap s\mathbb{N}| < \eta N/s$, including powers absent from G. For each base prime retain just its least marked power. The resulting set $\mathcal{B}$ is pairwise coprime. Put

$$T = \sum_{s \in \mathcal{B}} 1/s, \qquad V_s = \bigcup_{s \in \mathcal{B}} (s\mathbb{N} \cap [1, N]), \qquad P = G \setminus V_s.$$

Here T denotes a scalar sum, and the subscript s on V_s stands for the small-power scale, not a single member of $\mathcal{B}$.

Lemma 6.1 (Small-power health and cost)**.** *The marked powers satisfy $T = o(1)$. Every prime power $q \le Y$ not divisible by any $s \in \mathcal{B}$ occurs in P, with*

$$|P \cap q\mathbb{N}| \ge \eta N/(3q). \tag{6.3}$$

The actual lcm $Q = \operatorname{lcm}(P)$ contains every such power. Moreover

$$|F \cap V_s| \ge (1 - \eta - o(1))NT - o(N\delta). \tag{6.4}$$

Proof. The initial template F_0 leaves at least $(1-e^{-1})N/s - O(1)$ multiples of s outside itself. For $s \le Y$ the floor error is $O(L^{-B}N/s)$. The count of multiples of s that lie also in large-prime pools is at most

$$\sum_{p>Y} \lfloor N/(ps) \rfloor \ll_B (N/s)\delta,$$

because all such primes are coprime to s. To mark s one must therefore lose at least $(1-e^{-1}-\eta-o(1))N/s$ multiples to $(F \setminus F_0) \cup E_0$. Summing over $s \in \mathcal{B}$ is safe if incidences are counted with multiplicity. For $g(n) = \sum_{s\in\mathcal{B}} \mathbf{1}_{s|n}$, pairwise coprimality gives

$$\sum_{n\le N} g(n)^2 \le N(T+T^2).$$

Cauchy–Schwarz bounds the total incidences in $(F \setminus F_0) \cup E_0$ by $\sqrt{o(N)\,N(T+T^2)}$. Consequently, writing $M = |(F \setminus F_0) \cup E_0| = o(N)$, we have $cNT \le \sqrt{MN(T+T^2)}$ for a fixed $c > 0$. If $T > 0$, squaring and dividing by N^2T yields $(c^2 - M/N)T \le M/N$. Thus $T = o(1)$; no second-order bound on T has yet been used. This argument does not assume pointwise negligible exceptional losses in every small-power class.

If q is not divisible by a retained marked power, it was not marked, so $|G \cap q\mathbb{N}| \ge \eta N/q$. Removing the multiples of all other-base marked powers costs at most $(N/q)T$. A possible marked power of the same base prime is at least twice q, and has fewer than $\eta N/(2q)$ survivors in G. Removing these too leaves at least $(\eta/2 - T)N/q \ge \eta N/(3q)$ for large N. This proves (6.3). In particular, as $N/q \ge L^B$, all these powers appear in the actual lcm.

The union V_s has size $NT + O(NT^2 + YT)$, by Bonferroni and $|\mathcal{B}| \le YT$. Its intersection with G has at most ηNT elements. Its intersection with V_{all} has at most $O_B(NT\delta)$ elements, and the remaining exceptional count is at most $|E_0|$. Subtracting these counts from $|V_s|$ proves (6.4), because $T = o(1)$ and (6.2) holds. □

The pool P now meets the fine hypotheses of Lemma 2.2: it has size $(1-e^{-1}+o(1))N$ since the additional exclusions are $o(N)$, its smallest denominator is at least N/L^3, all its powers are at most Y, its prime-factor multiplicities satisfy the imposed bound, and every appearing power satisfies (6.3). We use its actual lcm Q throughout.

6.2 Large-prime pools and sparse repair

Fix small positive constants γ, ϵ with $1 + \gamma < B$; these will eventually tend to zero with B, η fixed. For each large prime p, put $x_p = \lfloor N/p \rfloor$ and $z = L^{1+\gamma}$. Call the pool healthy if at least ϵx_p cofactors $m \le x_p$ satisfy

$$pm \notin F \cup E_0 \cup V_s, \qquad P^+(m) > z. \tag{6.5}$$

Let $\mathcal{D}$ be the set of failed large primes. In a healthy pool, Lemma 3.2 gives a reserve C_p of size $O_\epsilon(x_p(\log x_p)^{-1/20})$ with cofactors at least $c_\epsilon x_p$. Indeed $x_p > z$, $p > x_p$, and

$$\log \operatorname{lcm}(C_p) \gg_\epsilon \frac{L^{1+\gamma}\log L}{(B\lambda)^{1/20}} \gg L$$

with ratio tending to infinity, so it covers every residue modulo p. Given any subset of the allowed p-multiples outside this reserve, we may therefore add a subset of pC_p to cancel its cofactor sum

modulo p. Because distinct large primes have disjoint multiple classes, these repairs never reuse any denominator. The total reserve size and reciprocal mass satisfy

$$\sum_{p \text{ healthy}} |C_p| = O_{B,\epsilon}(N\delta\lambda^{-1/20}) = o(N\delta), \qquad \sum_{p \text{ healthy}} R(pC_p) = o(\delta). \tag{6.6}$$

For the first bound use $x_p > L^{1+\gamma}$ and $\sum_{p>Y} x_p = O_B(N\delta)$; for the second use $pm \geq c_\epsilon p x_p \gg_\epsilon N$ for each reserved denominator. These are cardinality and mass bounds simultaneously, not just bounds on the cost of one congruence.

For every cofactor of an allowed denominator pm in a healthy pool, including those used in its reserve, all prime powers dividing m are at most $L^B < Y$. None is divisible by a retained marked power: otherwise $pm \in V_s$, contrary to the definition of the allowed denominators. They consequently divide Q by (6.3), hence $m \mid Q$. Also $p \nmid Q$. When the selected cofactors of a block are corrected to residue zero, the sum $Q\sum 1/m$ is an integer divisible by p. Dividing by p shows that the reciprocal mass of the corrected block pm belongs to $Q^{-1}\mathbb{Z}$. This lcm compatibility is crucial: coverage modulo p by itself does not imply a usable exact rational remainder.

6.3 The full Dickman ledger

Define

$$V_p = \bigcup_{p\in\mathcal{D}} (p\mathbb{N} \cap [1, N]), \quad V = V_s \cup V_p, \quad \omega = T + \sum_{p\in\mathcal{D}} p^{-1}, \quad d = 1 - \omega.$$

Again V_p denotes the union of failed pools. It is disjoint from the healthy reserves; $\omega = o(1)$ because $T = o(1)$ and the sum of the large-prime reciprocals is $O_B(\delta)$. We need two estimates, one for actual deletion costs and one for effective vertex counts.

Lemma 6.2 (Cost and effective rank)**.** *With all auxiliary constants fixed in the indicated order,*

$$|F \cap (V_p \setminus V_s)| \geq \sum_{p\in\mathcal{D}} \frac{N}{p} f(v_p) - O_B((\epsilon+\gamma)N\delta) - o(N\delta), \tag{6.7}$$

$$|V \cap [1, t]| = \omega t + o(N(T+\delta)) \qquad (0 \leq t \leq N), \tag{6.8}$$

uniformly in t. In particular

$$|F \cap V| - aN\omega \geq (1 - \eta - a - o(1))NT - \mathcal{C}(a)N\delta - O_B((\epsilon+\gamma)N\delta) - o(N\delta). \tag{6.9}$$

Proof. For each failed prime count all $m \leq x_p$ with $P^+(m) > z$. The uniform smooth-number formula gives a count

$$x_p f(v_p/(1+\gamma)) + o(x_p) \geq x_p\{f(v_p) - O_B(\gamma) - o(1)\}.$$

Here f is Lipschitz on fixed compact intervals (indeed $0 \leq f'(v) \leq 1$ for $v > 1$), and for $x_p \leq z$ the count is zero and $f(v_p) \leq O(\gamma) + o(1)$, making the lower estimate automatic. The replacement of $\log x_p/\lambda$ by v_p is uniform for the positive-count range, where x_p tends to infinity; outside this range the same zero bound applies. Unlike a count based solely on primes greater than $\sqrt{x_p}$, this counts the union of all nonsmooth integers without assuming their large prime factors are unique. Since the pool has failed, fewer than ϵx_p of those cofactors survive outside $F \cup E_0 \cup V_s$. Summing over the disjoint large-prime classes, the losses attributable to E_0 are at most $o(N\delta)$, and those in V_s at most $O_B(NT\delta) = o(N\delta)$. Floor errors contribute $O(\pi(N)) = o(N\delta)$. This proves (6.7) with a constant depending on fixed B.

For the rank count, the small-small overlaps are at most $O(NT^2)$, small-large overlaps at most $O_B(NT\delta)$, and large-large overlaps vanish. Floor errors in any prefix are bounded by $|\mathcal{B}| + \pi(N) \le YT + O(N/L)$. Since $T = o(1)$ and $Y/N = L^{-B}$, their sum is $o(N(T+\delta))$. This proves (6.8). Finally, sum (6.4) and (6.7), then subtract $aN\omega$. The contribution from failed primes is at least

$$-N \sum_{p>Y} \frac{(a - f(v_p))_+}{p} = -N\delta \int_0^B (a - f(v))_+ \, dv + o(N\delta).$$

By the choice $B > v_*$ the integral is exactly $\mathcal{C}(a)$. Thus (6.9) holds. In particular the coefficient of NT is strictly positive: we will not replace an uncontrolled $o(NT)$ term by $o(N\delta)$. □

6.4 Scalar capacity forcing

Let $h' = |F \setminus V|$ and let A be the effective surviving set after also excluding E_0 and every healthy reserve pC_p:

$$A = [1, N] \setminus \left(V \cup F \cup E_0 \cup \bigcup_{p \text{ healthy}} pC_p \right).$$

Its part supported on denominators with no large prime factor is exactly P. Its remaining denominators belong to healthy large-prime classes and lie outside the reserves. Let $A_p = A \cap p\mathbb{N}$ be this base part of a healthy class. Repair it by adding a subset of pC_p to cancel the cofactor sum modulo p (for an empty base part we may take the empty repair). Denote the disjoint union of corrected large-prime blocks by C. Section 6.2 proves $QR(C) \in \mathbb{Z}$, and $C \cap P = \varnothing$. Also

$$R(C) \le \sum_{p>Y} \frac{1}{p} H_{\lfloor N/p \rfloor} = O_B(\lambda\delta) = o(1). \tag{6.10}$$

This bound applies even to small original cofactors; it does not rely on the proportional location of the base denominators. All elements of $P \cup C$ avoid the actual deletion set F.

If $R(A) > 1 + 2L^{-3}$, then $u = 1 - R(C) > 0$ is on $Q^{-1}\mathbb{Z}$, by (6.10) and lcm compatibility. Since the corrected blocks contain all the original large-prime part of A,

$$R(P) - u \ge R(A) - 1 > 2L^{-3}.$$

The harmonic bound $R(P) \le 1 + L$ and $u > 1/2$ imply $u/R(P) \ge 1/(2(1+L)) \ge L^{-4}$. Also $(R(P) - u)/R(P) > 2L^{-3}/(1+L) \ge L^{-4}$ for large N. Thus both hypotheses of the fine completion lemma hold. The fine completion lemma supplies $S \subseteq P$ with $R(S) = u$. The disjoint set $S \cup C$ has reciprocal sum one and avoids F, contrary to transversality. Therefore any minimum transversal must satisfy

$$R(A) \le 1 + 2L^{-3}. \tag{6.11}$$

The excess tolerance $2L^{-3} = o(\delta)$ is important: a fixed positive mass margin would not yield a sharp second-order bound.

We next convert (6.11) into a sharp rank inequality. Before the actual deletions outside V, the effective set $[1, N] \setminus V$ has prefix count $dt + o(N(T+\delta))$ by (6.8). The total additional exclusions from E_0 and the reserves are $o(N\delta)$ by (6.2) and (6.6). Thus at most $k = h' + o(N\delta)$ effective vertices have been removed from it, where one may take the integer upper bound $k = h' + |E_0| + \sum |C_p|$. Here $h' = N/e + o(N)$, since $h = N/e + o(N)$ and $|V| = o(N)$. Among all choices of k deletions the total remaining reciprocal mass is minimized by removing the smallest effective denominators. The cutoff of the kth effective rank is $t_0 = k/d + o(N(T+\delta))$ by the uniform prefix estimate. In

particular t_0/N is bounded away from zero and one. Partial summation on $[t_0, N]$ converts the uniform count error to an $o(T+\delta)$ mass error; explicitly the error is at most a constant times the supremum of the prefix-count error divided by N. It follows that

$$R(A) \geq d \log(dN/h') - o(T+\delta). \tag{6.12}$$

Replacing k by h' here changes the right side by $o(\delta)$ since both are asymptotic to N/e. Combining (6.11) and (6.12) yields

$$h' \geq dN \exp(-1/d) - o(N(T+\delta)) = N/e - aN\omega - o(N(T+\delta)). \tag{6.13}$$

Indeed $(1-\omega)\exp(-1/(1-\omega)) = e^{-1} - (2/e)\omega + O(\omega^2)$ and $\omega = O_B(T+\delta) = o(1)$. Thus the quadratic remainder is $o(T+\delta)$, without requiring $T = O(\delta)$ in advance.

Finally, adding $|F \cap V|$ to (6.13) and using (6.9),

$$h \geq N/e - \mathcal{C}(a)N\delta + (1-\eta-a-o(1))NT - O_B((\epsilon+\gamma)N\delta) - o(N\delta).$$

Since $1-\eta-a > 0$, with $\eta, B, \epsilon, \gamma$ fixed all $o(NT)$ errors can be absorbed in this positive coefficient for sufficiently large N. In particular, dividing by $N\delta$ and dropping the nonnegative NT term gives

$$\liminf_{N\to\infty} \frac{h(N) - N/e}{N\delta} \geq -\mathcal{C}(a) - C_B(\epsilon+\gamma).$$

Only now let $\epsilon, \gamma \downarrow 0$ with η, B unchanged. This proves (6.1). Proposition 4.1 proves the matching bound on $E(N)$, and Lemma 5.1 proves its asserted first-order stability. This completes the proof of Theorem 1.3.

6.5 A stability consequence of the deletion ledger

The proof above also applies to any transversal of size $h(N) + o(N\delta)$, since only first-order stability and the cardinality assumption $h = N/e + o(N)$ were used in the converse. It yields a constraint on the distribution of arithmetic damage at the sharp scale.

Corollary 6.3 (Stability of the arithmetic cost profile)**.** *For an asymptotically optimal transversal F with $|F| = h(N) + o(N\delta)$, fix the same η, B as in the converse. Then the marked small-power sum satisfies $T = o(\delta)$. For the large-prime classification of Section 6.2, put*

$$\mathcal{E}_{\gamma,\epsilon}(N) = \sum_{\substack{p>Y,\ v_p<v_* \\ p\notin\mathcal{D}}} \frac{a - f(v_p)}{p} + \sum_{\substack{p>Y,\ v_p>v_* \\ p\in\mathcal{D}}} \frac{f(v_p) - a}{p}.$$

Then

$$\limsup_{N\to\infty} \frac{\mathcal{E}_{\gamma,\epsilon}(N)}{\delta} \ll_B \gamma + \epsilon.$$

In particular, in the iterated limit $N \to \infty$ followed by $\gamma, \epsilon \downarrow 0$, the failed-prime classification agrees with the cutoff $v_p < v_*$ up to $o(\delta)$ in prime reciprocal weight away from every fixed neighborhood of v_*. This is a cost-profile stability statement, not a description of the exact deleted vertices within each class.

Proof. Instead of discarding the slack in the optimization in Lemma 6.2, use the identity

$$\sum_{p\in\mathcal{D}} \frac{f(v_p)-a}{p} = -\sum_{p>Y} \frac{(a-f(v_p))_+}{p} + \mathcal{E}_{\gamma,\epsilon}(N).$$

The proof of (6.13) gives the same mass forcing, and the cost estimate now gives

$$|F| \geq N/e - \mathcal{C}(a)N\delta + (1-\eta-a-o(1))NT + N\mathcal{E}_{\gamma,\epsilon}(N) - O_B((\epsilon+\gamma)N\delta) - o(N\delta).$$

Theorem 1.3 and the near-optimality hypothesis therefore bound the two nonnegative slack terms by $O_B((\epsilon+\gamma)N\delta) + o(N\delta)$. The marked small-power sum is independent of ϵ, γ; taking these to zero after $N \to \infty$ proves $T = o(\delta)$. The displayed bound on the large-prime classification follows for each fixed pair of repair parameters. On the compact intervals of exponents separated by a positive distance from v_*, the weights $|a - f(v)|$ have a fixed positive lower bound by strict monotonicity for $v > 1$ (and $a > 0$ for $v \leq 1$). This proves the final assertion with the indicated order of limits. □

7 The near-logarithmic scale and companion entropy bounds

The Dickman law determines the deletion fraction at fixed exponents, but it is insensitive to the distinction between zero and a small number of required deletions at $v = 1$. The following elementary statement gives a pointwise bound in the linear logarithmic window.

Proposition 7.1 (Linear logarithmic window)**.** *For each fixed $0 < c < 1$, $\sigma(p, x) = 0$ for all $1 \leq x \leq c \log p$ when p is sufficiently large. For every fixed $A > 0$, uniformly for $1 \leq x \leq A \log p$,*

$$\sigma(p, x) \ll_A \frac{x}{\log_2 p}.$$

Here $\log_j$ denotes the j-fold iterated natural logarithm, for sufficiently large arguments.

Proof. The prime number theorem gives $\log \operatorname{lcm}(1, \ldots, \lfloor c \log p \rfloor) = (c + o(1)) \log p$. Hence this lcm times $H_{\lfloor c \log p \rfloor}$ is less than p, so every nonempty reciprocal numerator is positive and less than p. For the second claim, take a fixed sufficiently small $c_1 > 0$ as in the proof of the local upper bound and delete integers having a prime-power divisor above $c_1 \log p$. If $x \leq c_1 \log p$ no deletion is needed; otherwise the count is at most

$$x \sum_{\substack{c_1 \log p < q \leq A \log p \\ q \text{ prime}}} q^{-1} + O(x/(\log p)^{1/2}) \ll_A x/\log_2 p,$$

by Mertens' theorem and the prime-power tail. The retained lcm argument is exactly the same as before. □

For comparison with the above pointwise estimates, the companion manuscript [6] uses the distinct-sum entropy and its inverse

$$s(n) = \log |\{R(S) : S \subseteq [1, n]\}|, \qquad b(t) = \min\{n : s(n) \geq \log t\}.$$

It establishes $s(\lfloor tn \rfloor) \sim t s(n)$ for each fixed $t > 0$, and

$$b(p) \asymp \frac{\log p \log_2 p}{\mathcal{C}_4(p)},$$

where $\mathcal{C}_4(p)$ denotes the product of the iterated logarithms from $\log_4 p$ onward, stopped at a fixed cutoff greater than three as in that manuscript. In particular $b(p)/\log p \to \infty$. Its modular coverage results control the uniform law on *distinct rational values*, whereas its deletion-stable mixing theorem uses independent Bernoulli choices of denominators. The two laws need not coincide. We record the precise implications here separately: the following companion-dependent corollaries invoke those stated theorems and do not enter the proofs of Theorems 1.1 and 1.3.

Corollary 7.2 (Entropy bounds for the first zero sum, using N02)**.** *Assume the entropy and modular coverage results of [6]. For each fixed $0 < \varepsilon < 1$, on a relative density-one set of primes,*

$$\begin{aligned} &\sigma(p,x) = 0 \quad \text{if } x \le \lfloor (1-\varepsilon)b(p) \rfloor, \\ &\sigma(p,x) \ge 1 \quad \text{if } x \ge \lfloor (\tau+\varepsilon)b(p) \rfloor, \quad x < p, \end{aligned}$$

where $\tau \in (1,2)$ is the constant of the distinct-value full coverage theorem in [6]. In particular for every fixed $A > 0$ there are no nonempty zero sums up to $A \log p$ for density-one primes, once they are sufficiently large.

Proof. For the first assertion apply the companion modular coverage theorem below the entropy inverse threshold. Its deletion-mass formulation says that the only zero-residue reciprocal subset of $[1,q]$ is the empty one in this range: if a nonempty one existed, deleting its complement would have mass strictly smaller than the full harmonic mass, contradicting $d(p,q) = H_q$. For the second assertion use the companion conclusion on the first nonempty zero sum $z(p) \le (\tau+\varepsilon/2)b(p)$. As $b(p) \to \infty$, integer roundings are harmless. Both hold outside sets of primes of relative density zero, and monotonicity in x preserves the statements throughout the specified ranges. The last assertion follows since $b(p)/\log p \to \infty$. □

Corollary 7.3 (Deletion budget in the mixing window, using N02)**.** *Assume the deletion-stable Bernoulli mixing theorem of [6]. There are positive constants C, c such that for density-one primes and every*

$$C \log p \log_2 p \le x < p,$$

we have

$$\sigma(p,x) \gg \frac{\log p}{\sqrt{\log_2 p}\,\log_3 p}.$$

In fact after arbitrary deletions of at most $c \log p/(\sqrt{\log_2 p}\,\log_3 p)$ denominators, all residues are still represented by subsets of one pool inside $[1,x]$, with a nonempty representative for zero.

Proof. Use the rectangular block A_X and the prime block B_X of the companion theorem at scale $X < p \le 2X$, and write $L_X = \log X$, $U_X = \log L_X$, $V_X = \log U_X$. Their union lies in $[1, C_0 L_X U_X]$, $|A_X| \asymp_{C_0} L_X/V_X$, and $|B_X| \sim (C_0/10)L_X$. The mixing theorem permits deletion of a fixed small proportion of A_X and at most $c_0 L_X/(\sqrt{U_X}V_X)$ members of B_X. Thus any deletion set of total size at most $cL_X/(\sqrt{U_X}V_X)$ meets both requirements for sufficiently small fixed c. For nonexceptional primes the surviving pool has Bernoulli residue probabilities $p^{-1}(1+O(e^{-c_1 L_X/(\sqrt{U_X}V_X)}))$, uniformly in the deleted set and target. Take the sufficiently large fixed C_0 of the construction also with $(C_0/10)\log 2 > 1$; this is allowed in the companion construction. The number of subsets giving zero is then greater than one, since the surviving prime block alone has cardinality $(C_0/10 + o(1))L_X$, whereas $\log p = L_X + O(1)$. Hence at least one zero subset is nonempty. The exceptional count in each dyadic prime interval is $O(Xe^{-c_2 L_X/(\sqrt{U_X}V_X)}) = o(X/\log X)$. Enlarging the height constant absorbs the dyadic endpoints and rounding; the result follows for the full stated range of x. □

These estimates do not identify the exact onset of nonempty zero sums for every prime or the minimal deletion fraction at x comparable to $b(p)$. In particular distinct-value entropy near coverage is not automatically a deletion-stable fixed-target statement. The pointwise local Dickman theorem at fixed exponent $v > 1$ is unconditional with respect to the companion entropy theory.

8 Scope and further questions

The local results and the elementary global attack retain the original content of the Dickman note. The new matching global converse uses a full Dickman repair ledger, and the scalar carrier argument ensures that every corrected large-prime block is compatible with one common actual rational lattice. No matching conclusion is inferred merely from the local blocking threshold. The first-order result of Liu–Sawhney and the classical smooth-number estimate are credited explicitly. The fine-scale and coarse-scale exact completion arguments are reproduced here; only standard prime and sieve estimates enter them. The separate near-logarithmic companion corollaries require the entropy and mixing theorems as stated in [6].

The second-order converse and its cost-profile stability refinement concern the asymptotic extremal problem; the numerical evaluation of the coefficient is a separate computation. The related preprint [4] is cited for its local nonsmooth-set assertion, not as an input to the global proof. The companion-dependent statements above are not used in the main theorem. The exact all-primes onset in the $x \asymp b(p)$ range, a sharp remainder in the global expansion, and a classification of finite extremizers remain open here. In particular the first-order symmetric-difference stability does not imply that every optimal deletion pattern is a single initial segment plus the canonical large-prime attack at distance $o(N\delta_N)$.

References

[1] Y. P. Liu and M. Sawhney, On further questions regarding unit fractions, *Int. Math. Res. Not. IMRN* 2026, no. 2, rnaf382. doi:10.1093/imrn/rnaf382; arXiv:2404.07113 (we use the numbering of arXiv v1 for the first-order theorem and sieve lemma).

[2] P. Erdős and R. L. Graham, *Old and New Problems and Results in Combinatorial Number Theory*, Monographies de L'Enseignement Mathématique, vol. 28, 1980.

[3] E. S. Croot III, On a coloring conjecture about unit fractions, *Ann. of Math. (2)* **157** (2003), 545–556.

[4] W. Howe, G. Huffman, J. Kohlmeyer and L. Kruer, The sharp deficit in primitive signed reciprocal relations, Hexagon preprint 2611.00753v1, 21 November 2026, `https://hexagonmath.org/2611.00753v1`.

[5] A. Hildebrand and G. Tenenbaum, Integers without large prime factors, *J. Théorie Nombres Bordeaux* **5** (1993), 411–484. doi:10.5802/jtnb.101.

[6] P. Chen, *Recursive Entropy and Optimal Modular Coverage of Harmonic Subset Sums*, unpublished manuscript, 2026.