hexagon:2611.01299v1

OPTIMAL ENTROPY EXTRACTION AND EXACT ABSORPTION CAPACITIES OF RATIONAL SUMSETS

PENGCHENG CHEN

Abstract. We study the reciprocal mass needed to retain a prescribed fraction of the entropy of harmonic subset sums, and several exact-absorption capacities for ordinary subsets. A recursive prime-block decomposition yields an asymptotically exact variational formula for the leading cost coefficient at any fixed iterated logarithmic height. For every fixed positive entropy fraction the naive cardinality lower bound has a strict proportional gap, but the dilute-entropy limit has coefficient $1/\log 2$. Universal exact absorption and universal two-source addition obey exact recursive laws. A three-element weighted Fourier certificate strictly improves the two-source capacity; nevertheless, a dual trigonometric test on singleton components forces a fixed proportional gap from twice the entropy even for arbitrarily correlated probability measures. The constructible capacities yield full modular coverage at their entropy thresholds with optimal-order reciprocal mass, outside a strongly exceptional set of primes. Using the sharp rational entropy budget of the companion work, we obtain optimal-order mass under the same growing common-denominator budget, with fixed entropy slack. Finally, sparse exact repair obstructions delimit these methods. Neither a fixed-fraction limiting numerical cost constant nor directed coverage at the entropy threshold is asserted.

Contents

1. Introduction 1
2. Recursive components 3
3. The leading cost profile 4
4. Universal exact absorption 7
5. Weighted pair selection 8
6. Obstructions for arbitrary correlations 10
7. Modular transfer and costs of repair 11
7.1. Sparse signed representations 13
7.2. Why sparse exact repair can fail 14
8. Low-mass extraction under a growing denominator budget 15
9. Connection with square-layer losses and open questions 16
References 17

1. Introduction

For a finite set C of positive integers put

$$\mathcal{H}(C)=\sum_{a\in C}a^{-1},\qquad \mathscr{S}(C)=\{\mathcal{H}(A):A\subseteq C\},\qquad s(C)=\log|\mathscr{S}(C)|.$$

Date: 28 November 2026.
2020 *Mathematics Subject Classification.* 11B75, 11D68, 11T23, 42A16.
Key words and phrases. Harmonic subset sums, absorption, entropy, rational sumsets, Fourier majorants.

Write $\mathscr{S}_n = \mathscr{S}([1,n])$, $s(n) = s([1,n])$, $S(n) = e^{s(n)}$, $H_n = \mathcal{H}([1,n])$, and $\mathcal{L}(x) = \operatorname{lcm}(1,\dots,\lfloor x \rfloor)$. The empty sum is zero. All logarithms are natural; $\log_j$ denotes the j-fold logarithm. Let $b(P) = \min\{n : s(n) \ge \log P\}$. We count distinct rational values, not representations. The classical problem of counting distinct harmonic subsums is treated in [1, 2]; the companion work [4] develops the entropy and modular-cover framework. Here the questions are extremal and concern the cost and limitations of exact absorption.

For a finite C, define $F_{\mathrm{abs}}(C)$ as the maximum of $s(G)$ over $G \subseteq C$ for which there are pairwise disjoint sets $A_v \subseteq C \setminus G$, $v \in G$, with $\mathcal{H}(A_v) = 2/v$. Define

$$\Gamma(C) = \max_{A+B \subseteq \mathscr{S}(C)} \log(|A||B|), \qquad E(C) = \#\{(a,b) \in \mathscr{S}(C)^2 : a+b \in \mathscr{S}(C)\},$$

where the maximum is over nonempty finite rational sets A, B. For a finitely supported probability measure μ on $\mathbb{R}$ write $\widehat{\mu}(\theta) = \int e^{i\theta x}\, d\mu(x)$ and $\widetilde{\mu}$ for reflection. A feasible triple (μ, K_1, K_2) consists of finitely supported probability measures satisfying

$$\begin{gathered} \operatorname{supp}\mu \subseteq \mathscr{S}(C), \quad \operatorname{supp} K_i \subseteq \mathscr{S}(C) - \mathscr{S}(C), \quad \widehat{K_i} \ge 0, \\ |\widehat{\mu}| \le (\widehat{K_1}\widehat{K_2})^{1/2} \quad \text{on } \mathbb{R}. \end{gathered} \tag{1}$$

Let $\Omega(C)$ be the supremum of $-\log(K_1\{0\}K_2\{0\})$ over these triples. We abbreviate the capacities for $[1,n]$ by their argument n. A constructible capacity Ξ will be defined in Section 5; it depends on a fixed recursive decomposition.

Theorem 1.1 (Optimal reciprocal cost)**.** *For each fixed $0 < \varepsilon < 1$ there exist positive constants $c_\varepsilon, C_\varepsilon$ such that for sufficiently large n there is $G \subseteq [c_\varepsilon n, n]$ with*

$$s(G) \ge (1-\varepsilon)s(n), \qquad |G| \le C_\varepsilon s(n), \qquad \mathcal{H}(G) \le C_\varepsilon s(n)/n.$$

Any $G \subseteq [1,n]$ satisfying the entropy inequality has mass at least $(1-\varepsilon)s(n)/(n\log 2)$.

Theorem 1.2 (Absorption and majorant capacities)**.** *There are fixed positive constants $a, \eta, \eta_1, \eta_2, \eta_3$. For all sufficiently large n,*

$$as(n) \le F_{\mathrm{abs}}(n) \le (1-\eta)s(n), \qquad (1+\eta_1)s(n) \le \Gamma(n),$$

$$(1+\eta_2)\Gamma(n) \le \Xi(n) \le \Omega(n) \le (2-\eta_3)s(n).$$

The functions $F_{\mathrm{abs}}, \Gamma, \log E, \Xi$ are regularly varying of index one. In a fixed decomposition having ℓ singleton terminal components, the upper bound is more precisely $\Omega(n) \le 2s(n) - \ell\log(49/32)$.

The main distinction is between a Cartesian product of sources, a selected legal pair distribution, and an arbitrary feasible triple. The Fourier extremal question has a methodological connection with the classical Turán problem for supported positive definite functions [3]; here the extra majorization by the geometric mean of two transforms and the reciprocal component constraints are part of the definition, not consequences of the classical problem. None of their lower threshold bounds is asserted to be a necessary threshold for actual modular representation. The proofs of the structural theorems use only elementary rational rigidity, the prime number theorem and finite Fourier identities. The modular transfer additionally uses the elementary suppressor from [4], for which a proof is included. The cost profile in Section 3 reduces the leading coefficient to a finite variational problem at logarithmically smaller heights. Section 8 uses the sharp rational entropy budget of [5] as an explicit companion-manuscript input. No open directed-cover assertion is an input. The only other non-elementary input from the companion work is the already established all-prime fallback in the final square-layer application.

2. Recursive components

Fix an integer $D \geq 4$, a small positive constant c, and a sufficiently large fixed stopping height n_0. At any node of height $n > n_0$, set

$$y_n = n/(c \log n), \quad j_r = \lfloor n/r \rfloor, \quad V_r = r[1, j_r] \quad (y_n < r \leq n,\ r \text{ prime}),$$

and let C_n be the complement of the disjoint blocks V_r in $[1, n]$. Retain C_n as a core and recursively decompose the rescaled $[1, j_r]$; stop at heights at most n_0. All components are understood in the original scale. The choices of D, c, n_0 remain fixed in limits. When a larger stopping height is used for entropy extraction, keep D, c and the prime-block rule unchanged; this coarsens the tree. In the definition of Ξ below the reference stopping height is fixed once and for all.

Lemma 2.1 (Bounded-coefficient rigidity)**.** *Every relation $\sum_{a \leq n} c_a/a = 0$ with integral $|c_a| \leq D$ vanishes separately on each component of this tree. The same holds for any subset of the denominators by putting unused coefficients equal to zero.*

Proof. Take c so small that $D\mathcal{L}(j_r)H_{j_r} < r$ at every internal node; this follows from $\log \mathcal{L}(j) = O(j)$ and $j_r \leq c \log n$. Since $r^2 > n$, multiplying by r and reducing modulo r isolates the terms ra. The integer $\mathcal{L}(j_r) \sum_{a \leq j_r} c_{ra}/a$ is a multiple of r of absolute value less than r, hence zero. Subtraction gives the core relation, and induction gives the result in the descendants. □

Consequently the map from tuples of component subset-sum values to their total is a bijection onto the subset-sum set. The core common denominator divides $\prod_{r \leq y_n} r^{\lfloor \log n / \log r \rfloor}$, whose logarithm is $O_c(n/\log n)$ by the prime number theorem and the bound for higher powers. Counting integer numerators gives $s(C_n) = O_c(n/\log n)$. Thus

$$s(n) = s(C_n) + \sum_{r > y_n} s(j_r) = \frac{n}{\log n} \left(\sum_{j \leq \log n} \frac{s(j)}{j(j+1)} + O(1) \right). \tag{2}$$

For the second equality, group primes in $(n/(j+1), n/j]$. The classical prime number theorem with error term is uniform for $j \leq \log n$ and gives relative error $O(\log \log n / \log n)$ in replacing the prime count by $n/(j(j+1)\log n)$. The incomplete last group and change from $c \log n$ to $\log n$ cost $O(1)$ after normalization since $s(j) \leq j \log 2$. The same proof applies to any nonnegative $a(n) \asymp s(n)$ with this recursion and core contribution $O(n/\log n)$.

Using $s(j) \geq \pi(j) \log 2$ (distinct subsets of primes are distinguished by their prime denominators), and the trivial upper bound, (2) gives

$$\frac{n \log_3 n}{\log n} \ll s(n) \ll \frac{n \log_2 n}{\log n}. \tag{3}$$

The sum in parentheses diverges and changes by $o(1)$ on replacing n by any fixed multiple, so $s(tn) \sim ts(n)$. The same conclusion holds for a as above. In fact $a(n)$ is asymptotic to a nondecreasing function, namely $n/\log n$ times its positive truncated sum. Hence its first crossing of $v \log P$, for any fixed $v > 0$, has value $(v + o(1)) \log P$ and satisfies the inverse regular-variation law. To see this, sandwich a for all sufficiently large n between $(1 \pm \epsilon)$ times the nondecreasing equivalent; the inverse of that function has the stated law by comparing the values at $(t \pm \epsilon)n$. In particular, with $L = \log P, \lambda = \log L$,

$$b(P^t) \sim tb(P), \quad \log b(P) \sim \lambda, \quad L\lambda/\log \lambda \ll b(P) \ll L\lambda/\log\log\lambda = o(L\lambda). \tag{4}$$

Here the inverse bounds follow by evaluating (3) at constant multiples of the respective endpoints; $s(b(P)) = L + O(1)$ since each new denominator at most doubles the count.

Lemma 2.2 (Abundance of fixed terminal heights)**.** *For each fixed j and a sufficiently large fixed stopping height, the number $\ell_j(n)$ of terminal components of height j satisfies $\ell_j(n) \asymp_j s(n)$ as $n \to \infty$. One tree works for any prescribed finite list of heights.*

Proof. Let $V = \sum_m m/\log m$ over internal nodes with multiplicity. The sum of their core entropies is $O(V)$, and each internal node has $O(m/\log m)$ children, each terminal child contributing at most $n_0 \log 2$. Therefore $s(n) \ll V$. For fixed $j \le n_0$ and $c \log m > j+1$, primes in $(m/(j+1), m/j]$ give $\gg_j m/\log m$ direct terminal children of height j. Sum over nodes. Conversely $\ell_j(n)s(j) \le s(n)$ by rigidity. □

Proof of Theorem 1.1. Discard all internal cores, retain the terminal denominators, and use rigidity to add their entropies. At a node of height m the relative entropy discarded is $O(1/\log_3 m)$ by (3). On every descending path the next height is at most $\log m$; the sum of $1/\log_3 m$ on internal nodes of that path is $O(1/\log_3 n_0)$, by reverse exponential growth. Normalize the child entropy weights at every branching to sum to one; then the fraction of root entropy retained is an average of products of retention fractions along paths. Choose n_0 to make each such product at least $1-\varepsilon$. At a terminal height $j \le n_0$, one has $j \le C_{n_0} s(j)$, giving the cardinality bound by additivity. If the accumulated scale along a path is w, then $j = \lfloor n/w \rfloor$ by nested floors, so all retained denominators are at least $w > n/(n_0+1)$. This gives the mass bound. The universal lower bound follows from $s(G) \le |G| \log 2$ and $\mathcal{H}(G) \ge |G|/n$. □

3. The leading cost profile

Theorem 1.1 gives the correct order but does not specify its leading coefficient. We give a variational description of that coefficient at a logarithmically smaller height, and determine the coefficient in the small-entropy-fraction limit. Both arguments use the same bounded-coefficient rigidity, rather than an assumption that the limiting distribution of terminal types exists.

For $0 < \alpha < 1$ write

$$M_\alpha(n) = \min\{\mathcal{H}(G) : G \subseteq [1,n],\ s(G) \ge \alpha s(n)\}.$$

For real $x \ge 1$ and price $z \ge 0$, set

$$\Phi_z(x) = \max_{G \subseteq [1,\lfloor x \rfloor]} \{z s(G) - x\mathcal{H}(G)\},$$

$$A(x) = \int_1^x \frac{s(\lfloor u \rfloor)}{u^2}\, du, \qquad B_z(x) = \int_1^x \frac{\Phi_z(u)}{u^2}\, du.$$

In particular $0 \le \Phi_z(x) \le z s(\lfloor x \rfloor)$, and B_z is nondecreasing in x and convex in z. The integral $A(x)$ differs from the sum $\sum_{j \le x} s(j)/(j(j+1))$ by $O(1/x)$.

Finite lemma 1 (One-price relaxation and one-block rounding)**.** *Let I be a finite collection of blocks, each with a nonempty finite option set J_i. Option $j \in J_i$ has entropy $e_{ij} \ge 0$ and cost $c_{ij} \in [0, C]$. Assume each block has a zero-cost option, and there exists a choice with total entropy $> h \ge 0$. Set*

$$f_i(z) = \max_{j \in J_i}(z e_{ij} - c_{ij}), \qquad D(z) = zh - \sum_{i \in I} f_i(z) \quad (z \ge 0).$$

The minimum cost among convex combinations of the options in each block, with total entropy at least h, equals $\max_{z \ge 0} D(z)$. The minimum among integral choices lies in

$$[\max_{z \ge 0} D(z),\ \max_{z \ge 0} D(z) + C].$$

Proof. For any feasible convex combination, its cost is at least $D(z)$ for every $z \ge 0$, by summing the defining inequalities for f_i. The strict feasible choice implies $D(z) \to -\infty$ as $z \to \infty$. The continuous piecewise affine function D therefore attains a maximum at some $z_* \ge 0$.

Let e_i^- and e_i^+ be the least and greatest entropy among the options attaining $f_i(z_*)$. All inactive options have a strictly positive gap from the maximum; finiteness ensures that for sufficiently

small positive perturbations of z_* the active slopes suffice. For $z_* > 0$ the left and right slopes of f_i at z_* are respectively e_i^- and e_i^+. The maximum condition for D thus gives

$$\sum_i e_i^- \le h \le \sum_i e_i^+.$$

Start with a minimizing-entropy active option in each block. Change them in any order to the maximizing-entropy active option until the entropy reaches or first exceeds h, interpolating in just the crossing block to achieve equality (if already equal, no interpolation is needed). Every used option is active, so its cost equals $z_* e_{ij} - f_i(z_*)$. Summing, the interpolated choice costs exactly $D(z_*)$. Rounding the sole interpolated block to its greater-entropy active option preserves feasibility and raises the cost by at most C. For $z_* = 0$, the right slope condition gives $h \le \sum_i e_i^+$, and all maximum-entropy active options have minimum cost zero; they already give a feasible integral choice with cost $D(0) = 0$. This proves both claims using finite maxima and finite sums, without an appeal to general linear programming duality. □

Theorem 3.1 (Asymptotically exact variational cost)**.** *For each fixed $0 < \alpha < 1$ there is a finite constant Z_α such that for every fixed integer $k \ge 1$,*

$$\frac{nM_\alpha(n)}{s(n)} = \max_{0 \le z \le Z_\alpha} \left\{ z\alpha - \frac{B_z(\log_k n)}{A(\log_k n)} \right\} + o(1). \tag{5}$$

In particular the leading coefficient, up to $o(1)$, is determined by finite problems at heights at most $\lfloor \log_k n \rfloor$. This formula does not assert the existence of a constant limit when α is fixed.

Proof. First consider the profit at the original height. For any $z \ge 0$,

$$nM_\alpha(n) \ge z\alpha s(n) - \Phi_z(n).$$

Fix $\alpha' \in (\alpha, 1)$. By the proof of Theorem 1.1, a pool with entropy at least $\alpha' s(n)$ and cost at most $C_\alpha s(n)$ in the normalized units $n\mathcal{H}$ can be obtained from terminal blocks of a tree with a sufficiently large fixed stopping height M_0. Use the same prime-block rule but stop at $M \ge M_0$. This coarser tree retains every denominator in that pool. Denote its terminal blocks by T, and let e_M be a bound for the discarded fraction of entropy, with $e_M \to 0$ as $M \to \infty$; the path argument in the proof of Theorem 1.1 supplies such a bound. By rigidity, the profit separates exactly over the cores and terminal blocks. If $\Phi_z^{(M)}(n)$ is the maximum profit using only terminal denominators of this coarser tree, then

$$0 \le \Phi_z(n) - \Phi_z^{(M)}(n) \le z e_M s(n). \tag{6}$$

Each terminal block has the form $w[1, j]$ with $j = \lfloor n/w \rfloor \le M$. Its choices are the finitely many pairs of entropy and cost $(s(G), n\mathcal{H}(G))$, $G \subseteq w[1, j]$. Each cost is at most $(M+1)H_M$. Apply Finite lemma 1 to the choices in the terminal blocks, with $h = \alpha s(n)$ and $C = (M+1)H_M$. Its strict feasibility hypothesis follows from the pool for α', and the empty choice has zero cost. The optimal relaxed cost is exactly

$$\max_{z \ge 0} \{ z\alpha s(n) - \Phi_z^{(M)}(n) \},$$

and one-block rounding costs at most $(M+1)H_M$. The maximum may be restricted to $0 \le z \le Z_\alpha$, for example $Z_\alpha = 1 + C_\alpha/(\alpha' - \alpha)$: the pool for α' bounds the dual objective from above by $(C_\alpha - z(\alpha' - \alpha))s(n)$, and $z = 0$ gives zero. The same restriction works for the full profit $\Phi_z(n)$. Combining weak duality at height n, the rounded feasible terminal choice and (6), we obtain

$$0 \le nM_\alpha(n) - \max_{0 \le z \le Z_\alpha} \{ z\alpha s(n) - \Phi_z(n) \} \le Z_\alpha e_M s(n) + (M+1)H_M.$$

Let $n \to \infty$ and then $M \to \infty$. This establishes the asymptotically exact dual formula at height n.

For the reduction of height, the exact decomposition at $m = \lfloor x \rfloor$ gives

$$\Phi_z(x) = \max_{G \subseteq C_m} \{zs(G) - x\mathcal{H}(G)\} + \sum_{r > y_m} \Phi_z(x/r).$$

The core profit is $O_Z(m/\log m)$ for $0 \le z \le Z$. Uniformly on this compact price interval, $\Phi_z(u)$ has bounded variation $O_Z(Y \log(2Y))$ on $[1, Y]$: on a unit interval it is the maximum of finitely many affine functions of slopes in $[-H_{\lfloor u \rfloor}, 0]$, and at an integer its upward jump is at most $z \log 2$. The classical PNT error and partial summation therefore replace the prime sum by $\int_{y_m}^{x} \Phi_z(x/v)\, dv/\log v$ with error $o_Z(x/\log x)$. All its arguments x/v are at most $O(\log x)$, so the variation bound is more than sufficient for this error. Under $u = x/v$ we can replace $\log(x/u)$ by $\log x$, incurring $o_Z(x/\log x)$ since $\Phi_z(u) \le Zu \log 2$. Extending the upper integral limit from $x/y_m \sim c \log x$ to $\log x$ costs $O_Z(x/\log x)$. Hence

$$\Phi_z(x) = \frac{x}{\log x}(B_z(\log x) + O_Z(1)). \tag{7}$$

In the same notation (2) is $s(\lfloor x \rfloor) = x(A(\log x) + O(1))/\log x$, including real x. As $A(x) \to \infty$, these formulas give

$$\frac{\Phi_z(n)}{s(n)} = \frac{B_z(\log n)}{A(\log n)} + o(1)$$

uniformly on bounded price intervals. This proves (5) for $k = 1$.

We record explicitly why it can be iterated, rather than assume a limit for the ratios. The entropy recursion and monotonicity of A give

$$A(\log X) \ll A(\log_2 X) \log_3 X, \qquad A(X) \gg A(\log_2 X) \log_2 X$$

for large X. For the lower bound integrate over $\log X \le u \le X$, using $A(\log u) \ge A(\log_2 X) \to \infty$; for the upper bound integrate the recursion up to $\log X$, using $A(\log u) \le A(\log_2 X)$ and treating bounded heights separately. Thus $A(\log X)/A(X) \to 0$. For fixed $0 < \theta < 1$, integration on $[X^\theta, X]$ gives $A(X) - A(X^\theta) = O_\theta(A(\log X)) = o(A(X))$. It follows that

$$B_z(X) - B_z(X^\theta) \le Z(A(X) - A(X^\theta)) = o_Z(A(X))$$

uniformly for $z \le Z$. Integrating (7), on the interval $[\exp((\log X)^\theta), X]$ the function $B_z(\log u)$ differs from $B_z(\log X)$ by $o_Z(A(\log X))$. On the complementary lower interval the contribution is bounded by $O_Z(1) + Z\theta \log_2 X\, A(\log X)$, plus $O_Z(\log_2 X)$ from the recursion error. First let $X \to \infty$ and then $\theta \downarrow 0$. We obtain

$$B_z(X) = \log_2 X\, B_z(\log X) + o_Z(\log_2 X\, A(\log X)), \qquad A(X) \sim \log_2 X\, A(\log X).$$

The latter assertion follows by the identical integral argument with s in place of Φ_z. Consequently the normalized ratios at X and $\log X$ differ by $o_Z(1)$. Iterating a fixed number of times gives the asserted reduction. The number of iterations is not allowed to grow with n in this argument. □

Lemma 3.2 (Entropy near the top endpoint)**.** *Put $I_\delta(n) = [(1-\delta)n, n] \cap \mathbb{Z}$. Then*

$$\lim_{\delta \downarrow 0} \limsup_{n \to \infty} \frac{s(I_\delta(n))}{s(n)} = 0.$$

For each fixed $0 < \delta < 1/3$, a suitable fixed recursive tree has at least $c_\delta s(n)$ singleton terminal denominators in $I_\delta(n)$, for all sufficiently large n, with $c_\delta > 0$.

Proof. Stop the tree at M sufficiently large and put $V_M = \sum m/\log m$ over its internal nodes. The proof of Lemma 2.2 gives $s(n) \ll_M V_M$. Also $V_M \ll s(n)$ with an absolute implied constant for all large M: each internal node of height m contributes $\gg m/\log m$ direct singleton children using primes in $(m/2, m]$, and these children are disjoint terminal components. The entropy of all discarded cores is at most $e_M s(n)$ with $e_M \to 0$, as above.

At an internal node with accumulated scale w, put $x = n/w \in [m, m+1)$, where $m = \lfloor x \rfloor$. Every terminal child denominator in the indicated band has the form wra with $1 \le a \le \min(M, c \log m)$ and prime $r \in [(1-\delta)x/a, x/a]$. For $\delta \le 1/2$ the PNT with its classical error bounds the number of such pairs by

$$\ll \sum_{a \le \min(M, c\log m)} \left(\frac{\delta m}{a \log m} + \frac{m}{a} e^{-c_1\sqrt{\log m}} + 1 \right) \ll (\delta \log(2M) + \eta_M) \frac{m}{\log m},$$

where $\eta_M \to 0$ uniformly for $m > M$. Indeed all endpoints have logarithm $\log m + O(\log_2 m)$, and the normalized error is bounded by a constant times $\log m(1 + \log_2 m)e^{-c_1\sqrt{\log m}} + (\log m)^2/m$. Counting pairs may overcount but is an upper bound. Summing over nodes, and adding the core entropy, gives

$$s(I_\delta(n)) \le \big(e_M + C(\delta \log(2M) + \eta_M)\big) s(n).$$

Choose large M first and then small δ. This proves the first claim. For the lower claim fix δ and choose M large enough that each interval $[(1-\delta)x, m]$ contains $\gg \delta m/\log m$ primes by the PNT, uniformly for $x \in [m, m+1)$. These primes are $> m/2$ and $> y_m$, and yield singleton children whose denominators belong to the band. Thus their total count is $\gg_\delta V_M \gg_{M,\delta} s(n)$. Rigidity gives independence of their binary rational subset sums. □

Corollary 3.3 (Strict cost gap and the dilute-entropy constant)**.** *For each fixed $0 < \alpha < 1$ there exists $\kappa_\alpha > 0$ such that for all sufficiently large n,*

$$M_\alpha(n) \ge (1 + \kappa_\alpha) \frac{\alpha s(n)}{n \log 2}.$$

Nevertheless the optimal dilute-entropy coefficient is exactly $1/\log 2$:

$$\lim_{\alpha \downarrow 0} \liminf_{n\to\infty} \frac{nM_\alpha(n)}{\alpha s(n)} = \lim_{\alpha \downarrow 0} \limsup_{n\to\infty} \frac{nM_\alpha(n)}{\alpha s(n)} = \frac{1}{\log 2}.$$

Proof. Choose fixed $\delta > 0$ so small that $s(I_\delta(n)) \le \alpha s(n)/2$ eventually. Since the entropy of a union of disjoint denominator sets is at most the sum of their entropies, any feasible G has $s(G \setminus I_\delta(n)) \ge \alpha s(n)/2$. Its part outside the band has at least $\alpha s(n)/(2 \log 2)$ elements. All elements cost at least one in the normalized units $n\mathcal{H}$, and each outside the band costs at least $1/(1-\delta)$. Together with $|G| \ge \alpha s(n)/\log 2$, this proves the strict gap, for instance with $\kappa_\alpha = \delta/(2(1-\delta))$.

Conversely, for a fixed small δ choose the singleton pool of Lemma 3.2. When $\alpha < c_\delta \log 2$, select $\lceil \alpha s(n)/\log 2 \rceil$ of these denominators. Their sums are distinct, so this is feasible, with cost at most $(\alpha s(n)/\log 2 + 1)/((1-\delta)n)$. For these α the normalized upper limit is at most $1/((1-\delta)\log 2)$. Combine with the universal lower bound and let $\delta \downarrow 0$. This proves the two iterated limits, not a uniform joint limit for arbitrary $\alpha = \alpha(n)$. □

Remark 3.4. For a fixed α, proving convergence of the normalized recursive profit functions $B_z(X)/A(X)$ on $[0, Z_\alpha]$ would give a limiting coefficient in (5). Pointwise convergence on that interval would suffice by the uniform Lipschitz bound $|B_z(X) - B_{z'}(X)| \le |z - z'|A(X)$. Regular variation of s alone does not supply this convergence. Formula (5) and the strict gap are unconditional; a numerical fixed-fraction limit is left open.

4. Universal exact absorption

Theorem 4.1 (Exact component laws)**.** *For each of F_{abs} and Γ and $\log E$, the capacity of a set of denominators equals the sum of the capacities of its intersections with the recursive components of an enclosing interval. In particular*

$$a(n) = a(C_n) + \sum_{r > y_n} a(j_r)$$

for these three functions a. Moreover

$$\begin{aligned} \ell_2(n)\log 2 \le F_{\mathrm{abs}}(n) &\le s(n) - \ell_1(n)\log 2, \\ s(n) + \ell_6(n)\log(49/13) \le \Gamma(n) &\le 2s(n) - \ell_1(n)\log 2, \\ s(n) + \ell_1(n)\log(3/2) \le \log E([1,n]) &\le 2s(n) - \ell_1(n)\log(4/3). \end{aligned}$$

Proof. A matching $\mathcal{H}(A_v) = 2/v$ is confined to the component of v by Lemma 2.1 and positivity. Therefore the matching optimization separates exactly. Singleton components cannot be used. In a two-element component $\{w, 2w\}$ use $2w$ as the source and w as its matching set. The source and matching denominators have total size $O(s(n))$ and mass $O(s(n)/n)$ by the nested-floor argument.

For the two-source law write T_i for the component value sets and $\pi_i(z)$ for the unique coordinate of $z \in \mathscr{S}(C)$. Given $A + B \subseteq \mathscr{S}(C)$ fix a_0, b_0. Define $\alpha_i(a) = \pi_i(a + b_0)$ and $\beta_i(b) = \pi_i(a_0 + b) - \pi_i(a_0 + b_0)$. The four-term relation among $a + b, a + b_0, a_0 + b, a_0 + b_0$ is componentwise by rigidity. Hence $\alpha_i(a) + \beta_i(b) \in T_i$. Both maps to the full coordinate tuple are injective, giving the upper product bound. For the reverse, take optimal local sources, translate oppositely so that $A_i \subseteq T_i$, $B_i \subseteq T_i - T_i$. Form all component sums in each source. They are distinct by rigidity with coefficient bound two; all paired sums are legal. This proves exact addition. The maximum is finite and attained since the same translations confine the sources to finite sets. Always $s(C) \le \Gamma(C) \le 2s(C)$ by fixing one element of either source.

The singleton value set has capacity $\log 2$, using $|A + B| \ge |A| + |B| - 1$ in the ordered rational group. For height six, $\{2, 3, 6\}$ gives all $j/6$, $0 \le j \le 6$. Twice this interval of sums can be encoded using also denominator 1. Adding denominators 4 and 5 to the two sources separately gives two sets of size 14 with legal sums. Scaling by 60, the weights are $60, 30, 20, 15, 12, 10$; modulo 5 the weight 12 is identified, and then modulo 10 the weight 15 is identified. The other weights generate $10j$, $0 \le j \le 12$. Thus $S(6) = 52$ and the gain is $\log(196/52)$. A two-element alternative uses $\{0, 1/2\} + \{0, 1/2, 1\}$, gaining $\log(3/2)$ over $s(2)$.

Finally, the condition $a + b = c$ for three subset-sum values holds coordinatewise by rigidity, so legal ordered pair counts multiply. Each component has between S and S^2 legal pairs; a singleton has exactly three. This gives the last line. □

Lemma 2.2 gives the fixed proportional gaps in Theorem 1.2. Each of the three capacities in Theorem 4.1 is comparable to s, has bounded core term, and therefore is regularly varying by the argument after (2).

5. Weighted pair selection

Lemma 5.1 (A three-element certificate)**.** *For $C = [1, 3]$, one has $S(3) = 8$, $E(C) = 30$, $\Gamma(C) = \log 12$. There is a feasible triple with $K_1 = K_2 = K$ and $K\{0\} = 57/200$, giving a strict gain $g_3 = \log(10000/9747) > 0$ over $\Gamma(C)$.*

Proof. Scale all values by 6. The value set $T = \{0, 2, 3, 5, 6, 8, 9, 11\} = \{0, 3, 6, 9\} + \{0, 2\}$. By opposite translation any two legal rational sources may be made integer sets. They cannot both occupy at least two residue classes mod 3, since their sums would occupy all three. If A occupies one class, each occupied class B_i of B satisfies $|A| + |B_i| - 1 \le 4$, hence $|A||B| \le 2|A|(5 - |A|) \le 12$. The sets $\{0, 3\}$ and $\{0, 2, 3, 5, 6, 8\}$ attain this product.

z	0	2	3	5	6	8	9	11
$\#\{(a,b) \in T^2 : a + b = z\}$	1	2	2	4	3	6	4	8
$2000\mu(z)$	155	170	328	347	347	328	170	155

The table gives a probability law on legal sums; distribute its mass uniformly among the pairs with each sum. Then $\widehat{\mu}(\theta) = e^{11i\theta/2}G(\theta)$ where

$$\begin{aligned} 1000G(\theta) &= 347\cos(\theta/2) + 328\cos(5\theta/2) \\ &\quad + 170\cos(7\theta/2) + 155\cos(11\theta/2), \\ 1000F(\theta) &= 285 + 70\cos\theta + 154\cos 2\theta + 315\cos 3\theta \\ &\quad + 93\cos 5\theta + 83\cos 6\theta. \end{aligned}$$

We claim $F \geq |G|$. Put $x = \cos(\theta/2)$. Chebyshev expansion gives $F - G = (1-x)P(x)/250$, with

$$\begin{aligned} P(x) = {} & -42496x^{11} - 2816x^{10} + 112768x^9 + 3648x^8 \\ & - 110016x^7 + 1824x^6 + 47632x^5 - 3556x^4 \\ & - 8214x^3 + 1051x^2 + 238x + 11. \end{aligned}$$

Here is a purely algebraic positivity certificate. Write $P(x) = \sum_{k=0}^{11} p_k x^k$ with the integer coefficients displayed above. For integers $a < b$ and $d > 0$ define

$$c_j(a,b,d) = \sum_{k=0}^{11} p_k d^{11-k} \sum_{h=0}^{k} \binom{k}{h} a^{k-h} b^h \binom{11-k}{j-h}, \quad 0 \leq j \leq 11, \tag{8}$$

where out-of-range binomial coefficients are zero. By the binomial theorem,

$$d^{11}P(x) = \sum_{j=0}^{11} c_j(a,b,d) u^{11-j} v^j, \quad u = \frac{b-dx}{b-a}, \quad v = \frac{dx-a}{b-a}. \tag{9}$$

Indeed $u + v = 1$ and $au + bv = dx$, and the right side expands $\sum_k p_k d^{11-k}(au+bv)^k(u+v)^{11-k}$. The following table gives the exact least coefficient in each row of the certificate, calculated by the finite integer sums in (8):

a	b	d	$\min_{0\leq j\leq 11} c_j(a,b,d)$
-16	-15	16	190036955449856
-15	-14	16	115102532501504
-7	-6	8	41530949632
-3	-2	4	20278784
-2	-1	4	385532416
-2	-1	8	98370760192
-1	0	8	94489280512
0	1	4	46137344
2	3	8	43544217088
3	4	8	43544217088
2	3	4	49453568
3	4	4	49453568

The intervals $[a/d, b/d]$ form a partition of $[-1, 1]$ with shared endpoints. On each interval $u, v \geq 0$, and at least one is positive. Every coefficient in its row is strictly positive, so (9) proves $P(x) > 0$ everywhere on $[-1, 1]$, including the endpoints. These are $12 \cdot 12 = 144$ finite integer positivity checks, which can be verified without real-root algorithms.

For a redundant independent check, the unnormalized Sturm sequence starting with P, P' and continuing with negative Euclidean remainders has endpoint signs

-1	+	−	+	−	+	−	+	+	+	+	+	+
1	+	−	−	−	−	−	−	+	−	+	−	+

with six changes at either endpoint; $P(-1) = 250$ and $P(1) = 74$. The sign lists for the monic normalization of P are their simultaneous negatives. The positivity argument above does not require this optional Sturm check. Since F is even and G odd in x, replacing x by $-x$ also gives $F + G \geq 0$. The nonnegative Fourier coefficients of F sum to one, with support $\{0, \pm 1, \pm 2, \pm 3, \pm 5, \pm 6\} \subseteq T - T$. Thus $F = \widehat{K} \geq 0$ for a probability measure K, with zero mass $\alpha = 57/200$. Since $12\alpha^2 = 9747/10000 < 1$, the gain follows. Divide the support by six to recover the original scale. The homogeneous certificate, optional endpoint signs and all finite counts can be reproduced by integer binomial sums, exact rational Euclidean remainders and the displayed value sets; no numerical positivity test is used. □

For any legal sources A, B, their uniform measures give the feasible triple $\mu_A * \mu_B, \mu_A * \widetilde{\mu}_A, \mu_B * \widetilde{\mu}_B$ of capacity $\log |A||B|$. In general, averaging over a common rational period and using Cauchy–Schwarz gives

$$|\mathscr{S}(C)|^{-1} \leq \left\langle |\widehat{\mu}|^2 \right\rangle \leq \langle |\widehat{\mu}| \rangle \leq \sqrt{K_1\{0\}K_2\{0\}}. \tag{10}$$

Thus the zero masses are positive and $\Gamma \leq \Omega \leq 2s$.

Definition 5.2. Fix the tree of Section 2 with terminal heights including $1, 2, 3, 6$. Use the three-element certificate on every height-three terminal component and optimal Cartesian sources on all other components. The capacity of this construction is denoted $\Xi(n)$.

The triples combine by convolution. Their Fourier inequalities multiply and their supports remain legal. The zero atom of each global majorant is the product of the local zero atoms, because a relation among component differences is componentwise zero. Consequently

$$\Xi(n) = \Gamma(n) + g_3\ell_3(n) \geq (1 + \eta_2)\Gamma(n), \qquad \Xi(n) \leq \Omega(n). \tag{11}$$

This function has the same recursion with core term $\Gamma(C_n)$ and changed height-three initial value. The argument following (2) proves its regular variation, without assuming monotonicity. The first-crossing inverse of Ξ can therefore be used safely.

6. Obstructions for arbitrary correlations

We now allow all feasible triples, without imposing independence between terminal components. Let $v_1, \ldots, v_\ell$ denote the singleton denominators and C' the other denominators. Every ordinary sum has unique coordinates $x = y + \sum_j \epsilon_j/v_j$, $y \in \mathscr{S}(C')$, $\epsilon_j \in \{0, 1\}$, and $S(n) = 2^\ell |\mathscr{S}(C')|$.

Theorem 6.1 (Dual majorant obstruction)**.** *For the fixed tree, $\Omega(n) \leq 2s(n) - \ell \log(49/32)$. If the support condition on K_i in* (1) *is removed (still finitely supported probability measures), then even the resulting $\Omega_{\mathrm{free}}(n)$ satisfies $\Omega_{\mathrm{free}}(n) \leq 2s(n) - \ell \log(9/8)$.*

Proof. Use the trigonometric polynomials

$$w(\theta) = 1 - \cos 2\theta, \qquad h(\theta) = \frac{1 + e^{-i\theta}}{\sqrt{2}}(1 - \tfrac{1}{2}\cos\theta), \qquad g = wh.$$

Here $w \geq 0$. For $x = |\cos(\theta/2)|$ we have $|h| = \sqrt{2}x(3/2 - x^2)$ and the exact square certificate

$$1 - 2x^2(3/2 - x^2)^2 = (1 - x^2/2)(2x^2 - 1)^2 \geq 0. \tag{12}$$

Thus $|h| \leq 1$ without a calculus argument. The Laurent expansions, writing $z = e^{i\theta}$, are

$$\begin{aligned} 4\sqrt{2}h &= -z + 3 + 3z^{-1} - z^{-2}, \\ 8\sqrt{2}g &= z^3 - 3z^2 - 5z + 7 + 7z^{-1} \\ &\qquad - 5z^{-2} - 3z^{-3} + z^{-4}. \end{aligned} \tag{13}$$

Consequently the coefficients of g at $0, -1$ are both $a = 7/(8\sqrt{2})$, and its frequencies lie in $[-4, 3]$. For the finite-period averaging used below, if the frequencies are all in $D^{-1}\mathbb{Z}$, normalized integration over $[0, 2\pi D]$ extracts precisely the constant coefficient. For arbitrary finite real frequencies the same conclusion follows from the mean over $[-T, T]$ as $T \to \infty$, because the mean of $e^{it\omega}$ tends to zero for each fixed $\omega \neq 0$. Put $W(\theta) = \prod_j w(\theta/v_j)$. With normalized averaging over a common rational period, rigidity gives $\left\langle W\widehat{K}\right\rangle = K\{0\}$ for every K supported on the difference set: local difference coefficients are $-1, 0, 1$, whereas the test adds $-2, 0, 2$, so at each singleton the only possible cancellation has both zero. The remaining difference must vanish as well.

For arbitrary μ, choose a slice $y = y_0$ with probability at least $|\mathscr{S}(C')|^{-1}$, and test with $G(\theta) = e^{-i\theta y_0} \prod_j g(\theta/v_j)$. Rigidity with coefficient bound four gives

$$\langle G\widehat{\mu}\rangle = a^\ell \mu\{y = y_0\}, \qquad |G| \le W.$$

Thus $\langle W|\widehat{\mu}|\rangle \ge (2a)^\ell/S(n)$. Apply (1) and weighted Cauchy–Schwarz to bound the left side by $\sqrt{K_1\{0\}K_2\{0\}}$. Squaring gives $49/32$ per singleton.

For free supports use h instead of g, and $W = 1$. The two relevant coefficients of h are $3/(4\sqrt{2})$, so the same test gives $(3/(2\sqrt{2}))^\ell/S(n)$. Averaging the possibly irrational majorants over $[-T, T]$ and taking $T \to \infty$ extracts their zero atoms (finite support). Cauchy–Schwarz then proves the second assertion. □

Together with (11) and Lemma 2.2, this completes Theorem 1.2. For each fixed $J \ge 1$ one can strengthen the loss per detected leaf: replace w by

$$w_J(\theta) = \frac{1}{J+1}\left|\sum_{k=0}^{J}(-1)^k e^{2ik\theta}\right|^2.$$

Its coefficients at ± 2 are $-J/(J+1)$, odd coefficients vanish, and the constant coefficient is one. The test $w_J h$ has coefficients $0, -1$ equal to $(1 - 1/(4(J+1)))/\sqrt{2}$. With rigidity bound at least $2J+3$ and a correspondingly chosen fixed tree, this gives a loss per leaf of $\log[2(1-1/(4(J+1)))^2]$, tending to $\log 2$ as $J \to \infty$. The fraction of detected leaves depends on the tree; we do not let J grow with n in this argument.

Proposition 6.2 (Independent binary model)**.** *For $T = \{0,1\}^\ell \subset \mathbb{Z}^\ell$, using independent Fourier coordinates and support $T - T$ for the majorants, the exact capacity is $\ell \log 2$.*

Proof. The uniform distribution on T and its self-correlation as first majorant, with δ_0 as second, attain the bound. Conversely take each Fourier coordinate independently equal to $\pm\pi/2$. Multiplying $\widehat{\mu}$ by $e^{-i\sum\theta_j/2}$ and averaging gives $2^{-\ell/2}$ for any probability on T. Each majorant averages to its zero atom, since its other characters have a coordinate ± 1. The feasibility inequality and Cauchy–Schwarz prove optimality. □

The binary result is not an assertion that the rational components have independent real frequencies. The finite tests in Theorem 6.1 are needed to transfer a loss to the actual rational set. No exact recursive formula or limiting ratio for Ω is assumed.

7. Modular transfer and costs of repair

For a prime p and $q < p$ define

$$d(p,q) = \min\{\mathcal{H}(D) : D \subseteq [1,q],\ \mathcal{H}(D) \equiv H_q \pmod p\}.$$

Let $g(p)$ be the least height allowing all residues to be represented by ordinary subsets of mass at most one (excluding multiples of p). For $a = \Gamma, \Xi$ write $b_a(P) = \min\{n : a(n) \ge 2\log P\}$. The recurrence and capacity bounds show $b_a(P) \asymp b(P)$ and $b_a(P^t) \sim t b_a(P)$. For Ω, the same

pointwise bounds give a strict gap above $b(P)$ for its first crossing, without assuming regular variation of Ω.

Lemma 7.1 (A small suppressor)**.** *Put $L = \log P$, $\lambda = \log L$, $K = \lceil 20 \log \lambda \rceil$ and*

$$\mathcal{D} = \{ra : 100L < r \le 200L,\ r \text{ prime},\ 1 \le a \le K\}.$$

This set has height $o(b(P))$, cardinality $o(L)$ and mass $O(\log K/\lambda)$. Let σ_p be the mod-p distribution of the sum of a Bernoulli subset of $\mathcal{D}$. Apart from at most $P \exp(-cL \log \lambda/\lambda)$ primes in $[P, 2P]$, $\max_{k \neq 0} |\widehat{\sigma}_p(k)| \le e^{-2\lambda^2}$, where the finite Fourier convention uses $e^{2\pi ikh/p}$.

Proof. Suppose $\sum_{d \in \mathcal{D}} \|kd^{-1}/p\|^2 \le \lambda^2$ for $1 \le k < p$. Set $\epsilon = \lambda^2/\sqrt{L}$. At most L/λ^2 of the phases exceed ϵ. Call r good when none of its multiples in $\mathcal{D}$ does. Write $dz_d = k + pu_d$ with z_d a centered residue. For good r, $|u_{ra} + k/p| \le ra\epsilon$ and $u_{ra} \equiv u_r \pmod r$; as $(K+1)\epsilon = o(1)$ these integers coincide. Thus all u_r are in one class modulo $\mathcal{L}(K)$, of size between λ^{10} and λ^{40}. Fixing this class, each u_r has at most $\sqrt{r}\lambda^{-7}$ possibilities for large P. Select good r until their product R first exceeds $4P^2$; there are $\gg L/\lambda$ of them and $R \le 800LP^2$. Their u_r specify k/p uniquely by cross multiplication, since a nonzero cross difference has magnitude less than R. The number of possible pairs is at most

$$2^{O(L/\lambda)} \lambda^{40} \sqrt{800LP^2}\, \lambda^{-c_1 L/\lambda} \ll P e^{-c_2 L \log \lambda/\lambda}.$$

Finally $|\cos(\pi u)| \le e^{-2\|u\|^2}$ gives the Fourier bound. The size and mass bounds follow by the prime number theorem and (4). □

Lemma 7.2 (Prime-averaged majorants)**.** *Suppose K is a probability measure on $\mathcal{L}(n)^{-1}\mathbb{Z}$ with nonnegative Fourier transform and support contained in $[-2H_n, 2H_n]$. Assume $n \asymp b(P)$ and $K\{0\} \le P^{-1-\delta}$ for fixed $\delta > 0$. Then*

$$\sum_{P \le p \le 2P} \sum_{k \bmod p} \widehat{K}_p(k) \ll P\lambda.$$

In particular outside $O(P\lambda e^{-\lambda^2})$ primes the inner sum is at most e^{λ^2}.

Proof. The inner sum is $pK\{x : x \equiv 0 \pmod p\}$ by orthogonality. Each nonzero support point has a nonzero integer numerator after multiplication by $\mathcal{L}(n)$, of logarithmic height $O(n)$; at most $O(n/L)$ target primes divide it. Sum over primes first; the zero atom contributes $O(P^2K\{0\}/L)$. Markov applies because the inner sum is nonnegative. All real-frequency majorant inequalities reduce modulo p: for $x = u/\mathcal{L}(n)$ take $\theta = 2\pi k\mathcal{L}(n)\overline{\mathcal{L}(n)}/p$. □

Theorem 7.3 (Constructible cover transfer)**.** *Fix $\epsilon > 0$, take $a = \Gamma$ or Ξ and $n = \lfloor (1+\epsilon) b_a(P) \rfloor$. Outside $O_\epsilon(Pe^{-c_\epsilon \lambda^2})$ primes there is a distribution ρ_p supported on ordinary subset sums in $[1, n]$ of mass $o(1)$ with*

$$\max_{h \in \mathbb{F}_p} |p\rho_p(h) - 1| \le e^{-\lambda^2}.$$

In particular $g(p) \le (1+\epsilon) b_a(p)$ outside the same exceptional set. For some fixed $\beta < 1$, $b_\Xi(P) \le \beta b_\Gamma(P)$ for all sufficiently large P.

Proof. The total constructible capacity is $(2 + 2\epsilon + o(1))L$. Discard the root core and all top-level blocks with marker in $(100L, 200L]$, reserving those markers for Lemma 7.1. The discarded capacity is $O(n/\log n) = o(L)$, since a marker in that range contributes at most $2\lfloor n/r \rfloor \log 2$ and their number is $O(L/\lambda)$. All other top-level blocks have legal local triples. Interchange their two majorants as needed to balance the sums of negative logarithms of the zero atoms; the final imbalance is at most the largest single block capacity, $O(\log n) = o(L)$. Convolution and rigidity then give both global zero masses at most $P^{-1-\epsilon/2}$. The support of the sum distribution

μ lies in the retained blocks, whose total reciprocal mass is $O((\log\log n)^2/\log n) = o(1)$ by partial summation over $r > y_n$. Lemmas 7.1 and 7.2 and finite Fourier inversion for $\rho_p = \mu_p * \sigma_p$ yield

$$|p\rho_p(h) - 1| \le e^{-2\lambda^2} \sum_{k\neq 0} \sqrt{\widehat{K}_{1,p}(k)\widehat{K}_{2,p}(k)} \le e^{-\lambda^2}.$$

All sums can be encoded with disjoint ordinary denominator blocks, and the suppressor adds only $o(1)$ mass. Finally (11) and regular variation give $b_\Xi \le \beta b_\Gamma$ for every fixed β between $(1+\eta_2)^{-1}$ and one, eventually. Monotonicity of the first crossings in P gives the asserted bound for p. □

In particular this supplies a strict improvement of the *constructible* two-source threshold. Both first crossings have a fixed gap above $b(P)$ and below $2b(P)$ by Theorem 1.2; this is not a necessary gap for actual $g(p)$.

Corollary 7.4 (Number of legal label sequences)**.** *With $a = \Xi$ in Theorem 7.3, every residue for a good prime has at least $p^{-1}\exp(\Gamma(n) - o(L))$ legal pair-selection sequences (including the suppressor subset). In particular the bound exceeds $P^{0.9}$ for large P.*

Proof. At a height-three leaf the largest individual pair probability in the table is $41/500 < 1/12$. At any other component uniform Cartesian sources give maximum probability $e^{-\Gamma(C)}$. Multiply these probabilities over the retained components; the discarded Γ costs $o(L)$. Each sequence has probability at most $\exp(-\Gamma(n) + o(L))$, whereas the mass of a residue is $(1 + o(1))/p$. Moreover $\Xi(n) = \Gamma(n) + g_3\ell_3(n) \le (1 + g_3/\log 12)\Gamma(n)$, and $\Xi(n) \sim (2 + 2\epsilon)L$. This proves the count. Different sequences can encode the same rational value and even the same ordinary subset; no multiplicity conclusion about distinct deletion sets follows. □

Corollary 7.5 (Full coverage at the constructible threshold with optimal-order mass)**.** *Under the assumptions of Theorem 7.3, the covering distribution can also be supported on the subset sums of a single pool (independent of the target prime) with $O_\epsilon(L)$ denominators and total mass $O_\epsilon(L/b(P))$. The same exceptional-set estimate and Fourier conclusion hold. In particular on good primes, each residue has an ordinary representation of that mass and cardinality at the same height threshold $b_a(P)$ with fixed relative slack.*

Proof. Keep the reference tree defining Ξ fixed. Coarsen it at a large fixed height M, retaining all denominators of its coarse terminal blocks; discard the higher cores. For $a = \Gamma$ the component law, and for $a = \Xi$ its recursive construction, give a loss of at most $2e_M s(n)$ in capacity, because each discarded core has capacity at most twice its entropy. Within each coarse terminal block we retain the original local construction, including its lower internal cores if any. Since $s(n) = O_\epsilon(L)$, take M large enough that this loss is less than $\epsilon L/2$. The remaining capacity still exceeds $(2 + \epsilon)L$ for sufficiently large P. The denominators have total cardinality $O_M(s(n))$ and mass $O_M(s(n)/n)$, by the same nested-floor estimate as in Theorem 1.1; all exceed $n/(M + 1)$. The suppressor of Lemma 7.1 lies below this bound for large P, hence the pools are disjoint. Its size is $o(L)$ and its mass is $O(\log K/\lambda) = O(L/b(P))$ by (3). Interchange the two local majorants as needed among coarse terminal blocks. Each block has capacity at most $2M\log 2$, so the imbalance is bounded, and each resulting global zero atom is at most $P^{-1-\epsilon/3}$ eventually. The convolution proof of Theorem 7.3 applies unchanged. This is a separate low-mass construction; the near-full label-sequence count above concerns the original construction with only $o(L)$ capacity discarded. □

7.1. **Sparse signed representations.** For fixed $t > 1$ choose $1 < t_1 < t$, $m = \lfloor t_1 b(P)\rfloor$, and apply Theorem 1.1 with ε small enough that $s(G_m) > (1 + \delta)L$. The suppressor of Lemma 7.1 lies below $c_\varepsilon m$, hence is disjoint from G_m. The uniform law on distinct sums of G_m, convolved with the suppressor, has nonzero Fourier square sum at most $e^{-3\lambda^2}$ outside the strong exceptional set, by Lemma 7.2 applied to its self-correlation. Parseval shows that the fraction of missed residues is

$o(1)$: each missed residue contributes one to $|p\rho_p(h) - 1|^2$. A subset of $\mathbb{F}_p$ of size greater than $p/2$ has full difference set. Cancelling common denominators therefore proves the following statement.

Proposition 7.6. *For fixed $t > 1$, apart from $O_t(Pe^{-c_t\lambda^2})$ primes, every residue modulo $p \in [P, 2P]$ has a representation $\mathcal{H}(U) - \mathcal{H}(V)$ by disjoint subsets of $[1, \lfloor tb(P)\rfloor]$ with total cardinality $O_t(L)$ and mass $O_t(L/b(P))$. If $t > 2$, ordinary representations with the same cost exist.*

Remaining details. The mass bound for the suppressor follows from (3): $L/b(P) \gg \log_3 b(P)/\lambda \asymp \log\log\lambda/\lambda$, and $\log K = O(\log\log\lambda)$. For ordinary coverage at $t > 2$, choose $2 < t_1 < t$ and ε so that $s(G_m) > (2+\delta)L$. Partition G_m by its top-level components into two disjoint pools each with entropy exceeding $(1+\delta/4)L$ (individual block entropies are $O(\log m) = o(L)$). Use their autocorrelations as the two majorants in the proof of Theorem 7.3; this gives the same mass bound. To enforce positive mass at least three times negative mass in a signed representation one may reserve a denominator $a \asymp b(P)/L$ below both pools with $3W \le 1/a \le 4W$, where W is the total resource mass, and represent the translated target first. This last optional construction requires reserving that denominator outside the suppressor, which lies above $100L$, and outside G_m. □

7.2. **Why sparse exact repair can fail.**

Theorem 7.7 (Nested exact obstruction)**.** *Fix $1 < t_1 < t$ and $1 \le \kappa < 2$, set $m = \lfloor t_1 b(P)\rfloor$, $q = \lfloor tb(P)\rfloor$, $n = \lfloor \kappa q\rfloor$. For density-one primes in $[P, 2P]$ there exists a signed representation $\mathcal{H}(D_+) - \mathcal{H}(D_-) \equiv H_q \pmod p$ of mass $O(L/b(P))$, such that the multiset formed from $[1, q]$ by deleting D_+ and adding a second copy of every element of D_- has no equal rational sum among the ordinary subsets of $[1, n]$.*

Proof. Put $J = \lfloor\sqrt{\log q}\rfloor$. Choose a prime $r \asymp q/J$ and a prime s in $(\max(m/r, n/(2r)), q/r]$; both choices follow from the prime number theorem since $t_1/t < 1$, $\kappa/2 < 1$. Let $v = rs$, so $m < v \le q$, $r^2 > n$, and $s > \lfloor n/r\rfloor/2$. Use Proposition 7.6 at height m to represent $H_q + 1/v$ by $U - V$; put $D_+ = U, D_- = V \cup \{v\}$. The multiset has v twice. If an ordinary subset gave exactly the same sum, its multiplicity difference c_a would be bounded by two. Multiplication by r and reduction modulo r isolate the multiples ra; multiplying by $\mathcal{L}(\lfloor n/r\rfloor)$ gives an integer of absolute value $\exp(O(J)) < r$, hence zero. Reducing the resulting block relation at s isolates $c_v \in \{1, 2\}$ up to an overall sign, which cannot vanish modulo the growing prime s. This proves impossibility of exact repair of this particular representation, not impossibility of some other representation of the target. □

Theorem 7.8 (Sparse modular rigidity and repair entropy)**.** *Let $n \le Cb(P)$, with $C > 0$ fixed, and $D \ge 1$ fixed. For fixed $0 < \delta < 1$, outside $P^{1-\delta+o(1)}$ target primes, every bounded integral relation supported on at most $(1-\delta)L/\log_3 P$ denominators of $[1, n]$ is a rational identity if it is zero modulo p.*

For $0 < \delta < 1/2$, outside $P^{1-2\delta+o(1)}$ primes, any change of multiplicities from $\{0, 1, 2\}$ to $\{0, 1\}$ that preserves the residue but changes the rational sum must modify a collection of top-level blocks of total entropy greater than $(1/2-\delta)L$. The core can be changed freely. For changes consisting only of deleting a single copy per modified denominator, the lower bound becomes $(1-\delta)L$, with exceptional count $P^{1-\delta+o(1)}$.

Proof. With $K = \lfloor(1-\delta)L/\log_3 P\rfloor$, the number of vectors with coefficient magnitude at most D and support size at most K is bounded by $\sum_{j\le K}\binom{n}{j}(2D)^j$. By (4), its logarithm is at most $K\log(en/K) + O(K + \log n) \le (1 - \delta + o(1))L$ (if $n < K$, use the trivial vector count or enlarge n to $\lfloor Cb(P)\rfloor$). For every nonzero rational value, multiplication by $\mathcal{L}(n)$ gives a nonzero integer of logarithmic height $O(n)$, with at most $O(n/L) = P^{o(1)}$ target prime divisors. The union bound proves the first assertion.

The differences of multiplicities belong, up to sign, to $\{-1,0,1,2\}$. After adding one, they can be written as $u+2v$ with $u,v\in\{0,1\}$. Thus a changed block of height j_r produces at most $S(j_r)^2$ distinct values. The core produces $\exp(O(n/\log n))=P^{o(1)}$ possible values by its common denominator and the bounded coefficient condition. The choices of changed blocks contribute at most $2^{O(n/\log n)}=P^{o(1)}$. If their entropy sum is at most $(1/2-\delta)L$, all nonzero possible differences have cardinality at most $P^{1-2\delta+o(1)}$. Apply the same prime-divisor union bound. For single-copy deletions a block gives at most $S(j_r)$ values instead. This proves both statements, including arbitrary core changes. □

As an immediate application, for fixed $t>0$ and $0<\delta<1$, on a density-one prime set

$$d(p,\lfloor tb(p)\rfloor)\geq(1-\delta)\frac{\log p}{\lfloor tb(p)\rfloor\log_3 p}.$$

Indeed a smaller deletion has support below the sparse counting threshold with slack $\delta/2$, and the retained sum is nonzero. Count all such deletion sets and all possible heights $q\leq\lfloor tb(2P)\rfloor$; the latter increases the count only by $P^{o(1)}$. Apply the divisor bound to the retained sums, not to the deletion sum alone. Proposition 7.6 gives the upper bound $O_t(L/b(P))$ for $t>2$. More generally Corollary 7.5 gives this mass order at any height above the constructible threshold with fixed relative slack. A possible sufficient exact repair uses pairwise disjoint $A_v\subset U$ with $\mathcal{H}(A_v)=2/v$ for each $v\in V$: then $(U\setminus\bigcup A_v)\cup V$ has rational sum $\mathcal{H}(U)-\mathcal{H}(V)$. A predesigned matching pool outside the signed source must be included in the positive mass before representation, with a corresponding target shift.

8. Low-mass extraction under a growing denominator budget

We next connect the cost theorem to the exact rational entropy budget of the companion work [5]. This section uses that work's sharp growing-denominator theorem as a stated input, not its unresolved modular folding problem. The following argument is separate from the absorption and Fourier obstruction theorems.

For a prime p let $L=\log p$, $U=\log_2 p$, $V=\log_3 p$. With a fixed cutoff $a_*>3$, let $\mathcal{C}_j(p)$ be the product of the iterated logarithms $\log_i p$ for $i\geq j$ up to the last one at least a_*. Put $F_*=\mathcal{C}_4(p)$. The height and budget parameters of [5] are

$$c_*LU/F_*\leq Q\leq C_*LU,\qquad 1\leq K\leq\frac{Q}{L\exp(V^\gamma)},$$

where $0<c_*<C_*<\infty$ and $0<\gamma<1$ are fixed, Q is integral, and K is real. Write

$$\mathcal{F}_K=\frac{QA(V)}{U}\log\left(\frac{\log(QV/L)}{\log(Q/(KL))}\right).$$

Replacing the truncated sum in [5] by the integral A here changes this expression by a factor $1+o(1)$, since $A(V)\to\infty$. Its theorem gives $\log|\mathcal{R}_K(Q,p)|\sim\mathcal{F}_K$ for the set of all distinct rational values of reduced denominator at most p^K at height Q, and the same asymptotic for the maximal entropy of a pool with common denominator at most p^K. The lower construction in that theorem is a union of rows $r[1,M_r]$, $M_r=\lfloor Q/r\rfloor$, indexed by primes r larger than all their local heights, with all distinct local rational values injective modulo r, common denominator at most p^K, and total entropy $\sum_r s(M_r)\sim\mathcal{F}_K$.

Theorem 8.1 (Low-mass realization of the sharp rational budget)**.** *For every fixed $0<\varepsilon<1$, uniformly in the height and budget range just specified, there exists a pool $G\subseteq[c_\varepsilon Q,Q]$ with common denominator at most p^K such that*

$$s(G)\geq(1-\varepsilon)\mathcal{F}_K,\qquad|G|\ll_\varepsilon\mathcal{F}_K,\qquad\mathcal{H}(G)\ll_\varepsilon\frac{\mathcal{F}_K}{Q}.$$

Every pool in $[1, Q]$ *attaining entropy at least* $(1-\varepsilon)\mathcal{F}_K$ *has mass at least* $(1-\varepsilon)\mathcal{F}_K/(Q\log 2)$. *For* $K = 1$ *the constructed rational values are also distinct modulo* p.

Proof. Apply Theorem 1.1 in every row of the lower construction of [5], with entropy fraction $1-\varepsilon/3$. The theorem applies uniformly as all $M_r \to \infty$ in that construction (the smallest is at least $\exp(V^\gamma) - 1$); for bounded heights one could instead retain the entire row and enlarge the constants. It gives a subset $G_r \subseteq [cM_r, M_r]$ with entropy at least $(1-\varepsilon/3)s(M_r)$, cardinality $O_\varepsilon(s(M_r))$ and mass $O_\varepsilon(s(M_r)/M_r)$. Form $G = \bigcup_r rG_r$. Subsets preserve the original row-wise injectivity and the common-denominator bound. Equality of two global sums, reduced at r after clearing denominators, forces equality of the two local rational values, so entropies add exactly. Also $rM_r \geq Q/2$ and $rM_r \leq Q$. Consequently the global entropy is at least $(1-\varepsilon)\mathcal{F}_K$ for large p, the cardinality is $O_\varepsilon(\sum_r s(M_r))$, and the mass is $O_\varepsilon(\sum_r s(M_r)/Q)$. The lower bound uses $s(G) \leq |G|\log 2$ and each denominator at most Q. For $K = 1$ the lower construction of [5] already has no modular folding, a property inherited by the smaller pool. Alternatively its common denominator is strictly less than p and its mass is $o(1)$, so no nonzero difference can have integer numerator divisible by p. □

Corollary 8.2 (Budget threshold with optimal-order resource mass)**.** *Fix* $t > 1$, *take* $Q = \lfloor tb(p) \rfloor$, *and fix* $0 < \beta < 1 - 1/t$. *There are pools of entropy at least* L, *mass* $O_{t,\beta}(L/Q)$ *and cardinality* $O_{t,\beta}(L)$ *with common denominator at most* p^{K_β}, *where*

$$K_\beta = \frac{Q}{L\exp(V^\beta)}.$$

For each fixed $1-1/t < \beta < 1$, *no pool (even without a mass restriction) with common denominator at most* p^{K_β} *has entropy* L *at this height for large* p. *Thus the budget-exponent threshold* $1-1/t$ *remains the same when optimal-order reciprocal mass is required, with constants allowed to depend on the fixed slack in the exponent.*

Proof. By the exact budget theorem and its entropy-scale corollary in [5], $\mathcal{F}_{K_\beta} \sim t(1-\beta)L$. For the first assertion take fixed ε small enough that $(1-\varepsilon)t(1-\beta) > 1$ and apply Theorem 8.1. For the second, the unrestricted upper bound is already strictly below L. The resource lower bound for entropy L is $L/(Q\log 2)$, so the order of the mass in the first assertion is optimal. No claim about an optimal uniform constant as the slack tends to zero, or a full modular cover for $K_\beta > 1$, is made. □

9. Connection with square-layer losses and open questions

For N large let $B = b(\sqrt{N})$, $B_a = b_a(\sqrt{N})$ and $\Delta_2(N)$ be the sum of $p^{-2}d(p, \lfloor N/p^2 \rfloor)$ over $N^{1/3} < p \leq \sqrt{N}$. Theorem 7.3 improves the constructible bound to

$$(2-o(1))\frac{\sqrt{B}\log B}{\sqrt{N}\log N} \leq \Delta_2(N) \leq (2+o(1))\frac{\sqrt{B_a}\log B_a}{\sqrt{N}\log N}.$$

For completeness, the upper bound follows by splitting at $p = \sqrt{N/Z}$ with $Z = \lfloor (1+\epsilon)B_a \rfloor$. Above the split use $d(p,q) \leq H_q \leq \log Z + O(1)$ and partial summation of $1/p^2$. Below the split scale a full-cover representation of height at most Z by $v = \lfloor q/Z \rfloor$, giving $d(p,q) \leq 2Z/q$ for good primes. Its sum is $O(\sqrt{Z}/(\sqrt{N}\log N))$. The exceptional primes are controlled by the uniform all-prime bound from [4]: $d(p,q)/p^2 \ll Z_0 \log Z_0/N$, where $Z_0 = \lceil (\log N)^2 \rceil$. The number of exceptions on the whole square-layer interval is $O(\sqrt{N}\exp(-c(\log_2 N)^2))$ by dyadic summation. Let $\epsilon \downarrow 0$. The lower bound is the subcritical entropy argument in [4]: on each fixed ratio interval $p = v\sqrt{N/B}$ with $v > 1$, the distinct nonzero retained sums at the largest relevant height have fewer than $P^{1-\delta}$ integerized values for some $\delta > 0$. Almost all such primes divide none of them, giving $d = H_q$. Integrating v^{-2} over compact subintervals of $(1, \infty)$ gives the constant two. By Theorem 7.3, the

upper scale using B_{Ξ} improves the one using B_{Γ} by a fixed factor. No comparison with an optimal possible uniform-source bound is claimed.

The open questions include the limiting fixed-fraction cost in Theorem 3.1, the exact limiting ratios of these capacities, whether Ω obeys any useful recursive law, and the directed deletion estimate $d(p, \lfloor tb(p) \rfloor)/\log b(p) \to 0$ in prime density for every fixed $t > 1$. The last would yield the constant two by [4]; the absolute-majorant obstruction here is not its negation. The new cost profile, dilute-entropy constant, and low-mass realizations concern distinct rational values and the stated constructible covers. They do not resolve the directed threshold or the optimal all-primes modular folding problem. The finite-block certificates and proofs above are research results subject to independent review, not an originality or formal-verification claim.

References

[1] M. N. Bleicher and P. Erdős, The number of distinct subsums of $\sum_1^N 1/i$, *Math. Comp.* **29** (1975), 29–42, doi:10.2307/2005459.

[2] S. Bettin, L. Grenié, G. Molteni and C. Sanna, A lower bound for the number of Egyptian fractions, *Math. Comp.*, published online 2026, doi:10.1090/mcom/4190; arXiv:2509.10030.

[3] M. N. Kolountzakis and S. Gy. Révész, Turán's extremal problem for positive definite functions on groups, *J. Lond. Math. Soc.* (2) **74** (2006), 475–496, doi:10.1112/S0024610706023234.

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

[5] P. Chen, *Pointwise Expansion and Convolution Flattening of Modular Reciprocal Sumsets*, unpublished manuscript, 2026.