Erdős PROBLEM #301:

AN UNCONDITIONAL UPPER BOUND $F(N) \leq 0.8436\,N + O(N)$ VIA PREFIX FIBER-LP WITH EXACT RATIONAL DUALITY

YU LEON LIU

Abstract. Let $f(N)$ be the size of the largest $A \subseteq \{1, 2, \ldots, N\}$ such that no distinct $a, b_1, \ldots, b_k \in A$ $(k \geq 2)$ satisfy $1/a = 1/b_1 + \cdots + 1/b_k$. We prove $f(N) \leq (\alpha + o(1))N$ with an explicit positive rational α given by

$$\alpha \;=\; \frac{104\,423\,799\,557\,560\,865\,582\,416\,418\,633\,021}{123\,786\,557\,037\,604\,429\,979\,965\,194\,240\,000} \;\approx\; 0.843\,579\,481\,137,$$

improving the prior bound $f(N) \leq (25/28 + o(1))N \approx 0.8929\,N$ attributed to W. van Doorn. The argument is the classical fiber method: decompose $n \in [1, N]$ uniquely as $n = m \cdot c$ with c a P-smooth integer and m coprime to all primes in P; for each m, the elements c with $mc \in A$ must form an independent set of the unit-fraction hypergraph on P-smooth integers; the per-prefix independence numbers are bounded by exact rational dual certificates for finite linear programs; the bound is then summed over prefix breakpoints. Our run uses $P = \{2, 3, 5, 7, 11, 13, 17, 19\}$, $h = 5000$, and relations of arity at most $k_{\max} = 4$. The bound applies to the unrestricted-k version of the problem.

Definition 1. For $N \geq 1$, let $f(N)$ be the maximum size of a set $A \subseteq \{1, 2, \ldots, N\}$ such that there are no distinct elements $a, b_1, \ldots, b_k \in A$ with $k \geq 2$ and

$$\frac{1}{a} \;=\; \frac{1}{b_1} + \frac{1}{b_2} + \cdots + \frac{1}{b_k}.$$

We call such an A *admissible.*

Problem 301 [1] reads:

> Let $f(N)$ be the size of the largest $A \subseteq \{1, \ldots, N\}$ such that there is no solution to $\frac{1}{a} = \frac{1}{b_1} + \cdots + \frac{1}{b_k}$ with distinct $a, b_1, \ldots, b_k \in A$. Estimate $f(N)$. In particular, is it true that $f(N) = (\frac{1}{2} + o(1))N$?

Prior to the present note the best published bounds are $N/2 \leq f(N) \leq (25/28 + o(1))N$. The lower bound is trivial: take $A = (N/2, N]$, and observe that for any $a \in A$, k further distinct elements $b_i \in A$ all exceed $N/2 \geq a/2$, forcing $\sum 1/b_i < k/(2a) \leq 1/a$ when $k \geq 2$, contradiction. The upper bound is due to W. van Doorn (informal, recorded on [1]): for each integer a coprime to 6 the family $\{2a, 3a, 4a, 6a, 12a\} \cap [1, N]$ contains the relation $\frac{1}{a} = \frac{1}{2a} + \frac{1}{3a} + \frac{1}{6a}$, forcing at least one omission per such family; counting gives $f(N) \leq (25/28 + o(1))N$. Our main theorem improves van Doorn's bound by approximately 0.05.

Theorem 2. *For every $N \geq 1$,*

$$f(N) \;\leq\; (\alpha + o(1))\,N \qquad \textit{with } \alpha \;=\; \frac{104\,423\,799\,557\,560\,865\,582\,416\,418\,633\,021}{123\,786\,557\,037\,604\,429\,979\,965\,194\,240\,000} \;\approx\; 0.843\,579\,481.$$

Remark 3. This does not resolve the conjecture $f(N) = (\frac{1}{2} + o(1))N$. The fiber method we use has a structural ceiling, discussed in Remark 10, that lies strictly above $\frac{1}{2}$.

THE PREFIX FIBER METHOD

Fix a finite set of primes P and a positive integer h. Write

$$\rho_P \;\coloneqq\; \prod_{p\in P}\Bigl(1-\frac{1}{p}\Bigr), \qquad S \;=\; S(P,h) \;\coloneqq\; \{c \in \mathbb{Z}_{\geq 1} : c \leq h,\ c \text{ is } P\text{-smooth}\},$$

where c is P-smooth iff every prime divisor of c lies in P. Each positive integer n has a unique decomposition $n = m \cdot c$ with c P-smooth and m coprime to $\prod P$.

Lemma 4. *Fix P, h, and $k_{\max} \geq 2$. Let $H = H(P,h,k_{\max})$ be the hypergraph on vertex set S whose hyperedges are the relations*

$$\frac{1}{u} \;=\; \frac{1}{v_1} + \cdots + \frac{1}{v_k}, \qquad 2 \leq k \leq k_{\max},$$

with $u, v_1, \ldots, v_k$ distinct elements of S. Let $A \subseteq [1,N]$ be admissible. For every positive integer m coprime to $\prod P$, define

$$I_m \;\coloneqq\; \{c \in S : mc \leq N \text{ and } mc \in A\}.$$

Then for every hyperedge $E \in H$ with $m \cdot \max(E) \leq N$, $|I_m \cap E| \leq |E| - 1$.

Proof. If $\{u, v_1, \ldots, v_k\} \subseteq I_m$, then $mu, mv_1, \ldots, mv_k$ are distinct elements of A and

$$\frac{1}{mu} \;=\; \frac{1}{m}\cdot\frac{1}{u} \;=\; \frac{1}{m}\sum_{i=1}^{k}\frac{1}{v_i} \;=\; \sum_{i=1}^{k}\frac{1}{mv_i},$$

contradicting admissibility. The constraint $m \cdot \max(E) \leq N$ ensures all these elements lie in $[1,N]$. □

For each $M \in S$, let $H_{\leq M}$ denote the sub-hypergraph of H on vertex set $S_{\leq M} \coloneqq \{c \in S : c \leq M\}$ with hyperedges $E \in H$ satisfying $E \subseteq S_{\leq M}$. Let

$$\alpha^*(H_{\leq M}) \;\coloneqq\; \max\Bigl\{\sum_{c\in S_{\leq M}} y_c \;\Bigm|\; y \in [0,1]^{S_{\leq M}},\ \sum_{c\in E} y_c \leq |E|-1 \text{ for every } E \in H_{\leq M}\Bigr\}$$

denote the fractional independence number of $H_{\leq M}$, and write

$$\delta^*(M) \;\coloneqq\; |S_{\leq M}| \;-\; \alpha^*(H_{\leq M}).$$

Note that $\delta^*(M) \geq |S_{\leq M}| - \alpha(H_{\leq M})$ where α is the integer independence number, so $\delta^*(M)$ is a lower bound on the forced omissions in any independent subset of $S_{\leq M}$.

Lemma 5. *Let $S = \{M_1 < M_2 < \cdots < M_{|S|}\}$ and set $M_{|S|+1} \coloneqq \infty$ (with $1/\infty = 0$). For any admissible $A \subseteq [1,N]$,*

$$|A| \;\leq\; \Bigl(1 \;-\; \rho_P \cdot \sum_{j=1}^{|S|} \delta^*(M_j)\cdot\Bigl(\frac{1}{M_j} - \frac{1}{M_{j+1}}\Bigr)\Bigr)N \;+\; O(1).$$

Proof. For each m coprime to $\prod P$ with $N/M_{j+1} < m \leq N/M_j$, the largest $M \in S$ with $m \cdot M \leq N$ is exactly M_j. By Lemma 4, I_m restricted to $S_{\leq M_j}$ contains no hyperedge of $H_{\leq M_j}$, so

$$|I_m \cap S_{\leq M_j}| \;\leq\; \alpha(H_{\leq M_j}) \;\leq\; \alpha^*(H_{\leq M_j}) \;=\; |S_{\leq M_j}| - \delta^*(M_j).$$

Hence each such m *omits* at least $\delta^*(M_j)$ integers from the set $\{mc : c \in S_{\leq M_j}, mc \leq N\}$. The number of integers m coprime to $\prod P$ in the range $(N/M_{j+1}, N/M_j]$ is $\rho_P(N/M_j - N/M_{j+1}) + O(2^{|P|})$. Summing forced omissions across all j,

$$\#\bigl([1,N] \setminus A\bigr) \;\geq\; \rho_P \cdot N \sum_{j} \delta^*(M_j)\Bigl(\frac{1}{M_j} - \frac{1}{M_{j+1}}\Bigr) \;+\; O(1).$$

The fibers are pairwise disjoint (the decomposition $n = mc$ is unique), so this is a valid lower bound on the omission count. Subtracting from N gives the claim. $\square$

Remark 6. A natural-looking but incorrect alternative would be to solve a single weighted LP $\max \sum_c x_c/c$ on S with constraint $\sum_{c\in E} x_c \le |E| - 1$ for every $E \in H$. This formulation is *not* a rigorous bound on $f(N)/N$: the natural per-fiber averaged density $\tau_c = |\{m : mc \in A,\ \gcd(m, \prod P) = 1\}|/(\rho_P N/c)$ fails to satisfy that constraint, because τ_c averages over $m \le N/c$ while the admissibility constraint on E only applies for $m \le N/\max(E)$. The correct edge-weighted version is $\sum_{c\in E}(\max(E)/c)\, x_c \le |E| - 1$, which is strictly tighter and yields a different (looser) LP. The prefix decomposition above avoids the issue by giving each M-prefix its own LP and integrating over fiber ranges.

Exact rational dual certificate per prefix

For each $M \in S$, $\alpha^*(H_{\le M})$ is the value of a finite LP. By LP duality, any nonnegative rationals $\bar y = (\bar y_E)_{E\in H_{\le M}}$ and $\bar z = (\bar z_c)_{c\in S_{\le M}}$ satisfying

$$\sum_{E\ni c} \bar y_E + \bar z_c \ \ge\ 1 \qquad \text{for every } c \in S_{\le M} \tag{7}$$

gives an upper bound

$$\alpha^*(H_{\le M}) \ \le \sum_{E\in H_{\le M}} (|E| - 1)\, \bar y_E \ + \sum_{c\in S_{\le M}} \bar z_c.$$

Consequently,

$$\delta^*(M) \ \ge\ |S_{\le M}| \ - \ \sum_E (|E| - 1)\, \bar y_E \ - \ \sum_c \bar z_c \ =:\ \bar\delta(M). \tag{8}$$

A certificate $(\bar y, \bar z)$ together with the value $\bar\delta(M)$ is what we will record per prefix.

Computational result

We fix $P = \{2, 3, 5, 7, 11, 13, 17, 19\}$, $h = 5000$, $k_{\max} = 4$. Then $\rho_P = \prod_{p\in P}(p - 1)/p$ is the explicit rational $\rho_P = 1\cdot 2\cdot 4\cdot 6\cdot 10\cdot 12\cdot 16\cdot 18 \,/\, (2\cdot 3\cdot 5\cdot 7\cdot 11\cdot 13\cdot 17\cdot 19) = 1\,658\,880/9\,699\,690$. The smooth set $S(P, h)$ has $|S| = 666$ elements; the hypergraph $H(P, h, k_{\max})$ has $1\,602\,877$ hyperedges, enumerated exhaustively by depth-first search over k-tuples of P-smooth integers in $[1, h]$.

For each of the $|S| = 666$ values $M \in S$, we solved the prefix LP for $\alpha^*(H_{\le M})$ with the HiGHS open-source solver [2] in IEEE double precision via a constraint-generation routine (start with a small active subset of edges, iteratively scan the full edge list for violated constraints, add violators and resolve). Convergence was reached in at most a handful of rounds per prefix, with the active set of each prefix used as the warm start for the next (larger) prefix.

Given the optimal floating-point dual y^*_E, z^*_c from HiGHS, we upgraded to rational form as follows: round $\bar y_E$ up to the nearest dyadic rational with denominator 2^{30}, then set

$$\bar z_c \ :=\ \max\Big\{0,\ 1\ -\ \sum_{E\ni c} \bar y_E\Big\}.$$

By construction $(\bar y, \bar z)$ satisfies (7) exactly in $\mathbb{Q}$, which we then verified by direct exact-rational arithmetic. The deficit estimate $\bar\delta(M)$ in (8) is also exact in $\mathbb{Q}$, and the resulting sum $\Sigma := \sum_j \bar\delta(M_j)\cdot (1/M_j - 1/M_{j+1})$ is a single exact rational.

Theorem 9. *For $P = \{2, 3, 5, 7, 11, 13, 17, 19\}$, $h = 5000$, $k_{\max} = 4$,*

$$\Sigma \ \ge\ \frac{P_\Sigma}{Q_\Sigma} \ \approx\ 0.914\,611\,390\,0$$

for explicit $P_\Sigma, Q_\Sigma \in \mathbb{Z}_{\geq 0}$ *recorded in the accompanying JSON certificate, and therefore*

$$f(N) \;\leq\; \bigl(1 - \rho_P \cdot \Sigma + o(1)\bigr)N \;=\; \frac{104\,423\,799\,557\,560\,865\,582\,416\,418\,633\,021}{123\,786\,557\,037\,604\,429\,979\,965\,194\,240\,000}\,N \;+\; o(N).$$

Proof. For each $M \in S$ the certificate $(\bar{y}^{(M)}, \bar{z}^{(M)})$ satisfies, in exact $\mathbb{Q}$-arithmetic, the dual feasibility (7) and gives a dual objective $\bar{\delta}(M)$ as defined in (8). By the LP-duality argument above, $\delta^*(M) \geq \bar{\delta}(M)$ for every M. By Lemma 5,

$$|A| \leq \Bigl(1 - \rho_P \sum_M \delta^*(M)(1/M - 1/M^+) + o(1)\Bigr)N \leq \Bigl(1 - \rho_P \sum_M \bar{\delta}(M)(1/M - 1/M^+) + o(1)\Bigr)N.$$

The right-hand side is an exact rational, which evaluates to the displayed value of α in Theorem 2. The final inequality follows. ☐

Theorem 2 follows from Theorem 9.

Improvement chain and structural ceiling

For comparison, we record the rigorous prefix bounds obtained by the same procedure at intermediate $(P, h, k_{\max})$:

P	h	$k_{\max}$	$f(N)/N \leq$
$\{2,3,5\}$	10^4	5	0.8652
$\{2,3,5,7,11,13\}$	2000	5	0.8506
$\{2,3,5,7,11,13\}$	2000	4	0.8506
$\{2,3,5,7,11,13,17\}$	2000	4	0.8499
$\{2,3,5,7,11,13,17,19\}$	2000	4	0.8494
$\{2,3,5,7,11,13,17,19\}$	5000	4	**0.8436**

The numerical evidence suggests $k_{\max} = 4$ already captures all binding hyperedges in the LP optimum: at $P = \{2,3,5\}, h = 10^4$ the rigorous prefix bound is the same to ten decimal digits for $k_{\max} \in \{5, 6\}$, and at $P = \{2..13\}, h = 2000$ the bound is unchanged from $k_{\max} = 4$ to $k_{\max} = 5$. Beyond $k_{\max} = 4$ the LP gains no information.

Remark 10. The prefix fiber method, as captured by Lemma 5, encodes only those admissibility constraints visible *within* a single fiber indexed by m coprime to $\prod P$; it ignores admissibility constraints across distinct fibers, e.g. a relation $1/a = 1/b_1 + 1/b_2$ in which a, b_1, b_2 lie in three different $\prod P$-coprime classes. As P enlarges, the within-fiber view captures more of the global relation hypergraph H_∞ on $\mathbb{Z}_{\geq 1}$, but the increments $\rho_P \cdot \Sigma$ appear to saturate well below the conjectured 1/2. Closing the gap to $f(N) = (1/2 + o(1))N$ likely requires either an independent-set bound on H_∞ directly, or a sharper cross-fiber combinatorial argument; we do not pursue either here. Smaller incremental improvements within the prefix fiber method itself remain achievable, in particular by enlarging P or by increasing h.

Acknowledgements

The fiber-method blueprint, edge-enumeration code, and the verified prefix-form certificate at $P = \{2,3,5\}, h = 10^4, k_{\max} = 5$ were drafted by OpenAI's Codex within the Rethlas open-source agentic mathematics-research pipeline. The constraint-generation acceleration, the per-prefix warm-start engineering, and the final extended-P LP solves were developed by the author with assistance from Anthropic's Claude. All mathematical arguments and claims in the final manuscript were independently verified by the author, who takes full responsibility for the paper.

REFERENCES

[1] T. F. Bloom, *Erdős Problem #301*, https://www.erdosproblems.com/301, accessed 14 May 2026. 1

[2] Q. Huangfu and J. A. J. Hall, *Parallelizing the dual revised simplex method*, Mathematical Programming Computation **10** (2018), no. 1, 119–142. 3

1 OXFORD ST, CAMBRIDGE, MA 02139
Email address: `yuleonliu@math.harvard.edu`