

SUMS OF DISTINCT UNIT FRACTIONS

PAUL ERDÖS AND SHERMAN STEIN

We shall consider the representation of numbers as the sum of distinct unit fractions; in particular we will answer two questions recently raised by Herbert S. Wilf.

A sequence of positive integers $S = \{n_1, n_2, \dots\}$ with $n_1 < n_2 < \dots$ is an R -basis if every positive integer is the sum of distinct reciprocals of finitely many integers of S . In Research Problem 6 [1, p. 457], Herbert S. Wilf raises several questions about R -bases, including: Does an R -basis necessarily have a positive density? If S consists of all positive integers and $f(n)$ is the least number required to represent n , what, in some average sense, is the growth of $f(n)$? These two questions are answered by Theorems 1 and 5 below. Theorem 4 is a "best-possible" strengthening of Theorem 1.

THEOREM 1. *There exists a sequence S of density zero such that every positive rational is the sum of a finite number of reciprocals of distinct terms of S .*

The proof depends on two lemmas.

LEMMA 1. *Let r be real, $0 < r < 1$ and a_1, a_2, \dots integers defined inductively by*

$$a_1 = \text{smallest integer } n, r - \frac{1}{n} \geq 0,$$

$$\begin{array}{l} a_2 = \text{smallest integer } n, r - \frac{1}{a_1} - \frac{1}{n} \geq 0, \\ \vdots \\ \vdots \end{array}$$

$$a_k = \text{smallest integer } n, r - \frac{1}{a_1} - \frac{1}{a_2} - \dots - \frac{1}{a_{k-1}} - \frac{1}{n} \geq 0.$$

Then $a_{i+1} > a_i(a_i - 1)$ for each i . Also if r is rational the sequence terminates at some k , that is $r = \sum_{i=1}^k 1/a_i$.

Lemma 1 is due to Sylvester [2]. It provides a canonical representation for each positive real less than 1 which we will call the Sylvester representation.

Presented to the Society January 29, 1962; received by the editors January 29, 1962.

LEMMA 2. If r is a positive rational and A a positive integer then there exists a finite set of integers $S(r, A) = \{n_1, n_2, \dots, n_k\}$, $n_1 < n_2 < \dots < n_k$ such that

$$\begin{aligned} r &= \sum_{i=1}^k \frac{1}{n_i}, \\ n_1 &\geq A, \\ n_{i+1} - n_i &\geq A \qquad 1 \leq i \leq k-1. \end{aligned}$$

PROOF. Since the harmonic series diverges, there is an integer m such that

$$r - \left(\frac{1}{A} + \frac{1}{2A} + \dots + \frac{1}{3A} + \dots + \frac{1}{mA} \right) < \frac{1}{(m+1)A}.$$

Now applying Lemma 1 to

$$r - \left(\frac{1}{A} + \frac{1}{2A} + \frac{1}{3A} + \dots + \frac{1}{mA} \right)$$

we conclude that there are integers $m_1 < m_2 < \dots < m_s$ such that

$$r - \left(\frac{1}{A} + \frac{1}{2A} + \dots + \frac{1}{mA} \right) = \sum_{i=1}^s \frac{1}{m_i}.$$

By our choice of m we see that $m_1 > (m+1)A$. Moreover Lemma 1 assures us that $m_{i+1} - m_i > A$. Then

$$\{A, 2A, \dots, mA, m_1, m_2, \dots, m_s\}$$

serves as $S(r, A)$.

Now the proof of Theorem 1 is immediate. Order the rationals r_1, r_2, r_3, \dots . Let S_1 be an $S(r_1, 1)$. Let b_1 be the largest element of $S(r_1, 1)$. Let S_2 be an $S(r_2, 2b_1)$. Having defined S_1, S_2, \dots, S_k defines S_{k+1} as follows. Let b_k be the largest element of S_k . Let S_{k+1} be an $S(r_{k+1}, 2b_k)$.

Then since S_k 's are disjoint, there is a monotonically increasing bijection $S: (1, 2, 3, \dots) \rightarrow \bigcup_{k=1}^{\infty} S_k$ which satisfies the demands of Theorem 1.

In fact S does more than Theorem 1 asserted. It is possible to represent all the positive rationals by sums of reciprocals of terms in the S constructed so that each such reciprocal appears in the representation of precisely one rational. Similar reasoning proves

THEOREM 2. The set of unit fractions $\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$ can be partitioned

into disjoint finite subsets S_1, S_2, \dots such that each positive rational is the sum of the elements of precisely one S_i .

Theorem 2 remains true if the phrase "each positive rational," is replaced by "each positive integer." It would be interesting to know the necessary and sufficient condition that a sequence of rationals r_1, r_2, r_3, \dots corresponds to the sums of a partition of the set of unit fractions into disjoint finite subsets.

THEOREM 3. *If n_1, n_2, n_3, \dots , is a sequence of positive integers with (1) $n_{k+1} \geq n_k(n_k - 1) + 1$, for $k = 1, 2, 3, \dots$ and (2) for an infinity of k , $n_{k+1} > n_k(n_k - 1) + 1$ then $\sum_{k=1}^{\infty} 1/n_k$ is irrational.¹*

PROOF. Observe first that if a_1, a_2, \dots is a sequence of positive integers with $a_{k+1} = a_k(a_k - 1) + 1$ for $k = 1, 2, 3, \dots$, and $a_1 > 1$, then $\sum_{k=1}^{\infty} 1/a_k = 1/(a_1 - 1)$. By assumption (2) there is h such that $n_h > 1$. From the observation we see that for any integer i ,

$$\frac{1}{n_{i+1}} < \sum_{k=h}^{\infty} \frac{1}{n_k} - \sum_{n=h}^i \frac{1}{n_k} < \frac{1}{n_{i+1} - 1}.$$

Thus the Sylvester representation of $\sum_{k=h}^{\infty} 1/n_k$ is $1/n_h + 1/n_{h+1} + 1/n_{h+2} + \dots$. Since the Sylvester representation of $\sum_{n=h}^{\infty} 1/n_k$ has an infinite number of terms, we see by Theorem 1 that $\sum_{n=h}^{\infty} 1/n_k$ is irrational. Hence so is $\sum_{n=1}^{\infty} 1/n_k$ irrational.

We will soon strengthen Theorem 1 by Theorem 4 for which we will need

LEMMA 3. *The number of integers in $(x, 2x)$ all of whose prime factors are $\leq x^{1/2}$ is greater than $x/10$ for $x > x_0$.*

PROOF. The number of these integers is at least $x - \sum_{p_i} (x/p_i)$, where the summation extends over the primes $x^{1/2} < p_i < 2x$. From the fact that $\sum_{p < y} 1/p = \log \log y + c + o(1)$ Lemma 3 easily follows.

THEOREM 4. *Let $0 < a_1 < a_2 < \dots$ be a sequence A of integers with $\sum_{n=1}^{\infty} 1/a_n = \infty$. Then there exists a sequence $B: b_1 < b_2 < \dots$ of integers satisfying $a_n < b_n$, $1 \leq n < \infty$, such that every positive rational is the sum of the reciprocals of finitely many distinct b 's.*

PROOF. Set $A(x) = \sum_{a_i < x} 1$. We omit from A all the a_i , $2^k \leq a_i < 2^{k+1}$ for which

$$(1) \quad A(2^{k+1}) - A(2^k) < 2^k/k^2.$$

Thus we obtain a subsequence A' of A , $a'_1 < a'_2 < \dots$. Clearly $\sum_{n=1}^{\infty} 1/a'_n = \infty$, since, by (1), the reciprocals of the omitted a 's con-

¹ Added in proof. A similar result is to be found in [2].

verges.

Set $A'(x) = \sum_{a_i < x} 1$. Denote by $k_1 < k_2 < \dots$ the integers for which

$$(2) \quad t_{k_i} = A'(2^{k_i+1}) - A'(2^{k_i}) \geq 2^{k_i}/k_i^2.$$

By (2), if $m \neq k_i$ then $A'(2^{m+1}) = A'(2^m)$.

By Lemma 3 there are at least $(t_{k_i})/10$ integers in $(2^{k_i+1}, 2^{k_i+2})$ all of whose prime factors are less than $2^{(k_i+1)/2}$. Denote such a set of integers by $b_1^{(i)} < b_2^{(i)} < \dots < b_{q_i}^{(i)}$ where q_i is, say, the first integer larger than $t_{k_i}/10$. Clearly

$$\sum_{r=1}^{q_i} 1/b_r^{(i)} > (1/40) \sum 1/a_j' \quad (2^{k_i} < a_j' < 2^{k_i+1}).$$

Thus from $\sum 1/a_i' = \infty$ we have

$$(3) \quad \sum_{i=1}^{\infty} \sum_{r=1}^{q_i} 1/b_r^{(i)} = \infty.$$

Clearly $b_{q_i}^{(i)} < b_1^{(i+1)}$; thus all the b 's can be written in an increasing sequence $D: d_1 < d_2 < \dots$.

Now let $u_1/v_1, u_2/v_2, \dots$ be a well-ordering of the positive rationals. Suppose we have already constructed $b_1 < b_2 < \dots < b_{m_n}$ so that $a_i' < b_i, 1 \leq i \leq m_n$ and that $u_r/v_r, 1 \leq r < n$, are the sums of reciprocals of distinct b 's. Choose

$$(4) \quad 2^{k_i} > \max\{v_n, b_{m_n}, a'_{m_n} + 1\}$$

and let $d_{j_i+1} < d_{j_i+2} < \dots$ be the d 's greater than 2^{k_i+1} . By (3) and (4) there is an $s_i > j_i$ such that

$$(5) \quad \sum_{r=j_i+1}^{s_i} 1/d_r < u_n/v_n \leq \sum_{r=j_i+1}^{1+s_i} 1/d_r.$$

By (5)

$$(6) \quad 0 < u_n/v_n - \sum_{j_i+1}^{s_i} 1/d_r = C_n/D_n < 1/d_{s_i}.$$

Let x be the integer such that $2^x < d_{s_i} \leq 2^{x+1}$; then $x = k_{s_i+1}$ for some $s \geq i$ (by definition of the d 's). Since, by definition, all the prime factors of $d_r, j_i \leq r \leq q_i$ are less than $2^{(x+1)/2}$ we have

$$(7) \quad D_n \leq v_n [d_{j_i+1}, d_{j_i+2}, \dots, d_{s_i}] < v_n (2^{x+1})^{2^{(x+1)/2}} < 2^x (2^{x+1})^{2^{(x+1)/2}} < 2^{2^{x/2}}$$

for $x > x_0$.

Now

$$(8) \quad \frac{C_n}{D_n} = \frac{1}{y_1} + \cdots + \frac{1}{y_f}, \quad f < C^* \log D_n < C2^{2x/3}$$

with, clearly, $d_{s_i} < y_1 < \cdots < y_f$ (by [3]).

Define

$$b_{m_n+t} = d_{j_i+t} \quad \text{for } t = 1, \cdots, s_i - j_i, \\ \bar{b}_{m_n+s_i-j_i+t'} = y_{t'} \quad \text{for } 1 \leq t' \leq f.$$

By (8) the b 's are distinct. Clearly $b_{m_n+t} > a_{m_n+t}$ for $t = 1, \cdots, s_i - j_i$ since $b_{m_n+t} = d_{j_i+t}$, and the d 's are greater than the corresponding a 's, which in turn are greater than the a 's. By (8) the y 's do not change the situation. Their number is at most $C2^{2x/3}$. But by (2) there are at least

$$2^{k_s}/k_s^2 > 2^{x-1}/x^2, \quad x = k_s + 1$$

a_i 's in $(2^{k_s}, 2^{k_s+1})$ and by definition to more than half of them there does not correspond any d_i ; thus to those a_i 's to which no d corresponds we can make correspond the $f < C2^{2x/3}$ y 's since clearly $C2^{2x/3} < 2^{x-1}/x^2$, if $x > x_0$.

The proof is then completed as for Theorem 1. Note that each b_i is used in the representation of only one rational number.

Theorem 4 is a best possible result since if $\sum_{i=1}^{\infty} 1/a_i < \infty$ the conclusion could not possibly hold.

In the next theorem γ is Euler's constant.

THEOREM 5. $\lim_{n \rightarrow \infty} f(n)e^{-n} = e^{-\gamma}$.

PROOF. Define $g(n)$ by

$$\frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{g(n)} < n < \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{g(n)} + \frac{1}{g(n)+1}.$$

Then $n - \sum_{i=1}^{g(n)} 1/i$ is a rational number less than 1 which we denote a_n and which can be expressed in the form

$$a_n = \frac{A}{[1, 2, \cdots, g(n)]}$$

for some integer A .

Now, $0 < u/v < 1$ can be represented as the sum of less than

$$\frac{c \log v}{\log \log v}$$

distinct unit fractions [3].

Thus a_n is the sum of fewer than

$$\frac{c \log [1, 2, \dots, g(n)]}{\log \log [1, 2, \dots, g(n)]}$$

unit fractions (each less than $1/g(n)$). The expression $\log [1, 2, \dots, g(n)]$ is asymptotic to $g(n)$ [4, p. 362]. Thus for large n , a_n is the sum of fewer than

$$\frac{cg(n)}{\log g(n)}$$

distinct unit fractions.

Hence

$$g(n) < f(n) < g(n) + \frac{cg(n)}{\log g(n)}.$$

Thus

$$\lim_{n \rightarrow \infty} f(n)/g(n) = 1.$$

From the equation

$$n = \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{g(n)} + a_n = \log g(n) + \epsilon_n + a_n + \gamma$$

with $\lim_{n \rightarrow \infty} \epsilon_n = 0$ and $\lim_{n \rightarrow \infty} a_n = 0$, it follows that $g(n)$ is asymptotic to e^n/e^γ .

This proves Theorem 5.

REFERENCES

1. Herbert S. Wilf, *Reciprocal bases for integers*, Bull. Amer. Math. Soc. **67** (1961), 456.
2. J. J. Sylvester, *On a point in the theory of vulgar fractions*, Amer. J. Math. **3** (1880), 332-335, 388-389.
3. Paul Erdős, *The solution in whole numbers of the equation: $1/x_1 + 1/x_2 + 1/x_3 + \dots + 1/x_N = a/b$* , Mat. Lapok **1** (1950), 192-210.
4. G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 3rd. ed., Oxford Univ. Press, New York, 1954.

TECHNION, HAIFA AND
UNIVERSITY OF CALIFORNIA AT DAVIS