Sums of Distinct Unit Fractions

Paul Erdős 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 \( S = \{n_1, n_2, \ldots \} \) with \( n_1 < n_2 < \ldots \) 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 \( a_1, a_2, \ldots \) integers defined inductively by

\[
\begin{align*}
a_1 &= \text{smallest integer } n, \quad r - \frac{1}{n} \geq 0, \\
a_2 &= \text{smallest integer } n, \quad r - \frac{1}{a_1} - \frac{1}{n} \geq 0, \\
&\vdots \\
a_k &= \text{smallest integer } n, \quad r - \frac{1}{a_1} - \frac{1}{a_2} - \ldots - \frac{1}{a_{k-1}} - \frac{1}{n} \geq 0.
\end{align*}
\]

Then \( a_{k+1} > a_i(a_i - 1) \) 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.

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

$$r = \sum_{i=1}^{k} \frac{1}{n_i},$$

$$n_i \geq A,$$

$$n_{i+1} - n_i \geq A \quad 1 \leq i \leq k - 1.$$

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

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

Now applying Lemma 1 to

$$r - \left(\frac{1}{A} + \frac{1}{2A} + \cdots + \frac{1}{mA}\right) = \sum_{i=1}^{k} \frac{1}{n_i}.$$

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

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

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

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

serves as $S(r, A)$.

Now the proof of Theorem 1 is immediate. Order the rationals $r_1, r_2, r_3, \ldots$. Let $S_1$ be an $S(r_1, 1)$. Let $b_1$ be the largest element of $S(r_2, 2b_1)$. Having defined $S_1, S_2, \ldots, S_k$ 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 $S: (1, 2, 3, \ldots) \rightarrow \bigcup_{k=1}^{n} S_k$ 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 $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \ldots$ can be partitioned
into disjoint finite subsets $S_1, S_2, \ldots$ 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 $r_1, r_2, r_3, \ldots$ corresponds to the sums of a partition of the set of unit fractions into disjoint finite subsets.

**Theorem 3.** If $n_1, n_2, n_3, \ldots$, is a sequence of positive integers with

1. $n_{k+1} \geq n_k(n_k-1) + 1$, for $k = 1, 2, 3, \ldots$ and
2. for an infinity of $k$, $n_{k+1} > n_k(n_k-1) + 1$ then $\sum_{k=1}^{n} 1/n_k$ is irrational.1

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

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

Thus the Sylvester representation of $\sum_{h=n}^{i} 1/n_h$ is $1/n_h + 1/n_{h+1} + 1/n_{h+2} + \cdots$. Since the Sylvester representation of $\sum_{n=2}^{\infty} 1/n_k$ has an infinite number of terms, we see by Theorem 1 that $\sum_{n=2}^{\infty} 1/n_k$ 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 $x - \sum_{p \leq x} (x/p)$, where the summation extends over the primes $x^{1/2} < p < 2x$. From the fact that $\sum_{p \leq x} 1/p = \log \log x + c + o(1)$ Lemma 3 easily follows.

**Theorem 4.** Let $0 < a_1 < a_2 < \cdots$ be a sequence $A$ of integers with $\sum_{n=1}^{\infty} 1/a_n = \infty$. Then there exists a sequence $B: b_1 < b_2 < \cdots$ 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 $A(x) = \sum_{a \leq x} 1$. 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 $A, a_1' < a_2' < \cdots$. 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 \( A'(x) = \sum u_i < x \). Denote by \( k_1 < k_2 < \cdots \) 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 \(2(2k_i+1)/2\). Denote such a set of integers by \( b_1^{(i)} < b_2^{(i)} < \cdots < b_{q_i}^{(i)} \) 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 \( b_1^{(i)} < b_{(i+1)}^{(i)} \); thus all the \( b_i \)'s can be written in an increasing sequence

\( d_1 < d_2 < \cdots \).

Now let \( u_1/v_1, u_2/v_2, \cdots \) be a well-ordering of the positive rationals. Suppose we have already constructed \( b_1 < b_2 < \cdots < b_m \) so that

\( a_i < b_i, 1 \leq i \leq m \), and that \( u_r/v_r, 1 \leq r < n \), are the sums of reciprocals of distinct \( b_i \)'s. Choose

\[
(4) \quad 2^{k_i} > \max\{v_n, b_m, a_m + 1\}
\]

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

\[
(5) \quad \sum_{r=j_i+1}^{q_i} 1/d_r < u_n/v_n \leq \sum_{r=j_{i+1}}^{1+q_i} 1/d_r.
\]

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 \( x = k_{x+1} \) for some \( s \geq i \) (by definition of the \( d_i \)'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

\[
(7) \quad D_n \leq v_n[d_{j_{i+1}}, d_{j_{i+2}}, \cdots, d_{s_i}] < v_n(2^{x+1})^{2^{(x+1)/2}} < 2^x(2^{x+1})^{2^{(x+1)/2}} < 2^{2x/2}
\]

for \( x > x_0 \).
Now
\[
\frac{C_n}{D_n} = \frac{1}{y_1} + \cdots + \frac{1}{y_r}, \quad f < C^* \log D_n < C 2^{2s/3}
\]
with, clearly, \( d_s < y_1 < \cdots < y_r \) (by [3]).

Define
\[
b_{m+n+t} = d_{j_i+t} \quad \text{for } t = 1, \ldots, s_i - j_i,
\]
\[
b_{m+n+s_i-j+t'} = y_{t'} \quad \text{for } 1 \leq t' \leq f.
\]

By (8) the \( b \)'s are distinct. Clearly \( b_{m+n+t} > a_{m+n+t} \) for \( t = 1, \ldots, s_i - j_i \)
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 \( C 2^{2x/3} \). But by (2) there are at least
\[
2^{k_1^2}/k_2 > 2^{x-1}/x^2, \quad x = k_x + 1
\]

\( a_i \)'s in \((2^{k_1}, 2^{k_1+1})\) 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 \( f < C 2^{2s/3} \) \( y \)'s since clearly \( C 2^{2s/3} < 2^{x-1}/x^2 \), 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 \( \gamma \) is Euler's constant.

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

**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

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

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
\[
\frac{c \log [1, 2, \ldots, g(n)]}{\log \log [1, 2, \ldots, g(n)]}
\]
unit fractions (each less than \( 1/g(n) \)). The expression
\[
\log [1, 2, \ldots, g(n)]
\]
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
\[
\lim_{n \to \infty} f(n)/g(n) = 1.
\]
From the equation
\[
n = \frac{1}{1} + \frac{1}{2} + \cdots + \frac{1}{g(n)} + a_n = \log g(n) + \epsilon_n + a_n + \gamma
\]
with \( \lim_{n \to \infty} \epsilon_n = 0 \) and \( \lim_{n \to \infty} a_n = 0 \), it follows that \( g(n) \) is asymptotic
to \( e^\gamma \).
This proves Theorem 5.

REFERENCES

3. Paul Erdös, The solution in whole numbers of the equation: \( 1/x_1+1/x_3+1/x_1+\cdots+1/x_N = a/b \), Mat. Lapok 1 (1950), 192–210.

TECHNION, HAIFA AND
UNIVERSITY OF CALIFORNIA AT DAVIS