Journal of Number Theory **89**, 193–211 (2001)
doi:10.1006/jnth.2000.2646, available online at http://www.idealibrary.com on IDEAL®

Representing Powers of Numbers as Subset Sums of Small Sets

David Petrie Moulton

Center For Communications Research, 29 Thanet Rd., Princeton, New Jersey 08540
E-mail: moulton@idaccr.org

Communicated by T. Y. Lam

Received July 20, 1999; published online April 9, 2001

We investigate a number of questions concerning representations of a set of numbers as sums of subsets of some other set. In particular, we obtain several results on the possible sizes of the second set when the first set consists of a geometric sequence of integers, partially answering a generalisation of a question of Gerry Myerson. © 2001 Academic Press

Key Words: number theory; combinatorics; subset sums; representations of numbers.

1. INTRODUCTION

At the 1997 West Coast Number Theory Conference at Asilomar, Gerry Myerson asked the following question in the problem session [2, Problem 97:16, includes the history below]: Is there a set B of n numbers, such that each power of 2 from 1 to 2^n is the sum of the elements of some subset of B?

Myerson pointed out that an easy induction argument shows that at least one of the numbers must be negative and that straightforward checking shows that n must be at least 4. Just after Myerson posed this problem, Peter Montgomery answered in the affirmative by giving the set $\{-5, 1, 7, 9\}$, representing 1, 2, 4, 8, 16, and he later found the additional example $\{-3, 4, 5, 11\}$. After seeing Montgomery's solution, Myerson asked how many fewer elements a set could have than a set of powers of 2 that it so represented. The author then pointed out that one could use Montgomery's solution to make this difference arbitrarily large by using the set

$$\{-5 \cdot 32^i, 32^i, 7 \cdot 32^i, 9 \cdot 32^i \mid 0 \leqslant i \leqslant m-1\}.$$

This set has size $4m$, and each of the $5m$ powers of 2 from 1 to 2^{5m-1} is the sum of the elements of one of its subsets. Upon hearing this, Myerson

0022-314X/01 $35.00

asked what one could say in general about how many numbers it takes to represent the first n powers of 2.

We first generalise to powers of other integers and prove some simple facts about such representations, including Myerson's observations above, which he stated without proof. We then study subset-sum representations of sets of powers of 2 and provide lower and upper bounds on the size of a smallest representing set for powers of arbitrary natural numbers. We also investigate representations of general sets, finding, for instance, bounds on their number and on the size of their elements. Finally, we provide a number of directions for further research.

2. DEFINITIONS AND PRELIMINARY RESULTS

Throughout this paper, we will use the word "set" to refer to a set of real numbers, and such a set will be finite unless otherwise indicated. We let the *span* of a set B, denoted by $\mathrm{sp}(B)$, be the set of all sums of subsets of B. That is, we define

$$\mathrm{sp}(B) = \left\{ \sum_{b \in A} b \;\middle|\; A \subseteq B \right\}.$$

Thus the span of B is the set of linear combinations of elements of B with each coefficient equal either to 0 or to 1.

We say that a set B *represents* a set P when P is a subset of the span of B, that is, when each element of P is the sum of the elements of some subset of B. For any set P we let the *rank* of P, denoted by $\mathrm{rk}(P)$, be the smallest size of a set representing P. Often when referring to the span or rank of an explicit set, we leave out the curly brackets surrounding its elements for ease of reading. We call any set representing P and having size equal to the rank of P *optimal*. Notice that every set represents itself, so we have the trivial bound $\mathrm{rk}(P) \leqslant |P|$. We say that P is *independent* when we have equality in this bound and *dependent* otherwise. We are interested in whether a given set is independent or dependent and in how much smaller its rank can be than its size.

We remark that allowing B to be a multiset in the above (so that some of its elements may be repeated, that is, so that certain numbers are allowed to be used more than once in representing elements of P) would not change the rank of P. More precisely, we claim that the smallest size of a multiset representing P is equal to the rank of P. For suppose a multiset B represents P and uses the number b twice. Then replacing one of the occurrences of b by $2b$ gives a new multiset that represents everything B

does, including, in particular, P. An easy induction now shows that after a finite number of such replacements, we obtain a set (that is, a multiset with no repeated elements) of the same size that still represents P. Thus if P is representable by a multiset of a given size, then it must be representable by a set of the same size. As the converse is clearly true, this proves our assertion.

Recall that, given two sets P and Q, we can form their (elementwise, not to be confused with direct!) product $PQ = \{pq \mid p \in P, q \in Q\}$. When one of the sets is a singleton, we suppress the curly brackets surrounding its element. This definition clearly extends to arbitrary (finite) collections of sets. In the proof of the following lemma, we will consider such a product as a multiset, counting the multiplicity with which each number occurs as a product, but elsewhere we will view products of sets simply as sets. We begin by obtaining bounds on the ranks of unions and products of sets.

LEMMA 1. *The rank of the union of a collection of sets is at most the sum of the ranks of the sets. The rank of the product of a collection of sets is at most the product of the ranks of the sets.*

Proof. Let the given sets be $P_1, ..., P_k$, and for $1 \leqslant i \leqslant k$, let B_i be a set of size $\mathrm{rk}(P_i)$ representing P_i. Then it is easy to see that the set $\bigcup_{i=1}^{k} B_i$ and the multiset $\prod_{i=1}^{k} B_i$ represent $\bigcup_{i=1}^{k} P_i$ and $\prod_{i=1}^{k} P_i$, respectively. (In the latter case, multiply out the representations of the factors of an element of the product set $\prod_{i=1}^{k} P_i$ to get a sum of elements of $\prod_{i=1}^{k} B_i$.) As the set and the multiset have at most $\sum_{i=1}^{k} \mathrm{rk}(P_i)$ and exactly $\prod_{i=1}^{k} \mathrm{rk}(P_i)$ elements, respectively, the above remark on multisets proves the lemma. ∎

In particular, this lemma shows that adjoining k elements to a set, which cannot decrease the rank, will increase it by at most k.

Recall that we showed in the introduction that the difference between the rank and the size of a set of powers of 2 can be arbitrarily large. In our family of examples, however, the ratio of these two quantities was constant. It is natural, therefore, to ask what can be said about the ratio of the rank to the size of a set of powers of 2. More generally, we are interested in the ratio of $\mathrm{rk}(P)$ to $|P|$ when P is a finite geometric sequence.

By Lemma 1, for any real number r and any positive integers m and n, we have

$$\begin{aligned}
&\mathrm{rk}(1, r, ..., r^{m+n-1})\\
&\qquad \leqslant \mathrm{rk}(1, r, ..., r^{m-1}) + \mathrm{rk}(r^m, r^{m+1}, ..., r^{m+n-1})\\
&\qquad \leqslant \mathrm{rk}(1, r, ..., r^{m-1}) + \mathrm{rk}(1, r, ..., r^{n-1}).
\end{aligned}$$

Since these ranks are nonnegative, it follows from a standard real-analysis argument that the limit

$$\lim_{n\to\infty} \frac{\operatorname{rk}(1, r, ..., r^{n-1})}{n}$$

exists and is equal to the infimum of the defining ratios. We define $\rho(r)$ to be this limit. Our trivial bound on the rank of a set gives $\rho(r) \leqslant 1$ for any real number r, and we will later prove that this inequality is always strict when r is an integer. In fact, we will show that, for integral values of r, the limit is never achieved by any of its defining ratios. Notice that we have $\rho(0) = \rho(1) = 0$ and that Montgomery's example gives $\rho(2) \leqslant 4/5$.

Our next result, a generalisation of one of Myerson's observations, gives a necessary condition on representations of certain sets of positive numbers by smaller sets.

Proposition 1 (based on Myerson [2, Problem 97:16]). *If $P = \{p_1, ..., p_n\}$ is a set of positive numbers satisfying*

$$p_k > \sum_{i<k} p_i$$

for $1 \leqslant k \leqslant n$, then any set of size less than n that represents P must contain a negative number.

Proof. Let B be a set of nonnegative numbers representing P, and label the elements of B in increasing order as $b_1, b_2, ...$ We must show that B has at least n elements. To do this, we prove by induction on k that for $1 \leqslant k \leqslant n$, the set B has at least k elements, and b_k is at most p_k. Assume that this holds for all smaller values of k. Then we have

$$\begin{aligned} p_k &> \sum_{i<k} p_i \\ &\geqslant \sum_{i<k} b_i. \end{aligned}$$

Since the numbers b_i are all nonnegative, this inequality shows that p_k cannot equal the sum of any subset of the set $\{b_i \mid i<k\}$. Thus when p_k is written as a sum of the terms b_i for some set of values of i, this sum must use some term b_j with $j \geqslant k$. In particular, B must have at least k elements. Again, by nonnegativity, the term b_j must be at most p_k, and by the ordering, it is at least as large as b_k. Hence we have $b_k \leqslant p_k$, and our inductive hypothesis is satisfied. Thus, by induction, B has at least n elements, which proves the proposition. ∎

3. SPECIFIC EXAMPLES FOR POWERS OF 2

Here we examine representing sets for sequences of consecutive powers of 2, the sequences in which Myerson was originally interested. First we show that the rank of the set of powers of 2 through 2^{n-1} is n for values of n up through 4, and then we consider some larger values of n.

PROPOSITION 2 (Myerson [2, Problem 97:16]). *For* $n \in \{0, 1, 2, 3, 4\}$, *the set* $\{2^i \mid 0 \leqslant i \leqslant n-1\}$ *is independent.*

Proof. This is clear for $n \leqslant 2$, so we consider the cases $n=3$ and $n=4$. First suppose that $\{1, 2, 4\}$ could be represented by two numbers, say a and b. As the only nonzero sums of subsets of $\{a, b\}$ are a, b, $a+b$, these numbers would have to be the numbers 1, 2, 4 in some order. But then one of 1, 2, 4 would have to be the sum of the other two, which is false. Hence the case $n=3$ is proved.

So suppose that we have three numbers a, b, c that represent the set $P = \{1, 2, 4, 8\}$. Notice that, by the uniqueness of binary representations, no two different subsets of P have the same sum. Now each element of P is equal to one of the following: a, b, c, $a+b$, $a+c$, $b+c$, $a+b+c$; and, for instance, not all of a, $b+c$, $a+b+c$ can be used, or some element of P would be the sum of some of the others. It is straightforward to check that, in order to prevent this phenomenon, we have, after interchanging a, b, c if necessary, only four possible representations of P by a, b, c. Namely, we can have

$$P = \{a, b, a+c, b+c\} \qquad \text{or} \qquad P = \{a, a+b, a+c, b+c\} \qquad \text{or}$$

$$P = \{a, a+b, a+c, a+b+c\} \qquad \text{or} \qquad P = \{a+b, a+c, b+c, a+b+c\}.$$

But in the first and third cases, we would have two elements of P whose sum is the same as the sum of the other two elements of P, which cannot occur, and in the other two cases we would have the equalities

$$2(a) + (b+c) = (a+b) + (a+c)$$

and

$$(a+b) + (a+c) + (b+c) = 2(a+b+c),$$

respectively. Each of these two equalities would also imply an impossible equation among the elements of P. Thus all four cases are impossible, and P cannot be so represented. This contradicts the existence of appropriate a, b, c and proves the result. ∎

Now we consider some larger sets of powers of 2. As we saw from Montgomery's examples, $\{1, 2, 4, 8, 16\}$ is representable by a set of size 4. In fact, a search using the ideas from the proof of the previous proposition shows that there are exactly 19 optimal representing sets of $\{1, 2, 4, 8, 16\}$. Ten of these are (listed without their curly brackets for clarity)

$$\begin{array}{ccc}
-5, 1, 7, 9 & -5, 6, 7, 9 & -5, -8, 7, 9 \\
-5, -3, 7, 9 & -5, -1, 7, 9 & -6, 1, 7, 9 \\
-5, 1, 7, 8 & -5, 6, 7, 3 & 3, -8, 7, 9 \\
-1, 2, 5, 9; & &
\end{array}$$

and each of the others can be obtained from one of the first nine above by multiplying all elements by 4 and taking the absolutely least residues modulo 31. (It is not immediately obvious why such relationships among solutions should hold!) The orders chosen for the sets and for their elements are intended to show the similarities among them.

Since the set of powers of 2 through 8 has rank 4, and the set of powers of 2 through 16 does as well, we see that 4 is the smallest value of m such that a set of size m can represent the first $m+1$ powers of 2. We might also ask, for each natural number k, what the smallest value of m is such that a set of size m can represent the first $m+k$ powers of 2. Notice that this is just a reformulation of the problem of finding the rank of the first n powers of 2, but it focuses attention on the most interesting cases. For instance, given optimal representing sets corresponding to these values of m, we can obtain an optimal set for the first n powers of 2 for any natural number n, simply by adjoining some powers of 2 to one of the given optimal sets.

By checking all 19 of the sets representing $\{1, 2, 4, 8, 16\}$ described above, we can see that none of them also represent 32, so that no set of four numbers represents the powers of 2 through 32. Hence for the choice $k=2$ in the previous paragraph, the minimum possible value of m would be 5.

It turns out that this minimum can indeed be achieved; in fact we found by hand two (optimal) sets of size 5 representing $\{1, 2, 4, 8, 16, 32, 64\}$, namely $\{-20, -15, 17, 19, 28\}$ and $\{-27, -5, 7, 28, 36\}$, and we found ten others later with a computer search. We suspect that these are the only ones, but have not yet been able to prove this. As with the case of representing five powers of 2, these 12 sets come in pairs. Six of them are

$$\begin{array}{ccc}
-21, -5, 7, 27, 30 & -27, -5, 7, 28, 36 & -26, -3, 5, 30, 37 \\
-29, -13, 14, 19, 31 & -18, -11, 14, 19, 31 & -17, -12, 13, 20, 31.
\end{array}$$

Three additional optimal sets can be obtained by multiplying the elements of the first three of these sets by 4 and taking the absolutely least residues modulo 127, and three more result from multiplying the other three of the above sets by 8 and taking the absolutely least residues modulo 127.

Notice that the four previous paragraphs and the remark immediately following the proof of Lemma 1 show that $\mathrm{rk}(1, 2, 4, ..., 2^{n-1})$ is equal to 4 for $n=5$, to 5 for $n=6$, and to 5 for $n=7$. If we do, indeed, have all of the optimal representing sets for $n=7$, then, since none of them represent 128, it would follow that $\mathrm{rk}(1, 2, 4, ..., 128)$ is 6. In addition, we have run long computer searches looking for six numbers representing $\{1, 2, 4, ..., 256\}$, but have been unsuccessful so far, so we suspect that $\mathrm{rk}(1, 2, 4, ..., 256)$ is 7. This would imply further, by the following paragraph, that $\mathrm{rk}(1, 2, 4, ..., 512)$ also is 7 and that the value of m corresponding to $k=3$ above is 7.

We can use the second set from the above list of representing sets for $\{1, 2, 4, ..., 64\}$ to obtain, for many values of m, sets of size $2m+1$ representing the first $3m+1$ powers of 2. The idea is that if we have a set representing $\{1, 2, 4, ..., 2^{n-1}\}$ and also -2^{n-2}, then we can adjoin the two numbers 2^n+2^{n-2} and $2^{n+1}+2^{n-2}$ to get a new set that represents not only 2^n and 2^{n+1}, but also $2^{n+2}=2^{n-1}+(2^n+2^{n-2})+(2^{n+1}+2^{n-2})$. For instance, $\{-27, -5, 7, 28, 36, 160, 288\}$ represents $\{1, 2, 4, ..., 512\}$. After we do this, we can often modify the new set so that it also represents -2^{n+1}, and then we can repeat the process. Sometimes slight variations of these procedures are needed.

For instance, in the representing set from the previous paragraph, replace 7 by -281. Since in the original set, we never had to use 7 and 288 at the same time, we may now use both -281 and 288 where previously we used 7, and we can still represent all the same powers of 2. Furthermore, we can adjoin $1310=2^{10}+281+5$ and $2334=2^{11}+281+5$ to represent not only 2^{10} and 2^{11}, but also $2^{12}=-27-5+36+160+288+1310+2334$. Continuing along these lines, we have found a set of 15 integers representing $\{1, 2, 4, ..., 2^{21}\}$. This gives the bound $\rho(2)\leqslant 15/22$ on the limit ratio $\rho(2)$ defined in the previous section.

It seems that these constructions can be carried out indefinitely, which would imply the following upper bound on $\rho(2)$:

Conjecture 1. The limit $\rho(2)=\lim_{n\to\infty}\mathrm{rk}(1, 2, 4, ..., 2^{n-1})/n$ is at most 2/3.

We also record here the slightly stronger form indicated by the examples discussed above:

Conjecture 2. For any integer $m\geqslant 2$, we have $\mathrm{rk}(1, 2, 4, ..., 2^{3m})\leqslant 2m+1$.

In fact, it seems that there is a lot of flexibility in constructing sets of $2m+1$ integers representing the first $3m+1$ powers of 2, and so we suspect that perhaps $\rho(2)$ is actually strictly less than 2/3.

4. GENERAL RESULTS FOR POWERS OF INTEGERS

Here we obtain both nontrivial lower and nontrivial upper bounds on ranks of sets of consecutive powers of an arbitrary integer. First we show that the rank of $\{1, r, ..., r^{n-1}\}$ is at least $n/\log_r(rn-n)$, and then we show that sufficiently long sequences of consecutive powers of any integer are dependent. This latter result will allow us to conclude that for any integer r, the limiting ratio $\rho(r)$ of rank to length for sequences of powers of r can never be achieved by any finite sequence. Of course, these results also all apply to geometric sequences with common ratio r.

The best lower bound we can achieve for the rank of an arbitrary set of size n is $\lceil \log_2 n \rceil$, obtained by counting subsets of the representing set (the set of integers from 0 to $n-1$ shows that this bound is sharp). But we can get a significantly better bound for geometric sequences by using the ideas from the proof of Proposition 2.

Theorem 1. *For any integers $r \geqslant 2$ and $n \geqslant 2$, we have*

$$\operatorname{rk}(1, r, ..., r^{n-1}) \geqslant \frac{n}{\log_r(rn-n)}.$$

Proof. Let $B = \{b_1, ..., b_d\}$ be any set representing $\{1, r, ..., r^{n-1}\}$; it will be enough to show that d is at least $n/\log_r(rn-n)$. For $0 \leqslant j \leqslant n-1$, we can write

$$r^j = \sum_{i=1}^{d} b_i c_{ij},$$

with each coefficient c_{ij} being either 0 or 1. We let $\mathbf{b}$ be the vector $[b_1 \cdots b_d]$ and $\mathbf{c}_j$ be the vector $[c_{1j} \cdots c_{dj}]$, so that the inner product of $\mathbf{b}$ and $\mathbf{c}_j$ is r^j.

Suppose that two linear combinations $\sum_{j=0}^{n-1} \mathbf{c}_j u_j$ and $\sum_{j=0}^{n-1} \mathbf{c}_j v_j$, with all of the coefficients u_j and v_j being natural numbers less than r, are equal. Taking inner products with $\mathbf{b}$ shows that the expressions $\sum_{j=0}^{n-1} r^j u_j$ and $\sum_{j=0}^{n-1} r^j v_j$ must be equal. But by the uniqueness of representations of natural numbers in base r, this implies that each coefficient u_j must equal the corresponding coefficient v_j. Thus, as the variables u_j range independently from 0 to $r-1$, the sums $\sum_{j=0}^{n-1} \mathbf{c}_j u_j$ must all be distinct.

As there are r possible values for each of the n variables u_j, the total number of such linear combinations is r^n. If we omit the one that has every variable u_j equal to $r-1$, then we have $r^n - 1$ left, each of which has $u_j < r-1$ for some value of j. On the other hand, since every vector $\mathbf{c}_j$ is a zero-one vector, each of these $r^n - 1$ linear combinations has d coordinates, all of which must be natural numbers less than $(r-1)\,n$. Hence there are at most $(rn-n)^d$ possibilities for these linear combinations.

Notice that if all of these vectors did occur, then, in particular, each vector with a single entry equal to 1 and the rest equal to 0 would occur and would have to equal some vector $\mathbf{c}_j$. This would mean that each power r^j would be equal to some representing element b_i, giving $d \geqslant n$, which would give the desired result, as $rn-n$ is at least r. Thus we may assume that not all of the possible vectors actually occur, so that there are at most $(rn-n)^d - 1$ possible values for the $r^n - 1$ linear combinations. In order for them to be distinct, we must have

$$(rn-n)^d - 1 \geqslant r^n - 1$$

$$(rn-n)^d \geqslant r^n$$

$$d \log_r (rn-n) \geqslant n$$

$$d \geqslant \frac{n}{\log_r (rn-n)},$$

and this proves the theorem. ▮

In particular, this gives us the bounds $\mathrm{rk}(1, 2, 4, ..., 2^{n-1}) \geqslant n/\log_2 n$ and $\mathrm{rk}(1, r, ..., r^{n-1}) > n/(1+\log_r n)$. In the above proof, we achieved a slightly better result than we would have otherwise by omitting one of the linear combinations before counting and then arguing that not all vectors could occur. (Without these steps, we would have had a shorter proof, but would have only obtained the lower bound of $n/\log_r (rn-n+1)$.) We can, in fact, strengthen the bound a bit more by excluding more linear combinations. For instance, for $r=2$ and odd $n \geqslant 3$, we can omit half of them and get the bound $\mathrm{rk}(1, 2, 4, ..., 2^{n-1}) > (n-1)/(\log_2(n+1)-1)$. This, however, is a minor improvement and still seems to be some distance from the truth.

We now show that sufficiently long sequences of consecutive powers of an arbitrary integer are dependent. Note that even powers of a negative number are also powers of its square, which is positive, and that adding terms to a dependent set gives another dependent set. Hence in order to prove the above assertion, it is sufficient to consider powers of a positive integer. And by the examples from the previous section, we may restrict to powers of numbers greater than 2. For these we have a uniform construction.

THEOREM 2. *For any integer $r \geqslant 3$, we have*

$$\operatorname{rk}(1, r, ..., r^{2r-1}) \leqslant 2r-1.$$

Proof. First write

$$s = r^{2r-2} - r^{2r-5} - \sum_{i=0}^{r-2} r^{2i},$$

and then define

$$B = \{-s, 1, r^2, r^4, ..., r^{2r-6}, s+r, s+r^3, ..., s+r^{2r-3}, s+r^{2r-4}\}.$$

Note that B has $2r-1$ elements. Now all even powers of r through r^{2r-6} are in B, and all odd powers of r through r^{2r-3}, as well as r^{2r-4}, are the sum of $-s$ and another element of B. Hence, in order to show that B represents $\{1, r, ..., r^{2r-1}\}$, which will prove the theorem, we just need to verify that r^{2r-2} and r^{2r-1} are in the span of B. We have

$$\begin{aligned} r^{2r-2} &= s + r^{2r-5} + \sum_{i=0}^{r-2} r^{2i} \\ &= (s + r^{2r-5}) + (s + r^{2r-4}) - s + \sum_{i=0}^{r-3} r^{2i} \\ &\in \operatorname{sp}(B) \\ r^{2r-1} &= rr^{2r-2} \\ &= rs + r^{2r-4} + \sum_{i=0}^{r-2} r^{2i+1} \\ &= (s + r^{2r-4}) + \sum_{i=0}^{r-2} (s + r^{2i+1}) \\ &\in \operatorname{sp}(B), \end{aligned}$$

and this gives the result. ▌

For instance, taking $r = 3$ in the above proof gives $s = 68$, and we obtain the set $\{-68, 1, 71, 95, 77\}$ representing $\{1, 3, 9, 27, 81, 243\}$.

Notice that this already shows that for any positive integer $r \geqslant 3$, the limiting ratio $\rho(r)$ is at most $(2r-1)/2r$; in particular, it is less than 1. In fact, since 1 is always an element of the set B in the proof of Theorem 2, we can get a slightly better ratio of the size of a set of powers of r to its rank.

Corollary 1. *For any integers $r \geqslant 3$ and $m \geqslant 0$, we have*

$$\mathrm{rk}(1, r, ..., r^{m(2r-1)}) \leqslant m(2r-2)+1.$$

Proof. This inequality will follow from the fact that, with B as in the proof of Theorem 2, all powers of r through $r^{m(2r-1)}$ are in the span of the set $\{1\} \cup \{1, r^{2r-1}, ..., r^{(m-1)(2r-1)}\}(B \backslash \{1\})$. (Recall here our notation for set products.) To check this fact, we observe that all powers of r in $r^{k(2r-1)}\{1, r, ..., r^{2r-1}\}$ are represented by $r^{k(2r-1)}B$ as before, except that when k is positive, instead of using the element $r^{k(2r-1)}$, which has been removed, we use the numbers that summed to it in representing the set $r^{(k-1)(2r-1)}\{1, r, ..., r^{2r-1}\}$. ▮

Taking the limit as m goes to infinity now gives us the better bound on the ratio function $\rho(r)$ to which we alluded above.

Corollary 2. *For any integer $r \geqslant 3$, we have $\rho(r) \leqslant (2r-2)/(2r-1)$.*

Our first conjecture at the end of the previous section asserts that this result is also valid for $r=2$.

We also use the theorem to show that the optimal ratio indicated by $\rho(r)$ is never achieved.

Corollary 3. *For any integers r and $n \geqslant 1$, we have*

$$\mathrm{rk}(1, r, ..., r^{n-1})/n > \rho(r).$$

Proof. For simplicity, we restrict our attention to the case when r is at least 3; slight variations of the proof will deal with the other values of r. By Theorem 2, applied to r^n, we have $\mathrm{rk}(1, r^n, r^{2n}, ..., r^{n(2r^n-1)}) \leqslant 2r^n - 1$, so that Lemma 1 gives us

$$\begin{aligned}\mathrm{rk}(1, r, ..., r^{2nr^n-1}) &= \mathrm{rk}(\{1, r^n, r^{2n}, ..., r^{n(2r^n-1)}\}\{1, r, ..., r^{n-1}\}) \\ &\leqslant \mathrm{rk}(1, r^n, r^{2n}, ..., r^{n(2r^n-1)})\ \mathrm{rk}(1, r, ..., r^{n-1}) \\ &\leqslant (2r^n-1)\ \mathrm{rk}(1, r, ..., r^{n-1}) \\ &< 2r^n\ \mathrm{rk}(1, r, ..., r^{n-1}).\end{aligned}$$

Using the fact that $\rho(r)$ is the infimum of its defining ratios now allows us to conclude

$$\begin{aligned}\rho(r) &\leqslant \mathrm{rk}(1, r, ..., r^{2nr^n-1})/2nr^n \\ &< \mathrm{rk}(1, r, ..., r^{n-1})/n,\end{aligned}$$

proving the corollary. ▮

For instance, this corollary and the discussion on ranks of sets of powers of 2 give the strict inequality $\rho(2) < 15/22$.

Finally, Corollary 3 and the definition of $\rho(r)$ immediately produce the additional result:

COROLLARY 4. *Given integers r and $n \geqslant 1$, for any sufficiently large integer m, we have*

$$\operatorname{rk}(1, r, ..., r^{m-1})/m < \operatorname{rk}(1, r, ..., r^{n-1})/n.$$

5. REPRESENTATIONS OF GENERAL SETS

Here we prove several results concerning the representations of arbitrary sets. First we show that any finite set has only finitely many optimal representing sets, and we obtain a bound on the size of the elements of these optimal sets. Then we obtain some large independent sets.

THEOREM 3. *Any set of size n and rank d has at most*

$$\binom{n}{d}\binom{2^d-1}{d}$$

optimal representing multisets.

Proof. Let $P = \{p_1, ..., p_n\}$ be a finite set of rank d and $B = \{b_1, ..., b_d\}$ be any multiset of size d representing P. Then we have a $d \times n$ zero-one matrix $C = [c_{ij}]$ with entries satisfying $\sum_{i=1}^{d} b_i c_{ij} = p_j$ for $1 \leqslant j \leqslant n$. Fix a particular such matrix C and consider all possible corresponding multisets B. The system of linear equations $\sum_{i=1}^{d} b_i c_{ij} = p_j$, which gives possible values for each variable b_i, is consistent and so has either one solution or infinitely many solutions.

If this system had infinitely many solutions, then we could find a solution with some variable b_i equal to 0. But this would give a set of size d containing 0 and representing P, and then leaving out 0 would produce a set of size $d-1$ also representing P, contradicting the definition of rank. Hence the system has a unique solution, and there is some subsystem of d of the n equations that determines the d variables $b_1, ..., b_d$ uniquely. Let D be the $d \times d$ matrix of coefficients of this subsystem (so D is a nonsingular submatrix of C).

Now the elements of B are determined by the d elements of P that appear in this subsystem and by the collection of vectors occurring as rows of the matrix D (permuting these rows will only permute the elements of B, not change them). There are $\binom{n}{d}$ choices for the subset of P used, and there are at most $\binom{2^d-1}{d}$ choices for the rows of the matrix D, since it is a $d\times d$ zero-one matrix with distinct nonzero rows. Hence the total number of possible representing multisets B is at most $\binom{n}{d}\binom{2^d-1}{d}$, which is what we wanted to prove. ∎

This bound is far from sharp. For instance, if n equals 2^d-1 and P does not contain 0, then all nonzero zero-one vectors occur as columns of C, so we may insist that D be a permutation matrix. This removes the second factor from our calculations and gives the upper bound of $\binom{n}{d}=\binom{2^d-1}{d}$ optimal representing multisets.

Corollary 5. *Any finite set has only finitely many optimal representing sets.*

We can also use the proof of this theorem to get an upper bound on the size of elements of an optimal representing set. Throughout the rest of this section, we will use the notation Δ_k to denote the maximum determinant of $k\times k$ zero-one matrices.

Corollary 6. *For any set P of rank d, the absolute value of any element of an optimal representing set for P is at most Δ_{d-1} times the sum of the d largest absolute values of the elements of P. If P consists of integers, then any element of an optimal representing set of P is a rational number with numerator at most the above bound in absolute value and with positive denominator at most Δ_d.*

Proof. Let $B=\{b_1, ..., b_d\}$ be such a representing set, and take the numbers c_{ij} as in the proof of Theorem 3. That proof shows that we can solve for the elements b_i from a subset of d of the equations $\sum_{i=1}^{d} b_i c_{ij}=p_j$. Using Cramer's rule now gives the bounds. ∎

Notice that, since the entries of any zero-one matrix have absolute values at most 1, the Hadamard bound implies $\Delta_k \leqslant k^{k/2}$. In fact, it can be shown, using this Hadamard bound, that any $k\times k$ matrix (for $k\geqslant 2$) that has all entries nonnegative and at most 1 has determinant at most $2^{1-k}k^{k/2}(k-1)$ in absolute value. Hence this expression can be used for Δ_k in applying Corollary 6 and elsewhere.

Finally, the proof of Theorem 3 provides a finite algorithm to determine the rank of any finite set. Namely, to decide whether a set P of size n has rank at most d, run through all possible $d\times n$ zero-one matrices C and, for each one, determine whether the corresponding system of linear equations from the proof of the theorem is consistent. The set P will have rank at most d if and only if one of these linear systems is consistent, and solving such a consistent system will provide a representing set for P of size d. Repeating this procedure for each value of d from 0 to n will yield the rank of P.

Recall that a finite set of numbers with rank equal to its size is called independent, and one with rank less than its size is called dependent. We extend these definitions to infinite sets by calling an infinite set independent if all of its finite subsets are independent and by calling it dependent otherwise. At the 1997 conference, Seva Lev asked for "reasonable" conditions for a (finite) set of integers to be independent [2, Problem 97:19].

It is natural to ask whether there are arbitrarily large independent sets of integers. In Theorem 2, we showed that any sufficiently long geometric sequence of integers is dependent (with the length needed depending on the common ratio between successive terms), and there might conceivably be some number n such that any geometric sequence of integers of length n is dependent. In fact, we will show that there are arbitrarily long independent geometric sequences, which, in particular, will answer the question above, and we will give an infinite independent set of integers. The geometric sequences are due to Seva Lev [1], while the existence of an infinite independent set is due to the author via a different, but similar, method. The examples and proofs given here are based on a hybrid of Lev's and the author's methods.

First we need another characterisation of the rank of a set, which we get by generalising an observation of Lev [1] (who gave the case $d=n$).

LEMMA 2. *The rank of a set $P=\{p_1, ..., p_n\}$ is the smallest natural number d such that the vector $[p_1 \cdots p_n]$ is a linear combination of d zero-one vectors.*

Proof. As in the proof of Theorem 3, P is represented by a set $B=\{b_1, ..., b_d\}$ of size d if and only if there is a $d \times n$ zero-one matrix $C=[c_{ij}]$ with entries satisfying $\sum_{i=1}^{d} b_i c_{ij} = p_j$. But this is the same as saying that $[p_1 \cdots p_n]$ is a linear combination of the d rows of C with the numbers b_i as coefficients. Thus the smallest value of d such that P is represented by a set of size d is the same as the smallest value of d such that $[p_1 \cdots p_n]$ is a linear combination of d zero-one vectors, proving the lemma. ∎

In order to produce large independent sets, we use the following lemma, which allows us to add elements to sets without destroying independence. Recall the definition of Δ_k preceding Corollary 6 (and the discussion of it following the proof of that corollary).

LEMMA 3. *If $P=\{p_1, ..., p_n\}$ is a set such that $P\backslash\{p_n\}$ is independent and p_n satisfies*

$$|p_n| > \varDelta_{n-1} \sum_{i=1}^{n-1} |p_i|,$$

then P is independent.

Proof. Let $P=\{p_1, ..., p_n\}$ be a dependent set with p_n satisfying the inequality of the hypothesis; we will show that $P\backslash\{p_n\}$ is dependent.

By Lemma 2, the vector $[p_1 \cdots p_n]$ is a linear combination of $n-1$ zero-one vectors. Thus we have an $(n-1)\times n$ zero-one matrix $C=[c_{ij}]$ with $[p_1 \cdots p_n]$ a linear combination of the rows of C. This also means that $[p_1 \cdots p_{n-1}]$ is a linear combination of the rows of the matrix D obtained by deleting the last column of C. The first linear dependence further gives us the equation

$$\begin{vmatrix} c_{1,1} & \cdots & c_{1,n} \\ \vdots & \ddots & \vdots \\ c_{n-1,1} & \cdots & c_{n-1,n} \\ p_1 & \cdots & p_n \end{vmatrix} = 0.$$

Expanding this determinant by minors in the last row gives

$$\sum_{j=1}^{n} (-1)^{n+j} C_j p_j = 0,$$

where C_j is the determinant of the submatrix of C obtained by deleting the jth column. Since this submatrix is an $(n-1)\times(n-1)$ zero-one matrix, its determinant C_j has absolute value at most $\varDelta_{n-1}$, and we get

$$\begin{aligned} |C_n p_n| &= \left|\sum_{j=1}^{n-1} (-1)^j C_j p_j\right| \\ &\leqslant \sum_{j=1}^{n-1} |C_j|\,|p_j| \\ &\leqslant \sum_{j=1}^{n-1} \varDelta_{n-1}\,|p_j| \\ &= \varDelta_{n-1} \sum_{j=1}^{n-1} |p_j| \\ &< |p_n|. \end{aligned}$$

This inequality forces C_n to have absolute value less than 1, and as C_n is an integer, it must, therefore, equal 0. Thus the $n-1$ rows of the matrix D, which has determinant $C_n = 0$, are linearly dependent, and $[p_1 \cdots p_{n-1}]$ is a linear combination of $n-2$ of the rows of D. Now Lemma 2 shows that $P \backslash p_n$ is dependent. This proves the lemma. ▮

We now get the large independent sets to which we alluded above.

THEOREM 4 (Lev [1]). *For any natural number $n \geqslant 1$ and any natural number $r > \Delta_{n-1}$, the set of the first n powers of r is independent.*

Proof. For $1 \leqslant k \leqslant n$ we have

$$\begin{aligned} r^{k-1} &> (r-1)\frac{r^{k-1}-1}{r-1} \\ &\geqslant \Delta_{n-1} \sum_{i=0}^{k-2} r^i \\ &\geqslant \Delta_{k-1} \sum_{i=0}^{k-2} r^i. \end{aligned}$$

Now using Lemma 3 and induction on k shows that the geometric sequence $1, r, ..., r^{n-1}$ is independent. ▮

We should point out that a straightforward calculation using Theorem 1 gives this independence result for $r \geqslant n^{n-1}$. However, we need Lemma 3 anyway for the next theorem, and the result we have just proved is stronger than the one we would have obtained using Theorem 1, given our earlier remarks on the size of Δ_k.

THEOREM 5. *There is an infinite independent set of integers.*

Proof. Use $p_1, p_2, ..., p_n, ...,$ with each term p_n an integer larger than $\Delta_{n-1} \sum_{i=1}^{n-1} p_i$, and apply Lemma 3. ▮

Finally, we note another consequence of Lemma 2, which says, in some sense, that "most" finite sets of integers are independent.

THEOREM 6 (Lev [1]). *For every positive integer n, there is a positive rational number c_n such that the number of dependent sets of size n of integers from the interval $[-N, N]$ is $c_n N^{n-1}(1 + O(1/N))$ as $N \to \infty$.*

Proof. By Lemma 2, a vector P of n integers is dependent if and only if it is in a sublattice of $\mathbf{Z}^n$ spanned by some collection of $n-1$ zero-one vectors. Such a lattice, say having dimension d, grows like a rational constant times $N^d(1 + O(1/N))$. (By this, we mean that the number of points in it with all coordinates in the interval $[-N, N]$ is some rational constant times $N^d(1+O(1/N))$.) In fact, the intersection of any number of the lattices, being a lattice itself, also grows like a rational constant times $N^d(1+O(1/N))$ for appropriate d.

Since there are only finitely many n-dimensional zero-one vectors, there are only finitely many sublattices of $\mathbf{Z}^n$ spanned by $n-1$ such vectors. It follows by the inclusion-exclusion principle that there is some rational number c_n such that the union of all such sublattices grows like $c_n N^{n-1} + O(N^{n-2})$. And c_n is positive, because this union contains the lattice of all points with first coordinate 0. As the union of the sublattices corresponds to the collection of dependent sets, the result follows. ▮

6. FURTHER QUESTIONS

Finally, we offer some questions for further research.

1. Is it true that $\rho(2)$, the limit as n goes to infinity of the ratio of the rank of the first n powers of 2 to n, is at most 2/3? Can some better bound be proven? Can the upper bounds we give for $\rho(r)$ for other integers r be improved?

2. Is $\rho(2)$ greater than 0? If so, can some explicit lower bound be proved? If $\rho(2)$ equals 0, can some explicit sub-linear upper bound be put on the growth of the rank of the first n powers of 2? And can some better lower bound than that from Theorem 1 be obtained? Again, the same questions can be asked for other integers besides 2.

3. For any integer r, let $\mu(r)$ be the smallest value of n such that the set $\{1, r, ..., r^{n-1}\}$ is dependent. (This minimum exists by Theorem 2 and the discussion preceding it.) What can we say about $\mu(r)$? Proposition 2 and Montgomery's example give $\mu(2)=5$, while for $r \geqslant 3$, Theorems 4 and 2 give $\Delta_{\mu(r)-1} \geqslant r$ and $\mu(r) \leqslant 2r$, respectively.

4. Are there any optimal representing sets of $\{1, 2, 4, ..., 64\}$ besides the ones we give?

5. Are the patterns we observe among the optimal representations of $\{1, 2, 4, 8, 16\}$ and among those of $\{1, 2, 4, ..., 64\}$ special cases of some more general phenomenon?

6. Continuing in the vein of the last three questions, what are the ranks and optimal representing sets of particular small sets of powers of 2 (and of sets of powers of other small integers)?

7. Can independent sequences of integers be found that grow more slowly than those given in the previous section? For instance, is the set of

factorials independent? What is the smallest maximum of a set of n independent natural numbers? What is the smallest natural number p with $\{1, 2, 4, 8, p\}$ independent? Montgomery's first example shows that this p must be at least 18, and Myerson has checked that it must be at least 26. More generally, if we define p_n recursively to be the smallest natural number with $\{p_1, p_2, ..., p_n\}$ independent (p_n exists by Lemma 3), then what can we say about the sequence $p_1, p_2, ...$? Since every natural number k less than p_n is in the span of some optimal representing set for $\{p_1, p_2, ..., p_{n-1}\}$, adjoining p_n to this representing set gives an optimal representing set for $\{p_1, p_2, ..., p_n\}$ that also represents p_n+k. Thus p_{n+1} must be at least $2p_n$.

8. Is it true that for any rational number r, all sufficiently long sequences of consecutive powers of r are dependent?

9. All of the examples of optimal representing sets given above are sets of integers. But the set $\{-1/2, 3/2, 5/2\}$ represents $\{1, 2, 4\}$, and no set of size 2 does, so $\{1, 2, 4\}$ has a nonintegral optimal representing set. It does, however, have many integral optimal representing sets. Is it true that for any positive integers r and n, the set $\{1, r, ..., r^{n-1}\}$ must have an optimal representing set consisting of integers? More generally, is there a set of integers that does not have any integral optimal representing sets? Notice that if $\{1, r, ..., r^{n-1}\}$ and $\{1, r, ..., r^n\}$ have the same rank (which, for a given value of r, happens infinitely often by the inequality $\rho(r) < 1$), then dividing the elements of an optimal representing set for $\{1, r, ..., r^n\}$ by r gives an optimal representing set for $\{1, r, ..., r^{n-1}\}$ that also represents $1/r$ and so cannot consist entirely of integers. Can something else be said about when $\{1, r, ..., r^{n-1}\}$ has a nonintegral optimal representing set? Corollary 6 gives a bound on the denominators of elements in an optimal representing set for $\{1, r, ..., r^{n-1}\}$, but is it true that every such denominator must divide r?

10. Corollary 6 gives a bound on the absolute values of elements in an optimal representing set for a set P, but this bound is rather large. Can a more reasonable bound be found? Notice that $\{-3, -1, 5\}$ optimally represents $\{1, 2, 4\}$, so neither the maximum absolute value of elements of P nor is the maximum difference of two elements is such a bound, even for a geometric sequence. Is the sum of the absolute values of the elements of P such a bound? (By Corollary 6 we can use, for instance, this sum times $\Delta_{\mathrm{rk}(P)-1}$.) Does every set P have at least one optimal representation using no numbers larger in absolute value than the largest absolute value of the elements of P? Does this hold if we restrict to geometric sequences?

11. In a similar vein, can the bound from Theorem 3 on the number of optimal generating sets of a set be improved? Can it be improved for geometric sequences?

12. What can be said about the proportion of negative numbers in optimal representing sets for positive geometric sequences? The example two paragraphs ago shows that there may be more negative than positive numbers. What about the proportion of odd numbers in an optimal representing set for powers of 2?

13. What can be said about the ranks of other sets? As the geometric sequences that we study above satisfy first-order linear recurrences, it is natural to turn to sequences satisfying higher-order recurrences. For example, what bounds can be put on the rank of the set of the first n distinct nonzero Fibonacci numbers ($F_1 = F_2 = 1$ through F_{n+1})? The m numbers F_1, F_3, ..., F_{2m-1} represent the $2m-1$ nonzero Fibonacci numbers through F_{2m}. Hence the analogue of Myerson's original problem on powers of 2 is whether there are m numbers representing the $2m$ nonzero Fibonacci numbers through F_{2m+1}. Again, an easy check shows that m must be at least 4, and a variation of Proposition 1 shows that at least one negative representing number must be used.

14. Is there a polynomial-time algorithm for determining the ranks of arbitrary sets of integers? Is finding the ranks of sets of integers NP-hard?

15. Finally, we have only considered forming linear combinations with all coefficients taken from $\{0, 1\}$. What can be said when other possible coefficients are allowed?

ACKNOWLEDGMENTS

The research herein was conducted while the author was a Van Vleck Visiting Assistant Professor at the University of Wisconsin in Madison, Wisconsin. We thank Gerry Myerson for the original question and Peter Montgomery for his solution to it, which inspired us to work on the problem. Also, thanks to Seva Lev for his work, which he kindly allowed us to incorporate here. Finally, thanks to the several people who offered their advice and comments on this paper.

REFERENCES

1. Seva Lev, personal communication, 1998.
2. Gerry Myerson (Ed.), "West Coast Number Theory Conference Problems, December 18 & 21, 1997," West Coast Number Theory Conference, Asilomar, CA, 1997.