SUMS OF LEXICOGRAPHICALLY ORDERED SETS

M.D. ATKINSON,[1] A. NEGRO[2] and N. SANTORO[1]
[1]*School of Computer Science, Carleton University, Ottawa, Canada*
[2]*Dipartimento di Informatica, Universita' di Salerno, Salerno, Italy*

Received 21 May 1987
Revised 11 March 1988

We consider the problem of determining finite integer sets which are knapsack-solvable in linear time (that is, it is possible to determine in linear time, for any integer b, whether b can be expressed as a sum of distinct elements of that set) and where the largest element is as small as possible. We study the condition that the k-subsets (for fixed k) when lexicographically ordered have increasing sums. We give an optimal construction of sets with this property, prove that it is unique, and give the asymptotic behaviour of the largest member. Using these results, we construct sequences of positive integers where the largest element is minimal, the subset sums are distinct and lexicographically ordered, and are knapsack-solvable in linear time.

1. Introduction

Consider a set $A = \{a_1, \ldots, a_n\}$ of n distinct positive integers, where $a_i < a_{i+1}$ $(1 \leq i < n)$. The *Knapsack problem* on A is the problem of determining for any integer b whether b can be expressed as a sum of distinct elements of A, and, if so, of identifying such elements. Given a polynomial $p(n)$, the set A is said to be *(Knapsack-)solvable* in $O(p(n))$ time if there exists an algorithm which solves the Knapsack problem on A in $O(p(n))$ time.

Because of its applicability to *public key codes* [4], the identification and characterization of such sets (and the associated algorithms) have been the object of several investigations. In particular, due to the nature of this application, two goals can be identified:

(i) find solvable sets where the degree of $p(n)$ is very small (possibly, linear);

(ii) determine (for a fixed degree of the polynomial p) solvable sets in which the largest integer a_n is as small as possible.

It is not difficult to see that the set A of powers of 2 ($a_i = 2^{i-1}$) is solvable in *linear* time; this is due to the fact that this set has the following property

$$\textit{superincreasing:}\quad a_i > \sum_{j<i} a_j \quad (1 \leq i \leq n).$$

Observe that any set with the superincreasing property is solvable in linear time; thus, any such set would meet goal (i) described above. As for goal (ii), the powers of 2 example has $a_n = 2^{n-1}$, and any superincreasing set has $a_n \geq 2^{n-1}$. The

0012-365X/90/$3.50 © 1990, Elsevier Science Publishers B.V. (North-Holland)

question open is whether there exists a family of sets which are solvable in linear time *and* satisfy $a_n < 2^{n-1}$. To this end, observe that the superincreasing property implies the following property:

distinct subset sum: the 2^n different subsets of A have all different sums.

The problem of finding sets of positive integers with the distinct subset sum property has been independently studied by researchers in number theory; in particular, Erdös [2] specifically asked for sets with the distinct subset sum property in which the largest integer a_n was as small as possible. A lower bound on a_n has been established by Erdös and Moser who observed that, for any set of positive integers with the distinct subset sum property

$$na_n = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n-1} (a_n - a_i) \geqslant 2^n - 1 + n(n-1)/2$$

since $\sum_{i=1}^{n} a_i$ is the largest of 2^n non-negative distinct sums and the bracketed terms are distinct positive integers. Thus

$$a_n \geqslant 2^n/n + (n-1)/2;$$

but it seems likely that this lower bound is far from the best possible. Conway and Guy [1] found some sets with the distinct subset sum property which have $a_n < 2^{n-2}$ and recently Lunnon has slightly improved their construction [3]. However, the distinct subset sum property alone does not guarantee solvability in linear time; in fact, with the exception of superincreasing sets, all other sets with distinct subset sums which have been considered in those investigations are not known to be even polynomially solvable.

We shall study another class of sets with the distinct subset sum property, first described in Petit [5]. It is based on the following natural ordering (the *complete lexicographic ordering*) on the set of subsets of $\{a_1, \ldots, a_n\}$. In this ordering all the k-subsets (subsets of cardinality k) come before the $(k+1)$-subsets ($k = 0, 1, \ldots, n-1$) and, for each k, the k-subsets are listed in *lexicographic order.* For example, the ordering for $n = 4$ is shown in Fig. 1.

The sets that we investigate have the characteristic property that their sums are increasing with respect to this ordering. The sets are not such good solutions to the Erdös problem as those of Conway, Guy and Lunnon because they satisfy a more restrictive property; however, they are better than the obvious solution with the powers of 2 in that $a_n < 2^{n-1}$. We quantify this by giving the asymptotic behaviour as $n \to \infty$ of a_n in *optimal* sets for our various problems, and provide a simple technique for constructing those optimal sets. Furthermore, like the superincreasing sets (and unlike the sets of Conway, Guy and Lunnon), they are *solvable in linear time.* Fuller details of the solution algorithm may be found in [5] and here we sketch only the most important ideas for finding the subset C of $A = \{a_1, \ldots, a_n\}$ whose sum s is given. Suppose that the partial sums $s_i = a_1 + a_2 + \cdots + a_i$ have been computed. Then $k = |C|$ is determined as $k =$

$$
\begin{gathered}
\{\ \}\\
\{a_1\}\\
\{a_2\}\\
\{a_3\}\\
\{a_4\}\\
\{a_1, a_2\}\\
\{a_1, a_3\}\\
\{a_1, a_4\}\\
\{a_2, a_3\}\\
\{a_2, a_4\}\\
\{a_3, a_4\}\\
\{a_1, a_2, a_3\}\\
\{a_1, a_2, a_4\}\\
\{a_1, a_3, a_4\}\\
\{a_2, a_3, a_4\}\\
\{a_1, a_2, a_3, a_4\}
\end{gathered}
$$

Fig. 1. Lexicographic ordering of the subsets of $\{a_1, a_2, a_3, a_4\}$.

$\min\{t : s < s_t\} - 1$. Now we successively determine whether $a_1 \in C$, $a_2 \in C, \ldots$ At the time we are determining whether $a_i \in C$ we shall know of $B = C \cap \{a_i, a_{i+1}, \ldots, a_n\}$ both $r = |B|$ and $s = \sum_{b \in B} b$. Then $a_i \in C$ if and only if $s < s_{i+r} - s_i = a_{i+1} + \cdots + a_{i+r}$. Having made this test, i is increased and both r and s are updated.

We obtain our results by studying a less restrictive condition, namely that the k-subsets of A (for fixed k) when lexicographically ordered have increasing sums. We give an optimal construction of sets with this property, prove that it is unique, and give the asymptotic behaviour of the largest member. Using these results, we construct sequences of positive integers with a_n minimal and, for each k, the k-subsets having lexicographically ordered sums. A simple trick then allows us to impose on these sequences the additional condition that, for each k, the $(k+1)$-sums exceed the k-sums; we thereby obtain the condition that the subset sums are increasing in the complete lexicographic order given above.

2. The lexicographic ordering on k-subsets

Let $A = \{a_1, \ldots, a_n\}$ be a set of n distinct positive integers, where $0 < a_1$ and $a_i < a_{i+1}$ $(1 \leq i < n)$, and let k be any integer with $1 \leq k \leq n$. To each k-subset S of A we can associate a vector $(a_{i(1)}, \ldots, a_{i(k)})$ whose components are the members of S listed in increasing order of subscript. If $(a_{i(1)}, \ldots, a_{i(k)})$ and $(a_{j(1)}, \ldots, a_{j(k)})$ are any two such vectors we say the first precedes the second if, for some r,

$$i(1)=j(1),\ i(2)=j(2),\ \dots,\ i(r-1)=j(r-1), \quad \text{and} \quad i(r)<j(r).$$

In this way we obtain an induced ordering on the k-subsets of A, the well-known *lexicographic ordering*.

A is said to have the lex(k) property (the lexicographically ordered subset sum property on k-subsets) if whenever B and C are two k-subsets of A with B preceding C then

$$\sum_{x\in B} x < \sum_{x\in C} x.$$

Lemma 1. *The* lex(k) *condition holds if and only if for all t and i, with $1\leqslant t\leqslant k$ and $k-t+1\leqslant i\leqslant n-t$, $a_i+\sum_{j=2}^{t} a_{n-t+j}<\sum_{j=1}^{t} a_{i+j}$.*

Proof. We show that the inequalities given in the statement of the lemma are precisely the conditions which ensure that the sum of a k-subset is less than the sum of its successor. Given any k-subset B which is not last in the order we let a_{n-t+1} be the first of the elements $a_n, a_{n-1}, a_{n-2}, \dots$ which is not in B. Of course $1\leqslant t\leqslant k$. We also define a_i to be the element of $B-\{a_{n-t+2}, a_{n-t+3}, \dots, a_n\}$ with largest subscript. Then $i\leqslant n-t$ since $a_{n-t+1}\notin B$. Moreover i is the largest of $k-t+1$ such subscripts and so $i\geqslant k-t+1$. Thus $B=\{a_j, a_k, \dots, a_i, a_{n-t+2}, a_{n-t+3}, \dots, a_n\}$ (increasing subscripts) and its successor $\{a_j, a_k, \dots, a_{i+1}, a_{i+2}, a_{i+3}, \dots, a_{i+t}\}$. Hence the lex($k$) condition is precisely that $a_j+a_k+\cdots+a_i+a_{n-t+2}+a_{n-t+3}+\cdots+a_n<a_j+a_k+\cdots+a_{i+1}+a_{i+2}+\cdots+a_{i+t}$ as required. □

Lemma 2. lex(k)$\Leftrightarrow$lex($n-k$).

Proof. It is sufficient to prove that lex($n-k$)$\Rightarrow$lex(k). So assume lex($n-k$) and let t, i satisfy $1\leqslant t\leqslant k$ and $k-t+1\leqslant i\leqslant n-t$. Put $t'=n-t+1-i$, $k'=n-k$, and $i'=i$ and use lex(k') to deduce (since $1\leqslant t'\leqslant k'$ and $k'-t'+1\leqslant i'\leqslant n-t'$) that $a_{i'}+\sum_{j=2}^{t'} a_{n-t'+j}<\sum_{j=1}^{t'} a_{i'+j}$, that is

$$a_i+a_{t+i+1}+a_{t+i+2}+\cdots+a_n<a_{i+1}+a_{i+2}+\cdots+a_{n-t+1}$$

and hence

$$\begin{aligned}&a_i+(a_{n-t+2}+a_{n-t+3}+\cdots+a_n)+(a_{i+1}+a_{i+2}+\cdots+a_{i+t})\\&\quad+(a_{i+t+1}+\cdots+a_n)\\&\quad<(a_{i+1}+a_{i+2}+\cdots+a_{n-t+1})+(a_{n-t+2}+a_{n-t+3}+\cdots+a_n)\\&\quad+(a_{i+1}+a_{i+2}+\cdots+a_{i+t})\end{aligned}$$

and hence

$$a_i + (a_{n-t+2} + a_{n-t+3} + \cdots + a_n) + (a_{i+1} + a_{i+2} + \cdots + a_n)$$
$$< (a_{i+1} + a_{i+2} + \cdots + a_{i+t}) + (a_{i+1} + a_{i+2} + \cdots + a_n)$$

giving the result. □

Proposition 1. *If* $k \leq \lceil n/2 \rceil$ *then* lex(k) *is equivalent to the following inequalities*

$$a_{n-2j+1} + \sum_{r=2}^{j} a_{n-j+r} < \sum_{r=2}^{j+1} a_{n-2j+r}, \quad 1 \leq j \leq k$$

$$a_{n-2j} + \sum_{r=2}^{j} a_{n-j+r} < \sum_{r=1}^{j} a_{n-2j+r}, \quad 1 \leq j \leq k$$

$$a_i + \sum_{r=2}^{k} a_{n-k+r} < \sum_{r=1}^{k} a_{i+r}, \quad 1 \leq i < n-2k.$$

Proof. The inequalities are clearly necessary. It is sufficient to show that they imply all the inequalities of the lex(k) condition. We show how to deduce a typical inequality

$$a_i + a_{n-t+2} + \cdots + a_n < a_{i+1} + \cdots + a_{i+t}, \ k-t+1 \leq t \leq k, \ 1 \leq i \leq n-t. \quad (*)$$

In (*) we may assume there are no common terms on the left and right hand sides; if so, they may be mutually cancelled thereby effectively reducing t. Thus we may assume that $i+t < n-t+2$, that is $i \leq n-2t+1$. If $i = n-2t+1$ or $i = n-2t$ then (*) is one of the inequalities in the statement of the proposition, so we assume $i < n-2t$. Put $i = n-2j$ or $i = n-2j+1$ (according to the parity of $n-i$) with $j > t$. We consider only the latter case since the former case follows by a similar argument. One of the inequalities above is

$$a_{n-2j+1} + a_{n-j+2} + a_{n-j+3} + \cdots + a_n < a_{n-2j+2} + \cdots + a_{n-j+1} \quad (**)$$

and hence

$$(a_{n-2j+2} + a_{n-2j+3} + \cdots + a_{n-2j+t+1}) - (a_{n-2j+1} + a_{n-t+2} + a_{n-t+3} + \cdots + a_n)$$
$$= (a_{n-2j+2} + \cdots + a_{n-2j+t+1} + a_{n-2j+t+2} + \cdots + a_{n-j+1})$$
$$- (a_{n-2j+t+2} + \cdots + a_{n-j+1})$$
$$- (a_{n-2j+1} + a_{n-j+2} + \cdots + a_{n-t+1} + a_{n-t+2} + \cdots + a_n)$$
$$+ (a_{n-j+2} + \cdots + a_{n-t+1})$$
$$> (a_{n-j+2} + \cdots + a_{n-t+1}) - (a_{n-2j+t+2} + \cdots + a_{n-j+1}) \geq 0$$

(the latter inequality following from $j > t$). Hence

$$(a_{n-2j+1} + a_{n-t+2} + a_{n-t+3} + \cdots + a_n) < (a_{n-2j+2} + a_{n-2j+3} + \cdots + a_{n-2j+t+1})$$

or

$$a_i + a_{n-t+2} + \cdots + a_n < a_{i+1} + \cdots + a_{i+t}. \quad \square$$

Corollary 1. *If* $k \leqslant \lceil n/2 \rceil$ *then* $\operatorname{lex}(k) \Rightarrow \operatorname{lex}(k-1)$.

Corollary 2. $\operatorname{lex}(\lceil n/2 \rceil) \Rightarrow \operatorname{lex}(k)$ *for all* k.

3. Constructing optimal sets with lexicographic ordering

We can use Proposition 1 to *construct* sets $\{a_1, \ldots, a_n\}$ with the lex(k) property. Notice that $\{a_1, \ldots, a_n\}$ has the lex(k) property if and only if $\{a_1+p, \ldots, a_n+p\}$ also has it, and hence we may translate any such set of integers to obtain one with positive elements. In fact our interest is in lex(k) sets in which a_n is minimal (and these sets are called *optimal* sets). Therefore we should try to produce sets in which $a_n - a_i$ is minimal and then translate them so that a_1 maps to 1.

Our strategy is to choose an arbitrary value for a_n and then maximise a_1. To do this we define a lex(k) sequence $b_1, \ldots, b_n$ with $b_1 = 1$ for which the inequalities of the proposition are as 'tight' as possible, namely

$$b_{n-2j+1} = \sum_{r=2}^{j+1} b_{n-2j+r} - \sum_{r=2}^{j} b_{n-j+r} - 1, \quad 1 \leqslant j \leqslant k$$

$$b_{n-2j} = \sum_{r=1}^{j} b_{n-2j+r} - \sum_{r=2}^{j} b_{n-j+r} - 1, \quad 1 \leqslant j \leqslant k$$

$$b_i = \sum_{r=1}^{k} b_{i+r} - \sum_{r=2}^{k} b_{n-k+r} - 1, \quad 1 \leqslant i < n-2k.$$

These equations define $b_{n-1}, b_{n-2}, \ldots$ in terms of b_n and the boundary condition $b_1 = 1$ determines b_n.

Proposition 2. *If* $\{a_1, \ldots, a_n\}$ *is a set of positive integers with the* lex(k) *property then* $a_i \geqslant b_i$, $1 \leqslant i \leqslant n$, *and, if* $a_n = b_n$, *then* $a_i = b_i$, $1 \leqslant i \leqslant n$. *Thus* $\{b_1, \ldots, b_n\}$ *is the unique optimal set of size n with the* lex(k) *property.*

Proof. Let $c_i = a_i - b_i$. The relations between the a_i and b_i imply that $c_n \geqslant c_{n-1} \geqslant \cdots \geqslant c_1$. But, since a_1 is positive, $c_1 \geqslant 0$. $\square$

Example 1. The optimal set of 9 positive integers whose 3-subsets have lexicographically increasing sums is 1, 38, 58, 69, 75, 78, 80, 81, 82.

Theorem 1. *$b_n = O(\lambda^n)$ where λ is the largest root of $x^{k+1} - x^k + 1 = 0$.*

Proof. The defining conditions for $\{b_1, \ldots, b_n\}$ may be rewritten in terms of $d_i = b_n - b_{n-i}$ as

$$d_i = \sum_{j=1}^{k} d_{i-j} - \sum_{j=1}^{k-2} d_j + 1 \leqslant \sum_{j=1}^{k} d_{i-j}$$

together with $2k$ equations which determine $d_1, \ldots, d_{2k}$. This recurrence has solution $d_i = O(\lambda^i)$ where λ is the largest root of $x^k - x^{k-1} - x^{k-2} - \cdots - x - 1 = 0$. Multiplying this equation by $x - 1$ gives $x^{k+1} - x^k + 1 = 0$. However $b_n - 1 = b_n - b_1 = d_{n-1} = O(\lambda^n)$. □

In the case $k = \lceil n/2 \rceil$, these results show how to obtain sets of positive integers whose k-subsets, for each k, are lexicographically ordered and which have a_n minimal. In fact, if we extend the lexicographic order to all sets, listing first the singletons, then the doubletons etc, we can ask for sets *all* of whose subsets have the lexicographically ordered sum property; as mentioned previously, this was the condition studied by Petit. We take an optimal set of $n + 1$ integers $a_0, a_1, \ldots, a_n$ with lex($\lceil (n+1)/2 \rceil$) and translate it so that $a_0 = 0$. Then, for each k, the inequality

$$a_0 + a_{n-k+1} + \cdots + a_n < a_1 + a_2 + \cdots + a_{k+1}$$

implies that the $(k + 1)$-subset sums exceed the k-subset sums.

The following sequence provides a very handy method of constructing these optimal lexicographic sets which are knapsack-solvable in linear time: 1, 1, 2, 3, 6, 11, 22, 42, ... The rule is

$$c_1 = c_2 = 1$$

$$c_{i+1} = 2c_i \text{ if } i \text{ is even}$$

$$c_{i+1} = 2c_i - c_j, \text{ where } j = [i/2] \text{ if } i \text{ is odd.}$$

Then, for any given n, we can derive the optimal lexicographic set $\{a_1, a_2, \ldots, a_n\}$ of n integers by the rules

$$a_n = c_1 + c_2 + \cdots + c_n$$

and

$$a_{n-i} = a_{n-i+1} - c_i, \quad i = 1, 2, \ldots, n - 1.$$

This follows from a straightforward but tedious check that the sequence satisfies the optimality conditions given above. If we let e_n denote the largest element a_n in the optimal lexicographic set on n elements then, from the recurrences above, we have

$$e_{n+1} = 2e_n - e_j, \quad \text{where } j = \left[\frac{n}{2} - \frac{1}{2}\right].$$

n	1	2	3	4	5	6	7	8	9
e_n	1	2	4	7	13	24	46	88	172

Fig. 2.

Example 2. The optimal 7-element set whose sums are lexicographically ordered is

$$22, 33, 39, 42, 44, 45, 46.$$

Numerical computations show that e_n satisfies the asymptotic formula

$$e_n/2^{n-1} = 0.6333683473\ldots \text{ as } n \to \infty.$$

The first few values of e_n are shown in Fig. 2.

References

[1] R.K. Guy, Sets of integers whose subsets have distinct sums, Annals of Discrete Maths. 12 (1982) 141–154.

[2] R.K. Guy, Unsolved problems in number theory, pp. 64–65, (Springer-Verlag 1981).

[3] W.F. Lunnon, Integer sets with distinct subset sums, to appear in Math. Comp.

[4] R.C. Merkle and M.E. Hellman, Hiding information and signatures in trapdoor knapsacks, IEEE Trans. on Info. Theory IT-24(5) (1978) 525–530.

[5] M. Petit, Étude mathématique de certains systèmes de chiffrement: les sacs à dos, Doctoral thesis, University of Rennes (1982).