SUMS OF LEXICOGRAPHICALLY ORDERED SETS

M.D. Atkinson[1], A. Negro[2] and
N. Santoro[1]

SCS-TR-100

May 1987

School of Computer Science
Carleton University
Ottawa, Ontario
CANADA K1S 5B6

[1]School of Computer Science, Carleton University, Ottawa, Canada
[2]Dipartimento di Informatica, Universita' di Salerno, Salerno, Italy

Sums of lexicographically ordered sets

M.D. Atkinson[1], A. Negro[2], N. Santoro[1]

[1]School of Computer Science, Carleton University, Ottawa, Canada
[2]Dipartimento di Informatica, Universita' di Salerno, Salerno, Italy

Abstract. We consider the problem of determining finite integer sets which are knapsack-solvable in linear time (i.e., 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 $A=\{a_1,....,a_n\}$ of n distinct positive integers, where $a_i<a_{i+1}$ $(1 \le i < n)$. 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 a algorithm *A* 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) has 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

1

is due to the fact that this set has the following property

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

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 $a_n \ge 2^{n-1}$. The question open is whether there exists a family of sets which are solvable in linear time *and* $a_n < 2^{n-1}$. 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,

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

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

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

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 even quadratically 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 $\{a_1,...,a_n\}$. In this ordering all the k-subsets (subsets of cardinality k) come before the (k+1)-subsets (k=0,1, ..., n-1) and, for each k, the k-subsets are listed in *lexicographic order*. For example, the ordering for n=4 is shown in Figure 1.

$$\{\ \}$$
$$\{a_1\}$$
$$\{a_2\}$$
$$\{a_3\}$$
$$\{a_4\}$$
$$\{a_1,a_2\}$$
$$\{a_1,a_3\}$$
$$\{a_1,a_4\}$$
$$\{a_2,a_3\}$$
$$\{a_2,a_4\}$$
$$\{a_3,a_4\}$$
$$\{a_1,a_2,a_3\}$$
$$\{a_1,a_2,a_4\}$$
$$\{a_1,a_3,a_4\}$$
$$\{a_2,a_3,a_4\}$$
$$\{a_1,a_2,a_3,a_4\}$$

Figure 1. Complete lexicographic ordering of the subsets of $\{a_1,a_2,a_3,a_4\}$

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 A of $\{a_1,....,a_n\}$ whose sum s is given. Suppose that the partial sums $s_i = a_1 + a_2 + ... + a_i$ have been computed. Then $k = |A|$ is determined as $k = \max\{t : s < s_t\} - 1$. Now we successively determine whether $a_1 \in A$, $a_2 \in A,....$ At the time that we are determining whether $a_i \in A$ we shall know of $B = A \cap \{a_i, a_{i+1}, ..., a_n\}$ both $r = |B|$ and $s = \sum_{b \in B} b$. Then $a_i \in A$ if and only if $s < s_{i+r} - s_i = a_{i+1} + ... + a_{i+r}$. 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. From these sequences a simple trick allows us to impose 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 $A=\{a_1,....,a_n\}$ 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 $(a_{i(1)}, ..., a_{i(k)})$ whose components are the members of S listed in increasing order of subscript. If $(a_{i(1)}, ..., a_{i(k)})$ and $(a_{j(1)}, ..., a_{j(k)})$ are any two such vectors we say the first precedes the second if, for some r,

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

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 $1\leq t\leq k$ and $k-t+1\leq i\leq n-t$, $\qquad a_i + \sum_{j=2,n} a_{n-t+j} < \sum_{j=1,t} a_{i+j}$

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}, ... which is not in B. Of course $1\leq t\leq k$. We also define a_i to be the element of $B-\{a_{n-t+2},a_{n-t+3},...,a_n\}$ with largest subscript. Then $i\leq n-t$ since $a_{n-t+1}\notin B$. Moreover i is the largest of $k-t+1$ such subscripts and so $i\geq k-t+1$. Thus $B =\{a_j,a_k,....,a_i,a_{n-t+2},a_{n-t+3},....,a_n\}$ (increasing subscripts) and its successor is $\{a_j,a_k,...,a_{i+1},a_{i+2},a_{i+3},...,a_{i+t}\}$. Hence the lex(k) condition is precisely that $a_j+ a_k+....+a_i+a_{n-t+2}+a_{n-t+3}+....+a_n < a_j+a_k+....+a_{i+1}+a_{i+2} ....+a_{i+t}$ as required. []

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

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

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

4

and hence

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

and hence

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

giving the result. []

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

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

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

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

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

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

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

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

and hence

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

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

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

or

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

Corollary 1. If $k \leq \lceil n/2 \rceil$ then lex(k) $\Rightarrow$ lex(k-1).

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

3. Constructing optimal sets with lexicographic ordering

We can use Proposition 1 to *construct* sets $\{a_1,....,a_n\}$ with the lex(k) property. Notice that $\{a_1,....,a_n\}$ has the lex(k) property if and only if $\{a_1+p, ..., a_n+p\}$ 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. Therefore we should try to produce sets in which a_n-a_1 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 sequence $b_1, ..., b_n$ with $b_1 = 1$ for which the inequalities of the proposition are as 'tight' as possible, namely

$$b_{n-2j+1} = \sum_{r=2,j+1} b_{n-2j+r} - \sum_{r=2,j} b_{n-j+r} - 1, \qquad 1 \le j \le k$$
$$b_{n-2j} = \sum_{r=1,j} b_{n-2j+r} - \sum_{r=2,j} b_{n-j+r} - 1, \qquad 1 \le j \le k$$
$$b_i = \sum_{r=1,k} b_{i+r} - \sum_{r=2,k} b_{n-k+r} - 1, \qquad 1 \le i < n-2k$$

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

Proposition 2   If $\{a_1,....,a_n\}$ is a set of positive integers with the lex(k) property then $a_i \ge b_i$ for $i = 1, 2, ..., n$.

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

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 $\{b_1, ..., b_n\}$ may be rewritten in terms of $d_i = b_n - b_{n-i}$ as

6

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

together with 2k equations which determine $d_1, ..., d_{2k}$. This recurrence relation has solution $d_i = O(\lambda^i)$ where λ is the largest root of $x^k - x^{k-1} + x^{k-2} + ... + x + 1 = 0$. 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 with a_n *minimal* and, for each k, the *k-subsets having lexicographically ordered sums*. In fact, if we extend the lexicographic order to the complete lexicographic order, 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 $a_0, a_1, ..., a_n$ with lex($\lceil (n+1)/2 \rceil$) and translate it so that $a_0 = 0$. Then, for each k, the inequality

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

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 knapsack-solvable in linear time*: 1,1,2,3,6,11,22,42,.... The rule is

$c_1 = c_2 = 1$

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

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

Then, for any given n, we can put

$$a_n = c_1 + c_2 + ... + c_n$$

and $a_{n-i} = a_{n-i+1} - c_i$, i=1,2,...,n-1.

This follows from the definitions and a certain amount of patience.

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

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

Numerical computations show that the largest element a_n satisfies the asymptotic formula

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

The first few values of the optimal a_n are shown in Figure 2.

7

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

Figure 2.

References

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

[2] Guy, R.K.: Unsolved problems in Number Theory, pp.64-65, Springer-Verlag (1981).

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

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

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

**School of Computer Science
Carleton University
Ottawa, Ontario, Canada
KIS 5B6**

SCS-TR-66   **On the Futility of Arbitrarily Increasing Memory Capabilities of Stochastic Learning Automata**
_____   B.J. Oommen, October 1984. Revised May 1985.

SCS-TR-67   **Heaps in Heaps**
_____   T. Strothotte, J.-R. Sack, November 1984. Revised April 1985.

SCS-TR-68
out-of-print   **Partial Orders and Comparison Problems**
M.D. Atkinson, November 1984. See Congressus Numerantium 47 ('86), 77-88

SCS-TR-69   **On the Expected Communication Complexity of Distributed Selection**
_____   N. Santoro, J.B. Sidney, S.J. Sidney, February 1985.

SCS-TR-70   **Features of Fifth Generation Languages: A Panoramic View**
_____   Wilf R. LaLonde, John R. Pugh, March 1985.

SCS-TR-71   **Actra: The Design of an Industrial Fifth Generation Smalltalk System**
_____   David A. Thomas, Wilf R. LaLonde, April 1985.

SCS-TR-72   **Minmaxheaps, Orderstatisticstrees and their Application to the Coursemarks Problem**
_____   M.D. Atkinson, J.-R. Sack, N. Santoro, T. Strothotte, March 1985.

SCS-TR-73   **Designing Communities of Data Types**
Wilf R. LaLonde, May 1985.
Replaced by SCS-TR-108

SCS-TR-74
out-of-print   **Absorbing and Ergodic Discretized Two Action Learning Automata**
B. John Oommen, May 1985. See IEEE Trans. on Systems, Man and Cybernetics, March/April 1986, pp. 282-293.

SCS-TR-75   **Optimal Parallel Merging Without Memory Conflicts**
_____   Selim Akl and Nicola Santoro, May 1985

SCS-TR-76   **List Organizing Strategies Using Stochastic Move-to-Front and Stochastic Move-to-Rear Operations**
_____   B. John Oommen, May 1985.

SCS-TR-77   **Linearizing the Directory Growth in Order Preserving Extendible Hashing**
_____   E.J. Otoo, July 1985.

SCS-TR-78   **Improving Semijoin Evaluation in Distributed Query Processing**
_____   E.J. Otoo, N. Santoro, D. Rotem, July 1985.

SCS-TR-79   **On the Problem of Translating an Elliptic Object Through a Workspace of Elliptic Obstacles**
B.J. Oommen, I. Reichstein, July 1985.

SCS-TR-80   **Smalltalk - Discovering the System**
W. LaLonde, J. Pugh, D. Thomas, October 1985.

SCS-TR-81   **A Learning Automation Solution to the Stochastic Minimum Spanning Circle Problem**
B.J. Oommen, October 1985.

SCS-TR-82   **Separability of Sets of Polygons**
Frank Dehne, Jörg-R. Sack, October 1985.

SCS-TR-83
out-of-print   **Extensions of Partial Orders of Bounded Width**
M.D. Atkinson and H.W. Chang, November 1985. See Congressus Numerantium, Vol. 52 (May 1986), pp. 21-35.

SCS-TR-84   **Deterministic Learning Automata Solutions to the Object Partitioning Problem**
B. John Oommen, D.C.Y. Ma, November 1985

SCS-TR-85
out-of-print   **Selecting Subsets of the Correct Density**
M.D. Atkinson, December 1985. To appear in Congressus Numerantium, Proceedings of the 1986 South-Eastern conference on Graph theory, combinatorics and Computing.

SCS-TR-86   **Robot Navigation in Unknown Terrains Using Learned Visiblity Graphs. Part I: The Disjoint Convex Obstacles Case**
B. J. Oommen, S.S. Iyengar, S.V.N. Rao, R.L. Kashyap, February 1986

SCS-TR-87   **Breaking Symmetry in Synchronous Networks**
Greg N. Frederickson, Nicola Santoro, April 1986

SCS-TR-88   **Data Structures and Data Types: An Object-Oriented Approach**
John R. Pugh, Wilf R. LaLonde and David A. Thomas, April 1986

SCS-TR-89   **Ergodic Learning Automata Capable of Incorporating Apriori Information**
B. J. Oommen, May 1986

SCS-TR-90   **Iterative Decomposition of Digital Systems and Its Applications**
Vaclav Dvorak, May 1986.

SCS-TR-91   **Actors in a Smalltalk Multiprocessor: A Case for Limited Parallelism**
Wilf R. LaLonde, Dave A. Thomas and John R. Pugh, May 1986

SCS-TR-92   **ACTRA - A Multitasking/Multiprocessing Smalltalk**
David A. Thomas, Wilf R. LaLonde, and John R. Pugh, May 1986

SCS-TR-93   **Why Exemplars are Better Than Classes**
Wilf R. LaLonde, May 1986

SCS-TR-94   **An Exemplar Based Smalltalk**
Wilf R. LaLonde, Dave A. Thomas and John R. Pugh, May 1986

SCS-TR-95   **Recognition of Noisy Subsequences Using Constrained Edit Distances**
B. John Oommen, June 1986

SCS-TR-96   **Guessing Games and Distributed Computations in Synchronous Networks**