JOURNAL OF COMBINATORIAL THEORY, Series A **77**, 344–348 (1997)
ARTICLE NO. TA972732

NOTE

On a Reconstruction Problem for Sequences*

I. Krasikov and Y. Roditty

School of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, Israel 69978; and Department of Mathematics, Beit–Berl College, Kfar–Saba, Israel 44905

Communicated by the Managing Editors

Received August 11, 1995

It is shown that any word of length n is uniquely determined by all its $\binom{n}{k}$ subwords of length k, provided $k \geqslant \lfloor \frac{16}{7}\sqrt{n} \rfloor + 5$. This improves the bound $k \geqslant \lfloor n/2 \rfloor$ given in B. Manvel *et al.* (*Discrete Math.* **94** (1991), 209–219).

1. INTRODUCTION

Given a word X of length n with terms from an alphabet Σ, define the *k-deck* of X, $D_k(X)$, to be the multiset of all $\binom{n}{k}$ k-subwords of X. The following reconstruction problem is due to Kalashnik [4],

When is X uniquely determined by $D_k(X)$?

We call X *k-reconstructible* when this is the case. We call words X and Y *k-equivalent* if $D_k(X) = D_k(Y)$ (We will only use this term for *distinct* words).

This question resembles the well-known vertex reconstruction problem (see Bondy [1]) but seems more tractable. The problem has been treated in [5] by Manvel *et. al.* They proved that any word of length n is k-reconstructible whenever $k \geqslant \lfloor n/2 \rfloor$. On the other hand, they gave a construction of nonreconstructible words for $k \leqslant \log_2 n$. It also has been shown in [5] that without loss of generality we can restrict the problem to words over the alphabet $\{0, 1\}$, in view of the following:

All words of length n with terms from an alphabet Σ are k-reconstructible if and only if all words of the same length with terms from $\{0, 1\}$ are k-reconstructible.

* Research supported in part by the Beit–Berl College Research Foundation.

0097-3165/97 $25.00

Here we establish k-reconstructibility of words of length n for $k \geqslant \lfloor \frac{16}{7} \sqrt{n} \rfloor + 5$. For, we will compare "average" words

$$\binom{n}{k}^{-1} \sum_{X \in D_k(X)} X, \qquad \binom{n}{k}^{-1} \sum_{Y \in D_k(Y)} Y.$$

It turned out that this leads to the Prouhet–Tarry–Escott problem of Diophantine analysis and our result immediately follows from the bound obtained in [3].

Many variations of the above problem are possible. For example, what happens if all words are given up to their complements? What can be said about reconstruction of cyclic words?

2. RESULTS

In what follows we will deal only with words over the alphabet $\{0, 1\}$. For convenience we use indices 0, 1, ..., $n-1$ for letters of words.

We start with the following counting lemma.

LEMMA 2.1. *Let $X = x_0 x_1 \cdots x_{n-1}$, and let $S_j(X)$ be the total number of 1's appearing in the j's place in all words of $D_k(X)$. Then*

$$S_j(X) = \sum_{i=0}^{n-1} \binom{i}{j} \binom{n-i-1}{k-j-1} x_i, \qquad j = 0, 1, ..., k-1.$$

Proof. If $x_i = 1$ then it contributes $\binom{i}{j}\binom{n-i-1}{k-j-1}$ ones to the j's places of the deck. That is since the first j letters from 0 to $j-1$ of a word in the deck are chosen from the first i letters of X, while the last $k-j-1$ letters starting from the $j+1$, are chosen from the last $n-i-1$ letters of X. Summing upon all the x_i we obtain the required result. ∎

Let X and Y be two k-equivalent words. Define the following vector with components from $\{-1, 0, 1\}$: $\delta = \delta_0 \delta_1 \cdots \delta_{n-1}$, where $\delta_j = x_j - y_j$. As an immediate corollary of the previous lemma we get:

LEMMA 2.2.

$$\sum_{i=0}^{n-1} \binom{i}{j} \binom{n-i-1}{k-j-1} \delta_i = 0, \qquad j = 0, 1, ..., k-1.$$

Consider now the following set of polynomials

$$f_j^{n,k}(t) = f_j(t) = \binom{t}{j} \binom{n-t-1}{k-j-1}, \qquad j = 0, 1, ..., k-1.$$

Note that $f_j(t)$ is a polynomial of degree $k-1$ and has distinct integer roots since $f_j(i)=0$ for $0 \leqslant i < j$ and $n-k+j < i \leqslant n-1$.

LEMMA 2.3. *For fixed integers n and k, n, $k \geqslant 1$, the set $\{f_0(t), ..., f_{k-1}(t)\}$ is a basis for the space of polynomials of degree $k-1$.*

Proof. Consider

$$\phi(i) = \sum_{j=0}^{k-1} \lambda_j f_j(i) = \sum_{j=0}^{k-1} \lambda_j \binom{i}{j}\binom{n-i-1}{k-j-1}.$$

It is enough to show that $\phi(i)$ is not identically zero whenever the λ_i are not all zero. Assume the contrary and let λ_t be the first nonzero coefficient. Then $\phi(t) = \lambda_t \binom{n-i-1}{k-t-1}$, since $\binom{t}{j}=0$ for $j > t$. Hence $\phi(t)=0$ implies that $\lambda_t = 0$, a contradiction. ∎

Combining the two previous lemmas we obtain the following necessary condition for nonreconstructibility of words of length n.

COROLLARY 2.4. *If X and Y are k-equivalent then for any polynomial $\phi(j)$ of degree at most $k-1$,*

$$\sum_{j=0}^{n-1} \delta_j \phi(j) = 0. \tag{1}$$

Proof.

$$\sum_{j=0}^{n-1} \delta_j \phi(j) = \sum_{j=0}^{n-1} \delta_j \sum_{i=0}^{k-1} \lambda_i f_i(j) = \sum_{i=0}^{k-1} \lambda_i \sum_{j=0}^{n-1} \delta_j f_i(j) = 0. \quad ∎$$

Observe that if X and Y have the same k-decks they have also the same number of ones. To see this just choose $\phi(i)=1$ in (1).

If we denote by u_i and w_i, $i=1, 2, ..., s$, the indices of ones of the words X and Y, respectively, then (1) is equivalent to the following system:

$$\begin{aligned} &u_1^h + u_2^h + \cdots + u_s^h = w_1^h + w_2^h + \cdots + w_s^h, \qquad h=1, ..., k-1, \\ &u_1 < u_2 < \cdots < u_s, \qquad w_1 < w_2 < \cdots < w_s, \end{aligned} \tag{2}$$

and u_i and w_i are integers from the interval $[0, n-1]$.

Of course, this system ever has a trivial solution $u_i = w_i$, $i=0, ..., s$. Thus, we obtain

THEOREM 1. *If X and Y are k-equivalent then* (2) *has a nontrivial solution with u_i, $w_i \in [0, n-1]$.*

A problem of finding two distinct sets of integers $\{u_i\}$ and $\{w_i\}$ satisfying (2) is a classic (more than 200 years old) problem of Diophantine analysis, usually referred as the Prouhet–Tarry–Escott problem (see [2] for extensive discussion). Recently Borwein, Erdelyi, and Kos proved that (2) has only trivial solutions whenever $k \geqslant \lfloor \frac{16}{7} \sqrt{n} \rfloor + 5$. They construct a polynomial $\phi(j)$ of degree $k-1 = \lfloor \frac{16}{7} \sqrt{n} \rfloor + 4$, such that $|\phi(0)| > \sum_{j=1}^{n-1} |\phi(j)|$, clearly contradicting (1). This immediately yields our main result.

MAIN THEOREM. *A word X of length n is reconstructible from $D_k(X)$ provided $k \geqslant \lfloor \frac{16}{7} \sqrt{n} \rfloor + 5$.*

Notice that simple counting arguments show that for

$$k < c \sqrt{n/\log_2 n}, \qquad c < 2,$$

(2) already has a nontrivial solution [2] (it is enough to consider all words with about $n/2$ ones). Thus, our method cannot yield an essentially better bound. Yet, to obtain (2) we have actually compared the average number of ones appearing at the jth place in $D_k(X)$ and $D_k(Y)$. It seems plausible that taking into consideration higher moments may provide a sharper result.

In conclusion, let us notice that (2) yields also some conditional results, if some information about the number of ones in X and Y is known. For instance, considering the words as binary vectors and introducing the usual Hamming distance, we obtain:

COROLLARY 2.5. *If $0 < \operatorname{dist}(X, Y) < 2k$ then $D_k(X) \neq D_k(Y)$.*

Proof. We may assume that X and Y have the same number s of ones. So $\operatorname{dist}(X, Y)$ is even. Cancelling if necessary, one can take in (2) just $s = \operatorname{dist}(X, Y)/2$. It is well known and easily follows from the theory of symmetric functions that for $k-1 \geqslant s$ any solution of (2) satisfies $u_i = w_i$, that is, $X = Y$. ∎

ACKNOWLEDGMENTS

The authors thank N. Alon, Y. Caro, M. Krivelevich, and S. Litsyn for fruitful discussions. We are also very indebted to the anonymous referee for pointing out the connection of our problem to the Prouhet–Tarry–Escott problem and establishing the references on the works of P. Borwein *et al.* which provided a better bound than that which appears in the original version of the paper.

REFERENCES

1. J. A. Bondy, A graph reconstructor's manual, "Surveys in Combinatorics" (A. D. Keedwell, Ed.), pp. 221–252.
2. P. Borwein and C. Engalls, The Prouhet–Tarry–Escott problem revisited, *Ens. Math.* **40** (1994), 3–27.
3. P. Borwein, T. Erdelyi, and G. Kos, Littewood-type problems on [0, 1], preprint.
4. L. I. Kalashnik, The reconstruction of a word from fragments, *in* "Numerical Mathematics and Computer Technology," pp. 56–57, Akad. Nauk Ukrain. SSR Inst. Mat., Preprint IV, (1973). [Russian]
5. B. Manvel, A. Meyerowitz, A. Schwenk, K. Smith, and P. Stockmeyer, Reconstruction of sequences, *Discrete Math.* **94** (1991), 209–219.