

## Planar Realizations of Nonlinear Davenport–Schinzel Sequences by Segments\*

Ady Wiernik<sup>1</sup> and Micha Sharir<sup>1,2</sup>

<sup>1</sup> School of Mathematical Sciences, Tel-Aviv University, 69978 Tel-Aviv, Israel

<sup>2</sup> Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, NY 10012, USA

**Abstract.** Let  $G = \{l_1, \dots, l_n\}$  be a collection of  $n$  segments in the plane, none of which is vertical. Viewing them as the graphs of partially defined linear functions of  $x$ , let  $Y_G$  be their lower envelope (i.e., pointwise minimum).  $Y_G$  is a piecewise linear function, whose graph consists of subsegments of the segments  $l_i$ . Hart and Sharir [7] have shown that  $Y_G$  consists of at most  $O(n\alpha(n))$  segments (where  $\alpha(n)$  is the extremely slowly growing inverse Ackermann's function). We present here a construction of a set  $G$  of  $n$  segments for which  $Y_G$  consists of  $\Omega(n\alpha(n))$  subsegments, proving that the Hart–Sharir bound is tight in the worst case.

Another interpretation of our result is in terms of Davenport–Schinzel sequences: the sequence  $E_G$  of indices of segments in  $G$  in the order in which they appear along  $Y_G$  is a Davenport–Schinzel sequence of order 3, i.e., no two adjacent elements of  $E_G$  are equal and  $E_G$  contains no subsequence of the form  $a \cdots b \cdots a \cdots b \cdots a$ . Hart and Sharir have shown that the maximal length of such a sequence composed of  $n$  symbols is  $\Theta(n\alpha(n))$ . Our result shows that the lower bound construction of Hart and Sharir can be realized by the lower envelope of  $n$  straight segments, thus settling one of the main open problems in this area.

### 1. Introduction

Let  $G$  be a set of  $n$  straight line segments in the plane, none of which is vertical. The lower envelope of  $G$ , denoted by  $Y_G$ , is the pointwise minimum of these

---

\* Work on this paper has been partially supported by Office of Naval Research Grant N00014-82-K-0381, National Science Foundation Grant No. NSF-DCR-83-20085, and by grants from the Digital Equipment Corporation, and the IBM Corporation. This paper is part of the first author's M.Sc. thesis prepared at Tel Aviv University under the supervision of the second author. A preliminary version of this paper has appeared in *Proceedings of the 27th IEEE Symposium on Foundations of Computer Science*, Toronto, 97–106, 1986.

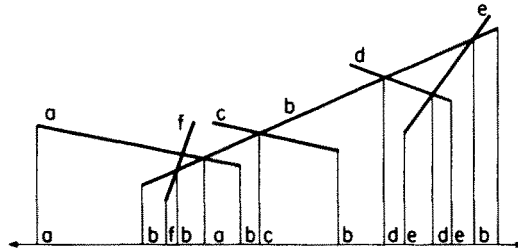


Fig. 1.1

segments when viewed as partially defined functions of  $x$ .  $Y_G$  is a (not necessarily continuous) piecewise linear function in  $x$  whose graph consists of portions of the segments of  $G$ . The main goal of this paper is to estimate the maximal number of segments that such  $Y_G$  can have. See Fig. 1.1 for an illustration.

This problem is a special case of the following general problem arising in computational geometry: let  $G$  be a set of  $n$  real-valued continuous functions  $f_1, \dots, f_n$  defined over the real axis. Suppose that for each  $i \neq j$  the functions  $f_i$  and  $f_j$  intersect in at most  $s$  points (e.g., this is the case for polynomials of fixed degree, or Chebyshev systems, and so on). Let  $Y_G(x) = \min\{f_i(x) : 1 \leq i \leq n\}$  be the *lower envelope* (i.e., the pointwise minimum) of the  $f_i$ 's. The graph of  $Y_G$  consists of (maximal) connected portions of the graphs of the functions  $f_i$ . Let  $\lambda_s(n)$  denote the maximum number of such portions, taken over all possible collections of  $n$  functions satisfying the above conditions.

An equivalent way of looking at the problem is as follows. Let  $U = U(f_1, \dots, f_n)$  denote the sequence of indices  $(u_1, u_2, \dots, u_m)$  so that the  $k$ th leftmost portion of the graph  $Y_G$  is a portion of the graph of the function  $f_{u_k}$ . Then  $U$  is easily seen to satisfy the following properties:

- (1)  $u_i \in \{1, \dots, n\}$  for each  $i$ .
- (2) For each  $i < m$  we have  $u_i \neq u_{i+1}$ .
- (3) There do not exist  $s+2$  indices  $1 \leq i_1 < i_2 < \dots < i_{s+2} \leq m$  such that  $u_{i_1} = u_{i_3} = u_{i_5} = \dots = a$  and  $u_{i_2} = u_{i_4} = u_{i_6} = \dots = b$ , for  $1 \leq a \neq b \leq n$ .

A sequence  $U = (u_1, \dots, u_m)$  satisfying these properties is called an  $(n, s)$  *Davenport-Schinzel sequence* (a  $DS(n, s)$  sequence for short). These sequences were introduced by Davenport and Schinzel in [6] and have been analyzed later in [5], [12], [15], [2], [7], [13], and [14].

Thus  $U(f_1, \dots, f_n)$  is a  $DS(n, s)$  sequence. Moreover, it is known (see [2]) that for any  $DS(n, s)$  sequence  $U$  one can construct a set  $f_1, \dots, f_n$  of continuous functions, each pair of which intersect in at most  $s$  points, for which  $U(f_1, \dots, f_n) = U$ . Such a set of functions is said to *represent* or *realize* the sequence  $U$ .

The following bounds are known for  $\lambda_s(n)$ :

$$\lambda_1(n) = n \quad (\text{Trivial}),$$

$$\lambda_2(n) = 2n - 1 \quad [2],$$

$$\lambda_3(n) = \Theta(n\alpha(n)) \quad [7],$$

$$\lambda_s(n) = O(n\alpha(n)^{O(\alpha(n)^{s-3})}) \quad \text{for } s > 3 \quad [13],$$

$$\lambda_{2s+1}(n) = \Omega(n\alpha^s(n)) \quad [14]$$

( $\alpha(n)$  is the functional inverse of Ackermann’s function and is very slowly growing).

If the functions  $f_1, \dots, f_n$  are partially defined linear functions (i.e., their graphs are straight segments), then one can extend them continuously to totally defined functions so that each pair of extended functions intersect in at most three points (see [2] and [7]).

Hence the maximum number of segments composing  $Y_G$ , for a collection  $G$  of  $n$  segments, is  $\lambda_3(n) = O(n\alpha(n))$ . However, even though there exist  $\text{DS}(n, 3)$  sequences  $U$  of length  $\Omega(n\alpha(n))$  [7], the construction (e.g., in [2]) that realizes such sequences  $U$  by collections of functions yields rather irregularly shaped functions, and it was an open problem whether  $U$  can also be realized by collections of functions of simple shape, in particular by collections of segments.

In this paper we show that this is indeed the case. Specifically, we take the  $\text{DS}(n, 3)$  sequences of nonlinear size produced by Hart and Sharir and realize them by collections of segments. Our construction is fairly involved, and is based on a doubly inductive process which follows the inductive pattern in the construction of [7].

Since the construction in [7] is not explicit, but rather defined in terms of generalized path compressions on trees, we first provide, in Section 2, an explicit inductive construction of the sequences of [7] and analyze their structure (a similar construction has recently been obtained by Komjath [8]). Section 3 then proceeds to describe the inductive realization of these sequences by collections of segments. For the convenience of the reader, two appendices are provided, summarizing the basic definitions used in our sequence construction, and giving a few initial examples of these sequences.

Our results are useful in setting tight bounds on the complexity of various problems in computational geometry. The concluding section, Section 4, mentions some of these problems and also discusses some related open problems.

## 2. Generation of Nonlinear $\text{DS}(n, 3)$ Sequences

In this section we provide an explicit construction of  $\text{DS}(n, 3)$  sequences whose length is nonlinear in  $n$ . This construction is essentially identical to the tree-based

construction of [7] but avoids the use of auxiliary constructs such as generalized path compressions (used in [7]). It proceeds inductively, following the double induction scheme of [7], and constructs a collection of sequences  $S(k, m)$ , defined in terms of two positive integer parameters  $k$  and  $m$ . Each sequence  $S(k, m)$  is constructed from subsequences defined inductively for smaller values of  $k$  and  $m$ , which are then modified and merged to form  $S(k, m)$ .

### 2.1. The Functions $C_k$

In the construction we use a sequence of functions  $\{C_k\}$  (see [14]), which are similar to Ackermann's functions  $A_k$  (the "generalized exponentials" of [1]), but are more suitable for the purpose of our construction.

Let  $\mathbb{N} = \{1, 2, \dots\}$  and  $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ . Define inductively a sequence of functions  $\{C_k: \mathbb{N}_0 \rightarrow \mathbb{N}\}_{k=1}^{\infty}$  as follows.

$$C_1(m) = 1, \quad m \geq 0,$$

$$C_k(0) = 2, \quad k \geq 2,$$

$$C_k(m) = C_k(m-1) \cdot C_{k-1}(C_k(m-1)), \quad k \geq 2, \quad m \geq 1.$$

A table of some "small" values of  $C_k(m)$  is given in Appendix 1. (For comparison, the original Ackermann's functions  $\{A_k: \mathbb{N} \rightarrow \mathbb{N}\}_{k=1}^{\infty}$  are defined as:

$$A_1(m) = 2m, \quad m \geq 1,$$

$$A_k(1) = 2, \quad k \geq 2,$$

$$A_k(m) = A_{k-1}(A_k(m-1)), \quad k \geq 2, \quad m \geq 2.$$

A table of some "small" values of  $A_k(m)$  is given in Appendix 1.) It can easily be proved that (see also [14]):

- (1)  $C_2(m) = 2$  for  $m \geq 0$ .
- (2)  $C_3(m) = 2^{m+1}$  for  $m \geq 0$ .
- (3)  $C_4(m) \geq 2^{2^{\dots^2}}$ , with  $m+1$  two's in the exponential tower.
- (4)  $C_k(1) = 2 \cdot C_{k-1}(2)$ .
- (5)  $A_{k-1}(m) \leq C_k(m) \leq A_k(m+3)$  for  $k \geq 4$ ,  $m \geq 1$ , so the growth of the sequences of functions  $\{C_k\}$  and  $\{A_k\}$  are of the same order of magnitude.

In what follows we will often use the shorthand notations  $\bar{\alpha} = C_k(m-1)$ ,  $\bar{\beta} = C_{k-1}(C_k(m-1))$ , and  $\bar{\gamma} = C_k(m) = \bar{\alpha} \cdot \bar{\beta}$  (by definition).

## 2.2. Generation of Nonlinear DS( $n, 3$ ) Sequences

For each  $k, m \geq 1$  the sequence  $S(k, m)$  that we are going to construct will satisfy the following properties:

- (a)  $S(k, m)$  is composed of  $N_k(m) = m \cdot C_k(m)$  distinct symbols. (These symbols are named  $(d, \gamma)$ , for  $d = 1, \dots, m, \gamma = 1, \dots, \bar{\gamma}$ , and are ordered so that  $(d, \gamma) < (d', \gamma')$  if  $\gamma < \gamma'$  or  $\gamma = \gamma'$  and  $d < d'$ .)
- (b) Define a *fan* to be a contiguous subsequence of the form

$$(1, \gamma) (2, \gamma) \cdots (m, \gamma)$$

such that all its elements are the leftmost appearances of the corresponding symbols in the entire sequence. Then  $S(k, m)$  contains  $C_k(m)$  fans of size  $m$  for  $\gamma = 1, 2, \dots, C_k(m)$ . These fans are pairwise disjoint and cover the entire set of characters in  $S$  (according to (a) above). Thus each symbol of  $S$  appears in *one* (and exactly one) fan of  $S$ . The naming scheme of the symbols of  $S(k, m)$  can be interpreted as assigning to each symbol the index  $\gamma$  of the fan in which it appears, and its index  $d$  within that fan.

**Definition.** Let  $U = (u_1, \dots, u_m)$  be a sequence whose symbols belong to some totally ordered alphabet. A *chain*  $c = (u_{i_0}, \dots, u_{i_1})$  is a maximal contiguous decreasing subsequence of  $U$ , i.e.,

$$(u_{i_0-1} <) u_{i_0} > u_{i_0+1} > \cdots > u_{i_1} (< u_{i_1+1}).$$

Chains are obviously disjoint and their union is the entire sequence  $U$ .

## 2.3. Inductive Construction of $S(k, m)$

The construction proceeds by double induction on  $k$  and  $m$  as follows.

- (1)  $k = 1$ . The sequence is a single fan of size  $m$ :

$$S = \langle (1, 1) (2, 1) \cdots (m, 1) \rangle.$$

Properties (a) and (b) clearly hold here ( $C_1(m) = 1$ ).

- (2)  $k = 2$ . The sequence contains a pair of disjoint fans of size  $m$ , with a chain of size  $m$  following each of these fans (and starting at the last element of the fan), and  $m - 1$  singleton chains following at the end of the sequence. Specifically,

$$\begin{aligned} S = \langle & (1, 1) (2, 1) \cdots (m-1, 1) (m, 1) (m-1, 1) \cdots (1, 1) \\ & (1, 2) (2, 2) \cdots (m-1, 2) (m, 2) (m-1, 2) \cdots (1, 2) \\ & (2, 2) (3, 2) \cdots (m, 2) \rangle. \end{aligned}$$

And indeed it contains  $C_2(m) = 2$  fans and is composed of  $2m$  symbols. (It can be easily checked that  $S(2, m)$  can also be obtained by the general construction steps 3 and 4 below; however, for exposition sake, we treat the case  $k = 2$  explicitly, both here and in Section 3 below.)

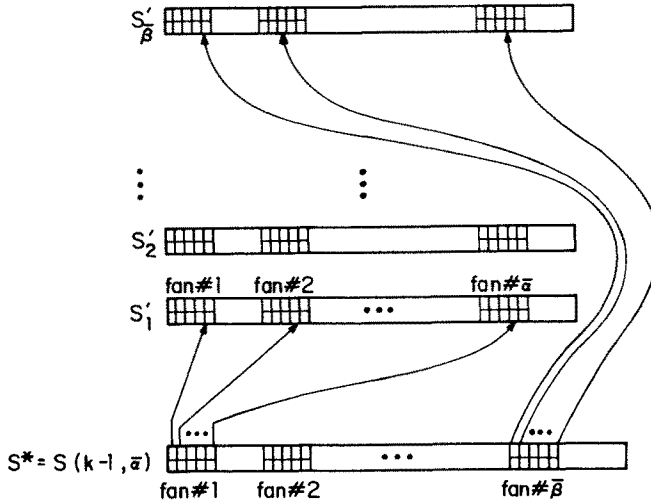


Fig. 2.1

- (3)  $k \geq 3$ ,  $m = 1$ . The sequence is identical to the sequence for  $k' = k - 1$  and  $m' = 2$ , except for renaming its symbols and fans:  $S(k-1, 2)$  contains  $C_{k-1}(2) = C_k(1)/2$  fans, each of which consists of two symbols. We regard each fan of  $S(k-1, 2)$  as two singleton fans in  $S(k, 1)$ . The  $i$ th symbol of the  $\beta$ th fan of  $S(k-1, 2)$  is named  $(i, \beta)$  in  $S(k-1, 2)$  and is renamed as  $(1, 2\beta - 2 + i)$  in  $S(k, 1)$  to reflect this fan splitting. Properties (a) and (b) clearly hold.
- (4) The general case  $k \geq 3$ ,  $m > 1$  (see Fig. 2.1 for an illustration of this step).
- (i) Generate inductively the sequence  $S' = S(k, m-1)$ ; by induction, it contains  $\bar{\alpha}$  fans of size  $m-1$  each and is composed of  $(m-1) \cdot \bar{\alpha}$  symbols whose names are  $(d, \alpha)$  ( $1 \leq d \leq m-1$  is the index of the symbol in the fan of  $S'$  containing it, and  $1 \leq \alpha \leq \bar{\alpha}$  is the index of this fan in  $S'$ ).
  - (ii) Create  $\bar{\beta}$  copies of  $S'$ . For each  $1 \leq \beta \leq \bar{\beta}$ , rename each symbol  $(d, \alpha)$  in the  $\beta$ th copy  $S'_\beta$  of  $S'$  as  $(d, \alpha, \beta)$ .
  - (iii) Generate inductively the sequence  $S^* = S(k-1, \bar{\alpha})$ ; by induction it contains  $\bar{\beta}$  fans of size  $\bar{\alpha}$  each. Rename the symbols  $(\alpha, \beta)$  of  $S^*$  as  $(m, \alpha, \beta)$ , where  $\alpha$  is the index of that symbol within its fan, and  $\beta$  is the index of that fan in  $S^*$ . Duplicate the last element  $(m, \bar{\alpha}, \beta)$  in each of the  $\bar{\beta}$  fans of  $S^*$ .
  - (iv) For each  $1 \leq \alpha \leq \bar{\alpha}$ ,  $1 \leq \beta \leq \bar{\beta}$ , extend the  $\alpha$ th fan of  $S'_\beta$  by duplicating its last element  $(m-1, \alpha, \beta)$ , and by inserting the corresponding symbol  $(m, \alpha, \beta)$  of  $S^*$  between these duplicated appearances of  $(m-1, \alpha, \beta)$ . This process extends the  $(m-1)$ -fans of  $S'_\beta$  into  $m$ -fans and adds a new element after each extended fan.
  - (v) Finally construct the desired sequence  $S(k, m)$  by merging the  $\bar{\beta}$  copies  $S'_\beta$  of  $S'$  with the sequence  $S^*$ . This is done by replacing, for

each  $1 \leq \beta \leq \beta$ , the  $\beta$ th fan of  $S^*$  by the corresponding copy  $S'_\beta$  of  $S'$ , as modified in (iv) above. Note that the duplicated copy of the last element in each fan of  $S^*$  (formed in step (iii) above) appears now after the copy  $S'_\beta$  that replaces this fan. Finally, the symbols of  $S(k, m)$  are renamed so that symbol  $(d, \alpha, \beta)$  becomes  $(d, \gamma)$ , where  $\gamma = \alpha + \bar{\alpha}(\beta - 1)$ .

To establish property (a), note that  $S(k, m)$  consists of

$$\begin{aligned} N_k(m) &= \bar{\beta} \cdot (m-1)C_k(m-1) + \bar{\alpha}C_{k-1}(\bar{\alpha}) \\ &= C_{k-1}(C_k(m-1))C_k(m-1)(m-1) + C_{k-1}(C_k(m-1))C_k(m-1) \\ &= mC_k(m) \end{aligned}$$

different symbols. Property (b) is trivial by induction, because the fans of  $S(k, m)$  are precisely the extended fans of the copies  $S'_\beta$  of  $S'$ .

#### 2.4. Structure and Length Theorems

We now establish several important properties of the sequences  $S(k, m)$ . The proof of these properties will be inductive and there is some interdependence between them, for which reason it is convenient to group these properties into the following single theorem.

**Theorem 2.1.** *For each  $k, m \geq 1$  the sequence  $S = S(k, m)$  satisfies the following properties:*

- (a)  $S$  is a  $DS(N_k(m), 3)$  sequence.
- (b) Each symbol of  $S$  appears in precisely one fan and makes there its first (leftmost) appearance in  $S$ .
- (c) For  $k \geq 2$  and for each  $\gamma \leq \bar{\gamma}$ , the last element  $(m, \gamma)$  of the  $\gamma$ th fan of  $S$  forms the beginning of a contiguous subsequence that is the reverse of that fan:

$$(m, \gamma) (m-1, \gamma) \cdots (2, \gamma) (1, \gamma).$$

(Note that this sequence is the initial portion of a chain of  $S$ .)

- (d) For each chain  $c$  of  $S$ , let  $f$  be the rightmost fan preceding  $c$  or overlapping  $c$  (in the latter case,  $c$  is either a singleton chain contained in the fan  $f$  or the chain which starts with the last element of  $f$ ). Let  $c_1, c_2, \dots, c_t$  be the chains appearing in  $S$  between  $f$  and  $c$ , for some  $t \geq 0$ . Let  $a$  be the first (leftmost) element of  $c$ ; then either this appearance of  $a$  is within  $f$  (in which case there are no intermediate chains  $c_i$ ), or else  $a$  must also appear in one of the preceding chains  $c_i$ .

**Notes.** (1) For each  $\gamma \leq \bar{\gamma}$  and each  $d < m$ , the element  $(d, \gamma)$  in the  $\gamma$ th fan of  $S$  forms a 1-element chain. Note that property (d) is trivially correct for these singleton chains.

- (2) Property (b) implies in particular that  $S$  starts with a fan.

(3) Unless  $c$  is one of the singleton chains mentioned in (1) above, the first chain  $c_1$  in property (d) is the chain mentioned in (c) (whose initial portion is the reverse of the fan  $f$ ). Note that property (d) clearly holds for the case  $c = c_1$ .

(4) While properties (a)-(c) describe the inherent structure of the sequences, property (d) is used later (in Section 3.5.4.3) only to prove some (crucial) properties of the geometric construction.

*Proof.* The proof proceeds by double induction on  $k$  and  $m$ . The base case  $k = 1$  is trivial:  $S(1, m)$  is plainly a  $DS(m, 3)$  sequence, (b) and (d) are trivial, and (c) is vacuous in this case.

The case  $k = 2$  is also easy. Here  $\bar{\gamma} = 2$  and  $S(2, m)$  is obviously a  $DS(2m, 3)$  sequence, so (a) follows. Properties (b), (c), and (d) are also immediate.

Next consider the case  $k > 2, m = 1$ . Here  $S(k, 1) = S(k-1, 2)$  (with its symbols being renamed), so property (a) holds by induction. Property (b) is also trivial because the only change in the fan structure between  $S(k-1, 2)$  and  $S(k, 1)$  is that each fan is split into two subfans. Since each fan is now of size 1, property (c) is trivial too. Finally, since the chain structure in  $S(k, 1)$  is identical to that in  $S(k-1, 2)$ , (d) also follows immediately by induction.

Finally consider the general case  $k > 2, m > 1$ . We first prove property (a). First note that no two adjacent elements of  $S = S(k, m)$  are equal: indeed, by induction hypothesis, no two adjacent elements in either  $S^*$  or in any  $S'_\beta$  are equal; all these sequences have pairwise disjoint sets of symbols (after renaming), and the merging process clearly does not introduce any new duplication of adjacent elements in  $S$  (all duplicated elements get separated by some other symbol).

Now suppose to the contrary that  $S(k, m)$  contains a subsequence of the form  $a \cdots b \cdots a \cdots b \cdots a$ , for some pair of distinct symbols  $a$  and  $b$ . By induction hypothesis, it is clearly impossible for both  $a$  and  $b$  to belong to  $S^*$  or to belong to the same  $S'_\beta$ . Moreover, for each  $\beta_1 < \beta_2$  the entire copy  $S'_{\beta_1}$  of  $S'$  lies in  $S$  to the left of  $S'_{\beta_2}$ ; thus  $a$  and  $b$  cannot belong to different copies  $S'_{\beta_1}, S'_{\beta_2}$  of  $S'$ .

Thus either  $a$  belongs to  $S^*$  and  $b$  to some  $S'_\beta$  or vice versa. In the first case, by the induction hypothesis (b) on  $S^*$ , the first appearance of  $a$  is in the appropriate fan of  $S^*$ . Because of the replacement process, only this first appearance will be inserted into some  $S'_\beta$  (and thus appear in  $S$  between two elements of  $S'_\beta$ ); Therefore, the subsequence  $a \cdots b \cdots a \cdots b \cdots a$  cannot appear in  $S$ . The same argument also rules out the second case in which  $b$  belongs to  $S^*$  and  $a$  to some  $S'_\beta$ .

Thus  $S(k, m)$  is a  $DS(m \cdot \bar{\gamma}, 3)$  sequence and property (a) holds.

Property (b) readily follows from the construction: all the symbols in each  $S'_\beta$  satisfy this property by induction hypothesis. In addition, each symbol of  $S^*$  appears (in  $S^*$ ) in precisely one fan, where this appearance is its leftmost one in  $S^*$ . Since each such first appearance in  $S^*$  is inserted into a fan of some  $S'_\beta$ , which then becomes a fan of  $S$ , property (b) holds also for the symbols of  $S^*$ .

As to property (c), let  $f$  be the  $\gamma$ th fan of  $S$  (where we write  $\gamma = (\alpha, \beta)$ ). Then by construction  $f$  is the concatenation of the  $\alpha$ th fan  $f'$  of  $S'_\beta$  and the corresponding element  $(m, \alpha, \beta)$  of  $S^*$ . By induction hypothesis, the portion of  $S'_\beta$  starting at



the last element of  $f'$  begins with the subsequence  $(m-1, \gamma) \cdots (1, \gamma)$ . The construction then implies that the portion of  $S$  from the last element of  $f$  must begin with the extended subsequence  $(m, \gamma)(m-1, \gamma) \cdots (1, \gamma)$ , thus (c) holds for  $S$ .

To prove (d), note first that no chain of  $S^*$  or of any  $S'_\beta$  was broken by the construction. Actually, it is easily checked that, by the way in which symbols of  $S^*$  and of the  $S'_\beta$ 's are renumbered in  $S$ , the only changes in chains caused by our construction are the extension of chains immediately following fans of the copies  $S'_\beta$  by the corresponding element of  $S^*$  inserted at the end of such a fan, and the “detachment” of each chain  $c$  of  $S^*$  that immediately follows a fan  $f$  of that sequence from  $f$ . (Note that  $c$  itself has not changed in this latter case, and that  $f$  is no longer a fan of  $S$ , but is distributed among the fans of some  $S'_\beta$ .)

Now let  $c$  be a chain of  $S$ , and let  $c_1, \dots, c_t$  be the sequence of chains lying in  $S$  between  $c$  and its preceding fan  $f$ . If  $c$  is not one of the special chains mentioned above, then  $c$  must be a chain of  $S^*$  or of some  $S'_\beta$ . In either case the (nonempty!) collection  $c_1, \dots, c_t$  of chains must contain all chains of  $S^*$  (resp. of  $S'_\beta$ ) lying between  $c$  and the fan immediately preceding  $c$  in  $S^*$  (resp. in  $S'_\beta$ ), where the first such chain may have been extended in  $S$ . Thus in this case (d) follows by induction hypothesis.

Next, suppose  $c$  is an extended chain of some  $S'_\beta$ . But then (d) is trivial, because the (new) first element of  $c$  is also the last element of the preceding fan in  $S$ .

Finally, suppose  $c$  is a chain of  $S^*$  that has been detached from its preceding fan  $f$ . Suppose  $f$  is the  $\beta$ th fan of  $S^*$ . Then in this case the first element  $q$  of  $c$  is equal to the last element of  $f$ , which has been inserted into the *last* fan  $f'$  of  $S'_\beta$ . It thus follows from the construction that the fan of  $S$  immediately preceding  $c$  is  $f'$  extended by  $q$ , and that  $q$  belongs to the first chain of  $S$  following that fan; thus (d) clearly holds for these chains too.

This completes the induction step and thus proves the theorem  $\square$

Next we prove that for each  $k, m \geq 1$  the length  $\sigma(k, m)$  of  $S(k, m)$  satisfies

$$\sigma(k, m) > (km - 2) \cdot C_k(m) + 1$$

(e.g.,  $\sigma(2, m) > (2m - 2) \cdot 2 + 1 = 4m - 3$ , which indeed holds, because  $\sigma(2, m) = 5m - 3$  as is easily checked).

The following recurrence formulas for  $\sigma(k, m)$  follow easily from the inductive construction of  $S(k, m)$ :

$$\sigma(1, m) = m,$$

$$\sigma(2, m) = 5m - 3,$$

$$\sigma(k, 1) = \sigma(k-1, 2), \quad k > 2,$$

$$\sigma(k, m) = \bar{\beta} \cdot \sigma(k, m-1) + \sigma(k-1, \bar{\alpha}) + \bar{\alpha} \cdot \bar{\beta} + \bar{\beta}, \quad k > 2, \quad m > 1.$$

(To see the last equality, note that the first term on the right-hand side is the total length of the  $\bar{\beta}$  copies  $S'_\beta$  of  $S'$ , and that the second term is the length of the sequence  $S^*$ . The third term is due to the fact that for each of the  $\bar{\alpha} \cdot \bar{\beta}$  symbols  $(m, \alpha, \beta)$  of  $S^*$ , its first appearance in  $S^*$  is inserted into the  $\alpha$ th fan of  $S'_\beta$  and causes the last element of that fan to be duplicated. Finally, the fourth term is due to the fact that the last element in each of the  $\bar{\beta}$  fans of  $S^*$  is also duplicated in  $S$ .)

**Theorem 2.2.** *Define  $Z_k(m) = \sigma(k, m) / C_k(m)$ ; then for each  $k, m \geq 1$  we have*

$$Z_k(m) > km - 2 + \frac{1}{C_k(m)}$$

(which clearly implies the lower bound on  $\sigma(k, m)$  stated above).

*Proof.* We proceed by the standard double induction on  $k$  and  $m$ . For  $k = 1$  we have

$$Z_1(m) = m > 1 \cdot m - 2 + \frac{1}{C_1(m)} = m - 1.$$

For  $k = 2$  we have

$$Z_2(m) = \frac{5m-3}{2} = \frac{5}{2}m - \frac{3}{2} > 2 \cdot m - 2 + \frac{1}{C_2(m)} = 2m - \frac{3}{2}.$$

For  $m = 1$ ,  $C_k(1) = 2C_{k-1}(2)$  and thus by induction hypothesis

$$\begin{aligned} Z_k(1) &= \frac{\sigma(k, 1)}{C_k(1)} = \frac{\sigma(k-1, 2)}{2C_{k-1}(2)} = \frac{Z_{k-1}(2)}{2} > \frac{2(k-1) - 2 + (1/C_{k-1}(2))}{2} \\ &= k \cdot 1 - 2 + \frac{1}{C_k(1)}. \end{aligned}$$

Finally, for  $k > 2$  and  $m > 1$  we have

$$\begin{aligned} Z_k(m) &= \frac{\sigma(k, m)}{C_k(m)} = \frac{\bar{\beta} \cdot \sigma(k, m-1) + \sigma(k-1, \bar{\alpha}) + \bar{\alpha} \cdot \bar{\beta} + \bar{\beta}}{\bar{\gamma}} \\ &= \frac{\sigma(k, m-1)}{\bar{\alpha}} + \frac{\sigma(k-1, \bar{\alpha}) + \bar{\beta}}{\bar{\alpha} \cdot \bar{\beta}} + 1 \\ &= Z_k(m-1) + \frac{Z_{k-1}(C_k(m-1)) + 1}{C_k(m-1)} + 1. \end{aligned}$$

Thus, by induction hypothesis

$$\begin{aligned}
 Z_k(m) &> k(m-1) - 2 + \frac{1}{C_k(m-1)} \\
 &\quad + \frac{(k-1)C_k(m-1) - 2 + (1/C_{k-1}(C_k(m-1))) + 1}{C_k(m-1)} + 1 \\
 &= [k(m-1) + (k-1) + 1] - 2 + \frac{1}{C_k(m-1) \cdot C_{k-1}(C_k(m-1))} \\
 &= km - 2 + \frac{1}{C_k(m)}.
 \end{aligned}$$

And this proves the theorem.  $\square$

**Theorem 2.3.** *There exists a set of increasing integers  $\{m_k\}_{k=1}^\infty$  such that*

$$\sigma(k, m_k) = \Omega(N_k(m_k)\alpha(N_k(m_k))).$$

*Proof.* For a given  $k$ , choose  $m_k = C_{k+1}(k-3)$ . Then define

$$n_k = N_k(m_k),$$

and since  $N_k(m) = m \cdot C_k(m)$  we now have

$$n_k = C_{k+1}(k-3) \cdot C_k(C_{k+1}(k-3)) = C_{k+1}(k-2) \leq A_{k+1}(k+1)$$

which implies

$$\alpha(n_k) \leq k+1,$$

and so

$$\begin{aligned}
 \sigma(k, m_k) &\geq kn_k - 2C_k(m_k) \geq (k-2)n_k \\
 &\geq (\alpha(n_k) - 3) \cdot n_k.
 \end{aligned}$$

Thus  $\sigma(k, m_k)$  is of the same order of magnitude as  $n_k \cdot \alpha(n_k)$ .  $\square$

### 3. Conversion of the Sequences $S(k, m)$ to Collections of Segments

#### 3.1. Basic Definitions and Notations

The following definitions refer to a collection  $G = G(k, m)$  of closed straight segments in the plane, none of which is vertical (in fact in our construction all segments in  $G$  have positive slope). We further assume that no endpoint of any segment in  $G$  lies on any other segment in  $G$ . Using the notation of the previous section, we write

$$G(k, m) = \{l_{d,\gamma} \mid 1 \leq d \leq m, 1 \leq \gamma \leq \bar{\gamma}\}$$

so that the segment  $l_{d,\gamma}$  will realize the symbol  $(d, \gamma)$  of  $S(k, m)$  as described below.

For each nonvertical segment its *starting point* is its left endpoint and its *ending point* is its right endpoint.

A point  $P$  is *hidden* by a segment  $l$  if the vertical line passing through  $P$  intersects  $l$  at a point lying below  $P$ . A segment is *hidden* by  $l$  if all its points are hidden by  $l$ .

A point  $P$  is *seen* (below  $G$ ) if it is not hidden by any segment of  $G$ . A segment is *seen* if one of its points is seen.

The *lower envelope*  $Y_G$  of  $G$  is the pointwise minimum of the segments of  $G$  when we view these segments as graphs of partially defined linear functions over the  $x$ -axis.  $Y_G$  is a piecewise linear function consisting of subsegments of the segments in  $G$ ;  $Y_G$  is not necessarily continuous.

$G$  is *contiguous* if the domain of definition of  $Y_G$  is some closed interval. Thus  $G$  is contiguous iff no vertical line can separate  $G$  into two parts without crossing it. Note that if  $G_1$  and  $G_2$  are two contiguous sets of segments that are both crossed by the same vertical line then  $G_1 \cup G_2$  is contiguous as well.

The *lower envelope sequence*  $E_G$  of  $G$  is the sequence of indices of the segments of  $G$  as they appear in its lower envelope from left to right. As noted in [7],  $E_G$  is a  $DS(n, 3)$  sequence, because any appearance of a (not necessarily contiguous) subsequence of the form

$$\langle a \cdots b \cdots a \cdots b \cdots a \rangle$$

in  $E_G$  implies that the two distinct segments  $l_a$  and  $l_b$  cross each other twice; it is also clear that no two adjacent elements of  $E_G$  are equal.

The “transition” points in which the current segment in  $Y_G$  changes are of three possible types:

- (1) Changes caused by the start of a new segment below any other segment. These include the leftmost start of a segment of  $G$  or a start of any segment after a region in which  $Y_G$  is undefined. These points will be denoted as *S-points*.
- (2) Changes caused by the presently lowest segment of  $G$  crossing another segment (which will thus become lowest). These points are denoted as *C-points*.
- (3) Changes caused by the presently lowest segment of  $G$  reaching its ending point when there is no segment below it. These points are denoted as *E-points*.

The set  $\Pi_X(G)$  (where  $X$  is any combination of S, C, and E) is the set of the  $x$ -coordinates of the transition points of type in  $X$ . For example,  $\Pi_{CE}(G)$  is the set of all  $x$ -coordinates of ending and crossing points in  $Y_G$ .

A *graph fan* in  $G(k, m)$  is a subset of  $m$  line segments  $\{l_{d,\gamma}\}_{d=1}^m$  for some  $\gamma$  (these  $m$  line segments are used to represent the  $\gamma$ th fan of  $S(k, m)$ ).

A *tube*  $T = (x_L, x_H, y_L, y_H, \{t_i\}_{i=1}^r)$  is a set of relatively open straight line segments  $t_i$  (with left endpoint  $(x_i^l, y_i^l)$  and right endpoint  $(x_i^h, y_i^h)$ ) of positive slope contained in the rectangle  $[x_L, x_H] \times [y_L, y_H]$  (so  $x_L < x_i^l < x_i^h < x_H$  and  $y_L < y_i^l < y_i^h < y_H$  for each  $1 \leq i \leq r$ ). The rectangle is the tube’s *frame*.

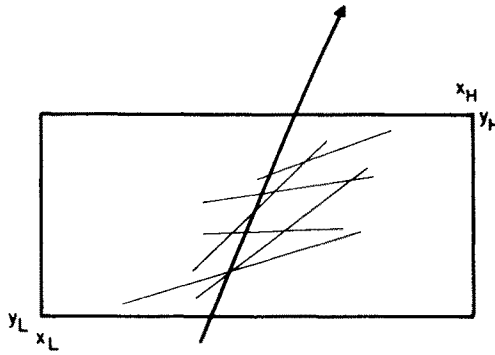


Fig. 3.1

A straight line  $\tilde{l}$  defined by  $y = ax + b$  is said to *pass through* a tube  $T$  (written  $\tilde{l} \in T$ ) if  $\tilde{l}$  crosses each  $t_i \in T$  from below, so that  $y_i^l > a \cdot x_i^l + b$  and  $y_i^h < a \cdot x_i^h + b$  (see Fig. 3.1). The slope of  $\tilde{l}$  is obviously greater than that of any  $t_i$  in  $T$  (and is thus positive). The segments  $(t_i)$  are *not* part of  $G$ , and are used only to control the lines which pass through  $T$  so that they pass in a specific range of positions in the plane.

A segment  $l$  is said to *pass through*  $T$  when the (infinite) line containing  $l$  passes through  $T$  (even though  $l$  itself need not cross any  $t_i$  in  $T$ ). A line  $\tilde{l}$  passing through a tube can cross the outer frame in any two sides excluding combinations which force  $\tilde{l}$  to have negative slope.

A tube  $T$  is *valid* if there exists at least one line  $\tilde{l} \in T$ .

Two lines  $\tilde{l}_1$  and  $\tilde{l}_2$  that pass through a tube  $T$  *cross inside* the tube  $T$  if their crossing point is contained inside the tube's frame but not on its boundary (see Fig. 3.2).

A line  $\tilde{l}_1$  is *above* a line  $\tilde{l}_2$  *relative to* a tube  $T$  if both lines pass through  $T$ ,

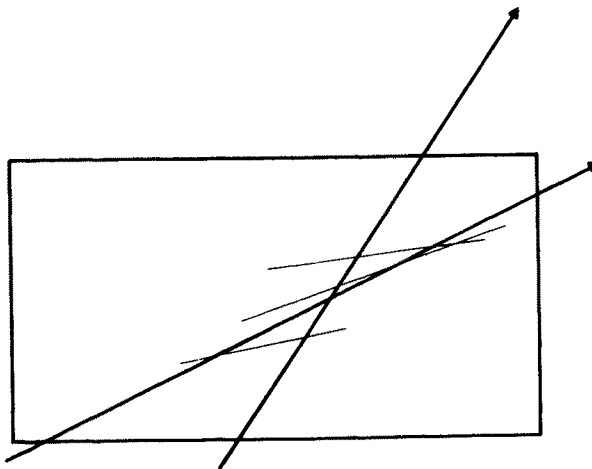


Fig. 3.2

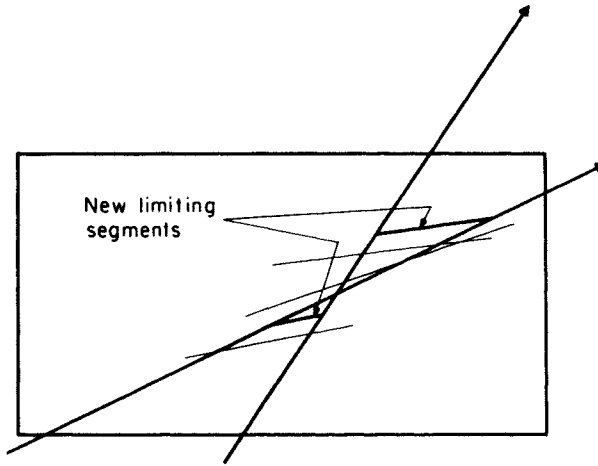


Fig. 3.3

they cross inside it, and  $\tilde{l}_1$ 's slope is greater than that of  $\tilde{l}_2$ . It is obvious that for any  $x > x_H$ ,  $\tilde{l}_1(x) > \tilde{l}_2(x)$  (looking at  $\tilde{l}_i$  as linear functions). We also say that  $\tilde{l}_1$  is above  $\tilde{l}_2$  for  $x > x_0$  if  $\tilde{l}_1(x) > \tilde{l}_2(x)$  for  $x > x_0$  (again, looking at  $\tilde{l}_i$  as linear functions).

If two lines  $\tilde{l}_1$  and  $\tilde{l}_2$  pass through a tube  $T$  and cross inside it, the tube can be limited by  $\tilde{l}_1$  and  $\tilde{l}_2$  to produce a modified tube  $T'$ . The modification is done by adding to the set of segments of  $T$  two new segments (inside the frame boundary) so that any line passing through the modified tube will fall between  $\tilde{l}_1$  and  $\tilde{l}_2$  (see Fig. 3.3).

It can easily be shown that the resulting tube  $T'$  is valid, that every line  $\tilde{l} \in T'$  also passes through the original  $T$  and falls between  $\tilde{l}_1$  and  $\tilde{l}_2$  outside the frame and that the slope of every line  $\tilde{l} \in T'$  is between the slopes of  $\tilde{l}_1$  and  $\tilde{l}_2$ . Note that some lines passing between  $\tilde{l}_1$  and  $\tilde{l}_2$  might not pass through  $T'$ .

The limiting operation is not uniquely defined—any operation which forces lines passing through  $T'$  to be between  $\tilde{l}_1$  and  $\tilde{l}_2$  would be fine.

Note that since the segments  $\{t_i\}$  defining a tube are relatively open, a tube defines an open domain of lines passing through it—in particular there is no line with maximal or minimal slope among these lines, although the appropriate supremum and infimum do exist.

A tube is essentially an open, convex structure of lines—if  $\tilde{l}_1, \tilde{l}_2 \in T$  cross inside  $T$  then any line containing their crossing point and having a slope between those of the two lines also passes through  $T$ . Thus for such a pair  $\tilde{l}_1, \tilde{l}_2$  there is always a line  $\tilde{l} \in T$  which lies between them.

A tube  $T$  is equivalent to a pair of convex polygons between which any line  $\tilde{l} \in T$  must pass. These are respectively the convex hulls of the right and of the left endpoints of the segments  $t_i$  and are thus contained in the frame of  $T$ . If a tube is valid these polygons will be disjoint and arranged so that any line separating them has positive slope (any line passing through the tube can serve as a separator between them).

Adding constraint segments to the tube while keeping it valid (e.g., when limiting a tube) is equivalent to enlarging both  $r$ -gons while keeping them convex and disjoint.

Two tubes  $T_1, T_2$  are *parallel* (written  $T_1 \parallel T_2$ ) if  $T_2$  is a planar translate of  $T_1$  by some vector  $u$ . Thus for every  $\tilde{l}_1 \in T_1$  there is a *matching* parallel line  $\tilde{l}_2 \in T_2$  created by translating  $\tilde{l}_1$  by the same vector  $u$ .

A tube  $T_1$  is *steeper* than  $T_2$  (written  $T_1 > T_2$ ) if for every  $\tilde{l}_1 \in T_1$  and  $\tilde{l}_2 \in T_2$  the slope of  $\tilde{l}_1$  is greater than that of  $\tilde{l}_2$ . The reverse relation  $T_1 < T_2$  is defined in a similar way.

### 3.2. The Induction Base

We will now define a valid constraint  $C(k, m)$  and show that for any  $C(k, m)$  we can construct a matching set of line segments  $G(k, m)$  which satisfies the constraint  $C$ . We will recursively construct  $G(k, m)$  from  $G(k, m-1)$  and  $G(k-1, \bar{\alpha})$ , which have to satisfy certain properties so that they will fit together. This will be achieved by generating induced constraints  $C(k, m-1)$  and  $C(k-1, \bar{\alpha})$  from  $C(k, m)$ , and require  $G(k, m-1)$  and  $G(k-1, \bar{\alpha})$  to satisfy these respective constraints. This is the point in which the combinatorial construction of  $S(k, m)$  differs from the construction of  $G(k, m)$ .

Specifically, a *valid constraint*  $C(k, m) = (\bar{T}, x_L, x_H, y_L, y_H)$  over the  $m$  segments in a single graph fan of  $G(k, m)$  consists of a rectangular *outer frame*  $[x_L, x_H] \times [y_L, y_H]$  and of a set of  $m$  valid tubes  $\bar{T} = \{T_j\}_{j=1}^m$ , which satisfy the following properties:

- V1. The tube frames are ordered “from northwest to southeast” (herein referred to as “NW to SE” order) inside the outer frame so that

$$\begin{aligned} x_L &< x_L(T_1), \\ x_H(T_i) &< x_L(T_{i+1}) \quad \text{for } 1 \leq i \leq m-1, \\ x_H(T_m) &< x_H, \\ y_H &> y_H(T_1), \\ y_L(T_i) &> y_H(T_{i+1}) \quad \text{for } 1 \leq i \leq m-1, \\ y_L(T_m) &> y_L \end{aligned}$$

(thus all the segments defining each of the tubes are contained in the outer frame).

- V2.  $T_j \parallel T_p$  for any  $1 \leq j, p \leq m-1$ .  
V3.  $T_m > T_{m-1}$ .  
V4. There exist lines  $u \in T_{m-1}$  and  $v \in T_m$  which cross each other at a point to the right of  $x_H$ .

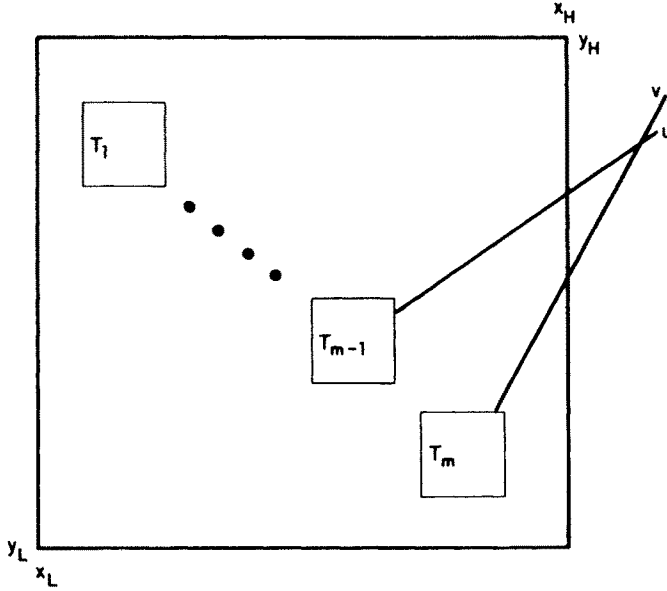


Fig. 3.4

**Notes.** (1) See Fig. 3.4 for an illustration of a constraint.

(2) Conditions V1-V4 are actually independent of  $k$ . The dependence of  $C(k, m)$  on  $k$  is implicit in the recursive construction given below, and is reflected in the inner structure of the tubes  $T_j$ .

**Lemma 3.1.** *Let  $C = C(k, m)$  be a valid constraint and let  $M = \{l_i \in T_i\}_{i=1}^m$  be any set of  $m$  segments such that the starting point of each  $l_i$  is inside the frame of its corresponding tube  $T_i$ . Furthermore, assume that  $\Pi_{CE}(M)$  is outside  $[x_L, x_H]$ . Take the portions of these segments lying inside the  $x$  range  $[x_L, x_H]$  to create  $M^*$ . Then the sequence  $R = E_{M^*}$  (in which each appearance of  $l_i$  is encoded as  $i$ ) is  $\langle 1 \ 2 \ \dots \ m \rangle$ .*

*Proof.* For each  $1 \leq i < j \leq m$  the tube  $T_i$  lies NW to  $T_j$ . Thus  $l_i$  cannot pass below the starting point of  $l_j$  and hide it as this would force  $l_i$  to have a negative slope. Thus  $R$  contains the (not necessarily contiguous) subsequence  $\langle 1 \ 2 \ \dots \ m \rangle$ . Furthermore, if  $R$  contains a subsequence of the form  $\langle x \ \dots \ y \ \dots \ x \rangle$  then there must exist a point in  $\Pi_{CE}(M)$  inside  $[x_L, x_H]$ , contrary to assumption. Thus  $R = \langle 1 \ 2 \ \dots \ m \rangle$ . □

### 3.3. The Induction Scheme

In order to convert the sequence of symbols  $S(k, m)$  into the corresponding collection of segments  $G(k, m)$ , we use a doubly inductive scheme that follows the inductive construction of  $S(k, m)$ . Our construction shows that, given a valid



constraint  $C = C(k, m)$ , there exists a set  $G(k, m)$  of segments satisfying certain properties (P1–P5 below), related to the constraint  $C$ ; the most important of these properties is  $E_{G(k, m)} = S(k, m)$ . Before presenting the formal (and rather complex) construction, we offer an intuitive description of the inductive step for the general case ( $k \geq 3, m > 1$ ); the other base cases are considerably simpler.

Here is a brief summary of the properties that  $G = G(k, m)$  has to satisfy. This set has to consist of  $\bar{\gamma}$  disjoint subsets (fans) of  $m$  segments each (property P2) and each fan has to meet the constraints imposed on  $G$  by an appropriate copy of  $C$ , where these  $\bar{\gamma}$  copies of  $C$  are placed in the plane from NW to SE (property P1). Specifically, for each  $1 \leq d \leq m, 1 \leq \gamma \leq \bar{\gamma}$  the segment  $l_{d, \gamma}$  in  $G$  (the  $d$ th segment in the  $\gamma$ th fan) corresponds to the  $d$ th tube  $T_d$  in the  $\gamma$ th copy of  $C$ , in the sense that the starting point of this segment is inside the tube frame of  $T_d$  (property P2A) and the segment passes through the tube (property P2B) (this last property forces all segments of  $G$  to have positive slope).

The transition points along  $Y_G$ —S-, C-, and E-points—have to satisfy the following properties. The starting points of the segments of  $G$  are all S-points (i.e., each segment starts below all other segments in  $G$ ). (This follows from the properties that each segment starts inside a tube frame, that there is an NW to SE order among the tube frames and among the outer frames of copies of  $C$ , and that all the segments of  $G$  have positive slope.) We furthermore require that all other transition points (in  $\Pi_{CE}(G)$ ) lie outside the  $x$ -ranges of the outer frames of the copies of  $C$  (property P3A).

These properties are used in the inductive merging process described below. Some additional properties (P3B, P4A, and P5) are used to establish correctness of our inductive process. Finally, property P4B states that  $G(k, m)$  is indeed the desired set of segments, i.e.,  $E_{G(k, m)} = S(k, m)$ .

The inductive pattern of our construction for the general case is as follows:

- (1) We construct from  $C(k, m)$  a modified constraint  $C'(k, m - 1)$ , by discarding the last tube  $T_m$  and a few additional modifications.
- (2) On top of this constraint we build inductively a set of segments  $G' = G(k, m - 1)$ . The construction duplicates the original constraint  $C$   $\bar{\alpha}$  times.
- (3) We use these copies of  $C$  to construct a new constraint  $C^* = C^*(k - 1, \bar{\alpha})$ . This  $C^*$  is based on the  $\bar{\alpha}$  copies of the discarded tubes  $T_m$  (one in each of the  $\bar{\alpha}$  copies of  $C$ ).
- (4) We then construct inductively another set of segments  $G^* = G(k - 1, \bar{\alpha})$  matching the constraint  $C^*$ . This construction creates  $\bar{\beta}$  copies of  $C^*$ .
- (5)  $G$  is then constructed as the union of  $G^*$  with  $\bar{\beta}$  translated copies of  $G'$ , each lying within a corresponding copy of the constraint  $C^*$  (each copy of  $C^*$  encompasses a single copy of  $G'$ ).

The intuitive reason why the constraints  $C(k, m)$  are needed is that when we construct  $G^*$  inductively we already have a set  $G'$ , which constrain  $G^*$  by requiring that each of its segments “pierce” an appropriate copy of  $G'$  at exactly one place, i.e., at the end of an appropriate fan. The tubes serve this purpose in a twofold manner; first, they specify the location for the starting points of the segments of  $G^*$  (at which the required piercing has to take place), and, secondly,

they constrain the slopes of the segments of  $G^*$  so as to ensure that none of them appears again after this initial piercing below any of the copies of  $G'$ .

### 3.4. The Conversion Theorem

Having presented this “bird’s view” of our construction, we next proceed to describe it formally.

**Theorem 3.2.** *Given integers  $k, m \geq 1$  and a valid constraint  $C = C(k, m)$ , one can construct a set  $G(k, m)$  of segments so that:*

- P1.  $\gamma = C_k(m)$  disjoint translates of  $C$  are created and placed in the plane so that their outer frames are ordered from NW to SE.
- P2.  $G(k, m)$  consists of  $\bar{\gamma} \cdot m$  segments  $l_{d,\gamma}$ ,  $1 \leq d \leq m$ ,  $1 \leq \gamma \leq \bar{\gamma}$ ; these segments are grouped into  $\bar{\gamma}$  disjoint subsets  $(l_{1,\gamma}, \dots, l_{m,\gamma})_{\gamma=1}^{\bar{\gamma}}$ , each forming a fan of size  $m$ , and satisfy:
  - P2A. The starting point of each  $l_{d,\gamma}$  is inside the corresponding  $d$ th tube  $T_{d,\gamma}$  in the  $\gamma$ th copy of  $C$ .
  - P2B. The segment  $l_{d,\gamma}$  passes through the tube  $T_{d,\gamma}$ .
- P3. Segment properties:
  - P3A. The set  $\Pi_{CE}(G(k, m))$  is disjoint from the  $x$ -range  $[x_L, x_H]$  of the outer frame of any copy of  $C$ .
  - P3B. Any other crossing or ending point of segments in  $G(k, m)$  may have  $x$ -coordinate lying inside an  $x$ -range  $[x_L, x_H]$  of some outer frame, but then the point must lie above this frame (i.e., above  $y_H$ ).
- P4. General properties:
  - P4A.  $G(k, m)$  is contiguous.
  - P4B.  $E_G(k, m) = S(k, m)$ .
- P5. In each portion of  $Y_G$  representing a chain  $c$  of  $S(k, m)$  of more than one element, each segment  $l$  has a slope lower than the segment  $l'$  preceding it in  $c$  and  $l$  crosses  $l'$  from above (at a point in  $\Pi_C(G)$ ). (In other words,  $c$  is represented by a convex portion of  $Y_G$ ; see Fig. 3.6.)

**Notes.** (1) In our inductive construction, we will follow the notations of Section 2 and re-index each segment  $l_{d,\gamma}$  of  $G(k, m)$  as  $l_{d,\alpha,\beta}$ , where  $1 \leq \alpha \leq \bar{\alpha}$ ,  $1 \leq \beta \leq \bar{\beta}$ , and  $\gamma = \alpha + \bar{\alpha}(\beta - 1)$ .

(2) Note that properties P1 and P2 imply that after leaving the outer frame in which it starts, a segment of  $G$  can never enter another outer frame or pass below it.

### 3.5. Proof of Theorem 3.2

3.5.1. *The Case  $k=1$ .* See Fig. 3.5. The sequence  $S(1, m)$  to be converted is

$$\langle (1, 1) (2, 1) \cdots (m, 1) \rangle.$$

In this case  $\bar{\gamma} = 1$  and thus no additional copies of  $C(1, m)$  need be created. Take  $u \in T_{m-1}$  and  $v \in T_m$  as given by V4 and let  $x_0$  be the  $x$ -coordinate of their crossing point. For each  $i < m$  take  $l_{i,1}$  to be a segment contained in the line passing

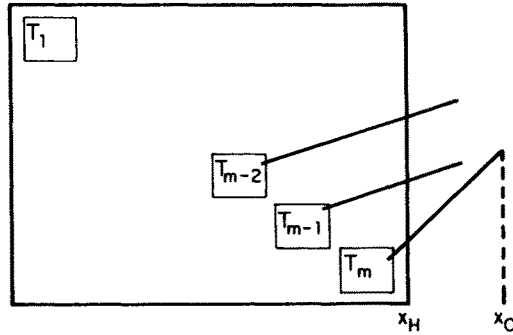


Fig. 3.5

through  $T_i$  which matches the line containing  $u$ , so that  $l_i$  starts inside the tube frame of  $T_i$  and ends at some  $x \in (x_H, x_0)$ . Take  $l_{m,1}$  to be a segment contained in  $v$ , starting inside the frame of  $T_m$  and ending at  $x = x_0$ .

The theorem clearly holds for  $G(1, m) = \{l_{i,1}\}_{i=1}^m$ :

- P1. Trivial ( $\bar{\gamma} = 1$ ).
- P2. Both P2A and P2B are trivial.
- P3. To prove P3A note that the segments do not cross at all since  $\{l_{i,1}\}_{i=1}^{m-1}$  are parallel and end before crossing  $l_{m,1}$ . Their ending points all lie to the right of  $x_H$ . Thus all points in  $\Pi_{CE}(G)$  lie to the right of  $[x_L, x_H]$ . P3B is trivial.
- P4. P4A is trivial, and P4B is a direct consequence of Lemma 3.1.
- P5. Inapplicable—there is no multielement chain.

3.5.2. *The Case  $k=2$ .* As mentioned in Section 2, the case  $k=2$  is treated separately only for exposition sake. It can actually be obtained inductively, as the case of general  $k$  mentioned in Subsections 3.5.3 and 3.5.4 below. However, the explicit constructive given here is somewhat simpler than the inductive construction. This case is illustrated by Fig. 3.7. The sequence  $S(2, m)$  to be converted is

$$\begin{aligned} &\langle (1, 1) (2, 1) \cdots (m-1, 1) (m, 1) (m-1, 1) \cdots (1, 1) \\ &(1, 2) (2, 2) \cdots (m-1, 2) (m, 2) (m-1, 2) \cdots (1, 2) \\ &(2, 2) (3, 2) \cdots (m, 2) \rangle. \end{aligned}$$

Here  $\bar{\gamma} = 2$  so we have to create two copies of  $C(2, m)$ .

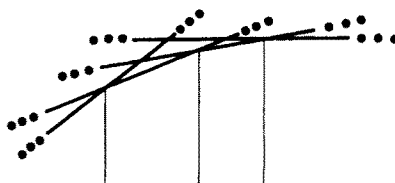


Fig. 3.6

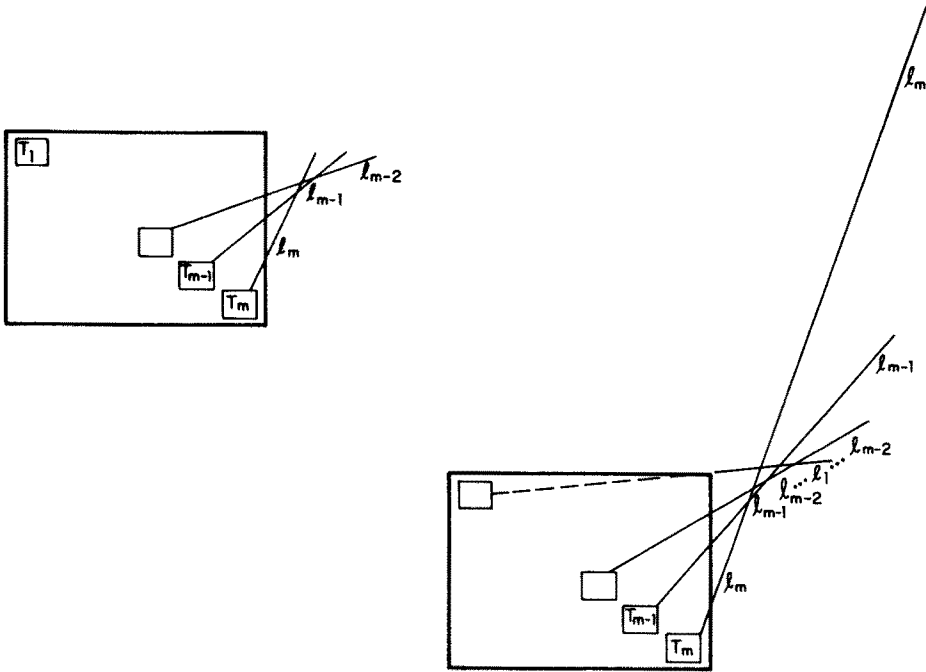


Fig. 3.7

Construct  $G_1$ , a set of  $m$  line segments, as follows: take lines  $u \in T_{m-1}$  and  $v \in T_m$  that cross to the right of  $x_H$  as given by V4. Use them as the lines  $\bar{l}_{m-1,1}, \bar{l}_{m,1}$  containing  $l_{m-1,1}$  and  $l_{m,1}$ , respectively. Continue by adding more lines from  $\bar{l}_{m-2,1}$  backward. Each  $\bar{l}_{i,1}$  will pass through  $T_i$ , will have a slope smaller by some (sufficiently small)  $\epsilon$  than the slope of the previous line ( $\bar{l}_{i+1,1}$ ), and will cross  $\bar{l}_{i+1,1}$  to the right of the previous crossing point (between  $\bar{l}_{i+1,1}$  and  $\bar{l}_{i+2,1}$ ). Define each segment  $l_{i,1}$  to be contained in  $\bar{l}_{i,1}$ , to start inside  $T_i$ , and to end just after the crossing point with  $\bar{l}_{i-1,1}$  (before the next crossing point). Terminate  $l_{i,1}$  to the right at some point after  $l_{2,1}$  is terminated.

The existence of such a set  $G_1$  is guaranteed by V4, and by the properties that  $T_1$  to  $T_{m-1}$  are parallel, and that a tube defines an open domain of lines passing through it. It is easy to check that

$$E_{G_1} = \langle (1, 1) (2, 1) \cdots (m-1, 1) (m, 1) (m-1, 1) \cdots (1, 1) \rangle.$$

In order to create the full  $G$  make a copy  $G_2 = \{l_{1,2}, \dots, l_{m,2}\}$  of  $G_1$  by translating  $G_1$  sufficiently far in an SE direction, and extending the segments so that they realize the last  $m-1$  symbols as well (as shown in Fig. 3.7). It is easy to check that

$$E_{G_2} = \langle (1, 1) (2, 1) \cdots (m-1, 1) (m, 1) (m-1, 1) \cdots (1, 1) \\ (2, 2) (3, 2) \cdots (m, 2) \rangle.$$

Extend the first segment in the first fan  $G_1$  to the right so that it extends just after the starting point of the segment  $l_{1,2}$  (in the second copy). Take  $G = G(2, m)$  to be the union of  $G_1$  (modified this way) and  $G_2$ .

The desired properties hold for  $G$ :

- P1. Trivial ( $\bar{\gamma} = 2$ ).
- P2. Both P2A and P2B are trivial.
- P3. To prove P3A note that the  $x$ -coordinates of all crossing and ending points of the segments of  $G$  lie outside the  $x$ -ranges of both copies of  $C$ , with a single exception: the ending point of  $l_1$  in the first (NW) copy  $G_1$  is above the second outer frame. Since this endpoint is not in  $\Pi_{CE}(G)$  both P3A and P3B follow.
- P4. Each  $G_i$  is contiguous by construction and the extension of  $l_1$  ensures that the union  $G$  is contiguous as well (proving P3A). P4B follows immediately from Lemma 3.1 and the structure of the construction.
- P5. Each  $G_i$  contains one multielement chain, and the construction makes it clear that both chains satisfy P5 (the chains at the end of  $G_2$  are singleton chains).

*3.5.3. The Case  $k > 2$  and  $m = 1$ .* In this case  $S(k, 1) = S(k - 1, 2)$ , so we will use an appropriately constructed  $G(k - 1, 2)$  for our desired  $G(k, 1)$ , making sure that this  $G(k - 1, 2)$  will match the (degenerate) constraint  $C(k, 1)$  and satisfy properties P1–P5 of the theorem. The valid constraint  $C = C(k, 1)$  has a single tube  $T_1$  inside an outer frame (note that in this case the conditions V1–V4 of the constraint are mostly vacuous).

Construct a new constraint  $C' = C(k - 1, 2)$  from  $C$  as follows: create a translate  $C_1$  of  $C$  far enough in the SE direction so that the outer frames of both copies have disjoint  $x$ -ranges and disjoint  $y$ -ranges. Take a line  $w \in T_1(C)$  and its corresponding translate  $w' \in T_1(C_1)$  in  $C_1$ . Limit the tube of  $C$  to lines passing below the line  $w$  and the tube of  $C_1$  to lines passing above the line  $w'$ . Rename the limited tubes  $T'_1 = T_1(C)$  and  $T'_2 = T_1(C_1)$  and use them as the two tubes of  $C'$ . For an outer frame of  $C'$  take any rectangle which contains the outer frames of both  $C$  and  $C_1$ .

$C'$  is clearly valid: V1–V3 are trivially correct and V4 is correct since we can always choose lines  $u \in T'_1$  and  $v \in T'_2$  sufficiently close to  $w$  and  $w'$ , respectively, so that their crossing point lies as far to the right as desired.

By the induction hypothesis there exists a set  $G' = G(k - 1, 2)$  based on  $\bar{\gamma}' = C_{k-1}(2) = C_k(1)/2$  copies of  $C'$  and consisting of  $\bar{\gamma}' \cdot 2 = \bar{\gamma}$  segments satisfying properties P1–P5 for  $(k - 1, 2)$ . We now take  $G = G(k, 1)$  to be this set, renaming each segment  $l_{d,\gamma}$  ( $d = 1, 2$ ) as  $l_{1,2\gamma+d-2}$ , and indeed P1–P5 (for  $(k, 1)$ ) hold for  $G$ :

- P1. As noted above,  $\bar{\gamma} = \bar{\gamma}' \cdot 2$ ; the NW to SE order follows immediately from our construction and from P1 over  $G'$ .
- P2. Both are trivial from P2 over  $G'$ .
- P3. Both P3A and P3B are direct consequence of P3A and P3B over  $G'$ .
- P4. P4A is trivial by induction. P4B is a consequence of P4B over  $G'$ :  $E_{G(k,1)} = E_{G(k-1,2)} = S(k - 1, 2) = S(k, 1)$ .

P5. Trivial from P5 over  $G'$  (the renaming did not change the chain structure of  $G'$ ).

Thus the theorem holds for  $(k, 1)$ .

3.5.4. *The General Case  $k > 2$  and  $m > 1$ .* See Fig. 3.8. We first construct a modified valid constraint  $C' = C(k, m - 1)$  as follows. By V4 there exist  $u \in T_{m-1}$  and  $v \in T_m$  which cross each other at a point right of  $x_H$  (call it  $A$ ). Since  $T_{m-2} \parallel T_{m-1}$  there is a line  $w \in T_{m-2}$  that is a matching translate of  $u$ . This line is above  $u$  and thus  $B$ , its crossing point with  $v$ , is to the right of and above  $A$ .

Take two lines  $u_1, u_2 \in T_{m-1}$  so that the slope of  $u_2$  is slightly greater than the slope of  $u_1$  which is slightly greater than that of  $u$ , and such that  $u_2$  is slightly above  $u_1$  which is slightly above  $u$  (relative to  $T_{m-1}$ ). These lines cross  $v$  at points  $A_1, A_2$ , respectively, which are ordered along  $v$  between  $A$  and  $B$ .  $u_1$  and  $u_2$  cross  $w$  at two respective points,  $B_1, B_2$  ordered along  $w$  in the order  $B, B_2, B_1$ .

Take a line  $w_1 \in T_{m-2}$  (below  $w$  and with slightly lower slope) which crosses  $v$  at  $C$ ,  $u_2$  at  $C_2$ , and  $u_1$  at  $C_1$  (in that order).  $w_1$  should be close enough to  $w$  so that  $C_2$  is to the right of  $B$ .

Take a line  $v_1 \in T_m$  which lies slightly above  $v$  and has slope slightly greater

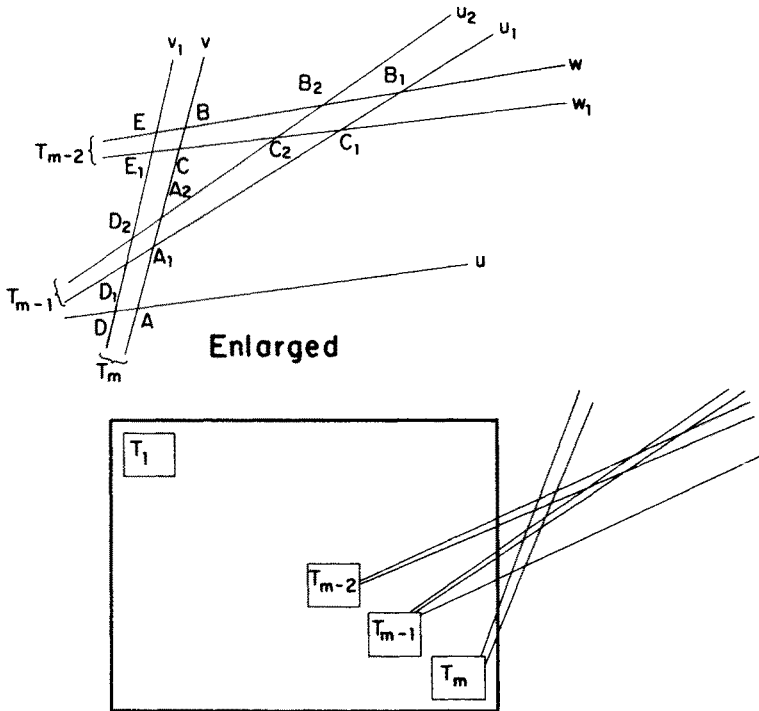


Fig. 3.8

than that of  $v$ . This line should cross the vertical line  $x = x_H$  and then  $u, u_1, u_2, w_1, w$  in that order. Call these crossing points from left to right  $D, D_1, D_2, E_1, E$ . These points are adjacent along their lines to the points  $A, A_1, A_2, C, B$ , respectively.

3.5.4.1. First Induction over  $k, m - 1$ . Define  $C'$  (over  $(k, m - 1)$ ) as follows:

- (1) Its set of tubes  $\bar{T}'$  consists of  $m - 1$  tubes  $T'_i$ . For each  $i < m - 1$ ,  $T'_i$  is the tube  $T_i$  limited by the two lines that are the corresponding translates of  $w$  and  $w_1$  in  $T_{m-2}$ .  $T'_{m-1}$  is defined as  $T_{m-1}$  limited by  $u_1$  and  $u_2$ . Also limit the tube  $T_m$  by  $v$  and  $v_1$  to create  $T'_m$ . This last definition is used only later in the second induction step.
- (2) Any pair of lines passing through  $T'_{m-2}$  and  $T'_{m-1}$  will thus have to cross inside the quadrangle  $B_1B_2C_2C_1$ . Any pair of lines passing through  $T'_{m-1}$  and  $T'_m$  will have to cross inside the quadrangle  $A_1A_2D_2D_1$ .
- (3) The outer frame of  $C'$  has the same  $x_L, y_L$  as that of  $C$ , but its right upper corner has coordinates  $x'_H = x(B)$  and  $y'_H = \max(y(B), y_H)$  (i.e., the outer frame is extended so as to include  $B$  too).

$C'$  is valid since:

- V1. The  $m - 1$  tubes  $T'_i$  are clearly contained in the outer frame of  $C'$  and are ordered there from NW to SE by the original V1 over  $C$ .
- V2. All  $T'_i$  for  $i < m - 1$  are parallel since the original  $T_i$  (for  $i < m - 1$ ) were parallel and matching limitations were added.
- V3.  $T'_{m-1} > T'_{m-2}$  because the slopes of  $u_1, u_2$  are greater than those of  $w, w_1$ .
- V4. There exist  $u' \in T'_{m-2}$  and  $v' \in T'_{m-1}$  crossing to the right of the new  $x'_H$ ; indeed, by construction any pair of lines  $u' \in T'_{m-2}, v' \in T'_{m-1}$  cross inside the region  $B_1B_2C_2C_1$  which lies to the right of  $B$ .

By induction there exists a set of segments  $G' = G(k, m - 1)$  based on  $\bar{\alpha}$  copies of  $C'$  and satisfying properties P1-P5 for  $k$  and  $m - 1$ . By P4 on  $G'$ ,  $E_{G'}$  is the sequence  $S' = S(k, m - 1)$  used in the construction of  $S(k, m)$  given in Section 2. In  $S'$  the fans do not include the “last” segments  $l_m$  and have the form

$$\langle (1, \alpha) (2, \alpha) \cdots (m - 2, \alpha) (m - 1, \alpha) \rangle.$$

The lower envelope sequence associated with the part of  $G'$  inside the  $x$ -range of the outer frame of each copy  $C'_\alpha$  of  $C'$  (for  $1 \leq \alpha \leq \bar{\alpha}$ ) coincides with the corresponding fan in  $S'$  (by P1-P3 and Lemma 3.1).

3.5.4.2. Second Induction over  $k - 1, \bar{\alpha}$ . Next define a new valid constraint  $C^*(k - 1, \bar{\alpha})$  as follows:

- (1) The induction step over  $C'$  has created  $\bar{\alpha}$  copies of the original  $C$  which are ordered in the plane from NW to SE. Let  $C'_\alpha$  denote the  $\alpha$ th copy of  $C'$ . Each of these copies contains a copy of  $T'_m$ . The set of tubes  $\bar{T}^*$  for  $C^*$  then consists of these  $\bar{\alpha}$  copies of  $T'_m$ , each limited by taking any line

$w^* \in T_{\bar{\alpha}}^*$  (the  $\bar{\alpha}$ th copy of  $T'_m$ ), by generating the corresponding translates  $w_\alpha$  of  $w^*$  in each  $T_\alpha^*$ , and by limiting  $T_\alpha^*$  for  $\alpha < \bar{\alpha}$  to admit only lines below  $w_\alpha$  and by limiting  $T_{\bar{\alpha}}^*$  to admit only lines above  $w^*$ .

- (2) The outer frame of  $C^*$  can be taken to be any sufficiently large rectangle  $[x_L, x_H] \times [y_L, y_H]$  that contains all the outer frames of all copies of  $C'$  and all the segments of  $G'$ .

$C^*$  is valid since:

- V1. The  $\bar{\alpha}$  tubes of  $C^*$  are properly ordered as each  $T_\alpha^*$  is contained in the outer frame of the corresponding  $C'_\alpha$  and these outer frames are ordered NW to SE by P1 over  $G'$ . These tubes are contained in the outer frame of  $C^*$  by (2) above.
- V2. All tubes  $T_\alpha^*$  for  $\alpha < \bar{\alpha}$  are parallel by (1) above (because of the matching limitations added to each).
- V3.  $T_{\bar{\alpha}}^* > T_{\bar{\alpha}-1}^*$  because of the  $w^*$ -limitations in (1).
- V4. We can always choose lines  $w' \in T_{\bar{\alpha}-1}^*$ ,  $w'' \in T_{\bar{\alpha}}^*$  sufficiently close to  $w_{\bar{\alpha}-1}$  (as defined in (1) above) and to  $w^*$ , respectively, so that their crossing point lies as far to the right as desired, so V4 follows immediately.

Again by induction hypothesis there exists a set  $G^*$  based on  $\bar{\beta}$  copies of  $C^*$  and consisting of  $\bar{\gamma} = \bar{\alpha} \cdot \bar{\beta}$  segments which satisfy properties P1–P5 for  $k-1$  and  $\bar{\alpha}$ .

**3.5.4.3. The Merging Step.** We now construct the desired set  $G$  as follows. For each  $1 \leq \beta \leq \bar{\beta}$  let  $G'_\beta$  denote the translated copy of  $G'$  inside the  $\beta$ th copy of  $C^*$  created by the construction of  $G^*$ .  $G$  is then defined as the union of  $G^*$  and of all these copies  $G'_\beta$  of  $G'$  for  $1 \leq \beta \leq \bar{\beta}$ . We also number the segments of  $G^*$  and of the  $G'_\beta$ 's using the same renumbering as in Section 2.

The following lemma is a consequence of Theorem 2.1(d), and establishes the crucial property of the merged set  $\mathcal{G}$ , namely, that the interaction of  $G^*$  with each  $G'_\beta$  matches exactly the interaction between  $S^*$  and the corresponding copy  $S'_\beta$  of  $S'$ , as described in Section 2:

**Lemma 3.3.** *Let  $1 \leq \beta \leq \bar{\beta}$ . For each  $1 \leq \alpha \leq \bar{\alpha}$  the segment  $l_\alpha^* = l_{m,\alpha,\beta}$  of  $G^*$  appears below  $Y_{G'_\beta}$  in exactly one interval which is contained in the interior of the range corresponding to the first appearance of  $l_{m-1,\alpha,\beta}$  in  $Y_{G'_\beta}$ .*

*Proof.* First, we claim that the slope of  $l_\alpha^*$  is larger than the slope of any segment in  $G'_\beta$ . Indeed,  $l_\alpha^*$  lies on a line that passes through a limited version of a copy of the original  $m$ th tube  $T_m$  of  $C$ ; similarly, any segment of  $G'_\beta$  lies on a line that passes through a copy of some original tube  $T_d$  for  $d < m$ . Thus V3 and V4 over  $C$  imply that the slope of  $l_\alpha^*$  is larger than that of any segment of  $G'_\beta$ .

Let  $\bar{S}$  be the portion of the sequence  $E_{G'_\beta} = S'_\beta$  realized by the portion of  $Y_{G'_\beta}$  lying to the right of the starting point of  $l_\alpha^*$ . Let  $x_1$  be the  $x$ -coordinate of this



starting point, and let  $x_2$  be the  $x$ -coordinate of the crossing point of  $l_\alpha^*$  and of  $l_{m-1,\alpha,\beta}$  (recall that by our construction this crossing point lies inside the outer frame of  $C'_\alpha$ ).

Associate with  $\bar{S}$  a sequence  $\bar{I}$  of intervals on the  $x$ -axis so that the  $j$ th interval  $I_j$  is the projection of the portion of  $Y_{G'_\beta}$  corresponding to the  $j$ th element of  $\bar{S}$ , with the exception that the first interval  $I_1$  of  $\bar{I}$  starts at  $x_2$ .

For  $x_1 \leq x \leq x_2$ ,  $l_\alpha^*$  lies below  $l_{m-1,\alpha,\beta}$  and  $l_{m-1,\alpha,\beta}$  is the lowest segment of  $G'_\beta$  there. Indeed, each preceding segment  $l$  of  $G'_\beta$  either starts inside a preceding copy  $C'_{\alpha_1}$  of  $C'$  or starts at the present copy  $C'_\alpha$  of  $C'$ . In the first case  $C'_{\alpha_1}$  lies NW to  $C'_\alpha$  and since  $l$  has positive slope it must pass above the outer frame of  $C'_\alpha$  and thus above  $l_{m-1,\alpha,\beta}$ . In the second case  $l$  must also be above  $l_{m-1,\alpha,\beta}$  because their crossing point must be to the right of  $C'_\alpha$  (by P3A over  $G'$ ).

Thus immediately after  $x_2$  the lower envelope  $Y_{G'_\beta \cup \{l_\alpha^*\}}$  is attained by  $l_{m-1,\alpha,\beta}$  and  $l_\alpha^*$  lies above it. The proof now proceeds by induction on the elements in the sequence  $\bar{S}$  (the first of which is clearly  $(m-1, \alpha, \beta)$ ) and shows that for each element  $s_j$  in  $\bar{S}$ , the segment  $l_{s_j}$  of  $G'_\beta$  lies below  $l_\alpha^*$  over the corresponding interval  $I_j$ . The base case  $j=1$  has just been observed, because  $s_1 = (m-1, \alpha, \beta)$  and  $I_1$  starts at  $x_2$ . Suppose the claim holds for all elements in  $\bar{S}$  preceding the  $j$ th one, and let the  $j$ th element of  $\bar{S}$  be the index  $\sigma$  of some segment  $l_\sigma$  of  $G'_\beta$ .  $\bar{S}$  is the tail of  $S' = S(k, m-1)$ , and the proof depends on the place in  $S'$  in which the current appearance of  $\sigma$  takes place.

- (1) If this appearance is at a fan of  $S'$  then it comes after the  $\alpha$ th fan, and by construction and the NW to SE order of the  $C'_\alpha$ -frames,  $l_\sigma$  must appear below  $l_\alpha^*$  over the corresponding interval  $I_j$  (because  $l_\sigma$  starts at a point lying SE to the starting point of  $l_\alpha^*$ , and  $l_\alpha^*$  has a slope greater than that of  $l_\sigma$ ).
- (2) If this appearance of  $\sigma$  is at some later chain of  $S'$  such that  $\sigma$  is not the first symbol of that chain, then let the preceding symbol in that chain be  $\sigma'$ . By induction hypothesis,  $l_{\sigma'}$  lies below  $l_\alpha^*$  over the corresponding interval  $I_{j-1}$ , and since  $l_{\sigma'}$  crosses  $l_\sigma$  from below (by P5 over  $G'$ ) it follows that  $l_\sigma$  must lie below  $l_\alpha^*$  over  $I_j$ .
- (3) Finally, suppose that the present appearance of  $\sigma$  is as the first symbol of some later chain which is not contained in a fan of  $S'$ . Then by Theorem 2.1(d),  $\sigma$  also appears in  $S'$  before its current appearance so that no fan of  $S'$  lies between these two occurrences. In particular,  $\sigma$  has a preceding appearance in  $\bar{S}$  as its  $i$ th element for some  $i < j$ . By induction,  $l_\sigma$  lies below  $l_\alpha^*$  over  $I_i$ , and since  $l_\alpha^*$  has slope larger than that of  $l_\sigma$ , it continues to lie above  $l_\sigma$  also over  $I_j$ .

This completes the inductive proof of the lemma. □

We can now establish P1–P5 for  $G$ :

- P1.  $\bar{\gamma} = \bar{\alpha} \cdot \bar{\beta}$  copies of  $C$  are created in the double induction process. They are ordered NW to SE because each of the copies of  $C$  is ordered NW

to SE inside each copy of  $C^*$ , which are themselves ordered NW to SE by P1 over  $G^*$ .

P2.  $G$  consists of  $\bar{\beta} \cdot (m-1)C_k(m-1) + \bar{\alpha}C_{k-1}(\bar{\alpha}) = m \cdot C_k(m)$  segments (see also Section 2) which are grouped into  $\bar{\alpha} \cdot \bar{\beta} = \bar{\gamma}$  fans.

P2A. By induction the starting point of each of the segments of  $G$  is inside its corresponding tube frame—the first induction substep proves it for segments  $l_{d,\gamma}$ ,  $d < m$ ; the second substep proves it for segments  $l_{m,\gamma}$ . Note that the construction did not change the tube frames in the original  $C(k, m)$ .

P2B. Each segment  $l_{d,\gamma}$  for  $d < m$  lies on a line that passes through the corresponding tube by the first induction substep and each segment  $l_{m,\gamma}$  lies on a line that passes through its corresponding tube by the second induction substep. These tubes are just limited versions of the tubes in  $C(k, m)$ , so all these segments lie on lines that pass through the corresponding original tubes.

P3. By P3 over  $G'_\beta$  all points in  $\Pi_{CE}(G'_\beta)$  lie outside the  $x$  range of any of the  $\bar{\alpha}$  copies of the extended outer frame of  $C'$  and other crossing/ending points of segments of  $G'_\beta$  lie either above these extended frames or outside their  $x$ -ranges. Thus, in particular, P3 also holds for the points in each  $\Pi_{CE}(G'_\beta)$  with respect to the outer frames of the copies of  $C$ .

By P3 over  $G^*$  all points in  $\Pi_{CE}(G^*)$  lie outside the  $\bar{\beta}$  copies of the outer frame of  $C^*$  and any other crossing/ending points of segments of  $G^*$  lie either above these outer frames or outside their  $x$ -ranges. Thus, again, these properties also hold for the outer frames of the copies of  $C$ .

Thus the only type of points that still need to be considered are points in which segments in  $G^*$  cross segments in some  $G'_\beta$  (call these  $X$ -points). However, by construction of  $C^*$  and of  $G'$  there is only one such  $X$ -point inside the outer frame of each  $C'_\alpha$ —namely, the point in which  $l_{m,\alpha,\beta}$  crosses  $l_{m-1,\alpha,\beta}$ . However, this point lies inside the corresponding quadrangle  $A_1A_2D_2D_1$  and thus outside the frame of the copy of  $C(k, m)$  from which this  $C'_\alpha$  is constructed (see Fig. 3.8).

From this point on, the segment  $l_{m,\alpha,\beta}$  of  $G^*$  does not cross  $Y_{G'_\beta}$  again (because of Lemma 3.3) and thus  $l_{m,\alpha,\beta}$  does not pass below  $Y_{G'_\beta}$ . Also, since such an  $l_{m,\alpha,\beta}$  in  $G^*$  lies outside the outer frame of any copy of  $C^*$  other than the  $\beta$ th copy and above all such subsequent copies,  $l_{m,\alpha,\beta}$  cannot cross segments from any other  $G'_{\beta_1}$  (each contained inside one such frame).

P4. General properties:

P4A.  $G$  is contiguous since  $G^*$  is contiguous (by P4A over  $G^*$ ), each  $G'_\beta$  is contiguous (by P4A over  $G'$ ), and each  $G'_\beta$  has points common to  $G^*$ , e.g., the crossing point of  $l_{m,\alpha,\beta}$  with  $l_{m-1,\alpha,\beta}$  for any  $\alpha$ .

P4B.  $E_{G(k,m)} = S(k, m)$ :  $E_{G'} = S'$  by P4B over  $G'$ , and  $E_{G^*} = S^*$  by P4B over  $G^*$ . By Lemma 3.3 and the construction of  $G^*$ , for each  $1 \leq \beta \leq \bar{\beta}$  the parts of  $G^*$  seen inside the  $x$  range of  $Y_{G'_\beta}$  are only the segments  $l_{m,\alpha,\beta}$  (for  $1 \leq \alpha \leq \bar{\alpha}$ ), each making a single appearance over an interval “inserted into” the interior of the range of the initial

appearance of its corresponding segment  $l_{m-1,\alpha,\beta}$  (in the fan containing  $l_{m-1,\alpha,\beta}$ ). Thus, it is easily checked that the lower envelope  $Y_{G(k,m)}$  is such that:

- (1) Outside the outer frames of the copies of  $C^*$  (more precisely, outside the  $x$ -ranges of the copies  $G'_\beta$  of  $G'$ )  $Y_G$  coincides with  $Y_{G^*}$ .
- (2) Within the  $x$ -range of a copy  $G'_\beta$  of  $G'$ , Lemma 3.3 implies that  $Y_G$  is equal to  $Y_{G'_\beta}$ , with each of the additional segments  $l_{m,\alpha,\beta}$  making a single appearance in the “end” of the corresponding fan, thereby splitting the range of the first appearance of the corresponding  $l_{m-1,\alpha,\beta}$  into two subintervals. It is also easily checked, using P3 over  $G^*$ , that the segment of  $G^*$  seen immediately to the right of the  $x$ -range of  $G'_\beta$  is the segment  $l_{m,\bar{\alpha},\beta}$ . Similarly, the segment of  $G^*$  seen immediately to the left of the  $x$ -range of  $G'_\beta$  is the segment immediately preceding (in  $Y_{G^*}$ ) the  $\beta$ th fan of  $G^*$  (in particular, no other segment of  $G^*$  has a lower  $y$  value for the  $x$ -coordinate of the left edge of the outer frame of the  $\beta$ th copy  $C_\beta^*$  of  $C^*$ , and  $G'_\beta$  is still not defined for this  $x$ -value).

It therefore follows that  $E_{G(k,m)}$  is constructed from  $E_{G'} = S'$  and  $E_{G^*} = S^*$  in exactly the same way in which  $S(k, m)$  is constructed from these subsequences, as described in Section 2. Hence  $E_{G(k,m)} = S(k, m)$ .

- P5. In the first chain after each fan of  $G$  the slopes of the segments from  $l_{m-1,\alpha,\beta}$  to  $l_{1,\alpha,\beta}$  are arranged in decreasing order (by P5 over  $G'$ ).  $l_{m,\alpha,\beta}$  has a slope larger than that of  $l_{m-1,\alpha,\beta}$  because of P2B over  $G^*$  (each segment lies on a line passing through its tube) and V3 over  $C$  ( $T_m > T_{m-1}$ ).

$Y_G$  can be partitioned into (modified) copies  $Y_{G'_\beta}$  of  $Y_{G'}$ , and chains of  $G^*$ . By Lemma 3.3, the only change in the chain structure of the various  $G'_\beta$  is the extension of each chain immediately succeeding each of their fans. The preceding argument establishes P5 for such chains, and the induction hypothesis establishes P5 for each of the remaining chains of  $G$ .

See Fig. 3.9 for details. This completes the proof of the main theorem.  $\square$

### 3.6. Actual Graph Construction

In order to create  $G(k, m)$  we actually need to start the above inductive construction with some initial valid base constraint of size  $m$ . It is quite easy to construct such an initial constraint and we leave details to the reader (Fig. 3.4 illustrates such a possible initial constraint). By the proof there is a  $G(k, m)$  built on top of copies of this valid constraint, which satisfies, among other properties, the desired equality  $E_{G(k,m)} = S(k, m)$ .

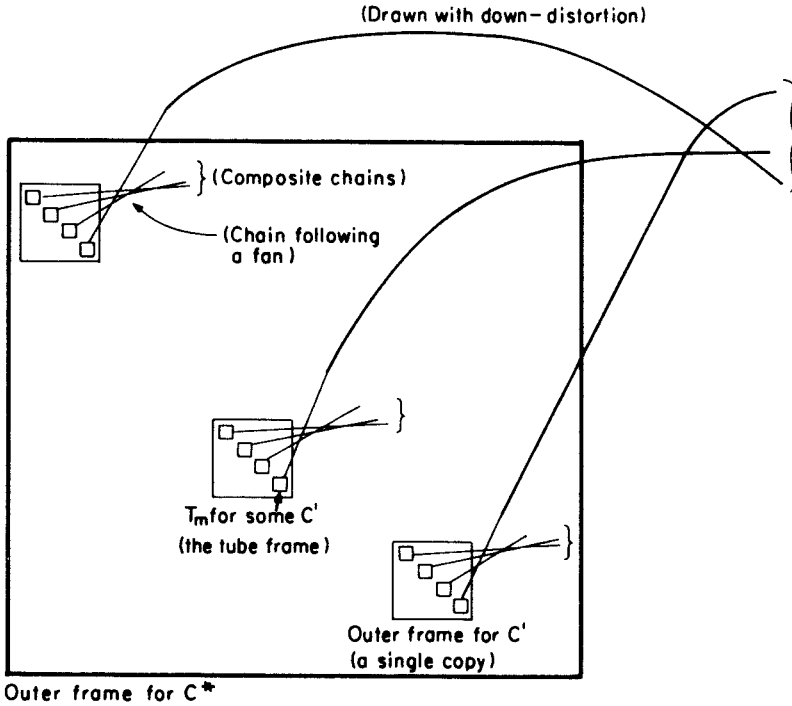


Fig. 3.9

### 3.7. Anchored Graph Construction

It is rather easy to modify the construction so that the starting points of the segments all lie on the line  $l(x) = -x$  (or, for that matter, an arbitrary line  $l$  with negative slope). We will call this construction *anchored realization* of the sequence using the *anchor*  $l$ .

To do this, the definition of a valid constraint has to be modified so that  $l$  will have to cross the tubes of  $C(k, m)$ . The segments will be selected so as to start on the anchor  $l$ . Since the case  $k=2$  is not essential (and was only included for exposition), we will just have to modify the case  $m=1$  so that the two copies of  $C$  created for the inductive generation of  $G(k-1, 2)$  will be placed in the plane so that  $l$  will pass in the tubes in both constraints.

This construction will also work for any segment or ray (half-line). After creating the anchored realization whose anchor is the line  $l'$  containing  $l$ , use a linear mapping to "pack"  $G(k, m)$  into the  $x$ -range of  $l$ .

Once this is done, we can construct a realization of  $S(k, m)$  whose anchor is an arbitrary curve  $y = f(x)$  provided that the curve can be enclosed in the strip between two parallel nonvertical lines. This is done by using a linear transformation to "squash" the curve so that it is contained in an  $\varepsilon$ -wide strip around a segment, ray, or line  $l$ ; building on top of  $l$  an anchored realization, and then

performing the reverse linear transformation. The construction is feasible if  $\varepsilon$  is small enough compared with the size of the tubes.

#### 4. Conclusions

Our result solves an open problem in the theory of Davenport–Schinzel sequences, and also sets a tight bound on the complexity of several problems in computational geometry whose solutions involve calculations of the lower envelope of collections of segments. Some of these problems are:

- (1) Given  $n$  segments in the plane, let  $C$  be the unbounded component (or, more generally, any single connected component) of the complement of the union of the segments (i.e.,  $C$  consists of points reachable from infinity (or from any fixed given point) without having to cross any of the given segments). It has been shown recently in [11] that the (polygonal) boundary of  $C$  consists of at most  $O(n\alpha(n))$  segments, and our results clearly imply that this bound is tight in the worst case.
- (2) Given a polyhedral terrain  $\sigma$  having  $n$  faces in 3-space and a point  $a$  lying above it, we wish to calculate the “horizon” as seen from  $a$ , i.e., the lower envelope of all the rays emerging from  $a$  and not penetrating into  $\sigma$ . It has been recently shown in [4] that the combinatorial complexity of such a horizon is  $O(n\alpha(n))$ , and our construction can be easily adapted to yield a terrain  $\sigma$  for which the horizon as seen from some given point has  $\Omega(n\alpha(n))$  complexity, proving that the above bound is tight in the worst case.
- (3) Given  $m$  bivariate piecewise-linear functions having overall  $n$  (triangular) faces, their pointwise minimum is also a piecewise-linear function. Recently it was shown in [16] that the number of faces in this minimum is at most  $O(n^2\alpha(n))$ . Our construction can be easily modified to yield such a collection of piecewise-linear functions whose lower envelope has  $\Omega(n^2\alpha(n))$  faces. Indeed, take our construction in the  $x$ - $z$  plane and sweep each segment in the  $y$ -direction to obtain  $n$  rectangular faces; add  $n$  sharp parallel wedges pointing downward, whose bottom edges are all parallel to the  $x$ -axis and are at the same height, such that each of them cuts the entire collection of the first kind of faces. It is then easy to check that the lower envelope of the resulting collection consists of  $\Omega(n^2\alpha(n))$  subfaces. This problem arises in the design of efficient algorithms for planning a purely translational collision-free motion of a convex polyhedral object amidst polyhedral obstacles.
- (4) Let  $T$  be a set of  $n$  triangles in  $E^2$ ; call a line  $t$  a *transversal* of  $T$  if it intersects each triangle of  $T$ . Let the *stabbing region*  $S(T)$  of  $T$  be the set of (dual) points  $p = (\pi_1, \pi_2)$  such that the lines

$$D(p) : y = \pi_1 x - \pi_2$$

are transversals of  $T$ . It is known that the stabbing region for a single

triangle is the ‘‘corridor’’ bounded between the upper envelope and the lower envelope of the three lines that are dual to the vertices of the triangle. It follows that  $S(T)$  is the region lying below the pointwise minimum of these individual upper envelopes and above the pointwise maximum of the lower envelopes. This easily implies that the combinatorial complexity of the stabbing region is  $O(n\alpha(n))$ , and our construction can be used to construct a set of  $n$  triangles whose stabbing region has complexity  $\Omega(n\alpha(n))$ , showing that this bound is tight in the worst case.

There are many problems related to our result that are still open. For example, can one realize nonlinear  $DS(n, 3)$  sequences by collections of cubic polynomials? As mentioned above, Sharir [14] has recently obtained an improved lower bound of the form  $\lambda_{2s+1}(n) = \Omega(n\alpha^s(n))$  for each  $s \geq 1$ . What is the simplest type of function, each pair of which intersect in at most  $2s + 1$  points, which can be used to realize this lower bound?

Also, many recent algorithms in computational geometry and motion planning involve calculation of lower envelopes of collections of functions, and study these envelopes using Davenport–Schinzel sequences (see [3], [9], and [10], for example). However, the functions that arise in these problems are generally not piecewise linear. It is therefore of interest to attempt to extend our results to realizations of nonlinear Davenport–Schinzel sequences using the specific types of functions arising in these problems.

Another open problem is related to the results in [11]. Given two simple polygons  $P$  and  $Q$ , having respectively  $m$  and  $n$  edges, let  $\gamma$  denote the boundary of the unbounded connected component of the complement of the difference  $P - Q = \{x - y : x \in P, y \in Q\}$ . Calculation of  $\gamma$  is used in [11] for planning a purely translational motion of  $Q$  which separates it from  $P$ . It is shown in [11] that  $\gamma$  is a polygonal curve consisting of  $O(mn\alpha(mn))$  segments. An interesting open problem is whether our construction can be modified to show that this bound is tight in the worst case.

### Acknowledgment

The authors wish to thank Herbert Edelsbrunner for a careful reading of the paper and for many valuable comments that helped us improve its presentation.

### Appendix 1. Some Values of $C_k(m)$ and $A_k(m)$

The functions  $\{C_k: \mathbb{N}_0 \rightarrow \mathbb{N}\}_{k=1}^\infty$  are defined as

$$\begin{aligned} C_1(m) &= 1, & m \geq 0, \\ C_k(0) &= 2, & k \geq 2, \\ C_k(m) &= C_k(m-1) \cdot C_{k-1}(C_k(m-1)), & k \geq 2, \quad m \geq 1, \end{aligned}$$

and a few initial values of these functions are given below. In the preceding

sections, the values of  $C_k(m)$  for  $m = 0$  are usually ignored, and the (equivalent) definition

$$C_k(1) = 2 \cdot C_{k-1}(2), \quad k \geq 2,$$

is used instead. The following three shorthand notations are also used:

$$\begin{aligned} \bar{\alpha} &= C_k(m-1), \\ \bar{\beta} &= C_{k-1}(C_k(m-1)), \\ \bar{\gamma} &= C_k(m) = \bar{\alpha} \cdot \bar{\beta} \quad (\text{by definition}). \end{aligned}$$

		$C_k(m)$							
		$m$							
$k$	0	1	2	3	4	5	6	7	
1	1	1	1	1	1	1	1	1	
2	2	2	2	2	2	2	2	2	
3	2	4	8	16	32	64	128	256	
4	2	16	$2^{21}$	$2^{2^{21}+22}$					
5	2	$2^{22}$							
6	2								
7	2								

The functions  $\{A_k: \mathbb{N} \rightarrow \mathbb{N}\}_{k=1}^\infty$  are defined as

$$\begin{aligned} A_1(m) &= 2m, \quad m \geq 1, \\ A_k(1) &= 2, \quad k \geq 1, \\ A_k(m) &= A_{k-1}(A_k(m-1)), \quad k \geq 2, \quad m \geq 2, \end{aligned}$$

and a few of their initial values are given below:

		$A_k(m)$						
		$m$						
$k$	1	2	3	4	5	6	7	
1	2	4	6	8	10	12	14	
2	2	4	8	16	32	64	128	
3	2	4	16	$2^{16}$	<i>tower(5)</i>	<i>tower(6)</i>		
4	2	4	$2^{16}$	<i>tower(2<sup>16</sup>)</i>				
5	2							

where *tower*( $n$ ) is a power tower of the form  $2^{2^{\dots^2}}$  with  $n$  two's.

## Appendix 2. Some Examples of the Sequences $S(k, m)$

Following are values for a few small sequences. The fan elements are shown in boldface. The sequences for  $k = 1$  are

$$S(1, 1) = (\mathbf{1}, \mathbf{1}),$$

$$S(1, 2) = (\mathbf{1}, \mathbf{1})(\mathbf{2}, \mathbf{1}),$$

$$S(1, 3) = (\mathbf{1}, \mathbf{1})(\mathbf{2}, \mathbf{1})(\mathbf{3}, \mathbf{1}),$$

and so on. The entire sequence consists of just a single fan. The sequences for  $k = 2$  are

$$S(2, 1) = (\mathbf{1}, \mathbf{1}) (\mathbf{1}, \mathbf{2}),$$

$$S(2, 2) = (\mathbf{1}, \mathbf{1})(\mathbf{2}, \mathbf{1})(\mathbf{1}, \mathbf{1}) (\mathbf{1}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{1}, \mathbf{2})(\mathbf{2}, \mathbf{2}),$$

$$S(2, 3) = (\mathbf{1}, \mathbf{1})(\mathbf{2}, \mathbf{1})(\mathbf{3}, \mathbf{1})(\mathbf{2}, \mathbf{1})(\mathbf{1}, \mathbf{1})$$

$$(\mathbf{1}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{3}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{1}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{3}, \mathbf{2}),$$

$$S(2, 4) = (\mathbf{1}, \mathbf{1})(\mathbf{2}, \mathbf{1})(\mathbf{3}, \mathbf{1})(\mathbf{4}, \mathbf{1})(\mathbf{3}, \mathbf{1})(\mathbf{2}, \mathbf{1})(\mathbf{1}, \mathbf{1})$$

$$(\mathbf{1}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{3}, \mathbf{2})(\mathbf{4}, \mathbf{2})(\mathbf{3}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{1}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{3}, \mathbf{2})(\mathbf{4}, \mathbf{2}),$$

and so on. For  $k = 3$  it will only be possible to show the first two sequences:

$$S(3, 1) = (\mathbf{1}, \mathbf{1}) (\mathbf{1}, \mathbf{2}) (\mathbf{1}, \mathbf{1}) (\mathbf{1}, \mathbf{3}) (\mathbf{1}, \mathbf{4}) (\mathbf{1}, \mathbf{3})(\mathbf{1}, \mathbf{4}),$$

(this is just  $S(2, 2)$  with its symbols renamed).  $S(3, 2)$  is constructed from  $\bar{\beta} = 2$  copies of  $S' = S(3, 1)$  (over disjoint alphabets) which are merged with  $S^* = S(2, 4)$ ; the sequence has eight fans. After symbol renaming, the final result is

$$S(3, 2) = (\mathbf{1}, \mathbf{1})(\mathbf{2}, \mathbf{1})(\mathbf{1}, \mathbf{1}) (\mathbf{1}, \mathbf{2})(\mathbf{2}, \mathbf{2})(\mathbf{1}, \mathbf{2}) (\mathbf{1}, \mathbf{1})$$

$$(\mathbf{1}, \mathbf{3})(\mathbf{2}, \mathbf{3})(\mathbf{1}, \mathbf{3}) (\mathbf{1}, \mathbf{4})(\mathbf{2}, \mathbf{4})(\mathbf{1}, \mathbf{4}) (\mathbf{1}, \mathbf{3})(\mathbf{1}, \mathbf{4}) (\mathbf{2}, \mathbf{4})(\mathbf{2}, \mathbf{3})(\mathbf{2}, \mathbf{2})(\mathbf{2}, \mathbf{1})$$

$$(\mathbf{1}, \mathbf{5})(\mathbf{2}, \mathbf{5})(\mathbf{1}, \mathbf{5}) (\mathbf{1}, \mathbf{6})(\mathbf{2}, \mathbf{6})(\mathbf{1}, \mathbf{6}) (\mathbf{1}, \mathbf{5})$$

$$(\mathbf{1}, \mathbf{7})(\mathbf{2}, \mathbf{7})(\mathbf{1}, \mathbf{7}) (\mathbf{1}, \mathbf{8})(\mathbf{2}, \mathbf{8})(\mathbf{1}, \mathbf{8}) (\mathbf{1}, \mathbf{7})(\mathbf{1}, \mathbf{8}) (\mathbf{2}, \mathbf{8})(\mathbf{2}, \mathbf{7})(\mathbf{2}, \mathbf{6})(\mathbf{2}, \mathbf{5})$$

$$(\mathbf{2}, \mathbf{6})(\mathbf{2}, \mathbf{7})(\mathbf{2}, \mathbf{8})$$

The next sequences are too complex to be shown.

## References

1. W. Ackermann, Zum Hilbertschen Aufbau der reellen Zahlen, *Math. Ann* **99** (1928), 118–133.
2. M. Atallah, Dynamic computational geometry, *Proceedings of the 24th Symposium on Foundations of Computer Science*, 92–99, 1983. (Also in *Comput. Math. Appl.* **11** (1985), 1171–1181.)
3. A. Baltan and M. Sharir, On Shortest Paths Between Two Convex Polyhedra, Technical Report No. 180, Computer Science Department, New York University, 1985. (To appear in *J. Assoc. Comput. Mach.*)



4. R. Cole and M. Sharir, Visibility of a Polyhedral Surface from a Point, Technical Report No. 266, Computer Science Department, New York University, 1986. (To appear in *J. Symbolic Comput.*)
5. H. Davenport, A combinatorial problem connected with differential equations, II, *Acta Arith.* **17** (1971), 363-372.
6. H. Davenport and A. Schinzel, A combinatorial problem connected with differential equations, *Amer. J. Math.* **87** (1965), 684-694.
7. S. Hart and M. Sharir, Nonlinearity of Davenport-Schinzel sequences and of generalized path compression schemes, *Combinatorica* **6** (1986), 151-177.
8. P. Komjath, private communication.
9. D. Leven and M. Sharir, On the Number of Critical Free Contacts of a Convex Polygonal Object Moving in Two-Dimensional Polygonal Space, Technical Report No. 187, Computer Science Department, New York University, 1985. (Also in *Discrete Comput. Geom.* **2** (1987), 255-270.)
10. C. Ó'Dúnlaing, M. Sharir, and C. K. Yap, Generalized Voronoi diagrams for a ladder, II: Efficient construction of the diagram, *Algorithmica* **2** (1987), 27-59.
11. R. Pollack, M. Sharir, and S. Sifrony, Separating Two Simple Polygons by a Sequence of Translations, Technical Report No. 215, Computer Science Department, New York University, 1986. (To appear in *Discrete Comput. Geom.*.)
12. D. P. Roselle and R. G. Stanton, Some properties of Davenport-Schinzel sequences, *Acta Arith.* **17** (1971), 355-362.
13. M. Sharir, Almost Linear Upper Bounds on the Length of General Davenport-Schinzel Sequences, *Combinatorica* **7** (1987), 131-143.
14. M. Sharir, Improved Lower Bound on the Length of Davenport-Schinzel Sequences, Technical Report No. 204, Computer Science Department, New York University, 1986. (To appear in *Combinatorica.*)
15. E. Szemerédi, On a problem of Davenport and Schinzel, *Acta Arith.* **25** (1974), 213-224.
16. J. Pach and M. Sharir, The Upper Envelope of Piecewise Linear Functions and the Boundary of a Region Enclosed by Convex Plates: Combinatorial Analysis, Technical Report No. 279, Computer Science Department, New York University, 1987.

Received May 19, 1986, and in revised form June 23, 1987.