PROVING LOWER BOUNDS
FOR LINEAR DECISION TREES

Marc Snir

New York University
Department of Computer Science
Courant Institute of Mathematical Sciences
New York, N. Y. 10012

1. INTRODUCTION

The worst case behavior of decision algorithms using comparisons between inputs has been extensively studied for various sorting-type problems. Two principal methods have been used: the information theoretic method, and the adversary method. When the inputs are numbers, one can consider the use of more general comparisons. Many authors have tried to extend the lower bounds known for simple comparisons to decision trees using linear comparisons [1,2,10-14]. Whereas information theoretic arguments carry through ad verbatim, the generalization of adversary arguments is not done so easily (see [12] for example).

Decision trees using linear comparisons are a natural computation model for problems which are defined by linear inequalities. This includes linear optimization problems such as minimal spanning trees for graphs, shortest path problems, minimal assignment problems, and many others.

It is therefore interesting to develop new methods for proving lower bounds on the complexity of linear decision trees. This has been done to some extent in an ad hoc (and often cumbersome) manner by authors interested in specific problems. It is our hope that by focusing on general methods rather than on specific results, a better understanding of the geometry underlying linear decision problems can be achieved.

2. PRELIMINARIES

We recall the following terminology from convex geometry (see for example [3]): A set $U \subseteq \mathbb{R}^n$ is *affine* if it is obtained by translating a linear subspace; $\dim U$, the *dimension* of U, is defined to be the dimension of that subspace. A *hyperplane* H in $\mathbb{R}^n$ is an affine set of dimension $n-1$, that is, the set of solutions to a nontrivial linear equation $\sum_{i=1}^{n} a_i x_i = b$. The hyperplane H cuts $\mathbb{R}^n$ into two (closed) halfspaces defined by $\Sigma a_i x_i \geq b$, and $\Sigma a_i x_i \leq b$. A *convex polyhedral set* is a set obtained from the intersection of finitely many half spaces. If C is obtained by intersecting the half spaces defined by the hyperplanes $\{H_i : i \in I\}$ then for each subset $I' \subseteq I$ the set $C \cap \left(\bigcap_{i \in I'} H_i\right)$ is a *face* of C.

The following definitions will be used:

A set U is *polyhedral* if it is the union of finitely many convex polyhedral sets; dim U, the dimension of U, is defined to be the least k such that U is contained in the union of finitely many affine sets of dimension k.

A family $\{D_i\}$ of polyhedral sets forms a *polyhedral partition* of $\mathbb{R}^n$ if the sets D_i have disjoint interiors and cover $\mathbb{R}^n$. The partition $\{E_i\}$ is a *refinement* of the partition $\{D_j\}$ if each set E_i is contained in a set D_j.

A *linear decision problem* in $\mathbb{R}^n$ is given by a polyhedral partition $\{D_i\}$ of $\mathbb{R}^n$. A decision procedure for that problem consists of an algorithm that computes for each input $\bar{x} = \langle x_1, \ldots, x_n \rangle$ the index of a set D_i containing it. Quite often, such a procedure can be represented by a *linear decision tree* T: T is a labeled binary tree. Each internal node of T is labeled with a comparison $f(\bar{x}) : 0$, where $f(\bar{x}) = \Sigma a_i x_i + b$, and the two edges leaving it are labeled with the mutually exclusive outcomes $f(\bar{x}) > 0$, $f(\bar{x}) \leq 0$.

The tree T represents a procedure for testing inputs from $\mathbb{R}^n$, doing linear comparisons. We perform the test at the root, choose according to the result one of the branches, and iterate until a leaf is reached. Thus an input $\bar{x}$ *reaches* leaf v if it fulfills all the inequalities labeling the path to v. T *solves* the problem defined by the partition $\{D_i\}$ if each leaf v of T can be associated with a set $D_{i(v)}$ such that $\bar{x}$ reaches v iff $\bar{x} \in D_{i(v)}$.

Let I(v) be the set of inputs reaching the leaf v in T. With T is associated a polyhedral partition $P(T) = \{\overline{I(v)}: v \text{ is a leaf of } T\}$ of $\mathbb{R}^n$. T solves the problem defined by $\{D_i\}$ iff P(T) refines $\{D_i\}$.

Thus, solving a linear decision problem has the geometrical meaning of building a refinement to given polyhedral partition of $\mathbb{R}^n$ by a succession of binary splittings.

The *linear complexity* of a problem P is equal to the minimal depth of a linear decision tree solving P.

Two other models of decision trees are frequently encountered in the literature: one can consider *ternary trees* where each comparison has three outcomes, $<$, $=$ and $>$; one can also consider binary decision trees where each comparison has two outcomes $<$ and $>$. For the sake of completeness we add a third model, where each comparison has two outcomes $\leq$ and $\geq$ (an input reaches more than one leaf if equality obtains). In all cases the relation "tree T solves problem P" is defined as before: T solves P if for each leaf v of T the set I(v) of inputs reaching v is contained in one of the sets defining P. It is easy to see that a decision tree of one type solving a problem P can be transformed into a decision tree of any other type, with the same depth, solving problem P. Also, we can delete from a partition $\{D_i\}$ any set with empty interior, without changing the complexity of the problem.

3. THE INFORMATION THEORETIC ARGUMENT AND ITS EXTENSIONS

The most general argument used to prove lower bounds on decision trees is the information theoretic one: if the problem P is defined by m sets then any tree solving P has at least m leaves, and therefore, has at least depth log m. It is sometimes possible to extend this argument to problems with few outcomes, by proving that one cannot solve that problem without solving some refinement of it, which has many outcomes. We introduce the following definitions:

A set $U \subseteq \mathbb{R}^n$ is *weakly connected* if there exists a set V such that $\dim V < n-1$, and $U \setminus V$ is disconnected; it is *strongly connected* otherwise.

Any convex set with nonempty interior is strongly connected. Any open polyhedral set in $\mathbb{R}^n$ is equal to the finite union of maximal strongly connected subsets. This decomposition into *strong components* is unique. We have:

LEMMA 3.1. *Let* $\{D_i\}$ *be a polyhedral partition of* $\mathbb{R}^n$*, and let for each* i, D_{ij} , $j = 1,\dots,j_i$ *be the strongly connected components of* int D_i . *Let* $\{E_i\}$ *be a partition of* $\mathbb{R}^n$ *into convex sets with nonempty interiors. Then, if the partition* $\{E_i\}$ *is finer than the partition* $\{D_i\}$ *, it is also finer than the partition* $\{\bar{D}_{ij}\}$.

PROOF: If $E_r \subseteq D_j$ then the sets int $E_r \cap D_{ij}$ partition int E_r into strong components. Since E_r is strongly connected, it follows that $E_r \subset \bar{D}_{ij}$ for some j.

COROLLARY 3.2. *Let* $\{D_i\}$, $\{D_{ij}\}$ *be defined as above. Then a linear decision tree (with strong inequalities) solves the problem defined by* $\{D_i\}$ *iff it solves the problem defined by* $\{D_{ij}\}$.

We list three applications to this theorem.

(i) Any linear decision tree that distinguishes inputs ordered according to an odd permutation from those permuted by an even permutation, solves the sorting problem as well [5, §5.3.1, Ex. 29].

(ii) Any linear decision tree finding the k-th element of $x_1,\ldots,x_n$, also finds which k-1 elements are smaller than it [5, §5.3.3, Ex. 2].

(iii) Any linear decision tree finding the set of maximal points out of n points in the plane, also sorts the maxima by their x (and y) coordinates [9,4].

4. FACE COUNTING ARGUMENT

We have shown in the previous chapter how to count correctly the number of components of a partition. We want now to take into consideration the complexity of each component, as embodied in its face structure. We shall restrict ourselves in this chapter to partitions consisting of convex polyhedral sets. The extension to the general case is straightforward but tedious.

We omit the proof to the following geometrical lemma.

LEMMA 4.1. *Let* $\{D_i\}$ *be a partition of the convex polyhedral set* D *into convex polyhedral subsets. Then each* k*-dimensional face of* D *contains a* k*-dimensional face of some subset* D_i.

Thus, if a partition $\{E_j\}$ is finer than the partition $\{D_i\}$, it also refines the face structure of $\{D_i\}$. This can be used to prove lower bounds based on a face counting argument. We have:

LEMMA 4.2. *Let* T *be a binary decision tree of depth* d, *with strong inequalities in* $\mathbb{R}^n$. *The sets in the partition* P(T) *defined by the leaves of* T *have at most* $\binom{d}{k}2^{d-k}$ n-k *dimensional faces.*

PROOF: We assume without loss of generality that T does not contain redundant comparisons. We associate each leaf v of T with a string of 1's and 2's representing the outcomes of the comparisons on the path leading to v. Each string thus obtained has length at most d and no string is a prefix of another. Each n-k dimensional face of I(v), the set of inputs reaching v, is obtained by replacing k inequalities on the path to v by equalities. We encode that face by putting n-k zeros in the corresponding positions of the string associated with v. It is now clear that the total number of n-k dimensional faces is bounded by the number of strings of length d over {0,1,2} containing exactly k zeros, which is $\binom{d}{k} 2^{d-k}$.

THEOREM 4.3. *Let* P *be the problem defined by the convex polyhedral sets* $\{D_i\}$ *in* $\mathbb{R}^n$. *Let* $F(P,k)$ *be the sum of the number of* k*-dimensional faces of each set* D_i . *Then, if* T *is a linear decision tree of depth* d *solving* P , *we have for* $0 \le k \le n$

$$\binom{d}{k} 2^{d-k} \ge F(P,n-k) \tag{4.1}$$

PROOF: Follows immediately from the previous two lemmas.

Note that for $k = 0$ we have the standard information theoretic inequality

$$2^d \ge \# \text{ of sets defining P.}$$

Inequality (4.1) also implies that the depth of T is at least as large as the maximum degree of a face. This degree argument implies for example that at least n-1 comparisons are needed to find the maximum of n elements.

It is interesting to note that Theorem 4.3 is proved in [13], for the particular case of a two-outcomes problem, by an adversary argument.

5. INVARIANCE PRINCIPLE

We wish to introduce in this chapter a new principle, which is familiar from other fields of mathematics. A way of showing that some problem has a simple structure is to show that it is invariant under a large family of transformations. If this is the case then we can in general restrict the search for an efficient solution to algorithms which are "invariant" under the "dual" family of transformations. To pick a very simple example, if x_n does not occur in the definition of a problem P in $\mathbb{R}^n$, a fact that can be expressed by saying that the sets $\{D_i\}$ defining P are invariant under translations along the x_n axis, then comparisons involving x_n do not help in solving P.

We first need a few facts from linear algebra. Let L^n be the space of *affine functionals* on $\mathbb{R}^n$, and A^n be the space of *affine transformations* in $\mathbb{R}^n$. Each functional $f \in L^n$ has the form $f(\bar{x}) = (\bar{a} \cdot \bar{x}) + b$, and we associate f with the n+1 vector $\bar{v}_f = [\bar{a};b]$. Each transformation $f \in A^n$ has the form $f(\bar{x}) = A\bar{x} + \bar{b}$, and we associate f with the $(n+1) \times (n+1)$ matrix

$$M_f = \begin{bmatrix} A & \bar{b} \\ 0 & 1 \end{bmatrix} . \tag{5.1}$$

Let $f \in L^n$, $F \in A^n$. The functional f is (weakly) *invariant under* F if $\forall\ \bar{x}$, $f(\bar{x}) \geq 0 \Leftrightarrow f(F\bar{x}) \geq 0$. This holds true iff $\bar{v}_f M_F = \lambda \bar{v}_f$, with $\lambda > 0$, that is, iff $\bar{v}_f$ is an eigenvector of M_F with positive eigenvalue. Note that f is invariant under F iff F maps each of the half spaces $[f(\bar{x}) \geq 0]$, $[f(\bar{x}) \leq 0]$ into itself.

A polyhedral partition $\{D_i\}$ is *invariant under* F if F preserves the sets $\{D_i\}$ and their faces. Let $\{f_1,\ldots,f_k\}$ be a minimal set of functionals such that the sets D_i are in the algebra of sets generated by the sets $[f_i(\bar{x}) \geq 0]$, $i = 1,\ldots,k$. (Each set D_i can be defined by some propositional combination of the terms $f_i(\bar{x}) \geq 0$, $i = 1,\ldots,k$). $\{f_1,\ldots,f_k\}$ is defined to be a *basis* for $\{D_i\}$. The partition $\{D_i\}$ is invariant under F iff f_i is invariant under F, for $i = 1,\ldots,k$.

The main theorem of this section is given below:

THEOREM 5.1. *Let* P *be the problem defined by the partition* $\{D_i\}$ *and let* G *be the set of affine transformations that keep that partition invariant. There exists a minimal depth linear decision tree that solves* P *using only comparisons invariant under* G.

The proof of that theorem will involve repeated applications of the following lemma.

LEMMA 5.2. *Let* T *be a decision tree (with strong inequalities) that solves the problem defined by the partition* $\{D_i\}$. *Let* T' *be the decision tree obtained from* T *by replacing each comparison* $f(\bar{x}) : 0$ *by the comparison* $f'(\bar{x}) : 0$. *Let* G *be a set of affine transformations that fulfills the following conditions:*

(i) $\forall F \in G$, $\bar{x} \in D_i$ *iff* $F\bar{x} \in D_i$
  (P *is invariant under* G);

(ii) *For any* f' *in* T' *and* $F \in G$
  $f'(F\bar{x}) > 0$ *iff* $f'(\bar{x}) > 0$
  (T' *is invariant under* F);

(iii) $\forall \bar{x} \in \mathbb{R}^n$ $\exists F \in G$ *such that*
  $f'(F\bar{x}) > 0$ *iff* $f(F\bar{x}) > 0$ *for any* f *in* T .

Then T' *solves* P .

PROOF: We shall show that if $\bar{x}$ reaches in T' the leaf v, which is associated in T with D_i , then $\bar{x} \in D_i$. Indeed, by (iii) there exist $F \in G$ such that $F\bar{x}$ reaches the same leaf in T and T'. By (ii), $F\bar{x}$ reaches leaf v in T', so that $F\bar{x}$ reaches leaf v in T, and $F\bar{x} \in D_i$. By (i), it follows that $\bar{x} \in D_i$.

THEOREM 5.3. *Let* $\{f_1,\ldots,f_k\}$ *be a basis for the partition* $\{D_i\}$. *There exists a minimal depth linear decision tree solving the problem defined by* $\{D_i\}$ *that uses only comparisons that are linear combinations of* $f_1,\ldots,f_k$.

PROOF: Let T be a minimal decision tree for P . We distinguish two cases:

(i) Any linear homogeneous functional is spanned by $\{f_1,\ldots,f_k\}$. Replace in T each nonhomogeneous comparison $(\bar{u}\cdot\bar{x}) + a : 0$, with $a \neq 0$, by the comparison $a : 0$. Apply Lemma 5.2, with G being the set of expansions $\bar{x} \to \lambda\bar{x}$, $\lambda > 0$.

(ii) Otherwise, there exists a vector $\bar{w} \in \mathbb{R}^n$ such that $\bar{w} \neq 0$ and $f_i(\bar{w}) = 0$, $i = 1,\ldots,k$. Let $V = sp\{f_1,\ldots,f_k\}$ and let $V^{\perp}$ be the orthogonal complement of V. Each functional $f \in L^n$ can be decomposed into $f = f_1 + f_2$, where $f_1 \in V$, $f_2 \in V^{\perp}$. Replace each comparison $f(\bar{x}) : 0$ by the comparison $f_2(\bar{w}) : 0$, if $f_2 \neq 0$. Apply Lemma 5.2, with G being the set of translations $\bar{x} \to \bar{x} + \lambda\bar{w}$, $\lambda > 0$.

PROOF OF THEOREM 5.1: Let T be a minimal depth decision tree for P. By the previous theorem we can assume without loss of generality that T uses only comparisons of the form $f(\bar{x}) : 0$, with $f \in V = sp\{f_1,\ldots,f_k\}$. Note that $[\bar{0},1] \notin V$. V can be decomposed into the direct sum of $s \leq k$ linear subspaces, $V = V_1 \oplus \ldots \oplus V_s$, such that f is invariant under G iff $f \in V_i$ for some i. Conversely, any affine transformation that preserves the functionals in $\cup V_j$ is in G. Let $1 > \lambda_1 > \lambda_2 > \ldots > \lambda_s$. Define a matrix $B = M_F$ of the form (5.1) such that $\bar{v}B = \lambda_i\bar{v}$ if $\bar{v} \in V_i$, $\bar{v}B = 0$ if $\bar{v} \notin V$. If $\bar{v} = \sum_{i=j}^{s} a_i\bar{v}_i$, with $\bar{v}_i \in V_i$, $\|\bar{v}_i\| = 1$, then $\bar{v}B^m = \sum_{i=j}^{s} a_i\lambda_i^m\bar{v}_i^m$, and $\bar{v}B^m/\|\bar{v}B^m\|$ converges to the vector $\bar{v}_j \in V_j$. Note that the transformations associated with the matrices B^m are in G, and that the limit $\bar{v}_j$ is a vector associated with a functional invariant under G.

Replace each comparison $f(\bar{x}) : 0$ in T by the comparison $f'(\bar{x}) : 0$, where $\bar{v}_{f'} = \lim \bar{v}_f B^m/\|\bar{v}_f B^m\|$. The resulting tree T' uses only comparisons invariant under G, and solves the problem P by Lemma 5.2.

Note that, unless T were of the prescribed form, some comparisons would be made redundant while T' is built.

Let us present some applications of this theorem. We have

COROLLARY 5.4 [14]. *If P is a problem defined by homogeneous inequalities then there exists a minimal depth decision tree solving P which uses only homogeneous comparisons.*

COROLLARY 5.5 [14]. *If P is a problem defined by inequalities of* *the form* $x_k < x_j$ *, then there exists a minimal depth decision tree which solves* P *using only comparisons of the form* $\Sigma a_i x_i : 0$ *, where* $\Sigma a_i = 0$.

We would like to conclude this section with an interesting combined application of the invariance principle with the adversary method. First let us introduce a few definitions.

Let P_1 and P_2 be problems defined on inputs $x_1,\ldots,x_m$ and $y_1,\ldots,y_m$ respectively by the families of sets $\{D_i\}$ and $\{E_j\}$. The *Cartesian product* of P_1 and P_2 , $P_1 \times P_2$ is defined on inputs $x_1,\ldots,x_m,y_1,\ldots,y_n$ by the sets $\{D_i \times E_j\}$. It consists of deciding concurrently to which of the sets D_i $\langle x_1,\ldots,x_m\rangle$ belongs, and to which of the sets E_j $\langle y_1,\ldots,y_n\rangle$ belongs. It is therefore obvious that the (linear) complexity of $P_1 \times P_2$ is at most equal to the sum of the complexities of P_1 and P_2 . Surprisingly, there are problems for which equality does not hold (for a result with the same flavor, see [7]). The following example is due to M. Rabin [9].

Let P_1 be the problem defined in $\mathbb{R}$ by the 5 sets $x \leq 0$, $i \leq x \leq (i+1)$ for $i = 0,1,2$ and $i \geq 3$. Let P_2 be the problem defined in $\mathbb{R}$ by the 9 sets $y \leq 0$, $i \leq y \leq (i+1)$ for $i = 0,\ldots,6$ and $y \geq 7$. The problem P_1 has 5 outcomes, and 3 comparisons are required to solve it, the problem P_2 has 9 outcomes and 4 comparisons are required to solve it. Yet the problem $P_1 \times P_2$ can be solved using only 6 linear comparisons: Figure 5.1 illustrates a partition of the xy plane that can be obtained in 4 comparisons. Each region of the partition intersects with at most 4 of the rectangular sets of the partition, and two more comparisons can distinguish each one.

The problems used in the previous example are defined by nonhomogeneous equations. It turns out that this is essential:

THEOREM 5.6. *If* P_1 *and* P_2 *are defined by homogeneous inequalities then the linear complexity of* $P_1 \times P_2$ *is equal to the sum of the complexities of* P_1 *and* P_2.

The proof of this theorem consists of two parts: we first show that one can assume without loss of generality that each step in the decision algorithm either involves the x_i or involves the y_j , but not both. This is done using the invariance principle. We then prove the theorem using an adversary argument.

LEMMA 5.7. *If* P_1 *and* P_2 *are defined by homogeneous inequalities then there is a minimal decision tree solving* $P_1 \times P_2$ *using only comparisons of the form* $f(\bar{x}) : 0$ *or* $g(\bar{y}) : 0$.

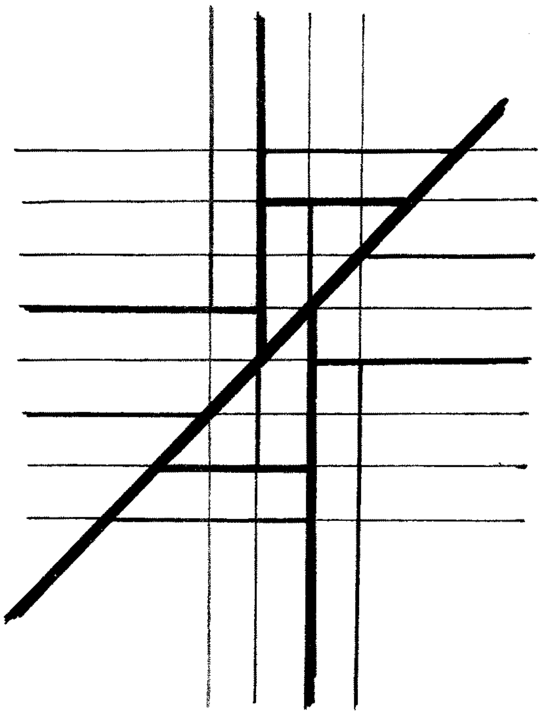

Figure 5.1

PROOF: The problem $P_1 \times P_2$ is invariant under the group of transformations of the form $T(\bar{x},\bar{y}) = (r\bar{x},s\bar{y})$, $r,s > 0$. A nontrivial affine functional is invariant under these transformations iff it depends only on $\bar{x}$ or depends only on $\bar{y}$.

PROOF OF THEOREM 4.6: Let us first define formally what is the adversary argument. We associate with the problem P to be solved a game played between *decider* and *adversary*. Alternately decider chooses a comparison on $x_1,\ldots,x_n$ and adversary chooses a consistent answer to this comparison. The game ends if the results of the comparisons determine a unique answer to problem P. The complexity of problem P is c iff the adversary has a strategy forcing c moves in the game.

Now let G_1 and G_2 be the games corresponding to problems P_1 and P_2 respectively, and c_1 and c_2 their respective complexities. By Lemma 5.7 we can restrict our attention in the solution of $P_1 \times P_2$ to unmixed comparisons involving only x_i's or y_j's. The complexity of $P_1 \times P_2$ is therefore equal to the maximal number of moves the adversary can force in the Cartesian product $G_1 \times G_2$ of the games G_1 and G_2, where moves are either moves of G_1 or moves of G_2, and this is c_1+c_2 [4].

6. CONCLUSIONS

We have exhibited several general methods that can be used to obtain lower bounds for linear decision trees. Applications were scarcely given, and the interested reader can find other results in [11], and in the other papers mentioned in the bibliography.

Some of these methods seem to generalize to more powerful comparisons (quadratic, polynomial (?)), particularly, the face counting argument. One must however, in order to do so, leave the well trodden fields of linear algebra and linear geometry, and enter into the more exotic realm of (real) algebraic geometry.

REFERENCES

1. M. Friedman and B. Weide, On the complexity of computing the measure of $\cup[a_i,b_i]$, *Comm. ACM 21* (1978), 540-544.

2. F. Fussenegger and H. Gabow, Using comparison trees to derive lower bounds for selection problems, *Proceedings IEEE Ann. Symp. on Foundations of Computer Science* (1976), 178-182.

3. B. Grunbaum, *Convex Polytopes*, Wiley, London, 1967.

4. J. C. Holladay, Cartesian products of termination games, in: M. Dresher, A. W. Tucker, P. Wolfe (Eds.), *Contributions to the Theory of Games, Vol. 3,* Annals of Math. Studies No. 39, Princeton University Press, Princeton, N. J., 1957, 189-199.

5. D. E. Knuth, *The Art of Computer Programming, Vol. 3: Sorting and Searching,* Addison-Wesley, Reading, Mass., 1973.

6. H. T. Kung, F. L. Luccio and F. P. Preparata, On finding the maxima of a set of vectors, *J. ACM 22* (1975), 469-476.

7. W. J. Paul, Realizing Boolean functions on disjoint sets of variables, *Th. Comp. Sci. 2* (1976), 383-396.

8. M. O. Rabin, Proving simultaneous positivity of linear forms, *J. Comput. Sys. Sci. 6* (1972), 639-650.

9. M. O. Rabin, Private communication.

10. E. M. Reingold, Computing the maxima and the median, *Proc. IEEE 12th Ann. Symp. on Switching and Automata Theory* (1971), 216-218.

11. M. Snir, Comparisons between linear forms can help, *Tech. Rep. CSR-60-80,* Comp. Sci. Dept., Univ. of Edinburgh, 1980. Also to be published in *Th. Comp. Sci.*

12. A. C. Yao, On the complexity of comparison problems using linear functions, *Proc. IEEE 16th. Ann. Symp. on Foundations of Computer Science* (1975), 85-89.

13. A. C. Yao and R. Rivest, On the polyhedral decision problem, *SIAM J. Comput. 9* (1980), 343-347.

14. C. K. Yap, On lifted problems, *Proc. IEEE 19th Ann. Symp. on Foundations of Computer Science* (1978), 267-279.