ON SELECTING THE k LARGEST WITH MEDIAN TESTS

Andrew Chi-Chih Yao

CS-TR-138-88

March 1988

On Selecting the k Largest with Median Tests[1]

Andrew Chi-Chih Yao

Department of Computer Science
Princeton University
Princeton, New Jersey 08544

Abstract

Let $W_k(n)$ be the minimax complexity of selecting the k largest elements of n numbers $x_1, x_2, \ldots, x_n$ by pairwise comparisons $x_i : x_j$. It is well known that $W_2(n) = n - 2 + \lceil \lg n \rceil$, and $W_k(n) = n + (k-1)\lg n + O(1)$ for all fixed $k \geq 3$. In this paper we study $W_k'(n)$, the minimax complexity of selecting the k largest, when tests of the form "Is x_i the median of $\{x_i, x_j, x_t\}$?" are also allowed. It is proved that $W_2'(n) = n - 2 + \lceil \lg n \rceil$, and $W_k'(n) = n + (k-1)\lg_2 n + O(1)$ for all fixed $k \geq 3$.

[1]This research was supported in part by the National Science Foundation under grant number DCR-8308109.

1 Introduction

The problem of finding the k largest elements, including their individual rankings, of n distinct real numbers, and other variants of the selection problem, have been studied extensively (see e.g. Knuth [Kn, Section 5.3]). Let $W_k(n)$ denote the worst-case complexity in the decision tree model when only comparisons of the form $x_i : x_j$ are allowed. It is well known that $W_1(n) = n - 1$, $W_2(n) = n - 2 + \lceil \lg n \rceil$ (Kislytsyn [Kis]), and, for large n, $W_k(n) = n + (k-1) \lg n + O(1)$ for any fixed $k \geq 3$ (Yao [Y3], Pratt and Yao [PY], Hyafil [H], Kirkpatrick [Kir]). When comparisons $f(x_1, x_2, \ldots, x_n) : 0$ with more general f are allowed, the corresponding complexity is less understood. It is known that, when linear functions f are permitted, the corresponding complexity satisfies $\hat{W}_1(n) = n - 1$ (Reingold [Re]), $\hat{W}_2(n) = n - 2 + \lceil \lg n \rceil$ (Yao [Y1]), and $\hat{W}_k(n) = n + (k-1) \lg n + O(1)$ for fixed $k \geq 3$ (Fusseneger and Gabow [FG]). However, when higher degree polynomials f are employed, it is only known (Rabin [Ra]) that $n-1$ comparisons are necessary and sufficient to find the largest element of n numbers. In particular, it is not even known whether, for some constant c, $n + c$ comparisons $f(x_1, x_2, \ldots, x_n) : 0$ with quadratic polynomials f are sufficient to determine the two largest of n real numbers.

In this paper, we study the complexity $W'_k(n)$ of finding the k largest of $x_1, x_2, \ldots, x_n$ using comparisons $x_i : x_j$ and a special type of quadratic tests $(x_j - x_i)(x_i - x_t) : 0$ (i.e. "Is x_i the median of $\{x_i, x_j, x_t\}$?").

To be precise, an *algorithm* A for finding the k largest elements is a binary decision tree, in which each internal node u contains a test of either the form "$x_i - x_j : 0$" or "$(x_r - x_s)(x_s - x_t) : 0$", where $i \neq j$ and $r \neq s \neq t$, and has two outgoing branches labeled respectively by " $<$ " and " $>$ "; each leaf ℓ of A contains an *output*, which is a k-tuple of integers $(\psi_1(\ell), \psi_2(\ell), \ldots, \psi_k(\ell))$. Let R_0^n be the set of $\tilde{x} = (x_1, x_2, \ldots, x_n)$, where x_i are distinct real numbers. For any input $\tilde{x} = (x_1, x_2, \ldots, x_n) \in R_0^n$, one can traverse a unique path in A from the root down, testing and branching at internal nodes encountered, until a leaf ℓ is reached; for A to be an algorithm, it is required that, for all $1 \leq i \leq k$, $x_{\psi_i(\ell)}$ must be the i-th largest among $x_1, x_2, \ldots, x_n$. Denote by *cost*$(A, \tilde{x})$ the number of internal nodes along the path. Let $C(A) = \max\{\text{cost}(A, \tilde{x}) | \tilde{x} \in R_0^n\}$. Let $\mathcal{A}_{n,k}$ be the family of all algorithms for finding the k largest of n distinct real numbers. Define the *complexity* by $W'_k(n) = \min\{C(A) | A \in \mathcal{A}_{n,k}\}$.

The main result of this paper is the next theorem, which essentially states that, for fixed k, the allowance of median tests changes the asymptotic complexity by at most an additive constant.

Theorem 1 $W'_k(n) \geq n - k + \sum_{1 \leq i \leq k-1} \lg(n - i + 1)$ for all $n > k \geq 2$.

Corollary $W'_2(n) = n - 2 + \lceil \lg n \rceil$ for $n \geq 2$, and for all fixed $k \geq 3$, $W'_k(n) = n + (k-1) \lg n + O(1)$ as $n \to \infty$.

In the corollary, the constant in the $O(1)$ term depends on the value of k. Note that the corollary follows from the theorem, the fact $W'_k(n) \leq W_k(n)$, and the known bounds on $W_k(n)$ mentioned earlier.

The complexity $V'_k(n)$ of the related problem of selecting *only* the k-th largest element of n numbers seems harder to determine. It would be of interest to prove a result analogous to Theorem 1. The next result states that $V'_2(n) \leq V_2(n) - 1$ for infinitely many n, where $V_2(n)$ is the complexity of selecting the second largest element when only direct comparisons $x_i : x_j$ can be used. ($V_2(n) = W_2(n) = n - 2 + \lceil \lg n \rceil$; see [Kn, Section 5.3].) That is, one can save at least one test infinitely often, if median tests are allowed. No corresponding phenomenon is known for the problem of selecting *all* the k largest.

Theorem 2 $V'_2(n) \leq n - 3 + \lceil \lg\ n \rceil$ for $n = 2^k + 1$ for all positive integers k.

In Section 2, a useful auxiliary theorem is derived; this result is also of independent interest. In Section 3 we prove Theroems 1 and 2. Some open problems are mentioned in Section 4.

2 A Geometric Theorem

We prove in this section a result (Theorem 3 below) with a geometric flavor, which will be needed to prove Theorems 1. For any set of real-valued functions G in R^n, let $S_G = \{\tilde{x} | g(\tilde{x}) > 0 \ \forall\ g \in G\}$; let $S_G = R^n$ when $G = \emptyset$. Let H be a set of real-valued functions in R^n. We will say that G is a *certificate for* H if $S_G \cap R^n_0 \neq \emptyset$ and $S_G \cap R^n_0 \subseteq S_H$. Thus, if a point $\tilde{x} \in R^n_0$ is known to satisfy the constraints $g(\tilde{x}) > 0$ for all $g \in G$, then $\tilde{x}$ must satisfy the constraints $h(\tilde{x}) > 0$ for all $h \in H$.

Let L_n denote the set of all functions of the form $\sum_{1 \leq i \leq n} \lambda_i x_i$, where all λ_i are real and at least one λ_i is nonzero. For any $H \subseteq L_n$, let *rank* (H) be the maximum number of linearly independent functions in H. Let $L_n^{(j)}$ denote the set of all functions of the form $p_1(\tilde{x}) \cdot p_2(\tilde{x}) \ \ldots \ p_j(\tilde{x})$, where $p_i(\tilde{x}) \in L_n$ for all i.

Theorem 3 Let $G \subseteq L_n \cup L_n^{(2)}$ and $H \subseteq L_n$ be two finite sets of functions, where $n \geq 2$. If G is a certificate for H, then $|G| \geq \text{rank}\ (H)$.

The rest of this section is devoted to a proof of Theorem 3. We will consider R^n as a vector space over the reals. For any $0 \leq \ell \leq n$, let $\mathcal{V}_{n,\ell}$ denote the set of all linear subspaces of R^n with dimension ℓ. For any $J \subseteq L_n^{(2)}$, let $N_J = \{\tilde{x} | g(\tilde{x}) = 0 \text{ for some } g \in J\}$; let $N_J = \emptyset$ when $J = \emptyset$. It is clear that $S_J \cap N_J = \emptyset$.

Lemma 1 Let $0 \leq m < n$, $J \subseteq L_n^{(2)}$ with $|J| = m$. If $\tilde{y} \in S_J$, then there exists $V \in \mathcal{V}_{n,n-m}$ such that $\tilde{y} \in V$ and $V - N_J \subseteq S_J$.

Proof We prove the lemma by induction on $m \geq 0$. If $m = 0$, we can satisfy the lemma by taking $V = R^n$. In the inductive step, let $0 < m_0 < n$, and assume that we have proved the lemma for all $m < m_0$. We will prove it for $m = m_0$.

Let $J = \{f_1, f_2, \ldots, f_{m_0}\}$. By the induction hypothesis, there exists $V_1 \in \mathcal{V}_{n,n-m_0+1}$ such that $\tilde{y} \in V_1$ and $V_1 - N_{J_1} \subseteq S_{J_1}$, where $J_1 = \{f_1, f_2, \ldots, f_{m_0-1}\}$.

Write $f_{m_0}(\tilde{x}) = p(\tilde{x}) \cdot q(\tilde{x})$, where $p, q \in L_n$. Let $Q = \{\tilde{x} | p(\tilde{x}) = 0 \ , \ q(\tilde{x}) = 0\}$, and $T = V_1 \cap Q$. Then T is a linear space of dimension at least $(n - m_0 + 1) - 2 = n - m_0 - 1$. Let $W \subseteq T$ be any linear subspace of T of dimension $n - m_0 - 1$. Define $V = \{\tilde{x} + \lambda\tilde{y} | \tilde{x} \in W \ , \ -\infty < \lambda < \infty\}$. We need to verify that V satisfies the requirements as stated in the lemma.

As $\tilde{y} \in S_J$, we have $\tilde{y} \notin Q$ and hence $\tilde{y} \notin W$. This implies $V \in \mathcal{V}_{n-m_0}$. Also it is clear that $\tilde{y} \in V$. It remains to show that $V - N_J \subseteq S_J$. First, $V - N_J \subseteq V_1 - N_{J_1} \subseteq S_{J_1}$. Secondly, for every $\tilde{z} \in V - N_J$, we have $\tilde{z} \in V - Q \subseteq V - T \subseteq V - W$, and thus $\tilde{z} = \tilde{x} + \lambda\tilde{y}$ where $\tilde{x} \in W$ and $\lambda \neq 0$, which in turn implies that $f_{m_0}(\tilde{z}) = p(\tilde{z}) \cdot q(\tilde{z}) = \lambda^2 p(\tilde{y}) q(\tilde{y}) > 0$; therefore $V - N_J \subseteq \{\tilde{x} | f_{m_0}(\tilde{x}) > 0\}$. From the above discussions, we conclude that $V - N_J \subseteq S_{J_1} \cap \{\tilde{x} | f_{m_0}(\tilde{x}) > 0\} = S_J$. This completes the inductive step of the proof. □

Lemma 2 Let $X \subseteq L_n$, $H \subseteq L_n$ be two finite sets of linear functions. Let $V \in \mathcal{V}_{n,\ell}$, and $Y = \cup_{1 \leq i \leq t} Y_i$ where $Y_i \in \mathcal{V}_{n,\ell_i}$, with $0 \leq \ell_i < \ell$, $0 < \ell \leq n$ and t any non-negative integer. If $S_X \cap (V - Y) \neq \emptyset$ and $S_X \cap (V - Y) \subseteq S_H$, then $\text{rank}(X) + (n - \ell) \geq \text{rank}(H)$.

Proof Let $V = \{\tilde{x} | p_i(\tilde{x}) = 0, \quad 1 \leq i \leq n - \ell\}$, where $p_i \in L_n$. For any set $B \subseteq R^n$, let $\bar{B}$ denote the closure of B under the standard topology on R^n (induced by e.g. the Euclidean metric). It is elementary that $\overline{V - Y} = V$ and that $\bar{S}_H = \{h(\tilde{x}) \geq 0 | h \in H\}$. It follows that $S_X \cap V = S_X \cap (\overline{V - Y}) \subseteq \bar{S}_H = \{h(\tilde{x}) \geq 0 | h \in H\}$. By the well-known Farkas' Lemma (see e.g. [SW]), we can write for each $h \in H$, $h = \sum_{f \in X} \lambda_f \cdot f + \sum_{1 \leq i \leq n-\ell} \mu_i p_i$ for some constants $\lambda_f \geq 0$ and arbitrary μ_i. This immediately implies $\text{rank}(H) \leq \text{rank}(X) + (n - \ell)$. □

We will now prove Theorem 3. We can assume that $|G| < n$, as otherwise $\text{rank}(H) \leq n \leq |G|$ is obviously true. Suppose $G = \{g_1, g_2, \ldots, g_{|G|}\}$ with $g_i \in L_n^{(2)}$ for $0 < i \leq m$ and $g_j \in L_n$ for $m < j \leq |G|$, where $0 \leq m \leq |G|$. Let $J = \{g_1, g_2, \ldots, g_m\}$. Choose any $\tilde{y} \in S_G \cap R_0^n$. Then $\tilde{y} \in S_J$, and by Lemma 1, there exists a $V \in \mathcal{V}_{n,n-m}$ such that $\tilde{y} \in V$ and $V - N_J \subseteq S_J$. Define $Y_i = V \cap \{\tilde{x} | g_i(\tilde{x}) = 0\}$ for $1 \leq i \leq m$. Let $t = m + \binom{n}{2}$; let Y_ℓ, $m < \ell \leq t$, be the $\binom{n}{2}$ linear spaces of the form $V \cap \{\tilde{x} = (x_1, x_2, \ldots, x_n) | x_i = x_j\}$ where $i < j$. Then each Y_i, $1 \leq i \leq t$, is a linear space of dimension one less than the dimension of V, since $\tilde{y} \in V - Y_i$. Let $Y = \cup_{1 \leq i \leq t} Y_i$. Then $N_J \subseteq Y$.

Now, let $X = \{g_i | m < i \leq |G|\}$. Clearly, $S_X \cap (V - Y) \neq \emptyset$, as $\tilde{y} \in S_X \cap (V - Y)$. Furthermore, $S_X \cap (V - Y) \subseteq S_X \cap (V - N_J) \subseteq S_X \cap S_J = S_H$. By Lemma 2, $\text{rank}(X) + n - (n - m) \geq \text{rank}(H)$, which implies $|X| + m \geq \text{rank}(H)$, i.e. $|G| \geq \text{rank}(H)$. This proves Theorem 3.

3 Proof of Theorems 1 and 2

We first prove Theorem 1. The general approach of the proof extends that used in Fussenegger and Gabow [FG]. Given any algorithm $A \in \mathcal{A}_{n,k}$, we will classify its leaves into $\prod_{1 \le i \le k-1}(n-i+1)$ classes, and show that each class must contain at least 2^{n-k} leaves. This gives a lower bound of $2^{n-k}\prod_{1 \le i \le k-1}(n-i+1)$ to the total number of leaves in A. Taking the logarithm gives a lower bound on the cost of the algorithm. Before doing that, we use Theorem 3 from the previous section to establish a result (Lemma 4 below) concerning algorithms for finding the maximum of n numbers.

Let G and H be finite sets of functions on R^n. The next lemma states that any certificate for x_1 being the maximum of $x_1, x_2, \ldots, x_n$ must have cardinality at least $n-1$.

Lemma 3 Let $G \subseteq L_n \cup L_n^{(2)}$ be a certificate for $H = \{x_1 - x_i | 2 \le i \le n\}$, where $n \ge 2$. Then $|G| \ge n-1$.

Corollary Any $A \in \mathcal{A}_{n,1}$ must have at least 2^{n-1} leaves.

Proof The lemma follows from Theorem 3, since $\text{rank}(H) = n-1$. We now prove the corollary. Let $A \in \mathcal{A}_{n,1}$. Without loss of generality, we can assume that there are no redundant tests, i.e. each branch of any internal node is traversed by some input $\tilde{x} \in R_0^n$. By the lemma, no node at a distance $j \le n-2$ from the root can be a leaf. It follows that there are 2^{n-2} internal nodes u at a distance $n-2$ from the root. Since each such u has at least two leaves among its descendants, the corollary follows. □

We need to consider a class of decision tree algorithms more general than $\mathcal{A}_{n,1}$. Consider input vectors of n distinct components $\tilde{x} = (x_1, x_2, \ldots, x_n)$. Let B be a decision tree, each of whose internal nodes u contains a test $f(\tilde{x}) : 0$, and has two outgoing branches labeled by "<" and ">", where f is either c, $c(x_i - x_j)$, or $(x_i - x_j)(x_j - x_r)$ for nonzero constants c; every leaf of B is associated with an integer output $\eta(\ell)$. We will say "B selects the largest element of n numbers", if for every input $\tilde{x} = (x_1, x_2, \ldots, x_n) \in R_0^n$, the leaf ℓ reached will give the correct output, i.e. $x_{\eta(\ell)}$ is the largest of all the x's. Let $\mathcal{B}_n$ be the family of all such B.

For any $B \in \mathcal{B}_n$, let T_B denote the set of leaves that can be reached for at least one input $\tilde{x} \in R_0^n$.

Lemma 4 Let $n \ge 2$. For any $B \in \mathcal{B}_n$, $|T_B| \ge 2^{n-1}$.

Proof For any $B \in \mathcal{B}_n$, one can prune away all the branches and nodes that cannot be reached by any input $\tilde{x} \in R_0^n$, and obtain a $B' \in \mathcal{B}_n$ such that the number of leaves in B' is no greater than $|T_B|$. To establish the lemma, it is sufficient to show that the number of leaves in B' is at least 2^{n-1}.

We construct from B' a modified decision tree. We process the internal nodes one at a time (in any order). For each internal node u with test $f : 0$, we perform the following modification: if $f = c(x_i - x_j)$ and c is positive, we replace the test by $x_i - x_j : 0$; if $f = c(x_i - x_j)$ and c is negative, we replace the test by $x_j - x_i : 0$; if $f = (x_i - x_j)(x_j - x_t)$, we change nothing; if $f = c$, then u has only one son, and we will just erase the node u (i.e. connect the parent node of u directly to the son of u if u is not the root, or in case that u is the root, delete u and make its son the new root). It is clear that the modified decision tree A has $|T_B|$ leaves, and is in $\mathcal{A}_{n,1}$. By the coollary to Lemma 3, A has at least 2^{n-1} leaves. It follows that $|T_B| \geq 2^{n-1}$.□

To prove Theorem 1, let $A \in \mathcal{A}_{n,k}$; we will prove

$$C(A) \geq n - k + \sum_{1 \leq i \leq k-1} \lg(n - i + 1). \tag{1}$$

Let $N(A)$ be the number of leaves in A. It suffices to show that

$$N(A) \geq 2^{n-k} \cdot \prod_{1 \leq i \leq k-1} (n - i + 1). \tag{2}$$

For any $(k-1)$-tuple $I = (i_1, i_2, \ldots, i_{k-1})$ of $k-1$ distinct integers between 1 and n, let M_I denote the set of input $\tilde{x} = \{x_1, x_2, \ldots, x_n\} \in R_0^n$ such that x_{i_j} is the j-th largest element of $\{x_1, x_2, \ldots, x_n\}$ for $1 \leq j \leq k-1$. Let L_I be the set of leaves for which the output is $(i_1, i_2, \ldots, i_{k-1}, m)$ for some m. Thus, if ℓ is reached when some $\tilde{x} \in M_I$ is input, then $\ell \in L_I$. To prove (2), it is sufficient to prove that $|L_I| \geq 2^{n-k}$ for all I, as L_I and L_J are disjoint if $I \neq J$. Without loss of generality, we will only prove

$$|L_I| \geq 2^{n-k}, \tag{3}$$

for $I = (n, n-1, \ldots, n-k+2)$.

Let $S = \{i \mid 1 \leq i \leq n-k+1\}$ and $S' = \{i \mid n-k+2 \leq i \leq n\}$. Let F denote the set of functions of the form $f = x_i - x_j$ or $f = (x_i - x_j)(x_j - x_r)$ on variables $x_1, x_2, \cdots, x_n$, and F' be the set of functions of the form $f = c$, $f = c(y_i - y_j)$, or $(y_i - y_j)(y_j - y_r)$ on variables $y_1, y_2, \ldots, y_{n-k+1}$, where c are constants. Define a mapping D from F to F' as follows: $D(x_i - x_j) = y_i - y_j$ if $i, j \in S$, and $D(x_i - x_j)$ is defined as the constant $i - j$ if at least one of i, j is not in S; let $D((x_i - x_j)(x_j - x_r)) = D(x_i - x_j)D(x_j - x_r)$.

We now construct from A a decision tree B by the following modifications: At each internal node, replace its test $f : 0$ by $D(f) : 0$; at each leaf in L_I, replace its output $(n, n-1, \ldots, n-k+1, m)$ by a single output m; for all other leaves, the outputs are set to be 1 (in fact any integer will do).

Lemma 5 $B \in \mathcal{B}_{n-k+1}$ and $|T_B| \leq |L_I|$.

Proof For each node (internal node or leaf) $u \in B$, let $\xi(u)$ be the unique corresponding node in A. For any $\tilde{y} = (y_1, y_2, \ldots, y_{n-k+1}) \in R_0^{n-k+1}$, let $\alpha(\tilde{y}) = (x_1, x_2, \ldots, x_n) \in R_0^n$, where $x_i = y_i$ for $i \in S$, and $x_j = y_j + \sum_{1 \le i \le n-k+1} |y_i|$ for $j \in S'$.

Fact 1 Let $1 \le i \le n$. Then x_i is the k-th largest element in $x_1, x_2, \ldots, x_n$ if and only if the following is true: $1 \le i \le n-k+1$ and y_i is the largest element of $y_1, y_2, \ldots, y_{n-k+1}$.

Fact 2 If for input $\tilde{y}$, $u_1, u_2, \ldots, u_t$ is the sequence of nodes traversed in B, then for input $\alpha(\tilde{y})$, $\xi(y_1), \xi(y_2), \ldots, \xi(y_t)$ is the sequence of nodes traversed in A.

Fact 1 is true, since $x_j > x_i = y_i$ for all $n-k+2 \le j \le n$ and $1 \le i \le n-k+1$. Fact 2 follows from the construction of B.

For any input $\tilde{y} = (y_1, y_2, \ldots, y_{n-k+1}) \in R_0^n$ to B, suppose that the traversed path ends in leaf u with output b. By Fact 2, if we feed input $\alpha(\tilde{y}) = (x_1, x_2, \ldots, x_n)$ to A, then the traversed path will end in $\xi(u)$. We now prove the two inequalities stated in the lemma.

Clearly, $\alpha(\tilde{y}) \in M_I$, and thus $\xi(u) \in L_I$. We conclude that $\xi(u) \in L_I$ for for every reachable leaf u in B; hence $|T_B| \le |L_I|$.

To prove that $B \in \mathcal{B}_{n-k+1}$, we need to show that y_b is the largest of the y_i's. Suppose that, in A, the output at $\xi(u)$ is $(i_1, i_2, \ldots, i_k)$. As $\xi(u) \in L_I$, we have $b = i_k$ from the construction of B. Since x_{i_k} is the k-th largest element in $x_1, x_2, \ldots, x_n$ by definition, we have, by Fact 1, y_b is the largest of $y_1, y_2, \ldots, y_{n-k+1}$. This proves $B \in \mathcal{B}_{n-k+1}$.□

It follows from Lemma 5 and Lemma 4 that $|L_I| \ge |T_B| \ge 2^{n-k}$. This proves (3), and completes the proof of Theorem 1.

We now turn to the proof of Theorem 2. Let $n = 2^k + 1$. We will give an algorithm A with $C(A) = n - 2 + k$, which identifies the second largest element of $x_1, x_2, \ldots, x_n$.

First perform a knockout balanced tournament using comparisons of the form $x_i - x_j : 0$ for each of the groups $\{x_1, x_2, \ldots, x_{2^{k-1}}\}$ and $\{x_{2^{k-1}}, \ldots, x_{n-1}\}$. This takes $2(2^{k-1} - 1) = n - 3$ tests. Now let the largest elements of the two groups be x_i, x_j and let S_1, S_2 be the set of x_t's directly defeated by x_i, x_j; clearly $|S_1| = |S_2| = k - 1$.

Now make one test "$(x_i - x_n)(x_n - x_j) : 0$."

CASE 1: If the answer is "$>$." then make one further test "$x_n - x_i : 0$;" this tells us whether $x_i < x_n < x_j$ or $x_j < x_n < x_i$. Without loss of generality, assume $x_i < x_n < x_j$ to be the case. We perform $k - 1$ tests to find the largest of $S_2 \cup \{x_n\}$, which clearly is the second largest of all x_ℓ's. The total number of tests is $(n-3) + 2 + (k-1) = n - 2 + k$.

CASE 2: If the answer is "$<$," then make one further test "$(x_i - x_j)(x_j - x_n) : 0$;" this tells us whether x_i or x_j is the median of $\{x_i, x_j, x_n\}$. Without loss of generality, assume the case x_j is median of $\{x_i, x_j, x_n\}$. It is easy to see that the maximum of the elements in $S_1 \cup \{x_j\}$ must be the second largest of all x_ℓ's. Thus, in $k-1$ further tests, we can find the desired output. The total number of tests is $(n-3)+2+(k-1) = n-2+k$.

This proves Theorem 2.

4 Conclusions

There are many interesting unresolved questions on this subject. The traditional region-counting technique for algebraic decision trees (e.g. Dobkin and Lipton [DL], Steele and Yao [SY], Ben-Or [B]) does not seem to yield nontrivial results to these problems. We list below a few open problems.

(a) Let $W_k^{(\ell)}(n)$ be the complexity of finding the k largest of n numbers, when polynomial tests of degree at most ℓ are allowed. Is there a constant $\mu > 0$ and a function $N(k,\ell)$ such that $W_k^{(\ell)}(n) - n \geq \mu\, k \log n$ for all $n \geq N(k,\ell)$? We conjecture that this is true at least for $k = \ell = 2$. The method used in Yao [Y2] for deriving lower bounds for the convex hull problem may be of some use in this special case.

(b) The complexity $V_k(n)$ of selecting *only* the k-th largest element is related to the complexity $W_k(n)$ of selecting all the k largest elements by an added constant term. Can we prove that the relation is still true, when median tests are allowed? Equivalently, can we prove that $V_k'(n) = n + (k-1)\lg n + O(1)$ for all fixed k and $n \to \infty$?

(c) Does Theorem 3 generalize to the case $G \subseteq L_n \cup L_n^{(2)} \cup \ldots \cup L_n^{(j)}$ for $j > 2$?

(d) Is there purely combinatorial proof of Lemma 3? The present proof involves geometric arguments, since it employs Theorem 3.

References

[B] M. Ben-Or, "Lower bounds for algebraic computation trees," *Proceedings of 15th ACM Symposium on Theory of Computing* (1983), 80-86.

[DL] D. Dobkin and R. J. Lipton, "A lower bound of $\frac{1}{2}n^2$ on linear search tree programs for the knapsack problems," *Journal of Computer and System Sciences* **16** (1978), 413-417.

[FG] F. Fussenegger and H.N. Gabow, "Using comparison trees to derive lower bounds for selection problems," *Journal of ACM* **26** (1979), 227-238.

[H] L. Hyafil, "Bounds for selection," *SIAM J. on Computing* **5** (1976), 109-114.

[Kir] D.G. Kirkpatrick, "Topics in the complexity of combinatorial algorithms," Computer Science Department Report TR 74(1974), University of Toronto.

[Kis] S.S. Kislytsyn, "On the selection of the k-th element of an ordered set by pairwise comparisons," Sibirsk. Mat. Zh. **5** (1964), 557-564.

[Kn] D.E. Knuth, *The Art of Computer Programming, Vol. 3: Searching and Sorting*, Addison-Wesley, 1973.

[PY] V.R. Pratt and F.F. Yao, "On lower bounds for computing the i-th largest element," *Proc. 14th IEEE Symp. on Switching and Automata Theory* (1973), 70-81.

[Ra] M. Rabin, "Proving simultaneous positivity of linear forms," *Journal of Computer System Sciences* **6**, (1972), 639-350.

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

[SW] J. Stoer and C. Witzgall, *Convexity and Optimization in Finite Dimensions I*, Springer-Verlag, 1970.

[SY] J.M. Steele and A.C. Yao, "Lower bounds for algebraic decision trees," *Journal of Algorithms* **3** (1982), 1-8.

[Y1] A.C. Yao, "On the complexity of comparison problems using linear functions," *Proceedings 16th Annual IEEE Symposium on Foundations of Computer Sciences*, 1975, 85-89.

[Y2] A.C. Yao, "A lower bound to finding convex hulls," *Journal of ACM* **28** (1981), 780-787.

[Y3] F.F. Yao, "On lower bounds for selection problems," Ph.D. thesis, Massachusetts Institute of Technology, 1973.