ON SELECTING THE SECOND LARGEST WITH MEDIAN TESTS

Andrew Chi-Chih Yao

CS-TR-121-87

November 1987

On Selecting the Second Largest with Median Tests[1]

Andrew Chi-Chih Yao

Department of Computer Science
Princeton University
Princeton, New Jersey 08544

Abstract

Let $V_k(n)$ be the minimax complexity of selecting the $k-th$ largest of n numbers $x_1, x_2, \ldots, x_n$ by pairwise comparisons $x_i : x_j$. It is well known that $V_2(n) = n - 2 + \lceil lg\, n \rceil$. In this paper we study $V_2'(n)$, the minimax complexity of selecting the second largest, when tests of the form "Is x_i the median of $\{x_i, x_j, x_k\}$?" are also allowed. It is proved that $n - 3 + \lceil lg\, n \rceil \leq V_2'(n) \leq n - 2 + \lceil lg\, n \rceil$. Furthermore, both upper and lower bounds are achieved for infinitely many n .

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

1 Introduction

The problem of finding the $k-th$ largest element of n distinct real numbers, also known as the tennis tournament problem, has been studied extensively (see e.g. Knuth [Kn]). Let $V_k(n)$ 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 (see [Kn]) that $V_1(n) = n-1$, $V_2(n) = n-2+\lceil lg\, n \rceil$ (Kislytsyn [Kl]), and, for large n, $V_k(n) = n + k\, lg\, n + O(1)$ for any fixed $k \geq 3$ (Yao [Y3], Pratt and Yao [PY], Hyafil [H], Kirkpatrick [K]). 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 $\hat{V}_1(n) = n-1$ (Reingold [Re]), $\hat{V}_2(n) = n-2+\lceil lg\, n \rceil$ (Yao [Y1]), and $\hat{V}_k(n) = n + k\, lg\, n + O(1)$ 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 second largest of n real numbers.

In this paper, we study the complexity $V'_2(n)$ of finding the second largest of $x_1, x_2, \ldots, x_n$ using comparisons $x_i : x_j$ and a special type of quadratic tests $(x_j - x_i)(x_i - x_k) : 0$ (i.e. "Is x_i the median of $\{x_i, x_j, x_k\}$?").

To be precise, an *algorithm* A for finding the k-th largest element is a binary decision tree in which each internal node v contains a test of either the form "$x_i - x_j : 0$" or "$(x_j - x_i)(x_i - x_t) : 0$"; the two outgoing branches of v are labeled as " $<$ " and " $>$ "; each leaf ℓ of A contains an integer a_ℓ. Let R_0^n be the set of $\tilde{x} = (x_1, x_2, \ldots, x_n)$ with all x_i being 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 x_{a_ℓ} must be the k-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 $C(A) = max\{cost(A, \tilde{x}) | \tilde{x} \in R_0^n\}$. Let $\mathcal{A}_{n,k}$ be the family of all algorithms for finding the k-th largest of n distinct real numbers. Define the *complexity* by $V'_k(n) = min\{C(A) | A \in \mathcal{A}_{n,k}\}$.

It is of interest to note that $V_2(3) = 3$ while $V'_2(3) = 2$, since one can determine the second largest of x, y, z by asking two questions: " Is x the median of $\{x, y, z\}$?" and " Is y the median of $\{x, y, z\}$?". The purpose of this paper is to prove two theorems on $V'_2(n)$. Recall that $V_2(n) = n-2+\lceil lg\, n \rceil$. The first theorem states that at most one comparison can be saved by using this additional primitive.

Theorem 1 *$V'_2(n) \geq n-3+\lceil lg\, n \rceil$ for all $n \geq 3$.*

The second theorem shows that the above lower bound and the upper bound provided by $V'_2(n) \leq$

$V_2(n) = n - 2 + \lceil lg\, n \rceil$ can be achieved infinitely often.

Theorem 2 *$V_2'(n) = n - 3 + \lceil lg\ n \rceil$ for $n = 2^k + 1$, and $V_2'(n) = n - 2 + \lceil lg\, n \rceil$ for $n = 2^k$ for all positive integers k.*

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 and 2. 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 $S_G \cap R_0^n \neq \emptyset$ and $S_G \cap R_0^n \subseteq S_H$. Thus, if a point $\tilde{x} \in R_0^n$ 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 $|G| \geq$ 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 of Lemma 1

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 $W \subseteq T$ by 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 $Y = \cup_{1 \le i \le t} Y_i$ where $Y_i \in \mathcal{V}_{n,\ell_i}$, with $0 \le \ell_i < \ell$, $0 < \ell \le n$ and t any non-negative integer. If $S_X \cap (V - Y) \neq \emptyset$ and $S_X \cap (V - Y) \subseteq S_H$, then $rank(X) + (n - \ell) \ge rank(H)$.*

Proof of Lemma 2

Let $V = \{\tilde{x} | p_i(\tilde{x}) = 0 \quad 1 \le i \le n - \ell\}$ 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 $\bar{S}_H = \{h(\tilde{x}) \ge 0 | h \in H\}$. It follows that $S_X \cap V = S_X \cap (\overline{V - Y}) \subseteq \bar{S}_H = \{h(\tilde{x}) \ge 0 | h \in H\}$. By the well-known Farkas' Lemma (see e.g. [SW]), we can write for each $h \in H$, $h = \sum_{f \in X} \lambda_f \cdot f + \sum_{1 \le i \le n-\ell} \mu_i p_i$ for some constants $\lambda_f \ge 0$ and arbitrary μ_i. This immediately implies $rank(H) \le rank(X) + (n - \ell)$. □

We will now prove Theorem 3. We can assume that $|G| < n$, as otherwise $rank(H) \le n \le |G|$ is obviously true. Suppose $G = \{g_1, g_2, \ldots, g_{|G|}\}$ with $g_i \in L_n^{(2)}$ for $0 < i \le m$ and $g_j \in L_n$ for $m < j \le |G|$, where $0 \le m \le |G|$. 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 $1 \le i \le m$. Let $t = m + \binom{n}{2}$; let Y_ℓ, $m < \ell \le t$, 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, $1 \le i \le t$, is a linear space of dimension one less than the dimension of V, since $\tilde{y} \in V - Y_i$. Let $Y = \cup_{1 \le i \le t} Y_i$. Then $N_J \subseteq Y$.

Now, let $X = \{g_i | m < i \le |G|\}$. 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, $rank(X) + n - (n - m) \ge rank(H)$, which implies $|X| + m \ge rank(H)$, i.e. $|G| \ge rank(H)$. This proves Theorem 3.

3 Proof of Theorem 1

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 $n - 1$. Let $n \ge 3$.

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

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

Proof: The lemma follows from Theorem 3, since $rank(H) = n - 1$. We now prove the corollary. Let $A \in \mathcal{A}_{n,1}$. Without loss of generality, we can assume that each branch of any internal node is traversed by some input $\tilde{x} \in R_0^n$. By the lemma, each node at a distance $j \leq n - 2$ from the root is an internal node, and hence has two descendants. It follows that there are 2^{n-2} internal nodes at a distance $n - 2$ from the root, and each has at least two leaves as its descendants. This proves the corollary. □

Definition 1. For any set G of functions on R^n, let M_G denote the set of all j for which there exist $(x_1, x_2, \ldots, x_n) \in S_G \cap R_0^n$ with $x_j = max\{x_i | 1 \leq i \leq n\}$.

Definition 2. Let D_n denote the set of $(x_1, x_2, \ldots, x_n) \in R_0^n$ for which x_1 is the second largest of the x_i's.

Lemma 4 *Let $G \subseteq L_n \cup L_n^{(2)}$ such that $\emptyset \neq S_G \cap R_0^n \subseteq D_n$. Then $1 \leq |M_G| \leq 2$.*

Proof: Clearly $|M_G| \geq 1$. We need to prove $|M_G| \leq 2$. Let $G' \subseteq G$ be the subset of functions with a dependency on x_1. We partition G' into J_0, J_1, J_2, J_3. Let $J_0 = \{i | x_i - x_1 \in G'\}$, $J_1 = \{i | x_1 - x_i \in G'\}$, $J_2 = \{i | (x_1 - x_i)(x_1 - x_j) \in G' \; for\ some\ j \neq 1, i\}$, $J_3 = \{i | (x_i - x_1)(x_1 - x_j) \in G' \; for\ some\ j \neq 1, i\}$. If $|J_0| \neq 0$, then clearly $|J_0| = 1$ and $M_G = J_0$; in this case $|M_G| = 1$. We can thus assume that $|J_0| = 0$.

Claim: $M_G \subseteq J_3$.

To prove the claim, we first note that $i \in M_G$ must be in $J_1 \cup J_2 \cup J_3$; otherwise take a point $(y_1, y_2, \ldots, y_n) \in S_G \cap R_0^n$ with $y_i > y_1 > y_j$ for all $j \neq 1$, and let $\tilde{x} = (x_1, x_2, \ldots, x_n)$ with $x_1 = y_i$, $x_i = y_1$, $x_j = y_j \; for\ all\ j \neq 1, i$, then $\tilde{x} \in S_G \cap R_0^n$, but $\tilde{x} \notin D_n$. If $i \in J_1$, then for all $\tilde{x} = (x_1, x_2, \ldots, x_n) \in S_G \cap R_0^n$, $x_1 > x_i$; hence $i \notin M_G$. If $i \in J_2$, then for all $\tilde{x} = (x_1, x_2, \ldots, x_n) \in S_G \cap R_0^n$, $(x_1 - x_i)(x_1 - x_j) > 0$ for some $j \neq 1, i$; that implies $x_1 > x_i$ and $x_1 > x_j$, as otherwise $x_1 < x_i$, $x_1 < x_j$ which would contradict $\tilde{x} \in D_n$. Thus, $i \in J_2$ also implies $i \notin M_G$. This proves $M_G \subseteq J_3$.

If $|J_3| \leq 2$, then $|M_G| \leq 2$, and the lemma is clearly valid. If $|J_3| > 2$, then there exist two distinct $(x_i - x_1)(x_1 - x_j)$, $(x_s - x_1)(x_1 - x_t) \in G'$. If $\{i, j\} \cap \{s, t\} = \emptyset$, then, for all $\tilde{x} = (x_1, x_2, \ldots, x_n) \in S_G \cap R_0^n$, $x_1 < max\{x_i, x_j\}$ and $x_1 < max\{x_s, x_t\}$, and hence x_1 cannot be the second largest of $x_1, x_2, \ldots, x_n$, contradicting $\tilde{x} \in D_n$. If $\{i, j\} \cap \{s, t\} \neq \emptyset$, then without loss of generality, we can assume that $s = j$ and $t \neq i$. Thus, for all $\tilde{x} = (x_1, x_2, \ldots, x_n) \in S_G \cap R_0^n$, either $x_t, x_i < x_1 < x_j$ or $x_t, x_i > x_1 > x_j$; as $\tilde{x} \in D_n$, we can only have $x_t, x_i < x_1 < x_j$, which implies that $M = \{j\}$, and hence $|M_G| = 1$. We have thus proved $|M_G| \leq 2$ in all cases. □

To prove Theorem 1, let $A \in \mathcal{A}_{n,2}$. We need to prove that $C(A) \geq n - 3 + \lceil lg\, n \rceil$. Without loss of generality, we assume that each leaf can be reached by some input $\tilde{x} \in R_0^n$. Let $\sharp(A)$ be the number of leaves. For any leaf ℓ in A, let the constraints along the path from the root of A to ℓ be $\{f > 0 | f \in G_\ell\}$, where $G_\ell \subseteq L_n \cup L_n^{(2)}$. By Lemma 4, $1 \leq |M_{G_\ell}| \leq 2$.

Let us make one further comparison $x_i - x_j : 0$ at each leaf ℓ with $M_{G_\ell} = \{i, j\}$ and $i \neq j$, which clearly determines the $t \in M_{G_\ell}$ with $x_t = max\{x_1, x_2, \ldots, x_n\}$. This gives an algorithm $A' \in \mathcal{A}_{n,2}$ such that

$$\sharp(A') \leq 2\, \sharp(A) \;, \tag{1}$$

and each leaf ℓ' in A' has $|M_{G_{\ell'}}| = 1$. Let $L_i = \{\ell' | M_{G_{\ell'}} = \{i\}\}$. Clearly,

$$\sharp(A') = \sum_{1 \leq i \leq n} |L_i| \;. \tag{2}$$

We will now prove, for each $1 \leq i \leq n$,

$$|L_i| \geq 2^{n-2} \;. \tag{3}$$

This will complete the proof of Theorem 1, since (1), (2) and (3) imply

$$\begin{aligned} \sharp(A) &\geq \frac{1}{2}\sharp(A') \\ &\geq \frac{1}{2} n \cdot 2^{n-2} \;, \end{aligned} \tag{4}$$

and hence $C(A) \geq \lceil lg\, \sharp(A) \rceil = n - 3 + \lceil lg\, n \rceil$.

It remains to prove (3). Without loss of generality, assume $i = 1$. We will trim the branches of A' to make it an algorithm A'' in $\mathcal{A}_{n-1,1}$ and with $\sharp(A'') \leq |L_1|$. But by the corollary to Lemma 3, $\sharp(A'') \geq 2^{n-2}$, and thus $|L_1| \geq 2^{n-2}$.

For any internal node $v \in A'$, if the comparison is $x_i - x_1 : 0$, or $(x_i - x_1)(x_1 - x_j) : 0$, we delete the " $>$" branch outgoing from v, and then remove v; if the comparison is $x_1 - x_i : 0$, we delete the " $<$" branch, and then remove v; if the comparison is $(x_1 - x_i)(x_i - x_j) : 0$, or $(x_j - x_i)(x_i - x_1) : 0$, we replace the comparison by $x_i - x_j : 0$. At each leaf $\ell' \in A'$, the output j remains the same. It is easy to see that this gives an algorithm $B \in \mathcal{A}_{n-1,1}$ for input $\tilde{y} = (x_2, x_3, \ldots, x_n) \in R_0^{n-1}$, since the path traversed by $\tilde{y}$ in B is the same path as traversed by $(\infty, x_2, x_3, \ldots, x_n)$ in A'. Now delete from B all nodes, branches and leaves that are not traversed by any such $(x_2, x_3, \ldots, x_n)$, and call the resulted algorithm A''. Clearly, all leaves not in L_1 are deleted, and we have $\sharp(A'') \leq |L_1|$. This completes the proof of (3), and Theorem 1.

4 Proof of Theorem 2

A. Let $n = 2^k + 1$. To prove $V_2'(n) = n - 3 + \lceil lg\, n \rceil$, we only need, in view of Theorem 1, to give an algorithm $A \in \mathcal{A}_{n,2}$ with $C(A) = n - 2 + k$. 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 check 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 $(n - 3) + 2 + (k - 1) = n - 2 + k$.

We have proved $V_2'(n) = n - 3 + \lceil lg\, n \rceil$ for $n = 2^k + 1$.

B. Let $n = 2^k$, and $A \in \mathcal{A}_{n,2}$. We will prove $C(A) \geq n - 2 + k$ by modifying the arguments used in the proof of Theorem 1. This will establish $V_2'(n) \geq n - 2 + \lceil lg\, n \rceil$.

Case 1: The comparison at the root is $x_i - x_j : 0$. In the proof of Theorem 1, if we can show that inequality (1) can be replaced by

$$\sharp(A') < 2\sharp(A) \ , \tag{5}$$

then the derivation gives instead of (4),

$$\sharp(A) > \frac{1}{2}\, n \cdot 2^{n-2} \ ,$$

which implies that

$$C(A) > \lceil lg\, (\frac{1}{2}\, n\, 2^{n-2}) \rceil = n - 3 + k \ .$$

This would prove $C(A) \geq n - 2 + k$. It remains to prove (5). It suffices to prove that there is at least one leaf ℓ in A with $|M_{G_\ell}| = 1$. Consider any input $\tilde{x} = (x_1, x_2, \ldots, x_n) \in R_0^n$ with $x_i > x_j > x_t \ \forall\, t \neq i, j$, and let ℓ be the leaf $\tilde{x}$ reaches. Since the output a_ℓ must be j, and the comparison at the root forces $x_i > x_j$, we have $|M_{G_\ell}| = |\{i\}| = 1$.

Case 2: The comparison at the root is $(x_j - x_i)(x_i - x_t) : 0$. Again, define A' obtained from A exactly as in the proof of Theorem 1. Now, consider only the left branch at the root of A (the branch corresponding to $(x_j - x_i)(x_i - x_t) < 0$), and let L_s denote the set of leaves ℓ in this branch that have $M_{G_\ell} = \{s\}$. Then,

$$C(A) \geq C(A') - 1 \ , \tag{6}$$

and

$$C(A') \geq 1 + \lceil lg(\sum_{1 \leq s \leq n} |L_s|) \rceil \ . \tag{7}$$

We will now prove that

$$\sum_{1 \leq s \leq n} |L_s| \geq (n+1)\, 2^{n-3} \ . \tag{8}$$

It would then follow then from (6), (7), and (8) that

$$\begin{aligned} C(A) &\geq n - 3 + \lceil lg(n+1) \rceil \\ &= n - 2 + k \ . \end{aligned}$$

To prove (8), first we obtain $|L_i| \geq 2^{n-2}$ by pruning A' to produce B and A'', as in the proof of Theorem 1, and observing that there is no leaf of L_i in the right branch of A'. For $s \neq i$, we will produce from A' an algorithm A''_s with $\sharp(A''_s) \leq |L_s|$ that computes the largest of $n-2$ distinct input numbers $\{x_r | r \neq i, s\}$. This then implies by the Corollary for Lemma 3 that $|L_s| \geq 2^{n-3}$. To obtain A''_s, we set $x_s = +\infty$, $x_i = -\infty$, and prune away from A' all branches, nodes, and leaves not reachable by any input of $n-2$ distinct real numbers $\{x_r | r \neq i, s\}$. In particular, the entire right branch of A' is removed. This gives A''_s, with only leaves in L_s left. This proves (8).

We have proved $V'_2(n) \geq n - 2 + \lceil lg\, n \rceil$ for $n = 2^k$. The upper bound $V'_2(n) \leq V_2(n) = n - 2 + \lceil lg\, n \rceil$ is obviously true. This completes the proof of Theorem 2.

5 Conclusions

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

(a) Determine $V'_2(n)$ for all n.

(b) What is the complexity of finding the second largest, when tests $f : 0$ with $f \in L_n^{(3)}$ are allowed?

(c) Let $V_k^{(\ell)}(n)$ be the complexity of finding the k-th largest of n numbers, when polynomial tests of degree at most ℓ are allowed. Is there a constant $\lambda > 0$ and a function $N(k, \ell)$ such that $V_k^{(\ell)}(n) - n \geq \lambda\, k\, log\, n$ 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.

(d) Let $W_2(n)$ be the minimum number of comparisons $x_i : x_j$ needed to identify individually both the largest and the second largest among n numbers $x_1, x_2, \ldots, x_n$. It is well known that $W_2(n) = n - 2 + \lceil lg\, n \rceil$ (see [Kn]). Let $W_2'(n)$ be the analogous complexity, when median tests discussed here are allowed. It is easy to modify the arguments in this paper to show that $W_2'(n) = W_2(n) = n - 2 + \lceil lg\, n \rceil$. Is it true that we still need $n - 2 + \lceil lg\, n \rceil$ tests even if we allow any polynomial tests?

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

(f) 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.

[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.

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

[Kl] 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**, 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.