A Note on Search Trees

Jianer Chen* Iyad A. Kanj† Ge Xia*

Lemma 0.1 *Let f be a differentiable function in $(0, \infty)$ such that $f(r) = 0$ for $r \in (0, \infty)$ implies that $f'(r) > 0$. Suppose further that f' is continuous in $(0, \infty)$. Then f has at most one root in $(0, \infty)$.*

Proof. Proceed by contradiction. Suppose that f has more than one root in $(0, \infty)$. Let r_1 and r_2, with $r_1 < r_2$, be two consecutive such roots (note that we can always find two consecutive roots r_1 and r_2 because $f'(r) > 0$ for any root r). From the hypothesis, we have $f'(r_1) > 0$ and $f'(r_2) > 0$. Since f' is continuous at r_1, there exists an $0 < \epsilon < (r_2 - r_1)$ such that $f'(x) > 0$ in the interval $(r_1, r_1 + \epsilon)$. This shows that f is strictly increasing on $(r_1, r_1 + \epsilon)$, and by the choice of r_1 and r_2, it follows that $f(x) > 0$ for all $x \in (r_1, r_2)$. Similarly, since $f'(r_2) > 0$, and by continuity of f' at r_2, there exists $0 < \delta < (r_2 - r_1)$ such that $f'(x) > 0$ in $(r_2 - \delta, r_2 + \delta)$, and hence, f is strictly increasing in the interval $(r_2 - \delta, r_2 + \delta)$. Let $a \in (r_2 - \delta, r_2 + \delta)$, then $f(a) < f(r_2) = 0$. But $a \in (r_1, r_2)$, and $f(x) > 0$ for all $x \in (r_1, r_2)$, a contradiction. This completes the proof. □

Definition A branching vector, or simply, a branch, is an r-tuple $(\alpha_1, \cdots, \alpha_r)$, where $r > 1$ and $\alpha_1, \cdots, \alpha_r$ are positive real numbers with $\alpha_1 \leq \alpha_2 \leq \cdots \leq \alpha_r$.

In the literature, the notion of a branching vector, or a branch, always involves positive integers as the coordinates of the branching vector. The reason being that these coordinates represent the reduction of the parameter along each side of the branch. In this paper, we unconventionally use positive real numbers instead. It will be justified later how these coordinates are interpreted when it comes to associating them with the parameter reduction in each side of the branch.

Definition Let $(\alpha_1, \cdots, \alpha_r)$ be a branch.

(a) The characteristic function of $(\alpha_1, \cdots, \alpha_r)$ is: $x^{\alpha_r} - x^{\alpha_r - \alpha_{r-1}} - \cdots - x^{\alpha_r - \alpha_1} - 1$.

(b) The auxiliary function of $(\alpha_1, \cdots, \alpha_r)$ is: $x^{-\alpha_r} + x^{-\alpha_{r-1}} + \cdots + x^{-\alpha_1} - 1$.

Theorem 0.2 *Let $(\alpha_1, \cdots, \alpha_r)$ be a branch. The following is true.*

(a) In the interval $(0, \infty)$, the branching function $f(x) = x^{\alpha_r} - x^{\alpha_r - \alpha_{r-1}} - \cdots - x^{\alpha_r - \alpha_1} - 1$ of $(\alpha_1, \cdots, \alpha_r)$ has a unique root x_0, and $f(x) < 0$ if and only if $x < x_0$.

*Department of Computer Science, Texas A&M University, College Station, TX 77843-3112. Email: {chen, gexia}@cs.tamu.edu. Supported in part by NSF under the grants CCR-0311590 and CCF-0430683.

†The corresponding author. School of Computer Science, Telecommunications, and Information Systems, DePaul University, 243 S. Wabash Avenue, Chicago, IL 60604-2301. Email: ikanj@cs.depaul.edu. This research was supported in part by DePaul University Competitive Research Grant.

(b) *In the interval* $(0,\infty)$, *the auxiliary function* $A(x) = x^{-\alpha_r} + x^{-\alpha_{r-1}} + \cdots + x^{-\alpha_1} - 1$ *of* $(\alpha_1, \cdots, \alpha_r)$ *has a unique root* x_0, *where* x_0 *is the root of* $f(x)$, *and* $A(x) > 0$ *if and only if* $x < x_0$.

PROOF. To prove part (a) we will prove a more general statement. We will show that the statement in (a) is true for any function $g(x)$ of the form $x^h - \Sigma_{i=0}^r a_i x^{\beta_i}$, where h is a positive real number, β_i, for $i = 0 \cdots r$, are non-negative real numbers smaller than h with $\beta_0 = 0$, and a_i, $i = 0, \cdots, r$ are positive integers. Since $f(x)$ satisfies this form, the statement in (a) will follow.

Let $g(x)$ be a function as above, and suppose that s is a root of g in $(0,\infty)$. Note that g is a differentiable function in the specified interval. Moreover, $g'(x)$ is a continuous function in $(0,\infty)$. Since $g(s) = 0$, we have

$$s^h \quad = \quad \Sigma_{i=0}^r a_i s^{\beta_i}. \tag{1}$$

Now $g'(x) = hx^{h-1} - \Sigma_{i=0}^r a_i \beta_i x^{\beta_i - 1} = hx^{h-1} - \Sigma_{i=1}^r a_i \beta_i x^{\beta_i - 1}$ since $\beta_0 = 0$. Therefore $xg'(x) = hx^h - \Sigma_{i=1}^r a_i \beta_i x^{\beta_i}$. Now

$$\begin{aligned} sg'(s) &= hs^h - \Sigma_{i=1}^r a_i \beta_i s^{\beta_i} \\ &= h\Sigma_{i=0}^r a_i s^{\beta_i} - \Sigma_{i=1}^r a_i \beta_i s^{\beta_i} \qquad (2) \\ &= h[\Sigma_{i=1}^r a_i s^{\beta_i} + a_0] - \Sigma_{i=1}^r a_i \beta_i s^{\beta_i} \\ &= [\Sigma_{i=1}^r a_i (h - \beta_i) s^{\beta_i}] + ha_0. \qquad (3) \end{aligned}$$

Note that we have used equality (1) to obtain equality (2).

Since $h > \beta_i$ for every i, $s \in (0,\infty)$, and a_i is a positive integer for every i, it follows from (3) that $g'(s)$ is a positive integer. Now s was an arbitrarily chosen root of $g(x)$. This shows that $g'(s) > 0$ for every root $s \in (0,\infty)$ of g. Since the function $g(x)$ satisfies the condition of Lemma 0.1, it follows that $g(x)$ has at most one root in $(0,\infty)$. Now $g(0) = -a_0 < 0$, and since $h > \beta_i$ for every i, $g(x)$ is positive for sufficiently large x. By continuity of $g(x)$, $g(x)$ has a root in $(0,\infty)$. Consequently, g has a unique root x' in $(0,\infty)$, and satisfies $g(x) < 0$ if and only if $x < x'$. It follows that the branching function $f(x)$ has a unique root x_0 in $(0,\infty)$, and satisfies $f(x) < 0$ if and only if $x < x_0$.

The proof of part (b) follows from the fact that $A(x) = -x^{-\alpha_r} f(x)$. □

Definition Let $(a_1, \cdots, a_t)$ and $(\alpha_1, \cdots, \alpha_r)$ be two branches. The branch $(a_1, \cdots, a_t)$ is said to be not worse than $(\alpha_1, \cdots, \alpha_r)$ if the root of the characteristic function of $(a_1, \cdots, a_t)$ is not larger than that of $(\alpha_1, \cdots, \alpha_r)$.

Let $(a_1, \cdots, a_t)$ and $(\alpha_1, \cdots, \alpha_r)$ be two branches such that $(a_1, \cdots, a_t)$ is not worse than $(\alpha_1, \cdots, \alpha_r)$. We first show that there exists a unique real number s, where $0 \le s < a_1$, such that the root of the characteristic function of $(a_1 - s, \cdots, a_t - s)$ is the same as that of $(\alpha_1, \cdots, \alpha_r)$.

Define the two-variable function $f(s', x) = x^{-a_t + s'} + x^{-a_{t-1} + s'} + \cdots + x^{-a_1 + s'} - 1$, where $s' \in [0, a_1[$ and $x \in (0,\infty)$. Note that for every fixed value s_0 of s', the function $f(s_0, x)$ is the auxiliary function of the branch $(a_1 - s_0, a_2 - s_0, \cdots, a_t - s_0)$, and by part (b) of Theorem 0.2, $f(s_0, x)$ has a unique root $x_0 \in (0,\infty)$ such that $f(s_0, x) < 0$ if and only if $x > x_0$. Define the function

$z(s') : [0, a_1) \longrightarrow (0, \infty)$ that for every $s_0 \in [0, a_1)$ associates $z(s_0)$, the unique root of $f(s_0, x)$. Then by part (b) of Theorem 0.2, the function $z(s')$ is well defined. Moreover, it is straightforward to verify that $z(s')$ is strictly increasing. We show next that $z(s')$ is a continuous function. Let s_0 be a fixed point in $(0, a_1)$. We show that $z(s')$ is continuous at s_0. We only prove that $z(s')$ is continuous on the left of s_0 (i.e., at s_0^-). The proof that the function is continuous on the right of s_0 is very similar. Also, the proof that the function is continuous on the right of 0 is similar. We proceed by contradiction. Suppose that $z(s')$ is not continuous at the left of s_0. Then we can find a strictly increasing sequence $(s_i)_{i=1}^{\infty}$ with $s_i < s_0$ for every i, such that $(s_i)_{i=1}^{\infty}$ converges to s_0, but $(z(s_i))_{i=1}^{\infty}$ does not converge to $z(s_0)$. Since $z(s')$ is a strictly increasing function, and since $s_i < s_0$ for every i, it follows that the sequence $(z(s_i))_{i=1}^{\infty}$ is increasing and bounded by $z(s_0)$, and hence is convergent to a point $y \neq z(s_0)$. Now $f(s_i, z(s_i)) = 0$ for all i. By continuity of f, $f(s_0, y) = 0$. But $f(s_0, z(s_0)) = 0$ and $f(s_0, x)$ has a unique root in $(0, \infty)$. It follows that $y = z(s_0)$, a contradiction. Therefore the function $z(s')$ is a continuous function on $(0, a_1)$.

By looking at the function $f(s', x)$, it is not difficult to see that as s' approaches a_1, the root $z(s')$ of $f(s', x)$ approaches (positive) infinity, and hence $z(s')$ is unbounded. Therefore, there exists a point $s_b \in [0, a_1)$ such that $z(s_b) > r_0$, where r_0 is the root of the characteristic function of the branch $(\alpha_1, \cdots, \alpha_r)$. Moreover, since the branch $(a_1, \cdots, a_t)$ is not worse than $(\alpha_1, \cdots, \alpha_r)$, $z(s' = 0) \leq r_0$. By continuity of $z(s')$, and since $z(s')$ is strictly increasing, there exists a unique point $s \in [0, s_b[$ (and also in $[0, a_1)$) such that $z(s) = r_0$.

This shows that there exists a unique real number s, where $0 \leq s < a_1$, such that the root of the characteristic function of $(a_1 - s, \cdots, a_t - s)$ is the same as that of $(\alpha_1, \cdots, \alpha_r)$. Now we are ready to define the following notion.

Definition Let $(a_1, \cdots, a_t)$ and $(\alpha_1, \cdots, \alpha_r)$ be two branches such that $(a_1, \cdots, a_t)$ is not worse than $(\alpha_1, \cdots, \alpha_r)$. Let r_0 be the root of the characteristic function of $(\alpha_1, \cdots, \alpha_r)$. The surplus of the branch $(a_1, \cdots, a_t)$ relative to the branch $(\alpha_1, \cdots, \alpha_r)$ is the unique value $s \in [0, a_1)$ such that the root of the characteristic function of $(a_1, \cdots, a_t)$ is equal to r_0.

The reason behind defining the above notion of the surplus is twofold. First, the notion of a branch being better than another branch in the sense that the root of its characteristic polynomial (or function) is smaller than that of the other branch, can be very misleading when it comes to combining the two branches with a third branch. As an example to illustrate this point, consider the two branches $(1, 11)$ and $(4, 4)$. The branch $(1, 11)$ is better than $(4, 4)$ because the root of its characteristic polynomial is smaller than 1.185 whereas that of $(4, 4)$ is larger than 1.185. Now consider a situation where we branch with a $(1, 6)$ branch and on the 1-side of the branch we follow that with a $(1, 11)$ branch. We get the combined branch $(2, 6, 12)$ with the root of its characteristic polynomial being larger than 1.240. One intuitively is tempted to believe that since $(4, 4)$ is worse than $(1, 11)$, branching with a $(4, 4)$ branch on the 1-side of the $(1, 6)$ branch should result in a worse branch than $(2, 6, 12)$. However, this is not the case. The branch $(5, 5, 6)$ is better than $(2, 6, 12)$ with the root of its characteristic polynomial being smaller than 1.240. So if one's goal is to show that every branch in a certain algorithm is not worse than a $(2, 5)$ whose root is smaller than 1.237, and uses the $(4, 4)$ branch to combine it with the 1-side of the $(1, 6)$ branch rather than the $(1, 11)$ in the belief that the $(4, 4)$ corresponds to a worst-case scenario, he will be obtaining a $(5, 5, 6)$ branch which has a root of 1.229 better than the root of the $(2, 5)$ branch, whereas the worst-case scenario corresponding to the $(2, 6, 12)$ branch has a root that is larger than that of the $(2, 5)$ branch!

The reason why this phenomenon has taken place can be readily seen. The branch $(1, 11)$ is much more "skewed" than the balanced $(4,4)$ branch, and when combined with some other branches it may result in a worse branch. To avoid this problem, we use the notion of the surplus. When we use the notion of the surplus, the surplus of the $(4,4)$ branch is larger than the $(1,11)$ branch relative to the $(2,5)$ branch. The $(2,5)$ branch will be referred to as a *base branch.* We will show next that if the surplus of a branch is not smaller than that of another branch relative to a certain base branch, then when the first branch is combined with any third branch it results in a branch that is not worse than that when the second branch is combined with this third branch, all being relative to the base branch. This will show that when combining branches, using the surplus is safe.

The second benefit behind the notion of the surplus is that instead of combining a branch with another branch, we can simply add its surplus to the other branch. This makes the "arithmetic" of branching easier to carry out.

Theorem 0.3 *Let $(\alpha_1,\cdots,\alpha_r)$ be a base branch. Let $(a_1,\cdots,a_t)$ and $(b_1,\cdots,b_p)$ be two branches such that $(a_1,\cdots,a_t)$ has a surplus of s_1 relative to $(\alpha_1,\cdots,\alpha_r)$, and $(b_1,\cdots,b_p)$ has a surplus of s_2 relative to $(\alpha_1,\cdots,\alpha_r)$ with $s_1 \geq s_2$. Let $(c_1,\cdots,c_q)$ be any branch. Let B_1 be the branch resulting by branching with the branch $(a_1,\cdots,a_t)$ on the c_i-side of the branch $(c_1,\cdots,c_q)$, where $i \in \{1,\cdots,q\}$, and B_2 that resulting by branching with $(b_1,\cdots,b_p)$ on the c_i-side of the branch $(c_1,\cdots,c_q)$. If B_2 is not worse than $(\alpha_1,\cdots,\alpha_r)$ than neither is B_1.*

Proof. Let r_0 be the root of the characteristic function of $(\alpha_1,\cdots,\alpha_r)$. Let $A_1(x) = x^{-c_q} + x^{-c_{i+1}} + x^{-c_{i-1}} + \cdots + x^{-c_1} - 1 + x^{-c_i-a_t} + \cdots + x^{-c_i-a_1}$ be the auxiliary function of B_1 and $A_2(x) = x^{-c_q} + x^{-c_{i+1}} + x^{-c_{i-1}} + x^{-c_{i-1}} + \cdots + x^{-c_1} - 1 + x^{-c_i-b_p} + \cdots + x^{-c_i-b_1}$ be the auxiliary function of B_2. Let r_1 and r_2 be the roots of A_1 and A_2, respectively. From the hypothesis, we have $r_2 \leq r_0$. From part (b) in Theorem 0.2, it follows that:

$$r_0^{-c_q} + r_0^{-c_{i+1}} + r_0^{-c_{i-1}} + \cdots + r_0^{-c_1} - 1 + r_0^{-c_i-b_p} + \cdots + r_0^{-c_i-b_1} \leq 0. \tag{4}$$

To prove the theorem, we only need to show that:

$$r_0^{-c_q} + r_0^{-c_{i+1}} + r_0^{-c_{i-1}} + \cdots + r_0^{-c_1} - 1 + r_0^{-c_i-a_t} + \cdots + r_0^{-c_i-a_1} \leq 0. \tag{5}$$

From inequality (4), it suffices to show that $r_0^{-c_i-a_t} + \cdots + r_0^{-c_i-a_1} \leq r_0^{-c_i-b_p} + \cdots + r_0^{-c_i-b_1}$, or equivalently,

$$r_0^{-a_t} + \cdots + r_0^{-a_1} \leq r_0^{-b_p} + \cdots + r_0^{-b_1}. \tag{6}$$

Since $(a_1,\cdots,a_t)$ has a surplus s_1 relative to the base branch $(\alpha_1,\cdots,\alpha_r)$, we have $r_0^{-a_t+s_1} + \cdots + r_0^{-a_1+s_1} - 1 = 0$. Similarly, since $(b_1,\cdots,b_p)$ has a surplus of s_2 relative to the base branch, we have $r_0^{-b_p+s_2} + \cdots + r_0^{-b_1+s_2} - 1 = 0$. Therefore, $r_0^{-a_t+s_1} + \cdots + r_0^{-a_1+s_1} = r_0^{-b_p+s_2} + \cdots + r_0^{b_1+s_2}$. Now $s_1 \geq s_2$, and hence, $r_0^{-a_t+s_2} + \cdots + r_0^{-a_1+s_2} \leq r_0^{-a_t+s_1} + \cdots + r_0^{-a_1+s_1} = r_0^{-b_p+s_2} + \cdots + r_0^{-b_1+s_2}$. Consequently, $r_0^{-a_t} + \cdots + r_0^{-a_1} \leq r_0^{-b_p} + \cdots + r_0^{-b_1}$ and inequality (6) follows, and so does the theorem. $\square$

The above theorem shows that the notion of surplus captures the the intuitive notion of branch comparison. Since computing the actual value of the surplus of a branch relative to another branch, or in particular, to the base branch, is not an easy task, the following corollary provides a feasible tool for comparing branches. The proof of the corollary follows a very similar approach to that of the Theorem 0.3 with a minor tweak at the end. We leave the proof to the interested reader.

Corollary 0.4 *Let $(\alpha_1, \cdots, \alpha_r)$ be a base branch. Let $(a_1, \cdots, a_t)$ and $(b_1, \cdots, b_p)$ be two branches such that $(a_1, \cdots, a_t)$ has a surplus of value greater or equal to s_1 relative to $(\alpha_1, \cdots, \alpha_r)$, and $(b_1, \cdots, b_p)$ has a surplus smaller than s_1 relative to $(\alpha_1, \cdots, \alpha_r)$. Let $(c_1, \cdots, c_q)$ be any branch. Let B_1 be the branch resulting by branching with the branch $(a_1, \cdots, a_t)$ on the c_i-side of the branch $(c_1, \cdots, c_q)$, where $i \in \{1, \cdots, q\}$, and B_2 that resulting by branching with $(b_1, \cdots, b_p)$ on the c_i-side of the branch $(c_1, \cdots, c_q)$. If B_2 is not worse than $(\alpha_1, \cdots, \alpha_r)$ than neither is B_1.*

Corollary 0.4 provides a feasible way for showing that the branch $(a_1, \cdots, a_t)$ is less skewed than $(b_1, \cdots, b_p)$. This can be done by computing a lower bound on the surplus of $(a_1, \cdots, a_t)$ relative to the base branch—which can be done by few trial and errors on the characteristic function of the $(a_1 - s, \cdots, a_t - s)$ for different values of s using the monotonicity of this function and part (b) in Theorem 0.2–such that this lower bound is an upper bound on the surplus of $(b_1, \cdots, b_p)$ relative to the base branch, which can also be showed in a similar fashion.

One thing is left to be explained. When we have a branch of the form $(a_1, \cdots, a_t)$, where $a_1, \cdots, a_t$ are positive real numbers, how do we interpret this branch in conjunction with the notion of a search tree? In the standard notion of a branch, the coordinates of a branch correspond to the reduction of the parameter along each side of the branch. This notion makes sense when these coordinates are positive integers. However, this may no longer be the case. To illustrate how such a branch can be interpreted, consider a branch of the form $(n_1 + s, n_2, \cdots, n_p)$, where $n_1, \cdots, n_p$ are positive integers, and s is a real positive number which is the surplus resulting from branching with a branch $(b_1, \cdots, b_q)$ of surplus at least s (relative to a standard branch $(\alpha_1, \cdots, \alpha_r)$), on the n_1-side of the branch $(n_1, n_2, \cdots, n_p)$. When it comes to interpreting the branch $(n_1 + s, n_2, \cdots, n_p)$ with respect to the notion of the search tree and the reduction in the parameter, we look at this branch as an $(n_1, n_2, \cdots, n_p)$ branch followed by a $(b_1, \cdots, b_q)$ branch on the n_1-side. Now to calculate an upper bound on the size of the search tree, we can simply use the root of the characteristic function of the branch $(n_1 + s, n_2, \cdots, n_p)$. The reason being that, by using a similar proof to that of Theorem 0.3, it can be shown that the root of this characteristic function is not smaller than that of the characteristic polynomial of the two combined integer-coordinate branches. So an upper bound on the root of this characteristic function implies an upper bound on the root of the characteristic polynomial of the combined branches. Therefore, hypothetically, we can look at the branch $(a_1, \cdots, a_t)$ as a branch with real-valued coordinates and use the root of its characteristic function to compute an upper bound on the size of the search tree. However, such a branch should be understood as a standard branch with integer coordinates, with the accumulation of certain surplus along some of its sides.