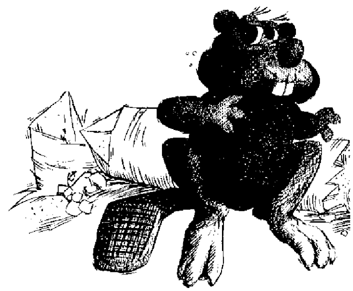
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT
UNIVERSITY OF WATERLOO COMPUTER SCIENCE DEPARTMENT



Expected Behaviour Analysis of AVL Trees

Gaston H. Gonnet
Nivio Ziviani

CS-82-18

June, 1982

Expected Behaviour Analysis of AVL Trees

Gaston H. Gonnet
Nivio Ziviani †

Department of Computer Science
University of Waterloo
Waterloo, Ontario
Canada

ABSTRACT

In this paper we use a fringe analysis method based on a new way of describing the composition of a fringe in terms of tree collections. We present a closed tree collection of AVL trees containing three types and obtain bounds on the expected number of rotations per insertion and on the expected number of balanced nodes. A new way of handling larger tree collections that are not closed is presented. An inherent difficulty posed by the transformations necessary to keep the AVL tree balanced makes its analysis difficult when using fringe analysis methods. We derive a technique to cope with this difficulty and again obtain bounds on the expected number of rotations per insertion and on the expected number of balanced nodes.

Key phrases: Analysis of algorithms, fringe analysis, AVL trees, number of rotations per insertion, number of balanced nodes.

The work of the first author was supported by a Natural Sciences and Engineering Research Council of Canada Grant No. A-3353, and the second by a Brazilian Coordenação do Aperfeicoamento de Pessoal de Nivel Superior Contract No. 4799/77 and by the University of Waterloo.

† Permanent address:

Departamento de Ciência da Computação, UFMG, Belo Horizonte MG 30000, Brazil

Expected Behaviour Analysis of AVL Trees

Gaston H. Gonnet
Nivio Ziviani

1. Introduction

Balanced tree structures are efficient ways of storing information. They provide an excellent solution for the dictionary data structure problem. For a linear list of length N the operations find, insert, and delete can be done in $O(\log N)$ units of time. The most popular are AVL trees.

AVL trees were introduced by Adel'son-Vel'skii and Landis (1962). A binary search tree is AVL if the height of the subtrees at each node differ by at most one. A balance field in each node can indicate this with two bits: $+1$, higher right subtree; 0, equal heights; -1, higher left subtree.

The process of insertion of a new key consists of three parts:

(i) Follow the search path until it is verified that the key is not in the tree (i.e., find the place of insertion).

(ii) Insert the new node and set its balance field to 0.

(iii) Retreat along the search path and check the balance field at each node. At this point a transformation may be necessary, as described below.

In phase 3 rebalancing occurs if the balance field indicates that the node becomes more unbalanced with the insertion (occurs when the direction of the search path and the present balance coincide). In this case a single or double rotation occurs, depending on the balance field of the node and on the balance field of its son, which is along the search path. Figure 3.2.2, in Section 3.2, illustrates the AVL tree transformations. As the height of the rotated subtree is the same as the height of the subtree before the insertion, at most one rotation per insertion is necessary. Of course if the balance field indicates that the subtree becomes less unbalanced, a modification of the balance field is sufficient.

The first valuable attempt to analyse a balanced search tree was performed by Yao (1978). In his work Yao presented a method which he used to obtain a partial analysis of 2-3 trees and B-trees. The method used by Yao (1978) was later used by Brown (1979) to obtain a partial analysis of AVL trees. In his analysis Brown considered the collection of AVL subtrees with three or less leaves and called it the fringe of the AVL tree. By analysing the fringe of large AVL trees Brown was able to derive bounds on the expected number of balanced nodes in the whole tree.

An improvement on Brown's results for AVL trees was obtained by

Mehlhorn (1979a), through the study of 1-2 brother trees (Ottmann and Six, 1976). The main technical contribution of Mehlhorn's paper is a method for analysing the behaviour of 1-2 brother tree schemes where the rebalancing operations require knowledge about the brother of a node. Using the close relationship between 1-2 brother trees and AVL trees (Ottmann and Wood, 1979), Mehlhorn was able to improve the bounds on the expected number of balanced nodes in AVL trees. Mehlhorn (1979b) presented a fringe analysis of AVL trees under random insertions and deletions.

To improve the results on AVL trees we need larger tree collections. However, the use of larger AVL tree collections represents a complex problem. An inherent difficulty posed by the transformations necessary to maintain the AVL trees balanced makes its fringe analysis quite difficult. (cf. Section 4.) In Section 5 we present a technique to cope with this difficulty which permits us to obtain bounds on the expected number of balanced nodes and the expected number of rotations per insertion.

Consider an AVL tree T with N keys and consequently $N+1$ external nodes. These N keys divide all possible key values into $N+1$ intervals. An insertion into T is said to be a *random insertion* if it has an equal probability of being in any of the $N+1$ intervals defined above. A *random AVL tree* with N keys is an AVL tree constructed by making N successive random insertions into an initially empty tree. In this paper we assume that all trees are random trees.

We now define certain complexity measures:

(i) Let $\bar{b}(N)$ be the expected number of balanced nodes in an AVL tree after the random insertion of N keys into the initially empty tree;

(ii) Let $r(N)$ be the expected number of rotations required during the insertion of the $(N+1)^{st}$ key into a random AVL tree with N keys;

(iii) Let $Pr\{no\ rotation\}$ be the probability that no rotation occurs during the $(N+1)^{st}$ random insertion into a random AVL tree with N keys;

(iv) Let $m(N)$ be the maximum number of rotations that may occur outside the fringe of an AVL tree during the insertion of the $(N+1)^{st}$ key into a random AVL tree with N keys;

(v) Let $\bar{u}(N)$ be the expected number of unbalanced nodes in an AVL tree after the random insertion of N keys into the initially empty tree;

(vi) Let $\bar{f}(N)$ be the expected number of nodes in the fringe of an AVL tree after the random insertion of N keys into the initially empty tree.

In Section 2 we present the fringe analysis technique used to analyse AVL trees. In Section 3 we present a new closed AVL tree collection which improves the lower bound on the expected number of balanced nodes. In

Section 4 we study weakly-closed AVL tree collections, and in Section 5 we present a technique to deal with weakly-closed AVL tree collections. In Section 6 we present larger weakly-closed AVL tree collections and discuss the problems involved in their analyses. Finally we present some equivalent results for weight-balanced trees.

Table 1.1 shows the summary of the results related to AVL trees.

Tree Collection		$\mathcal{J}(N)$	$r(N)$	$\frac{\bar{\delta}(N)}{N}$
Size	Characteristic			
2	closed	$0.57N$ for $N \geq 6$	$[0.29, 0.86]$ for $N \geq 6$	$[0.48+0.48/N, 0.86-0.14/N]$ for $N \geq 6$
3	closed ambiguous	$0.66N$ for $N \geq 6$	$[0.29, 0.86]$ for $N \geq 6$	$[0.51+0.51/N, 0.86-0.14/N]$ for $N \geq 6$
4 †	weakly-closed ambiguous	$0.69N$	$[0.29, 0.81]$	$[0.51+0.51/N, 0.81-0.19/N]$

† Results are approximated to $O(N^{-11/3})$

Table 1.1 Summary of AVL tree results

2. Fringe Analysis Technique

In the first part of this section we introduce the concepts and the definitions necessary to describe the Markov chain used to model the insertion process in search trees. In the second part we study the matrix recurrence relation involved in the Markov process.

2.1. The Markov Process

Let us define a *tree collection* as a finite collection $C = \{T_1, \cdots, T_m\}$ of trees. The collection of AVL trees with three leaves or fewer forms a tree collection, as shown in Figure 2.1.1.

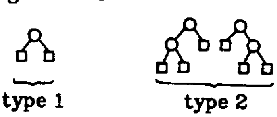


Fig. 2.1.1 Tree collection of AVL trees with three leaves or fewer

The *fringe* of a tree consists of one or more subtrees that are isomorphic to members of a tree collection C. Typically, the fringe will contain all subtrees that meet this definition; for example the fringe of an AVL tree that corresponds to the tree collection of Figure 2.1.1 is obtained by deleting all nodes at a distance greater than 2 from the leaves. Figure 2.1.2 shows an instance of an AVL tree with eleven keys in which the fringe that corresponds to the tree collection of Figure 2.1 is encircled.

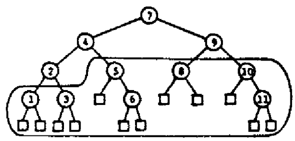


Fig. 2.1.2 An AVL tree and its fringe that corresponds to the tree collection of Figure 2.1.1

The composition of the fringe can be described in several ways. One possible way is to consider the probability that a randomly chosen leaf of the tree belongs to each of the members of the corresponding tree collection. In other words, the probability p is

$$p_i(N) = \frac{\textit{Expected number of leaves of type } i \textit{ in a } N\textit{-key tree}}{N+1} \qquad (1)$$

Yao (1978) describes the fringe in a different way. His description of the composition of the fringe considers the expected number of trees of type i, while we describe it in terms of leaves as in Eq.(1). As we shall see our description of the composition of the fringe simplifies the notation necessary to present the fringe analysis technique, and also makes easier the task of finding which complexity measures can be obtained from the analysis of each search tree.

In fringe analysis problems we always deal with a collection $C = \{T_1, \ldots, T_m\}$ of trees. We now introduce some concepts about the fringes of search trees.

Def. 2.1.1. A tree collection $C = \{T_1, \ldots, T_m\}$ is *weakly-closed* if for all $j \in [1,...,m]$ an insertion into T_j always leads to one or more T_i, $i \in [1,...,m]$.

Def. 2.1.2. A tree collection $C = \{T_1, \ldots, T_m\}$ is *closed* when (i) C is weakly-closed and (ii) the effect of an insertion on the composition of the fringe is determined only by the subtree of the fringe where the insertion is performed.

The tree collection of Figure 2.1.1 is an example of a closed tree collection. Brown (1979) proved that this tree collection is closed. On the other hand the collection of AVL trees with more than 2 and fewer than 7 leaves (see Figure 2.1.3) is not closed. This is because an insertion into a type 2 tree of Figure 2.1.3, when the type 2 tree is part of the fringe of an AVL tree, may cause a rotation higher in the tree, and the composition of the fringe depends on this rotation at the higher level. Figure 2.1.4 shows an instance of an AVL tree where an insertion into a type 2 tree does not lead to a type 3 tree as expected.

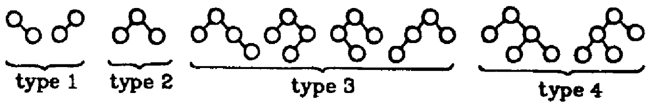


Fig. 2.1.3 Tree collection of AVL trees with more than 2 and less than 7 leaves (leaves not shown)

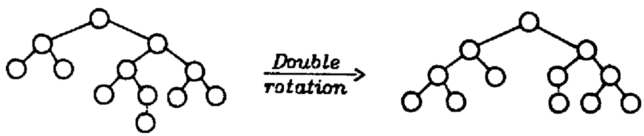


Fig. 2.1.4 Example of an insertion that unexpectedly changes the fringe of an AVL tree (dotted edge shows the point of insertion)

Def. 2.1.3. A tree collection $C = \{T_1, \ldots, T_m\}$ is *ambiguous* when a tree in C appears as a subtree of another tree in C. Figure 2.1.5 shows an AVL tree collection that is ambiguous, since a tree of type 1 is a subtree of trees of type 3.

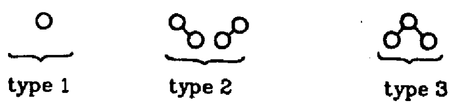


Fig. 2.1.4 Tree collection of AVL trees with more than 1 and less than 5 leaves (leaves not shown)

The transitions between trees of a tree collection can be used to model the insertion process. In an insertion of a key into the type 1 tree shown in Figure 2.1.1 two leaves of type 1 are lost and three leaves of type 2 are obtained. In an insertion of a key into the type 2 tree three leaves of the type 2 are lost and four leaves of the type 1 tree are obtained.

Clearly the probability that an insertion in one type of a tree collection C leads to another type of C depends only on the two types involved, and so the process is a Markov process (cf. Cox and Miller, 1965; Feller, 1968). A sequence $\{X_N\} = \{X_0, X_1, \cdots\}$ of random variables taking values on a state space S is a Markov chain if

$$Pr\{X_N = i \mid X_{N-1} = j, X_{N-2} = j_1, \cdots, X_0 = j_{N-1}\} = Pr\{X_N = i \mid X_{N-1} = j\}$$

for all $i, j, j_1, \cdots, j_{N-1} \in S$. The current value of X_N depends on the history of the process only through the most recent value X_{N-1}.

To illustrate this fact consider the tree collection of AVL trees shown in Figure 2.1.1. In this context, let X_N and Y_N be respectively the numbers of type 1 and type 2 leaves after the N^{th} insertion. Since the tree collection is closed, the value of X_N depends only on the value of X_{N-1} and as a consequence $\{X_N\}$ (or equivalently $\{Y_N\}$) is a Markov chain.

The transition probabilities of the chain $\{X_N\}$ are given by

$$Pr\{X_N = i \mid X_{N-1} = j\} = \begin{cases} \dfrac{j}{N} & i = j-2 \\ \dfrac{N-j}{N} & i = j+4 \end{cases}$$

while those of Y_N are

$$Pr\{Y_N = i \mid Y_{N-1} = j\} = \begin{cases} \dfrac{j}{N} & i = j-3 \\ \dfrac{N-j}{N} & i = j+3 \end{cases}$$

Let $j_N = E(X_N)$ and $k_N = E(Y_N)$. Then

$$\begin{aligned} j_N = E(X_N) &= E[E(X_N \mid X_{N-1}, Y_{N-1})] \\ &= E\left[\frac{X_{N-1}}{N}(X_{N-1}-2) + \frac{Y_{N-1}}{N}(X_{N-1}+4)\right] \\ &= j_{N-1} - \frac{2}{N} j_{N-1} + \frac{4}{N} k_{N-1} \end{aligned}$$

and similarly

$$k_N = k_{N-1} - \frac{3}{N} k_{N-1} + \frac{3}{N} j_{N-1} .$$

But, by definition

$$j_{N-1} = N p_1(N-1); \quad j_N = (N+1) p_1(N);$$

$k_{N-1} = Np_2(N-1). \quad k_N = (N+1)p_2(N).$

Substituting these equations into the previous equations we get

$$p_1(N) = \frac{(N-2)p_1(N-1)+4p_2(N-1)}{N+1}$$

and

$$p_2(N) = \frac{3p_1(N-1)+(N-3)p_2(N-1)}{N+1}$$

In matrix notation

$$\begin{bmatrix} p_1(N) \\ p_2(N) \end{bmatrix} = \begin{bmatrix} \frac{N-2}{N+1} & \frac{4}{N+1} \\ \frac{3}{N+1} & \frac{N-3}{N+1} \end{bmatrix} \begin{bmatrix} p_1(N-1) \\ p_2(N-1) \end{bmatrix}$$

or

$$\begin{bmatrix} p_1(N) \\ p_2(N) \end{bmatrix} = \left[I + \frac{H}{N+1} \right] \begin{bmatrix} p_1(N-1) \\ p_2(N-1) \end{bmatrix}$$

where $H = \begin{bmatrix} -3 & 4 \\ 3 & -4 \end{bmatrix}$ and $I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

Thus the probability of an insertion occurring in each of the subtrees of the fringe can be obtained from the steady state solution of a matrix recurrence relation in a Markov chain. In general, let $p(N)$ be an m-component column vector containing $p_i(N)$. Then

$$p(N) = \left[I + \frac{H}{N+1} \right] p(N-1) \tag{2}$$

where I is the $m \times m$ identity matrix, and H is the transition matrix.

Extensions to other tree collections with more than two types requires consideration of a vector process $\{\vec{X}_N\}$ where X_{jN} is equal to the number of type j leaves at time N.

2.2. The Matrix Recurrence Relation

In fringe analysis problems we always deal with a tree collection $C = \{T_1, \ldots, T_m\}$ of trees. Let L_i be the number of leaves of T_i. An insertion into the k^{th} leaf, $k \in [1, \ldots, L_j]$, of T_j will generate $l_{ij}(k)$ leaves of type T_i. Let $p_i(N)$ be defined as in Eq.2.1-1. Then Eq.2.1-2 can be written as

$$p(N) = \left[I + \frac{H_2 - H_1 - I}{N+1}\right] p(N-1) \qquad (1)$$

where

$$H_2 = \left[\frac{1}{L_j} \sum_{k=1}^{L_j} l_{ij}(k)\right]_{1 \le i,j \le m}, \qquad H_1 = \mathrm{diag}(L_1, \ldots, L_m),$$

and I is the $m \times m$ identity matrix.

Def. 2.2.2. Consider a fringe analysis problem. Eq.(1) is the associated recursion equation, where $H = H_2 - H_1 - I = (h_{ij})$ is its transformation matrix. We have

$$h_{ij} = \frac{1}{L_j} \sum_{k=1}^{L_j} l_{ij}(k) - \delta_{ij}(L_j + 1)$$

where δ_{ij} is the Kronecker symbol.

Intuitively, the elements in the diagonal of H represent the number of leaves lost due to an insertion minus one, and off diagonal elements represent the number of leaves obtained for each type times the probability that each type is reached in a transition.

Def. 2.2.3. A fringe analysis is *connected* if there is an $l \in [1 \ldots m]$ such that $\det(H_{ll}) \ne 0$, where H_{ll} is matrix H with the l^{th} column and l^{th} row deleted.

The following theorem shows that the real part of the eigenvalues of the transition matrix H are non-positive. The proof of this theorem and all the following theorems may be found in Ziviani (1982), or in Eisenbarth, Ziviani, Gonnet, Mehlhorn and Wood (1982).

Theorem 2.2.1. Consider a connected fringe analysis problem with a $m \times m$ transition matrix H as in Definition 2.2.2. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of H. We can order them so that $\lambda_1 = 0$ and $0 > \mathrm{Re}\lambda_2 \ge \mathrm{Re}\lambda_3 \ge \cdots \ge \mathrm{Re}\lambda_m$.

Def. 2.2.4. Let $T_j \to T_i$ if $\sum_{k=1}^{L_j} l_{ij}(k) > 0$, i.e. T_j can produce T_i. The symbol $\overset{*}{\to}$ is the reflexive transitive closure of $\to$.

The following theorem describes a test for connectedness.

Theorem 2.2.2. A fringe is connected if and only if there is a T_i such that $T_j \overset{*}{\to} T_i$ for all $j \in [1 \ldots m]$.

It remains to solve Eq.(1) for connected fringe analysis problems. In a previous version of the proof of the convergence of the matrix recurrence relation (Gonnet, Ziviani, and Wood, 1981, Lemma 2.1, p.4) the eigenvalues of the transition matrix are assumed to be pairwise distinct. The following theorem extends the proof to the general case.

Theorem 2.2.3. Let H be the $m \times m$ transition matrix of a connected fringe analysis problem. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of H, where $\lambda_1 = 0 > \mathrm{Re}\lambda_2 \geq \mathrm{Re}\lambda_3 \geq \ldots \geq \mathrm{Re}\lambda_m$, and let q be the right eigenvector of H corresponding to $\lambda_1 = 0$. Then for every vector $p(0)$ there is a c such that

$$|p(N) - cq| = O(N^{\mathrm{Re}\lambda_2})$$

where $p(N)$ is defined by Eq.(1).

It is important to note that:

(i) Consider an $m \times m$ transition matrix H of a connected fringe analysis problem. Theorem 2.2.3 says that $p(N)$, the m-component column vector solution of Eq.(1), converges to the solution of

$$Hq = 0\,, \quad as\ N \to \infty \tag{2}$$

where q is also an m-component column vector that is independent of N, and

$$p(N) = \alpha_1 q + O(N^{\mathrm{Re}\lambda_2}) \tag{3}$$

q is the right eigenvector of H corresponding to eigenvalue $\lambda_1 = 0$. Furthermore, the eigenvalues of H do not need to be pairwise distinct.

(ii) Let $A_i(N)$ be the expected number of trees of type i in a random search tree with N keys. Let L_i be the number of leaves of the type i tree. We observe that Eq.2.1-1 can be written as

$$p_i(N) = \frac{A_i(N) L_i}{N+1} \tag{4}$$

3. Closed AVL Tree Collections

The only previously known closed tree collection for AVL trees is the one composed of trees with three leaves or fewer. This tree collection is studied in Section 3.1. In Section 3.2 we present a new closed tree collection for AVL trees composed of trees with four leaves or fewer.

3.1. Tree Collection of AVL Trees with Three Leaves or Less

The tree collection of AVL trees with three leaves or fewer is shown in Figure 3.1.1. Brown (1979) proved that this tree collection is closed, obtained bounds on the expected number of balanced nodes, and gave a lower bound on the expected number of rotations. For completeness we derive again the results obtained by Brown and present also an upper bound on the expected number of rotations.

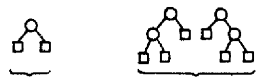


Fig. 3.1.1 Tree collection of AVL trees with three leaves or fewer

For the AVL tree collection shown in Figure 3.1.1 $H = \begin{pmatrix} -3 & 4 \\ 3 & -4 \end{pmatrix}$. From Eq.2.2-2 we have $H\,p(N) = 0$, and therefore $p_1(\infty) = 4/7$, and $p_2(\infty) = 3/7$. Since the eigenvalues of H are 0 and -7, we observe that $p_1(N) = 4/7$, and $p_2(N) = 3/7$, for $N \geq 6$. To simplify notation $p_i(N)$ is written as p_i throughout the remainder of this paper.

Lemma 3.1.1. The expected number of rotations in a random AVL tree with N keys is bounded above by

(*i*) $r(N) = 1 - Pr\{no\ rotation\}$

and

(*ii*) $r(N) \leq r(N)$ in the fringe $+\, m(N)$

Proof : For case (i) it is known that the maximum number of rotations per insertion in an AVL tree is 1. For case (ii) $r(N)$ must be less than or equal to the number of rotations per insertion in the fringe plus all possible rotations per insertion that may occur outside the fringe. ▪

Theorem 3.1.2. The expected number of rotations in a random AVL tree with N keys is bounded by

(*i*) $\frac{2}{3} p_2 \leq r(N) \leq 1 - \frac{1}{3} p_2$ for $N \geq 1$

and

(*ii*) $\frac{2}{3} p_2 \leq r(N) \leq \frac{2}{3} p_2 + p_1$ for $N \geq 1$

Proof : The left hand side of (i) and (ii) are obtained by observing Figure 3.1.1. The right hand side of (i) and (ii) are obtained by using Lemma 3.1.1. ▪

Corollary. $\frac{2}{7} \le r(N) \le \frac{6}{7}$ for $N \ge 6$

Lemma 3.1.3. The expected number of single rotations ($sr(N)$) in a random AVL tree with N keys is bounded by

$$\frac{1}{3}p_2 \le sr(N) \le 1-\frac{1}{3}p_2 \quad \text{for } N \ge 1$$

Proof : The above expression can be obtained by observing Figure 3.1.1 and by using Lemma 3.1.1. ▪

Corollary. $\frac{1}{7} \le sr(N) \le \frac{6}{7}$ for $N \ge 6$

Lemma 3.1.4. The expected number of double rotations ($dr(N)$) in a random AVL tree with N keys is bounded by

$$\frac{1}{3}p_2 \le dr(N) \le 1-\frac{1}{3}p_2 \quad \text{for } N \ge 1$$

Proof : Similar to the proof of Lemma 3.1.3. ▪

Corollary. $\frac{1}{7} \le dr(N) \le \frac{6}{7}$ for $N \ge 6$

Lemma 3.1.5. The expected number of nodes in the fringe of an AVL tree with N keys that corresponds to the tree collection of Figure 3.1.1 is

$$\bar{f}(N) = \left(\frac{p_1}{L_1}+2\frac{p_2}{L_2}\right)(N+1) \quad \text{for } N \ge 1$$

Proof : From Eq.2.2-4 we have $\bar{f}(N) = A_1(N)+2A_2(N)$. ▪

Corollary. $\bar{f}(N) = \frac{4}{7}N+\frac{4}{7}$ for $N \ge 6$

Lemma 3.1.6. The expected number of balanced nodes in a random AVL tree with N keys is bounded above by

(i) $\bar{b}(N) = N-\bar{u}(N)$ for $N \ge 1$

and

(ii) $\bar{b}(N) \le \bar{b}(N)$ in the fringe + $[N-\bar{f}(N)]$ for $N \ge 1$

Proof : For case (i) $\bar{b}(N)+\bar{u}(N) = N$. For case (ii) $\bar{b}(N)$ must be less than or equal to the number of balanced nodes in the fringe plus all nodes outside the fringe. ▪

Theorem 3.1.7. The expected number of balanced nodes in a random AVL tree with N keys is bounded by

$$\left(\frac{p_1}{L_1}+\frac{p_2}{L_2}\right)(N+1) \le \bar{b}(N) \le N-\frac{p_2}{L_2}(N+1) \quad \text{for } N\ge 1$$

Proof : The left hand side is obtained by observing Figure 3.1.1 and by using Eq.2.2-4. The right hand side is obtained by using Lemma 3.1.6, by observing Figure 3.1.1, and by using Eq.2.2-4. ▪

Corollary. $\frac{3}{7}+\frac{3}{7N} \le \frac{\bar{b}(N)}{N} \le \frac{6}{7}-\frac{1}{7N}$ for $N\ge 6$

Brown (1979, p.40) showed that an improvement on the lower bound of the result of Theorem 3.1.7 can be obtained by observing that, when the number of type 1 trees is greater than the number of type 2 trees, then at least $(\frac{p_1}{L_1}-\frac{p_2}{L_2})(N+1)/3$ balanced nodes lie outside the fringe. Thus

$$\bar{b}(N) \ge \left(\frac{p_1}{L_1}+\frac{p_2}{L_2}\right)(N+1)+\frac{1}{3}\left(\frac{p_1}{L_1}-\frac{p_2}{L_2}\right)(N+1) \quad \text{for } N\ge 1$$

or

$$\bar{b}(N) \ge \frac{10}{21}N+\frac{10}{21}$$

3.2. Tree Collection of AVL Trees with Four Leaves or Less

To improve the results obtained in the previous section we need larger tree collections. A tree collection with three types is shown in Figure 3.2.1. The first step necessary to perform the analysis is to show that the AVL tree collection of Figure 3.2.1 is closed. (cf. Definition 2.1.2.)

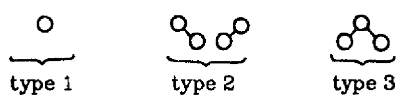
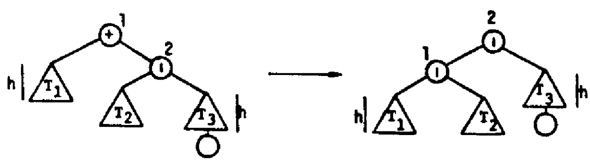


Fig. 3.2.1 Tree collection of AVL trees with four leaves or fewer

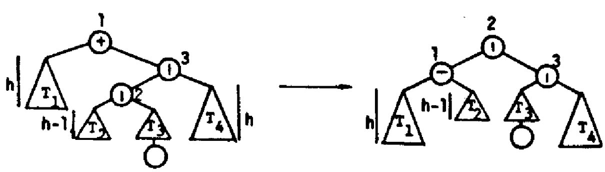
Theorem 3.2.1. The AVL tree collection shown in Figure 3.2.1 is closed.

Proof : An insertion into the type 1 tree always leads to a type 2 tree, and an insertion into the type 2 tree always leads to a type 3 tree. An insertion into the type 3 tree may cause a transformation higher in the tree, since the root of a type 3 tree is balanced. By inspecting Figure 3.2.2 we can see that a transformation has no effect on the nodes which are outside the transformed subtree. Furthermore, if the fringe of the transformed subtree is entirely contained in the subtrees T_1, T_2, T_3 and T_4 of Figure 3.2.2(b) then the transformation has no effect on the composition of the fringe. (i.e. T_1, T_2, T_3 and T_4 are moved without change by the transformation.)

However, there are six cases in which the fringe of the transformed subtree is moved with change by the transformation, as shown in Figure 3.2.3. In all six cases the number of type 3 trees decreases by one and the number of type 1 and type 2 trees increases by one. Note that each one of the three transformed trees shown in Figure 3.2.3(a and b) contains one 3-nodes subtree which is not considered as a type 3 tree, but as a subtree composed of two type 1 trees. ▪



(a) Single rotation



(b) Double rotation

Fig.3.2.2 AVL tree transformations (symmetric transformations occur)

Theorem 3.2.1 says that the transitions in the tree collection of Figure 3.2.1 are well-defined, so the theorems of Section 2 can be applied. Thus

$$H = \begin{pmatrix} -3 & 0 & 2 \\ 3 & -4 & 3 \\ 0 & 4 & -5 \end{pmatrix}.$$

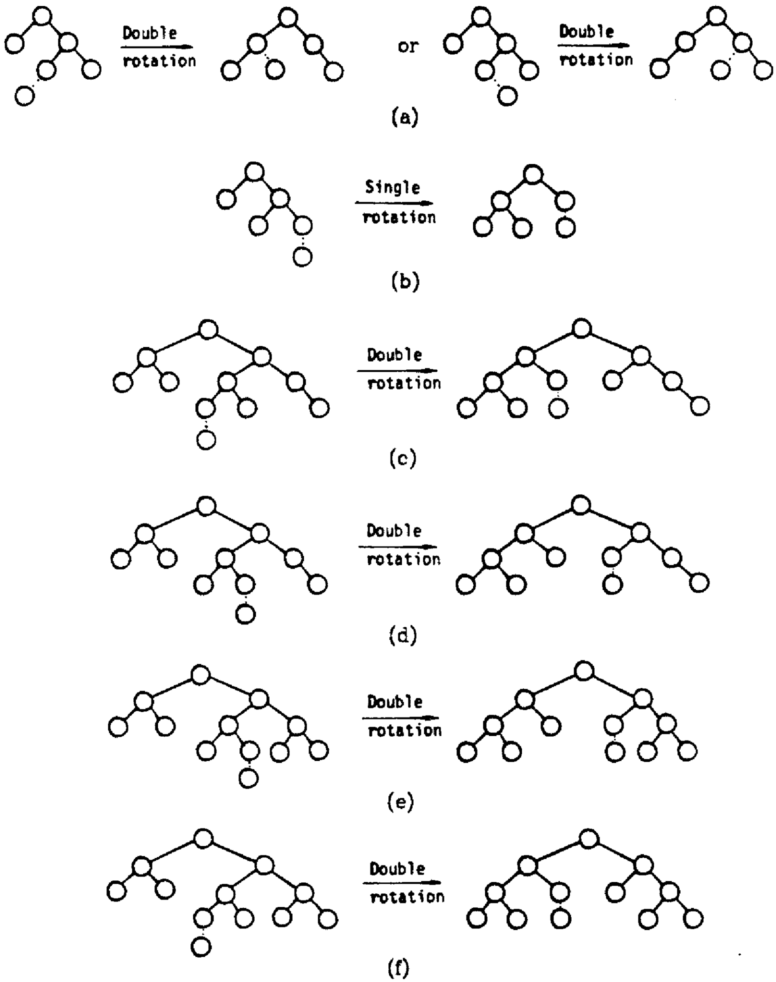


Fig.3.2.3 Cases in which the fringe of the transformed subtree is moved with change (symmetric transformations occur)

From Eq.2.2.-2 we have $Hp(N) = 0$, and therefore $p_1(\infty) = 8/35$, $p_2(\infty) = 15/35$, and $p_3(\infty) = 12/35$. Since the eigenvalues of H are 0, −5, and −7, we observe that $p_1(N) = 8/35$, $p_2(N) = 15/35$, and $p_3(N) = 12/35$, for $N \geq 6$.

Theorem 3.2.2. The expected number of balanced nodes in a random AVL tree with N keys is bounded by

$$\left(\frac{p_1}{L_1} + \frac{p_2}{L_2} + 3\frac{p_3}{L_3}\right)(N+1) \leq \bar{b}(N) \leq N - \frac{p_2}{L_2}(N+1)$$

Proof : The left hand side is obtained by observing Figure 3.2.1 and by using Eq.2.2-4. The right hand side is obtained by using Lemma 3.1.6, by observing Figure 3.2.1, and by using Eq.2.2-4. ▪

Corollary. $\frac{18}{35} + \frac{18}{35N} \leq \frac{\bar{b}(N)}{N} \leq \frac{6}{7} - \frac{1}{7N}$, for $N \geq 6$

Lemma 3.2.3. The expected number of nodes in the fringe of an AVL tree with N keys that corresponds to the tree collection of Figure 3.2.1 is

$$\bar{f}(N) = \left(\frac{p_1}{L_1} + 2\frac{p_2}{L_2} + 3\frac{p_3}{L_3}\right)(N+1)$$

Proof : From Eq.2.2-4 we have $\bar{f}(N) = A_1(N) + 2A_2(N) + 3A_3(N)$. ▪

Corollary. $\bar{f}(N) = \frac{23}{35}(N+1)$, for $N \geq 6$

The results on the expected number of rotations derived in the previous section cannot be improved by the use of this tree collection. This tree collection corresponds to the tree collection used in the previous section augmented by the type 3 tree, and the type 3 tree does not contain any information about rotations.

4. Weakly-closed AVL Tree Collections

If the effect of an insertion on the composition of the fringe is determined not only by the subtree of the fringe where the insertion is performed, but by some other transformation that may happen outside the fringe, then the tree collection is weakly-closed (Definition 2.1.2). We will show that the tree collection of AVL trees with five or less leaves shown in Figure 4.1 is not closed.

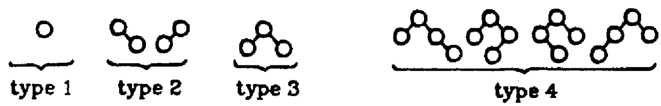


Fig. 4.1 Tree collection of AVL trees with five or less leaves

Lemma 4.1. If the trees shown in Figure 4.1 form the fringe of a random AVL tree with N keys and $N \to \infty$, then an insertion into a leaf of a type 3 tree (i) decreases by one the number of type 3 trees and increases by one the number of type 4 trees; or (ii) decreases by one the number of type 3 trees and increases by one the number of type 1 and type 2 trees; or (iii) decreases by one the number of type 1 trees and increases by one the number of type 2 trees.

Proof : We will denote the probability of the second of these alternatives by s_N, the probability of the third by t_N, and the probability of the first by $1-s_N-t_N$.

Case (i): This case is obvious: the type 3 tree is transformed in a type 4 tree. If there is no transformation higher in the tree then this is the transition.

Cases $(ii)$and(iii): If a transformation takes place higher in the tree, which is possible since the root of a type 3 tree is balanced, side-effects on the composition of the fringe will occur. By inspecting Figure 3.2.2 we can see that a transformation has no effect on the nodes which are outside of the transformed subtree. Furthermore, if the fringe of the transformed subtree is entirely contained in the subtrees T_1, T_2, T_3, and T_4 of Figure 3.2.2 then the transformation has no effect on the composition of the fringe, because T_1, T_2, T_3, and T_4 are moved without change by the transformation.

Figure 4.2 shows the five cases in which the fringe of the tree to be transformed is moved with change by the transformation, and this change produces side-effects on the composition of the fringe.

In cases (a) and (b) of Figure 4.2 the number of type 1 trees decreases by one and the number of type 2 trees increases by one. This case occurs with an unknown probability we call t_N.

In cases (c), (d), and (e) of Figure 4.2 the number of type 3 trees decreases by one and the number of type 1 and type 2 trees increases by one. This case occurs with an unknown probability we call s_N. ▪

Lemma 4.1 tells us that any AVL tree collection that contains the types 3 and 4 shown in Figure 4.1 is not closed. (i.e. it is weakly-closed.) In fact it

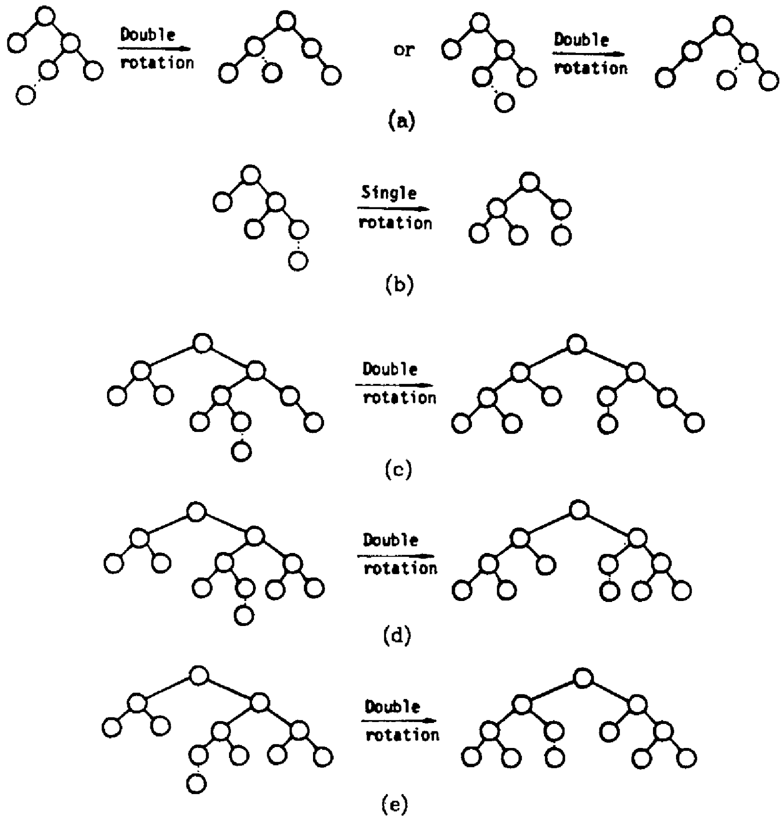


Fig. 4.2 Cases in which transformations change the fringe
(Symmetric transformations occur)

is not difficult to show that every AVL tree type that contains more than one internal node and has its root node balanced suffers from the same type of misbehaviour that occurs with type 3 (i.e. consider the AVL tree with six nodes). Consequently an AVL tree collection that contains a tree type with the root node balanced and has more than three types is weakly-closed.

We know from Lemma 4.1 that an insertion into the type 3 tree shown in Figure 4.1, when it belongs to a fringe of an AVL tree with N keys, produces a transition that is not well defined: the transition depends on two unknown probabilities s_N and t_N which also depend on N. First of all let us give a more precise meaning to s_N and t_N. Let I be the expected number of leaves in an AVL tree with N keys such that an insertion in one of the I leaves causes one of the three transformations shown in Figure 4.2(c, d, and e). In a similar way let J be the expected number of leaves such that an insertion in one of the J leaves causes one of the two transformations shown in Figure 4.2(a and b). Thus

$$s_N = \frac{I}{N+1}$$

and

$$t_N = \frac{J}{N+1} .$$

Although the probabilities s_N and t_N are unknown they cannot assume arbitrary values between 0 and 1.

Lemma 4.2. The probability t_N is bounded by $0 \le t_N \le \frac{1}{3}$

Proof :

Case (i): Let q_1 be the probability that an insertion is made into any of the subtrees of Figure 4.3. Let $1-q_1$ be the probability that an insertion is made into any of the subtrees of Figure 4.4.

Consider a N-key AVL tree with all subtrees in the fringe being of the type shown in Figure 4.3, the type shown in Figure 4.4, or a mixture of the two. Let us consider one tree of Figure 4.3 and one tree of Figure 4.4, as shown in Figure 4.5. The arcs show the probabilities of two possible transitions. Then

$$\frac{3(1-q_1)}{5N} = \frac{q_1}{N}$$

or $q_1 = \frac{3}{8}$.

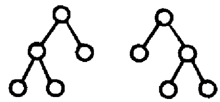


Fig. 4.3

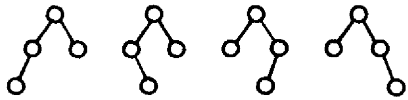


Fig. 4.4

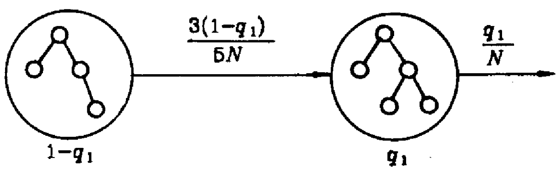


Fig. 4.5

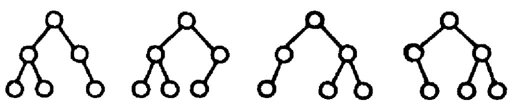


Fig.4.6

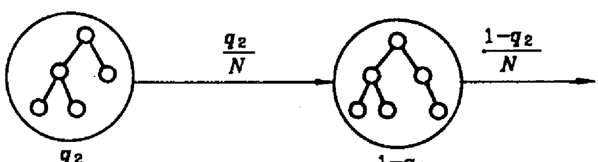


Fig. 4.7

If $\Delta t_N\uparrow$ is the increment in t_N then

$$\Delta t_N\uparrow \le \frac{3-8q_1}{5N},$$

so t_N cannot increment beyond $q_1 = 3/8$ and $q_1 = 3/8$ is the maximum value for q_1. By definition $t_N = \frac{2q_1}{3}$, which gives

$$0 \le t_N \le \frac{1}{4}.$$

Case (*ii*): Let q_2 be the probability that an insertion happens into any of the subtrees of Figure 4.3. Let $1-q_2$ be the probability that an insertion happens into any of the subtrees of Figure 4.6.

Consider a N-key AVL tree with all subtrees in the fringe being of the type shown in Figure 4.3, the type shown in Figure 4.6, or a mixture of the two. Let us consider one tree of Figure 4.3 and one tree of Figure 4.6, as shown in Figure 4.7. The arcs show the probabilities of two possible transitions. Then

$$\frac{q_2}{N} = \frac{1-q_2}{N}$$

or $q_2 = \frac{1}{2}$, where $q_2 = 1/2$ is the maximum value for q_2. By definition $t_N = \frac{2q_2}{3}$, which gives

$$0 \le t_N \le \frac{1}{3} \quad \blacksquare$$

Lemma 4.3. The probability s_N is bounded by $0 \le s_N \le \frac{1}{6}$

Proof :

Case (*i*): Let r_1 be the probability that an insertion happens into a tree of the type shown in Figure 4.8. Let $1-r_1$ be the probability that an insertion happens into any tree of the types shown in Figure 4.9. Notice that an insertion into a tree of Figure 4.9 gives a tree of Figure 4.8 with probability 3/11. Furthermore, it is not difficult to see that the trees of Figure 4.9 represent the main source of subtrees that under a new insertion are transformed into a tree of the type shown in Figure 4.8. (The trees of Figure 4.8 may be obtained from other sources by performing rotations on larger subtrees, but the probabilities in these cases are smaller than the probabilities related to the trees shown in Figure 4.9.)

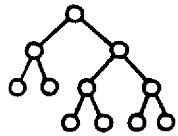


Fig. 4.8 (Symmetric cases occur)

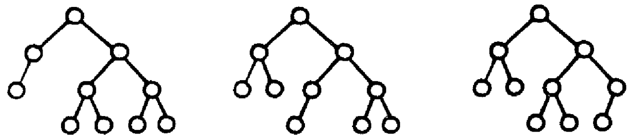


Fig. 4.9 (Symmetric cases occur)

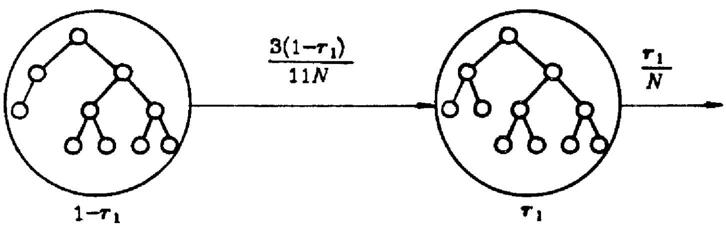


Fig. 4.10

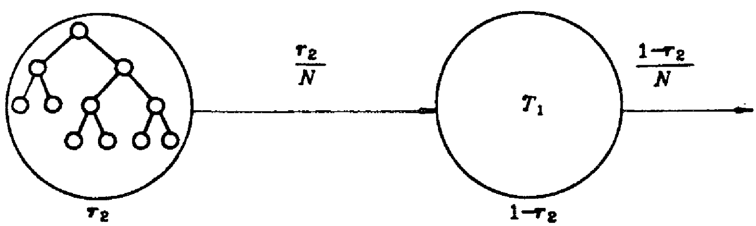


Fig. 4.11 (T_1 is a tree obtained from an insertion into a tree of Figure 4.8)

Consider a N-key AVL tree with all subtrees in the fringe being of the type shown in Figure 4.8, type shown in Figure 4.9, or a mixture of the two. Let us consider one tree of Figure 4.8 and one tree of Figure 4.9, as shown in Figure 4.10. The arcs show the probabilities of two possible transitions. Then

$$\frac{3(1-r_1)}{11N} = \frac{r_1}{N}$$

or $r_1 = \frac{3}{14}$.

If $\Delta s_N\uparrow$ is the increment in s_N then

$$\Delta s_N\uparrow \le \frac{3-14r_1}{11n},$$

so s_N cannot increment beyond $r_1 = 3/14$ and $r_1 = 3/14$ is the maximum value for r_1. By definition $s_N = \frac{r_1}{3}$, which gives

$$0 \le s_N \le \frac{1}{14}.$$

Case (ii): Let r_2 be the probability that an insertion happens into any of the trees of Figure 4.8. Let $1-r_2$ be the probability that an insertion happens in one of the trees one may obtain by inserting into a tree of Figure 4.8.

Consider a N-key AVL tree with all subtrees in the fringe being of the type shown in Figure 4.8, the type one may obtain by inserting into a tree of Figure 4.8, or a mixture of the two. Let us consider the two trees shown in Figure 4.11. The arcs show the probabilities of two possible transitions. Then

$$\frac{r_2}{N} = \frac{1-r_2}{N}$$

or $r_2 = \frac{1}{2}$, where $r_2 = 1/2$ is the maximum value for r_2. By definition $s_N = \frac{r_2}{3}$, which gives

$$0 \le s_N \le \frac{1}{6} \quad ▪$$

In the following section we present a technique to deal with weakly-closed tree collections, in which unknown probabilities appear in the transition matrix.

5. Coping with Weakly-closed AVL Tree Collections

Consider again the tree collection of AVL trees with five or fewer leaves shown in Figure 4.1. As shown in Section 4 this tree collection is weakly-closed. In this section we present a technique to deal with weakly-closed AVL tree collections.

From the results of Lemma 4.1 we can examine the insertion process and obtain

$$H(s_N,t_N) = \begin{bmatrix} -3 & 0 & 2(s_N-t_N) & 6/5 \\ 3 & -4 & 3(s_N+t_N) & 12/5 \\ 0 & 4 & -5+4t_N & 12/5 \\ 0 & 0 & 5(1-s_N-t_N) & -6 \end{bmatrix} \qquad (1)$$

where s_N and t_N depend on N. Figure 5.1 shows how the values of column three of $H(s_N,t_N)$ in Eq.(1) were obtained.

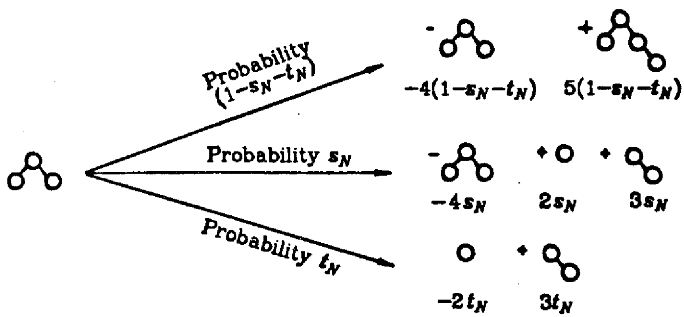


Fig. 5.4.1

The characteristic polynomial of $H(s_N,t_N)$ is

$$\det\big(H(s_N,t_N)-\lambda I\big) = \lambda^4+(18-4t_N)\lambda^3+(107-52t_N)\lambda^2+(210-168t_N)\lambda,$$

the eigenvalues are $\lambda_1 = 0$, $\lambda_2 = -5+4t_N$, $\lambda_3 = -6$, $\lambda_4 = -7$,

and the eigenvectors are

$$x_1(s_N,t_N) = \frac{1}{35-28t_N}\begin{bmatrix} 4(1+s_N-3t_N) \\ 3(3+2s_N-2t_N) \\ 12 \\ 10(1-s_N-t_N \end{bmatrix}, \text{ considering } p_1+p_2+p_3+p_4=1$$

$$\boldsymbol{x}_2(s_N,t_N) = \frac{1}{1+4t_N}\begin{bmatrix} -3+2(s_N-t_N) \\ 3(-1+s_N+t_N) \\ 1+4t_N \\ 5(1-s_N-t_N \end{bmatrix}$$

$$\boldsymbol{x}_3 = \frac{1}{5}\begin{bmatrix} 2 \\ 3 \\ 0 \\ -5 \end{bmatrix}$$

$$\boldsymbol{x}_4(s_N,t_N) = \frac{1}{2}\begin{bmatrix} 3-4s_N-2t_N \\ 5-6s_N-8t_N \\ 2 \\ 10(-1+s_N+t_N) \end{bmatrix}$$

where $\boldsymbol{x}_1(s_N,t_N)$, $\boldsymbol{x}_2(s_N,t_N)$, $\boldsymbol{x}_3$, and $\boldsymbol{x}_4(s_N,t_N)$ correspond to the eigenvalues λ_1, $\lambda_2(s_N,t_N)$, λ_3, and λ_4 respectively.

If the matrix $H = H(s_N,t_N)$ is independent of N, has one eigenvalue equal to zero, and the others have negative real part, then $\boldsymbol{p}(N)$, the solution of Eq.2.2-1 converges to the solution of $Hq = 0$ (cf. Theorem 2.2.3). However the matrix $H(s_N,t_N)$ in Eq.(1) contains the unknown probabilities s_N and t_N that depend on N, and consequently $H(s_N,t_N)$ depends on N. For this reason we have to prove the following result:

Theorem 5.1. Let $\boldsymbol{p}(N)$ be defined by

$$\boldsymbol{p}(N) = \left[I + \frac{H(s_N,t_N)}{N+1}\right]\boldsymbol{p}(N-1), \quad N>4 \tag{2}$$

and $\boldsymbol{p}(4) = (0,0,0,1)^T$ (an AVL tree with four nodes is the type 4 tree shown in Figure 4.1, with probability 1), where $\{s_N,t_N\}_{N>4}$ is a given sequence of probabilities. Then there exists a sequence $\{s_N',t_N'\}_{N>4}$, such that $\boldsymbol{p}(N)$ converges to $q(N)$, the solution of

$$H(s_N',t_N')q(N) = 0 \cdot \tag{3}$$

Proof : We will construct a sequence $\{s_N',t_N'\}$ and, in each iteration, express $\boldsymbol{p}(N)$ in the basis of eigenvectors of $H(s_N',t_N')$. In this basis

$$\boldsymbol{p}(N) = \alpha_1(N)\boldsymbol{x}_1(s_N,t_N)+\alpha_2(N)\boldsymbol{x}_2(s_N,t_N)+\alpha_3(N)\boldsymbol{x}_3+\alpha_4(N)\boldsymbol{x}_4(s_N,t_N)$$

where $\alpha_1(N) = 1$. Because $p_1+p_2+p_3+p_4 = 1$ the components of $\boldsymbol{x}_1(s_N,t_N)$ add to 1. The initial vector is

$$\begin{bmatrix} \alpha_1(4) \\ \alpha_2(4) \\ \alpha_3(4) \\ \alpha_4(4) \end{bmatrix} = E^{-1}(s_4',t_4')p(4)$$

where $E^{-1}(s_4',t_4')$ is the matrix that produces the spectral decomposition for $H(s_4',t_4')$. ($E(s,t)$ is the matrix of eigenvectors of $H(s,t)$.)

We want to prove that $(\alpha_1(N), \alpha_2(N), \alpha_3(N), \alpha_4(N))^T$ converges to $(1, 0, 0, 0)^T$, as $N \to \infty$.

Assume at step $N-1$ we have $(\alpha_1(N-1), \alpha_2(N-1), \alpha_3(N-1), \alpha_4(N-1))^T$ and we have already constructed $\{s_4',t_4'; \cdots ;s_{N-1}',t_{N-1}'\}$. In the next step we compute the effect of applying $\left(I+\frac{H(s_N,t_N)}{N+1}\right)$ to $(\alpha_1(N-1), \alpha_2(N-1), \alpha_3(N-1), \alpha_4(N-1))^T$ and at the same time express the new probability vector in a different basis of eigenvectors.

This is equivalent to the effect of one random insertion into the tree, i.e. going from $N-1$ to N nodes. To compute this we apply $\left(I+\frac{H(s_N,t_N)}{N+1}\right)$ to $x_1(s_{N-1}',t_{N-1}')$, $x_2(s_{N-1}',t_{N-1}')$, x_3, and $x_4(s_{N-1}',t_{N-1}')$, and obtain the spectral decomposition in each case. Then

$$\begin{bmatrix} \alpha_1(N) \\ \alpha_2(N) \\ \alpha_3(N) \\ \alpha_4(N) \end{bmatrix} = C(s_N,t_N;s_{N-1}',t_{N-1}') \begin{bmatrix} \alpha_1(N-1) \\ \alpha_2(N-1) \\ \alpha_3(N-1) \\ \alpha_4(N-1) \end{bmatrix}$$

where $C(s_N,t_N;s_{N-1}',t_{N-1}')$ is the matrix that operates the transformation with parameters s_N and t_N due to one insertion on the basis $x_1(s_{N-1}',t_{N-1}')$, $x_2(s_{N-1}',t_{N-1}')$, x_3, $x_4(s_{N-1}',t_{N-1}')$. Then

$$C(s_N,t_N;s_{N-1}',t_{N-1}') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ C[2,1] & 1-\frac{5-4t_N}{N+1} & 0 & C[2,4] \\ C[3,1] & C[3,2] & 1-\frac{6}{N+1} & C[3,4] \\ 0 & 0 & 0 & 1-\frac{7}{N+1} \end{bmatrix}$$

where

$$C[2,1] = \frac{-48(t_N-t_{N-1}')}{7(4t_{N-1}'-5)(N+1)}$$

$$C[2,4] = \frac{4(t_N-t_{N-1}')}{N+1}$$

$$C[3,1] = \frac{60(4s'_{N-1}t_N - 5t_N - 4s_N t'_{N-1} + 5t'_{N-1} - s_N + s'_{N-1})}{7(4t'_{N-1}+1)(4t'_{N-1}-5)(N+1)}$$

$$C[3,2] = \frac{-5(4s'_{N-1}t_N - 5t_N - 4s_N t'_{N-1} + 5t'_{N-1} - s_N + s'_{N-1})}{(4t'_{N-1}+1)(N+1)}$$

$$C[3,4] = \frac{-5(4s'_{N-1}t_N - 5t_N - 4s_N t'_{N-1} + 5t'_{N-1} - s_N + s'_{N-1})}{(4t'_{N-1}+1)(N+1)}$$

At this point the new probability vector is still expressed in the basis of eigenvectors of $H(s'_{N-1}, t'_{N-1})$. Now we will change to a new basis of eigenvectors, for suitably chosen s'_N, t'_N. Let $B(s'_N, t'_N; s'_{N-1}, t'_{N-1})$ be the matrix that changes basis. In this case

$$\begin{bmatrix} \alpha_1(N) \\ \alpha_2(N) \\ \alpha_3(N) \\ \alpha_4(N) \end{bmatrix} = B(s'_N, t'_N; s'_{N-1}, t'_{N-1}) \begin{bmatrix} \alpha_1(N-1) \\ \alpha_2(N-1) \\ \alpha_3(N-1) \\ \alpha_4(N-1) \end{bmatrix}$$

Notice again that $\alpha_1(N) = \alpha_1(N-1) = 1$. Then the matrix B is

$$B(s'_N, t'_N; s'_{N-1}, t'_{N-1}) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ B[2,1] & 1 & 0 & B[2,4] \\ B[3,1] & B[3,2] & 1 & B[3,4] \\ 0 & 0 & 0 & B[4,4] \end{bmatrix}$$

where

$$B[2,1] = \frac{-48(t'_N - t'_{N-1})}{7(4t'_N-5)(4t'_{N-1}-5)}$$

$$B[2,4] = \frac{2(t'_N - t'_{N-1})}{2t'_N+1}$$

$$B[3,1] = \frac{-10(4s'_{N-1}t'_N - 5t'_N - 4s'_N t'_{N-1} + 5t'_{N-1} - s'_N + s'_{N-1})}{7(4t'_N+1)(4t'_{N-1}-5)}$$

$$B[3,2] = \frac{5(4s'_{N-1}t'_N - 5t'_N - 4s'_N t'_{N-1} + 5t'_{N-1} - s'_N + s'_{N-1})}{(4t'_N+1)(4t'_{N-1}+1)}$$

$$B[3,4] = \frac{-5(4s'_{N-1}t'_N - 5t'_N - 4s'_N t'_{N-1} + 5t'_{N-1} - s'_N + s'_{N-1})}{4t'_N+1}$$

$$B[4,4] = \frac{2t'_{N-1}+1}{2t'_N+1}$$

The combined effect of these two transformations is the matrix

$$A(s_N',t_N';s_N,t_N;s_{N-1}',t_{N-1}') = B(s_N',t_N';s_{N-1}',t_{N-1}')\times C(s_N,t_N;s_{N-1}',t_{N-1}')$$

Then

$$A(s_N',t_N';s_N,t_N;s_{N-1}',t_{N-1}') = \begin{bmatrix} 1 & 0 & 0 & 0 \\ A[2,1] & 1-\dfrac{5-4t_N}{N+1} & 0 & A[2,4] \\ A[3,1] & A[3,2] & 1-\dfrac{6}{N+1} & A[3,4] \\ 0 & 0 & 0 & A[4,4] \end{bmatrix}$$

where

$$A[2,1] = \frac{-48(t_N'-t_{N-1}')}{7(4t_N'-5)(4t_{N-1}'-5)} + \frac{-48(t_N-t_N')}{7(4t_{N-1}'-5)(N+1)}$$

$$A[2,4] = \frac{4(t_N-t_{N-1}')}{N+1} + \left[1-\frac{7}{N+1}\right]\left[\frac{2(t_N'-t_{N-1}')}{2t_N'}\right]$$

$$A[3,1] = \frac{-10(4s_{N-1}'t_N'-5t_N'-4s_N't_{N-1}'+5t_{N-1}'-s_N'+s_{N-1}')}{7(4t_N'+1)(4t_{N-1}'-5)}$$
$$+\frac{-48(t_N-t_{N-1}')}{7(4t_{N-1}'-5)(N+1)}\left[\frac{5(4s_{N-1}'t_N'-5t_N'-4s_N't_{N-1}'+5t_{N-1}'-s_N'+s_{N-1}')}{(4t_N'+1)(4t_{N-1}'+1)}\right]$$
$$+\frac{60(4s_{N-1}'t_N-5t_N-4s_Nt_{N-1}'+5t_{N-1}'-s_N+s_{N-1}')}{7(4t_{N-1}'+1)(4t_{N-1}'-5)(N+1)}$$

$$A[3,2] = \frac{-5(4s_{N-1}'t_N-5t_N-4s_Nt_{N-1}'+5t_{N-1}'-s_N+s_{N-1}')}{(4t_{N-1}'+1)(N+1)}$$
$$+\left[1-\frac{5-4t_N}{N+1}\right]\left[\frac{5(4s_{N-1}'t_N'-5t_N'-4s_N't_{N-1}'+5t_{N-1}'-s_N'+s_{N-1}')}{(4t_N'+1)(4t_{N-1}'+1)}\right]$$

$$A[3,4] = \frac{-5(4s_{N-1}'t_N-5t_N-4s_Nt_{N-1}'+5t_{N-1}'-s_N+s_{N-1}')}{(4t_{N-1}'+1)(N+1)}$$
$$+\frac{4(t_N-t_{N-1}')}{N+1}\left[\frac{5(4s_{N-1}'t_N'-5t_N'-4s_N't_{N-1}'+5t_{N-1}'-s_N'+s_{N-1}')}{(4t_N'+1)(4t_{N-1}'+1)}\right]$$
$$+\left[1-\frac{7}{N+1}\right]\left[\frac{-5(4s_{N-1}'t_N'-5t_N'-4s_N't_{N-1}'+5t_{N-1}'-s_N'+s_{N-1}')}{4t_N'+1}\right]$$

$$A[4,4] = \left[1-\frac{7}{N+1}\right]\left[\frac{2t_N+1}{2t_N'}\right]$$

Let us show what happens when we go step by step. Recall that s'_N, t'_N represent the current values of the auxiliary sequence, and s_N, t_N represent the current values of the unknown probabilities in the transition matrix. Then

$$\begin{bmatrix}1\\ \alpha_2(N)\\ \alpha_3(N)\\ \alpha_4(N)\end{bmatrix} = \begin{bmatrix}\alpha_1(N)\\ \alpha_2(N)\\ \alpha_3(N)\\ \alpha_4(N)\end{bmatrix} = A(s'_N,t'_N;s_N,t_N;s'_{N-1},t'_{N-1})\times\cdots\times$$

$$A(s'_6,t'_6;s_6,t_6;s'_5,t'_5)\times A(s'_5,t'_5;s_5,t_5;s'_4,t'_4)E^{-1}(s'_4,t'_4)p(4)\ .$$

To prove that $(\alpha_1(N),\alpha_2(N),\alpha_3(N),\alpha_4(N))^T$ converges to $(1,0,0,0)^T$ it is enough to prove that

$$\lim_{N\to\infty}\prod_{k=5}^{N} H(s'_k,t'_k;s_k,t_k;s'_{k-1},t'_{k-1}) = \begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}. \qquad (4)$$

This would mean that

$$\begin{bmatrix}1\\ \alpha_2(N)\\ \alpha_3(N)\\ \alpha_4(N)\end{bmatrix} = \begin{bmatrix}1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0\end{bmatrix}\cdot E^{-1}(s'_4,t'_4)p(4) = \begin{bmatrix}1\\0\\0\\0\end{bmatrix},$$

or that the probabilities converge to an eigenvector associated to $\lambda = 0$ of $H(s'_N,t'_N)$. Notice that this is independent of $p(4)$ and of the choice of s'_4,t'_4.

The entries $A[2,1]$ and $A[3,1]$ are the critical ones for the convergence of the $\prod_N A(s'_N,t'_N;s_N,t_N;s'_{N-1},t'_{N-1})$ to the matrix (4). Let us solve $A[2,1] = 0$ for t'_N. Then

$$t'_N = \frac{(N+1)t'_{N-1}+5(t_N-t'_{N-1})}{N+1+4(t_N-t'_{N-1})} \qquad (5)$$

Now we substitute t'_N in $A[3,1] = 0$ and solve for s'_N. Then

$$s'_N = \frac{5(t_N-t'_{N-1})+6(s_N-s'_{N-1})+(N+1)s'_{N-1}}{N+1+4(t_N-t'_{N-1})} \qquad (6)$$

Before we go ahead to compute the $\prod_N A(s'_N,t'_N;s_N,t_N;s'_{N-1},t'_{N-1})$ with $A[2,1] = 0$ and $A[3,1] = 0$ we have to show that t'_N in Eq.(5) and s'_N in Eq.(6) are bounded.

Proposition 5.2. For $0 \le t'_{N-1},t_N \le \frac{1}{3}$ the value of t'_N in Eq.(5) is bounded by

$$0 \le t'_N \le \frac{1}{3}$$

Proof :

Case (i): $t'_N \ge 0$ is equivalent to

$$(N+1)t'_{N-1} + 5(t_N - t'_{N-1}) \ge 0 \quad ,$$

which is true.

Case (ii): $t'_N \le \frac{1}{3}$ is equivalent to

$$(N+1)t'_{N-1} + 5(t_N - t'_{N-1}) \le \frac{1}{3}[N+1+4(t_N - t'_{N-1})]$$

or

$$\left(N+1-5+\frac{4}{3}\right)t'_{N-1} + \left(5-\frac{4}{3}\right)t_N \le \frac{1}{3}(N+1)$$

The left hand side of the above expression is maximum when $t_N = t'_{N-1} = \frac{1}{3}$. Then

$$\frac{1}{3}(N+1) \le \frac{1}{3}(N+1) \quad \bullet$$

Proposition 5.3. For $0 \le s_N \le \frac{1}{6}$, $-\frac{5}{14} \le s'_{N-1} \le \frac{4}{11}$, and $0 \le t'_{N-1}, t_N \le \frac{1}{3}$, the value of s'_N in Eq.(6) is bounded by

$$-\frac{5}{14} \le s'_N \le \frac{4}{11}$$

Proof :

Case (i): $s'_N \ge -\frac{5}{14}$ is equivalent to

$$5(t_N - t'_{N-1}) + 6(s_N - s'_{N-1}) + (N+1)s'_{N-1} \ge -\frac{5}{14}[N+1+4(t_N - t'_{N-1})]$$

or

$$\frac{90}{14}(t_N - t'_{N-1}) + (N-5)s'_{N-1} + 6s_N \ge -\frac{5}{14}(N+1) \quad .$$

The left hand side of the above expression is minimum when $t_N = 0$, $t'_{N-1} = \frac{1}{3}$, $s_N = 0$, and $s'_{N-1} = -\frac{5}{14}$. Then

$$-\frac{5}{14}(N+1) \geq -\frac{5}{14}(N+1)$$

Case (*ii*): $s_N' \leq \frac{4}{11}$ is equivalent to

$$5(t_N-t_{N-1}')+6(s_N-s_{N-1}')+(N+1)s_{N-1}' \leq \frac{4}{11}[N+1+4(t_N-t_{N-1}')]$$

or

$$\frac{39}{11}(t_N-t_{N-1}')+(N-5)s_{N-1}'+6s_N \leq \frac{4}{11}(N+1) \cdot$$

The left hand side of the above expression is maximum when $t_N = \frac{1}{3}$, $t_{N-1}' = 0$, $s_{N-1}' = \frac{4}{11}$, and $s_N = \frac{1}{6}$. Then

$$\frac{4}{11}(N+1) \leq \frac{4}{11}(N+1) \quad \blacksquare$$

Now we compute $\prod_N A(s_N', t_N'; s_N, t_N; s_{N-1}', t_{N-1}')$:

$$A(s_N', t_N'; s_N, t_N; s_{N-1}', t_{N-1}') \times \prod_{k=5}^{N-1} A(s_k', t_k'; s_k, t_k; s_{k-1}', t_{k-1}') =$$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & u_N & 0 & v_N \\ 0 & w_N & x_N & y_N \\ 0 & 0 & 0 & z_N \end{bmatrix} \times \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & a_{N-1} & 0 & b_{N-1} \\ 0 & c_{N-1} & d_{N-1} & e_{N-1} \\ 0 & 0 & 0 & f_{N-1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & a_N & 0 & b_N \\ 0 & c_N & d_N & e_N \\ 0 & 0 & 0 & f_N \end{bmatrix}$$

where

$$a_N = a_{N-1}u_N$$

$$b_N = b_{N-1}u_N + f_{N-1}v_N$$

$$c_N = c_{N-1}x_N + a_{N-1}w_N$$

$$d_N = d_{N-1}x_N$$

$$e_N = b_{N-1}w_N + e_{N-1}x_N + f_{N-1}y_N$$

$$f_N = f_{N-1}z_N$$

and

$$u_N = 1 - \frac{5-4t_N}{N+1}$$

$$v_N = A[2,4]$$

$$w_N = A[3,2]$$

$$x_N = 1 - \frac{6}{N+1}$$

$$y_N = A[3,4]$$

$$z_N = A[4,4]$$

Notice that:

(i) $d_N = 0$ for $N \geq 5$

(ii) $f_N = 0$ for $N \geq 6$

(iii) $a_N = \prod_{k=5}^{N} \left(1 - \frac{5 - 4t_k}{k+1} \right) \leq \prod_{k=5}^{N} \left(1 - \frac{5 - \frac{4}{3}}{k+1} \right) = O(N^{-11/3})$

(iv) $b_N = b_{N-1} u_N$ since $f_N = 0$ for $N \geq 6$. Then b_N is like a_N, or $b_N = O(N^{-11/3})$.

The recurrences for c_N and e_N are the remaining ones. Considering that $f_N = 0$ for $N \geq 6$, and that a_N and b_N have the same type of recurrence we conclude that c_N and e_N are similar.

Let us look at c_N. The solution for c_N can be found in Sedgewick (1975, pp. 297-298). Then

$$c_N = \prod_{j=7}^{N} x_j \left(c_6 + \sum_{k=7}^{N} w_k a_{k-1} \prod_{j=1}^{k} \frac{1}{x_j} \right)$$

$$= \left(\prod_{j=7}^{N} x_j \right) c_6 + \sum_{k=7}^{N} w_k a_{k-1} \left(\prod_{j=k+1}^{N} x_j \right)$$

Considering that

$$\prod_{j=7}^{N} x_j = \Theta(N^{-6}) ,$$

$$c_6 = O(1) ,$$

$$a_N = O(N^{-11/3}) ,$$

$$w_N = O(1),$$

$$\prod_{j=k+1}^{N} x_j = O(1),$$

then

$$c_N = O(N^{-8/3}) \, .$$

We conclude that if the value of t_N' is selected according to Eq.(5) and the value of s_N' is selected according to Eq.(6) then $\prod_N A(s_N',t_N';s_N,t_N;s_{N-1}',t_{N-1}')$ converges to $\begin{pmatrix} 1&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \\ 0&0&0&0 \end{pmatrix}$. ∎

The theorem we just proved tell us that the solution of Eq.(2) converges to the solution of $H(s_N,t_N)p(N) = 0$, where $H(s_N,t_N)$ is as in Eq.(1). Then

$$p_1 = \frac{4(1+s_N-3t_N)}{35-28t_N}$$

$$p_2 = \frac{3(3+2s_N-2t_N)}{35-28t_N}$$

$$p_3 = \frac{12}{35-28t_N}$$

$$p_4 = \frac{10(1-s_N-t_N)}{35-28t_N}$$

for some value of s_N,t_N, according to Proposition 5.2 and Proposition 5.3. Since the eigenvalues of H are 0, $-5+4t_N$, -6, and -7, and considering that $0 \le t_N \le 1/3$ (cf. Proposition 5.4.2), using Eq.2.2-2 the asymptotic values of $p(N)$ obtained from Eq.2.2-3 are approximated to the $O(N^{-11/3})$.

Theorem 5.4. The expected number of rotations in a random AVL tree with N keys is bounded by

$$(i) \;\; 2\frac{p_2}{L_2}+2\frac{p_4}{L_4} \le r(N) \le 1-\left(\frac{p_2}{L_2}+3\frac{p_4}{L_4}\right)$$

and

$$(ii) \;\; 2\frac{p_2}{L_2}+2\frac{p_4}{L_4} \le r(N) \le 2\frac{p_2}{L_2}+2\frac{p_4}{L_4}+p_1+p_3$$

Proof : The left hand side of cases (i) and (ii) can be obtained by observing Figure 4.1. The right hand side of cases (i) and (ii) can be obtained by using Lemma 3.1.1 and by observing Figure 4.1. ∎

In the following corollary the bounds for $r(N)$ are obtained to hold for any values of s_N and t_N in the range given by Proposition 5.2 and Proposition 5.3:

Corollary. $\frac{2}{7}+O(N^{-11/3}) \le r(N) \le \frac{98}{121}+O(N^{-11/3})$

Lemma 5.5. The expected number of nodes in the fringe of an AVL tree with N keys corresponding to the tree collection of Figure 4.1 is

$$\bar{f}(N) = \left(\frac{p_1}{L_1}+2\frac{p_2}{L_2}+3\frac{p_3}{L_3}+4\frac{p_4}{L_4}\right)(N+1)$$

Proof : The above expression can be obtained by observing Figure 4.1 and by using Eq.2.2-4. ▪

In the following corollary the value for $\bar{f}(N)$ is obtained to hold for any values of s_N and t_N in the range given by Proposition 5.2 and Proposition 5.3:

Corollary. $\bar{f}(N) = \frac{267}{385}(N+1)+O(N^{-11/3})$

Lemma 5.6. The expected number of unbalanced nodes outside the fringe of a random AVL tree with N keys is at least $\frac{p_1}{L_1}(N+1)$.

Proof : The above expression is obtained as follows: a type 1 tree shown in Figure 4.1 must always have a type 3 tree as brother, otherwise it constitutes a type 3 or a type 4 tree. Thus the father node of a type 1 tree is always unbalanced, and the number of trees in this situation is $\frac{p_1}{L_1}$. ▪

Theorem 5.7. The expected number of balanced nodes in a random AVL tree with N keys is bounded by

$$\left(\frac{p_1}{L_1}+\frac{p_2}{L_2}+3\frac{p_3}{L_3}+2\frac{p_4}{L_4}\right)(N+1)\le\bar{b}(N)\le N-\left(\frac{p_1}{L_1}+\frac{p_2}{L_2}+2\frac{p_4}{L_4}\right)(N+1)$$

Proof : The left hand side of the above expression is obtained by observing Figure 4.1 and by using Eq.2.2-4. The right hand side is obtained by using Lemma 3.1.6, Lemma 5.5, and Lemma 5.6. ▪

In the following corollary the bounds for $\bar{b}(N)$ are obtained to hold for any values of s_N and t_N in the range given by Proposition 5.2 and Proposition 5.3:

Corollary. $\frac{18}{35}+\frac{18}{35N}+O(N^{-11/3})\le\frac{\bar{b}(N)}{N}\le\frac{62}{77}-\frac{15}{77N}+O(N^{-11/3})$

Experimental results show that $r(N)\approx 0.47$ (Ziviani and Tompa, 1980), and $\bar{b}(N)\approx 0.68N$ (Knuth, 1973).

6. Larger Weakly-closed AVL Tree Collections

In Section 4 we showed that any AVL tree collection that contains a tree type with its root node balanced and has more than three types is weakly-closed. This happens because every AVL tree type that contains more than one internal node and has its root node balanced suffers from the same type of misbehaviour that occurs with type 3 of Figure 4.2, as described in Lemma 4.1.

It is easy to prove a lemma similar to Lemma 4.1 for the tree collection shown in Figure 6.1. The only difference in the proof of such lemma is that the trees shown in Figure 4.2(a and b) do not occur, and consequently the unknown probability t_N does not exist. The transition matrix corresponding to the tree collection shown in Figure 6.1 involves one unknown probability s_N, as follows

$$H(s_N) = \begin{bmatrix} -4 & 3s_N & 12/5 & 3 \\ 4 & -5-4s_N & 0 & 4 \\ 0 & 5(1-s_N) & -6 & 0 \\ 0 & 6s_N & 18/5 & -7 \end{bmatrix} \qquad (1)$$

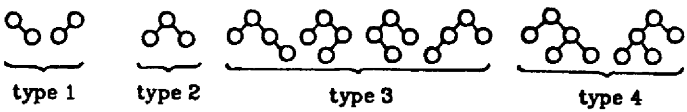


Fig. 6.1 Tree collection of AVL trees with more than 2 and less than 7 leaves (leaves not shown)

The transition matrix in Eq.(1) contains only one unknown probability, and the corresponding tree collection shown in Figure 6.1 contains more information than the tree collection used in the previous section. Now comes the question: Is it possible to apply Theorem 5.1 to this tree collection? Unfortunately we were not able to show convergence in this case. We feel that a similar proof may exist for the tree collection shown in Figure 6.1. In fact the ideal situation is to prove a general theorem about matrix recurrence relations involving unknown probabilities, but it seems too difficult to obtain.

What can we say about s as a function of N ?

Unfortunately we cannot say much about s_N. For trees of size $N = 10$ and $N = 11$ we are able to obtain s_N exactly (5/154 and 3/77, respectively). Table 6.1 shows simulation results for larger trees, obtained with a 95% confidence interval. From Table 6.1 the value of s_N seems to converge to

8/900 when N is large, but we are not able to prove it. Moreover s_N may oscillate smoothly, in such a way that simulations cannot detect. (e.g. consider $s_N = \cos(\ln N)/100$.)

Tree Size	Number of Trees	s_N (percent)		
		Fig.4.4(a)	Fig.4.4(b-c)	Total
49	10000	0.4204 ± 0.0250	0.5088 ± 0.0388	0.9292 ± 0.0454
99	5000	0.4368 ± 0.0257	0.4960 ± 0.0363	0.9328 ± 0.0459
499	5000	0.4201 ± 0.0113	0.4936 ± 0.0173	0.9138 ± 0.0204
999	2000	0.4212 ± 0.0125	0.4944 ± 0.0193	0.9156 ± 0.0230
2999	1500	0.4154 ± 0.0083	0.4740 ± 0.0127	0.8893 ± 0.0152
4999	1000	0.4101 ± 0.0076	0.4810 ± 0.0119	0.8910 ± 0.0141
9999	1000	0.4132 ± 0.0055	0.4806 ± 0.0086	0.8938 ± 0.0101
14999	200	0.4092 ± 0.0106	0.4773 ± 0.0153	0.8865 ± 0.0183
19999	300	0.4086 ± 0.0070	0.4836 ± 0.0109	0.8922 ± 0.0126

Table 6.1 Results for s_N

It is also possible to prove a lemma similar to Lemma 4.1 for the tree collection containing 10 types shown in Figure 6.2. The corresponding transition matrix, which involves eight unknown probabilities s, s_1, s_2, s_3, s_4, s_5, t and u, is shown in Figure 6.3.

When the number of unknown probabilities involved in the transition matrix is greater than one the problem of dealing with these unknown probabilities becomes a mathematical programming problem. This fact is important because the bounds for any complexity measure are obtained from the minimum over all possible values of the unknown probabilities in the transition matrix.

Assuming that a convergence theorem exists for (i) the AVL tree collection containing 4 types shown in Figure 6.1, (ii) the AVL tree collection containing 10 types shown in Figure 6.2, and (iii) the AVL tree collection containing 15 types shown in Figure 6.4, then the solution of

$$p(N) = \left[I + \frac{H(N)}{N+1}\right] p(N-1)$$

converges to the solution of

$$H(N)p(N) = 0. \tag{2}$$

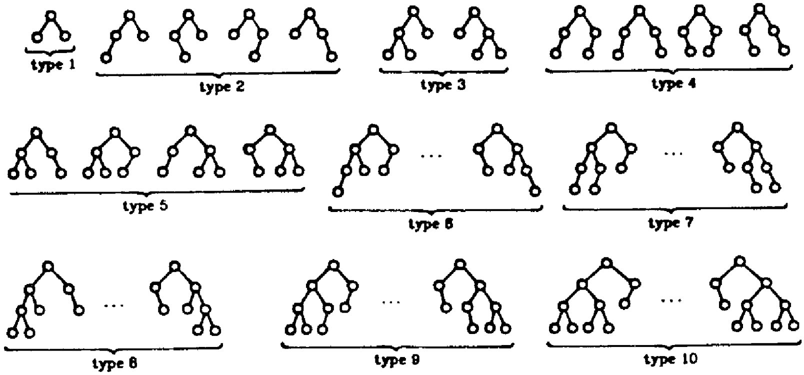


Fig. 6.2 Tree collection of AVL trees with 10 types (leaves not shown)

$$\begin{bmatrix}
-5+8u & & & & \frac{24}{7}-\frac{32}{7}ss_5-\frac{16}{7}t & \frac{3}{2} & \frac{4}{3} & \frac{4}{3} & \frac{6}{5} & \frac{36}{11} \\
5(1-u) & -6 & & & \frac{20}{7}(s-ss_1)+\frac{15}{7}t & \frac{15}{8} & & & 2 & \frac{30}{11} \\
6u & \frac{18}{5} & -7 & & -\frac{24}{7}ss_2+\frac{6}{7}t & & & 2 & & \frac{12}{11} \\
 & \frac{12}{5} & & -7 & -\frac{24}{7}ss_3+\frac{6}{7}t & & 2 & & \frac{12}{7} & \frac{12}{11} \\
7u & & 7 & 7 & -8-4ss_4+3t & & & & \frac{21}{10} & \frac{42}{11} \\
 & & & & \frac{32}{7}(1-s+ss_1-t) & -9 & & & & \\
 & & & & \frac{36}{7}ss_3 & \frac{9}{4} & -10 & & & \\
 & & & & \frac{36}{7}ss_2 & \frac{27}{8} & & -10 & & \\
 & & & & \frac{40}{7}ss_4 & & \frac{20}{3} & \frac{20}{3} & -11 & \\
 & & & & \frac{44}{7}ss_5 & & & & \frac{33}{10} & -12
\end{bmatrix}$$

Fig. 6.3 Transition matrix corresponding to the tree collection of AVL trees shown in Figure 5.5.2

Solving Eq.(2) for the three AVL tree collections just mentioned, and taking the minimum over all possible values of the unknown probabilities for each complexity measure considered, we obtain the results shown in Table 6.1.

Tree Collection		$\overline{f}(N)$	$\overline{r}(N)$	$\frac{\overline{\delta}(N)}{N}$
Size	Characteristic			
4	weakly-closed	$0.75N$	$[0.38, 0.74]$	$[0.53 + 0.53/N, 0.78 - 0.22/N]$
10	weakly-closed	$0.83N$	$[0.40, 0.72]$	$[0.58 + 0.58/N, 0.76 - 0.24/N]$
15	weakly-closed	$0.86N$	$[0.43, -]$	$[0.60 + 0.60/N, 0.74 - 0.26/N]$

Table 6.1

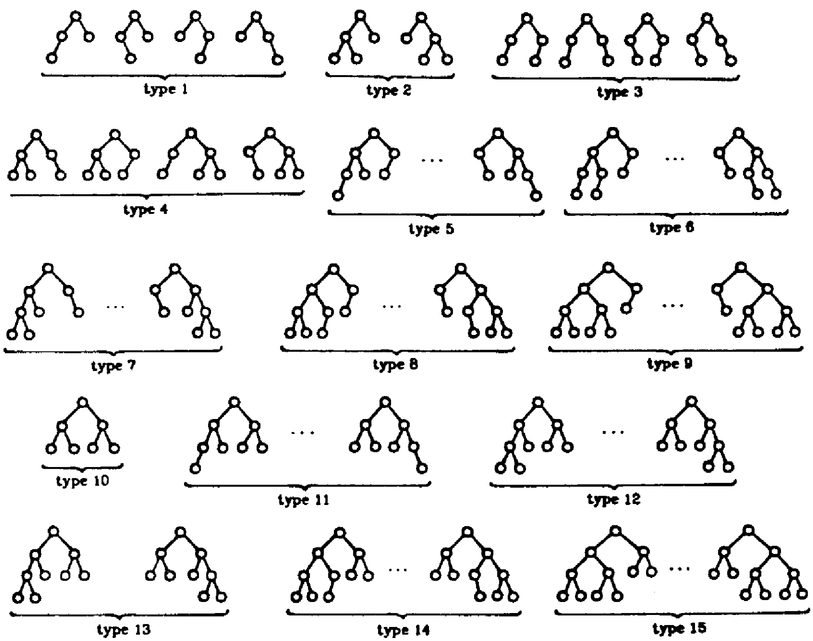


Fig. 6.4 Tree collection of AVL trees with 15 types (leaves not shown)

7. Application to Other Binary Search Trees

Weight-balanced trees ($WB[\alpha]$) were introduced by Nievergelt and Reingold (1973). A binary search tree is $WB[\alpha]$ if the number of leaves in the left subtree of the root node over the total number of leaves in the tree is in the interval $[\alpha, 1-\alpha]$. The root balance α of a complete binary search tree is 1/2. Like AVL trees, $WB[\alpha]$ trees are balanced by single and double rotations.

Another class of weight-balanced trees were introduced by Baer (1975) and also Gonnet (1982). They derived an algorithm that can be described as a counterpart of the AVL trees: perform single or double rotations whenever these rotations can reduce the total internal path of the subtree.

The closed AVL tree collections of Figure 3.1.1 and Figure 3.2.1 are also closed weight-balanced tree collections. Consequently, the AVL results shown in Table 1.1 for these tree collections are exactly the same results one would obtain in the analysis of weight-balanced trees using these same tree collections.

8. Conclusions

In Section 2 we present the fringe analysis technique. We show that the matrix recurrence relation related to fringe analysis problems converges to the solution of a linear system involving the transition matrix, even when the transition matrix has eigenvalues with multiplicity greater than one (i.e., the eigenvalues of the transition matrix do not need to be pairwise distinct).

In Section 3 we present a closed AVL tree collection containing three types. In Section 4 we show that an AVL tree collection containing four types is not closed. An inherent difficult posed by the rotations necessary to keep the AVL tree balanced forces the introduction of two unknown probabilities s_N and t_N into the transition matrix. In the main theorem of Section 5 we prove convergence of the matrix recurrence relation involving the unknown probabilities s_N and t_N.

Like AVL trees, weight-balanced trees are balanced by single and double rotations (Knuth, 1973, § 6.2.3). For this reason only small tree collections of weight-balanced trees are closed. For large tree collections we find the same type of difficulties showed for AVL trees. Consequently, the technique presented for the analysis of AVL trees is also suitable for the analysis of weight-balanced trees.

9. Acknowledgement

We wish to acknowledge the many helpful suggestions from Patricio Poblete.

10. References

Adel'son-Vel'skii, G.M. and Landis, E.M. "An Algorithm for the Organization of Information," *Doklady Akademia Nauk* USSR 146, 2 (1962), 263-266. English translation in *Soviet Math. Doklay* 3 (1962), 1259-1263.

Baer, J.-L. "Weight Balanced Trees," *Proc. AFIPS National Computer Conference 44* (1975), 467-472.

Brown, M. "A Partial Analysis of Random Height-Balanced Trees," *SIAM Journal of Computing 8*, 1 (Feb 1979), 33-41.

Cox, D.R. and Miller, H.D. *The Theory of Stochastic Processes* (London: Chapman and Hall Ltd., 1965).

Eisenbarth, B., Ziviani, N., Gonnet, G.H., Mehlhorn, K. and Wood, D. "The Theory of Fringe Analysis and Its Application to 2-3 Trees and B-trees," *Report CS-82-17*, Department of Computer Science, University of Waterloo, Jun 1982.

Feller, W. *An Introduction to Probability Theory and Its Applications*, Vol. 1 (New York: John Wiley and Sons, 1968).

Gonnet, G.H. "Balancing Binary Trees by Internal Path Reduction," *Report CS-82-07*, Department of Computer Science, University of Waterloo, Waterloo, Canada, 1982.

Gonnet, G.H., Ziviani, N. and Wood, D. "An Analysis of 2-3 Trees and B-trees," *Report CS-81-21*, Department of Computer Science, University of Waterloo, Canada, Jun 1981.

Knuth, D.E. *The Art of Computer Programming*, Vol. 3 (Reading, Mass.: Addison-Wesley, 1973).

Mehlhorn, k. "A Partial Analysis of Height-Balanced Trees," *Report A 79/13*, Universität des Saarlandes, Saarbrucken, West Germany, 1979a.

Mehlhorn, K. "A Partial Analysis of Height-Balanced Trees under Random Insertions and Deletions," *Report A 79/21*, Universität des Saarlandes, Saarbrucken, West Germany, 1979b.

Nievergelt, J. and Reingold, E.M. "Binary Search Trees of Bounded Balance," *SIAM Journal of Computing 2*, 1 (Mar 1973), 33-43.

Ottmann, Th. and Six, H.-W. "Eine neue Klasse von ausgeglichenen Binärbäumen," *Augewandte Informatik 9* (1976), 395-400.

Ottmann, Th. and Wood, D. "1-2 Brother Trees or AVL Trees Revisited," *The Computer Journal 23*, 3 (1980), 248-255.

Sedgewick, R. "Quicksort," *Report STAN-CS-492*, Computer Science Department, Stanford University, May 1975.

Yao, A. "On Random 2-3 Trees," *Acta Informatica 9*, (1978), 159-170.

Ziviani, N. "The Fringe Analysis of Search Trees," *Report CS-82-15*, Ph. D. Dissertation, Department of Computer Science, University of Waterloo, May 1982.

Ziviani, N. and Tompa, F. "A Look at Symmetric Binary B-trees," *Report CS-80-51*, Department of Computer Science, University of Waterloo, Canada, Nov 1980 (to appear in *INFOR - Canadian Journal of Operational Research and Information Processing*,).