

# Caracterización de árboles AVL de hojas mínimas

Universidad de Buenos Aires, Facultad de Ingeniería  
75.04 Algoritmos y Programación II  
Lucio Santi <lsanti at dc.uba.ar>

23 de junio de 2011

## 1. Introducción

Dado un árbol binario  $t$  de  $n$  nodos, es claro que su cantidad de hojas  $\lambda(t)$  es una  $O(n)$ . No obstante, sin restricciones adicionales también es cierto que  $\lambda(t) \in \Omega(1)$ . Al concentrar la atención en los árboles AVL, surge la curiosidad de determinar si esta última cota asintótica se mantiene o si, por otro lado, es aún mayor. El objetivo de este apunte es no sólo responder a este interrogante, sino también formular una caracterización completa de los árboles AVL de hojas mínimas para una cantidad de nodos dada. Esto posibilitará, adicionalmente, conocer la cota inferior exacta de  $\lambda(t)$ .

Comenzaremos discutiendo un acercamiento intuitivo al problema, a través del cual sentaremos las bases para la formulación matemática de los conceptos involucrados. Una vez hecho esto, demostraremos que la caracterización propuesta es correcta y completa, y luego derivaremos de los resultados encontrados un algoritmo para generar un árbol AVL de hojas mínimas para un  $n \in \mathbb{N}$  dado.

## 2. Caracterización

### 2.1. Acercamiento intuitivo










El problema que pretendemos resolver es el de encontrar una cota inferior para la cantidad de hojas de un árbol AVL, en función de su cantidad de nodos. Identifiquemos, entonces, cuál es el árbol AVL de menor cantidad de hojas para un tamaño  $n$  dado, comenzando desde  $n = 1$  y continuando incrementalmente, con la idea de desprender de esta estrategia una generalización.

Es claro que si  $n = 1$ , el AVL que buscamos es el único árbol de un nodo (que también es hoja). Para  $n = 2$ , podemos tomar cualquiera de los árboles de dos nodos; particularmente el que tiene la hoja a izquierda. Cuando  $n = 3$ , las reglas de balance de los árboles AVL nos restringen a considerar sólo el árbol completo de tres nodos. Para  $n = 4$ , el árbol que buscamos es el que se obtiene colgando una hoja al árbol completo de tres nodos.

Siguiendo esta línea, estaremos en condiciones de conjeturar que, dado un  $n$ , el árbol AVL de mínima cantidad de hojas y  $n$  nodos puede formarse tomando como árbol izquierdo el AVL de mínimas hojas e  $i$  nodos y tomando como árbol derecho el AVL con mínimas hojas y  $j$  nodos, de manera tal que  $n = i + j + 1$  y que la restricción sobre la altura se respete. El cuadro 2.1 muestra los resultados de poner en práctica esta estrategia, para  $n = 1, \dots, 9$ .  $\lambda(t)$  indica la cantidad de hojas de  $t$ .

Una primera observación de esta tabla permite concluir que, dado  $t_n$ , su cantidad de nodos es  $n$  y su cantidad de hojas es  $\lceil \frac{n+1}{3} \rceil$ . Sin embargo, también se observa que la conjetura formulada es falsa: podemos formar, por ejemplo, un árbol AVL  $t'_8$  de 8 nodos tomando como subárbol izquierdo a  $t_3$  y como subárbol

derecho a  $t_4$ . Es claro que la condición de balance no se viola y que su cantidad de hojas es 4, mientras que  $t_8$  posee 3 hojas.

$n$	1	2	3	4	5	6	7	8	9
$t_n$									
$\lambda(t_n)$	1	1	2	2	2	3	3	3	4

Cuadro 1: Árboles AVL de hojas mínimas con  $n$  nodos

Haciendo un análisis sobre la formas de combinar los subárboles de forma tal que la cantidad resultante de hojas sea precisamente  $\lceil \frac{n+1}{3} \rceil$ , se puede ver que se requiere que  $i \equiv 2(3)$ , o bien que  $j \equiv 2(3)$ , o bien que  $i \equiv j \equiv 1(3)$ <sup>1</sup>. Este hallazgo motiva la idea de modificar nuestra caracterización de manera de adicionar una cláusula más que refleje estas congruencias. En la siguiente sección, de hecho, daremos una definición formal de la conjetura actualizada.

## 2.2. Formalización

En esta sección nos proponemos poner en términos formales las ideas discutidas previamente con el objetivo de poder desarrollar una demostración matemática de que la caracterización sugerida es correcta y completa.

### 2.2.1. Definiciones

En primera instancia, daremos algunas definiciones básicas que nos serán de utilidad a lo largo del proceso.

**Definición 1.** Sea  $t$  un árbol binario. Los enteros  $|t|$ ,  $\lambda(t)$  y  $h(t)$  indican la cantidad de nodos, la cantidad de hojas y la altura de  $t$ , mientras que los árboles binarios  $\text{left}(t)$  y  $\text{right}(t)$  representan los subárboles izquierdo y derecho de  $t$  respectivamente.

**Definición 2.** **AVL** es el conjunto de todos los árboles binarios que satisfacen el invariante de los árboles AVL.

**Definición 3.** Sea  $n \in \mathbb{N}$ . Definimos a  $\mathbf{AVL}_n$  como el conjunto de árboles AVL de  $n$  nodos:

$$\mathbf{AVL}_n = \{t \in \mathbf{AVL} : |t| = n\}$$

Las siguientes definiciones corresponden a los conjuntos con los que estaremos trabajando.

**Definición 4.** Sea  $n \in \mathbb{N}$ .  $\mathcal{L}_n$  es el conjunto de árboles AVL de  $n$  nodos y cantidad mínima de hojas:

$$\mathcal{L}_n = \{t \in \mathbf{AVL}_n : \lambda(t) \leq \lambda(z) \forall z \in \mathbf{AVL}_n\}$$

**Definición 5.** Sea  $n \in \mathbb{N}$ .  $\mathcal{M}_n$  es el conjunto de árboles AVL de  $n$  nodos y  $\lceil \frac{n+1}{3} \rceil$  hojas:

$$\mathcal{M}_n = \left\{ t \in \mathbf{AVL}_n : \lambda(t) = \left\lceil \frac{n+1}{3} \right\rceil \right\}$$

<sup>1</sup> $a \equiv b(m)$  sii el resto de dividir a  $a$  por  $m$  es el mismo que el de dividir a  $b$  por  $m$ .

El próximo conjunto representa la caracterización que pretendemos probar correcta y completa. Observar que la definición del mismo sigue las líneas de lo discutido informalmente en las secciones anteriores.

**Definición 6.** Sea  $n \in \mathbb{N}$ . El conjunto  $\mathbb{L}_n$  se define recursivamente de la siguiente manera:

$$\mathbb{L}_n = \begin{cases} \{\bullet\} & \text{si } n = 1 \\ \{\bullet, \bullet\} & \text{si } n = 2 \\ \left\{ t \in \mathbf{AVL}_n : \exists i, j \in \mathbb{N} : n = i + j + 1 \wedge \right. \\ \quad \left. \text{left}(t) \in \mathbb{L}_i \wedge \text{right}(t) \in \mathbb{L}_j \wedge \right. \\ \quad \left. (i \equiv 2(3) \vee j \equiv 2(3) \vee i \equiv j \equiv 1(3)) \right\} & \text{si } n \geq 3 \end{cases}$$

### 2.2.2. Demostración de correctitud y completitud

La prueba está dividida en una serie de proposiciones que ahora pasamos a enunciar y demostrar.

La primera de ellas será utilizada, esencialmente, para afirmar que  $\mathbb{L}_n \neq \emptyset$ . A su vez, este resultado se usará en forma auxiliar para probar otras proposiciones.

**Proposición 1.** Si  $n \geq 2$ ,  $\exists t \in \mathbb{L}_n : h(t) = 2 + \min \{e \in \mathbb{N}_0 : n \leq 3 \cdot 2^e\}$ .

**Demostración** Por inducción en  $n$ .

Si  $n = 2$ , se observa que cualquier árbol en  $\mathbb{L}_2$  tiene altura 2. Además, es claro que el mínimo exponente entero  $e$  para el cual  $2 \leq 3 \cdot 2^e$  es  $e = 0$ . De esto se concluye la validez del enunciado cuando  $n = 2$ , y puede seguirse un razonamiento similar para cuando  $n = 3$  y  $n = 4$  (en este caso, no obstante, se tiene que  $e = 1$  y la altura es 3).

Supongamos, pues, que  $n \geq 5$  y definamos:

- $i = \lfloor \frac{n-1}{2} \rfloor - k$
- $j = \lceil \frac{n-1}{2} \rceil + k$

con

$$k = \min \left\{ m \in \mathbb{N}_0 : \left\lfloor \frac{n-1}{2} \right\rfloor - m \equiv 2(3) \vee \left\lceil \frac{n-1}{2} \right\rceil + m \equiv 2(3) \vee \left\lfloor \frac{n-1}{2} \right\rfloor - m \equiv \left\lceil \frac{n-1}{2} \right\rceil + m \equiv 1(3) \right\}$$

Observemos, por un lado, que  $i + j + 1 = n$  y que, además,  $0 \leq k \leq 1^2$ . A partir de la hipótesis inductiva, sabemos que:

- $\exists t_1 \in \mathbb{L}_i : h(t_1) = 2 + \min \{e \in \mathbb{N}_0 : i \leq 3 \cdot 2^e\} = 2 + e_1$ , y
- $\exists t_2 \in \mathbb{L}_j : h(t_2) = 2 + \min \{e \in \mathbb{N}_0 : j \leq 3 \cdot 2^e\} = 2 + e_2$

Consideremos un árbol binario  $t$  de  $n$  nodos tal que  $\text{left}(t) = t_1$  y  $\text{right}(t) = t_2$ . El objetivo será mostrar que  $t \in \mathbf{AVL}_n$  (la única condición restante para poder asegurar que  $t \in \mathbb{L}_n$ ) y que  $h(t) = 2 + \min \{e \in \mathbb{N}_0 : n \leq 3 \cdot 2^e\}$ . Para esto, analicemos los posibles valores de  $k$ .

Si  $k = 0$ , puede suceder que  $i = j$  o bien que  $j = i + 1$ . En el primero de estos casos, está claro que  $h(t_1) = h(t_2)$ . Para el restante, esta igualdad también será cierta: como  $i \equiv 1(3)$  o  $i \equiv 2(3)$ ,  $i < 3 \cdot 2^{e_1}$ ,

<sup>2</sup>Sólo puede ser  $k = 1$  cuando  $\lfloor \frac{n-1}{2} \rfloor \equiv 0(3) \Rightarrow \lfloor \frac{n-1}{2} \rfloor - 1 \equiv 2(3)$

de manera que  $j = i + 1 \leq 3 \cdot 2^{e_1}$ . Además,  $j > 3 \cdot 2^{e_1-1}$  ya que  $i$  verifica esto mismo. Luego,  $e_2 = e_1$  y por ende  $h(t_1) = h(t_2)$ . De esto se desprende que  $t \in \mathbf{AVL}_n$ .

Como ya hemos dicho,  $i < 3 \cdot 2^{e_1}$  y, entonces,

$$\begin{aligned} n &= i + j + 1 \\ &< 3 \cdot 2^{e_1} + 3 \cdot 2^{e_1} + 1 \\ &= 2 \cdot 3 \cdot 2^{e_1} + 1 \\ &= 3 \cdot 2^{e_1+1} + 1 \end{aligned}$$

Desde ya, esto quiere decir que  $n \leq 3 \cdot 2^{e_1+1}$ . Veamos ahora que ningún exponente más chico que  $e_1 + 1$  satisface la desigualdad. Para ello, partamos del hecho que  $e_1$  es el mínimo exponente que la verifica para  $i$  y  $j$ . Entonces,

$$\begin{aligned} n &= i + j + 1 \\ &> 3 \cdot 2^{e_1-1} + 3 \cdot 2^{e_1-1} + 1 \\ &= 2 \cdot 3 \cdot 2^{e_1-1} + 1 \\ &= 3 \cdot 2^{e_1} + 1 \\ &> 3 \cdot 2^{e_1} \end{aligned}$$

De esto sigue que cualquier exponente inferior a  $e_1$  tampoco servirá. Por lo tanto,

$$\begin{aligned} h(t) &\stackrel{\text{Def. } h}{=} 1 + \max(h(t_1), h(t_2)) \\ &= 1 + \max(2 + e_1, 2 + e_1) \\ &= 2 + (1 + e_1) \\ &= 2 + \min\{e \in \mathbb{N}_0 : n \leq 3 \cdot 2^e\} \end{aligned}$$

Cuando  $k = 1$ , puede suceder que:

$$(a) \left\lfloor \frac{n-1}{2} \right\rfloor \equiv \left\lceil \frac{n-1}{2} \right\rceil \equiv 0(3) \Rightarrow i \equiv 2(3) \text{ y } j = i + 2 \equiv 1(3), \text{ o bien}$$

$$(b) \left\lfloor \frac{n-1}{2} \right\rfloor \equiv 0(3) \text{ y } \left\lceil \frac{n-1}{2} \right\rceil \equiv 1(3) \Rightarrow i \equiv j = i + 3 \equiv 2(3)$$

En el escenario (a), primero supongamos que  $i < 3 \cdot 2^{e_1} - 1$ . Luego,  $j = i + 2 < 3 \cdot 2^{e_1} + 1 \Rightarrow j \leq 3 \cdot 2^{e_1}$ . Como  $e_2$  es el mínimo exponente  $e$  tal que  $j \leq 3 \cdot 2^e$ , se tiene que  $e_2 \leq e_1$ . Por otra parte, dado que  $i < j$ , debe ocurrir que  $e_1 \leq e_2$ . En consecuencia,  $e_1 = e_2$ , lo cual implica que  $h(t_1) = h(t_2)$ . A partir de esto, se puede seguir un razonamiento análogo al caso donde  $k = 0$  para verificar que se cumple lo buscado.

Si  $i = 3 \cdot 2^{e_1} - 1$ , se tiene que  $j = 3 \cdot 2^{e_1} + 1$ . Luego,  $e_2 = e_1 + 1$  y

$$\begin{aligned} n &= i + j + 1 \\ &= (3 \cdot 2^{e_1} - 1) + (3 \cdot 2^{e_1} + 1) + 1 \\ &= 2 \cdot 3 \cdot 2^{e_1} + 1 \\ &= 3 \cdot 2^{e_1+1} + 1 \\ &\leq 3 \cdot 2^{e_1+2} \end{aligned}$$

Además, es claro que  $n > 3 \cdot 2^{e_1+1}$ . Por ende,  $2 + e_1 = \min\{e \in \mathbb{N}_0 : n \leq 3 \cdot 2^e\}$  y entonces,

$$\begin{aligned} h(t) &\stackrel{\text{Def. } h}{=} 1 + \max(h(t_1), h(t_2)) \\ &= 1 + \max(2 + e_1, 2 + (e_1 + 1)) \\ &= 2 + (2 + e_1) \\ &= 2 + \min\{e \in \mathbb{N}_0 : n \leq 3 \cdot 2^e\} \end{aligned}$$

Notar, además, que  $h(t_2) = 1 + (2 + e_1) = 1 + h(t_1)$ , con lo cual  $t \in \mathbf{AVL}_n$ .

Para el escenario (b), un razonamiento similar al (a) permite arribar a las mismas conclusiones.

Concluyendo, se demostró que, en cualquiera de los casos posibles, el árbol  $t$  propuesto está en  $\mathbb{L}_n$  y que, además, verifica que  $h(t) = 2 + \min \{e \in \mathbb{N}_0 : n \leq 3 \cdot 2^e\}$

□

Una propiedad adicional de los árboles en  $\mathbb{L}_n$ , que dejamos sin demostrar, es que la mínima altura que pueden alcanzar es, precisamente,  $2 + \min \{e \in \mathbb{N}_0 : n \leq 3 \cdot 2^e\}$ .

**Corolario 1.**  $\mathbb{L}_n \neq \emptyset \forall n \in \mathbb{N}$

**Demostración** Para  $n = 1$ , el resultado es trivial por definición de  $\mathbb{L}_1$ . Por otro lado, si  $n \geq 2$ , la Proposición anterior asegura que habrá al menos un árbol en  $\mathbb{L}_n$ .

□

Esta Proposición afirma que todo árbol de  $n$  nodos de nuestra caracterización posee exactamente  $\lceil \frac{n+1}{3} \rceil$  hojas. Luego veremos que esta cantidad es, en efecto, mínima.

**Proposición 2.**  $\mathbb{L}_n \subseteq \mathcal{M}_n \forall n \in \mathbb{N}$

**Demostración** Por inducción en  $n$ .

Si  $n \leq 2$ , el resultado es trivial puesto que cualquier  $t \in \mathbb{L}_n$  verifica que  $\lambda(t) = 1 = \lceil \frac{n+1}{3} \rceil$ .

Supongamos ahora que  $n \geq 3$  y sea  $t \in \mathbb{L}_n$ . Luego,  $t$  verifica lo siguiente:

- $t_1 = \text{left}(t) \in \mathbb{L}_i$ , con  $i = |t_1|$
- $t_2 = \text{right}(t) \in \mathbb{L}_j$ , con  $j = |t_2|$ , y
- $i \equiv 2(3)$ , o bien  $j \equiv 2(3)$  o bien  $i \equiv j \equiv 1(3)$ .

Estudiaremos cada uno de estos tres casos por separado para concluir que en cada uno de ellos se cumple lo que buscamos. Antes de esto, notemos que podemos escribir a  $i$  y  $j$  a partir de los respectivos cocientes y restos en sus divisiones por 3:

- $i = 3q_i + r_i$ , y
- $j = 3q_j + r_j$

donde  $0 \leq r_i, r_j \leq 2$ . A partir de esto y de la hipótesis inductiva sobre  $t_1$  y  $t_2$ , tenemos:

$$\begin{aligned}
 \lambda(t) &= \lambda(t_1) + \lambda(t_2) \\
 &= \left\lceil \frac{i+1}{3} \right\rceil + \left\lceil \frac{j+1}{3} \right\rceil \\
 &= \left\lceil \frac{3q_i + r_i + 1}{3} \right\rceil + \left\lceil \frac{3q_j + r_j + 1}{3} \right\rceil \\
 &= \left\lceil q_i + \frac{r_i + 1}{3} \right\rceil + \left\lceil q_j + \frac{r_j + 1}{3} \right\rceil \\
 &\stackrel{r_i, r_j \leq 2}{=} (q_i + 1) + (q_j + 1) \\
 &= q_i + q_j + 2
 \end{aligned}$$

Para cada caso anterior, entonces, probaremos que  $\lceil \frac{n+1}{3} \rceil = q_i + q_j + 2$ .

- $i \equiv 2(3)$ , i.e.,  $r_i = 2$

$$\begin{aligned} \left\lceil \frac{n+1}{3} \right\rceil &= \left\lceil \frac{3(q_i + q_j) + 2 + r_j + 1 + 1}{3} \right\rceil \\ &= \left\lceil q_i + q_j + \frac{r_j + 4}{3} \right\rceil \\ &= \left\lceil q_i + q_j + 1 + \frac{r_j + 1}{3} \right\rceil \\ &\stackrel{r_j \leq 2}{=} q_i + q_j + 2 \end{aligned}$$

- $j \equiv 2(3)$ , i.e.,  $r_j = 2$

Este escenario es análogo al anterior.

- $i \equiv j \equiv 1(3)$ , i.e.,  $r_i = r_j = 1$

$$\begin{aligned} \left\lceil \frac{n+1}{3} \right\rceil &= \left\lceil \frac{3(q_i + q_j) + 1 + 1 + 1 + 1}{3} \right\rceil \\ &= \left\lceil q_i + q_j + \frac{4}{3} \right\rceil \\ &= \left\lceil q_i + q_j + 1 + \frac{1}{3} \right\rceil \\ &= q_i + q_j + 2 \end{aligned}$$

Todo esto permite concluir que  $\lambda(t) = \left\lceil \frac{n+1}{3} \right\rceil$ , con lo cual  $t \in \mathcal{M}_n$ . □

La Proposición que sigue garantiza que la caracterización es correcta: todo árbol AVL de  $n$  nodos en  $\mathbb{L}_n$  es, de hecho, de hojas mínimas.

**Proposición 3.**  $\mathbb{L}_n \subseteq \mathcal{L}_n \forall n \in \mathbb{N}$

**Demostración** Por inducción en  $n$ .

Si  $n \leq 2$ , el resultado es trivial: cualquier  $t$  en  $\mathbb{L}_n$  tiene exactamente una hoja, que es la cantidad mínima para un árbol binario con  $n$  nodos.

Veamos qué sucede cuando  $n \geq 3$ . Dado  $t \in \mathbb{L}_n$ , el objetivo será concluir que  $t \in \mathcal{L}_n$ . Para ello, tomemos un árbol cualquiera  $z \in \mathbf{AVL}_n$  y comprobemos que  $\lambda(t) \leq \lambda(z)$ . Un tal  $z$  verificará que:

- $z_1 = \text{left}(z) \in \mathbf{AVL}_{k_1}$ , con  $k_1 = |z_1|$
- $z_2 = \text{right}(z) \in \mathbf{AVL}_{k_2}$ , con  $k_2 = |z_2|$ , y
- $n = k_1 + k_2 + 1$

Observemos además que  $k_1, k_2 > 0$ . Imaginemos qué ocurriría en caso contrario. Si  $k_1 \leq 0$  o bien  $k_2 \leq 0$ , podemos suponer, sin pérdida de generalidad, que  $k_1 \leq 0$ . Dado que la cantidad de nodos de un árbol no puede ser negativa, debe ser  $k_1 = 0$ . Consecuentemente,  $k_2 = n - k_1 - 1 = n - 1 \geq 2$  pues  $n \geq 3$ . Por lo tanto, se tiene que  $h(z_2) \geq 2$  y, como  $h(z_1) = 0$ ,  $|h(z_1) - h(z_2)| \not\leq 1$ , que contradice la hipótesis de que  $z \in \mathbf{AVL}_n$ .

A través de este hecho y del Corolario 1, podemos afirmar que  $\mathbb{L}_{k_1} \neq \emptyset$  y que  $\mathbb{L}_{k_2} \neq \emptyset$ . Sean, pues,  $w_1 \in \mathbb{L}_{k_1}$  y  $w_2 \in \mathbb{L}_{k_2}$ . Dado que  $k_1 < n$  y  $k_2 < n$ , la hipótesis inductiva nos asegura que  $w_1 \in \mathcal{L}_{k_1}$  y que  $w_2 \in \mathcal{L}_{k_2}$ . A partir de esta observación y de la Proposición 2, tenemos lo siguiente:

$$\begin{aligned} q_{k_1} + q_{k_2} + 2 &= \left\lceil \frac{k_1 + 1}{3} \right\rceil + \left\lceil \frac{k_2 + 1}{3} \right\rceil \\ &= \lambda(w_1) + \lambda(w_2) \\ &\leq \lambda(z_1) + \lambda(z_2) \\ &= \lambda(z) \end{aligned} \tag{1}$$

en donde  $q_{k_1}, q_{k_2} \in \mathbb{N}$  son tales que  $k_1 = 3q_{k_1} + r_{k_1}$  y  $k_2 = 3q_{k_2} + r_{k_2}$ , con  $0 \leq r_{k_1}, r_{k_2} \leq 2$ .

Ahora bien, dado que  $t \in \mathbb{L}_n$ , por la Proposición 2 también tenemos que

$$\begin{aligned} \lambda(t) &= \left\lceil \frac{n+1}{3} \right\rceil \\ &= \left\lceil \frac{3(q_{k_1} + q_{k_2}) + r_{k_1} + r_{k_2} + 1 + 1}{3} \right\rceil \\ &= \left\lceil q_{k_1} + q_{k_2} + \frac{r_{k_1} + r_{k_2} + 2}{3} \right\rceil \\ &\stackrel{r_{k_1} + r_{k_2} \leq 4}{\leq} \left\lceil q_{k_1} + q_{k_2} + 2 \right\rceil \\ &= q_{k_1} + q_{k_2} + 2 \end{aligned} \tag{2}$$

A partir de (2) y (1), se concluye de inmediato que  $\lambda(t) \leq \lambda(z)$ . En otras palabras,  $t \in \mathcal{L}_n$ . □

**Corolario 2.**  $\mathcal{L}_n = \mathcal{M}_n \forall n \in \mathbb{N}$

**Demostración** Por el Corolario 1, sabemos que  $\mathbb{L}_n \neq \emptyset$ . Sea entonces  $z \in \mathbb{L}_n$ .

Dado un  $t \in \mathcal{L}_n$ , por ser  $t$  de hojas mínimas para  $n$  nodos, tenemos que  $\lambda(t) \leq \lambda(z)$ . Pero, por la Proposición anterior, también es cierto que  $\lambda(z) \leq \lambda(t)$  por ser  $z$  también de hojas mínimas. Así, y aplicando nuevamente la Proposición 2, se tiene que  $\lambda(t) = \lambda(z) = \left\lceil \frac{n+1}{3} \right\rceil$ , con lo cual  $t \in \mathcal{M}_n$ .

Por otra parte, dado un  $t \in \mathcal{M}_n$  y un  $w \in \mathbf{AVL}_n$ , y usando otra vez las mencionadas Proposiciones, se cumple que

$$\lambda(t) = \left\lceil \frac{n+1}{3} \right\rceil = \lambda(z) \leq \lambda(w)$$

de lo que es inmediato concluir que  $t \in \mathcal{L}_n$ . □

A través de la próxima Proposición, llegaremos a la conclusión de que la caracterización es también completa. Esto es, cualquier árbol AVL de  $n$  nodos con mínima cantidad de hojas está a su vez en  $\mathbb{L}_n$ .

**Corolario 3.**  $\mathcal{M}_n \subseteq \mathbb{L}_n \forall n \in \mathbb{N}$

**Demostración** Por inducción en  $n$ .

Si  $n = 1$ , la inclusión es trivialmente cierta pues todo árbol AVL de un nodo y  $\left\lceil \frac{1+1}{3} \right\rceil = 1$  hoja está en  $\mathbb{L}_1$ . Lo mismo vale para  $n = 2$ .

Sea entonces  $n \geq 3$ , y sea  $t \in \mathcal{M}_n$ . Esto nos dice lo siguiente sobre  $t$ :

- $t_1 = \text{left}(t) \in \text{AVL}_i$ , con  $i = |t_1|$
- $t_2 = \text{right}(t) \in \text{AVL}_j$ , con  $j = |t_2|$ ,
- $n = i + j + 1$ , y
- $\lambda(t) = \lambda(t_1) + \lambda(t_2) = \lceil \frac{n+1}{3} \rceil$

Tal como hicimos en demostraciones previas, escribamos a  $i$  y  $j$  a partir de sus respectivos cocientes y restos en sus divisiones por 3:

- $i = 3q_i + r_i$ , y
- $j = 3q_j + r_j$

con  $0 \leq r_i, r_j \leq 2$ .

Supongamos ahora que  $i \equiv 2(3) \vee j \equiv 2(3) \vee i \equiv j \equiv 1(3)$ . Bajo estas condiciones, puede verse fácilmente<sup>3</sup> que

$$\lambda(t) = q_i + q_j + 2 = \left\lceil \frac{i+1}{3} \right\rceil + \left\lceil \frac{j+1}{3} \right\rceil$$

Para que se verifique simultáneamente que  $\lambda(t_1) \neq \lceil \frac{i+1}{3} \rceil$  y  $\lambda(t_2) \neq \lceil \frac{j+1}{3} \rceil$ , debería existir  $k \in \mathbb{N}$  tal que  $\lambda(t_1) = \lceil \frac{i+1}{3} \rceil + k$  y  $\lambda(t_2) = \lceil \frac{j+1}{3} \rceil - k$ . No obstante, por la Proposición anterior, sabemos que la mínima cantidad de hojas para un AVL de  $j$  nodos es  $\lceil \frac{j+1}{3} \rceil$ , de manera que tal  $k$  no puede existir. De esto se deduce que  $\lambda(t_1) = \lceil \frac{i+1}{3} \rceil$  y  $\lambda(t_2) = \lceil \frac{j+1}{3} \rceil$ . Dado que  $i, j < n$ , la hipótesis inductiva garantiza que  $t_1 \in \mathbb{L}_i$  y  $t_2 \in \mathbb{L}_j$ . Esto completa las condiciones necesarias requeridas para que  $t \in \mathbb{L}_n$ .

Finalmente, veamos que nunca podrá darse alguno de los casos restantes. Éstos se muestran a continuación:

- $i \equiv 1(3) \wedge j \equiv 0(3)$ ,
- $j \equiv 1(3) \wedge i \equiv 0(3)$ , o bien
- $i \equiv j \equiv 0(3)$

En cualquiera de ellos, se obtiene que  $\lambda(t) = q_i + q_j + 1$ . Por la Proposición anterior, sabemos que la mínima cantidad de hojas de un AVL de  $i$  nodos es  $\lceil \frac{i+1}{3} \rceil$  y la mínima cantidad de hojas de un AVL de  $j$  nodos es  $\lceil \frac{j+1}{3} \rceil$ . Luego,

$$\begin{aligned} \lambda(t) &= \left\lceil \frac{n+1}{3} \right\rceil \\ &= q_i + q_j + 1 \\ &= \lambda(t_1) + \lambda(t_2) \\ &\geq \left\lceil \frac{i+1}{3} \right\rceil + \left\lceil \frac{j+1}{3} \right\rceil \\ &= (q_i + 1) + (q_j + 1) \\ &= q_i + q_j + 2 \end{aligned}$$

De esto se desprende que  $q_i + q_j + 1 \geq q_i + q_j + 2$ , lo cual es un absurdo y proviene de suponer que puede ocurrir alguno de los casos mencionados anteriormente. □

**Corolario 4.**  $\mathbb{L}_n = \mathcal{M}_n = \mathcal{L}_n \forall n \in \mathbb{N}$

**Demostración** Inmediato a partir de los resultados anteriores. □

<sup>3</sup>Recurrir a la demostración de la Proposición 2 para más detalles.



## 2.3. Algoritmo de generación de árboles AVL de hojas mínimas

Una consecuencia interesante de los resultados de la sección 2.2 es que otorgan sustento formal a un algoritmo divide & conquer sencillo para generar un árbol AVL de  $n$  nodos y cantidad mínima de hojas, a partir de un natural  $n$  dado. La figura 1 muestra los lineamientos generales del mismo.

```

AVLHOJASMÍNIMAS( $n$ ):
  Si  $n = 1$ : devolver  $\bullet$ .
  Si  $n = 2$ : devolver  $\bullet \rightarrow \bullet$ .
  Sea  $i := \lfloor \frac{n-1}{2} \rfloor$ ,  $j := \lceil \frac{n-1}{2} \rceil$ .
  Si  $i \equiv j \equiv 0(3) \vee (i \equiv 0(3) \wedge j \equiv 1(3))$ :
     $i := i - 1$ ,  $j := j + 1$ .
  Sea  $t_1 := \text{AVLHOJASMÍNIMAS}(i)$ .
  Sea  $t_2 := \text{AVLHOJASMÍNIMAS}(j)$ .
  Devolver  $t$  tal que  $\text{left}(t) = t_1$  y  $\text{right}(t) = t_2$ .

```

Figura 1: Algoritmo para generar árboles AVL de hojas mínimas

Es claro que, si  $n = 1$ , basta devolver el árbol  $\bullet$ . Si  $n = 2$ , es suficiente también con retornar cualquiera de los dos miembros de  $\mathbb{L}_2$ ; tomamos pues la decisión de utilizar  $\bullet \rightarrow \bullet$ .

Cuando  $n \geq 3$ , el algoritmo implementa la técnica desarrollada en la demostración de la Proposición 1. Las variables  $i$  y  $j$  definidas de tal manera aseguran que  $i \equiv 2(3)$ , o bien  $j \equiv 2(3)$  o bien  $i \equiv j \equiv 1(3)$  y, además, las llamadas recursivas reflejan la inducción utilizada en la demostración. De esta manera, nos aseguramos que  $t_1 \in \mathbb{L}_i$  y  $t_2 \in \mathbb{L}_j$ . El retorno de  $t$  en la última línea garantiza que  $t \in \mathbb{L}_n$  por todo lo dicho y por ser las alturas de  $t_1$  y  $t_2$  compatibles, como se analizó en la demostración mencionada.

Observemos que, si bien en la demostración el razonamiento inductivo parte desde  $n = 5$  (para asegurar la validez de la hipótesis inductiva), el algoritmo relaja esta situación partiendo desde  $n = 3$  en la porción recursiva del mismo. Esto se debe a que, siempre que  $n \in \{3, 4\}$ , los números de las llamadas recursivas quedan bien definidos en forma natural.

### 2.3.1. Complejidad temporal

Sea  $T(n)$  la cantidad de operaciones que realiza AVLHOJASMÍNIMAS en el peor caso, dado un  $n \in \mathbb{N}$ . Al tratarse de un algoritmo recursivo, podemos definir a  $T$  como una recurrencia. Para ello, notemos que el algoritmo, en cada invocación, realiza trabajo que podemos asumir de costo constante (definir  $i$  y  $j$ , tomar sus restos módulo 3, efectuar comparaciones, definir y retornar  $t$ , etc.) además de las dos llamadas recursivas. Dado que podemos acotar el tamaño de estas invocaciones por  $\frac{n}{2}$ , resulta

$$T(n) \leq 2 \cdot T\left(\frac{n}{2}\right) + O(1)$$

Es evidente entonces que  $T(n) \in O(n)$ .

### 3. Conclusiones

A partir del conjunto  $\mathcal{L}_n$  de árboles AVL de  $n$  nodos con cantidad mínima de hojas, hemos definido un conjunto  $\mathbb{L}_n$  con el objetivo de proveer una caracterización correcta y completa de  $\mathcal{L}_n$ . Se probó, en efecto, que  $\mathbb{L}_n = \mathcal{L}_n$ , lo cual valida nuestra propuesta. Además de esto, vimos que la cantidad mínima de hojas para un AVL  $t$  de  $n$  nodos es exactamente  $\lceil \frac{n+1}{3} \rceil$ , de lo cual se deduce que  $\lambda(t) \in \Omega(n)$  y, en consecuencia,  $\lambda(t) \in \Theta(n)$ .

Como corolario de los resultados encontrados, se formuló además un algoritmo que permite construir, tomando un  $n \in \mathbb{N}$ , un árbol AVL de  $n$  nodos con cantidad mínima de hojas. Se determinó, asimismo, que la complejidad temporal de este algoritmo es una  $O(n)$ .

Podría ser interesante, como trabajo a futuro, analizar el problema para árboles red-black. Al disponer de otras reglas para garantizar el balance, no está del todo claro a priori si los resultados serán equivalentes.