Discrete Mathematics and Theoretical Computer Science DMTCS vol. **20**:1, 2018, #11

Growing and Destroying Catalan–Stanley Trees

Benjamin Hackl[1] Helmut Prodinger[2]

[1] *Institut für Mathematik, Alpen-Adria-Universität Klagenfurt, Austria*
[2] *Department of Mathematical Sciences, Stellenbosch University, South Africa*

received 28th Sep. 2017, revised 31st Jan. 2018, accepted 14th Feb. 2018.

Stanley lists the class of Dyck paths where all returns to the axis are of odd length as one of the many objects enumerated by (shifted) Catalan numbers. By the standard bijection in this context, these special Dyck paths correspond to a class of rooted plane trees, so-called Catalan–Stanley trees.

This paper investigates a deterministic growth procedure for these trees by which any Catalan–Stanley tree can be grown from the tree of size one after some number of rounds; a parameter that will be referred to as the age of the tree. Asymptotic analyses are carried out for the age of a random Catalan–Stanley tree of given size as well as for the "speed" of the growth process by comparing the size of a given tree to the size of its ancestors.

Keywords: Planar trees, tree reductions, asymptotic analysis

1 Introduction

It is well-known that the nth Catalan number $C_n = \frac{1}{n+1}\binom{2n}{n}$ enumerates Dyck paths of length $2n$. In [9], Stanley lists a variety of other combinatorial interpretations of the Catalan numbers, one of them being the number of Dyck paths from $(0,0)$ to $(2n+2,0)$ such that any maximal sequence of consecutive $(1,-1)$ steps ending on the x-axis has odd length. At this point it is interesting to note that there are more subclasses of Dyck paths, also enumerated by Catalan numbers, that are defined via parity restrictions on the length of the returns to the x-axis as well (see, e.g., [1]). The height of the class of Dyck paths with odd-length returns to the origin has already been studied in [8] with the result that the main term of the height is equal to the main term of the height of general Dyck paths as investigated in [2].

By the well-known glove bijection, this special class of Dyck paths corresponds to a special class $\mathcal{S}$ of rooted plane trees, where the distance between the rightmost node in all branches attached to the root and the root is odd. This bijection is illustrated in Figure 1.

The trees in the combinatorial class $\mathcal{S}$ are the central object of study in this paper.

B. Hackl is supported by the Austrian Science Fund (FWF): P 24644-N26 and by the Karl Popper Kolleg "Modeling-Simulation-Optimization" funded by the Alpen-Adria-Universität Klagenfurt and by the Carinthian Economic Promotion Fund (KWF). This paper has been written while he was a visitor at Stellenbosch University.

H. Prodinger is supported by an incentive grant of the National Research Foundation of South Africa.

Email-addresses: benjamin.hackl@aau.at, hproding@sun.ac.za

ISSN 1365–8050

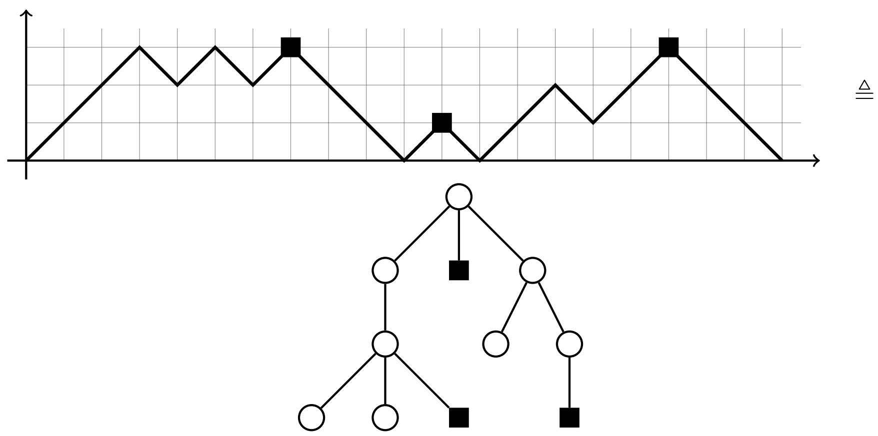

Fig. 1: Bijection between Dyck paths with odd returns to zero and Catalan–Stanley trees. ■ marks all peaks before a descent to the x-axis and all rightmost leaves in the branches attached to the root, respectively.

Definition. Let $\mathcal{S}$ be the combinatorial class of all rooted plane trees τ, where the rightmost leaves in all branches attached to the root of τ have an odd distance to the root. In particular, $\bigcirc$ itself, i.e., the tree consisting of just the root belongs to $\mathcal{S}$ as well. We call the trees in $\mathcal{S}$ *Catalan–Stanley* trees.

There are some recent approaches (see [6, 7]) in which classical tree parameters like the register function for binary trees are analyzed by, in a nutshell, finding a proper way to grow tree families in a way that the parameter of interest corresponds to the age of the tree within this (deterministic) growth process.

Following this idea, the aim of this paper is to define a "natural" growth process enabling us to grow any Catalan–Stanley tree from $\bigcirc$, and then to analyze the corresponding tree parameters.

In Section 2 we define such a growth process and analyze some properties of it. In particular, in Proposition 2.4 we characterize the family of trees that can be grown by applying a fixed number of growth iterations to some given tree family. This is then used to derive generating functions related to the parameters investigated in Sections 3 and 4.

Section 3 contains an analysis of the age of Catalan–Stanley trees, asymptotic expansions for the expected age among all trees of size n and the corresponding variance are given in Theorem 1.

Section 4 is devoted to the analysis of how fast trees of given size can be grown by investigating the size of the rth ancestor tree compared to the size of the original tree. This is characterized in Theorem 2.

We use the open-source computer mathematics software SageMath [10] with its included module for computing with asymptotic expansions [5] in order to carry out the computationally heavy parts of this paper. The corresponding worksheet can be found at `https://benjamin-hackl.at/publications/catalan-stanley/`.

2 Growing Catalan–Stanley Trees

We denote the combinatorial class of rooted plane trees with $\mathcal{T}$, and the corresponding generating function enumerating these trees with respect to their size by $T(z)$. For the sake of readability, we omit the argument of $T(z) = T$ throughout this paper. By means of the symbolic method [4, Chapter I], the combinatorial class $\mathcal{T}$ satisfies the construction $\mathcal{T} = \bigcirc \times \operatorname{SEQ}(\mathcal{T})$. It translates into the functional equation

$$T(z) = \frac{z}{1 - T(z)} \qquad \Longleftrightarrow \qquad z + T(z)^2 = T(z), \tag{1}$$

which will be used throughout the paper. Additionally, it is easy to see by solving the quadratic equation in (1) and choosing the correct branch of the solution, we have the well-known formula $T(z) = \frac{1-\sqrt{1-4z}}{2}$.

Proposition 2.1. *The generating function of the combinatorial class $\mathcal{S}$ of Catalan–Stanley trees, where t marks all the rightmost nodes in the branches attached to the root of the tree and z marks all other nodes, is given by*

$$S(z,t) = z + \frac{zt}{1 - t - T^2}. \tag{2}$$

In particular, there is one Catalan–Stanley tree of size 1 and C_{n-2} Catalan–Stanley trees of size n for $n \geq 2$.

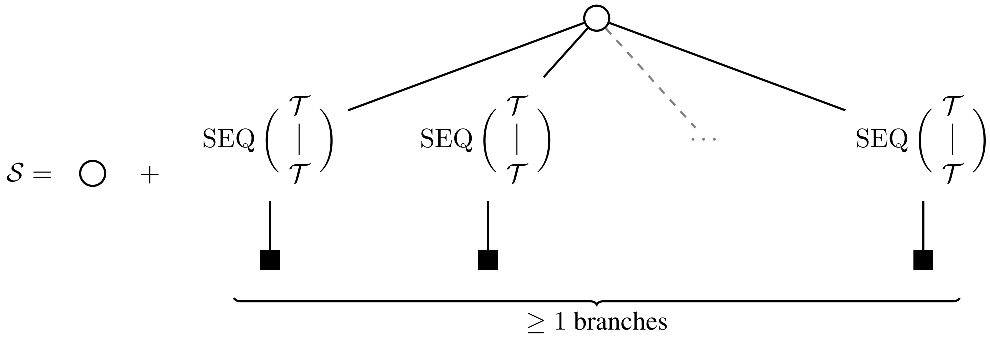

Fig. 2: Symbolic specification of the combinatorial class $\mathcal{S}$ of Catalan–Stanley trees. Nodes represented by ■ are marked by the variable t, all other nodes are marked by z.

Proof: By using the symbolic method [4, Chapter I], the symbolic representation of $\mathcal{S}$ given in Figure 2 (which is based on the decomposition of the rightmost path in each subtree of the root into a sequence of pairs of rooted plane trees and the final rightmost leaf ■) translates into the functional equation

$$S(z,t) = z + \frac{z\frac{t}{1-T^2}}{1 - \frac{t}{1-T^2}},$$

which simplifies to the equation given in (2).

In order to enumerate Catalan–Stanley trees with respect to their size, we consider $S(z,z)$, which simplifies to $z(T+1)$ and thus proves the statement. □

We want to describe how to grow all Catalan–Stanley trees beginning from the tree that has only one node, ○.

We consider the tree reduction $\rho\colon \mathcal{S} \to \mathcal{S}$ that operates on a given Catalan–Stanley tree τ (or just the root) as follows:

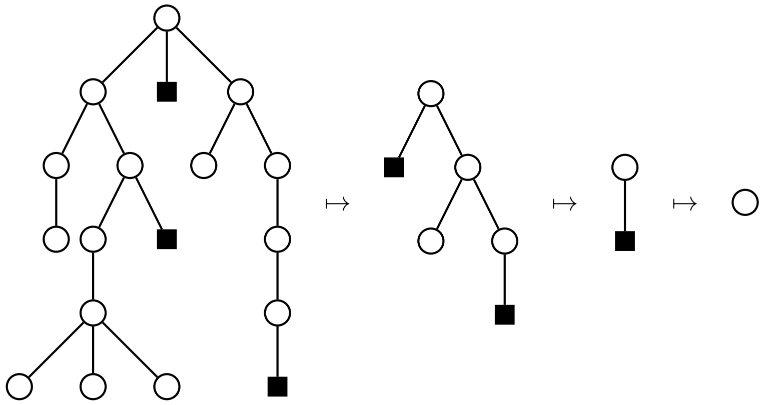

Fig. 3: Illustration of the reduction operator ρ, ■ marks the rightmost leaves in the branches attached to the root.

Start from all nodes that are represented by t, i.e. the rightmost leaves in the branches attached to the root: if the node is a child of the root, it is simply deleted. Otherwise we delete all subtrees of the grandparent of the node and mark the resulting leaf, i.e., the former grandparent, with t.

This tree reduction is illustrated in Figure 3. While the reduction ρ is certainly not injective as there are several trees with the same reduction $\tau \in \mathcal{S}$, it is easy to construct a tree reducing to some given $\tau \in \mathcal{S}$ by basically inserting chains of length 2 before all rightmost leaves in the branches attached to the root. This allows us to think of the operator ρ^{-1} mapping a given tree (or some family of trees) to the respective set of preimages as a *tree expansion operator*. In this context, we also want to define the *age* of a Catalan–Stanley tree.

Definition. Let $\tau \in \mathcal{S}$ be a Catalan–Stanley tree. Then we define $\alpha(\tau)$, the *age* of τ, to be the number of expansions required to grow τ from the tree of size one, $\bigcirc$. In particular, we want

$$\alpha(\tau) = r \qquad \Longleftrightarrow \qquad \tau \in (\rho^{-1})^r(\bigcirc) \text{ and } \tau \notin (\rho^{-1})^{r-1}(\bigcirc)$$

for $r \in \mathbb{Z}_{\geq 1}$, and we set $\alpha(\bigcirc) = 0$.

Remark. Naturally, the concept of the age of a tree strongly depends on the underlying reduction procedure. In particular, for the reduction procedure considered in this article we have $\alpha(\tau) = r$ if and only if the maximum depth of the rightmost leaves in the branches attached to the root is $2r - 1$.

Before we delve into the analysis of the age of Catalan–Stanley trees, we need to be able to translate the tree expansion given by ρ^{-1} into a suitable form so that we can actually use it in our analysis. The following proposition shows that ρ^{-1} can be expressed in the language of generating functions.

Proposition 2.2. *Let $\mathcal{F} \subseteq \mathcal{S}$ be a family of Catalan–Stanley trees with bivariate generating function $f(z,t)$, where t marks rightmost leaves in the branches attached to the root and z marks all other nodes. Then the generating function of $\rho^{-1}(\mathcal{F})$, the family of trees whose reduction is in $\mathcal{F}$, is given by*

$$\Phi(f(z,t)) = \frac{1}{1-t} f\Big(z, \frac{t}{1-t} T^2\Big). \tag{3}$$

Proof: From a combinatorial point of view it is obvious that the operator Φ has to be linear, meaning that we can focus on determining all possible expansions of some tree represented by the monomial $z^n t^k$, i.e.

a tree where the root has k children (and thus k different rightmost leaves in the branches attached to the root), and n other nodes.

In order to expand such a tree represented by $z^n t^k$ we begin by inserting a chain of length two before every rightmost leaf in order to ensure that the distance to the root is still odd. These newly inserted nodes can now be considered to be roots of some rooted plane trees, meaning that we actually insert two arbitrary rooted plane trees before every node represented by t. This corresponds to a factor of $t^k T^{2k}$.

In addition to this operation, we are also allowed to add new children to the root, i.e. we can add sequences of nodes represented by t before or after every child of the root. As observed above, the root has k children and thus there are $k+1$ positions where such a sequence can be attached. This corresponds to a factor of $(1-t)^{-(k+1)}$.

Finally, we observe that nodes that are represented by z are not expanded in any way, meaning that z^n remains as it is.

Putting everything together yields that

$$\Phi(z^n t^k) = \frac{1}{1-t} z^n \Big(\frac{tT^2}{1-t}\Big)^k,$$

which, by linearity of Φ, proves the statement. □

Corollary 2.3. *The generating function $S(z,t)$ satisfies the functional equation*

$$\Phi(S(z,t)) = S(z,t).$$

Proof: This follows immediately from the fact that the reduction operator ρ is surjective, as discussed above. □

Actually, in order to carry out a thorough analysis of this growth process for Catalan–Stanley trees we need to have more information on the iterated application of the expansion. In particular, we need a precise characterization of the family of Catalan–Stanley trees that can be grown from some given tree family by expanding it a fixed number of times.

Proposition 2.4. *Let $r \in \mathbb{Z}_{\geq 0}$ be fixed and $\mathcal{F} \subseteq \mathcal{S}$ be a family of Catalan–Stanley trees with bivariate generating function $f(z,t)$. Then the family of trees obtained by expanding the trees in $\mathcal{F}$ r times is enumerated by the generating function*

$$\Phi^r(f(z,t)) = \frac{1}{1 - t\frac{1-T^{2r}}{1-T^2}} f\Big(z, \frac{tT^{2r}}{1 - t\frac{1-T^{2r}}{1-T^2}}\Big). \tag{4}$$

Proof: By linearity, it is sufficient to determine the generating function for the family of trees obtained by expanding some tree represented by $z^n t^k$. Consider the closely related multiplicative operator Ψ with

$$\Psi(f(z,t)) = f\big(z, \frac{t}{1-t} T^2\big).$$

It is easy to see that we can write the r-fold application of Φ with the help of Ψ as

$$\Phi^r(f(z,t)) = \Psi^r(f(z,t)) \prod_{j=0}^{r-1} \frac{1}{1 - \Psi^j(t)}.$$

As Ψ is multiplicative, we have

$$\Psi^r(z^n t^k) = \Psi^r(z)^n \Psi^r(t)^k,$$

meaning that we only have to investigate the r-fold application of Ψ to z and to t.

We immediately see that $\Psi^r(z) = z$, as Ψ maps z to z itself. For $\Psi^r(t)$, we can prove by induction that the relation

$$\Psi^r(t) = \frac{tT^{2r}}{1 - t\frac{1-T^{2r}}{1-T^2}}$$

holds for $r \geq 0$. Finally, observe that for $j \geq 1$ we have

$$\Psi^j(t) = \frac{\Psi^{j-1}(t)}{1 - \Psi^{j-1}(t)} T^2, \tag{5}$$

and thus

$$\Psi^r(t) = \frac{\Psi^{r-1}(t)}{1 - \Psi^{r-1}(t)} T^2 = \frac{\Psi^{r-2}(t)}{(1 - \Psi^{r-2}(t))(1 - \Psi^{r-1}(t))} T^4 = \cdots = \frac{tT^{2r}}{\prod_{j=0}^{r-1}(1 - \Psi^j(t))}$$

by iteratively using (5) in the numerator. With our explicit formula for $\Psi^r(t)$ from above this yields

$$\prod_{j=0}^{r-1}(1 - \Psi^j(t)) = 1 - t\frac{1 - T^{2r}}{1 - T^2}$$

for $r \geq 1$. Putting everything together we obtain

$$\Phi^r(z^n t^k) = \frac{1}{1 - t\frac{1-T^{2r}}{1-T^2}} z^n \Psi^r(t)^k,$$

which proves (4) by linearity of Φ^r. □

From this characterization we immediately obtain the generating functions for all the classes of objects we will investigate in the following sections.

Corollary 2.5. *Let $r \in \mathbb{Z}_{\geq 0}$. The generating function $F_r^{\leq}(z,t)$ enumerating Catalan–Stanley trees of age less than or equal to r where t marks the rightmost leaves in the branches attached to the root and z marks all other nodes is given by*

$$F_r^{\leq}(z,t) = \frac{z}{1 - t\frac{1-T^{2r}}{1-T^2}}. \tag{6}$$

Proof: As we defined $\rho(\bigcirc) = \bigcirc$ we have $\bigcirc \in \rho^{-1}(\bigcirc)$, which implies $F_r^{\leq}(z,t)$ is given by $\Phi^r(z)$. □

Corollary 2.6. *Let $r \geq 0$. Then the generating function $G_r(z,v)$ enumerating Catalan–Stanley trees where z marks the tree size and v marks the size of the r-fold reduced tree, is given by*

$$G_r(z,v) = \Phi^r(S(zv,tv))|_{t=z} = \frac{1}{1 - z\frac{1-T^{2r}}{1-T^2}} S\biggl(zv, \frac{zT^{2r}}{1 - z\frac{1-T^{2r}}{1-T^2}} v\biggr). \tag{7}$$

Proof: Observe that the generating function $S(zv, tv)$ enumerates Catalan–Stanley trees with respect to the number of rightmost leaves in the branches attached to the root (marked by t), the number of other nodes (marked by z), and the size of the tree (marked by v). Applying the operator Φ^r to this generating function thus yields a generating function where v still marks the size of the tree, t and z however enumerate the number of rightmost leaves in the branches attached to the root and all other nodes of the r-fold expanded tree, respectively. After setting $t = z$, we obtain a generating function where v marks the size of the original tree and z the size of the r-fold expanded tree—which is equivalent to the formulation in the corollary. □

3 The Age of Catalan–Stanley Trees

In this section we want to give a proper analysis of the parameter α defined in the previous section. Formally, we do this by considering the random variable D_n modeling the age of a tree of size n, where all Catalan–Stanley trees of size n are equally likely.

Remark. It is noteworthy that in [7] it was shown that the well-known register function of a binary tree can also be obtained as the number of times some reduction can be applied to the binary tree until it degenerates. The age of a Catalan–Stanley tree can thus be seen as a "register function"-type parameter as well.

First of all, we are interested in the minimum and maximum age a tree of size n can have.

Proposition 3.1. *Let $n \in \mathbb{Z}_{\geq 2}$. Then the bounds*

$$1 \leq D_n \leq \left\lfloor \frac{n}{2} \right\rfloor \tag{8}$$

hold and are sharp, i.e. there are trees τ, $\tau' \in \mathcal{S}$ of size $n \geq 2$ such that $D_n(\tau) = 1$ and $D_n(\tau') = \lfloor n/2 \rfloor$ hold. The only tree of size 1 is $\bigcirc$, and it satisfies $D_1(\bigcirc) = 0$.

Proof: Note that only $\bigcirc$, the tree of size 1 has age 0, therefore the lower bound is certainly valid for trees of size $n \geq 2$. This lower bound is sharp, as the tree with $n - 1$ children attached to the root is a Catalan–Stanley tree and has age 1.

For the upper bound, first observe that given a tree of size $n \geq 3$ the reduction ρ always removes at least 2 nodes from the tree. If the tree is of size 2, then ρ only removes one node. Given an arbitrary Catalan–Stanley tree τ of age r and size n, this means that

$$1 = |\bigcirc| = |\rho^r(\tau)| \leq |\tau| - 2 \cdot (r - 1) - 1 = n - 2r + 1,$$

where $|\tau|$ denotes the size of the tree τ. This yields $r \leq n/2$, and as r is known to be an integer we may take the floor of the number on the right-hand side of the inequality. This proves that the upper bound in (8) is valid.

The upper bound is sharp because we can construct appropriate families of trees precisely reaching the upper bound: for even n, the chain of size n is a Catalan–Stanley tree of age $n/2$. For odd $n = 2m+1$ we consider the chain of size $2m$ and attach one node to the root of it. The resulting tree is a Catalan–Stanley tree of age $m = \lfloor n/2 \rfloor$, and thus proves that the bound is sharp. □

By investigating the generating functions obtained from Corollary 2.5 we can characterize the limiting distribution of the age of Catalan–Stanley trees when the size n tends to ∞.

Theorem 1. *Consider $n \to \infty$. Then the age of a (uniformly random) Catalan–Stanley tree of size n behaves according to a discrete limiting distribution where*

$$\begin{aligned}\mathbb{P}(D_n = r) = \Big(&\frac{4(4^r(3r-1)+1)}{(4^r+2)^2} - \frac{4(4^{r+1}(3r+2)+1)}{(4^{r+1}+2)^2}\Big) \\ - \Big(&\frac{6\cdot 64^r(2r^3-5r^2+4r-1) - 6\cdot 16^r(16r^3-24r^2+10r-1) + 24\cdot 4^r(2r^3-r^2)}{(4^r+2)^4} \\ -&\frac{6\cdot 64^{r+1}(2r^3+r^2) - 6\cdot 16^{r+1}(16r^3+24r^2+10r+1) + 24\cdot 4^{r+1}(2r^3+5r^2+4r+1)}{(4^{r+1}+2)^4}\Big)n^{-1} \\ &+ O\Big(\frac{r^5}{3^r}n^{-2}\Big) \qquad (9)\end{aligned}$$

for $r \in \mathbb{Z}_{\geq 1}$, and the O-term holds uniformly in r. Additionally, by setting

$$\begin{aligned} c_0 &= \sum_{r\geq 1} \frac{4^{r+1}(3r-1)+4}{(4^r+2)^2} \\ &= 2.71825364286795285266483619282193673445854356803 44\ldots, \\ c_1 &= -\sum_{r\geq 1} \frac{6\cdot 64^r(2r^3-5r^2+4r-1) - 6\cdot 16^r(16r^3-24r^2+10r-1) + 24\cdot 4^r(2r-1)r^2}{(4^r+2)^4} \\ &= -4.22209715101588408238218734776004780808164112104 06\ldots, \\ c_2 &= \sum_{r\geq 1}(2r-1)\frac{4^{r+1}(3r-1)+4}{(4^r+2)^2} - c_0^2 \\ &= 0.918456042143747973577971478140194965036889539339 67\ldots, \\ c_3 &= -\sum_{r\geq 1}\frac{(2r-1)}{(4^r+2)^4}\big(6\cdot 64^r(2r^3-5r^2+4r-1) - 6\cdot 16^r(16r^3-24r^2+10r-1) \\ &\qquad + 24\cdot 4^r(2r-1)r^2\big) - 2c_0c_1 \\ &= -9.16217532008362749969124365683102689885365345949 42\ldots, \end{aligned}$$

the expected age and the corresponding variance are given by the asymptotic expansions

$$\mathbb{E}D_n = c_0 + c_1 n^{-1} + O(n^{-2}), \qquad (10)$$

$$\mathbb{V}D_n = c_2 + c_3 n^{-1} + O(n^{-2}). \qquad (11)$$

Proof: For the sake of convenience we set $F_r^{\leq}(z) := F_r^{\leq}(z,z)$, where $F_r^{\leq}(z,t)$ is given in (6). This univariate generating function now enumerates Catalan–Stanley trees of age $\leq r$ with respect to the tree size.

We begin by observing that $F_r^{\geq}(z)$, the generating function enumerating Catalan–Stanley trees of age $\geq r$ with respect to the tree size is given by

$$F_r^{\geq}(z) = S(z,z) - F_{r-1}^{\leq}(z) = z(1+T) - \frac{z}{1 - z\frac{1-T^{2r-2}}{1-T^2}} = z(1+T)\frac{T^{2r-1}}{1+T^{2r-1}}, \tag{12}$$

where the last equation follows after some elementary manipulations and by using (1).

Now let $f_{n,r} := [z^n]F_r^{\geq}(z)$ denote the number of Catalan–Stanley trees of size n and age $\geq r$. As we consider all Catalan–Stanley trees of size n to be equally likely, we find

$$\mathbb{P}(D_n = r) = \mathbb{P}(D_n \geq r) - \mathbb{P}(D_n \geq r+1) = \frac{f_{n,r} - f_{n,r+1}}{C_{n-2}}.$$

We use singularity analysis (see [3] and [4, Chapter VI]) in order to obtain an asymptotic expansion for $f_{n,r}$. To do so, we first observe that $z = 1/4$ is the dominant singularity of T and thus also of $F_r^{\geq}(z)$. We then consider z to be in some Δ-domain at $1/4$ (see [4, Definition VI.1]). The task of expanding $F_r^{\geq}(z)$ for $z \to 1/4$ now largely consists of handling the term $\frac{T^{2r-1}}{1+T^{2r-1}}$. Observe that we can write

$$\begin{aligned}\frac{T^{2r-1}}{1+T^{2r-1}} &= \frac{1}{1+T^{1-2r}} = \frac{1}{1+2^{2r-1}(1-\sqrt{1-4z}\,)^{1-2r}},\\ &= \frac{1}{(1+2^{2r-1})\big(1+\frac{2^{2r-1}}{1+2^{2r-1}}\sum_{j\geq 1}\binom{2r+j-2}{j}(1-4z)^{j/2}\big)}\end{aligned}$$

which results in

$$\begin{aligned}\frac{T^{2r-1}}{1+T^{2r-1}} = \frac{2}{4^r+2} &- \frac{2\cdot 4^r(2r-1)}{(4^r+2)^2}(1-4z)^{1/2}\\ &+ \frac{2\cdot 4^r\,(4^r(r-1)-2r)(2r-1)}{(4^r+2)^3}(1-4z)\\ - \frac{2\cdot 4^r(16^r(2r^2-5r+3) - 4^{r+2}(r^2-r) + 8r^2+4r)(2r-1)}{3(4^r+2)^4}&(1-4z)^{3/2} + O\Big(\frac{r^4}{3^r}(1-4z)^2\Big),\end{aligned}$$

where the O-term holds uniformly in r. Multiplying this expansion with the expansion of $z(1+T)$ yields the expansion

$$\begin{aligned}F_r^{\geq}(z) = \frac{3}{4(4^r+2)} &- \frac{4^r(3r-1)+1}{2(4^r+2)^2}(1-4z)^{1/2}\\ &+ \frac{16^r(6r^2-7r-1) - 2\cdot 4^r(6r^2-5r+7) - 12}{4(4^r+2)^3}(1-4z)\\ - \frac{64^r(2r^3-5r^2+r) - 2\cdot 16^r(8r^3-12r^2+11r-2) + 4^{r+1}(2r^3-r^2-3r) - 4}{2(4^r+2)^4}&(1-4z)^{3/2}\\ &+ O\Big(\frac{r^4}{3^r}(1-4z)^2\Big).\end{aligned}$$

By means of singularity analysis we extract the nth coefficient and find

$$
\begin{aligned}
f_{n,r} = \frac{4^r(3r-1)+1}{4\sqrt{\pi}\,(4^r+2)^2}4^n n^{-3/2} - \Big(&\frac{3\cdot 64^r(8r^3-20r^2+r+1)}{32\sqrt{\pi}\,(4^r+2)^4} \\
&- \frac{3\cdot 16^r(64r^3-96r^2+100r-19)-12\cdot 4^r(8r^3-4r^2-15r)+60}{32\sqrt{\pi}\,(4^r+2)^4}\Big)4^n n^{-5/2} \\
&+ O\Big(\frac{r^5}{3^r}4^n n^{-7/2}\Big).
\end{aligned}
$$

Computing the difference $f_{n,r} - f_{n,r+1}$ and dividing by the Catalan number C_{n-2} then yields the expression for $\mathbb{P}(D_n = r)$ given in (9).

The expected value can then be computed with the help of the well-known formula

$$\mathbb{E}D_n = \sum_{r\geq 1}\mathbb{P}(D_n \geq r),$$

which proves (10). Finally, the variance can be obtained from $\mathbb{V}D_n = \mathbb{E}(D_n^2) - (\mathbb{E}D_n)^2$, where

$$\mathbb{E}(D_n^2) = \sum_{r\geq 1} r^2\mathbb{P}(D_n = r) = \sum_{r\geq 1}(2r-1)\mathbb{P}(D_n \geq r),$$

which proves (11). □

In addition to the asymptotic expansions given in Theorem 1 we can also determine an exact formula for the expected value $\mathbb{E}D_n$. The key tools in this context are Cauchy's integral formula as well as the substitution $z = \frac{u}{(1+u)^2}$.

Proposition 3.2. *Let $n \in \mathbb{Z}_{\geq 2}$. The expected age of the Catalan–Stanley trees of size n is given by*

$$\mathbb{E}D_n = \frac{1}{C_{n-2}}\sum_{k\geq 1}(-1)^{k+1}\sigma_0^{\mathrm{odd}}(k)\left(\binom{2n-4-k}{n-3} + \binom{2n-4-k}{n-2} - 2\binom{2n-4-k}{n-1}\right), \tag{13}$$

where $\sigma_0^{\mathrm{odd}}(k)$ denotes the number of odd divisors of k.

Proof: We begin by explicitly extracting the coefficient $[z^n]F_r^{\geq}(z)$. The expected value can then be obtained by summation over r and division by C_{n-2}.

With the help of the substitution $z = \frac{u}{(1+u)^2}$ we can bring $F_r^{\geq}(z)$ into the more suitable form

$$F_r^{\geq}(z) = \frac{(1+2u)u^{2r}}{(1+u)^3(u^{2r-1}+(1+u)^{2r-1})}.$$

We extract the coefficient of z^n now by means of Cauchy's integral formula. Let γ be a small contour winding around the origin once. Then we have

$$[z^n]F_r^{\geq}(z) = \frac{1}{2\pi i}\oint_\gamma \frac{F_r^{\geq}(z)}{z^{n+1}}\,dz = \frac{1}{2\pi i}\oint_{\tilde{\gamma}} \frac{(1+u)^{2n+2}}{u^{n+1}}\frac{(1+2u)u^{2r}}{(1+u)^3(u^{2r-1}+(1+u)^{2r-1})}\frac{1-u}{(1+u)^3}\,du$$

$$
\begin{aligned}
&= [u^{n-2r}](1+2u)(1-u)(1+u)^{2n-2r-3}\frac{1}{1+(\frac{u}{1+u})^{2r-1}}\\
&= [u^{n-2r}](1+u-2u^2)\sum_{j\ge 1}(-1)^{j-1}u^{(2r-1)(j-1)}(1+u)^{2n-4-j(2r-1)}\\
&= \sum_{j\ge 1}(-1)^{j-1}\Bigg(\binom{2n-4-j(2r-1)}{n-3}+\binom{2n-4-j(2r-1)}{n-2}\\
&\qquad\qquad -2\binom{2n-4-j(2r-1)}{n-1}\Bigg),
\end{aligned}
\tag{14}
$$

where $\tilde{\gamma}$, the integration contour of the second integral, is the transformation of γ under the transformation $z = u/(1+u)^2$ and is also a small contour winding around the origin once.

Now consider the auxiliary sum

$$\vartheta(k) := \sum_{\substack{j,r\ge 1\\ j(2r-1)=k}} (-1)^{j-1}.$$

It is easy to see by distinguishing between even and odd k that with the help of $\sigma_0^{\mathrm{odd}}(k)$, $\vartheta(k)$ can be written as $\vartheta(k) = (-1)^{k-1}\sigma_0^{\mathrm{odd}}(k)$.

Summing the expression from (14) over $r \ge 1$, simplifying the resulting double sum by means of the auxiliary sum ϑ, and finally dividing by C_{n-2} then proves (13). □

4 Analysis of Ancestors

In this section we focus on characterizing the effect of the (repeatedly applied) reduction ρ on a random Catalan–Stanley tree of size n. We are particularly interested in studying the size of the reduced tree. In the light of the fact that all Catalan–Stanley trees can be grown from ○ by means of the growth process induced by ρ, we can think of the rth reduction of some tree τ as the rth ancestor of τ.

In order to formally conduct this analysis, we consider the random variable $X_{n,r}$ modeling the size of the rth ancestor of some tree of size n, where all Catalan–Stanley trees of size n are equally likely.

Similar to our approach in Proposition 3.1 we can determine precise bounds for $X_{n,r}$ as well.

Proposition 4.1. *Let $n \in \mathbb{Z}_{\ge 2}$ and $r \in \mathbb{Z}_{\ge 1}$. Then the bounds*

$$1 \le X_{n,r} \le n - 2(r-1) - 1 \tag{15}$$

hold for $r \le \lfloor n/2 \rfloor$ and are sharp, i.e. there are trees τ, $\tau' \in \mathcal{S}$ of size $n \ge 2$ such that $X_{n,r}(\tau) = 1$ and $X_{n,r}(\tau') = n - 2(r-1) - 1$. For $r > \lfloor n/2 \rfloor$ the variable $X_{n,r}$ is deterministic with $X_{n,r} = 1$.

Proof: Assume that $r \le \lfloor n/2 \rfloor$. The lower bound is obvious as trees cannot reduce further than to ○, and as the first ancestor of the tree with $n-1$ children attached to the root already is ○ the lower bound is valid and sharp.

For the upper bound we follow the same argumentation as in the proof of Proposition 3.1 to arrive at

$$1 \le |\rho^r(\tau)| \le |\tau| - 2(r-1) - 1 = n - 2r + 1$$

for some Catalan–Stanley tree of size n, which proves that the upper bound is valid. Any tree τ of size n having the chain of length 2 as its $(r-1)$th ancestor satisfies $X_{n,r}(\tau) = n - 2(r-1) - 1$ and thus proves that the upper bound is sharp. This proves (15).

In the case of $r > \lfloor n/2 \rfloor$ we observe that as the $\lfloor n/2 \rfloor$th ancestor of any Catalan–Stanley tree of size n already is certain to be ○ by Proposition 3.1, the rth ancestor is ○ as well. □

With the generating function $G_r(z,v)$ enumerating Catalan–Stanley trees with respect to their size (marked by n) and the size of their rth ancestor (marked by v) from Corollary 2.6 we can write the probability generating function of $X_{n,r}$ as

$$\mathbb{E}v^{X_{n,r}} = \frac{1}{C_{n-2}}[z^n]G_r(z,v).$$

This allows us to extract parameters like the expected size of the rth ancestor and the corresponding variance.

Theorem 2. *Let $r \in \mathbb{Z}_{\geq 0}$ be fixed and consider $n \to \infty$. Then the expected value and the variance of the random variable $X_{n,r}$ modeling the size of the rth ancestor of a (uniformly random) Catalan–Stanley tree of size n are given by the asymptotic expansions*

$$\mathbb{E}X_{n,r} = \frac{1}{4^r}n + \frac{2\cdot 4^r - 2r^2 + r - 2}{2\cdot 4^r} + \frac{(2r+1)(2r-1)(r-3)r}{2\cdot 4^{r+1}}n^{-1} + O(n^{-3/2}), \tag{16}$$

$$\begin{aligned}\mathbb{V}X_{n,r} = \frac{(2^r+1)(2^r-1)}{16^r}n^2 - \frac{\sqrt{\pi}(4^r(3r+1)-1)}{3\cdot 16^r}n^{3/2} \\ + \frac{18\cdot 4^r r^2 + 3\cdot 4^r r - 38\cdot 4^r + 36r^2 - 42r + 38}{18\cdot 16^r}n \\ + \frac{5\sqrt{\pi}\,(4^r(3r+1)-1)}{8\cdot 16^r}n^{1/2} + O(1).\end{aligned} \tag{17}$$

Proof: The strategy behind this proof is to determine the first and second factorial moment of $X_{n,r}$ by extracting the coefficient of z^n in the derivatives $\frac{\partial^d}{\partial v^d}G_r(z,v)|_{v=1}$ for $d \in \{1,2\}$ and normalizing the result by dividing by C_{n-2}.

We begin with the expected value. With the help of SageMath [10] we find for $z \to 1/4$

$$\begin{aligned}\frac{\partial}{\partial v}G_r(z,v)|_{v=1} = \frac{1}{4^{r+2}}(1-4z)^{-1/2} + \frac{3\cdot 4^r - r}{2\cdot 4^{r+1}} - \frac{2\cdot 4^r - 2r^2 + r + 2}{4^{r+2}}(1-4z)^{1/2} \\ - \frac{9\cdot 4^r + 2r^3 - 3r^2 - 5r}{6\cdot 4^{r+1}}(1-4z) + O((1-4z)^{3/2}),\end{aligned}$$

where the O-constant depends implicitly on r. Extracting the coefficient of z^n and dividing by C_{n-2} yields the expansion given in (16).

Following the same approach for the second derivative yields the expansion

$$\frac{\partial^2}{\partial v^2}G_r(z,v)|_{v=1} = \frac{1}{2\cdot 4^{r+2}}(1-4z)^{-3/2} - \frac{4^r(3r+1)-1}{3\cdot 16^{r+1}}(1-4z)^{-1}$$

$$+\frac{4^r(18r^2+3r+7)-24r+2}{18\cdot 16^{r+1}}(1-4z)^{-1/2}+O(1),$$

such that after applying singularity analysis and division by C_{n-2} we obtain the expansion

$$\mathbb{E}X_{n,r}^{\underline{2}}=\frac{1}{4^r}n^2-\frac{\sqrt{\pi}\,(4^r(3r+1)-1)}{3\cdot 16^r}n^{3/2}+\frac{4^r(18r^2+3r-20)-24r+2}{18\cdot 16^r}n+\frac{5\sqrt{\pi}\,(4^r(3r+1)-1)}{8\cdot 16^r}n^{1/2}+O(1)$$

for the second factorial moment $\mathbb{E}X_{n,r}^{\underline{2}}$. Applying the well-known formula

$$\mathbb{V}X_{n,r}=\mathbb{E}X_{n,r}^{\underline{2}}+\mathbb{E}X_{n,r}-(\mathbb{E}X_{n,r})^2$$

then leads to the asymptotic expansion for the variance given in (17) and thus proves the statement. □

Besides the asymptotic expansion given in Theorem 2, we are also interested in finding an exact formula for the expected value $\mathbb{E}X_{n,r}$. We can do so by means of Cauchy's integral formula.

Proposition 4.2. *Let n, $r\in\mathbb{Z}_{\geq 1}$. Then the expected size of the rth ancestor of a random Catalan–Stanley tree of size n is given by*

$$\mathbb{E}X_{n,r}=\frac{1}{C_{n-2}}\binom{2n-2r-4}{n-2}+1. \tag{18}$$

Proof: We rewrite the derivative $g(z):=\frac{\partial}{\partial v}G_r(z,v)|_{v=1}$ into a more suitable form which makes it easier to extract the coefficients. To do so, we use the substitution $z=u/(1+u)^2$ again, allowing us to express the derivative as

$$g(z)=\frac{u^{2r+2}}{(1-u)(1+u)^{2r+3}}+\frac{(1+2u)u}{(1+u)^3}.$$

Note that as $T=\frac{u}{1+u}$, the summand $\frac{(1+2u)u}{(1+u)^3}$ actually represents $z(1+T)$, implying that the coefficient of z^n in this summand is given by C_{n-2}. Now let γ be a small contour winding around the origin once, so that with Cauchy's integral formula we obtain

$$\begin{aligned}[z^n]g(z)&=\frac{1}{2\pi i}\oint_\gamma\frac{g(z)}{z^{n+1}}\,dz=\frac{1}{2\pi i}\oint_{\tilde\gamma}\frac{(1+u)^{2n+2}}{u^{n+1}}\frac{u^{2r+2}}{(1-u)(1+u)^{2r+3}}\frac{1-u}{(1+u)^3}\,du+C_{n-2}\\&=[u^{n-2r-2}](1+u)^{2n-2r-4}+C_{n-2}=\binom{2n-2r-4}{n-2r-2}+C_{n-2},\end{aligned}$$

where $\tilde\gamma$ is the image of γ under the transformation (and is still a small contour winding around the origin once). Dividing by C_{n-2} then proves (18). □

References

[1] David Callan, *The 136th manifestation of C_n*, arXiv:math/0511010 [math.CO], 2005.

[2] Nicolaas G. de Bruijn, Donald E. Knuth, and Stephen O. Rice, *The average height of planted plane trees*, Graph theory and computing, Academic Press, New York, 1972, pp. 15–22. MR 0505710 (58 #21737)

[3] Philippe Flajolet and Andrew Odlyzko, *Singularity analysis of generating functions*, SIAM J. Discrete Math. **3** (1990), 216–240. MR MR1039294 (90m:05012)

[4] Philippe Flajolet and Robert Sedgewick, *Analytic combinatorics*, Cambridge University Press, Cambridge, 2009.

[5] Benjamin Hackl, Clemens Heuberger, and Daniel Krenn, *Asymptotic expansions in SageMath*, `http://trac.sagemath.org/17601`, 2015, module in SageMath 6.10.

[6] Benjamin Hackl, Clemens Heuberger, Sara Kropf, and Helmut Prodinger, *Fringe analysis of plane trees related to cutting and pruning*, Aequationes Math. (2018).

[7] Benjamin Hackl, Clemens Heuberger, and Helmut Prodinger, *Reductions of binary trees and lattice paths induced by the register function*, Theoret. Comput. Sci. **705** (2018), 31–57. MR 3721457

[8] Helmut Prodinger, *Dyck paths with parity restrictions for the final runs to the origin: a study of the height*, Fund. Inform. **117** (2012), no. 1–4, 279–285. MR 2977924

[9] Richard P. Stanley, *Catalan numbers*, Cambridge University Press, Cambridge, 2015.

[10] The SageMath Developers, *SageMath Mathematics Software (Version 7.6)*, 2017, `http://www.sagemath.org`.