DEEPEST NODES IN MARKED ORDERED TREES

HELMUT PRODINGER

ABSTRACT. A variation of ordered trees, where each rightmost edge might be marked or not, if it does not lead to an endnode, is investigated. These marked ordered trees were introduced by E. Deutsch et al. to model skew Dyck paths. We study the number of deepest nodes in such trees. Explicit generating functions are established and the average number of deepest nodes, which approaches $\frac{5}{3}$ when the number of nodes gets large. This is to be compared to standard ordered trees where the average number of deepest nodes approaches 2.

1. INTRODUCTION

In [2] we find the following variation of ordered trees: Each rightmost edge might be marked or not, if it does not lead to an endnode (leaf). They were introduced to model skew Dyck paths using trees.

We depict a marked edge by the red colour and draw all of them of size 4 (4 nodes) in a table at the end of this introductory section.

Now we move to a symbolic equation for the marked ordered trees:

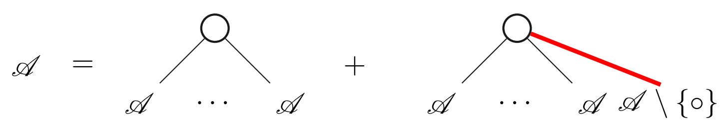


FIGURE 1. Symbolic equation for marked ordered trees.
$\mathscr{A} \cdots \mathscr{A}$ refers to ≥ 0 copies of $\mathscr{A}$.

Recall that ordered (plane, planted plane) trees are simpler and are given by deleting the last component with the red edge.

We also bring the notion of height into the game (length of longest chain of the root to a leaf, measured in the number of nodes). Let $\mathscr{A}_h$ denote the family of marked ordered trees with height $\leq h$. Then

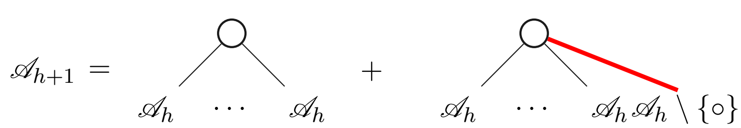


FIGURE 2. Symbolic equation for marked ordered trees of bounded height for height $h \geq 2$.

Date: February 14, 2022.
2020 *Mathematics Subject Classification.* 05A15.

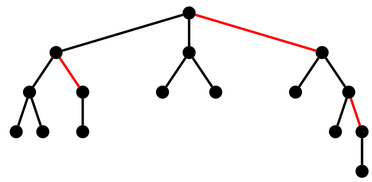


FIGURE 3. Example of a marked ordered tree.

The classical bijection between ordered trees and Dyck paths consists of walking around the tree, and recording an up-step when walking down and recording a down-step when walking up. This can be adapted to marked ordered trees to produce decorated Dyck paths. The additional rule is to record a red down-step when walking up a red (marked) edge.

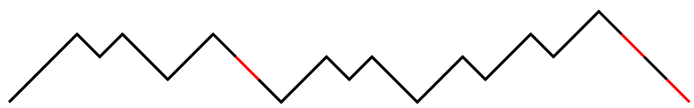


FIGURE 4. The decorated Dyck path corresponding to Figure 3.

Decorated Dyck paths are in (simple) bijection to skew Dyck path, by replacing each red down-step by a south-west ($= (-1,-1)$) step. The next tables show all marked treed of size 4 (4 nodes) and the corresponding objects.

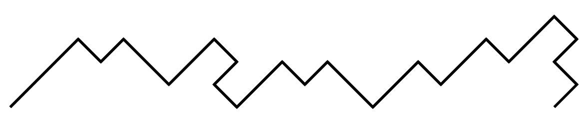


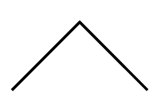
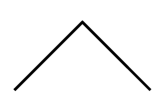
FIGURE 5. The skew Dyck path corresponding to Figure 4.

A representative example of trees and corresponding paths is in Figures 3, 4, 5.[1]

The main object of this paper is the analysis of the number of *deepest nodes*, i.e. the nodes defining the height of the tree.

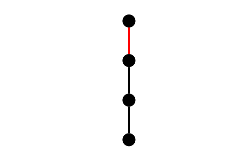
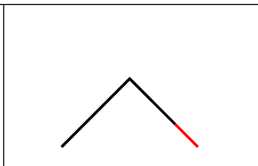
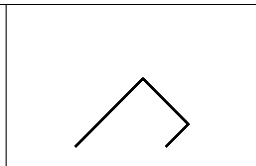
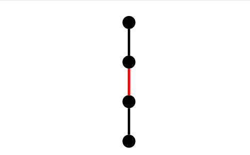
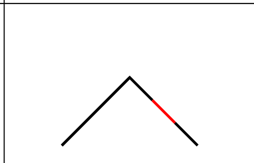
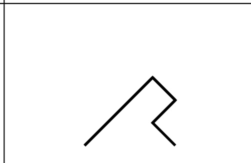
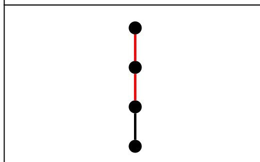
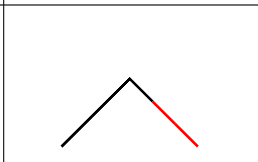
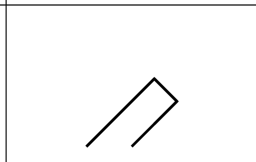
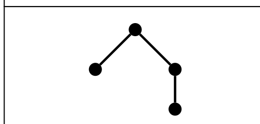
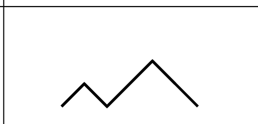
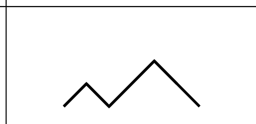
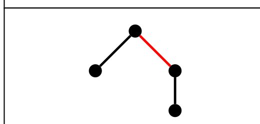
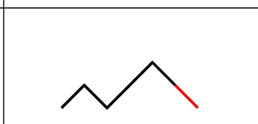
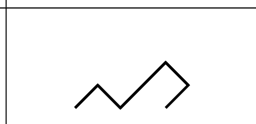
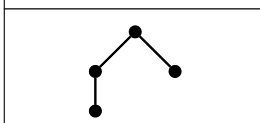
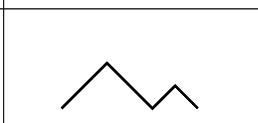
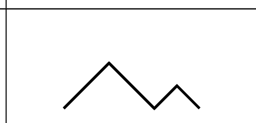
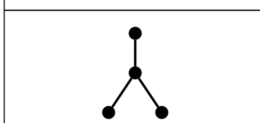
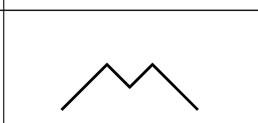
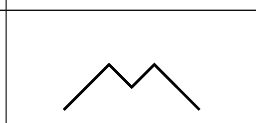
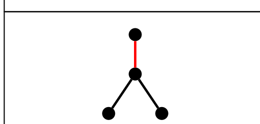
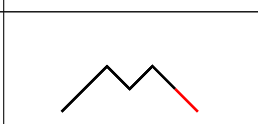
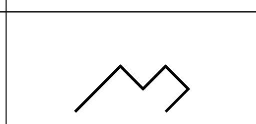
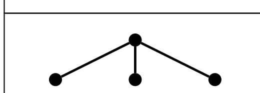
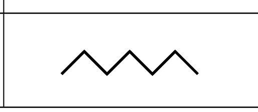
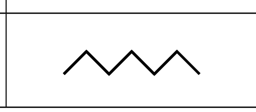
For ordered trees, this was investigated by Rainer Kemp [7], with important contributions provided by Volker Strehl [11].

A complete list of all 10 marked ordered trees with 4 nodes is provided for the benefit of the reader:[2]

			height 4, 1 node on bottom level

[1]Thanks are due to G. Feierabend for the drawings.

[2]G. Feierabend has compiled lists for all trees with up to 6 nodes https://www.math.tugraz.at/ prodinger/pdffiles/gregg.pdf.

			height 4, 1 node on bottom level
			height 4, 1 node on bottom level
			height 4, 1 node on bottom level
			height 3, 1 node on bottom level
			height 3, 1 node on bottom level
			height 3, 1 node on bottom level
			height 3, 2 nodes on bottom level
			height 3, 2 nodes on bottom level
			height 2, 3 nodes on bottom level

For completeness, it is mentioned that the average height of such marked ordered trees was already identified to be asymptotic to $\frac{2}{\sqrt{5}}\sqrt{\pi n}$ [8], which is slightly smaller than $\sqrt{\pi n}$ in the classical case [1].

2. Enumeration

We start by the enumerating the marked trees according to the number of nodes. Translating the symbolic equation,

$$A = \frac{z}{1-A} + \frac{z(A-z)}{1-A} = -z + \frac{z(2-z)}{1-A},$$

with the relevant solution

$$A(z) = \frac{1-z-\sqrt{1-6z+5z^2}}{2}$$
$$= z + z^2 + 3z^3 + 10z^4 + 36z^5 + 137z^6 + 543z^7 + 2219z^8 + \cdots,$$

and the sequence $1, 1, 3, 10, 36, \dots$ of coefficients is sequence A002212 in [10]. Next we enumerate the classes $\mathscr{A}_h$ according to the size. The treatment of deepest nodes will come a bit later. The enumerating sequence of $\mathscr{A}_h$ is defined to be $A_h = A_h(z) = \frac{f_h}{g_h}$. The recursion is

$$A_{h+1} = -z + \frac{z(2-z)}{1-A_h}, \quad A_1 = z.$$

We may set $f_1 = z$, $g_1 = 1$, $f_2 = z$, $g_2 = 1-z$. Then

$$f_{h+1} = zf_h + z(1-z)g_h, \quad g_{h+1} = g_h - f_h.$$

From this

$$g_{h+1} - g_{h+2} = zg_h - zg_{h+1} + z(1-z)g_h.$$

Solving the characteristic equation $X - X^2 = z - zX + z(1-z)$, we find the two roots

$$\lambda = \frac{1+z+\sqrt{1-6z+5z^2}}{2}, \quad \mu = \frac{1+z-\sqrt{1-6z+5z^2}}{2}.$$

The solution must be of the form

$$A_h = \frac{C_1\lambda^h - C_2\mu^h}{C_3\lambda^h - C_4\mu^h},$$

and an attractive form could be written using the substitution $z = \frac{v}{1+3v+v^2}$, since then $\frac{\lambda}{z} = 2 + v^{-1}$ and $\frac{\mu}{z} = 2 + v$. Then

$$A_h = z(1+v)\frac{(1+2v)^{h-1} - v^h(v+2)^{h-1}}{(1+2v)^{h-1} - v^{h+1}(v+2)^{h-1}},$$

which could be proved by induction as well.[3] It is also worthwhile to write

$$v = \frac{1-3z-\sqrt{1-6z+5z^2}}{2z}.$$

Note that $\frac{\mu}{z} = 2 + v$ and $\frac{\lambda}{z} = 2 + v^{-1}$. Further

$$\frac{\mu}{\lambda} = \frac{(1+z)}{(2-z)z}\mu - 1 = \frac{v(2+v)}{1+2v} =: q.$$

All these equivalent forms are useful somehow.

Now we count the deepest nodes, using a second variable t. We write $p_h = p_h(z,t)$, and $[z^n t^i]p_h(z,t)$ is the number of marked ordered trees with n nodes, height $\le h$, and i nodes on level h. For $i = 0$, this means that the tree has height $< h$, so $p_h(z,0) = p_{h-1}(z,1)$. The symbolic equation is used, but with a twist, since the recursion does not allow to compute p_2. We have

$$p_1 = zt, \quad p_2 = \frac{z}{1-p_1} = \frac{z}{1-zt}, \quad p_{h+1} = -z + \frac{z(2-z)}{1-p_h}, \text{ for } h \ge 2.$$

[3] G. Feierabend has worked out the details of such a proof https://www.math.tugraz.at/ prodinger/pdffiles/gregg.pdf.

Therefore

$$p_3 = -z + \frac{z(2-z)}{1-p_2} = -z + \frac{z(2-z)}{1 - \dfrac{z}{1-zt}},$$

$$p_4 = -z + \frac{z(2-z)}{1-p_3} = -z + \frac{z(2-z)}{1+z - \dfrac{z(2-z)}{1 - \dfrac{z}{1-zt}}},$$

$$p_5 = -z + \frac{z(2-z)}{1-p_4} = -z + \frac{z(2-z)}{1+z - \dfrac{z(2-z)}{1+z - \dfrac{z(2-z)}{1 - \dfrac{z}{1-zt}}}}, \quad \&c.$$

Expanding

$$\begin{aligned}
p_2 &= z + tz^2 + t^2z^3 + t^3z^4 + \cdots, \\
p_3 &= z + z^2 + (1+2t)z^3 + (1+3t+2t^2)z^4 + \cdots, \\
p_4 &= z + z^2 + 3z^3 + (6+4t)z^4 + \cdots, \\
p_5 &= z + z^2 + 3z^3 + 10z^4 + \cdots.
\end{aligned}$$

We look at the coefficient of z^4 and think about the list of 10 trees drawn earlier. For height ≤ 2, one such tree appears, and it has 3 deepest nodes. Next, 3 trees appear with one deepest node, and 2 with two deepest nodes. For height ≤ 4, four further trees appear, with one deepest node each.

With a lot of help from GFUN [9], we get for $h \ge 2$

$$\begin{aligned}
p_h &= z(1+v) \\
&\quad \times \frac{(v^2t - v^2 + vt - 3v - 1)(1+2v)^{h-2} - (-v^2 + vt - 3v + t - 1)v^{h-1}(2+v)^{h-2}}{(v^2t - v^2 + vt - 3v - 1)(1+2v)^{h-2} - (-v^2 + vt - 3v + t - 1)v^{h}(2+v)^{h-2}} \\
&= z(1+v)\frac{(v^2t - v^2 + vt - 3v - 1)\lambda^{h-2} - (-v^2 + vt - 3v + t - 1)v\mu^{h-2}}{(v^2t - v^2 + vt - 3v - 1)\lambda^{h-2} - (-v^2 + vt - 3v + t - 1)v^2\mu^{h-2}} \\
&= z(1+v)\frac{1 - Rvq^{h-2}}{1 - Rv^2q^{h-2}}
\end{aligned}$$

with

$$R = \frac{-v^2 + vt - 3v + t - 1}{v^2t - v^2 + vt - 3v - 1} = 1 + \frac{v-1}{v}\sum_{k\ge 1}(1+v)^k t^k z^k.$$

The representation

$$R = 1 + \frac{v-1}{v}\frac{t(1+v)z}{1 - t(1+v)z}$$

might be the most attractive. The reader can compare this for $t = 1$ with A_h given earlier.

The most interesting generating function is

$$G(z,t) := zt + \sum_{h\ge 2}\Big(p_h(z,t) - p_h(z,0)\Big);$$

the coefficient of $z^n t^i$ in $G(z,t)$ for $n, i \ge 1$ is the number of marked ordered trees with n nodes and i deepest nodes.

3. Continuing with exact analysis

First, note that $R|_{t=0} = \dfrac{-v^2-3v-1}{-v^2-3v-1} = 1$. Then

$$\begin{aligned}
\frac{p_h(z,t)-p_h(z,0)}{z(1+v)} &= \frac{1-Rvq^{h-2}}{1-Rv^2q^{h-2}} - \frac{1-vq^{h-2}}{1-v^2q^{h-2}}\\
&= \frac{(1-Rvq^{h-2})(1-v^2q^{h-2}) - (1-vq^{h-2})(1-Rv^2q^{h-2})}{(1-Rv^2q^{h-2})(1-v^2q^{h-2})}\\
&= \frac{(1-v)v(1-R)q^{h-2}}{(1-Rv^2q^{h-2})(1-v^2q^{h-2})}\\
&= (1-v)v\left[\frac{q^{h-2}}{1-v^2q^{h-2}} - \frac{Rq^{h-2}}{1-Rv^2q^{h-2}}\right],
\end{aligned}$$

or

$$p_h(z,t)-p_h(z,0) = \frac{z(1-v^2)}{v}\left[\frac{v^2q^{h-2}}{1-v^2q^{h-2}} - \frac{Rv^2q^{h-2}}{1-Rv^2q^{h-2}}\right].$$

Summing,

$$\begin{aligned}
\sum_{h\ge1}\Big(p_{h+1}(z,t)-p_{h+1}(z,0)\Big) &= \frac{z(1-v^2)}{v}\sum_{h\ge1}\left[\frac{v^2q^{h-1}}{1-v^2q^{h-1}} - \frac{Rv^2q^{h-1}}{1-Rv^2q^{h-1}}\right]\\
&= \frac{z(1-v^2)}{v}\sum_{h\ge1}\left[\frac{\delta q^h}{1-\delta q^h} - \frac{R\delta q^h}{1-R\delta q^h}\right]\\
&= \frac{z(1-v^2)}{v}\sum_{k\ge1}\left[\frac{\delta^k q^k}{1-q^k} - \frac{R^k\delta^k q^k}{1-q^k}\right]\\
&= \frac{z(1-v^2)}{v}\sum_{k\ge1}\frac{\delta^k q^k(1-R^k)}{1-q^k}
\end{aligned}$$

with

$$\delta = \frac{v(2v+1)}{v+2}.$$

Using the binomial theorem,

$$R^k - 1 = \sum_{i=1}^{k}\binom{k}{i}\left(\frac{v-1}{v}\,\frac{t(1+v)z}{1-t(1+v)z}\right)^i.$$

Putting things together,

$$\begin{aligned} G(z,t) - zt &= \sum_{h\ge1} \Big(p_{h+1}(z,t) - p_{h+1}(z,0)\Big) \\ &= \frac{z(v^2-1)}{v} \sum_{1\le i\le k} \binom{k}{i} \left(\frac{v-1}{v}\frac{t(1+v)z}{1-t(1+v)z}\right)^i \frac{\delta^k q^k}{1-q^k}. \end{aligned}$$

The generating function is now fully explicit.

Theorem 1. *The generating function $G(z,t)$ where the coefficient of $z^n t^i$ refers to the number of marked ordered trees with n nodes and i deepest nodes, has the explicit form*

$$G(z,t) = zt + \frac{z(v^2-1)}{v} \sum_{1\le i\le k} \binom{k}{i} \left(\frac{v-1}{v}\frac{t(1+v)z}{1-t(1+v)z}\right)^i \frac{\delta^k q^k}{1-q^k},$$

with $z = \dfrac{v}{1+3v+v^2}$, $q = \dfrac{v(v+2)}{2v+1}$, *and* $\delta = \dfrac{v(2v+1)}{v+2}$. □

Now we are interested in the average number of deepest nodes, assuming all trees of size n to be equally likely. For that, we have to differentiate $G(z,t)$ w.r.t. t, followed by $t=1$. We ignore the tree with one node and one deepest node. Only the quantity R contains the variable t:

$$\frac{d}{dt}(1-R^h)\Big|_{t=1} = \frac{(1-v^2)(1+3v+v^2)h(v+2)^{h-1}v^{h-1}}{(2v+1)^{h+1}} = \frac{(1-v^2)(1+3v+v^2)}{(v+2)v(2v+1)}hq^h.$$

Therefore

$$\begin{aligned} \frac{d}{dt}\sum_{h\ge1}\Big(p_{h+1}(z,t)-p_{h+1}(z,0)\Big)\Big|_{t=1} &= \frac{z(1-v^2)}{v}\sum_{k\ge1}\frac{\delta^k q^k \frac{(1-v^2)(1+3v+v^2)}{(v+2)v(2v+1)}kq^k}{1-q^k} \\ &= \frac{(1-v^2)^2}{(v+2)v(2v+1)}\sum_{k\ge1}\frac{k\delta^k q^{2k}}{1-q^k}. \end{aligned}$$

4. Asymptotics

For the following, we refer to [4] and use a hybrid approach, first the Mellin transform, to establish to local behaviour, and then singularity analysis to switch to the behaviour of the coefficients. The book [6] is of course also relevant here.

The goal is to find the behaviour of

$$\sum_{k\ge1}\frac{k\delta^k q^{2k}}{1-q^k} = \sum_{k\ge1}\frac{kv^{2k}q^k}{1-q^k} = \sum_{k,j\ge1} kv^{2k}q^{jk}$$

as $z\to\frac15$, or $v\to1$. First, we start with the simpler sum

$$\sum_{k\ge1}\frac{kq^{3k}}{1-q^k}$$

and discuss later that the difference is negligible. We set $q=e^{-w}$. Then we deal with

$$\sum_{k\ge1}\frac{ke^{-3kw}}{1-e^{-kw}} = \sum_{k\ge1}\sum_{j\ge3} ke^{-kjw}.$$

The Mellin transform of this is then $\Gamma(s)\zeta(s-1)\big(\zeta(s)-1-\frac{1}{2^s}\big)$, and the next step is to find the residues of $\Gamma(s)\zeta(s-1)(\zeta(s)-1-\frac{1}{2^s})w^{-s}$ left to the line $\Re s=\frac32$, say. We compute these residues at $s=1$ and $s=0$ (with a computer), with the (cumulative) result $-\frac{1}{2w}+\frac{5}{24}$. But $w=-\log(q)$, and we expand

$$\frac{(1-v^2)^2}{(v+2)v(2v+1)}\Big(\frac{1}{2\log(q)}+\frac{5}{24}\Big)$$

around $v=1$, with the result

$$-\frac13(1-v)-\frac{2}{27}(1-v)^2+\cdots.$$

One could from this translate to an expansion about $1-5z$, but it is not necessary, since the generating function of the marked ordered trees, established to be $z(1+v)$, which we need for normalization, is $\sim\frac25-\frac15(1-v)+\cdots$, and the quotient $(-\frac13)/(-\frac15)=\frac53$ is the average number of deepest nodes (leading term), when all trees of size n are considered to be equally likely and n gets large.

Theorem 2. *The average number of deepest nodes, when all marked ordered trees with n nodes are considered to be equally likely, approaches $\frac53$ as $n\to\infty$.* □

Nodes	Deepest nodes	Trees	Ratios
2	1	1	1.00000000000000
3	4	3	1.33333333333333
4	14	10	1.40000000000000
5	52	36	1.44444444444444
6	202	137	1.47445255474453
7	814	543	1.49907918968692
8	3367	2219	1.51735015772871
9	14224	9285	1.53193322563274
10	61122	39587	1.54399171445171
11	266336	171369	1.55416673960868
12	1174054	751236	1.56282978983968
13	5226196	3328218	1.57026853409242
14	23459020	14878455	1.57671075390556
15	106065578	67030785	1.58234127796653
16	482598675	304036170	1.58730678326858
17	2208111308	1387247580	1.59172114612736
18	10153335117	6363044315	1.59567254514713
19	46894469566	29323149825	1.59923029571739
20	217453338987	135700543190	1.60245002617664
21	1011990062528	630375241380	1.60537723580733
22	4725078802079	2938391049395	1.60804968523569
23	22127901099074	13739779184085	1.61049903368935
24	103910897639245	64430797069375	1.61275201247876
25	489188386162736	302934667061301	1.61483131299643
26	2308345828917289	1427763630578197	1.61675628898215
27	10915917653075084	6744284275226223	1.61854352628232
28	51723425586415104	31923955212096244	1.62020730961359
29	245539814027935212	151403298421257630	1.62176000515363

It remains to discuss that replacing v by q leads to the same main term. We have

$$v-q=\frac{v(v-1)}{1+2v},\quad v^2-q^2=\frac{3v^2(v-1)(1+v)}{(1+2v)^2},\quad v^3-q^3=\frac{v^3\,(v-1)\,(7v^2+13v+7)}{(1+2v)^3}$$

and so on; there is always a factor $v-1$ present, so the asymptotics of the difference of the two sums has an extra factor $v-1$, which means one order of magnitude smaller. This is perhaps easier to see when switching to the z-world: $1-v\sim\sqrt{5}\sqrt{1-5z}$, and for instance $(1-5z)^{3/2}$ leads already to coefficients that are smaller by a factor $\frac{1}{n}$. The necessary background information can be found in [5].

5. Conclusion

The continued fraction expression for $p_h=p_h(z,t)$ contains the variable t only at the bottom level. It would be desirable to have an equivalent representation where t appears only near the top of the continued fraction. This should then lead to an identity of the Kemp/Strehl type [7, 11]. Flajolet's paper [3] does not seem to be immediately applicable.

So far, we were not successful with this.

References

[1] N. G. De Bruijn, D. E. Knuth, and S. O. Rice. The average height of planted plane trees. In R. C. Read, editor, *Graph Theory and Computing*, pages 15–22. Academic Press, 1972.

[2] E. Deutsch, E. Munarini, and S. Rinaldi. Skew Dyck paths. *J. Stat. Plann. Infer.*, 140 (8) (2010) 2191–2203.

[3] P. Flajolet. Combinatorial aspects of continued fractions. *Discrete Math.*, 32 (1980) 125–161.

[4] P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: harmonic sums. Theoret. Comput. Sci. 144 (1995), 3–58.

[5] P. Flajolet and A. Odlyzko. Singularity Analysis of Generating Functions. SIAM J. Discrete Math., 3 216–240.

[6] P. Flajolet and R. Sedgewick. Analytic combinatorics. Cambridge University Press, Cambridge, 2009.

[7] R. Kemp. On the number of deepest nodes in ordered trees. *Discrete Math.*, 81 (1990) 247–258.

[8] H. Prodinger. A walk in my lattice path garden. preprint 2021. *https://arxiv.org/abs/2111.14797*.

[9] B. Salvy and P. Zimmermann, Gfun: a Maple package for the manipulation of generating and holonomic functions in one variable, ACM Transactions on Mathematical Software, vol. 20, no. 2, pp. 163–177, 1994.

[10] Neil J. A. Sloane and The OEIS Foundation Inc. The on-line encyclopedia of integer sequences, 2021.

[11] V. Strehl. Two short proofs of Kemp's identity for rooted plane trees. *European J. Combin.*, 5 (1984) 373–376.

Helmut Prodinger, Department of Mathematical Sciences, Stellenbosch University, 7602 Stellenbosch, South Africa, and NITheCS (National Institute for Theoretical and Computational Sciences), South Africa

Email address: `hproding@sun.ac.za`