INTRODUCTION TO PHILIPPE FLAJOLET'S WORK ON THE REGISTER FUNCTION AND RELATED TOPICS

HELMUT PRODINGER

Flajolet's research on the register function of binary trees as published in the preliminary [11] and the final article [12] was his first work in the analysis of algorithms. He was lucky to find this problem, as it is nontrivial but manageable, and one can learn/develop a lot while working on it.

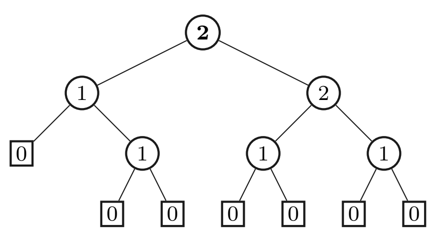
It deals with binary trees (counted by Catalan numbers) and the parameter reg, which associates to each binary tree (which is used to code an arithmetic expression, with data in the leaves and operators in the internal nodes) the minimal number of extra registers that is needed to evaluate the tree. The optimal strategy is to evaluate difficult subtrees first, and use one register to keep its value, which does not hurt, if the other subtree requires less registers. If both subtrees are equally difficult, then one more register is used, compared to the requirements of the subtrees.

There is a recursive description of this function: $\mathsf{reg}(\square) = 0$, and if tree t has subtrees t_1 and t_2, then

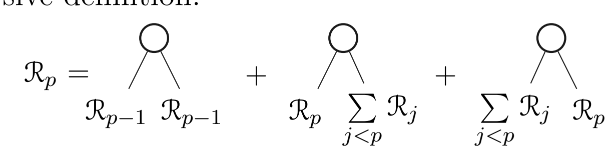
$$\mathsf{reg}(t) = \begin{cases} \max\{\mathsf{reg}(t_1), \mathsf{reg}(t_2)\} & \text{if } \mathsf{reg}(t_1) \neq \mathsf{reg}(t_2), \\ 1 + \mathsf{reg}(t_1) & \text{otherwise.} \end{cases}$$

The register function is also known as Horton-Strahler numbers in the study of the complexity of river networks. The original papers are [13, 21]; since then, many papers have been written about these numbers, but there is no space here to collect them all; a fair amount of them is cited in [16].

The recursive description attaches numbers to the nodes, starting with 0's at the leaves and then going up; the number appearing at the root is the register function of the tree.



Let $\mathcal{R}_p$ denote the family of trees with register function $= p$, then one gets immediately from the recursive definition:



Later in life, Flajolet would write such *symbolic equations* in a linearized form.

Date: April 16, 2012.

In terms of generating functions, these equations read as

$$R_p(z) = zR_{p-1}^2(z) + 2zR_p(z)\sum_{j<p} R_j(z);$$

the variable z is used to mark the size (i. e., the number of internal nodes) of the binary tree.

Amazingly, this can be solved explicitly! First, a trigonometric substitution was used, and this led to *Chebyshev polynomials*. However, eventually the substitution

$$z = \frac{u}{(1+u)^2}$$

that de Bruijn, Knuth, and Rice [2] also used, produced the nice expression

$$R_p(z) = \frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}}.$$

Of course, once this is *known*, it can be proved by induction, using the recursive formula.

Thanks to the trigonometric representation of $R_p(z)$, explicit forms for $[z^n]R_p(z)$, the number of binary trees of size n with register function $= p$, are available.

Reading off coefficients, the average number of registers requires to evaluate

$$\sum_{k\ge1} v_2(k)\left[\binom{2n}{n+1-k} - 2\binom{2n}{n-k} + \binom{2n}{n-1-k}\right] \tag{1}$$

with $v_2(k)$ being the number of trailing zeroes in the binary representation of k. The main result is: The average number of registers to evaluate a binary tree with n nodes is asymptotically given by

$$\log_4 n + D(\log_4 n) + o(1)$$

with

$$D(x) = \sum_{k\in\mathbb{Z}} d_k e^{2\pi ikx}$$

and

$$d_0 = \frac{1}{2} - \frac{\gamma}{2\log 2} - \frac{1}{\log 2} + \log_2 \pi,$$
$$d_k = \frac{1}{\log 2}\zeta(\chi_k)\Gamma(\tfrac{\chi_k}{2})(\chi_k - 1), \quad k \ne 0,$$

with $\chi_k = \frac{2\pi ik}{\log 2}$. The classical Gamma- and zeta-functions appear here. Because of the fast decay of the Gamma-function along the imaginary axis, the Fourier coefficients become small very quickly, and the periodic function $D(x)$ (oscillating around the value d_0) has small amplitude and is continuous; details are in the papers. This is the first periodic oscillation (given as a Fourier series) that appeared in Flajolet's work, with many more to follow.

To get this asymptotic equivalent, there are (at least) three methods, and Flajolet used (and developed) them all in his work.

- An elementary approach that uses Delange's [3] result about the sum-of-digits function. This is featured in [12].

- Approximation of the terms in the sum, and using the Mellin transform to evaluate the resulting series. This appears already in Flajolet's thesis [6]. He often mentioned that he learnt the technique from Rainer Kemp. A special case of the approach appears in Knuth [15] under the nickname Gamma-function method.[1,2]
- Mellin transform, directly on the level of generating functions, together with singularity analysis of generating functions.

Let us describe these methods in more detail. We start with the elementary approach. Summation leads to

$$\sum_{j\le k} v_2(j) = k - S_2(k),$$

where $S_2(k)$ is the number of ones in the binary expansion of k (sum of digits). It is known that

$$\sum_{m<n} S_2(m) = \frac{n\log_2 n}{2} + nF(\log_2 n).$$

This was shown by Delange [3] and was apparently mentioned to Flajolet directly by Delange. The periodic function $F(t)$ is fully explicit in terms of Fourier coefficients.

To use this for the evaluation of (1), one has to use partial summation (Abel's summation) twice. A negative side effect of this is that the second difference of binomial coefficients becomes a fourth difference. But this is no problem: approximations are available (Hermite polynomials):

$$\Delta^r \frac{\binom{2n}{n+k}}{\binom{2n}{n}} \sim e^{-k^2/n} H_r\Big(\frac{k}{\sqrt{n}}\Big)\frac{1}{n^{r/2}}.$$

It seems that Flajolet found this approximation by comparing the sum $\sum_{k\ge1} e^{-k^2/n}$ and its differences with the corresponding integral and its derivatives. This approximation gives the main term, and this is enough for the research on the register function. In [19], it is described how lower order terms can be produced if needed. Doing all this, one is left with the asymptotic evaluation (for $n\to\infty$) of

$$\sum_{k\ge1}\Big[\frac{k\log_2 k}{2} + kF(\log_2 k)\Big] H_4\Big(\frac{k}{\sqrt{n}}\Big)e^{-k^2/n}.$$

The asymptotic evaluation of this is manageable (Riemann sums, controlling the error), but it is a bit dry. Nevertheless, it is completely elementary!

Remark. To solve explicitly for R_p is somewhat crucial. Auber et al. [1] suggested a generalisation of the register function to t-ary trees, where no explicit formula is available, and Drmota and Prodinger [4] could only identify the leading $\log_4 n$ term! This generalisation interprets an internal node of outdegree t as a t-ary operation, and the recursive rule of the register function is accordingly defined as $\mathsf{reg}(T) = \max\{c_1, c_2+$

[1]Rainer Kemp solved the asymptotics of the register function independently in [14]. This important pioneer of the analysis of algorithms in Europe died in 2004, barely 55 years old. Among the three people who started the series of conferences on the analysis of algorithms in Dagstuhl, Germany, there is thus only the writer of these lines still around.

[2]Some time later, another paper appeared [17], which only had the main term $\log_4 n$ but no oscillations; the authors were unaware of the papers [12, 14].

$1,\dots,c_t+t-1\}$, where $c_1\ge c_2\ge\dots\ge c_t$ is the ordered list of the register functions of the t subtrees. The leading term $\log_4 n$ is independent of t. —

Delange's paper was extended and generalized into many different directions by many people.

Now we describe the second method: After Flajolet learnt about the Mellin transform, he attacked a sum like

$$\sum_{k\ge1} v_2(k)H_2(kt)e^{-k^2t^2}$$

(with $t=1/\sqrt{n}$) directly. This goes well, since $v_2(2k+1)=0$ and $v_2(2k)=1+v_2(k)$ and

$$\sum_{k\ge1}\frac{v_2(k)}{k^s}=\frac{\zeta(s)}{2^s-1}.$$

The sum is what he would later call a *harmonic sum*, and its Mellin transform splits. One factor is the Dirichlet series just mentioned, the other one a simple variation of the Gamma function.

And now to the third method: Flajolet became familiar with *singularity analysis of generating functions* through Odlyzko's pioneering paper [18] (written in 1978). Around this time, a fruitful collaboration between Flajolet and Odlyzko has started which culminated in the highly cited paper [8]. Before the name *singularity analysis* became really popular, Flajolet himself liked to say "à la Odlyzko", as is apparent in [9] and other writings. After that, he would consider

$$E(z)=\sum_{p\ge1}pR_p(z)=\sum_{p\ge1}p\frac{1-u^2}{u}\frac{u^{2^p}}{1-u^{2^{p+1}}}$$

and study it around the singularity $z=\frac14$ (equivalently $u=1$) with the Mellin transform!

But now the Mellin transform is used directly on the generating function, where a substitution $u=e^{-t}$ is used. In terms of singularity analysis, one must study the behaviour around $z=\frac14$ or equivalently $u\sim1$ or $t\sim0$. (The notation '$\sim$' refers to a "camembert-shaped" neighbourhood; the technical details can be taken from [8].) The essential part is then

$$\sum_{p\ge1}p\frac{e^{-t2^p}}{1-e^{-t2^{p+1}}}=\sum_{p\ge1,\,\lambda\ge0}pe^{-t2^p(1+2\lambda)}=\sum_{n\ge1}v_2(n)e^{-tn}.$$

This is a harmonic sum!

A local expansion around $t\sim0$ is thus found. It translates; since

$$\sqrt{1-4z}\sim 2t$$

one finds with

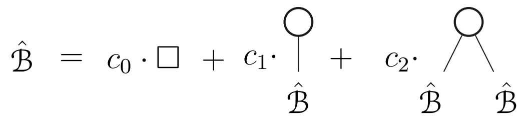
$$K=-1+\log_2\pi+\frac12+\frac{\gamma}{\log2},\quad r=\sqrt{1-4z}\quad\text{and}\quad c_k=\frac{1}{\log2}\zeta(\chi_k)\Gamma(\chi_k),$$

that

$$E(z)=2r\log_2 r+2(K+1)r+4\sum_{k\ne0}c_kr^{1-\chi_k}+\cdots.$$

This can then be translated into an asymptotic expansion of the coefficients.

This approach is exactly the one used in [9] where the register function is studied in the context of unary-binary trees. Such trees can be used to model not only binary operations but unary operations as well. For obvious reasons, the register function does not increase when going up to a unary node. The symbolic equation for the objects under consideration is now this (arbitrary non-negative weights have been introduced):



This family can be obtained from the family $\mathcal{B}$ of binary trees by means of substitutions:

Let $B(z)$ denote the generating function of binary trees (Catalan numbers). Then the generating function $yB(yz)$ marks internal nodes with the variable y and leaves with the variable z. The two substitutions mean on the level of generating functions

$$y \to \frac{c_0 y}{1 - c_1 z} \quad \text{and} \quad z \to \frac{c_2 z}{1 - c_1 z}.$$

Since the register function is not changed by performing these substitutions, the new function $\hat{R}_p$ is obtained from the already known R_p, thanks to these substitutions. For the *size* of unary-binary trees there are now two meaningful options: One can count both, leaves and internal nodes, or, count only internal nodes.

Eventually one gets for the average register function of unary-binary trees of size n a formula

$$\log_4 n + D(\log_4 n - \mathsf{constant}) - \mathsf{constant},$$

where the constant and periodic function is described in the paper.

The algorithm *odd-even merge* was analyzed by R. Sedgewick in [20] by means of the Mellin transform (the second method as discussed above). Flajolet wanted to show that the asymptotic expansion can also be obtained in an elementary fashion (akin to the first method from above). He submitted indeed “A note on Gray code and odd-even merge”; this original submission might still exist in his drawers. Published is however a much longer paper with the additional author L. Ramshaw [10].

To achieve his goal, Delange’s analysis had to be carried over first to the instance of the Gray code, which is a representation of integers with digits 0 and 1, where the pattern of the last digit is: 0110 0110 0110 ..., of the penultimate digit: 00111100 00111100 00111100 ..., and so on. Hence there is an explicit formula for the k-th digit of the Gray code representation of the integer n:

$$a_k(n) = \left\lfloor \frac{n}{2^{k+2}} + \frac{3}{4} \right\rfloor - \left\lfloor \frac{n}{2^{k+2}} + \frac{1}{4} \right\rfloor.$$

A Delange type approach works indeed, since

$$\left\lfloor \frac{n}{2^{k+2}} + \frac{3}{4} \right\rfloor = \int_n^{n+1} \left\lfloor \frac{t}{2^{k+2}} + \frac{3}{4} \right\rfloor dt.$$

The quantity that Sedgewick studied in his paper on odd-even merge is

$$1+(n+1)\sum_{i\geq 1}\beta(i)\frac{\binom{2n}{n+i+2}-3\binom{2n}{n+i+1}+3\binom{2n}{n+i}-\binom{2n}{n+i-1}}{\binom{2n}{n}},$$

with $\beta(i)=\sum_k a_k(i)$ being the number of ones in the Gray code representation of i.

One application of partial summation to this brings indeed the summatory function of $\beta(i)$ into the game, for which there is now an *explicit expression* available, thanks to the Delange approach. The rest is as in the instance of the register function.

The third approach is also feasible in the present instance; one would have to perform partial summation in the other direction, which leads to the sequence $\beta(i)-\beta(i-1)$; this sequence has a simple explicit Dirichlet generating function, which is expressible in terms of the Hurwitz zeta function.

In a paper with coauthors Grabner, Kirschenhofer, Prodinger, Tichy [7], he used the *Mellin-Perron* technique to deal with digital sums, but that is described in a different place in this volume.

The paper [5] does not contain new material; it describes the elementary approach to the register function and the odd-even merge (in French).

References

[1] D. Auber, J.-P. Domenger, M. Delest, P. Duchon, and J.-M. Fédou. New Strahler numbers for rooted plane trees. In *Mathematics and Computer Science III*, pages 203–215. Birkhäuser, Basel, 2004.

[2] 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.

[3] H. Delange. Sur la fonction sommatoire de la fonction "somme des chiffres". *Enseignement Mathématique*, 21:31–47, 1975.

[4] M. Drmota and H. Prodinger. The register function for t-ary trees. *ACM Trans. Algorithms*, 2(3):318–334, 2006.

[5] P. Flajolet. Deux problèmes d'analyse d'algorithmes. *Séminaire Delange-Pisot-Poitou. Théorie des nombres*, 20:1–10, 1978–79.

[6] P. Flajolet. Analyse d'algorithmes de manipulation d'arbres et de fichiers. *Cahiers du Bureau Universitaire de Recherche Opérationnelle*, 34/35:1–209, 1981.

[7] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R.F. Tichy. Mellin transforms and asymptotics: Digital sums. *Theoretical Computer Science*, 123:291–314, 1994.

[8] P. Flajolet and A. Odlyzko. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, 3:216–240, 1990.

[9] P. Flajolet and H. Prodinger. Register allocation for unary-binary trees. *SIAM J. Comput.*, 15:629–640, 1986.

[10] P. Flajolet and L. Ramshaw. A note on Gray code and odd-even merge. *SIAM J. Comput.*, 9(1):142–158, 1980.

[11] P. Flajolet, J.-C. Raoult, and J. Vuillemin. On the number of registers required for evaluating arithmetic expressions. *Proceedings of the 18th Annual Symposium on Foundations of Computer Science*, pages 196–205, 1977.

[12] P. Flajolet, J.-C. Raoult, and J. Vuillemin. The number of registers required for evaluating arithmetic expressions. *Theoretical Computer Science*, 9(1):99–125, 1979.

[13] R. E. Horton. Erosioned development of systems and their drainage basins. *Bulletin Geological Society of America*, 56:275–370, 1945.

[14] R. Kemp. The average number of registers needed to evaluate a binary tree optimally. *Acta Inform.*, 11(4):363–372, 1978/79.

[15] D. E. Knuth. *The Art of Computer Programming*, volume 3: Sorting and Searching. Addison-Wesley, 1973.

[16] G. Louchard and H. Prodinger. The register function for lattice paths. *Discrete Mathematics and Theoretical Computer Science*, proc. AG:139–152, 2008.

[17] A. Meir, J. W. Moon, and J. R. Pounder. On the order of random channel networks. *SIAM J. Algebraic Discrete Methods*, 1(1):25–33, 1980.

[18] A. Odlyzko. Periodic oscillations of coefficients of power series that satisfy functional equations. *Advances in Math.*, 44:180–205, 1982.

[19] H. Prodinger. Quelques techniques analytiques pour l'étude asymptotique des paramètres dans les arbres. In *Summer School Ile de Ré, English version in EATCS Bulletin 47 (1992), 180–199*, 1983.

[20] R. Sedgewick. Data movement in odd-even merging. *SIAM J. Comput.*, 7(3):239–272, 1978.

[21] A. N. Strahler. Hypsomic analysis of erosional topography. *Bulletin Geological Society of America*, 63:1117–1142, 1952.

Helmut Prodinger, Mathematics Department, Stellenbosch University, 7602 Stellenbosch, South Africa.

E-mail address: hproding@sun.ac.za