Average Case Analysis of Tree Rewriting Systems

Cyril Chabaud
LITP

June 12, 1995

[summary by Xavier Gourdon]

Abstract

A general technique is presented to easily compute the order of the average complexity of a tree rewriting system from its matrix representation. It can be used for example to prove that the average cost of the k-th differentiation is of order $n^{1+k/2}$.

1. Introduction

We aim at studying the order of the average complexity of regular tree rewriting systems. We deal with simple families of trees $\mathcal{T}$, in the sense of Meir and Moon [4]. The corresponding generating function (GF) is defined by $T(z) = z\phi(T(z))$ where $\phi(y)$ is a polynomial whose n-th coefficient is the number of constructors of arity n. For example, the GF of binary trees is defined by $T(z) = z(1 + T^2(z)) = z\phi(T(z))$ with $\phi(y) = 1 + y^2$.

Asymptotics of $T(z)$. We define $\tau > 0$ as the solution of $\tau\phi'(\tau) - \phi(\tau) = 0$ and we denote $\rho = \tau/\phi(\tau) = 1/\phi'(\tau)$. Then ρ is the dominant singularity of $T(z)$ with the Puiseux expansion

$$T(z) = \tau - \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}\left(1 - \frac{z}{\rho}\right)^{1/2} + \sum_{n\geq 2} d_n \left(1 - \frac{z}{\rho}\right)^{n/2}.$$

From singularity analysis, we deduce the estimate of the n-th coefficient of $T(z)$

$$[z^n]T(z) = \sqrt{\frac{\phi(\tau)}{2\pi\phi''(\tau)}}\rho^{-n} n^{-3/2}\left(1 + O\left(\frac{1}{n}\right)\right).$$

An example: differentiation of trees. A typical example of a tree rewriting system is formal differentiation. We describe the action of the differentiation and copy operators on trees constructed with a binary constructor $*$ and a variable a

$$\begin{aligned} \mathrm{d}(a) &\to a & \mathrm{cp}(a) &\to a \\ \mathrm{d}(u * v) &\to \mathrm{d}(u) * \mathrm{cp}(v) + \mathrm{d}(v) * \mathrm{cp}(u) & \mathrm{cp}(u * v) &\to \mathrm{cp}(u) * \mathrm{cp}(v) \end{aligned}$$

If $B(z)$ denotes the GF of binary trees, this translates in terms of cost generating functions in the form

$$\begin{aligned} C_{\mathrm{d}}(z) &= B(z) + 2zB(z)C_{\mathrm{d}}(z) + 2zB(z)C_{\mathrm{cp}}(z), \\ C_{\mathrm{cp}}(z) &= B(z) + 2zB(z)C_{\mathrm{cp}}(z). \end{aligned}$$

Equivalently, we have the matrix representation

$$\begin{pmatrix} C_{\mathrm{d}}(z) \\ C_{\mathrm{cp}}(z) \end{pmatrix} = \begin{pmatrix} 2zB(z) & 2zB(z) \\ 0 & 2zB(z) \end{pmatrix} \begin{pmatrix} cC_{\mathrm{d}}(z) \\ C_{\mathrm{cp}}(z) \end{pmatrix} + \begin{pmatrix} B(z) \\ B(z) \end{pmatrix}. \tag{1}$$

The solution is

$$C_{\mathrm{d}}(z) = \frac{B(z)}{(1-2zB(z))^2} = \frac{B(z)(1+B^2(z))}{(1-B(z))^2(1+B(z))^2}, \quad C_{\mathrm{cp}}(z) = \frac{B(z)(1+B^2(z))}{(1-B(z))(1+B(z))}.$$

Since $B(z) = z\phi(B(z))$ with $\phi(w) = 1 + w^2$, τ and ρ are easily determined: $\tau = 1$, $\rho = 1/2$.

Average complexity. The cost GF $C_{\mathrm{cp}}(z)$ writes as $F(B(z))$, where $F(w)$ is a rational function. The dominant pole of $F(w)$ is $w = \tau$ and it is simple. An application of transfer lemma of singularity analysis [2] then leads to the estimate $\overline{C_n^{\mathrm{cp}}} \sim c_1 n$ with $c_1 > 0$ for the cost of the copy operator over trees of size n. As for the cost GF $C_{\mathrm{d}}(z)$, it writes as a rational functional in $B(z)$ with the double dominant pole τ, and we deduce an average asymptotic value of the form $\overline{C_n^{\mathrm{d}}} \sim c_2 n^{3/2}$, $c_2 > 0$.

2. Regular rewriting systems

The matrix representation (1) for the cost GF's can be generalised for all regular rewriting systems [1].

THEOREM 1 (MATRIX REPRESENTATION FOR REGULAR REWRITING SYSTEMS). *The cost GF's of operators $f_1, \ldots, f_n$ of a regular system satisfy a system of the form*

$$\begin{pmatrix} C_{f_1}(z) \\ \vdots \\ C_{f_n}(z) \end{pmatrix} = M(z, T(z)) \begin{pmatrix} C_{f_1}(z) \\ \vdots \\ C_{f_n}(z) \end{pmatrix} + \begin{pmatrix} T^{r_1}(z) \\ \vdots \\ T^{r_n}(z) \end{pmatrix}, \tag{2}$$

where the r_i are the arities of the f_i's, and where the coefficient of the square matrix $M(z, T(z))$ are polynomials in z and $T(z)$ with non negative coefficients.

Thus, the expression of each cost GF is

$$C_{f_i}(z) = \frac{\det^{[i]}(\mathrm{Id} - M(z, T(z)))}{\det(\mathrm{Id} - M(z, T(z)))}, \tag{3}$$

where ${}^{[i]}A$ denotes the matrix in which the i-th column of A has been substituted by the rightmost vector of equation (2). We deduce, since $z = T(z)/\phi(T(z))$, that $C_{f_i}(z)$ writes as

$$C_{f_i}(z) = \frac{P_i(T(z))}{Q_i(T(z))},$$

where $P_i(w)$ and $Q_i(w)$ are polynomials. The average complexity of the operator f_i, defined by

$$\overline{C_n^{f_i}} = \frac{[z^n]C_{f_i}(z)}{[z^n]T^{r_i}(z)},$$

is determined by the relative position of ρ with respect to the smallest positive solution $\rho_{0,i}$ of $Q_i(T(\rho_{0,i})) = 0$ (see [1]).

THEOREM 2 (AVERAGE COST ESTIMATE). *The average cost satisfies*

(i) *If $Q_i(T(z))$ does not vanish on $(0, \rho]$, then $\overline{C_n^{f_i}} = c_1(1 + O(1/n))$;*
(ii) *if $\rho = \rho_{0,i}$, then $\overline{C_n^{f_i}} = c_2 n^{k/2}(1 + O(1/\sqrt{n}))$;*

(iii) *if* $\rho > \rho_{0,i}$, *then* $\overline{C_n^{f_i}} = c_3(\rho/\rho_{0,i})^n n^{q+1/2}(1 + O(1/\sqrt{n}))$,
with $c_j > 0$ *and* k, q *positive integers.*

In the case $\rho = \rho_{0,i}$, we have $k = s + 1$ where s is the multiplicity of the factor $(T(z) - \tau)$ in $Q_i(T(z))$.

3. Computation of the order of the average cost

It is possible to derive directly from the matrix $M(z, T(z))$ the order of the average cost of the operators of a regular system. The substance relies on Frobenius theory of matrices with nonnegative coefficients (see for instance [5]). The general technique proceeds as follows. First, decompose $M(z, T(z))$ into diagonal blocks of irreducible matrices (Definition 1), then work on each irreducible block.

3.1. Irreducible matrix case.

Definition 1. A square matrix M is *irreducible* if there does not exist any permutation matrix P such that $P^{-1}MP = \left(\begin{smallmatrix} A & 0 \\ B & C \end{smallmatrix}\right)$ with A, B and C square matrices.

In other terms, an irreducible matrix is associated to a strongly connected graph. If the matrix $M(z, T(z))$ is irreducible, the order of the average complexity of the operators are easily found.

Theorem 3. *Let* $\{f_i\}$ *be a set of operators of a regular rewriting system represented by an irreducible matrix* $M(z, T(z))$. *Then all the* $\rho_{0,i}$ *are equal to the smallest positive root* ρ_0 *of the equation* $\det(Id - M(z, T(z))) = 0$ *(take* $\rho_0 = +\infty$ *if there is no positive solution). The relative position of* ρ_0 *with respect to* ρ *is determined from the dominant eigenvalue* $r(\rho, \tau)$ *of* $M(\rho, \tau)$. *We have*

$$r(\rho,\tau) < 1 \text{ iff } \rho_0 > \rho, \qquad r(\rho,\tau) = 1 \text{ iff } \rho_0 = \rho, \qquad r(\rho,\tau) > 1 \text{ iff } \rho_0 < \rho.$$

When $r(\rho, \tau) = 1$, or equivalently $\rho_0 = \rho$, it is possible to get the exponent of n in the estimate (ii) of Theorem 1. This is the polynomial case.

Theorem 4. *Let* $\{f_i\}$ *be a set of operators of a regular rewriting system represented by an irreducible matrix* $M(z, T(z))$. *If the dominant eigenvalue of* $M(\rho, \tau)$ *is* 1, *then the* f_i*'s have an average complexity which is linear or of order* $n^{3/2}$.

The case $n^{3/2}$ occurs only in the degenerate case where $M(z, T(z))$ does not depend on $T(z)$.

3.2. General case. In the general case, we start by finding a permutation matrix P such that $P^{-1}MP$ writes as a block diagonal matrix, each block being of the form

$$B = \begin{pmatrix} A_1 & & 0 \\ \ddots & \ddots & \\ \ddots & \ddots & A_k \end{pmatrix},$$

where each $A_i = A_i(z, T(z))$ is an irreducible square block. We also need the constraint that for all $i < j$, the submatrix of B whose lines are those of A_j and columns are those of A_i is not zero. Considering the graph represented by the matrix M, this task can be achieved thanks to Tarjan algorithm on strongly connected components (see [3, pp. 441–448] for example).

Now, each block of the form B can be considered independently. Let $C_{f_j}(z)$ be a cost GF associated to an irreducible square block A_ℓ. Expression (3) together with Theorem 3 show that the position of $\rho_{0,j}$ is the smallest positive root of $\prod_{i=1}^{\ell} \det(\mathrm{Id} - A_i)$. Thus, if ρ_i denotes the smallest positive root of $\det(\mathrm{Id} - A_i)$, for each ℓ, we need to compare ρ_ℓ with $\min_{1 \le i \le \ell-1} \rho_i$ in order to get the order of the average complexity of the operators f_j. In fact, Theorem 3 asserts that this task can be achieved by comparing only the dominant eigenvalues $r_i(\rho,\tau)$ of the $A_i(\rho,\tau)$'s.

In the polynomial case, the multiplicity of the factor $(T(z)-\tau)$ in the denominator of the cost GF is obtained by adding the multiplicities of this factor in the determinants $\det(\mathrm{Id}-A_i)$, yielding the exponent of n in equation (ii) of theorem 2.

3.3. Examples.

Tree shuffle. We consider binary trees $\mathcal{B}=a+o(\mathcal{B},\mathcal{B})$ and operators f and g defined on $\mathcal{B}^2$ by

$$\begin{array}{lll} f(a,a)\to a & f(o(u,v),a)\to o(u,v) & f(a,o(u,v))\to g(u,v) \\ g(a,a)\to a & g(o(u,v),a)\to g(u,v) & g(a,o(u,v))\to g(u,v) \end{array}$$

$$\begin{array}{l} f(o(u1,v1),o(u2,v2))\to o(f(u1,u2),f(v1,v2)) \\ g(o(u1,v1),o(u2,v2))\to o(f(u1,u2),g(v1,v2)) \end{array}$$

The matrix representation of the shuffle is

$$M(z,T(z))=\begin{pmatrix} 2z^2T^2(z) & 2z^2 \\ z^2T^2(z) & z^2T^2(z)+2z^2 \end{pmatrix}.$$

This matrix is irreducible. The eigenvalues of $M(\rho,\tau)$ are 1 and $1/4$, thus we are in the polynomial case with a linear average complexity of the operators f and g.

Formal differentiation. The classical formal double differentiation on unary-binary trees $\mathcal{T}$ with constructors `*`, `exp` and a variable has the matrix representation

$$M(z,T(z))=\begin{pmatrix} z+2zT(z) & 0 & 0 \\ 2zT(z)+z & z+2zT(z) & 0 \\ 2zT(z)+2z & 4zT(z)+2z & z+2zT(z) \end{pmatrix}.$$

The diagonal coefficients of $M(\rho,\tau)$ are only 1's, thus we are in the polynomial case with three blocks of irreducible matrices on the diagonal, all giving a contribution to the order of the complexity. Thus, the average complexity of the double differentiation operator is cn^2. By induction on k, it can be proved that the average cost of the k-th differentiation is of order $n^{k/2+1}$.

Bibliography

[1] Choppy (Christine), Kaplan (Stéphane), and Soria (Michèle). – Complexity analysis of term rewriting systems. *Theoretical Computer Science*, vol. 67, 1989, pp. 261–282.

[2] Flajolet (Philippe) and Odlyzko (Andrew M.). – Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics*, vol. 3, n° 2, 1990, pp. 216–240.

[3] Froidevaux (Christine), Gaudel (Marie-Claude), and Soria (Michèle). – *Types de données et algorithmes.* – McGraw–Hill, Paris, 1990.

[4] Meir (A.) and Moon (J. W.). – On the altitude of nodes in random trees. *Canadian Journal of Mathematics*, vol. 30, 1978, pp. 997–1015.

[5] Minc (Henryk). – *Nonnegative matrices.* – J. Wiley and sons, New York, 1988, *Wiley interscience series in discrete mathematics and optimization.*