Multivariate Lagrange inversion formula and the cycle lemma[★]

Axel Bacher and Gilles Schaeffer

Université Paris Nord, and CNRS / École Polytechnique

Abstract. We give a multitype extension of the cycle lemma of (Dvoretzky and Motzkin 1947). This allows us to obtain a combinatorial proof of the multivariate Lagrange inversion formula that generalizes the celebrated proof of (Raney 1963) in the univariate case, and its extension in (Chottin 1981) to the two variable case.

Until now, only the alternative approach of (Joyal 1981) and (Labelle 1981) via labelled arborescences and endofunctions had been successfully extended to the multivariate case in (Gessel 1983), (Goulden and Kulkarni 1996), (Bousquet et al. 2003), and the extension of the cycle lemma to more than 2 variables was elusive.

The cycle lemma has found a lot of applications in combinatorics, so we expect our multivariate extension to be quite fruitful: as a first application we mention economical linear time exact random sampling for multispecies trees.

1 Introduction

For any power series $g(x)$ with $g(0) \neq 0$, there exists a unique power series $f(t)$ solution of the equation $f = tg(f)$. The Lagrange inversion formula says that the nth coefficient of $f(t)$ is $\frac{1}{n}[x^{n-1}]g(x)^n$. This formula is now known as a fundamental tool to derive tree enumeration results. The two simpler and most classical examples are:

- if $g(x) = \exp(x)$ then $f(t)$ is the exponential generating function of Cayley trees, so that the number of rooted Cayley trees with n nodes is equal to $\frac{1}{n}\left[\frac{x^{n-1}}{(n-1)!}\right]\exp(nx) = n^{n-2}$;
- if $g(x) = (1+x)^2$ then $f(t)$ is the ordinary generating function of binary trees, so that the number of rooted binary trees with n nodes is equal to $\frac{1}{n}[x^{n-1}](1+x)^{2n} = \frac{1}{n}\binom{2n}{n-1} = \frac{1}{n+1}\binom{2n}{n}$.

These examples are in a sense generic: the bijection between doubly rooted Cayley trees and endofunctions underlies Labelle's proof of the Lagrange inversion formula [11], while the cyclic lemma used in [7] to count ballot numbers underlies Raney's proof [16].

Our interest is in the multivariate extension of the Lagrange inversion formula. Fix an integer $k \geq 1$. Let bold letters denote k-dimensional vectors; write $\mathbf{x^n} = x_1^{n_1}\cdots x_k^{n_k}$ and $\mathbf{x^{n-1}} = x_1^{n_1-1}\cdots x_k^{n_k-1}$. Let $h(\mathbf{x})$ and $g_1(\mathbf{x}),\ldots,g_k(\mathbf{x})$ be power series in $\mathbf{x}$ such that for $i = 1,\ldots,k$, $g_i(\mathbf{0}) \neq 0$. Again there is a unique family of power series $\mathbf{f(t)}$ solution of the system of equations $f_i = t_i g_i(f_i)$ for $i = 1,\ldots,k$

---

and the multivariate Lagrange inversion formula admits the two equivalent following formulations (among several others):

$$[\mathbf{t^n}]h(\mathbf{f(t)}) = [\mathbf{x^{n-1}}]h(\mathbf{x})\mathbf{g(x)^n} \det\left(\delta_{i,j} - \frac{x_i}{g_j(\mathbf{x})}\frac{\partial g_j(\mathbf{x})}{\partial x_i}\right) \qquad (1)$$

$$= \frac{1}{n_1 \cdots n_k}[\mathbf{x^{n-1}}]\sum_T \partial_T(h, g_1^{n_1}, \ldots, g_k^{n_k}) \qquad (2)$$

where the sum is over oriented 0-rooted Cayley trees (non-plane trees with arcs going toward 0) with vertices $\{0, \ldots, k\}$ and the derivative ∂_G with respect to a directed graph G with vertex set $V = \{0, \ldots, k\}$ and edge set E is defined as

$$\partial_G(f_0(\mathbf{x}), \ldots, f_k(\mathbf{x})) = \prod_{j=0}^{k}\left(\left(\prod_{(i,j)\in E}\frac{\partial}{\partial x_i}\right)f_j(\mathbf{x})\right)$$

Several variants of (1) are given in [8] but (2) appeared more recently, implicitely in [9] and explicitely in [1]. As far as we know all combinatorial proofs of the multivariate Lagrange inversion extend Joyal and Labelle's approach [10, 11]: [8] proves another variant of (1), [9, 3] prove (2). A completely different approach was recently proposed in [2].

Chottin [4, 5] instead proposed a remarkable extension of Raney's strategy to prove the two variable Lagrange inversion formula. Yet he failed to move to three variables and the problem of proving the multivariate Lagrange inversion formula with the cycle lemma was considered as difficult. Apart from the theoretical interest of such a proof, an extension of the cycle lemma is desirable in view of its numerous applications, to tree and map enumeration [12, Chapter 11],[13, Chapter 9], probability [14], and random sampling [6]. We present such an extension in this paper.

2   Generalized Cycle Lemma

Following [5] and the modern accounts of Raney's proof, our combinatorial construction is in terms of encodings of rooted plane trees by sequences of nodes. To deal with the multivariate case, we introduce colored trees.

A *colored tree* is a plane, rooted tree in which all edges have a color in the set $\{1, \ldots, k\}$. A *colored bush* is a colored tree that can have "pending" edges with no node attached. We call such edges *free edges*; an edge is *occupied* if a node is attached to it.

Let ℓ be an integer with $0 \leq \ell \leq k$. An *ℓ-bush* is a colored bush such that occupied edges have color more than ℓ and free edges have color at most ℓ. In particular, a k-bush only has a single node, while a 0-bush is a colored tree. Therefore, ℓ-bushes can be seen as intermediate objects between colored nodes and colored trees.

Definition 1. Let $0 \leq \ell \leq k$. We denote by $\mathscr{S}_\ell$ the set of tuples of the form $S = (S_0, \ldots, S_\ell, e_1, \ldots, e_k)$, where $S_0, \ldots, S_\ell$ are sequences of ℓ-bushes and e_i is

an edge of color i in S (refered to as a *marked edge*), satisfying the following conditions.

1. The sequence S_0 has only one element; for $i = 1, \ldots, \ell$, the number of elements of the sequence S_i is equal to the number of edges of color i in S.
2. Let T be the graph with vertices $0, \ldots, \ell$ and an arc $i \rightarrow j$ if the edge e_j is in the sequence S_i. The graph T is a 0-rooted Cayley tree.

According to the previous remark, the objects of $\mathscr{S}_\ell$ may also be seen as intermediates between two objects. If $k = \ell$, then $S_0, \ldots, S_k$ are simply sequences of nodes. If $\ell = 0$, the unique element of the sequence S_0 is a colored tree.

Theorem 2 (generalized Cycle Lemma). *There is a bijection between the sets $\mathscr{S}_\ell$ and $\mathscr{S}_{\ell-1}$ that works by attaching the elements of the sequence S_ℓ to the edges of color ℓ.*

The actual description of the bijection is given in Section 3; an example is given in Figure 1. Observe that for $\ell = k = 1$, the statement is equivalent to the standard Cycle Lemma: there is a bijection between p-uples of rooted plane trees with one pointed node (represented here as a unique tree with an extra node of degree p at the root) and pairs formed of a node of degree p having a marked free edge and a sequence of nodes such that the number of nodes in the sequence equals the total number of free edges.
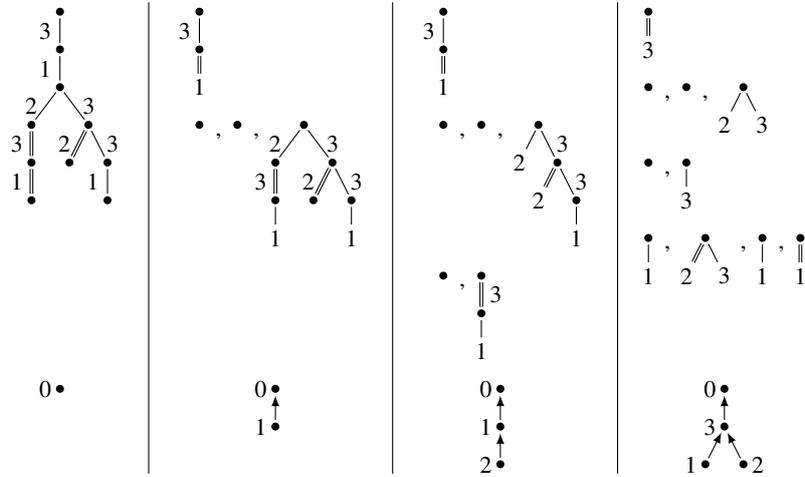

Fig. 1. An example with $k = 3$, a tree of $\mathscr{S}_0$ and the corresponding elements of $\mathscr{S}_1, \mathscr{S}_2$ and $\mathscr{S}_3$ (from left to right), with their associated Cayley trees (bottom). Marked edges are represented as double lines.

Iterating the bijection yields a bijection between $\mathscr{S}_0$ and $\mathscr{S}_k$. Given a Cayley tree T with vertices $\{0, \ldots, k\}$, the generating function of the associated subset of $\mathscr{S}_k$ is $h g_1^{n_1} \cdots g_k^{n_k}$ with the proper derivatives to mark edges (taking a derivative of $g_j^{n_j}$ with respect to x_i amounts to marking an edge of color i in

the sequence S_j). It can thus be seen that the formula (2) is a corollary of the generalized Cycle Lemma.

3 The bijection

We now describe without proof our bijection between the sets $\mathscr{S}_\ell$ and $\mathscr{S}_{\ell-1}$, for $1 \le \ell \le k$, illustrated in Figure 1. This bijection uses *Prüfer codes* of Cayley trees [15]. The Prüfer code of a 0-rooted Cayley tree T with vertices $0, \ldots, s$ is a sequence $p_1, \ldots, p_s$ with $p_s = 0$; there exists a permutation σ such that the parent vertex of j in T is $p_{\sigma(j)}$ for $j = 1, \ldots, s$. Moreover, every sequence corresponds to a unique tree.

Let $S = (S_0, \ldots, S_\ell, e_1, \ldots, e_k)$ be an element of $\mathscr{S}_\ell$; let T be the associated Cayley tree. In the following, we call ℓ-*edge* an edge of color ℓ. Since we are dealing with plane trees, there is a natural order on the set of ℓ-edges of S given by depth-first traversals of the ℓ-bushes.

Let r be the number of ℓ-edges in the sequences $S_0, \ldots, S_{\ell-1}$. Since T is a Cayley tree, it does not contain the arc $\ell \to \ell$, which means that the marked edge e_ℓ is not in the sequence S_ℓ. Thus, we have $r \ge 1$. Let u be the root of the first element of S_ℓ. By Definition 1, the sequence S_ℓ has exactly r more elements than it has ℓ-edges. We combine the elements of S_ℓ using the Cycle Lemma; we denote by $b_1, \ldots, b_r$ the resulting $\ell-1$-bushes, so that the node u is in b_1.

First case: the vertex ℓ is a leaf of T. In this case, we attach the bushes $b_1, \ldots, b_r$ to the remaining free ℓ-edges in order, with a cyclic permutation chosen so that the bush b_1 is attached to the marked edge e_ℓ.

Second case: the vertex ℓ is not a leaf of T. In the Cayley tree T, all arcs going toward ℓ correspond to marked edges in the sequence S_ℓ. Let $m_1, \ldots, m_s$ be the $\ell-1$-bushes that we just formed containing at least one marked edge.

We break up the Cayley tree T in the following manner. For $i = 1, \ldots, s$, let T_i be the forest composed of the colors of the marked edges in m_i and their descendants in T. Let T_0 be the tree composed of the remaining vertices of T with the vertex ℓ deleted; in other words, the vertices of T_0 are exactly the non-descendants of ℓ. We also assume that the order of the m_i's was chosen so that the T_i's are ordered according to their lowest label.

Now, attach the bushes $b_1, \ldots, b_r$ to the free ℓ-edges in order, with a cyclic permutation chosen so that b_s is attached to the marked edge e_ℓ. For $j = 1, \ldots, s$, let $p_j = i$ if m_j is attached to an edge in the sequence with an index in T_i. As the edge e_ℓ is in the sequence corresponding to the parent of ℓ in T, which is in T_0, we have $p_s = 0$.

We can therefore regard the sequence $p_1, \ldots, p_s$ as the Prüfer code of a 0-rooted Cayley tree $\widetilde{T}$ with labels $\{0, \ldots, s\}$. Let σ be the permutation associated to this tree and swap the bushes $m_1, \ldots, m_s$ according to the permutation σ.

Let $S'_0, \ldots, S'_{\ell-1}$ be the sequences resulting from this procedure. As no free ℓ-edges remain, every element of the sequences $S'_0, \ldots, S'_{\ell-1}$ is an $\ell-1$-bush. Finally, let e'_ℓ be the parent edge of the node u defined at the beginning. Let T' be the graph associated with the marked edges $e_1, \ldots, e_{\ell-1}$. In the first case

above, this is the tree T with the leaf ℓ deleted; in the second, it is a compound of the tree T_0 and the forests $T_1, \ldots, T_s$, arranged according to the tree $\tilde{T}$; it is therefore a Cayley tree. This shows that the tuple $(S'_0, \ldots, S'_{\ell-1}, e_1, \ldots, e'_\ell, \ldots, e_k)$ is an element of $\mathscr{S}_{\ell-1}$.

We conclude with some comments. We use Cayley trees and Prüfer codes in a manner that may seem needlessly complicated; however, in the first stage, we see how important the condition that the graph T contains no edge $\ell \to \ell$ is, which is implied by the fact that T is a Cayley tree. The construction aims at ensuring that the graph T' describing the marked edges $e_1, \ldots, e_{\ell-1}$ remains a Cayley tree.

References

1. E. A. Bender and L. B. Richmond. A multivariate Lagrange inversion formula for asymptotic calculations. *Electron. J. Combin.*, 5:Research Paper 33, 4 pp. (electronic), 1998.
2. O. Bernardi and A. H. Morales. Counting trees using symmetries. `arXiv:1206.0598`, 2013.
3. M. Bousquet, C. Chauve, G. Labelle, and P. Leroux. Two bijective proofs for the arborescent form of the Good-Lagrange formula and some applications to colored rooted trees and cacti. *Theoret. Comput. Sci.*, 307(2):277–302, 2003.
4. L. Chottin. Une démonstration combinatoire de la formule de Lagrange à deux variables. *Discrete Math.*, 13(3):215–224, 1975.
5. L. Chottin. Énumération d'arbres et formules d'inversion de séries formelles. *J. Combin. Theory Ser. B*, 31(1):23–45, 1981.
6. N. Dershowitz and S. Zaks. The cycle lemma and some applications. *European J. Combin.*, 11(1):35–40, 1990.
7. A. Dvoretzky and Th. Motzkin. A problem of arrangements. *Duke Math. J.*, 14:305–313, 1947.
8. Ira M. Gessel. A combinatorial proof of the multivariable Lagrange inversion formula. *J. Combin. Theory Ser. A*, 45(2):178–195, 1987.
9. I. P. Goulden and D. M. Kulkarni. Multivariable Lagrange inversion, Gessel-Viennot cancellation, and the matrix tree theorem. *J. Combin. Theory Ser. A*, 80(2):295–308, 1997.
10. A. Joyal. Une théorie combinatoire des séries formelles. *Advances in Mathematics*, 42:1–82, 1981.
11. G. Labelle. Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange. *Adv. in Math.*, 42(3):217–247, 1981.
12. M. Lothaire. *Combinatorics on words*. Cambridge University Press, 2nd edition, 1999.
13. M. Lothaire. *Applied Combinatorics on words*. Cambridge University Press, 2005.
14. J. Pitman. Enumerations of trees and forests related to branching processes and random walks. In D. Aldous and J. Propp, editors, *Microsurveys in Discrete Probability*, volume 41 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* Amer. Math. Soc., Providence, RI, 1998.
15. H. Prüfer. Neuer beweis eines satzes über permutationen. *Arch. Math. Phys.*, 27:742–744, 1918.
16. George N. Raney. Functional composition patterns and power series reversion. *Trans. Amer. Math. Soc.*, 94:441–451, 1960.