Context-free Grammars, Differential Operators and Formal Power Series

WILLIAM Y. C. CHEN

Department of Mathematics
Massachusetts Institute of Technology
Cambridge, MA 02139, USA

Abstract — In this paper, we propose the concept of formal functions over an alphabet and a formal derivative based on a set of substitution rules. We call such a set of rules a context-free grammar because these rules act much like a context-free grammar in the sense of a formal language. Given a context-free grammar, we can associate each formal function with an exponential formal power series. In this way, we obtain a *grammatical interpretation* of the operations addition, multiplication and functional composition of formal power series. A surprising fact about the grammatical calculus is that the composition of two formal power series has a very simple grammatical representation. We also apply this method to obtain a simple demenstration of Faà di Bruno's formula, Bell polynomials, Stirling numbers and symmetric functions. In particular, the Lagrange inversion formula has a simple grammatical representation. From this point of view, we can show that Cayley's formula on labeled trees is essentially equivalent to the Lagrange inversion formula.

1. Introduction

Let A be an alphabet whose letters are regarded as independent commutative indeterminates. A *formal function* over A is defined as follows:

1. Every letter in A is a formal function.
2. If u and v are formal functions, then $u+v$ and uv are also formal functions.
3. If $f(x)$ is an analytic function in x, and u is a formal function, then $f(u)$ is a formal function.
4. Every formal function is constructed as above in a finite number of steps.

We can also define the *formal derivative* of a letter or a formal function by a set of substitution rules. Such a set of substitution rules can be regarded as a context-free grammar in the sense of context-free grammars in the theory of formal langanges. In this paper, an alphabet is allowed to contain infinitely many letters. For this reason, J. Goldman introduced the term *formal schema* to distinguish context-free grammars having infinite alphabets from those having finite alphabets. Given a formal derivative and a formal function, we may associate an exponential formal power series. This is different from the well-known approach to formal languages which use the ordinary formal power series. It is interesting that the common operations on exponential formal power series have simple grammatical explanations. The Lagrange inversion formula has a very simple grammatical representation, which leads to a short combinatorial proof of this formula. In fact, we show that the Lagrange inversion formula is equivalent to Cayley's formula on labeled trees. We also give other examples of grammatical calculus including Bell polynomials, Stirling numbers, and some classical identities on symmetric functions.

2. Context-free Grammar and Formal Derivative

A *context-free* grammar G over A is defined as a set of substitution rules which replace a letter in A by a formal function over A. A rule in a context-free grammar is also called a *production* as in the theory of formal languages. For example, let $A = \{f, g, h\}$, then the following set of productions form a context-free grammar:

$$G = \{f \to 2fg, \quad g \to g\}.$$

We then consider an operator with respect to a context-free grammar G over A. Any formal function over A can be regarded as a function $h(a_1, a_2, \ldots, a_n)$, where a_1, a_2, ..., a_n are letters in A. Since all the letters are independent, we may treat them as abstract symbols for functions in variable x (where x is not a letter in A). Thus, the derivative of a letter in A could be defined as a formal function (we may even denote such a formal function by a new symbol) in order to make the common differential rules still work for formal functions. Thus, we have the following

Proposition 2.1 *The following operator D on formal functions over an alphabet A is well-defined:*

1. *For two formal functions u and v, we have*

$$D(u+v) = D(u) + D(v) \quad \text{and} \quad D(uv) = D(u)v + uD(v).$$

2. *For any analytic function $f(x)$, and any formal function w, we have*

$$Df(w) = \frac{\partial f(w)}{\partial w} Dw.$$

3. *For a letter v in A, if there is a production $v \to w$ in the grammar, where w is a forma[l] function, then $Dv = w$; otherwise $Dv = 0$ and we call such an element v a* constant *or* [a] terminal.

We call the above operator D the *formal derivative* with respect to the grammar G. It is clea[r] that Leibniz's formula still holds for a formal derivative:

$$D^n(fg) = \sum_{k=0} \binom{n}{k} D^k(f)\, D^{n-k}(g).$$

Let's consider a special case where the grammar G is a context-free grammar of a forma[l] language (i.e., every production is a substitution rule of replacing a letter by a word over th[e] alphabet). Let u and v be two words over A, then we must have

$$D(uv) = D(u)v + uD(v),$$

because the substitution must be done in either the word u or v. For example, let $A = \{a, b, c$ and

$$G = \{a \to ab, \quad b \to bc, \quad c \to ca\}.$$

Then we have

$$\begin{aligned} D(ab) &= ab^2 + abc, \\ D^2(ab) &= ab^3 + 3ab^2c + abc^2 + a^2bc. \end{aligned}$$

In the above definition of formal functions, we have assumed that the letters in the alphabet A are commutative indeterminates. However, we may similarly define the formal derivative for noncommutative algebras and define a formal function alternatively as a formal power series over alphabet A of non-commutative indeterminates. For convenience, we shall sometimes identify a letter a with the letter a_0, c with c_0, and so on.

Example 2.2 *Let $a_0, a_1, a_2, \ldots$ and $b_0, b_1, b_2, \ldots$ be two sequences. Then we have the following inversion pair:*

$$a_n = \sum_{k=0}^{n} \binom{n}{k} b_k \quad \text{and} \quad b_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} a_k. \tag{2.1}$$

Proof. Let G be the following grammar:

$$\{f \to f, \quad c_i \to c_{i+1}\}.$$

Denote b_i by $c_i f$. Then the first identity in (2.1) can be rewritten as $a_n = D^n(cf)$. Suppose it is true, then we have $Df^{-1} = -f^{-2}Df = -f^{-1}$, and

$$\begin{aligned} b_n &= f\, D^n(c) \\ &= f\, D^n(cff^{-1}) \\ &= f \sum_{k=0}^{n} \binom{n}{k} D^k(cf) D^{n-k}(f^{-1}) \\ &= \sum_{k=0}^{k} \binom{n}{k} (-1)^{n-k} a_k. \end{aligned}$$

The converse can be proved similarly. □

The next example will be a grammar which will be used throughout this paper:

$$\begin{aligned} f_i &\to f_{i+1} g_1, \\ g_i &\to g_{i+1}. \end{aligned}$$

We shall call this grammar the *Faà di Bruno grammar.* The next proposition gives a connection between this grammar and the lattice of partitions of a finite set.

Definition 2.3 (Type of a Partition) *Let $\pi = \{B_1, B_2, \ldots, B_k\}$ be a partition of an n-set. Suppose B_1 has i_1 elements, B_2 has i_2 elements, $\ldots$, B_k has i_k elements. Then we define the type of π by*

$$\lambda(\pi) = f_k\, g_{i_1}\, g_{i_2} \, \cdots \, g_{i_k}\,.$$

Proposition 2.4 *Let D be the formal derivative of the Faà di Bruno grammar and E be a set of n elements. Then $D^n(f)$ is the sum of types of all partitions of E.*

Proof. Consider a general term $T = f_k g_{i_1} g_{i_2} \cdots g_{i_k}$ in $D^n(f)$. Note that each g_i is obtained by a substitution on an f_j to get g_1, and $i-1$ substitutions on g_1. Thus, each g_i corresponds to an

i-subset of $\{1,2,\cdots,n\}$. When we substitute f_j by $f_{j+1}g_1$, we may always put g_1 at the end of the current term. For example, $D(f_4g_2g_1g_3g_2)$ contains the term $f_5g_2g_1g_3g_2g_1$. By this imposed order on g_i's, the above term T will always correspond to a partition $\{B_1,B_2,\ldots,B_k\}$ of $\{1,2,\ldots,n\}$ whose blocks are ordered in the increasing order of their minimum elements. Since any partition can be uniquely written in such a form, this completes the proof. □

We shall call the above proof the *"partition argument"*. It is easy to see that the number of partitions of $\{1,2,\cdots,n\}$ with k_1 1-blocks, k_2 2-blocks, ..., k_n n-blocks is

$$\frac{n!}{k_1!\,k_2!\cdots k_n!\,1!^{k_1}\,2!^{k_2}\cdots n!^{k_n}}.$$

Therefore, the above proposition can be restated as follows:

$$D^n(f)=\sum_{k=0}^{n} f_k \sum_{k_1,k_2,\cdots,k_n} \frac{n!}{k_1!\,k_2!\cdots k_n!\,1!^{k_1}\,2!^{k_2}\cdots n!^{k_n}}\, g_1^{k_1}g_2^{k_2}\cdots g_n^{k_n}\,, \tag{2.2}$$

where the second summation runs over all nonnegative integers k_1, k_2, ..., k_n such that $k_1+k_2+\ldots k_n=k$ and $k_1+2k_2+\ldots nk_n=n$.

Example 2.5 (Faà di Bruno's Formula) *Let $F(t)=f(g(t))$ be a composite function. Let D_u be the differential operator d/du and set*

$$F_n=D_t^nF(t),\quad f_k=D_u^k[f(u)]_{u=g(t)},\quad g_k=D_t^kg(t).$$

Then we have

$$F_n=\sum_{k=0}^{n} f_k \sum_{k_1,k_2,\cdots,k_n} \frac{n!}{k_1!\,k_2!\cdots k_n!\,1!^{k_1}\,2!^{k_2}\cdots n!^{k_n}}\, g_1^{k_1}g_2^{k_2}\cdots g_n^{k_n}\,,$$

where the range of the second summation is the same as in (2.2).

Proof. Since the Faà di Bruno grammar simulates the procedure to compute the nth derivative F_n, it follows that $D^n(f)$ has the same expression as F_n. □

The above proof can be easily extended to the generalized Faà di Bruno's formula for a function of several functions [2].

3. Formal Power Series

In this section, we shall consider the formal power series of a formal function with respect to a formal derivative. Let G be a context-free grammar on an alphabet A, and D be the formal derivative corresponding to the grammar G. For simplicity, if f is a formal function on an alphabet A and G is a context-free grammar on A, then we may say that f is a formal function on G.

Definition 3.1 (Evaluation of a Formal Function) *Let A be an alphabet and f be a formal function over A. An evaluation on f is a linear function which maps a letter to a real number. We shall use $|f|$ to denote an evaluation on f.* The regular evaluation *is the evaluation which maps every letter to 1.*

Let w be a formal function over an alphabet A and $|w|$ be an evaluation on w. Then we define

$$\begin{aligned} \mathrm{Gen}(w,t) &= \sum_{n\geq 0} D^n(w)\frac{t^n}{n!}, \\ \mathrm{gen}(w,t) &= \sum_{n\geq 0} |D^n(w)|\frac{t^n}{n!}, \\ \mathrm{Gen}^+(w,t) &= \sum_{n\geq 1} D^n(w)\frac{t^n}{n!}, \\ \mathrm{gen}^+(w,t) &= \sum_{n\geq 1} |D^n(w)|\frac{t^n}{n!}. \end{aligned}$$

The formal power series $\mathrm{Gen}^+(w,t)$ and $\mathrm{gen}^+(w,t)$ are called the *delta series* of w. Note that the variable t is not in the alphabet A, namely t is a constant with respect to the derivative defined by a context-free grammar. We shall use D_t to denote the differential operator in the variable t, for convenience, we shall use the common notation $'$ for D_t. For example, we may write $\mathrm{Gen}'(w,t)$ for $D_t(\mathrm{Gen}(w,t))$. The following proposition relates a formal derivative to the ordinary differentiation of a formal power series.

Proposition 3.2 *We have*

$$\begin{aligned} Gen'(w,t) &= Gen(D(w),t), \\ gen'(w,t) &= gen(D(w),t). \end{aligned}$$

We define an *integration* on a formal function as follows: Let w be a formal function over an alphabet A, and D be the formal derivative corresponding to a context-free grammar over A. If there exists a formal function u such that $D(u) = w$, then we say that u is an integration of w, denoted $u = \int w\, dG$. Note that if u is an integration of w, then $u + c$ is also an integration of w provided that c is a constant.

Proposition 3.3 *We have*

$$\begin{aligned} \int Gen(w,t)\, dt &= Gen(\int w\, dG, t), \\ \int gen(w,t)\, dt &= gen(\int w\, dG, t). \end{aligned}$$

Proposition 3.4 *We have*

$$\begin{aligned} Gen(u+v,t) &= Gen(u,t) + Gen(v,t), \\ Gen(uv,t) &= Gen(u,t)Gen(v,t). \end{aligned}$$

Definition 3.5 (Disjoint Grammars) *Let G_1 and G_2 be two context-free grammars on alphabets A and B. Then G_1 and G_2 are said to be disjoint if A and B are disjoint.*

Let G_1 and G_2 be two disjoint grammars. Let w be a formal function on G_2. We define the composition of G_1 and G_2 at w as follows:

Definition 3.6 (Composition of Grammars) *Let G_1 and G_2 be two disjoint context-free grammars on A and B. Let w be a formal function on G_2. Then we denote by $G_1D(w)$ the grammar obtained from G_1 by replacing every rule $u \to v$ in G_1 with $u \to vD(w)$, where D is the formal derivative corresponding to grammar G_2. Then the union of these two grammars (as the union of productions) $G_1D(w)$ and G_2 is called the composition of G_1 and G_2 at w, denoted by $G = G_1(G_2, w)$.*

Note that the above definition can also be stated as $G_1(G_2, w) = G_1 D(w) \cup G_2$. The followi proposition gives the relationship between the composition of two disjoint grammars and t composition of two formal power series.

Proposition 3.7 *Let G_1 and G_2 be two disjoint context-free grammars, f and g be two forn functions on G_1 and G_2 respectively. Let $H(t)$ be the composition of the formal power series oj and the delta series of g, i.e.,*

$$H(t) = Gen(f, Gen^+(g,t)).$$

Then $H(t)$ is the formal power series of f with respect to the grammar $G_1(G_2, g)$.

Proof. Let $F(t) = \mathrm{Gen}(f,t)$, $G(t) = \mathrm{Gen}^+(g,t)$ be the formal power series of f and g with respe to grammars G_1 and G_2. Then $H(t) = F(G(t))$. Let D be the formal derivative with respect the union of the two disjoint grammars G_1 and G_2. Set

$$F_n = \left.\frac{\partial^n F(u)}{\partial u^n}\right|_{u=G(t)}, \quad G_n = D_t^n(G(t)), \quad H_n = D_t^n(H(t)),$$

and

$$f_n = D^n(f), \quad g_n = D^n(g).$$

By the differentiation rules for formal power series, we know that H_n can be obtained as $E^n(I$ where E is the formal derivative with respect to the following grammar G:

$$\{F_i \to F_{i+1} G_1, \quad G_i \to G_{i+1}\}.$$

Since $G(0) = 0$, we have

$$\left.\frac{\partial^n F(u))}{\partial u^n}\right|_{u=G(t)=0} = \left.\frac{\partial^n F(t)}{\partial t^n}\right|_{t=0} = f_n.$$

We also have $g_n = G_n|_{t=0}$, $h_n = H_n|_{t=0}$. Therefore, h_n can be obtained as $h_n = D^n(f)$ in t following induced grammar from E by setting $t = 0$:

$$\{f_i \to f_{i+1} g_1, \quad g_i \to g_{i+1}\}.$$

Clearly the rules $f_i \to f_{i+1} g_1$ are equivalent to the grammar $G_1 D(g)$, and the rules $g_i \to g_{i+1}$ equivalent to the grammar G_2. Since G_1 and G_2 are disjoint, the proof is complete.

It should be noted that the above proposition and the "partition argument" for Faà di Bru grammar imply a combinatorial interpretation of the composition of two formal power series Joyal's theory of species. Given two formal power series

$$f(t) = \sum_{n \geq 0} f_n \frac{t^n}{n!} \quad \text{and} \quad g(t) = \sum_{n \geq 1} g_n \frac{t^n}{n!},$$

let

$$h(t) = f(g(t)) = \sum_{n \geq 0} h_n \frac{t^n}{n!}.$$

Then the above proposition implies that $h_n = D^n(f)$, where D is the formal derivative of the I di Bruno grammar.

Another consequence of the above proposition is a derivation of the formula (2.2) and the I di Bruno's formula for composite functions without using the "partition argument". Let D be formal derivative for the Faà di Bruno grammar. Then the above proposition gives the followi

$$\sum_{n \geq 0} D^n(f) \frac{t^n}{n!} = \sum_{k \geq 0} f_k \frac{(g_1 t + g_2 \frac{t^2}{2!} + g_3 \frac{t^3}{3!} + \ldots)^k}{k!}.$$

By expanding the above formal power series, the coefficient of $\frac{t^n}{n!}$ gives (2.2) and Faà di Bruno's formula.

Example 3.8 *Let*

$$e^{a(e^{bt}-1)} = \sum_{n\geq 0} Q_n \frac{t^n}{n!}.$$

Then we have the following recursion

$$Q_{n+1} = ab\sum_{k=0}^{n} \binom{n}{k} b^{n-k} Q_k. \tag{3.1}$$

Proof. Let G_1 be the grammar $\{f \to a\,f\}$, and G_2 be the grammar $\{g \to b\,g\}$. Then it is obvious that

$$\mathrm{Gen}(f,t) = fe^{at}, \qquad \mathrm{Gen}^+(g,t) = g\left(e^{bt}-1\right).$$

Thus the composition of $\mathrm{Gen}(f,t)$ and $\mathrm{Gen}^+(g,t)$ is $fe^{ag(e^{bt}-1)}$. The composition of G_1 and G_2 at g is

$$\{f \to abfg, \quad g \to bg\}.$$

It follows that

$$\begin{aligned} D^{n+1}(f) &= D^n(abfg) = ab\,D^n(fg) \\ &= ab\sum_{k=0}^{n} \binom{n}{k} D^k(f)\,D^{n-k}(g) \\ &= ab\sum_{k=0}^{n} \binom{n}{k} D^k(f)\,b^{n-k}\,g\,. \end{aligned}$$

Setting f and g to 1 in the above identity, we have (3.1) □

When $a = b = 1$, Q_n becomes the Bell number B_n, i.e., the number of partitions of an n-set. Thus, (3.1) gives the known recursion for B_n:

$$B_{n+1} = \sum_{k=0}^{n} \binom{n}{k} B_k.$$

Set $a = x$ and $b = 1$. Then Q_n will become the generalized Bell number $\phi_n(x)$ (see [19] or (4.2) for definition) and we have the following recursion for $\phi_n(x)$:

$$\phi_{n+1}(x) = x\sum_{k=0}^{n} \binom{n}{k} \phi_k(x)\,.$$

Example 3.9 *Let*

$$e^{e^{(e^t-1)}} - 1 = \sum_{n=0} T_n \frac{t^n}{n!}.$$

Then T_n satisfies the following recursion

$$T_{n+1} = \sum_{i+j+k=n} \binom{n}{i,j,k} T_i\,B_j, \tag{3.2}$$

where B_j is the Bell number.

Proof. The formal power series $e^{e^{(e^t-1)}}-1$ can be obtained as $f(g(h(t)))$, where $g(t)=h(t)=e^t-1$ and $f(t)=e^t$. By Proposition 3.7, $f(g(h(t)))$ is the formal power series of f for the following grammar:

$$\{f \rightarrow fgh, \quad g \rightarrow gh, \quad h \rightarrow h\}.$$

Thus we have

$$\begin{aligned} D^{n+1}(f) &= D^n(fgh) \\ &= \sum_{i+j+k=n} \binom{n}{i,j,k} D^i(f)\, D^j(g)\, D^k(h) \\ &= \sum_{i+j+k=n} \binom{n}{i,j,k} D^i(f)\, D^j(g)\, h\,. \end{aligned}$$

Since $T_n = |D^n(f)|$ and $B_n = |D^n(g)|$, it follows (3.2). □

We note that that T_n is the number of *double partitions* of an n-set. A double partition of a set S is a partition whose underlying set is a partition of S.

4. Examples

In this section we shall give some examples of the utility of the grammatical calculus in deriving certain combinatorial identities. We also give a simple combinatorial proof of the Lagrange inversion formula based on its grammatical representation and using Cayley's formula on the number of rooted trees with a given degree sequence.

4.1 Bell Polynomials

Recall that the Bell polynomials are defined as follows:

$$Y_n(y_1, y_2, \ldots, y_n) = \sum_{k_1,k_2,\cdots,k_n} \frac{n!}{k_1!\, k_2! \cdots k_n!\, 1!^{k_1}\, 2!^{k_2} \cdots n!^{k_n}}\, y_1^{k_1} y_2^{k_2} \cdots y_n^{k_n},$$

where the summation ranges over all nonnegative integers k_1, k_2, $\ldots$, k_n satisfying $k_1 + 2k_2 + \ldots nk_n = n$. Define the grammar G as

$$\{f \rightarrow fy_1, \quad y_i \rightarrow y_{i+1}\}.$$

Then it follows immediately from the partition argument that

$$D^n(f) = f\, Y_n(y_1, y_2, \ldots, y_n).$$

We shall use the evaluation on $D^n(f)$ by setting f to 1. We first give a grammatical proof of the following recursion for Bell polynomials:

$$Y_{n+1}(y_1, y_2, \ldots, y_{n+1}) = \sum_{k=0}^{n} \binom{n}{k} Y_{n-k}(y_1, y_2, \ldots, y_{n-k})\, y_{k+1}. \tag{4.1}$$

Proof. Since $Y_{n+1}(y_1, y_2, \ldots, y_{n+1}) = |D^{n+1}(f)| = |D^n(fy_1)|$, the above identity (4.1) follows immediately from the Leibniz formula. □

In his classic book [17], Riordan used a rather mysterious symbolic method invented in the last century to derive Faà di Bruno's formula. The idea of his symbolic proof is to establish a lifferential equation on symbols, then solve the equation by treating it as an ordinary differential equation. The symbolic calculus has proven to be a very efficient tool in invariant theory and combinatorial enumeration. A rigorous foundation of such symbolic calculus was first found by Rota in his theory of umbral calculus. By using Rota's general theory, Roman [18] eventually found a rigorous explanation of Riordan's symbolic proof of the Faà di Bruno formula. However, Roman's interpretation is not as simple as the symbolic computation itself, so it does not seem to have really explained why it should work. However, it is somehow surprising that our grammatical calculus can give a completely clear and faithful explanation of Riordan's symbolic computation. Let D be the formal derivative of the above grammar G, and

$$\mathrm{Gen}(f,t) = \sum_{n\geq 0} D^n(f)\frac{t^n}{n!} = f\sum_{n\geq 0} Y_n(y_1, y_2, \ldots, y_n)\frac{t^n}{n!}.$$

By differentiation, we have

$$\begin{aligned} D_t(\mathrm{Gen}(f,t)) &= \mathrm{Gen}(D(f),\, t)) \\ &= \mathrm{Gen}(fy_1,\, t) \\ &= \mathrm{Gen}(f,\, t)\,\mathrm{Gen}(y_1,\, t). \end{aligned}$$

That is,

$$D_t(\log \mathrm{Gen}(f,t)) = \mathrm{Gen}(y_1,t).$$

It follows that

$$\begin{aligned} \mathrm{Gen}(f,t) &= e^{\mathrm{Gen}(\int y_1 dG,\, t)+c} \\ &= e^{\mathrm{Gen}(y,\, t)+c} \\ &= e^{\mathrm{Gen}^+(y,\, t)+c}. \end{aligned}$$

By setting $t=0$, we have $f = e^c$ and

$$\mathrm{Gen}(f,t) = fe^{y_1 t + y_2\frac{t^2}{2!} + y_3\frac{t^3}{3!} + \cdots}.$$

Setting $f=1$, we get the formal power series of Bell polynomials, i.e.,

$$\sum_{n\geq 0} Y_n(y_1, y_2, \ldots, y_n)\frac{t^n}{n!} = \mathrm{gen}(f,t) = e^{y_1 t + y_2\frac{t^2}{2!} + y_3\frac{t^3}{3!} + \cdots}.$$

Note that the above grammatical proof involves neither the "partition argument" nor the proposition on composition of grammars. It also suggests the study of formal differential equations based on a context-free grammar.

.2 Stirling Numbers

Recall that the Stirling number $S(n,k)$ of the second kind is the number of partitions of $\{1,2,\ldots,n\}$ with k blocks, and the Stirling number $s(n,k)$ of the first kind is defined such that $(-1)^{n+k}s(n,k)$ the number of permutations on $\{1,2,\ldots,n\}$ with k cycles. We shall call $\phi_n(x)$ the generalized Bell number of order n which is defined as

$$\phi_n(x) = \sum_{k=0}^{n} S(n,k)x^k. \tag{4.2}$$

The following properties of Stirling numbers $S(n,k)$ can be proved grammatically.

$$S(n+1,k) = S(n,k-1)+kS(n,k), \tag{4.3}$$

$$S(n+1,k) = \sum_{j=0}^{n}\binom{n}{j}S(j,k-1), \tag{4.4}$$

$$\sum_{n=0}^{\infty} S(n,k)\frac{t^n}{n!} = \frac{(e^t-1)^k}{k!}, \tag{4.5}$$

$$\phi_n(x) = e^{-x}\sum_{k\geq 0}\frac{x^k k^n}{k!}. \tag{4.6}$$

$$\binom{i+j}{i}S(n,i+j) = \sum_{k=0}^{n}\binom{n}{k}S(k,i)\,S(n-k,j), \tag{4.7}$$

Proof. Let G be the following grammar:

$$\{f\to fg, \quad g\to g\}.$$

From the "partition argument", it follows that

$$D^n(f) = \sum_{k=0}^{n} S(n,k)fg^k. \tag{4.8}$$

Hence we have

$$\begin{aligned} D^{n+1}(f) &= D(D^n(f)) \\ &= D\left(\sum_{k=0}^{n} S(n,k)\,fg^k\right) \\ &= \sum_{k=0}^{n} S(n,k)(fg^{k+1}+kfg^k). \end{aligned}$$

Thus, (4.3) follows by comparing the coefficients of g^k. Also,

$$\begin{aligned} D^{n+1}(f) &= D^n(fg) \\ &= \sum_{j=0}^{n}\binom{n}{j}D^j(f)\,g \\ &= \sum_{j=0}^{n}\binom{n}{j}\sum_{l=0}^{j} S(j,l)\,fg^{l+1}. \end{aligned}$$

Comparing the coefficients of g^k, we may get (4.4). From the composition theorem for grammar we have

$$\mathrm{Gen}(f,t) = \sum_{k=0}^{\infty} D^n(f)\,\frac{t^n}{n!} = fe^{g(e^t-1)}. \tag{4.9}$$

Combining (4.8) and (4.9), (4.5) follows immediately by comparing the coefficients of g^k. Therefore, (4.6) is immediate from (4.9) by setting g to x and then expanding as follows

$$\begin{aligned} e^{x(e^t-1)} &= e^{-x}e^{xe^t} \\ &= e^{-x}\sum_{k\geq 0}\frac{x^k}{k!}e^{kt} \\ &= e^{-x}\sum_{n\geq 0}\left(\sum_{k\geq 0}\frac{x^k k^n}{k!}\right)\frac{t^n}{n!}. \end{aligned}$$

ıs is noted in [19], (4.7) follows from the fact that $\phi_n(x)$ is of binomial type. Here we shall give proof by using the following grammar.

$$G_{x+y} = G_x \cup G_y = \{f \to f\,(x+y), \quad x \to x, \quad y \to y\}.$$

)efine G_x as the grammar by replacing g with x in the above grammar G and define G_y similarly. ,et D_{x+y}, D_x and D_y be the corresponding derivatives of G_{x+y}, G_x and G_y. Since $D_{x+y} = D_x + D_y$, re have

$$\begin{aligned} D^n_{x+y}(f) &= \sum_{m=0}^{n} S(n,m) f (x+y)^m \\ &= \sum_{k=0}^{n} \binom{n}{k} D_x^k \, D_y^{n-k}(f) \\ &= \sum_{k=0}^{n} \binom{n}{k} f \sum_i S(k,p) x^i \sum_j S(n-k,q) y^j. \end{aligned}$$

'hen (4.7) follows by comparing the coefficients of $x^i y^j$. □

It would be interesting to compare the above grammatical proof with the more classical proofs rhich use generating functions and the umbral calculus (see [17, 19]). The identity (4.6) is called he generalized Dobinski's formula. From the operator identity $D^{m+n} = D^m D^n$, we may obtain n identity on Stirling numbers of the second kind which seems to be new. This identity unifies lentities (4.3) and (4.4).

'roposition 4.1 (Vandermonde Convolution for Stirling Numbers)

$$S(m+n,k) = \sum_{i+j \geq k} \binom{m}{j} i^{m-j} \, S(n,i) \, S(j,k-i).$$

Let G be the following grammar

$$\{f \to f\, g_1, \quad g_i \to -i\, g_{i+1}\}.$$

Ve define the evaluation $|D^n(f)|$ by setting g_i to g. By the "partition argument", it is easy to see hat

$$|D^n(f)| = \sum_{k=0}^{n} s(n,k)\, f g^k, \tag{4.10}$$

'here $s(n,k)$ the the Stirling number of the first kind. All the basic identities on $s(n,k)$ can be erived grammatically. We shall also use the above grammar to derive some classical identities on ymmetric functions.

:.3 Symmetric Functions

et's recall the following definitions of some basic symmetric functions:

$$\begin{aligned} a_n(x_1, x_2, \ldots, x_m) &= \sum_{1 \leq i_1 < i_2 < \ldots < i_n \leq m} x_{i_1} x_{i_2} \ldots x_{i_n}, \\ h_n(x_1, x_2, \ldots, x_m) &= \sum_{1 \leq i_1 \leq i_2 \leq \ldots \leq i_n \leq m} x_{i_1} x_{i_2} \ldots x_{i_n}, \\ s_n(x_1, x_2, \ldots, x_m) &= x_1^n + x_2^n + \ldots + x_m^n. \end{aligned}$$

We shall use M to denote a set of m variables $\{x, y, \ldots, z\}$. For a variable x in M, we shall associate it with a sequence of letters $x_0, x_1, x_2, \ldots$, and call the following grammar E_x the *Waring grammar of the first kind*:

$$\begin{aligned} f &\rightarrow f\, x_1, \\ x_i &\rightarrow -i\, x_{i+1}. \end{aligned}$$

The following grammar H_x is called the *Waring grammar of the second kind*:

$$\begin{aligned} f &\rightarrow f\, x_1, \\ x_i &\rightarrow i\, x_{i+1}. \end{aligned}$$

We define the *Waring evaluation* of a formal function by setting f to 1, x_i to x^i, y_i to y^i, and so on. Similar to the "partition argument" , it is easy to prove the following proposition combinatorially since we know that the number of permutations on $\{1, 2, \ldots, n\}$ is $n!$ and the number of even permutations is equal to the number of odd permutations on $\{1, 2, \ldots, n\}$ for $n > 1$.

Proposition 4.2 *In the Waring grammar E_x of the first kind, we have*

$$|D(f)| = x, \qquad |D^n(f)| = 0, \quad \textit{for} \quad n > 1. \tag{4.11}$$

In the Waring grammar H_x of the second kind, we have

$$|D^n(f)| = n!\, x^n. \tag{4.12}$$

From the above proposition, we can easily obtain a grammatical proof of Waring's formulas. From now on, we shall assume that the symmetric functions a_n, h_n and s_n are on the finite set M. It is not difficult to see that may make this assumption without loss of generality.

Proposition 4.3 (Waring's Formulas) *Let a_n, h_n and s_n be the symmetric functions on M as before. Then we have*

$$\sum_{n\geq 0} a_n\, t^n = e^{s_1 t - s_2 \frac{t^2}{2} + s_3 \frac{t^3}{3} - \cdots}, \tag{4.13}$$

$$\sum_{n\geq 0} h_n\, t^n = e^{s_1 t + s_2 \frac{t^2}{2} + s_3 \frac{t^3}{3} + \cdots}. \tag{4.14}$$

Proof. Let G_x, G_y, …, G_z be the Waring grammars of the first kind corresponding to variables x, y, …, z and let D_x, D_y, …, D_z be the formal derivatives with respect to G_x, G_y, …, G_z. Set

$$G_{x+y+\ldots z} = G_x \cup G_y \cup \ldots \cup G_z\,.$$

Denote by $D_{x+y+\ldots+z}$ the formal derivative for $G_{x+y+\ldots+z}$. Then it is clear that

$$D_{x+y+\ldots+z} = D_x + D_y + \ldots + D_z\,.$$

Since $|D_x^k(f)| = 0$ for any $k > 1$ and $|D_x(f)| = x$, it follows that

$$\begin{aligned} |D^n_{x+y+\ldots+z}(f)| &= \sum_{k_1+k_2+\ldots+k_m=n} \binom{n}{k_1, k_2, \ldots, k_m} |D_x^{k_1}(f)|\, |D_y^{k_2}(f)| \cdots |D_z^{k_m}(f)| \\ &= n!\, a_n(x, y, \ldots, z)\,. \end{aligned}$$

This proves (4.13). (4.14) can proved similarly. □

Newton's formulas can also be simply proved grammatically. A combinatorial proof of Newton's formulas has been given by Zeilberger [20].

Proposition 4.4 (Newton's Formulas) *Let a_n, h_n and s_n be the symmetric functions on M as before. Then we have*

$$(n+1)\,a_n = \sum_{k=0}^{n}(-1)^{n-k}\,a_k\,s_{n-k+1}, \tag{4.15}$$

$$(n+1)\,h_n = \sum_{k=0}^{n} h_k\,s_{n-k+1}. \tag{4.16}$$

Proof. First we prove (4.15). Let D be the formal derivative $D_{x+y+\ldots+z}$ for the Waring grammar $G_x \cup G_y \cup \ldots \cup G_z$ of the first kind. Thus,

$$\begin{aligned}
|D^{n+1}(f)| &= |D^n(f(x_1+y_1+\ldots+z_1))| \\
&= \sum_{k=0}^{n}\binom{n}{k}|D^k(f)|\,|D^{n-k}(x_1+y_1+\ldots+z_1)| \\
&= \sum_{k=0}^{n}\binom{n}{k}|D^k(f)|\,(-1)^{n-k}(n-k)!\,s_{n-k+1}\,.
\end{aligned}$$

From the Waring's formula we see that $n!\,a_n = |D^n(f)|$, therefore, we have (4.15). The grammatical proof of (4.16) is similar to that of (4.15). □

Newton's formulas are usually stated as follows

$$\begin{aligned}
s_n - a_1\,s_{n-1} + a_2\,s_{n-2} - a_3\,s_{n-3} + \ldots + (-1)^n\,n\,a_n &= 0, \\
s_n + h_1\,s_{n-1} + h_2\,s_{n-2} + h_3\,s_{n-3} + \ldots - n\,h_n &= 0.
\end{aligned}$$

4.4 The Lagrange Inversion Formula

The Lagrange inversion formula is an important technique in combinatorial enumeration. The first combinatorial proof of this formula was given by Raney[15]. Many other combinatorial proofs have been found since. Here we shall give a grammatical formulation of the Lagrange inversion formula, showing that it is essentially equivalent to Cayley's formula on labeled rooted trees. This leads to a simple combinatorial proof of the Lagrange inversion formula.

Proposition 4.5 (The Lagrange Inversion Formula) *Let $v(x)$ and $R(x)$ be two formal power series satisfying $v(x) = xR(v(x))$. Let*

$$v(x) = \sum_{n\geq 1} v_n \frac{x^n}{n!}.$$

Then we have for $n \geq 1$,

$$v_n = \textit{coefficient of } \frac{x^{n-1}}{(n-1)!} \textit{ in } R(x)^n.$$

We now give a grammatical formulation of the Lagrange inversion formula. Let A be the alphabet $\{v_1, v_2, v_3, \ldots, r_0, r_1, r_2, \ldots\}$, and S be the formal derivative with respect to the grammar:

$$\begin{aligned}
r_i &\rightarrow r_{i+1}, & i \geq 0, \\
v_i &\rightarrow v_{i+1}, & i \geq 1.
\end{aligned}$$

Let D be the formal derivative with respect to the Faà di Bruno grammar:

$$\begin{aligned}
r_i &\rightarrow r_{i+1}\,v_1, & i \geq 0, \\
v_i &\rightarrow v_{i+1}, & i \geq 1.
\end{aligned}$$

Then the Lagrange inversion formula is equivalent to the following form.

Proposition 4.6 (Grammatical Version of the Lagrange Inversion Formula) *Let S and D be the formal derivatives as above. Suppose $v_n = n\, D^{n-1}(r)$ for $n \geq 1$, then we must have*

$$v_n = S^{n-1}(r^n).$$

Now we need to recall some properties of labeled rooted trees. Let T be a rooted tree with vertex set $\{x_1, x_2, \ldots, x_n\}$. For any vertex x_i in T, we shall use d_i to denote the outdegree of x_i — the number of vertices covered by x_i. The type of T is defined as

$$\lambda(T) = \prod_{x_i \in T} r_{d_i}\,.$$

We shall use $\mathbf{R}_n$ to denote the set of all rooted trees on $X = \{x_1, x_2, \ldots, x_n\}$. By the Prüfer correspondence, or a modified version of Prüfer correspondence for rooted trees, it follows immediately that

$$\sum_{T \in \mathbf{R}_n} x_1^{d_1}\, x_2^{d_2} \ldots x_n^{d_n} = (x_1 + x_2 + \ldots + x_n)^{n-1}. \tag{4.17}$$

Equivalently, the number of rooted trees on X with outdegree sequence $(d_1, d_2, \ldots, d_n)$ is

$$\binom{n-1}{d_1,\ d_2,\ \ldots,\ d_n}.$$

If we treat r^n as a word $w = r\,r \ldots r$ (here r stands for r_0), then the derivative S becomes an operator which increases the index of a letter by 1. Suppose we always write a polynomial in $x_1, x_2, \ldots, x_n$ in the standard form $x_1^{i_1} x_2^{i_2} \ldots x_n^{i_n}$. Then the operation of multiplying a polynomial by $(x_1 + x_2 + \ldots + x_n)$ is the same as increasing the power of one of the x_i's by 1. Hence we obtain that $S^{n-1}(r^n)$ is the sum of types of all rooted trees on n vertices.

Proof of the Lagrange Inversion Formula. Since $v_1 = r_0$, we may assume that $v_k = S^{k-1}(r^k)$ for $k = 1, 2, \ldots, n$. Because we have the condition $v_{n+1} = (n+1)D^n(r)$, we need to show the following identity:

$$(n+1)D^n(r) = S^n(r^{n+1}). \tag{4.18}$$

The right hand side of (4.18) is the sum of types of all rooted trees on $\{x_1,\ x_2,\ \ldots,\ x_{n+1}\}$. Since there are $n+1$ ways to choose the root, it suffices to show that $D^n(r)$ is the sum of types of all rooted trees on $\{x_1,\ x_2,\ \ldots,\ x_{n+1}\}$ with root x_{n+1}. For a partition $\{X_1,\ X_2,\ \ldots,\ X_k\}$ of the vertex set $\{x_1,\ x_2,\ \ldots,\ x_n\}$, let T_i be a rooted tree on X_i for $1 \leq i \leq k$. From the rooted trees T_i $(1 \leq i \leq k)$, we may construct a rooted tree T by joining all the roots of T_i's to x_{n+1} and specify x_{n+1} as the root of T. Since the outdegree of x_{n+1} in T is k, it follows that

$$\lambda(T) = r_k\, \lambda(T_1)\, \lambda(T_2) \ldots \lambda(T_k)\,.$$

From the "partition argument', it follows that $D^n(r)$ is the sum of types of all the rooted trees on $\{x_1,\ x_2,\ \ldots,\ x_{n+1}\}$ with root x_{n+1}. This completes the proof. $\square$

Finally, we note that the above proof also shows that Cayley's formula (4.17) follows from the Lagrange inversion formula. Thus, we have shown that these two formulas are equivalent.

Acknowledgements

The author would like to thank M. Hawrylycz, A. Joyal, D. Khoshnevisan, J. A. Stein and J. Oliveira for stimulating discussions, and to thank J. R. Goldman and D. Zeilberger for very helpful comments.

..eferences

1] W. Y. C. Chen, A general bijective algorithm for trees, *Proc. Natl. Acad. Sci. USA*, 87(1990), 9635-9639.

2] L. S. Dederick, Successive derivatives of a function of several functions, *Ann. of Math.* 27(1925)-26), 385-394.

3] M.-P. Delest and G. Viennot, Algebraic languages and polyominoes enumeration, *Theoretical Computer Science*, 34(1984), 169-206.

4] Faà di Bruno, Note sur une nouvelle formule de calcul différentiel, *Quarterly J. Math.* 1(1857), 359-360.

5] L. E. Fraenkel, Formulae for high derivatives of composite functions, *Proc. Cambridge Philos. Soc.* 83(1978), 159-165.

6] I. Gessel, A factorization for formal Laurent series and lattice path enumeration, *J. Combinatorial Theory, Ser. A*, 32(1980), 321-337.

7] S. Ginsburg, The Mathematical Theory of Context-free Languages, McGraw Hill, New York, 1966.

8] J. R. Goldman, Formal languages and enumeration, *J. Combinatorial Theory, Series A*, 24(1978), 318-338.

9] ———, Formal languages and enumeration II, *Ann. of the New York Acad. of Sciences*, 319(1979), 234-241.

0] J. R. Goldman and T. Sundquist, Lattice path enumeration by formal schema, preprint.

] A. Joyal, Une théorie combinatoire des séries formelles, *Adv. in Math.* 42(1981), 1-82.

2] D. E. Knuth, The Art of Computer Programming, vol. 1, 2nd ed. Addison-Wesley, Reading, 1973.

3] G. Labelle, Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange, *Adv. in Math.* 42(1981), 217-247.

] M. Lothaire ed., Combinatorics on Words, Encyclopedia of Mathematics and Its Applications, (G.-C. Rota ed.) vol. 17, Addison-Wesley, London, 1983.

] N. G. Raney, Functional composition patterns and power series reversion, *Trans. Amer. Math. Soc.* 94(1960), 441-451.

] J. Riordan, Derivatives of composite functions, *Bull. Amer. Math. Soc.*, 52(1946), 664-667.

] ———, An Introduction to Combinatorial Analysis, John Wiley & Sons, New York, 1958.

] S. Roman, The formula of Faà di Bruno, *Amer. Math. Monthly*, 87(1980) 805-809.

] G.-C. Rota, Finite Operator Calculus, Academic Press, New York, 1975.

] D. Zeilberger, A combinatorial proof of Newton's identities, *Discrete Math.*, 49(2984), 319.

] ———, Toward a combinatorial proof of the Jacobian conjecture, Combintoire énumérative, *Lecture Notes in Math.* 1234(1986), 370-380.

] ———, Six etudes in generating functions, *Intern. J. Computer Math.* 29(1989), 201-215.