FORMAL CALCULUS, UMBRAL CALCULUS, AND BASIC AXIOMATICS OF VERTEX ALGEBRAS

BY THOMAS J. ROBINSON

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Mathematics

Written under the direction of

James Lepowsky

and approved by

New Brunswick, New Jersey

October, 2009

ABSTRACT OF THE DISSERTATION

Formal calculus, umbral calculus, and basic axiomatics of vertex algebras

by Thomas J. Robinson

Dissertation Director: James Lepowsky

The central subject of this thesis is formal calculus together with certain applications to vertex operator algebras and combinatorics. By formal calculus we mean mainly the formal calculus that has been used to describe vertex operator algebras and their modules as well as logarithmic tensor product theory, but we also mean the formal calculus known as umbral calculus. We shall exhibit and develop certain connections between these formal calculi. Among other things we lay out a technique for efficiently proving certain general formal Taylor theorems and we show how to recast much of the classical umbral calculus as stemming from a formal calculus argument that calculates the exponential generating function of the higher derivatives of a composite function. This formal calculus argument is analogous to an important calculation proving the associativity property of lattice vertex operators. We use some of our results to derive combinatorial identities. Finally, we apply other results to study some basic axiomatics of vertex (operator) algebras. In particular, we enhance well known formal calculus approaches to the axioms by introducing a new axiom, "weak skew-associativity," in order to exploit the $\mathcal{S}_3$-symmetric nature of the Jacobi identity axiom. In particular, we use this approach to give a simplified proof that the weak associativity and the Jacobi identity axioms for a module for a vertex algebra are equivalent, an important result

in the representation theory of vertex algebras.

Acknowledgements

I wish to thank James Lepowsky most particularly for his long standing support, patience and guidance.

I also wish to thank both Yi-Zhi Huang and Haisheng Li, who, among other things, each taught me large portions of the material relevant to this work.

In addition, my thanks go to Stephen Greenfield, Amy Cohen, Ovidiu Costin, Richard Lyons and Richard Wheeden for their help when I first arrived at Rutgers.

I would also like to thank Antun Milas, Robert Wilson, Doron Zeilberger and Louis Shapiro. Of course, I have also benefited greatly from conversations with other members of the faculty at Rutgers, but they are too numerous to name. I further wish to thank Mark Krusemeyer and Gail Nelson, my undergraduate professors, and also Mr. Suggs and Dr. Watson, my high school math teachers. And most certainly all the staff in the Rutgers Math Department who through the years have saved me from myself.

Finally, I am grateful for partial support from NSF grant PHY0901237.

Dedication

To my Mom and Dad.

Table of Contents

Chapter 1

Introduction

The central subject of this thesis is formal calculus together with certain applications to vertex operator algebras and combinatorics. By formal calculus we mean mainly the formal calculus that has been used to describe vertex operator algebras and their modules as well as logarithmic tensor product theory, but we also mean the formal calculus known as umbral calculus. We shall exhibit and develop certain connections between these formal calculi. Among other things we lay out a technique for efficiently proving certain general formal Taylor theorems and we show how to recast much of the classical umbral calculus as stemming from a formal calculus argument that calculates the exponential generating function of the higher derivatives of a composite function. This formal calculus argument is analogous to an important calculation proving the associativity property of lattice vertex operators. We use some of our results to derive combinatorial identities. Finally, we apply other results to study some basic axiomatics of vertex (operator) algebras. In particular, we enhance well known formal calculus approaches to the axioms by introducing a new axiom, "weak skew-associativity," in order to exploit the $\mathcal{S}_3$-symmetric nature of the Jacobi identity axiom. In particular, we use this approach to give a simplified proof that the weak associativity and the Jacobi identity axioms for a module for a vertex algebra are equivalent, an important result in the representation theory of vertex algebras.

In the next two paragraphs we very briefly review some highlights in the development of vertex operator algebras and umbral calculus; then we describe the evolution of the present work. Following that, we give a basic chapter-by-chapter outline, itself followed by a more extensive section-by-section outline of the main body of this work.

In mathematics, vertex operators first arose in the work of J. Lepowsky and R. Wilson, who were seeking concrete representations of affine Lie algebras [LW1], and using this work they were able to give a new, natural interpretation of the Rogers-Ramanujan partition identities [LW2]. Vertex operators were used by I. Frenkel, J. Lepowsky and A. Meurman [FLM1] to construct a natural "moonshine module," whose "character" is the modular function $J(q)$, for the largest sporadic finite simple group, the Fischer-Griess Monster group [G]. This proved a conjecture of J. McKay and J. Thompson and part of the Conway-Norton conjectures [CN]. R. Borcherds [Bor1] introduced the first mathematical definition of vertex algebra which, in particular, extended the relations for the vertex operators used to construct the moonshine module. Using this and other ideas, Borcherds [Bor2] proved the remaining Conway-Norton conjectures [CN] for the moonshine module. A variant of the notion of vertex algebra, that of vertex operator algebra, was introduced in [FLM2] using a "Jacobi identity," which was implicit in Borcherds's work, as the main axiom. Independently, physicists including A. Belavin, A.M. Polyakov and A.B. Zamolodchikov [BPZ], in conformal field theory and string theory, had been using structures that were essentially equivalent to vertex operator algebras, sometimes called chiral algebras. There has been a huge amount of work in the development of the field of vertex (operator) algebra theory. One deep and extensive study has been the braided tensor category theory of non-semisimple modules for vertex algebras, or logarithmic tensor product theory, developed in a monograph by Y.-Z. Huang, J. Lepowsky and L. Zhang [HLZ].

While there is a pre-history of umbral calculus, the first systematic treatment of the old-fashioned, quasi-rigorous methods appeared in a series of papers by Blissard [Bli]. A century later G.C. Rota and others put extensive earlier work on a rigorous footing in works such as [MR], [RKO], [Ga], [Rt], and [RR]. While the general principles of umbral calculus have far-reaching applications, we shall, in this work, be concerned with perhaps the most well-known application, which is for the study of Sheffer sequences, the classical umbral calculus.

This work had two starting points, which led to loosely related results. The first starting point was a simple self-posed homework exercise concerning a generalization of a logarithmic formal Taylor theorem which appeared in [HLZ]. When first taking a course on the subject taught by Professor Lepowsky, I was interested in working out a naïve generalization to include formal iterated logarithms. Since there was interesting classical combinatorics underlying the single logarithm case as discussed in [HLZ] I also checked the corresponding combinatorics in the generalized case. In this context it seemed natural, for certain purposes, to view formal Taylor theorems from a slightly different point of view than that taken in the past, although others had perhaps been implicitly using this point of view in their work as, for example, in [Mi]. This point of view, together with some of its consequences, is discussed in detail in [R2], which we essentially reproduce as part of Chapter 2 in this work.

The formal Taylor theorem shows how the operator $e^{y\frac{d}{dx}}$ acts as a formal translation operator. It is also important, at times, to be able to handle more general formal change of variable operators, such as $e^{yx\frac{d}{dx}}$. In trying to prove corresponding theorems for other operators, it seemed convenient to use substitution operators which could intertwine between the formal change-of-variable operators so that results could be transferred from one to another. In the process of establishing an intertwining substitution between $e^{y\frac{d}{dx}}$ and $e^{yx\frac{d}{dx}}$ on a space involving iterated logarithms it became clear that further generalization to include formal iterated exponential variables was natural for certain calculations and so I took this level of generality for this work.

Further routine generalization related to using substitution maps as intertwining substitutions and checking the underlying combinatorics led to considering more general subcases of what have been called hyperbinomial numbers (see e.g. [W2]) and this, in turn, led me to a rediscovery of the exponential Riordan group via an (incomplete but suggestive enough) combinatorial argument. It was only later that I realized that much of this work could be centered around a calculation of the higher derivatives of a composite function that appeared in [FLM2]. This calculation was interesting for those authors because it was analogous to an important argument establishing the "associativity" property of lattice vertex operators. This led in turn to recasting some of the classical umbral calculus, in which the exponential Riordan group plays a role, roughly speaking, as a matrix representation of the group of classical umbral operators.

During the course of that work, some more analogies between umbral calculus and vertex operator algebras were observed. All of this work was originally in one paper, which kept branching in different directions until it became clear that it should be split into three separate papers, with some overlap of material, namely [R1], [R2] and [R3]. That work is reproduced here in Chapters 2 and 3.

The other starting point for this work occurred, in fact, earlier than the first, although it was not worked out until later, for a specific reason. It was during a course given by Professor Haisheng Li based on [LL] that I initially began working on a naïve extension of material concerning the "weak commutativity" and "weak associativity" axioms for a vertex algebra. Roughly speaking, the presence of three variables in the Jacobi identity axiom for a vertex algebra suggested to me that there should be a third companion property in addition to weak commutativity and weak associativity, which I eventually called "weak skew-associativity." I soon stopped working on this extension because, as the treatment in [LL] shows, the two properties already worked out are enough to recover the full Jacobi identity, and so it seemed pointless to investigate a third companion property since the first two were clearly "enough." In addition, no doubt, saying that vertex algebras are analogous, from one point of view, to commutative associative algebras, as is discussed in [FLM2] and [LL], provided a psychological block to looking at a third property which has no particular importance in the classical case. It was only until a year or two had passed and I was again taking a class based on [LL] given by the other author of that work, Professor Lepowsky, that I again considered this problem, but now that I was more familiar with the material I had more motivation to attempt an easier proof of a certain important equivalence of associativity and the Jacobi identity as axioms for a module for a vertex algebra. This new proof was roughly based on exploiting the S_3-symmetry underlying the notion of vertex algebra, as discussed in [FHL]. For part of this work, which is treated in [R4] and is reproduced as Chapter 4 of this work, it was convenient to systematically rework some of the basic formal calculus for vertex algebras, and so another thread of formal calculus was involved in this work loosely related to the type of material from [R1], [R2] and [R3], presented in Chapters 2 and 3 of this work.

This thesis then combines the material in the papers [R1], [R2], [R3] and [R4]. The main body of work is divided into three parts as Chapters 2, 3 and 4 (where Chapter 2 contains the material from [R1] and [R2], Chapter 3 contains the material from [R3] and Chapter 4 contains the material from [R4]). Chapter 2 discusses the nature and handling of certain aspects of the class of formal Taylor theorems, mainly with a view towards further developing ease and flexibility in providing suitable generalization whenever desired in potential future contexts. Some generalizations together with certain applications to combinatorics are worked out. Chapter 3 begins with an application of the formal Taylor theorem to calculate the exponential generating function of the derivatives of a composite function and then observes that this result shows a connection between the classical umbral calculus and vertex operator algebra theory. Connections between the two areas are then developed further. Ample expository material is supplied setting the results in some historical context. Chapter 4 is only very loosely related to the first two Chapters. It consists of a rather thorough treatment of the basic axiomatic theory of vertex (operator) algebras. The main goal of this study is the further development of the $\mathcal{S}_3$-symmetric nature of the axioms for vertex (operator) algebras and their modules by the introduction of the weak skew-associativity axiom, a companion to the well-known axioms, weak commutativity and weak associativity. Among other things, we apply this property to get a new proof of the result, important in the representation theory of vertex algebras, that the Jacobi identity for a module is equivalent to weak associativity.

Each of the chapters may be read independently of the others and because of this there is perhaps some repetition of material, but the repeated material is written specifically to suit the purpose of each chapter. We note, however, that the final section of Chapter 2 gives the first hints of Chapter 3. We shall have more to say about this below. We give a thorough introduction to each chapter separately in the following sections.

1.1 Introduction to Chapter 2

Our subject is certain aspects of the formal calculus used, as presented in [FLM2], to describe vertex algebras, although we do not treat any issues concerning "expansions of zero," which is at the heart of the subject. An important basic result which we describe in detail is the formal Taylor theorem and this along with some variations is the topic we mostly consider. It is well known, and we recall the simple argument below, that if we let x and y be independent formal variables, then the formal exponentiated derivation $e^{y\frac{d}{dx}}$, defined by the expansion, $\sum_{k\geq 0} y^k \left(\frac{d}{dx}\right)^k /k!$, acts on a (complex) polynomial $p(x)$ as a formal translation in y. That is, we have

$$e^{y\frac{d}{dx}}p(x) = p(x+y). \tag{1.1.1}$$

Formulas of this type, where one shows how a formal exponentiated derivation acts as a formal translation over some suitable space, such as polynomials, are the content of the various versions of the formal Taylor theorem. In the standard literature on formal calculus, the expression $(x+y)^n$ is defined as a formal Taylor series given by the binomial series in nonnegative powers of the second-listed variable. This notational convention is called the "binomial expansion convention," as in [FLM2] and [LL]. (Such series expansions often display interesting underlying combinatorics, as we discuss below.) We note that there are really two issues in this notational definition. One is the relevant "expansion" of interest, which is easy but substantial mathematically. The other is purely a "convention," namely, deciding which listed variable should be expanded in nonnegative powers. Of course, one needs such a definition before even stating a formal Taylor theorem since one needs to know how to define what we mean when we have a formal function whose argument is $(x+y)$. The issue of how to define $\log(x+y)$ for use in the recently developed logarithmic formal calculus is parallel and is discussed in detail in Section 3 of [HLZ], where it was used in setting up some necessary language to handle the recently developed theory of braided tensor categories of non-semisimple modules for a vertex algebra. Actually, in [HLZ] the authors proved a more general formal Taylor theorem than they strictly needed, one involving general complex powers. We discuss this issue of the generality of exponents below. The standard approach therefore is to define the relevant expressions $p(x+y)$ via formal analytic expansions and to then prove the desired formal Taylor theorem. We argue that, in fact, for certain purposes it is more convenient to use formulas of the form (1.1.1) as the definition of $p(x+y)$ whenever we extend beyond the elementary case of polynomials, but most especially if one wishes to extend beyond the logarithmic formal calculus.

Actually, the necessary structure is contained in the "automorphism property," which for polynomials $p(x)$ and $q(x)$ says that

$$e^{y\frac{d}{dx}}(p(x)q(x)) = \left(e^{y\frac{d}{dx}}p(x)\right)\left(e^{y\frac{d}{dx}}q(x)\right).$$

The various formal Taylor theorems may then be interpreted as representations of the automorphism property which specialize properly in the easy polynomial case. We note that from this point of view the "expansion" part of the binomial expansion convention is not a definition but a consequence. (The "convention" part, which tells which listed variable should be expanded in the direction of nonnegative powers is, of course, retained in both approaches as the choice of notational convention.)

Whenever it was necessary to formulate more general formal Taylor theorems, such as in [HLZ], it was heuristically obvious that they could be properly formulated in the standard approach but as soon as one generalizes beyond the case of the logarithmic calculus then there may be some tedious details to work out. It is hoped that the approach presented here may in the future make such generalization more efficient. In particular, we show how to generalize to a space that involves formal logarithmic variables iterated an arbitrary number of times as an example to show how this approach may be applied to desired generalizations.

We noted that the traditional approach to proving generalized formal Taylor theorems via formal analytic expansions may be tedious, and while narrowly speaking this is true, it is also true that these expansions are themselves interesting. Indeed, once we have firmly established the algebra of the automorphism property and the formal Taylor theorem relevant to any given context we may calculate formal analytic expansions. If there is more than one way to perform this calculation we may equate the coefficients of the multiple expansions and find a combinatorial identity. We record certain such identities, which turn out to involve the well-known Stirling numbers of the first kind and thereby recover and generalize an identity similarly considered in Section 3 of [HLZ], which was part of the motivation for this work.

We are sometimes also interested in exponentiating derivations other than simply $\frac{d}{dx}$. For instance, in [HLZ] the authors needed to consider the operator $e^{yx\frac{d}{dx}}$. Such exponentiated derivations were considered in [FLM2], and in fact much more general derivations appearing in the exponent have been treated at length in [H], but we shall only consider a couple of very special cases like those mentioned already. We present what we call "intertwining substitutions," (although the reader should not confuse these with the completely different intertwining maps used in vertex algebra theory) which help us to transfer formulas involving one derivation to parallel formulas for a second one which can be interpreted as a type of representation of the first. The automorphism property holds true for all derivations, but the formal Taylor theorem becomes a parallel statement telling us that another formal exponentiated derivation acts as a formal change of variable other than translation. For example, for a polynomial $p(x)$, one may easily show that

$$e^{yx\frac{d}{dx}}p(x) = p\left(xe^{y}\right).$$

There is additional very interesting material which the automorphism property, the formal Taylor theorem and the notion of intertwining substitution lead to. For instance, it turns out that certain of the basic structures of the classical umbral calculus, which was studied by G.C. Rota, D. Kahaner, A. Odlyzko and S. Roman ([RKO], [Rt], [RR] and [Rm1]), and certain aspects of the exponential Riordan group, which was studied by L.W. Shapiro, S. Getu, W.-J. Woan and L. C. Woodson ([SGWW] and [Sh]), may be naturally formulated and recovered in a similar context to the one we are considering. In this chapter we only indicate this connection in a brief comment. Such material is treated in [R3] and Chapter 3 of this work.

In Section 2.1 we give an expository review of the traditional formulation of formal Taylor theorems. In Section 2.2 we reformulate the material of the previous section from the point of view that formal Taylor theorems may be regarded as representations of the automorphism property. In Section 2.3 we apply our point of view to a general formal Taylor theorem involving "iterated logarithmic" and "iterated exponential" formal variables, a case which would be prohibitively difficult using the traditional method, but which is easy using the point of view laid out in Section 2.2. In Sections 2.4 through 2.7 we show how to calculate various formal analytic expansions which we bypassed in section 2.3. In Section 2.5 we prove a recurrence which is useful for calculating certain formal analytic expansions, but which is also an amusing recurrence on its own. In Section 2.8 we record some underlying combinatorics recovering, in particular, a classical identity involving Stirling numbers of the first kind, which was rediscovered in [HLZ]. In Section 2.9 we consider a relation between the formal translation operator and a second formal change of variable operator. Finally, in Section 2.10 we briefly show a connection to Faà di Bruno's classical formula for the higher derivatives of a composite function following a proof given in [FLM2], as well as a related connection to the umbral calculus.

1.2 Introduction to Chapter 3

We present in this chapter a treatment, from first principles, of certain aspects of the classical umbral calculus, concluding with a connection to the Virasoro algebra. In fact, one of our main purposes is to show connections between the classical umbral calculus and certain central considerations in vertex operator algebra theory. Roughly speaking, until the mid 20-th century, umbral calculus consisted of an abundance of more or less quasi-rigorous tricks making heavy use of notational curiosities which literally resembled "shadow" versions of rigorous calculations. The first attempt to systematize some form of umbral calculus was made by Blissard in a series of papers [Bli]. A later independent treatment by Lucas [Lu] points to even earlier sources which he claimed contained at least some hints of umbral techniques (see also [B]). In the 20-th century, Riordan used an updated old-style umbral calculus to great effect in, for instance, his classic work [Ri2]. The fully rigorous period was ushered in by Gian-Carlo Rota. While there have been numerous papers and works on this subject, e.g. [MR], [RKO], [Ga], [Rt],

[RR] and [Rm1], to list but a few of those dating from the first generation or two of the rigorous period of the subject, we believe that our approach introduces certain novelties. For an extensive bibliography through 2000 as well as some additional history we refer the reader to [BL]. The general principle of umbral techniques reaches far beyond the classical umbral calculus and continues to be a subject of research (see e.g. [Z2]). Our treatment involves only certain portions of the classical umbral calculus of Sheffer sequences as developed in [Rm1].

This work began, unexpectedly, with certain considerations of the formal calculus developed to handle some of the algebraic, and ultimately, analytic, aspects of vertex operator algebra theory. In particular, in order to motivate the mechanism of a key part of their proof of a fundamental "associativity" property of the important structures called lattice vertex operator algebras, the authors of [FLM2] presented a very short generating-function proof calculating the exponential generating function of the higher derivatives of a composite function. See the introduction to Chapter 8 as well as Sections 8.3 and 8.4 of [FLM2] and in particular Proposition 8.3.4, formula (8.4.32) and the comment following it. For the authors of [FLM2], the fact that the short generating-function argument proved a well-known classical result about the higher derivatives of a composite function was an amusing byproduct of their proof of the associativity property of lattice vertex operator algebras. Generating-function ideas, of a variety of types, are fundamental to the development of vertex operator algebra theory.

The most famous formula for the higher derivatives of a composite function is called Faà di Bruno's formula. There are many proofs of this and related formulas dating back to at least the early 19th century (see [Jo] for a brief history, as well as [A], [B], [Bli], [F1], [F2], [Lu], [M], [Me], and [Sc]). Moreover, it is a result that seems basic enough to be prone to showing up in numerous unexpected places, such as in connection with vertex operator algebra theory and also, as I recently learned from Professor Robert Wilson, in the theory of divided power algebras, to give just one more example, where, for instance, a special case of Faà di Bruno's formula is related to, and in fact implies, the fact that certain coefficients are combinatorial and therefore integral, which is the point of interest since one wants a certain construction to work over fields of finite characteristic (see e.g. Lemma 1.3 [Wi]). The result is purely algebraic or combinatorial. For a couple of combinatorial proofs we refer the reader to [Z1] and [Ch]. Although we shall discuss some combinatorics in this chapter for expository purposes, our main thread of development is algebraic throughout and our interest in Faà di Bruno's formula is its long well-known connections with umbral calculus. Indeed there are several umbral style proofs of Faà di Bruno's formula. According to [Jo], an early one of these is due to Riordan [Ri1] using an argument later completely rigorized in [Rm2] and [Ch]. Perhaps even more important, though, is the point of view taken in Section 4.1.8 of [Rm1], where the author discusses what he calls the "generic associated sequence," which he relates to the Bell polynomials, which themselves are closely related to Faà di Bruno's formula. Indeed, the first part of this chapter may, very roughly, be regarded as showing a way to develop some of the classical umbral calculus beginning from such "generic" sequences. We also bring attention more fully to [Ch] in which the formalism of "grammars" and some of the techniques quite closely resemble our approach at this stage, as I recently became aware. We note here that what Roman [Rm1] calls "associated sequences," we shall call "attached sequences," the difference being only that the sequence attached to a given formal function is one and the same as the associated sequence of the compositional inverse of said formal function. The compositional inverse always exists for the formal functions under consideration. This shift in terminology, while mildly annoying, is naturally imposed by our point of view. There are, not unsurprisingly, many such terminological discrepancies in the literature.

We begin our treatment by recalling a special case of the argument used in [FLM2] which calculates the higher derivatives of a composite function. This concise argument has certain closely related predecessors dating to at least the mid 19-th century (see e.g. [Me], [A] and [Sc]). For a more complete history of Faà di Bruno's formula, we refer the reader to [Jo], which, however, neglects to mention the proof in [FLM2]. We note certain important differences between the early proofs and that given in [FLM2]. One is that the early proofs seem to all depend upon the formal Taylor theorem, whereas the proof in [FLM2], while making use of the formal Taylor theorem, is instead technically based on what those authors called the "automorphism property." These two properties are closely related, and the formal Taylor theorem may be regarded as one particular important representation of the automorphism property. Because of the more fundamental nature of the automorphism property the concise argument in [FLM2] is actually carried out in the much more general case where the derivative is replaced by a general derivation. We shall not need this generality. For further discussion of the relation of the formal Taylor theorem and the automorphism property, beyond the very brief treatment we give in Section 3.1, we refer the reader to both [R1] and Chapter 2 of this work. See also [H], where the author develops some of these ideas in a different direction in great generality as the algebraic basis for a far-reaching study that establishes a geometric interpretation of the notion of vertex operator algebra. We also note that Chen [Ch] uses a version of the automorphism property at certain points.

A more important difference for us, between the early proofs of Faà di Bruno's formula and that given in [FLM2], is that the authors of [FLM2] were concerned with the form of the answer, as are we, and so they naturally emphasized the exponential generating function of the higher derivatives as being more important than the coefficients themselves. Indeed, our present treatment is based on the exponential generating function of the higher derivatives of a composite function and although we shall include a calculation of the coefficients in Section 3.12 for the convenience of the reader, we shall never make any use of this calculation. This exponential generating function of the higher derivatives, very roughly speaking, resembles "half of" a vertex operator. By roughly regarding this resemblance as giving rise to an analogue, we shall arrive at certain formulas which may be regarded as umbral analogues of certain standard formulas appearing in Section 8.7 of [FLM2] concerning the action of the Virasoro algebra on certain vertex operators. The observation that such analogues might be playing a role was suggested by Professor James Lepowsky after looking at a preliminary version of this work.

After obtaining the exponential generating function for the higher derivatives of a composite function, we abstract the answer arriving at, roughly, the "generic" sequences referred to above that Roman [Rm1] discussed. Some of our work at this stage is also closely related to the "grammar" formalism in [Ch]. We then develop the main body of material, recalling some of the standard material of the classical umbral calculus, the related Riordan group(s) and a short section on a small piece of the related combinatorics. While we arrive at much of this material in what we believe is a somewhat novel approach, certain further material is included merely to provide more context.

We conclude the chapter by introducing a connection between the classical umbral shifts and the Virasoro algebra. This gives a second connection with vertex operator algebras, since every vertex operator algebra "contains" in a very special way a representation of the Virasoro algebra. The Virasoro algebra itself was studied in the characteristic 0 case in [GF] and the characteristic p analogue was introduced by R. Block in [Bl]. Over $\mathbb{C}$ it may be realized as a central extension of the complexified Lie algebra of polynomial vector fields on the circle, which is itself called the Witt algebra. A certain crucial operator representation was introduced by Virasoro in [V] with unpublished contributions made by J.H. Weis, and the operators of this representation play a well known and essential role in string theory and vertex operator algebra theory (cf. [FLM2]). Our connection with umbral calculus is made via one of these operators.

Since this work is interdisciplinary, relating ideas in vertex operator algebra theory and umbral calculus, we have included quite a lot of expository material, in an effort to make it accessible to readers who are not specialists in both of these fields.

We shall now outline the present work section-by-section. In Section 3.1, along with some basic preliminary material, we begin by presenting a special case of the concise calculation of the exponential generating function of the higher derivatives of a composite function which appeared in the proof of Proposition 8.3.4 in [FLM2]. Using this as our starting point, in Section 3.2 we then abstract this calculation and use the resulting abstract version to derive various results of the classical umbral calculus related to what Roman [Rm1] called associated Sheffer sequences. The umbral results we derive in this section essentially calculate certain adjoint operators, though in a somewhat disguised form. In Section 3.3, we then translate these "disguised" results into more familiar language using essentially the formalism of [Rm1].

In Section 3.4, we recall some further classical results to provide more context for the convenience of the reader. In addition to building on our present point of view, we also use a combination of generating function and operator theoretic techniques. The practice of using both techniques in this subject goes back at least to [Ga] in which Garsia briefly but effectively demonstrated how generating function techniques and operator theoretic techniques could be used to complement one another. We chose our approach for a number of reasons. First, our main reference, [Rm1], emphasized almost wholly operator theoretic techniques following such treatments as [MR], [Rt] and [RR]. We thought it would be instructive to follow Garsia's advice and (especially for the reader following along with [Rm1]) juxtapose the two types of technique, while anchoring our results with some of those of the standard literature. We also tried to hint at how the generating function techniques may be seen to almost transparently show why there ought to be a parallel operator theoretic story (see in particular Remark 3.4.3), although we do not pursue this observation in our exposition. We shall also note in this section how umbral shifts are defined as those operators satisfying what may be regarded as an umbral analogue of the $L(-1)$-bracket-derivative property (cf. formula (8.7.30) in [FLM2]).

In Section 3.5 we make an observation about umbral shifts which will be useful in the last phase of the chapter.

In Section 3.6, we return to our original point of view and show that if instead of beginning with only the higher derivatives of a composite function, we also initially multiply the composite function by a third function and then follow in parallel to our original steps, we arrive at the full Sheffer sequences instead of only the attached (or associated) Sheffer sequences.

Sections 3.7, 3.8 and 3.9 all deal with what Shapiro, Getu, Woan and Woodson [SGWW] have called the exponential Riordan group. As those authors discussed, they noticed, while reading [Rm1] and some of Rota's papers, that they could treat this group and related structures arising from the umbral calculus as free-standing objects of study. Our treatment also deals with the exponential Riordan group as a free-standing object, and aside from needing Definition 3.3.2 from Section 3.3 each one of these sections is largely self-contained. While these sections are essentially expository, a couple of remarks are in order. First, in Section 3.7, we try to make the case that from a purely algebraic point of view, having the umbral operators act on the Sheffer sequence generating functions seems a particularly convenient method to derive the group property. Second, in Section 3.8 we get to advertise the utility of the formal Taylor theorem. Third, it was the combinatorial group multiplication property which was for me the initial clue that I was looking at the umbral calculus at all. I was doing some calculations related to the logarithmic formal calculus as developed in [Mi], and much further developed in [HLZ] for the study of logarithmic tensor product theory for vertex operator algebra modules, when I stumbled across the connection, which was made via [SGWW] when I did an internet search on "group" and "Stirling." It was only later that I realized how I could use the proof of of Proposition 8.3.4 in [FLM2] to better organize this work. Of course, the ubiquitous Stirling numbers of various kinds are playing a role, but we shall not digress so far as to deal with them much in this chapter (see [R1] and [R2] and Chapter 2 of this work). Non-specialist readers who are interested in the combinatorics might wish to refer to works dealing with "the exponential formula" and related topics such as Chapter 3 of [W1], Chapter 1 of [St] or Chapter 8 of [Bo]. There is, of course, a wealth of further related combinatorics.

In Section 3.10 we begin the final phase of this chapter, in which we relate the classical umbral calculus to the Virasoro algebra of central charge 1. Here we recall the definition of the Virasoro algebra along with one special case of a standard "quadratic" representation; cf. Section 1.9 of [FLM2] for an exposition of this well-known quadratic representation. We then show how an operator which was central to our development of the classical umbral calculus is precisely the $L(-1)$ operator of this particular representation of the Virasoro algebra of central charge 1. Using a result which we obtain in Section 3.5, we show a relationship between the classical umbral shifts and the operator now identified as $L(-1)$ and we then introduce those operators which in a parallel sense correspond to $L(n)$ for $n \geq 0$. (Of course, by focusing on only those operators $L(n)$ with $n \geq -1$, which themselves span a Lie algebra, the full Virasoro algebra along with its central extension remain effectively invisible.) We conclude by showing a couple of characterizations of these new operators in parallel to characterizations we already had of the umbral shifts. In particular we also note how the second of these characterizations, formulated as Proposition 3.11.2, may be regarded as an umbral analogue of (8.7.37) in [FLM2], extending an analogue already noted concerning the $L(-1)$-bracket-derivative property.

We note also that Bernoulli polynomials have long had connections to umbral calculus (see e.g. [Mel]) and have recently appeared appeared in vertex algebra theory (see e.g. [L] and [DLM]). It might be interesting to investigate further connections between the two subjects that involve Bernoulli polynomials explicitly.

We conclude this chapter with two expository Sections. Section 3.12 completes the calculation of Faà di Bruno's formula, which was never technically needed for anything else in this work. Section 3.13 gives two similar elementary proofs that the "quadratic" operators recalled in Section 3.10 give a representation of the Virasoro algebra. These proofs closely follow an elementary exposition of this proof given in [FLM2] (see the proof of Theorem 1.9.6 of [FLM2]). Those authors also give a much deeper treatment of this result arising from the theory of vertex operator algebras. One difference in our elementary proofs (besides that we only deal with a special case) is the calculation of the central term. Those authors indicate three approaches, one of which they follow. One whose details they omit is what they call a certain "careful calculation," and in Section 3.13 we supply the details (in our special case) for the benefit of readers not specializing in vertex operator algebra theory.

1.3 Introduction to Chapter 4

This chapter gives an enhancement of certain axiomatic treatments of the notion of module for a vertex algebra, and the notion of vertex algebra itself. We note especially that we handle certain issues of the module theory that are more subtle than in the algebra theory alone, a point made clear in [Li1] (cf. [LL]); we shall discuss these issues in detail below. The notion of vertex algebra was first mathematically defined and considered by Borcherds in [Bor1]. Our treatment follows the formal calculus approach, which originally appeared in [FLM2] and was further developed in [FHL]. In particular, the Jacobi identity, implicit in Borcherds' definition, first appeared in [FLM2]. The original mathematical motivation for the formulation of the notion of vertex algebra and its variant notion of vertex operator algebra was related to work done to construct a natural "moonshine" module for the Monster group, a module conjectured to exist by J. McKay and J. Thompson and constructed in [FLM1] and [FLM2]. It was soon recognized that vertex operator algebras were essentially equivalent to chiral algebras in conformal field theory and string theory, as was discussed in [FLM2].

We take as our main axiom of vertex algebra the Jacobi identity, as in [FLM2] and [FHL]. It is well known that there are various replacement axioms for the Jacobi identity that are useful in the module and representation theory of and construction of vertex algebras. These replacement axioms are based on "commutativity" and "associativity" properties, as developed in [FLM2], [FHL], [DL], [Li1]. This theory is treated in detail in [LL]. In particular, each of the notions of weak commutativity and weak associativity together with other more minor properties may replace the Jacobi identity. For instance, weak commutativity, as well as the equivalence of weak commutativity together with certain minor axioms and the Jacobi identity, first appeared in [DL], in the setting of vertex operator algebras as well as in the much more general settings of generalized vertex algebras and abelian intertwining algebras. In the case of vertex operator algebras, this equivalence was then generalized in [Li1] (cf. [LL]) to handle the theory that does not require any gradings of the algebras and also to handle certain subtle and important issues concerning modules. In this chapter we also work in a setting without gradings and also discuss certain of these issues concerning modules; however, we do not handle the vertex superalgebra case (which is a mild generalization).

Our purpose in this chapter is twofold. First, we introduce the notion of "weak skew-associativity" to complement the properties of weak commutativity and weak associativity of a vertex algebra (cf. [LL]). This third property brings out more fully the $\mathcal{S}_3$-symmetric nature of the axioms for a vertex algebra, which is suggested by the $\mathcal{S}_3$-symmetry of the Jacobi identity presented in [FHL]. Just as weak commutativity and weak associativity may be thought of as vertex-algebraic analogues of the relations $a(bc) = b(ac)$ and $a(bc) = (ab)c$, respectively, for commutative associative algebras, weak skew-associativity is analogous to the third relation in a natural triangle: $b(ac) = (ab)c$. We take especial note that in each of these analogues, "c" always appears in the rightmost position, a point of importance for the module theory, as is discussed in Section 3.6 of [LL] and which is related to the main motivation for the work in this chapter. We show how using weak skew-associativity we may simplify certain proofs of the equivalence of axiom systems for a vertex algebra and for a module for a vertex algebra. In particular, in the final section we derive our main result, which says that, in the presence of certain minor axioms, the Jacobi identity for a module is equivalent to either weak associativity or weak skew-associativity. The equivalence of the Jacobi identity (for a module) with weak associativity (for a module) was shown in [Li1] (cf. Theorem 4.4.5 in [LL]) and enters into the proof of the (nontrivial) equivalence of the notions of representation of, and of module for, a vertex algebra ([Li1]; cf. Theorem 5.3.15 in [LL]).

Our second goal is to more fully check some of the dependencies among the minor properties of a vertex algebra. For instance, we avoid using the vacuum vector in our considerations as long as possible. Since the vacuum vector is analogous to an identity element, this approach is analogous to the study of rings without identity, sometimes known as "rngs." (However, we resist the temptation to call these vacuum-free vertex algebras "ertex algebras"). Although our motivation for this level of generality was not example-driven, we refer the reader to [BD] and [HL], where a vacuum-free setting appeared.

In Section 2, we set up some basic definitions and notation as well as summarize certain formal calculus results which we need. Almost all the relevant results may also be found in the fuller account presented in [FLM2] as well as in Chapter 2 of [LL].

In Section 3, we further develop the formal calculus, essentially redoing many calculations which are usually performed after the definition of vertex algebra is given. Our goal is to systematize these calculations and to demonstrate how they depend only on the formal calculus rather than on any of the particulars of the vertex algebras where they are applied. We note that a similar approach was taken in [Li2], where the reader should compare the statement of Lemma 2.1 in [Li2] with the statement of Proposition 3.3 of this chapter (the proof of Lemma 2.1 in [Li2] is related to the proofs of both Propositions 3.2 and 3.3 in this work); the latter result is an extension of the former. In particular, this extension is relevant for handling skew-associativity properties in addition to commutativity and associativity properties.

In Section 4, we define a vertex algebra without vacuum, which we call a vacuum-free vertex algebra. We then note how our formal calculus results in Section 3 may be immediately applied without further comment to show how the main axiom, the Jacobi identity, may be replaced in the definition of a vacuum-free vertex algebra by any two of weak commutativity, weak associativity and weak skew-associativity. We next formalize what we call vacuum-free skew-symmetry. This notion is often used in the literature when necessary but is not highlighted or named, mostly because with a vacuum vector one is guaranteed to have a $\mathcal{D}$ operator and therefore one may obtain skew-symmetry [Bor1] for a vertex algebra. Since we are trying to work with a minimum of assumptions, we shall not have such a $\mathcal{D}$ operator, at least at this stage in the development, so that we cannot even state skew-symmetry. We then show how the Jacobi identity may be replaced as an axiom by vacuum-free skew-symmetry together with any single one of weak commutativity, weak associativity or weak skew-associativity. The strategy employed is the same as that used in [LL], where the analogy between vertex algebras and commutative associative algebras provides classical guides.

In Section 5, we define the notion of module for a vacuum-free vertex algebra and show the parallel results concerning replacement of the (module) Jacobi identity.

In Section 6, we recall the definition of vertex algebra (with vacuum) so that we may exactly recover certain results considered in [LL]. In developing the minor properties related to the vacuum vector and the $\mathcal{D}$ operator, we make more prominent use of vacuum-free skew-symmetry than in other treatments, as far as the author is aware. We then develop various replacement axioms for the Jacobi identity and a couple of further minor results which will be useful in the final section. Whereas without the vacuum vector we have derived consequential properties from the Jacobi identity by "slicing" it with residues or using visible symbolic symmetry, with the vacuum vector at our disposal the strategy is to plug it into our known formulas, thereby specializing them. Then, as before, once we have derived the remaining minor properties, we attempt to piece combinations of them together, using classical guides when we can, in order to build back up to recover certain remaining properties. We note that in [Li2], the author considered certain generalizations of vertex algebras where, in place of the Jacobi identity, only weak associativity was assumed as an axiom. Because of this generalization, it was natural (or really necessary) for the author to examine more carefully certain dependencies. Some of our results are therefore, at least in aesthetic terms, developed in the same spirit. In particular, the reader should compare Proposition 2.6 and Corollary 2.7 in [Li2] with Propositions 4.5.7 and 4.5.8 of this work, where the former results were already stronger than ours, as we discuss in Remark 4.5.6 (see also Remark 4.5.8).

In Section 7, we define a module for a vertex algebra (with vacuum) and, after a couple of preliminary results, we present the main result of this chapter. Namely, we show that the module Jacobi identity may be replaced by either module weak associativity or module weak skew-associativity. The result that the module Jacobi identity may be replaced by module weak associativity was shown in [Li1] (cf. Theorem 4.4.5 in [LL]) and a corollary to it was used in [LL] to show, following [Li1], the equivalence between the notions of representation of and module for a vertex algebra (see Theorem 5.3.15 in [LL]).

There are many treatments of axiom systems for the notion of vertex (operator) algebra in the literature, involving the results of [Bor1], [FLM2], [FHL], [DL] and [Li1] mentioned above, but as far as we are aware, the results of this chapter that did not essentially appear in those works have not appeared before. This work is not intended as a survey of the many existing treatments of axioms. The reader may wish to consult the bibliography in [LL]. In any case, by introducing weak skew-associativity and viewing it as being on equal footing with weak commutativity and weak associativity, we are bringing to light the fuller $\mathcal{S}_3$-symmetric nature of the family of axiom systems, extending beyond and suggested by the $\mathcal{S}_3$-symmetry of the Jacobi identity [FHL].

Except for certain minor exceptions, the well-known results which we shall recall appeared in [FLM2]. However, for convenience, we shall use the (mostly expository) treatment in [LL] when we give specific references to basic results and definitions, etc.

Chapter 2

Exponentiated derivations, the formal Taylor theorem, and Faà di Bruno's formula

2.1 The formal Taylor theorem: a traditional approach

We begin by recalling some elementary aspects of formal calculus (cf. e.g. [FLM2]). We write $\mathbb{C}[x]$ for the algebra of polynomials in a single formal variable x over the complex numbers; we write $\mathbb{C}[[x]]$ for the algebra of formal power series in one formal variable x over the complex numbers, and we also use obvious natural notational extensions such as writing $\mathbb{C}[x][[y]]$ for the algebra of formal power series in one formal variable y over $\mathbb{C}[x]$. Further, we shall frequently use the notation e^w to refer to the formal exponential expansion, where w is any formal object for which such expansion makes sense. For instance, we have the linear operator $e^{y\frac{d}{dx}} : \mathbb{C}[x] \to \mathbb{C}[x][[y]]$:

$$e^{y\frac{d}{dx}} = \sum_{n\geq 0} \frac{y^n}{n!} \left(\frac{d}{dx}\right)^n .$$

Proposition 2.1.1. *(The "automorphism property") Let A be an algebra over $\mathbb{C}$. Let D be a derivation on A. That is, D is a linear map from A to itself which satisfies the product rule:*

$$D(ab) = (Da)b + a(Db) \text{ for all } a \text{ and } b \text{ in } A.$$

Then

$$e^{yD}(ab) = \left(e^{yD}a\right)\left(e^{yD}b\right).$$

Proof. Notice that

$$D^n ab = \sum_{n=0}^{r} \binom{r}{n} D^{r-n} a D^n b.$$

Then divide both sides by $n!$ and sum over y and the result follows. □

Proposition 2.1.2. *(The polynomial formal Taylor theorem) For* $p(x) \in \mathbb{C}[x]$*, we have*

$$e^{y\frac{d}{dx}}p(x) = p(x+y).$$

Proof. By linearity we need only check the case where $p(x) = x^m$, m a nonnegative integer. We simply calculate as follows:

$$\begin{aligned} e^{y\frac{d}{dx}}x^m &= \sum_{n\geq 0} \frac{y^n}{n!}\left(\frac{d}{dx}\right)^n x^m \\ &= \sum_{n\geq 0} \frac{y^n}{n!}(m)(m-1)\cdots(m-(n-1))x^{m-n} \\ &= \sum_{n\geq 0} \binom{m}{n} x^{m-n}y^n \\ &= (x+y)^m. \end{aligned}$$

$\square$

Here, so far, we are, of course, using only the simplest, combinatorially defined binomial coefficients, $\binom{m}{n}$ with $m, n \geq 0$. We observe that the only "difficult" point in the proof is knowing how to expand $(x+y)^m$ as an element in $\mathbb{C}[x][[y]]$. In other words, the classical binomial theorem is at the heart of the proof of the polynomial formal Taylor theorem as well as at the heart of the proof of the automorphism property.

In order to extend the polynomial formal Taylor theorem to handle the case of Laurent polynomials, we extend the binomial notation to include expressions $\binom{m}{n}$ with $m < 0$ and we also recall the binomial expansion convention:

Definition 2.1.1. We write

$$(x+y)^m = \sum_{n\geq 0} \binom{m}{n} x^{m-n}y^n \quad , m \in \mathbb{Z}, \tag{2.1.1}$$

where we assign to $\binom{m}{n}$ the algebraic (rather than combinatorial) meaning: for all $m \in \mathbb{Z}$ and n nonnegative integers

$$\binom{m}{n} = \frac{(m)(m-1)\cdots(m-(n-1))}{n!}. \tag{2.1.2}$$

Remark 2.1.1. In the above version of the binomial expansion convention we may obviously generalize to let $m \in \mathbb{C}$.

With our extended notation, as the reader may easily check, the above proof of Proposition 2.1.2 exactly extends to give:

Proposition 2.1.3. *(The Laurent polynomial formal Taylor theorem) For* $p(x) \in \mathbb{C}[x, x^{-1}]$, *we have*

$$e^{y\frac{d}{dx}} p(x) = p(x+y).$$

$\square$

Notation 2.1.1. We write $\mathbb{C}\{[x]\}$ for the algebra of finite sums of monomials of the form cx^r where c and $r \in \mathbb{C}$.

As the reader may easily check, the above proof of Proposition 2.1.2 exactly extends even further to give:

Proposition 2.1.4. *(The generalized Laurent polynomial formal Taylor theorem) For* $p(x) \in \mathbb{C}\{[x]\}$, *we have*

$$e^{y\frac{d}{dx}} p(x) = p(x+y).$$

Remark 2.1.2. There is an alternate approach to get the generalized Laurent polynomial formal Taylor theorem, an approach which has the advantage that no additional calculation is necessary in the final proof. The argument is simple. For $r \in \mathbb{C}$, we need to verify that

$$e^{y\frac{d}{dx}} x^r = (x+y)^r.$$

Now simply notice that both expressions lie in

$$\mathbb{C}x^r[x^{-1}][[y]]$$

with coefficients being polynomials in r. But the polynomials on matching monomials agree for r a nonnegative integer and so they must be identical. An argument in essentially this style appeared in [HLZ] to prove a logarithmic formal Taylor theorem (Theorem 3.6 of [HLZ]).

We now extend our considerations to a logarithmic case.

Definition 2.1.2. Let $\log x$ be a formal variable commuting with x and y such that $\frac{d}{dx}\log x = x^{-1}$.

We shall need to define expressions involving $\log(x+y)$. In parallel with (2.1.1) we shall define $(\log(x+y))^r$, $r \in \mathbb{C}$, by its formal analytic expansion:

Notation 2.1.2.

$$(\log(x+y))^r = \left(\log x + \log\left(1+\frac{y}{x}\right)\right)^r, \tag{2.1.3}$$

where we make a second use of the symbol "log" to mean the usual formal analytic expansion, namely

$$\log(1+X) = \sum_{i\geq 0} \frac{(-1)^{i-1}}{i} X^i,$$

and where we expand (2.1.3) according to the binomial expansion convention.

Remark 2.1.3. We note that (2.1.3) is a special case of the definition used in the treatment of logarithmic formal calculus in [HLZ]. Our special case avoids the complication of the generality, treated in [HLZ], of (uncountable, non-analytic) sums over $r \in \mathbb{C}$.

Remark 2.1.4. The reader will need to distinguish from context which use of "log" is meant.

Proposition 2.1.5. *(The generalized polynomial logarithmic formal Taylor theorem) For $p(x) \in \mathbb{C}\{[x, \log x]\}$, we have*

$$e^{y\frac{d}{dx}} p(x) = p(x+y).$$

Proof. By linearity and the automorphism property, we need only check the case $p(x) = (\log x)^r$, $r \in \mathbb{C}$. We could proceed by explicitly calculating

$$e^{y\frac{d}{dx}} (\log x)^r,$$

but this is somewhat involved. Instead we argue as in Remark 2.1.2 to reduce to the case $r = 1$. Even without explicitly calculating $e^{y\frac{d}{dx}}(\log x)^r$, it is not hard to see that it is in

$$\mathbb{C}[r](\log x)^r\mathbb{C}[(\log x)^{-1}, x^{-1}][[y]].$$

When we expand (2.1.3) we find that it is also in

$$\mathbb{C}[r](\log x)^r\mathbb{C}[(\log x)^{-1}, x^{-1}][[y]].$$

Thus we only need to check the case for r a positive integer. A second application of the automorphism property now shows that we only need the case where $r = 1$. This case is not difficult to calculate:

$$\begin{aligned} e^{y\frac{d}{dx}}\log x &= \log x + \sum_{i\geq 1}\frac{y^i}{i!}\left(\frac{d}{dx}\right)^i\log x \\ &= \log x + \sum_{i\geq 1}\frac{y^i}{i!}\left(\frac{d}{dx}\right)^{i-1}x^{-1} \\ &= \log x + \sum_{i\geq 1}\frac{y^i}{i}(-1)^{i-1}x^{-i} \\ &= \log x + \log\left(1+\frac{y}{x}\right). \end{aligned}$$

□

Remark 2.1.5. Although we are working in a more special case than that considered in [HLZ], the argument presented in the proof of Proposition 2.1.5 could be used as a replacement for much of the algebraic proof of Theorem 3.6 in [HLZ] as long as one is not concerned with calculating explicit formal analytic expansions and checking the corresponding combinatorics. These two approaches are very similar, however, the difference only being how much work is left implicit. In the next section we shall take a different point of view altogether.

2.2 The formal Taylor theorem from a different point of view

From the examples in Section 2.1 we see a common strategy for formulating a formal Taylor theorem:

1) Pick some reasonable space (e.g., $\mathbb{C}[x]$, $\mathbb{C}\{[x, \log x]\}$) on which $\frac{d}{dx}$ acts in a natural way. The space need not be an algebra, but in this chapter we shall only consider this case.

2) Choose a plausible formal analytic expansion of relevant expressions involving $x+y$ (e.g., $(x+y)^r$, $r \in \mathbb{C}$, $\log(x+y)$).

3) Consider the equality $e^{y\frac{d}{dx}}p(x) = p(x+y)$ and either directly expand both sides to show equality or if necessary use a trick like in Remark 2.1.2.

Step 2 is necessarily anticipatory and dependent on formal analytic expressions. Therefore it seems natural to replace Step 2 by simply defining expressions involving $x+y$ in terms of the operator $e^{y\frac{d}{dx}}$. Then the formal Taylor theorem is trivially true, being viewed now as a (plausible) representation of the underlying structure of the automorphism property. We redo the previous work from this point of view.

Proposition 2.2.1. *(The polynomial formal Taylor theorem) For $p(x) \in \mathbb{C}[x]$, we have*

$$e^{y\frac{d}{dx}}p(x) = p(x+y).$$

Proof. We have by the automorphism property:

$$e^{y\frac{d}{dx}}p(x) = p\left(e^{y\frac{d}{dx}}x\right) = p(x+y).$$

$\square$

Now for the replacement step:

Definition 2.2.1.

$$(x+y)^r = e^{y\frac{d}{dx}}x^r \qquad \textit{for} \quad r \in \mathbb{C}.$$

Remark 2.2.1. Of course, Definition 2.2.1 is equivalent to Definition 2.1.1 together with Remark 2.1.1. This definition immediately leads to the most convenient proofs of certain "expected" basic properties, instead of needing to wait (as is often done) to prove a formal Taylor theorem to officially obtain these proofs. For example, we have:

$$\begin{aligned}(x+y)^{r+s} &= e^{y\frac{d}{dx}}x^{r+s} \\ &= e^{y\frac{d}{dx}}(x^r x^s) \\ &= \left(e^{y\frac{d}{dx}}x^r\right)\left(e^{y\frac{d}{dx}}x^s\right) \\ &= (x+y)^r(x+y)^s.\end{aligned}$$

Proposition 2.2.2. *(The generalized Laurent polynomial formal Taylor theorem) For* $p(x) \in \mathbb{C}\{[x]\}$,

$$e^{y\frac{d}{dx}}p(x) = p(x+y).$$

Proof. This is trivial. $\square$

We also have this example of the replacement step:

Definition 2.2.2.

$$(\log(x+y))^r = e^{y\frac{d}{dx}}(\log x)^r \qquad r \in \mathbb{C}.$$

Proposition 2.2.3. *(The generalized polynomial logarithmic formal Taylor theorem) For* $p(x) \in \mathbb{C}\{[x, \log x]\}$,

$$e^{y\frac{d}{dx}}p(x) = p(x+y).$$

Proof. The result follows by considering the trivial cases $p(x) = x^r$ and $p(x) = (\log x)^r$ for $r \in \mathbb{C}$ and applying the automorphism property. $\square$

The formal analytic expansions are now viewed as calculations rather than definitions or conventions (that is except for the convention that we still list "y," the variable expanded in the direction of nonnegative powers, in the second position). So, for instance, we may calculate the expansions (2.1.1) and (2.1.3) as consequences rather than viewing them as definitions.

2.3 The formal Taylor theorem for iterated logarithms and exponentials

We demonstrate the utility, for the purposes of generalization, of our approach to the formal Taylor theorem, which we discussed in the previous section, with a more involved example. Let $\ell_n(x)$ be formal commuting variables for $n \in \mathbb{Z}$. We define an action of $\frac{d}{dx}$, a derivation, on

$$\mathbb{C}[\ldots, \ell_{-1}(x)^{\pm 1}, \ell_0(x)^{\pm 1}, \ell_1(x)^{\pm 1}, \ldots]$$

(which for short we denote by $\mathbb{C}[\ell^{\pm 1}]$) by

$$\begin{aligned} \frac{d}{dx}\ell_{-n}(x) &= \prod_{i=-1}^{-n} \ell_i(x), \\ \frac{d}{dx}\ell_n(x) &= \prod_{i=0}^{n-1} \ell_i(x)^{-1}, \\ \text{and} \qquad \frac{d}{dx}\ell_0(x) &= 1, \end{aligned}$$

for $n > 0$.

Remark 2.3.1. Secretly, $\ell_n(x)$ is the $(-n)$-th iterated exponential for $n < 0$ and the n-th iterated logarithm for $n > 0$ and $\ell_0(x)$ is x itself.

We make the following, by now typical, definition in order to obtain a formal Taylor theorem.

Definition 2.3.1. *Let*

$$\ell_n(x+y) = e^{y\frac{d}{dx}}\ell_n(x) \qquad \textit{for} \quad n \in \mathbb{Z}.$$

This gives:

Proposition 2.3.1. *(The iterated exponential/logarithmic formal Taylor theorem) For $p(x) \in \mathbb{C}[\ell^{\pm 1}]$ we have:*

$$e^{y\frac{d}{dx}}p(x) = p(x+y).$$

Proof. The result follows from the automorphism property. $\square$

We shall calculate formal analytic expansions of $\ell_n(x+y)$ later, where we shall see how unwieldy they are and how unsuited they are for developing the formal Taylor theorem along traditional lines in this level of generality, whereas our present approach works simply and smoothly.

2.4 Formal analytic expansions: warmup

In this section we begin to calculate formal analytic expansions of $\ell_N(x+y)^m$ for $N \in \mathbb{Z}$, $m \in \mathbb{C}$. Actually, we shall content ourselves with the cases $N \geq -2$ and leave further expansions to the reader. Indeed, one could continue to generalize interminably, but without further motivation this seems less than worthwhile. Recall Remark 2.3.1 for the "true meanings" of these objects. We shall begin by restricting our attention to the cases $N = -2, -1, 0, 1$. The cases $N = -1, 0$ are easy but somewhat illustrative so we shall work through them in detail. We shall use four different methods for each case.

2.4.1 Case $N = 0$, the case "x"

We would like to calculate the formal analytic expansion of $e^{\frac{d}{dx}}\ell_0(x)^n$ for $n \in \mathbb{C}$.

Method 1

We shall establish a recursion with initial condition having a unique solution which yields the coefficients of the expansion. We first must note what form the solution is up to coefficients. It is easy to see that we have

$$\left(\frac{d}{dx}\right)^m \ell_0(x)^n = \left(\frac{\partial}{\partial \ell_0(x)}\right)^m \ell_0(x)^n = c_m \ell_0(x)^{n-m},$$

where $c_m \in \mathbb{C}$. Then it is easy to see that

$$c_m = (n-(m-1))c_{m-1},$$

along with the initial condition that $c_0 = 1$. This recursion together with initial condition, of course, uniquely determines c_m for $m \geq 0$ so that we may solve for $e^{\frac{d}{dx}}\ell_0(x)^n$

in terms of the coefficients of the solution of this recursion. That is, we have

$$e^{\frac{d}{dx}}\ell_0(x)^n = \sum_{k\geq 0} \frac{c_k}{k!}\ell_0(x)^{n-k}y^k.$$

One could, of course, solve the recursion any way one likes, but we shall leave that sort of solution in this (very easy) and all further (sometimes a bit more difficult) cases to the interested reader and consider this method now complete.

Method 2

In this case this method is too trivial to mention separately.

Method 3

It is easy to see how the m-th power of $\frac{d}{dx}$ acts on $\ell_0(x)^n$ because $\frac{d}{dx}\ell_0(x)^n = n\ell_0(x)^{n-1}$, which is a monomial. This is the essential observation that this method is based on. Later we shall have to expand $\frac{d}{dx}$ as a sum of linear operators yielding such monomial results, but in this case it is already immediate that for $n \in \mathbb{C}$ and $m \geq 0$ that $\frac{d}{dx}^m\ell_0(x)^n = n(n-1)\cdots(n-m+1)\ell_0(x)^{n-m}$ so that

$$e^{y\frac{d}{dx}}\ell_0(x)^n = \sum_{k\geq 0} \frac{n(n-1)\cdots(n-k+1)}{k!}\ell_0(x)^{n-k}y^k.$$

Method 4

Here we first calculate $e^{y\frac{d}{dx}}\ell_0(x)$ and then using the automorphism property, or alternatively the formal Taylor theorem, we may find $e^{y\frac{d}{dx}}\ell_0(x)^n$ for $n \in \mathbb{N}$ by using a (nonnegative integral) binomial expansion. We have

$$e^{y\frac{d}{dx}}\ell_0(x)^n = (\ell_0(x) + y)^n = \sum_{k\geq 0} \binom{n}{k}\ell_0(x)^{n-k}y^k.$$

Then by explicitly expressing $\binom{n}{k}$ as a polynomial in n, using a formula like (2.1.2), it is easy to see by arguing as in Remark 2.1.2 that the result extends for all $n \in \mathbb{C}$.

Remark 2.4.1. Of course, considering the three methods and equating coefficients, we have thus calculated the binomial coefficients explicitly and also found them as the unique solution to a recurrence equation with initial condition. We shall later equate coefficients of more cases to record combinatorial identities, some of them classical identities involving Stirling numbers of the first and second kinds.

2.4.2 Case $N = -1$, the case "exp x"

We would like to calculate the formal analytic expansion of $e^{\frac{d}{dx}}\ell_{-1}(x)^n$ for $n \in \mathbb{C}$.

Method 1

It is easy to see that we have

$$\left(\frac{d}{dx}\right)^m \ell_{-1}(x)^n = \left(\ell_{-1}(x)\frac{\partial}{\partial \ell_{-1}(x)}\right)^m \ell_{-1}(x)^n = c_m \ell_{-1}(x)^n,$$

where $c_m \in \mathbb{C}$, and where

$$c_m = nc_{m-1},$$

along with the initial condition that $c_0 = 1$. This recursion together with initial condition, of course, uniquely determines c_m for $m \geq 0$ so that we may solve for $e^{\frac{d}{dx}}\ell_{-1}(x)^n$ in terms of the coefficients of the solution of this recursion. That is, we have

$$e^{\frac{d}{dx}}\ell_{-1}(x)^n = \sum_{k\geq 0} \frac{c_k}{k!}\ell_{-1}(x)^n y^k.$$

Method 2

We note that for our purposes $\frac{d}{dx} = \ell_{-1}(x)\frac{\partial}{\partial \ell_{-1}(x)}$, and that the m-th power for $m \geq 0$ may be normally ordered so that we have

$$\left(\ell_{-1}(x)\frac{\partial}{\partial \ell_{-1}(x)}\right)^m = \sum_{i=0}^{m-1} a_i \ell_{-1}(x)^{m-i}\left(\frac{\partial}{\partial \ell_{-1}(x)}\right)^{m-i},$$

for some $a_i \in \mathbb{N}$ which we leave undetermined. To see this, one need only observe that

$$\left[\frac{\partial}{\partial \ell_{-1}(x)}, \ell_{-1}(x)\right] = \left[\frac{d}{dx}, x\right] = 1.$$

Then we have

$$\begin{aligned}\left(\frac{d}{dx}\right)^m \ell_{-1}(x)^n &= \sum_{i=0}^{m-1} a_i \ell_{-1}(x)^{m-i}\left(\frac{\partial}{\partial \ell_{-1}(x)}\right)^{m-i}\ell_{-1}(x)^n \\ &= \sum_{i=0}^{m-1} a_i \ell_{-1}(x)^{m-i} n(n-1)\cdots(n-(m-i))\ell_{-1}(x)^{n-(m-i)} \\ &= \ell_{-1}(x)^n \sum_{i=0}^{m-1} a_i n(n-1)\cdots(n-(m-i)),\end{aligned}$$

giving in turn

$$e^{\frac{d}{dx}}\ell_{-1}(x)^n = \ell_{-1}(x)^n \sum_{k\geq 0}\frac{\sum_{i=0}^{k-1} a_i n(n-1)\cdots(n-(k-i))y^k}{k!}.$$

Method 3

For $n, m \geq 0$ it is easy to see how the m-th power of $\frac{d}{dx}$ acts on $\ell_{-1}(x)^n$ because $\frac{d}{dx}\ell_{-1}(x)^n = n\ell_{-1}(x)^n$ is a monomial. So we have $\frac{d}{dx}^m \ell_{-1}(x)^n = n^m \ell_{-1}(x)$ which gives

$$\begin{aligned}e^{y\frac{d}{dx}}\ell_{-1}(x)^n &= \sum_{k\geq 0}\frac{(ny)^k}{k!}\ell_{-1}(x)^n \\ &= \ell_{-1}(x)^n e^{ny}.\end{aligned}$$

Method 4

We have for $n \in \mathbb{N}$

$$\begin{aligned}e^{y\frac{d}{dx}}\ell_{-1}(x)^n &= \left(e^{y\frac{d}{dx}}\ell_{-1}(x)\right)^n \\ &= (\ell_{-1}(x)e^y)^n \\ &= (\ell_{-1}(x) + \ell_{-1}(x)(e^y-1))^n \\ &= \sum_{k\geq 0}\binom{n}{k}\ell_{-1}(x)^{n-k}\ell_{-1}(x)^k(e^y-1)^k \\ &= \ell_{-1}(x)^n \sum_{k\geq 0}\binom{n}{k}(e^y-1)^k.\end{aligned}$$

Now by arguing as in Remark 2.1.2 and by using a formula like (2.1.2) to extend the definition of $\binom{n}{k}$, we get that the result extends for all $n \in \mathbb{C}$. We may also write

$$e^{y\frac{d}{dx}}\ell_{-1}(x)^n = \ell_{-1}(x)^n\,(e^y)^n$$

for all $n \in \mathbb{C}$ where $(e^y)^n$ is interpreted as $(1 + (e^y - 1))^n$ expanded using the binomial expansion convention.

2.4.3 Case $N = 1$, the case "$\log x$"

We would like to calculate the formal analytic expansion of $e^{y\frac{d}{dx}}\ell_1(x)^n$ for $n \in \mathbb{C}$.

Method 1

We consider that for $n, k \in \mathbb{C}$ we have

$$\frac{d}{dx}\ell_1(x)^n\ell_0(x)^k = n\ell_1(x)^{n-1}\ell_0(x)^{k-1} + k\ell_1(x)^n\ell_0(x)^{k-1}.$$

Therefore we have

$$e^{y\frac{d}{dx}}\ell_1(x)^n = \sum_{i\geq 0}\sum_{j=0}^{i} c(i,j)\ell_1(x)^{n-j}\ell_0(x)^{-i}\frac{y^i}{i!},$$

where

$$c(i,j) = -(i-1)c(i-1,j) + (n-(j-1))c(i-1,j-1)$$

with initial conditions

$$c(0,j) = \begin{cases} 1 & j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

We note that this recurrence does depend on n, even though our notation suppresses this.

It will be convenient to consider the recurrence given by

$$s(i,j) = (i-1)s(i-1,j) + s(i-1,j-1) \tag{2.4.4}$$

with initial conditions

$$s(0,j) = \begin{cases} 1 & j = 0 \\ 0 & \text{otherwise.} \end{cases}$$

One may easily verify, say by induction, that

$$c(i,j) = (-1)^{i-j}\binom{n}{j}j!s(i,j).$$

Thus

$$e^{y\frac{d}{dx}}\ell_1(x)^n = \sum_{i\geq 0}\sum_{j=0}^{i}(-1)^{i-j}\binom{n}{j}j!s(i,j)\ell_1(x)^{n-j}\ell_0(x)^{-i}\frac{y^i}{i!}.$$

Method 2

We note that for our purposes $\frac{d}{dx} = \frac{\partial}{\partial \ell_0(x)} + \ell_0(x)^{-1}\frac{\partial}{\partial \ell_1(x)}$ and that the m-th power for $m \geq 0$ may be normally ordered so that we have

$$\left(\frac{\partial}{\partial \ell_0(x)} + \ell_0(x)^{-1}\frac{\partial}{\partial \ell_1(x)}\right)^m = \ell_0(x)^{-m}\sum_{0\leq i\leq m} M(i)\left(\frac{\partial}{\partial \ell_1(x)}\right)^i$$

for some $M(i) \in \mathbb{C}$, which we leave undetermined. This follows because

$$\left[x^{-1}, \frac{d}{dx}\right] = x^{-2}$$

and because any normally ordered term with $\frac{\partial}{\partial \ell_0(x)}$ vanishes when it acts against $\ell_1(x)$. Then

$$\begin{aligned}\left(\frac{d}{dx}\right)^m \ell_1(x)^n &= \ell_0(x)^{-m}\sum_{0\leq i\leq m} M(i)\left(\frac{\partial}{\partial \ell_1(x)}\right)^i \ell_1(x)^n \\ &= \ell_0(x)^{-m}\sum_{0\leq i\leq m} M(i)n(n-1)\cdots(n-(i-1))\ell_1(x)^{n-i}\end{aligned}$$

giving in turn

$$e^{y\frac{d}{dx}}\ell_1(x)^n = \sum_{m\geq 0}\frac{y^m}{m!}\ell_0(x)^{-m}\sum_{0\leq i\leq m} M(i)n(n-1)\cdots(n-(i-1))\ell_1(x)^{n-i}.$$

Method 3

We closely follow the argument leading to (3.15) in [HLZ]. We have

$$\frac{d}{dx}\ell_0(x)^n\ell_1(x)^m = n\ell_0(x)^{n-1}\ell_1(x)^m + m\ell_0(x)^{n-1}\ell_1(x)^{m-1}.$$

Then define two linear operators T_0 and T_1 on $\mathbb{C}[\ell_0(x), \ell_1(x)]$ by

$$\begin{aligned}T_0\ell_0(x)^n\ell_1(x)^m &= n\ell_0(x)^{n-1}\ell_1(x)^m \\ T_1\ell_0(x)^n\ell_1(x)^m &= m\ell_0(x)^{n-1}\ell_1(x)^{m-1}.\end{aligned}$$

Then

$$\frac{d}{dx}^{k}\ell_0(x)^n\ell_1(x)^m = (T_0+T_1)^k\ell_0(x)^n\ell_1(x)^m.$$

It is not hard to see that

$$\begin{aligned}\left(\frac{d}{dx}\right)^k \ell_0(x)^n\ell_1(x)^m = &\sum_{j=0}^{k} m(m-1)\cdots(m-j+1)\cdot \\ &\cdot\left(\sum_{0\le t_1<t_2<\cdots<t_{k-j}<k}(n-t_1)\cdots(n-t_{k-j})\right)\ell_0(x)^{n-k}\ell_1(x)^{m-j},\end{aligned}$$

where the reader should think of j as corresponding to the number of T_1's in a summand of the expansion of $(T_0+T_1)^k$ and the t_i's as corresponding to the positions of the T_0's. It is now easy to get that

$$\begin{aligned}e^{y\frac{d}{dx}}\ell_1(x)^m = &\sum_{k\ge 0}\left(\frac{y}{\ell_0(x)}\right)^k\sum_{j=0}^{k}\binom{m}{j}\ell_1(x)^{m-j}\frac{j!}{k!}\cdot \\ &\cdot\left(\sum_{0\le t_1<t_2<\cdots<t_{k-j}<k}(-t_1)\cdots(-t_{k-j})\right). \qquad (2.4.5)\end{aligned}$$

Method 4

We also have $\left(\frac{d}{dx}\right)^l\ell_1(x) = (-1)^{l-1}(l-1)!\ell_0(x)^{-l}$ for $l\ge 1$. Thus it is easy to get that for $m\in\mathbb{N}$:

$$\begin{aligned}e^{y\frac{d}{dx}}\ell_1(x)^m &= \left(e^{y\frac{d}{dx}}\ell_1(x)\right)^m \\ &= \left(\ell_1(x)+\log\left(1+\frac{y}{\ell_0(x)}\right)\right)^m \\ &= \sum_{j\ge 0}\binom{m}{j}\ell_1(x)^{m-j}\left(\sum_{l\ge 1}\frac{(-1)^{l-1}}{l}\left(\frac{y}{\ell_0(x)}\right)^l\right)^j \\ &= \sum_{j\ge 0}\binom{m}{j}\ell_1(x)^{m-j}\sum_{k\ge j}(-1)^{k-j}\sum_{\substack{k_1+\cdots+k_j=k\\ k_i\ge 1}}\frac{1}{k_1\cdots k_j}\left(\frac{y}{\ell_0(x)}\right)^k, \qquad (2.4.6)\end{aligned}$$

where we recall that

$$\log(1+X) = \sum_{i\ge 1}\frac{(-1)^{i-1}}{i}X^i.$$

Then by arguing as in Remark 2.1.2 and by using a formula like (2.1.2) to extend the definition of $\binom{m}{j}$, we get that the result extends for all $m\in\mathbb{C}$.

2.4.4 Case $N=-2$, the case "exp exp x"

We would like to calculate the formal analytic expansion of $e^{y\frac{d}{dx}}\ell_{-2}(x)^n$ for $n\in\mathbb{C}$.

Method 1

We consider that for $n,k\in\mathbb{C}$ we have

$$\frac{d}{dx}\ell_{-2}(x)^n\ell_{-1}(x)^k = n\ell_{-2}(x)^n\ell_{-1}(x)^{k+1} + k\ell_{-2}(x)^n\ell_{-1}(x)^k.$$

Therefore we have

$$e^{y\frac{d}{dx}}\ell_{-2}(x)^n = \sum_{i\geq 0}\sum_{j=0}^{i} C(i,j)\ell_{-2}(x)^n\ell_{-1}(x)^j\frac{y^i}{i!},$$

where

$$C(i,j) = jC(i-1,j) + nC(i-1,j-1)$$

with initial conditions

$$C(0,j) = \begin{cases} 1 & j=0 \\ 0 & \text{otherwise.} \end{cases}$$

We note that this recurrence does depend on n, even though our notation suppresses this.

It will be convenient to consider the recurrence given by

$$S(i,j) = jS(i-1,j) + S(i-1,j-1) \tag{2.4.7}$$

with initial conditions

$$S(0,j) = \begin{cases} 1 & j=0 \\ 0 & \text{otherwise.} \end{cases}$$

One may easily verify, say by induction, that

$$C(i,j) = n^j S(i,j).$$

Thus

$$e^{y\frac{d}{dx}}\ell_{-2}(x)^n = \sum_{i\geq 0}\sum_{j=0}^{i} n^j S(i,j)\ell_{-2}(x)^n\ell_{-1}(x)^j\frac{y^i}{i!},$$

Method 2

We note that for our purposes $\frac{d}{dx} = \ell_{-1}(x)\frac{\partial}{\partial \ell_{-1}(x)} + \ell_{-1}(x)\ell_{-2}(x)\frac{\partial}{\partial \ell_{-2}(x)}$ and that the m-th power for $m \geq 0$ may be normally ordered so that we have

$$\left(\ell_{-1}(x)\frac{\partial}{\partial \ell_{-1}(x)} + \ell_{-1}(x)\ell_{-2}(x)\frac{\partial}{\partial \ell_{-2}(x)}\right)^m = \sum_{0\leq j\leq i\leq m} N(i,j)\ell_{-1}(x)^i\ell_{-2}(x)^j\left(\frac{\partial}{\partial \ell_{-2}(x)}\right)^j$$

for some $N(i,j) \in \mathbb{N}$, which we leave undetermined. This follows because

$$\left[\frac{d}{dx}, x\right] = 1$$

and because any normally ordered term with $\frac{\partial}{\partial \ell_{-1}(x)}$ vanishes when it acts against $\ell_{-2}(x)$.

Then

$$\begin{aligned}
\left(\frac{d}{dx}\right)^m \ell_{-2}(x)^n &= \sum_{0\leq j\leq i\leq m} N(i,j)\ell_{-1}(x)^i\ell_{-2}(x)^j\left(\frac{\partial}{\partial \ell_{-2}(x)}\right)^j \ell_{-2}(x)^n \\
&= \sum_{0\leq j\leq i\leq m} N(i,j)\ell_{-1}(x)^i\ell_{-2}(x)^j n(n-1)\cdots(n-(j-1))\ell_{-2}(x)^{n-j} \\
&= \ell_{-2}(x)^n \sum_{0\leq j\leq i\leq m} N(i,j)\ell_{-1}(x)^i n(n-1)\cdots(n-(j-1)),
\end{aligned}$$

giving in turn

$$e^{y\frac{d}{dx}}\ell_{-2}(x)^n = \ell_{-2}(x)^n\sum_{m\geq 0}\frac{y^m}{m!}\sum_{0\leq j\leq i\leq m} N(i,j)n(n-1)\cdots(n-(j-1))\ell_{-1}(x)^i.$$

Method 3

We have for $n, m \in \mathbb{C}$

$$\frac{d}{dx}\ell_{-1}(x)^n\ell_{-2}(x)^m = n\ell_{-1}(x)^n\ell_{-2}(x)^m + m\ell_{-1}(x)^{n+1}\ell_{-2}(x)^m.$$

Then define two linear operators S_0 and S_1 on $\mathbb{C}[\ell_{-1}(x), \ell_{-2}(x)]$ by

$$S_0\ell_{-1}(x)^n\ell_{-2}(x)^m = n\ell_{-1}(x)^n\ell_{-2}(x)^m$$

$$S_1\ell_{-1}(x)^n\ell_{-2}(x)^m = m\ell_{-1}(x)^{n+1}\ell_{-2}(x)^m.$$

Then

$$\left(\frac{d}{dx}\right)^k \ell_{-1}(x)^n\ell_{-2}(x)^m = (S_0+S_1)^k\ell_{-1}(x)^n\ell_{-2}(x)^m.$$

It is not hard to see that

$$\left(\frac{d}{dx}\right)^k \ell_{-1}(x)^n \ell_{-2}(x)^m = \sum_{j=0}^{k} m^j \left(\sum_{0 \le t_1 \le t_2 \le \cdots \le t_{k-j} \le j} (n+t_1)\cdots(n+t_{k-j}) \right) \cdot$$
$$\cdot\, \ell_{-1}(x)^{n+j} \ell_{-2}(x)^m,$$

where the reader should think of j as corresponding to the number of S_1's in a summand of the expansion of $(S_0+S_1)^k$ and the t_i's as corresponding to the positions of the S_0's. It is now easy to get that

$$e^{y\frac{d}{dx}} \ell_{-2}(x)^m = \ell_{-2}(x)^m \sum_{k \ge 0} \frac{y^k}{k!} \sum_{j=0}^{k} (m\ell_{-1}(x))^j \sum_{0 \le t_1 \le t_2 \le \cdots \le t_{k-j} \le j} t_1 \cdots t_{k-j}. \qquad (2.4.8)$$

Method 4

In order to proceed as in the previous three examples we would like to be able to easily calculate $e^{y\frac{d}{dx}}\ell_{-2}(x)$ and then take the m-th power of the result. However, it is just as difficult to calculate $e^{y\frac{d}{dx}}\ell_{-2}(x)$ as $e^{y\frac{d}{dx}}\ell_{-2}(x)^m$, since the answer already involves two variables non-trivially. We shall therefore use a different strategy which we could have used in place of method 4 in the case $N=-1$ and also in a sense cases $N=0,1$ although these cases are roughly like initial cases. We shall prove, in the next section, a recursive formula for $\ell_n(x+y)$ in terms of $\ell_{n+1}(x+y)$ (and inversely in terms of $\ell_{n-1}(x+y)$), which we will use to calculate $e^{y\frac{d}{dx}}\ell_{-2}(x)$.

2.5 A recursion

In this section we establish a recursive identity for $\ell_n(x+y)$ in terms of $\ell_{n-1}(x+y)$ (and inversely in terms of $\ell_{n+1}(x+y)$) for all $n \in \mathbb{Z}$. We shall use this recursion to calculate a formal analytic expansion of $\ell_n(x+y)$ for $n \ge 0$ in the next section.

Our approach is based on the following identity:

$$\lim_{r\to 0}\left(\left(\frac{d}{dx}\right)^m \frac{(\ell_n(x))^r}{r}\right) = \left(\frac{d}{dx}\right)^m \ell_{n+1}(x) \quad (m \ge 1). \qquad (2.5.9)$$

But we shall need to define just what we mean by taking a limit in this context in order for the above expression to make precise sense. We first define a new space.

Definition 2.5.1. We let $F(\mathbb{Z}_+, \ell)$ be the complex vector space of functions from the positive integers into $\mathbb{C}\{[\dots, \ell_{-1}(x), \ell_0(x), \ell_1(x), \dots]\}$.

We may define a "lifting" of $\frac{d}{dx}$ on $F(\mathbb{Z}_+, \ell)$.

Definition 2.5.2. For f and $g \in F(\mathbb{Z}_+, \ell)$, we say that $g = \frac{d}{dx} f$ when $g(r) = \frac{d}{dx} f(r)$ for all $r \geq 0$.

Of course, $\frac{d}{dx} f$ may not exist for all $f \in F(\mathbb{Z}_+, \ell)$. We shall actually be interested in a subspace of $F(\mathbb{Z}_+, \ell)$, which we call $P(\mathbb{Z}_+, \ell)$ on which $\frac{d}{dx} f$ does always exist.

Definition 2.5.3. We let $P(\mathbb{Z}_+, \ell) \subset F(\mathbb{Z}_+, \ell)$ be the space of functions $f(r)$, from the nonzero natural numbers into $\mathbb{C}\{[\ell_0(x), \ell_1(x), \dots]\}$, which may be represented in the form

$$f(r) = \sum_{j=0}^{J} q_j(r) \prod_{i \in \mathbb{Z}} \ell_i(x)^{p_{i,j}(r)},$$

where $q_j(r), p_{i,j}(r)$ are complex polynomials in r for all $j \geq 0$, $i \in \mathbb{Z}$ and where we further require that for all $j \geq 0$ there exists some $I_j \geq 0$ such that $p_{i,j}(r) = 0$ when $|i| \geq I_j$. We call such a representation a *formal polynomial form* of the function. The function is given by the obvious substitution procedure for r in the formal polynomial form.

Definition 2.5.4. For $f(r) \in P(\mathbb{Z}_+, \ell)$ we say a formal polynomial form representation,

$$f(r) = \sum_{j=0}^{J} q_j(r) \prod_{i \in \mathbb{Z}} \ell_i(x)^{p_{i,j}(r)},$$

is in *reduced formal polynomial form* or *reduced form,* when for all $j \neq k$, $j, k \geq 0$ there is some $i \in \mathbb{Z}$ such that

$$p_{i,j}(r) \neq p_{i,k}(r).$$

Then we get

Proposition 2.5.1. *If $f(r) \in P(\mathbb{Z}_+, \ell)$, then it is uniquely expressible in reduced formal polynomial form.*

Proof. Let us say that

$$M(r) = q(r) \prod_{i \in \mathbb{Z}} \ell_i(x)^{p_i(r)}$$

is a monomial summand in one reduced formal polynomial form of $f(r)$. Then consider any other reduced formal polynomial form expression for $f(r)$. Since two formally unequal complex polynomials can only agree for a finite number of substitution values, it is not difficult to see that there must be a monomial summand in the second reduced polynomial form of the form

$$N(r) = \bar{q}(r) \prod_{i \in \mathbb{Z}} \ell_i(x)^{p_i(r)}.$$

But since our forms are reduced, then in fact $N(r)$ is the only monomial summand of this form in the second representation, and therefore $q(r) = \bar{q}(r)$. The result now obviously follows by induction. $\square$

It is now easy to define what is meant by $\lim_{r \to 0} f(r)$ when $f(r) \in P(\mathbb{Z}_+, \ell)$. One simply expresses $f(r)$ in its unique reduced formal polynomial expansion and substitutes 0 for r to yield a well-defined element of $\mathbb{C}\{[\ldots \ell_{-1}(x), \ell_0(x), \ell_1(x), \ldots]\}$.

Remark 2.5.1. In order to justify our use of the term "limit" we note that this process is very similar to what we teach introductory calculus students. When finding the limit of a function: first find a nice form and then substitute.

Before we return to the identity which we want we should note that $P(\mathbb{Z}_+, \ell)$ is obviously closed under $\frac{d}{dx}$. It is also necessary to prove one lemma.

Lemma 2.5.1. *For any $A_r(x) \in P(\mathbb{Z}_+, \ell)$ we have that*

$$\lim_{r \to 0} \frac{d}{dx} A_r(x) = \frac{d}{dx} \lim_{r \to 0} A_r(x).$$

Proof. Since $\frac{d}{dx}$ is linear we only have to consider the case where $A_r(x)$ is a monomial. For convenience we call $\lim_{r \to 0} A_r(x) = A_0(x)$. Let $A_r(x) = B_r(x) C_r(x)$. then

$$\begin{aligned} \lim_{r \to 0} \frac{d}{dx} A_r(x) &= \lim_{r \to 0} \frac{d}{dx} (B_r(x) C_r(x)) \\ &= (\lim_{r \to 0} \frac{d}{dx} B_r(x)) C_0(x) + B_0(x) \lim_{r \to 0} \frac{d}{dx} (C_r(x)) \end{aligned}$$

and

$$\begin{aligned}\frac{d}{dx}\lim_{r\to 0} A_r(x) &= \frac{d}{dx}\lim_{r\to 0}(B_r(x)C_r(x)) \\ &= (\frac{d}{dx}B_0(x))C_0(x)) + B_0(x)\frac{d}{dx}C_0(x)),\end{aligned}$$

which means that we only need consider the case where $A_r(x) = p(r)\ell_i(x)^{q(r)}$ where $i \in \mathbb{Z}$. Now we get:

$$\begin{aligned}\lim_{r\to 0}\frac{d}{dx}A_r(x) &= \lim_{r\to 0} q(r)p(r)\ell_i(x)^{q(r)-1}\frac{d}{dx}\ell_i(x) \\ &= q(0)p(0)\ell_i(x)^{q(0)-1}\frac{d}{dx}\ell_i(x) \\ &= \frac{d}{dx}p(0)\ell_i(x)^{q(0)} \\ &= \frac{d}{dx}\lim_{r\to 0} A_r(x).\end{aligned}$$

□

We now prove the desired identity (2.5.9).

Lemma 2.5.2. *For $m \geq 1$ and $n \in \mathbb{Z}$*

$$lim_{r\to 0}\left(\left(\frac{d}{dx}\right)^m \frac{(\ell_n(x))^r}{r}\right) = \left(\frac{d}{dx}\right)^m \ell_{n+1}(x).$$

Proof. First note that $\frac{d}{dx}\frac{(\ell_n(x))^r}{r} \in P(\mathbb{Z}_+, \ell)$. Then we may calculate to get:

$$\begin{aligned}&\lim_{r\to 0}\left(\left(\frac{d}{dx}\right)^m \frac{(\ell_n(x))^r}{r}\right) = \left(\frac{d}{dx}\right)^{m-1} \lim_{r\to 0}\left(\frac{d}{dx}\frac{(\ell_n(x))^r}{r}\right) \\ &= \left(\frac{d}{dx}\right)^{m-1} \lim_{r\to 0}\left(\ell_n(x)^{r-1}\frac{d}{dx}\ell_n(x)\right) \\ &= \left(\frac{d}{dx}\right)^{m-1}\left(\ell_n(x)^{-1}\frac{d}{dx}\ell_n(x)\right).\end{aligned}$$

And now we proceed in the two separate cases $n \geq 0$ and $n \leq -1$. First, when $n \geq 0$ we have,

$$\begin{aligned}\lim_{r\to 0}\left(\left(\frac{d}{dx}\right)^m \frac{(\ell_n(x))^r}{r}\right) &= \left(\frac{d}{dx}\right)^{m-1}\left(\ell_n(x)^{-1}\prod_{i=0}^{n-1}(\ell_i(x))^{-1}\right) \\ &= \left(\frac{d}{dx}\right)^{m-1}\prod_{i=0}^{n}(\ell_i(x))^{-1} \\ &= \left(\frac{d}{dx}\right)^m \ell_{n+1}(x).\end{aligned}$$

And second, when $n \leq -1$ we have,

$$\lim_{r\to 0}\left(\left(\frac{d}{dx}\right)^m \frac{(\ell_n(x))^r}{r}\right) = \left(\frac{d}{dx}\right)^{m-1}\left(\ell_n(x)^{-1}\prod_{i=-1}^{n}\ell_i(x)\right)$$
$$= \left(\frac{d}{dx}\right)^{m-1}\left(\prod_{i=-1}^{n+1}\ell_i(x)\right)$$
$$= \left(\frac{d}{dx}\right)^{m}\ell_{n+1}(x).$$

□

With some care, we now see that for $n \in \mathbb{Z}$

$$\lim_{r\to 0}\left(e^{y\frac{d}{dx}}\left(\frac{(\ell_n(x))^r}{r}\right) - \frac{\ell_n(x)^r}{r}\right) = e^{y\frac{d}{dx}}\ell_{n+1}(x) - \ell_{n+1}(x).$$

One must note that indeed $e^{y\frac{d}{dx}}\left(\frac{(\ell_n(x))^r}{r}\right) - \frac{\ell_n(x)^r}{r} \in P(\mathbb{Z}_+, \ell)$ because the first term of $e^{y\frac{d}{dx}}\left(\frac{(\ell_n(x))^r}{r}\right)$ cancels $\frac{\ell_n(x)^r}{r}$. Next we get for $n \in \mathbb{Z}$

$$\ell_{n+1}(x+y) = \ell_{n+1}(x) + \lim_{r\to 0}\left(\frac{\ell_n(x+y)^r - \ell_n(x)^r}{r}\right).$$

But we don't want the limit in the expression, so, recalling that r stands for a positive integer, we get:

$$\ell_{n+1}(x+y) = \ell_{n+1}(x) + \lim_{r\to 0}\left(\frac{(\ell_n(x) + (\ell_n(x+y) - \ell_n(x)))^r - \ell_n(x)^r}{r}\right)$$
$$= \ell_{n+1}(x) + \lim_{r\to 0}\sum_{p\geq 1}\frac{r(r-1)\cdots(r-(p-1))}{rp!}\ell_n(x)^{r-p}\left(\ell_n(x+y) - \ell_n(x)\right)^p$$
$$= \ell_{n+1}(x) + \sum_{p\geq 1}\frac{(-1)^{p-1}}{p}\left(\frac{\ell_n(x+y) - \ell_n(x)}{\ell_n(x)}\right)^p$$
$$= \ell_{n+1}(x) + \log\left(1 + \left(\frac{\ell_n(x+y) - \ell_n(x)}{\ell_n(x)}\right)\right),$$

where for a formal object X

$$\log(1+X) = \sum_{i\geq 1}\frac{(-1)^{i-1}}{i}X^i,$$

whenever well defined.

Theorem 2.5.1. *For $n \in \mathbb{Z}$ we have*

$$\ell_{n+1}(x+y) = \ell_{n+1}(x) + \log\left(1 + \left(\frac{\ell_n(x+y) - \ell_n(x)}{\ell_n(x)}\right)\right). \tag{2.5.10}$$

□

We note that we may solve for $\ell_n(x+y)$ in (2.5.10) to get for all $n \in \mathbb{Z}$

$$\ell_n(x+y) = \ell_n(x) e^{(\ell_{n+1}(x+y) - \ell_{n+1}(x))}, \tag{2.5.11}$$

where we used that $e^{\log(1+X)} = 1 + X$, which is perhaps checked most easily by calculating that $\frac{d}{dx}\left(e^{\log(1+X)}(1+X)^{-1}\right) = 0$ which gives $e^{\log(1+X)} = c(1+X)$ for some constant, c which is, in turn, solved for by substituting 0 for X or, in other words, checking the constant term.

Remark 2.5.2. Both (2.5.10) and (2.5.11) make sense heuristically as may be seen easily, when one recalls that $\ell_n(x)$ is "really" an iterated logarithm or exponential.

2.6 Formal analytic expansions: warmup continued and completed

Using our results from above we now calculate, in one final way, a formal analytic expansion of $e^{y\frac{d}{dx}}\ell_{-2}(x)^m$. We shall begin with the subcase, $m = 1$. By (2.5.11) we have

$$\begin{aligned}
\ell_{-2}(x+y) &= \ell_{-2}(x) e^{(\ell_{-1}(x+y) - \ell_{-1}(x))} \\
&= \ell_{-2}(x) \sum_{j \geq 0} \frac{\ell_{-1}(x)^j (e^y - 1)^j}{j!} \\
&= \ell_{-2}(x) \sum_{j \geq 0} \frac{\ell_{-1}(x)^j}{j!} \left(\sum_{s \geq 1} \frac{y^s}{s!} \right)^j \\
&= \ell_{-2}(x) \sum_{k \geq j \geq 0} \frac{\ell_{-1}(x)^j}{j!} \sum_{s_1 + \cdots s_j = k} \frac{y^k}{s_1! s_2! \cdots s_j!} \\
&= \ell_{-2}(x) \sum_{k \geq 0} \frac{y^k}{k!} \sum_{j=0}^{k} \frac{k! \ell_{-1}(x)^j}{j!} \sum_{s_1 + \cdots s_j = k} \frac{1}{s_1! s_2! \cdots s_j!}.
\end{aligned}$$

It is essentially trivial to generalize to the case for all $m \geq 1$ to get

$$e^{y\frac{d}{dx}}\ell_{-2}(x)^m = \ell_{-2}(x+y)^m \tag{2.6.12}$$

$$= \ell_{-2}(x)^m e^{m(\ell_{-1}(x+y) - \ell_{-1}(x))} \tag{2.6.13}$$

$$= \ell_{-2}(x)^m \sum_{k \geq 0} \frac{y^k}{k!} \sum_{j=0}^{k} \frac{k! (m\ell_{-1}(x))^j}{j!} \sum_{s_1 + \cdots s_j = k} \frac{1}{s_1! s_2! \cdots s_j!}. \tag{2.6.14}$$

Remark 2.6.1. We use here that $e^{mx} = (e^x)^m$, which one can easily prove from scratch or, alternatively, observe by comparing the expansions for $e^{y\frac{d}{dx}}\ell_{-1}(x)^m$ obtained by methods 3 and 4.

2.7 Formal analytic expansions: Cases $N \geq 0$, "iterated logarithms"

2.7.1 Method 1

We note that for the purpose of acting on $\ell_N(x)^m$ for $N \geq 0$ we have

$$\frac{d}{dx} = \frac{\partial}{\partial \ell_0(x)} + \ell_0(x)^{-1}\frac{\partial}{\partial \ell_1(x)} + \ell_0(x)^{-1}\ell_1(x)^{-1}\frac{\partial}{\partial \ell_2(x)} + \cdots \tag{2.7.15}$$

$$= \sum_{i\geq 0}\prod_{j=0}^{i-1}\ell_j(x)^{-1}\frac{\partial}{\partial \ell_i(x)}. \tag{2.7.16}$$

This shows that the coefficients in the expansion of $e^{y\frac{d}{dx}}\ell_N(x)^r$ are given by a linear recursion equation in $N+1$ variables. Indeed it is not hard to see that

$$\begin{aligned}\frac{d}{dx}^k \ell_N(x)^r = &\sum_{1\leq j_N\leq j_{N-1}\leq\cdots\leq j_1\leq j_0=k} B(j_N, j_{N-1}, \cdots, j_0)\cdot\\ &\cdot \ell_N(x)^{r-j_N}\ell_{N-1}(x)^{-j_{N-1}}\cdots\ell_0(x)^{-j_0},\end{aligned}$$

where

$$\begin{aligned}B(j_N, j_{N-1}, \cdots, j_0) = &(r-(j_N-1))B(j_N-1, j_{N-1}-1, \cdots, j_0-1)\\ &+(1-j_{N-1})B(j_N, j_{N-1}-1, \cdots, j_0-1)\\ &\vdots\\ &+(1-j_0)B(j_N, j_{N-1}, \cdots j_1, j_0-1),\end{aligned}$$

along with the initial conditions,

$$B(j_N, j_{N-1}, \cdots, j_1, 1) = \begin{cases} r & j_N = j_{N-1} = \cdots = j_1 = 1\\ 0 & \text{otherwise.}\end{cases}$$

It will be convenient, and is easy, to partially solve for $B(j_N, j_{N-1}, \cdots, j_0)$. Let $S(j_N, j_{N-1}, \cdots, j_0)$ be given by the following recursion:

$$\begin{aligned}S(j_N, j_{N-1}, \cdots, j_0) = {} & S(j_N - 1, j_{N-1} - 1, \cdots, j_0 - 1) \\ & + (j_{N-1} - 1) S(j_N, j_{N-1} - 1, \cdots, j_0 - 1) \\ & \vdots \\ & + (j_0 - 1) S(j_N, j_{N-1}, \cdots j_1, j_0 - 1),\end{aligned}$$

along with the initial conditions,

$$S(j_N, j_{N-1}, \cdots, j_1, 1) = \begin{cases} 1 & j_N = j_{N-1} = \cdots = j_1 = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then it is easy to verify (by induction) that

$$B(j_N, j_{N-1}, \cdots, j_1, j_0) = j_N! \binom{r}{j_N} (-1)^{j_0 + j_N} S(j_N, j_{N-1}, \cdots, j_1, j_0).$$

This gives us

Proposition 2.7.1. *For $r \in \mathbb{C}$,*

$$\begin{aligned}e^{y\frac{d}{dx}} \ell_N(x)^r &= \ell_N(x+y)^r \\ &= \sum_{k \geq 0} \frac{y^k}{k!} \sum_{1 \leq j_N \leq \cdots \leq j_1 \leq j_0 = k} j_N! \binom{r}{j_N} (-1)^{j_0 + j_N} S(j_N, \cdots, j_0) \cdot \\ &\quad \cdot \ell_N(x)^{r - j_N} \ell_{N-1}^{-j_{N-1}} \cdots \ell_0^{-k}.\end{aligned}$$

2.7.2 Method 2

We again note that the purpose of acting on $\ell_N(x)^m$ for $N \geq 0$ we have

$$\frac{d}{dx} = \frac{\partial}{\partial \ell_0(x)} + \ell_0(x)^{-1} \frac{\partial}{\partial \ell_1(x)} + \ell_0(x)^{-1} \ell_1(x)^{-1} \frac{\partial}{\partial \ell_2(x)} + \cdots \tag{2.7.17}$$

$$= \sum_{i \geq 0} \prod_{j=0}^{i-1} \ell_j(x)^{-1} \frac{\partial}{\partial \ell_i(x)} \tag{2.7.18}$$

and, in fact, we may further truncate the sum at $i = N$. The m-th power for $m \geq 0$ may be normally ordered giving

$$\begin{aligned}\left(\frac{d}{dx}\right)^m = \sum_{j_0 + j_1 + \cdots j_N = m} & \bar{M}(j_0, \ldots, j_N) \ell_0(x)^{-(j_1 + j_2 + \cdots + j_N) - j_0} \ell_1(x)^{-(j_2 + j_3 + \cdots + j_N) - j_1} \cdots \\ & \cdots \ell_{N-1}(x)^{-(j_N) - j_{N-1}} \left(\frac{\partial}{\partial \ell_N(x)}\right)^{j_N},\end{aligned}$$

where j_i corresponds to the number of occurrences of the term containing $\frac{\partial}{\partial \ell_i(x)}$ in a generic monomial in the expansion prior to performing any commutations.

For notational convenience we let

$$\alpha_i = j_i + \cdots + j_N,$$

for $0 \leq i \leq N$.

Then we have

$$\left(\frac{d}{dx}\right)^m = \sum_{0 \leq \alpha_N \leq \cdots \leq \alpha_1 \leq \alpha_0 = N} M(\alpha_0, \ldots, \alpha_N) \cdot \\ \cdot\, \ell_0(x)^{-\alpha_0} \ell_1(x)^{-\alpha_1} \cdots \ell_{N-1}(x)^{-\alpha_{N-1}} \left(\frac{\partial}{\partial \ell_N(x)}\right)^{\alpha_N}.$$

2.7.3 Method 3

We begin by defining the following operators:

$$T_i = \prod_{j=0}^{i-1} \ell_j(x)^{-1} \frac{\partial}{\partial \ell_i(x)} \qquad i \geq 0.$$

Then

$$\frac{d}{dx} = \sum_{i \geq 0} T_i.$$

Therefore,

$$\left(\frac{d}{dx}\right)^k = \sum_{i_1, i_2, \cdots, i_k \geq 0} T_{i_1} T_{i_2} \cdots T_{i_k}.$$

If we consider all the monomials with a fixed number of occurrences of each T_i, and call this fixed number j_i, then we can partially calculate to get

$$D^k \prod_{l=0}^{N} (\ell_l(x))^{c_l} = \sum_{\substack{j_0 + j_1 + \cdots + j_N = k \\ 0 \leq j_0, j_1, \cdots, j_N}} P(c) \prod_{i=0}^{N} \ell_i(x)^{c_i - \alpha_i},$$

where $P(c)$ is a certain sum of polynomials in the c_i each of degree j_i in c_i, and where, for notational convenience, we have let

$$\alpha_i = \sum_{l=i}^{N} j_l.$$

We shall describe $P(c)$ by using a combinatorial construction, a type of tableau. A tableau will consist of a specified number of columns of blank entries each of a specified length. We shall construct a tableau on any such "grid" of blanks by filling in each blank with nonnegative numbers beginning at the top of each column and moving down. Each new entry can be any non-negative number subject to two restrictions. First, the numbers must strictly ascend as one descends a column and second the entry in each column must be less than or equal to the number of entries above and to the right (not necessarily above). So for example,

$$\begin{array}{ccc} & & 0 \\ 0 & & 1 \\ 2 & 3 & 2 \\ 8 & 5 & 3 \\ 11 & 6 & 4 \end{array}$$

is a tableaux. But

$$\begin{array}{ccc} & & 0 \\ 0 & & 1 \\ 8 & 3 & 2 \\ 2 & 5 & 3 \\ 11 & 6 & 4 \end{array} \qquad \text{and} \qquad \begin{array}{ccc} & & 0 \\ 0 & & 1 \\ 2 & 3 & 2 \\ 8 & 5 & 3 \\ 11 & 20 & 4 \end{array}$$

are not.

We shall consider all tableaux of a particular shape and assign to that shape a polynomial in as many variables, x_i, as there are columns. We shall denote this polynomial by $[m_1, m_2, \cdots, m_n](x_i)$ where the shape is n columns of heights m_1 on the left followed by m_2 next to the right etc. The polynomial is found by summing over all the tableau of the given shape. Each summand is found by inserting "x_i-" in each entry of the i-th column and multiplying all entries. A moment's reflection yields that

$$D^k \prod_{l=0}^{N} (\ell_l(x))^{c_l} = \sum_{\substack{j_0+j_1+\cdots+j_N=k \\ 0\le j_0,j_1,\cdots,j_N}} [j_0, j_1, \cdots, j_n](c_i) \prod_{i=0}^{N} \ell_i(x)^{c_i-\alpha_i}. \tag{2.7.19}$$

We shall specialize this calculation, but shall first revisit the tableaux. Notice that the rightmost column of a tableau is completely determined by its length alone. Therefore the piece of the tableau polynomial due to the rightmost column may be factored out. If look at the remaining factor of $[m_1, m_2, \cdots, m_n](x_i)$ and set all the variables to zero we get an integer which we shall call $(m_1, m_2, \cdots, m_{n-1}; m_n)$. Notice that when constructing a tableau of fixed shape each column is labelled independently of the others. Thus we have

Proposition 2.7.2.

$$(m_1, m_2, \cdots m_{n-1}; m_n) = (m_1; m_2 + \cdots m_n) \cdots (m_{n-2}; m_{n-1} + m_n)(m_{n-1}; m_n).$$

It is easy to see that

$$(m; n) = (-1)^m \sum_{0 \le i_1 < i_2 < \cdots < i_m \le m+n-1} i_1 i_2 \cdots i_m.$$

Remark 2.7.1. The reader may now recognize a formula for the Stirling numbers of the first kind. Indeed we have

$$(m; n) = (-1)^m \begin{bmatrix} m+n \\ n \end{bmatrix},$$

a fact which will actually be reproven below, but for ease of notation we will anticipate the result.

Now we can specialize (2.7.19) to get

$$\begin{aligned} D^k \ell_n(x)^r &= \sum_{\substack{j_0+j_1+\cdots+j_n=k \\ 0 \le j_0, j_1, \cdots, j_n}} (r)(r-1)\cdots(r-(j_n-1))(j_0, \cdots j_{n-1}; j_n)\ell_n(x)^r \prod_{i=0}^{n} \ell_i(x)^{-\alpha_i} \\ &= \sum_{\substack{j_0+j_1+\cdots+j_n=k \\ 0 \le j_0, j_1, \cdots, j_n}} (r)(r-1)\cdots(r-(j_n-1))(-1)^{\alpha_0 - \alpha_n} \cdot \\ &\quad \cdot \prod_{i=0}^{n-1} \begin{bmatrix} j_i + \alpha_{i+1} \\ \alpha_{i+1} \end{bmatrix} \ell_n(x)^r \prod_{i=0}^{n} \ell_i(x)^{-\alpha_i}. \end{aligned}$$

Thus we get

Proposition 2.7.3. *For $r \in \mathbb{C}$,*

$$e^{yD}\ell_n(x)^r = \sum_{k\geq 0}\frac{y^k}{k!}\sum_{\substack{j_0+j_1+\cdots+j_n=k\\ 0\leq j_0,j_1,\cdots,j_n}} j_n!\binom{r}{j_n}(-1)^{\alpha_0-\alpha_n}\prod_{i=0}^{n-1}\begin{bmatrix} j_i+\alpha_{i+1}\\ \alpha_{i+1}\end{bmatrix}\ell_n(x)^r\prod_{i=0}^{n}\ell_i(x)^{-\alpha_i}.$$

2.7.4 Method 4

Letting $r \in \mathbb{N}$ (and noting the remark below to explain notation), we can calculate to get:

$$\begin{aligned}
\ell_n(x+y)^r &= \left(\ell_n(x)+\log\left(1+\left(\frac{\ell_{n-1}(x+y)-\ell_{n-1}(x)}{\ell_{n-1}(x)}\right)\right)\right)^r\\
&= \left(\ell_n(x)+\sum_{m\geq 1}\frac{(-1)^{m-1}}{m}\left(\frac{\ell_{n-1}(x+y)-\ell_{n-1}(x)}{\ell_{n-1}(x)}\right)^m\right)^r\\
&= \sum_{j\geq 0}\binom{r}{j}\ell_n(x)^{r-j}\left(\sum_{m\geq 1}\frac{(-1)^{m-1}}{m}\left(\frac{\ell_{n-1}(x+y)-\ell_{n-1}(x)}{\ell_{n-1}(x)}\right)^m\right)^j\\
&= \sum_{j\geq 0}\binom{r}{j}\ell_n(x)^{r-j}\sum_{k\geq 0}\frac{j!}{k!}\begin{bmatrix} k\\ j\end{bmatrix}(-1)^{k-j}\left(\frac{\ell_{n-1}(x+y)-\ell_{n-1}(x)}{\ell_{n-1}(x)}\right)^k\\
&= \sum_{k,j\geq 0}\binom{r}{j}\ell_n(x)^{r-j}\frac{j!}{k!}\begin{bmatrix} k\\ j\end{bmatrix}(-1)^{k-j}\ell_{n-1}(x)^{-k}\cdot\\
&\quad\cdot\sum_{l\geq 0}\binom{k}{l}\ell_{n-1}(x+y)^l(-\ell_{n-1}(x))^{k-l}\\
&= \sum_{k,j,l,\geq 0}\binom{r}{j}\frac{j!}{k!}\begin{bmatrix} k\\ j\end{bmatrix}\binom{k}{l}(-1)^{j+l}\ell_n(x)^{r-j}\ell_{n-1}(x)^{-l}\ell_{n-1}(x+y)^l.
\end{aligned}$$

where

$$\begin{bmatrix} k\\ j\end{bmatrix} = \frac{k!}{j!}\sum_{\substack{i_1+\cdots+i_j=k\\ i_l\geq 1}}\frac{1}{i_1\cdots i_j}. \tag{2.7.20}$$

Remark 2.7.2. The reader may notice that 2.7.20 is a formula for the unsigned Stirling numbers of the first kind, thereby justifying our notation. Indeed, a specialization of a formula we shall get below will relate this expression to the standard recursion for the Stirling numbers of the first kind thus reproving this fact.

Remark 2.7.3. Compare with the calculation leading to (3.16) in [HLZ].

Rewriting

$$\ell_n(x+y)^{r_n} = \sum_{k_{n-1},j_n,r_{n-1}\geq 0} \binom{r_n}{j_n}\binom{k_{n-1}}{r_{n-1}}\frac{j_n!}{k_{n-1}!}\begin{bmatrix} k_{n-1} \\ j_n \end{bmatrix}(-1)^{j_n+r_{n-1}} \cdot \ell_n(x)^{r_n-j_n}\ell_{n-1}(x)^{-r_{n-1}}\ell_{n-1}(x+y)^{r_{n-1}}$$

makes clear how, with a bit of care, we can iterate to get

$$\begin{aligned}
\ell_n(x+y)^{r_n} &= \sum_{\substack{k_i\geq r_i\geq j_i\geq 0\\ 0\leq i\leq n-1, j_n\geq 0}} \prod_{i=0}^{n-1}\left(\binom{r_i}{j_i}\binom{k_i}{r_i}\frac{j_i!}{k_i!}\begin{bmatrix} k_i \\ j_{i+1}\end{bmatrix}\right)(-1)^{\sum_{i=1}^{n}(j_i+r_{i-1})}\frac{j_n!}{j_0!}\cdot \\
&\quad \cdot\binom{r_n}{j_n}\ell_n(x)^{r_n}\prod_{i=0}^{n}\ell_i(x)^{-j_i}y^{j_0} \\
&= \sum_{\substack{k_i\geq j_i\geq 0\\ 0\leq i\leq n-1, j_n\geq 0}} \prod_{i=0}^{n-1}\left(\frac{j_i!}{k_i!}\begin{bmatrix} k_i \\ j_{i+1}\end{bmatrix}\right)(-1)^{\sum_{i=1}^{n} j_i}\frac{j_n!}{j_0!}\cdot \\
&\quad \cdot\binom{r_n}{j_n}\ell_n(x)^{r_n}\prod_{i=0}^{n}\ell_i(x)^{-j_i}y^{j_0} \\
&\quad \cdot\sum_{k_i\geq r_i\geq j_i}(-1)^{\sum_{i=1}^{n} r_{i-1}}\prod_{i=0}^{n-1}\binom{r_i}{j_i}\binom{k_i}{r_i}.
\end{aligned}$$

Now it is useful to consider the special case where $n = 1$, because in this case we symbolically expanded

$$\left(\frac{\ell_{n-1}(x+y)-\ell_{n-1}(x)}{\ell_{n-1}(x)}\right)^k.$$

But when $n = 1$ this is a trivial expansion. Therefore it will be useful to perform the following trivial expansion to get the easy combinatorial identity which results. We have

$$\begin{aligned}\left(\frac{y}{x}\right)^k &= \left(\frac{(x+y)-x}{x}\right)^k \\ &= x^{-k}\sum_{p\geq 0}\binom{k}{p}(x+y)^p(-x)^{m-p} \\ &= x^{-k}\sum_{p,q\geq 0}\binom{k}{p}\binom{p}{q}(-1)^{k-p}x^{p-q}y^q x^{k-p} \\ &= \sum_{p,q\geq 0}\binom{k}{p}\binom{p}{q}(-1)^{k-p}\left(\frac{y}{x}\right)^q,\end{aligned}$$

which implies

$$\sum_{p\geq 0}\binom{k}{p}\binom{p}{q}(-1)^{k-p} = \begin{cases} 0 & q \neq k \\ 1 & q = k. \end{cases}$$

This implies:

Proposition 2.7.4. *Let $r \in \mathbb{N}$, then*

$$\ell_n(x+y)^r = \sum_{j_0,\cdots,j_n\geq 0}\prod_{i=0}^{n-1}\left(\begin{bmatrix} j_i \\ j_{i+1}\end{bmatrix}\right)(-1)^{j_0+j_n}\frac{j_n!}{j_0!}\binom{r}{j_n}\ell_n(x)^r\prod_{i=0}^{n}\ell_i(x)^{-j_i}y^{j_0}.$$

Remark 2.7.4. The same type of argument used in Remark 2.1.2 shows that we may extend the above proposition to $r \in \mathbb{C}$.

2.8 Some combinatorics

We have calculated, by various methods certain formal analytic expansions. It is now routine to equate the coefficients of different expansions of the same object. Doing this one finds various classical identities involving Stirling numbers of the first and second kinds, and, in the case of the iterated logarithmic objects, perhaps some extensions of the usual identities. One of these identities, involving Stirling numbers of the first kind, was re-discovered in [HLZ] and was one of the motivations for this work. We shall mention only a couple of these identities here as an illustration. We leave the rest to the reader.

Equating the coefficients of (2.4.5) and (2.4.6) found by methods 3 and 4 of expanding $e^{y\frac{d}{dx}}\ell_1(x)^n$, we find

$$\sum_{0\leq t_1<t_2<\cdots<t_{k-j}<k} t_1\cdots t_{k-j} = \sum_{\substack{k_1+\cdots+k_j=k\\ k_i\geq 1}} \frac{1}{k_1\cdots k_j}. \tag{2.8.21}$$

Remark 2.8.1. The recurrence given by (2.4.4), the well-known recurrence for Stirling numbers of the first kind (unsigned), used for method 1 of expanding $e^{y\frac{d}{dx}}\ell_1(x)^n$ clearly shows how (2.8.21) is giving two expansions of the Stirling numbers of the first kind.

Remark 2.8.2. The identity (2.8.21) appeared in Section 3 of [HLZ] in the course of a "traditional-style" algebraic proof of a logarithmic formal Taylor theorem. See also Remark 3.8 in [HLZ], where this identity was observed to solve a problem posed by D. Lubell in the Problems and Solutions section of the American Mathematical Monthly [Lub].

In parallel fashion, equating the coefficients of (2.4.8) and (2.6.12) found by methods 3 and 4 of expanding $e^{y\frac{d}{dx}}\ell_{-2}(x)^n$, we find

$$\sum_{0\leq t_1\leq t_2\leq\cdots\leq t_{k-j}\leq j} t_1\cdots t_{k-j} = \frac{k!}{j!}\sum_{\substack{s_1+\cdots s_j=k\\ s_i\geq 1}} \frac{1}{s_1!s_2!\cdots s_j!}. \tag{2.8.22}$$

Remark 2.8.3. The recurrence given by (2.4.7), the well-known recurrence for Stirling numbers of the second kind, used for method 1 of expanding $e^{y\frac{d}{dx}}\ell_{-2}(x)^n$ clearly shows how (2.8.22) is giving two expansions of the Stirling numbers of the second kind.

Remark 2.8.4. Readers familiar with the Stirling numbers of the first and second kinds will note that the exponential generating functions of these numbers give a clue as to why they appear in the expansions we have considered.

2.9 Substitution maps for iterated logs and exponentials

Now consider the substitution map

$$\phi : \mathbb{C}[\ell^{\pm 1}] \to \mathbb{C}[\ell^{\pm 1}]$$

and its inverse defined by

$$\phi(\ell_n(x)) = \ell_{n+1}(x) \qquad \text{for} \quad n \in \mathbb{Z}$$
$$(\text{and} \qquad \phi^{-1}(\ell_n(x)) = \ell_{n-1}(x) \qquad \text{for} \quad n \in \mathbb{Z}).$$

Proposition 2.9.1. *We have*

$$\phi \circ \frac{d}{dx} = \ell_0(x)\frac{d}{dx} \circ \phi$$
$$\text{and} \qquad \phi^{-1} \circ \ell_0(x)\frac{d}{dx} = \frac{d}{dx} \circ \phi^{-1}.$$

This proposition makes clear that, on the appropriate space, $e^{y\frac{d}{dx}}$ and $e^{y\ell_0(x)\frac{d}{dx}}$ are simply shifted (in terms of the subscripts of $\ell_n(x)$) versions of each other.

Proof. Since $\frac{d}{dx}$ and $\ell_0(x)\frac{d}{dx}$ are derivations we need only check the action on $\ell_n(x)$ $n \in \mathbb{Z}$. The verification is routine calculation. For instance:

For $n > 1$

$$\begin{aligned} \ell_0(x)\frac{d}{dx}\phi\ell_{-n}(x) = \ell_0(x)\frac{d}{dx}\ell_{-n+1}(x) &= \ell_0(x)\prod_{i=-1}^{-n+1}\ell_i(x) \\ &= \prod_{i=0}^{-n+1}\ell_i(x) \\ &= \phi\prod_{i=-1}^{-n}\ell_i(x) \\ &= \phi\frac{d}{dx}\ell_{-n}(x). \end{aligned}$$

□

We then have the following two examples of the "lifting" process referred to in the introduction to this section:

$$\begin{aligned} e^{y\ell_0(x)\frac{d}{dx}}\ell_0(x) &= \phi \circ e^{y\frac{d}{dx}}\phi^{-1}(\ell_0(x)) \\ &= \phi \circ e^{y\frac{d}{dx}}\ell_{-1}(x) \\ &= \phi \circ \sum_{n\geq 0}\frac{y^n}{n!}\ell_{-1}(x) \\ &= \ell_0(x)e^y, \end{aligned}$$

and

$$
\begin{aligned}
e^{y\ell_0(x)\frac{d}{dx}}\ell_1(x) &= \phi \circ e^{y\frac{d}{dx}}\phi^{-1}(\ell_1(x)) \\
&= \phi \circ e^{y\frac{d}{dx}}\ell_0(x) \\
&= \phi(\ell_0(x) + y) \\
&= \ell_1(x) + y,
\end{aligned}
$$

which translate respectively to the following identities in more standard logarithmic notation:

$$
\begin{aligned}
e^{yx\frac{d}{dx}}x &= xe^y \\
e^{yx\frac{d}{dx}}\log x &= \log x + y.
\end{aligned}
$$

Remark 2.9.1. Of course, these examples can be obtained much more easily without resorting to this method but in more involved examples this approach is very useful as we shall see.

Remark 2.9.2. Although we do not give a precise definition here, it is maps like ϕ that we call intertwining substitutions.

2.10 A glimpse of Faà di Bruno and umbral calculus

There are some interesting variants of the notion of intertwining substitution. In fact, one such variant appears implicitly in the proof of Faà di Bruno's formula in Proposition 8.3.4 of [FLM2], an argument which is essentially the basis for proving the (highly-nontrivial) "associativity" property of lattice vertex operator algebras in a setting based on arbitrary rational lattices; see Sections 8.3 and 8.4 of [FLM2]. We present a special case of this argument next.

Let x, y, z be formal commuting variables. Let $f(x), g(x) \in \mathbb{C}[x]$. Then

$$\begin{aligned}
e^{y\frac{d}{dx}} f(g(x)) &= f(g(x+y)) \\
&= f(g(x) + (g(x+y) - g(x))) \\
&= e^{(g(x+y)-g(x))\frac{d}{dz}} f(z)|_{z=g(x)} \\
&= \sum_{n\geq 0} \frac{f^{(n)}(z)(g(x+y)-g(x))^n}{n!}|_{z=g(x)} \\
&= \sum_{n\geq 0} \frac{f^{(n)}(g(x))(g(x+y)-g(x))^n}{n!} \\
&= \sum_{n\geq 0} \frac{f^{(n)}(g(x))\left(e^{y\frac{d}{dx}}g(x) - g(x)\right)^n}{n!} \\
&= \sum_{n\geq 0} \frac{f^{(n)}(g(x))\left(\sum_{m\geq 1}\frac{y^m g^{(m)}(x)}{m!}\right)^n}{n!}.
\end{aligned} \tag{2.10.23}$$

Motivated by this, we consider the algebra $\mathbb{C}[y_0, y_1, y_2, \ldots, x_1, x_2, \ldots]$ where y_i, x_j for $i \geq 0$ and $j \geq 1$ are commuting formal variables. Let D be the unique derivation on $\mathbb{C}[y_0, y_1, y_2, \ldots, x_1, x_2, \ldots]$ satisfying the following:

$$\begin{aligned}
Dy_i &= y_{i+1}x_1 \qquad && i \geq 0 \\
Dx_j &= x_{j+1} \qquad && j \geq 1.
\end{aligned}$$

Then this question of calculating $e^{y\frac{d}{dx}} f(g(x))$ is seen to be essentially equivalent to calculating

$$e^{zD} y_0,$$

where we "secretly," loosely speaking, identify $\frac{d}{dx}$ with D, $f^{(n)}(g(x))$ with y_n and $g^{(m)}(x)$ with x_m (and y with z). The reader may note that we are now really dealing with, among other things, a certain sort of completion of the original problem, so that one may, for instance, wish to view $f(x)$ as a formal power series and $g(x)$ as a formal power series with zero constant term, and indeed we note that it was in this generality (and with even more general derivations) that the above argument was carried out in

[FLM2]. For a detailed description of this material, we refer the reader to [R3] and Chapter 3 of this work.

Before proceeding, we note that we may write an intermediate step of (2.10.23) as

$$e^{y\frac{d}{dx}}\phi(f(z)) = \phi\left(e^{(g(x+y)-g(x))\frac{d}{dz}}f(z)\right),$$

where $\phi : \mathbb{C}[z] \to \mathbb{C}[x]$ substitutes $g(x)$ for z. That is, we have a commutative diagram:

$$\begin{array}{ccc}
\mathbb{C}[z] & \xrightarrow{e^{(g(x+y)-g(x))\frac{d}{dz}}} & \mathbb{C}[x,y,z] \\
\downarrow \phi & & \downarrow \phi \\
\mathbb{C}[x] & \xrightarrow{e^{y\frac{d}{dx}}} & \mathbb{C}[x,y].
\end{array}$$

This commutative diagram shows how ϕ may be regarded as a sort of "global (i.e., exponentiated) intertwining substitution." In our new setting we might consider looking at "nonglobal" intertwining substitution of D. This turns out to be too restrictive, but a suitably loosened version of this question turns out to lead to interesting results.

Let ϕ_B be the substitution which sends y_j to 1 for $j \geq 0$ and sends x_i to xB_i for $i \geq 1$, where $B_i \in \mathbb{C}$ for $i \geq 1$ is a fixed, arbitrary sequence subject to the requirement that $B_1 \neq 0$.

Proposition 2.10.1. *There is a unique linear map $D_B : \mathbb{C}[x] \to \mathbb{C}[x]$ which satisfies the condition*

$$D_B^n \phi_B(y_0) = \phi_B(D^n(y_0)) \qquad n \geq 0. \tag{2.10.24}$$

Proof. It is easy to see that $\phi_B D^n(y_0)$ is a polynomial of degree n whose leading term is $B_1^n x^n$, where we recall that this coefficient is nonzero. Thus each required equality in turn (indexing by n) may be solved to obtain an equation of the form $D_B x^{n-1} = r(x)$, where $r(x)$ is a polynomial of degree n. Of course this recursive process solves for and completely determines $D_B x^n$ for all $n \geq 0$. $\square$

The maps D_B are what have been called umbral shifts, as in [Rm1]. For more on the connection to classical umbral calculus see [R3] and Chapter 3 of this work.

Chapter 3

Formal calculus and umbral calculus

3.1 Preliminaries

We set up some notation and recall some well-known and easy preliminary propositions in this section. For a more complete treatment, we refer the reader to the first three sections of Chapter 8 of [FLM2] (cf. Chapter 2 of [LL]), while noting that in this chapter we shall not need any of the material on "expansions of zero," the heart of the formal calculus treated in those works.

We shall write $t, u, v, w, x, y, z, x_n, y_m, z_n$ for commuting formal variables, where $n \geq 0$ and $m \in \mathbb{Z}$. All vector spaces will be over $\mathbb{C}$. Let V be a vector space. We use the following:

$$\mathbb{C}[[x]] = \left\{ \sum_{n\geq 0} c_n x^n | c_n \in \mathbb{C} \right\}$$

(formal power series), and

$$\mathbb{C}[x] = \left\{ \sum_{n\geq 0} c_n x^n | c_n \in \mathbb{C}, c_n = 0 \text{ for all but finitely many } n \right\}$$

(formal polynomials).

We denote by $\frac{d}{dx}$ the formal derivative acting on either $\mathbb{C}[x]$ or $\mathbb{C}[[x]]$. Further, we shall frequently use the notation $e^{\square}$ to refer to the formal exponential expansion, where $\square$ is any formal object for which such expansion makes sense. For instance, we have the linear operator $e^{w\frac{d}{dx}} : \mathbb{C}[[x, x^{-1}]] \to \mathbb{C}[[x, x^{-1}]][[w]]$:

$$e^{w\frac{d}{dx}} = \sum_{n\geq 0} \frac{w^n}{n!} \left(\frac{d}{dx}\right)^n.$$

We recall that a linear map D on an algebra A which satisfies

$$D(ab) = (Da)b + a(Db) \quad \text{for all} \quad a, b \in A$$

is called a *derivation*. Of course, the linear operator $\frac{d}{dx}$ when acting on either $\mathbb{C}[x]$ or $\mathbb{C}[[x]]$ is an example of a derivation.

It is a simple matter to verify, by induction for instance, the following "version" of the elementary binomial theorem.

Proposition 3.1.1. *Let A be an algebra with derivation D. Then for all a, $b \in A$, we have:*

$$1) \qquad D^n ab = \sum_{k+l=n} (k+l)! \frac{D^k a}{k!} \frac{D^l b}{l!}$$

$$2) \qquad e^{wD} ab = \left(e^{wD} a\right) \left(e^{wD} b\right).$$

□

We call part 2 of Proposition 3.1.1 the "automorphism property." Further, we separately state the following important special case of the automorphism property.

Corollary 3.1.1. *Let $f(x), g(x) \in \mathbb{C}[[x]]$. We have*

$$e^{w\frac{d}{dx}} f(x)g(x) = \left(e^{w\frac{d}{dx}} f(x)\right) \left(e^{w\frac{d}{dx}} g(x)\right).$$

□

The automorphism property shows how the operator $e^{w\frac{d}{dx}}$ may be regarded as a formal substitution, since, for $n \geq 0$, we have:

$$e^{w\frac{d}{dx}} x^n = \left(e^{w\frac{d}{dx}} x\right)^n = (x+w)^n.$$

Therefore, by linearity, we get the following polynomial formal Taylor theorem.

Proposition 3.1.2. *Let $p(x) \in \mathbb{C}[x]$. We have*

$$e^{w\frac{d}{dx}} p(x) = p(x+w).$$

□

Since the total degree of every term in $(x+w)^n$ is n, we see that $e^{w\frac{d}{dx}}$ preserves total degree. By equating terms with the same total degree we can therefore extend the previous proposition to get the following.

Proposition 3.1.3. *Let $f(x) \in \mathbb{C}[[x]]$. We have*

$$e^{w\frac{d}{dx}} f(x) = f(x+w).$$

$\square$

We have calculated the higher derivatives of a product of two polynomials using the automorphism property. We next follow (in a very special case, for the derivation $\frac{d}{dx}$), the argument given in Proposition 8.3.4 of [FLM2] to calculate the higher derivatives of the composition of two power series. Let $f(x), g(x) \in \mathbb{C}[[x]]$. We further require that $g(x)$ have zero constant term, so that, for instance, the composition $f(g(x))$ is always well defined. We shall approach the problem by calculating the exponential generating function of the higher derivatives of $f(g(x))$. We get

$$\begin{aligned} e^{w\frac{d}{dx}} f(g(x)) &= f(g(x+w)) \\ &= f(g(x) + (g(x+w) - g(x))) \\ &= \left(e^{(g(x+w)-g(x))\frac{d}{dz}} f(z) \right) |_{z=g(x)} \\ &= \sum_{n \geq 0} \frac{f^{(n)}(g(x))}{n!} \left(e^{w\frac{d}{dx}} g(x) - g(x) \right)^n \\ &= \sum_{n \geq 0} \frac{f^{(n)}(g(x))}{n!} \left(\sum_{m \geq 1} \frac{g^{(m)}(x)}{m!} w^m \right)^n . \end{aligned} \tag{3.1.1}$$

While our calculation of the higher derivatives is not, strictly speaking, complete at this stage (although all that remains is a little work to extract the coefficients in powers of w), it is in fact this formula which will be of importance to us, since, roughly speaking, many results of the classical umbral calculus follow because of it, and so we shall record it as a proposition.

Proposition 3.1.4. *Let $f(x)$ and $g(x) \in \mathbb{C}[[x]]$. Let $g(x)$ have zero constant term. Then we have*

$$e^{w\frac{d}{dx}} f(g(x)) = \sum_{n \geq 0} \frac{f^{(n)}(g(x))}{n!} \left(\sum_{m \geq 1} \frac{g^{(m)}(x)}{m!} w^m \right)^n . \tag{3.1.2}$$

$\square$

We include the rest of the calculation of the higher derivatives of the composition of two polynomials, which yields Faà di Bruno's classical formula, in Section 3.12.

Remark 3.1.1. The more general version of this calculation appeared in [FLM2] because it was related to a much more subtle and elaborate argument showing that vertex operators associated to lattices satisfied a certain associativity property (see [FLM2], Sections 8.3 and 8.4). The connection is due, at least in part, to the rough resemblance between the exponential generating function of the higher derivatives of a composite function in the special case $f(x) = e^x$ (see (3.1.3) below) and "half of" a vertex operator.

Noting that in (3.1.1) we treated $g(x+w) - g(x)$ as one, atomic object suggests a reorganization. Indeed by calling $g(x+w) - g(x) = v$ and $g(x) = u$, the second, third and fourth lines of (3.1.1) become

$$f(u+v) = e^{v\frac{d}{du}} f(u) = \sum_{n\geq 0} \frac{f^{(n)}(u)}{n!} v^n.$$

This is just the formal Taylor theorem, of course, and so we could have begun here and then re-substituted for u and v to get the result. This, according to [Jo], is how the proof of U. Meyer [Me] runs.

It is also interesting to specialize to the case where $f(x) = e^x$, as is commonly done, and indeed was the case which interested the authors of [FLM2] and will interest us in later sections. We have simply

$$e^{w\frac{d}{dx}} e^{g(x)} = e^{g(x+w)} = e^{g(x)} e^{g(x+w)-g(x)} = e^{g(x)} e^{\sum_{m\geq 1} \frac{g^{(m)}(x)}{m!} w^m}. \tag{3.1.3}$$

Remark 3.1.2. The generating function for what are called the Bell polynomials (cf. Chapter 12.3 and in particular (12.3.6) in [An]) easily follows from (3.1.3) using a sort of slightly unrigorous old-fashioned umbral argument replacing $g^{(m)}$ with g_m (see the proof of Proposition 3.2.1 for one way of handling such arguments). Of course, if we also set $g(x) = e^x - 1$, we get $e^{w\frac{d}{dx}} e^{e^x - 1} = e^{e^{x+w} - 1}$ and setting $x = 0$ is easily seen to give the well-known result that $e^{e^w - 1}$ is the generating function of the Bell numbers, which are themselves the Bell polynomials with all variables evaluated at 1.

For convenience we shall globally name three generic (up to the indicated restrictions) elements of $\mathbb{C}[[t]]$:

$$A(t) = \sum_{n\geq 0} A_n \frac{t^n}{n!}, \qquad B(t) = \sum_{n\geq 1} B_n \frac{t^n}{n!}. \qquad \text{and} \qquad C(t) = \sum_{n\geq 0} C_n \frac{t^n}{n!}, \tag{3.1.4}$$

where both $B_1 \neq 0$ and $C_0 \neq 0$ (and note the ranges of summation). We recall, and it is easy for the reader to check, that $B(t)$ has a compositional inverse, which we denote by $\overline{B}(t)$, and that $C(t)$ has a multiplicative inverse, $C(t)^{-1}$. We note further that since $\overline{B}(t)$ has zero constant term, $B'(\overline{B}(t))$ is well defined, and we shall denote it by $B^*(t)$. In addition, $p(x)$ will always be a formal polynomial and sometimes we shall feel free to use a different variable such as z in the argument of one of our generic series, so that $A(z)$ is the same type of series as $A(t)$, only with the name of the variable changed.

We shall also use the notation $A^{(n)}(t)$ for the derivatives of, in this case, $A(t)$, and it will be convenient to define this for all $n \in \mathbb{Z}$ to include anti-derivatives. Of course, to make that well-defined we need to choose particular integration constants and only one choice is useful for us, as it turns out.

Notation 3.1.1. For all $n \in \mathbb{Z}$, given a fixed sequence $A_m \in \mathbb{C}$ for all $m \in \mathbb{Z}$, we shall define

$$A^{(n)}(t) = \sum_{m\geq n} \frac{A_m t^{m-n}}{(m-n)!}.$$

3.2 A restatement of the problem and further developments

In the last section we considered the problem of calculating the higher formal derivatives of a composite function of two formal power series, $f(g(x))$, where we obtained an answer involving only expressions of the form $f^{(n)}(g(x))$ and $g^{(m)}(x)$. Because of the restricted form of the answer it is convenient to translate the result into a more abstract notation which retains only those properties needed for arriving at Proposition 3.1.4. This essential structure depends only on the observation that $\frac{d}{dx} f^{(n)}(g(x)) = f^{(n+1)}(g(x))(g^{(1)}(x))$ for $n \geq 0$ and that $\frac{d}{dx} g^{(m)}(x) = g^{(m+1)}(x)$ for $m \geq 1$.

Motivated by the above paragraph, we consider the algebra

$$\mathbb{C}[\ldots, y_{-2}, y_{-1}, y_0, y_1, \ldots, x_1, x_2, \ldots].$$

Then let D be the unique derivation on $\mathbb{C}[\ldots, y_{-2}, y_{-1}, y_0, y_1, y_2, \cdots, x_1, x_2, \cdots]$ satisfying

$$\begin{aligned} Dy_i &= y_{i+1}x_1 \quad & i \in \mathbb{Z} \\ Dx_j &= x_{j+1} \quad & j \geq 1. \end{aligned}$$

Then the question of calculating $e^{w\frac{d}{dx}}f(g(x))$ as in the last section is seen to be essentially equivalent to calculating

$$e^{wD}y_0,$$

where we "secretly" identify D with $\frac{d}{dx}$, $f^{(n)}(g(x))$ with y_n and $g^{(m)}(x)$ with x_m. We shall make this identification rigorous in the proof of the following proposition, while noting that the statement of said following proposition is already (unrigorously) clear, by the "secret" identification in conjunction with Proposition 3.1.4.

Proposition 3.2.1. *We have*

$$e^{wD}y_0 = \sum_{n\geq 0} \frac{y_n \left(\sum_{m\geq 1} \frac{w^m x_m}{m!}\right)^n}{n!}. \tag{3.2.1}$$

Proof. Let $f(x), g(x) \in \mathbb{C}[[x]]$ such that $g(x)$ has zero constant term as in Proposition 3.1.4. Consider the unique algebra homomorphism

$$\phi_{f,g} : \mathbb{C}[\ldots, y_{-2}, y_{-1}, y_0, y_1, \ldots, x_1, x_2, \ldots] \to \mathbb{C}[[x]]$$

satisfying

$$\begin{aligned} \phi_{f,g}y_i &= f^{(i)}(g(x)) \quad & i \in \mathbb{Z} \quad \text{and} \\ \phi_{f,g}x_i &= g^{(i)}(x) \quad & i \geq 1. \end{aligned}$$

Then we claim that we have

$$\phi_{f,g} \circ D = \frac{d}{dx} \circ \phi_{f,g}.$$

Since $\phi_{f,g}$ is a homomorphism and D is a derivation, it is clear that we need only check that these operators agree when acting on y_i for $i \in \mathbb{Z}$ and x_j for $j \geq 1$. We get

$$(\phi_{f,g} \circ D)\, y_i = \phi_{f,g}(y_{i+1}x_1) = f^{(i+1)}(g(x))g'(x) = \frac{d}{dx}f^{(i)}(g(x)) = \left(\frac{d}{dx} \circ \phi_{f,g}\right) y_i$$

and

$$(\phi_{f,g} \circ D)\, x_i = \phi_{f,g} x_{i+1} = g^{(i+1)}(x) = \frac{d}{dx} g^{(i)}(x) = \left(\frac{d}{dx} \circ \phi_{f,g}\right) x_i,$$

which gives us the claim. Then, using the obvious extension of $\phi_{f,g}$, by (3.1.2) we have

$$\phi_{f,g}\left(e^{wD} y_0\right) = e^{w\frac{d}{dx}} \phi_{f,g} y_0 = e^{w\frac{d}{dx}} f(g(x)) = \phi_{f,g}\left(\sum_{n\geq 0} \frac{y_n \left(\sum_{m\geq 1} \frac{w^m x_m}{m!}\right)^n}{n!}\right), \quad (3.2.2)$$

for all $f(x)$ and $g(x)$.

Next take the formal limit as $x \to 0$ of the first and last terms of (3.2.2). These identities clearly show that we get identities when we substitute $f^{(n)}(0)$ for y_n and $g^{(n)}(0)$ for x_n in (3.2.1). But $f^{(n)}(0)$ and $g^{(n)}(0)$ are arbitrary and since (3.2.1) amounts to a sequence of multinomial polynomial identities when equating the coefficients of w^n, we are done. $\square$

We observe that it would have been convenient in the previous proof if the maps $\phi_{f,g}$ had been invertible. We provide a second proof of Proposition 3.2.1 using such a set-up. This proof is closely based on a proof appearing in [Ch]. We hope the reader won't mind a little repetition.

Proof. (second proof of Proposition 3.2.1)

Let $F(x) = \sum_{n\geq 0} \frac{y_n x^n}{n!}$ and $G(x) = \sum_{n\geq 1} \frac{x_n x^n}{n!}$. Consider the unique algebra homomorphism

$$\psi : \mathbb{C}[\ldots, y_{-2}, y_{-1}, y_0, y_1, \ldots, x_1, x_2, \ldots] \to \mathbb{C}[\ldots, y_{-2}, y_{-1}, y_0, y_1, \ldots, x_1, x_2, \ldots][[x]]$$

satisfying

$$\begin{aligned} \psi(y_i) &= F^{(i)}(G(x)) \quad && i \in \mathbb{Z} \quad \text{and} \\ \psi(x_i) &= G^{(i)}(x) && i \geq 1. \end{aligned}$$

Then we claim that we have

$$\psi \circ D = \frac{d}{dx} \circ \psi.$$

Since ψ is a homomorphism and D is a derivation, it is clear that we need only check that these operators agree when acting on y_i for $i \in \mathbb{Z}$ and x_j for $j \geq 1$. We get

$$(\psi \circ D)\, y_i = \psi(y_{i+1} x_1) = F^{(i+1)}(G(x)) G'(x) = \frac{d}{dx} F^{(i)}(G(x)) = \left(\frac{d}{dx} \circ \psi\right) y_i$$

and

$$(\psi \circ D)\, x_i = \psi(x_{i+1}) = G^{(i+1)}(x) = \frac{d}{dx} G^{(i)}(x) = \left(\frac{d}{dx} \circ \psi \right) x_i,$$

which gives us the claim. Then, using the obvious extension of ψ, we have

$$\psi \left(e^{wD} y_0 \right) = e^{w\frac{d}{dx}} \psi(y_0) = e^{w\frac{d}{dx}} F(G(x)). \tag{3.2.3}$$

But now we get to note that ψ has a left inverse, namely setting $x = 0$, because $F^{(i)}(G(0)) = y_i$ for $i \in \mathbb{Z}$ and $G^{(i)}(0) = x_i$ for $i \geq 1$. Thus we get

$$e^{wD} y_0 = \left(e^{w\frac{d}{dx}} F(G(x)) \right) |_{x=0} = F(G(x + w))|_{x=0} = F(G(w)), \tag{3.2.4}$$

which is exactly what we want. □

We note that our second proof of Proposition 3.2.1 did not depend on Proposition 3.1.4. Completing a natural circle of reasoning, by using the first proof of Proposition 3.2.1, before invoking Proposition 3.1.4, we had from (3.2.2)

$$\phi_{f,g} \left(e^{wD} y_0 \right) = e^{w\frac{d}{dx}} \phi_{f,g} y_0 = e^{w\frac{d}{dx}} f(g(x)),$$

which by (3.2.4) gives

$$e^{w\frac{d}{dx}} f(g(x)) = \phi_{f,g}(F(G(w))),$$

which gives us back Proposition 3.1.4. Thus we have shown in a natural way how Propositions 3.1.4 and 3.2.1 are equivalent.

Remark 3.2.1. One nice aspect of our second proof of Proposition 3.2.1, based closely on a proof in [Ch], is that its key brings to the fore of the argument perhaps the most striking feature of the result, which is that the exponential generating function of higher derivatives of a composite function is itself in the form of a composite function. This, of course, is an old-fashioned umbral feature. Furthermore, it was the form of the answer, that it roughly resembled "half of a vertex operator," which was what interested the authors of [FLM2]. This feature is also central to what follows.

We may now clearly state the trick on which (from our point of view) much of the classical umbral calculus is based. It is clear that if we substitute A_n for y_n and B_n for x_n in (3.2.1) then the right-hand side will become $A(B(w))$. Actually, it will be more interesting to substitute xB_n for x_n. With this as motivation, we formally define two (for flexibility) substitution maps. Let $\chi_{B(t)}$ and $\psi_{A(t)}$ be the algebra homomorphisms uniquely defined by the following:

$$\chi_{B(t)} : \mathbb{C}[\ldots, y_{-1}, y_0, y_1, \ldots, x_1, x_2, \cdots] \to \mathbb{C}[\ldots, y_{-1}, y_0, y_1, \ldots, x]$$

with

$$\begin{aligned} \chi_{B(t)}(y_i) &= y_i && i \in \mathbb{Z} \\ \chi_{B(t)}(x_j) &= B_j x && j \geq 1. \end{aligned}$$

and

$$\psi_{A(t)} : \mathbb{C}[\ldots, y_{-1}, y_0, y_1, \ldots, x] \to \mathbb{C}[x]$$

with

$$\begin{aligned} \psi_{A(t)}(y_i) &= A_i && i \in \mathbb{Z} \\ \psi_{A(t)}(x) &= x. \end{aligned}$$

Then we have

$$\psi_{A(t)} \circ \chi_{B(t)} \left(e^{wD} y_0\right) = A(xB(w)). \tag{3.2.5}$$

To keep the notation from becoming cluttered, we shall sometimes abbreviate $A(t)$ by simply A and make other similar abbreviations when there should be no confusion.

We next note that it is not difficult to explicitly calculate the action of $\psi_A \circ \chi_B \circ e^{wD}$ on $\mathbb{C}[\ldots, y_{-1}, y_0, y_1, \ldots, x_1, x_2, \ldots]$. Indeed it is easy to see that we have

$$\psi_A \circ \chi_B \circ e^{wD} y_n = A^{(n)}(xB(w)) \qquad n \in \mathbb{Z} \tag{3.2.6}$$

$$\text{and} \quad \psi_A \circ \chi_B \circ e^{wD} x_n = xB^{(n)}(w) \qquad n \geq 1. \tag{3.2.7}$$

These identities determine the action completely because of the automorphism property satisfied by e^{wD}.

The following series of identities (one of which is (3.2.5) repeated) is immediate from what we have shown:

$$\psi_A \circ \chi_B \circ e^{wD} y_1 = A'(xB(w)) \tag{3.2.8}$$

$$\psi_{A'} \circ \chi_B \circ e^{wD} y_0 = A'(xB(w)) \tag{3.2.9}$$

$$\frac{\partial}{\partial x} \circ \psi_A \circ \chi_B \circ e^{wD} y_{-1} = A(xB(w))B(w) \tag{3.2.10}$$

$$\psi_{tA(t)} \circ \chi_B \circ e^{wD} y_0 = xB(w)A(xB(w)) \tag{3.2.11}$$

$$\psi_A \circ \chi_B \left(e^{wD} y_0\right) = A(xB(w)) \tag{3.2.12}$$

$$\psi_{A\circ B} \circ \chi_t \circ e^{wD} y_0 = A(B(xw)) \tag{3.2.13}$$

$$\frac{\partial}{\partial w} \circ \psi_A \circ \chi_B \circ e^{wD} y_0 = A'(x(B(w))xB'(w) \tag{3.2.14}$$

$$\psi_{B^*(t)A'(t)} \circ \chi_B \circ e^{wD} y_0 = B^*(xB(w))A'(xB(w)). \tag{3.2.15}$$

We can now easily get the following proposition.

Proposition 3.2.2. *We have*

1. $A'(B(w)) = \left(\psi_A \circ \chi_B \circ e^{wD} y_1\right)|_{x=1} = \left(\psi_{A'} \circ \chi_B \circ e^{wD} y_0\right)|_{x=1}$,
2. $A(B(w))B(w) = \left(\frac{\partial}{\partial x} \circ \psi_A \circ \chi_B \circ e^{wD} y_{-1}\right)|_{x=1} = \left(\psi_{tA(t)} \circ \chi_B \circ e^{wD} y_0\right)|_{x=1}$,
3. $A(B(w)) = \left(\psi_A \circ \chi_B \circ e^{wD} y_0\right)|_{x=1} = \left(\psi_{A\circ B} \circ \chi_t \circ e^{wD} y_0\right)|_{x=1}$, *and*
4. $A'(B(w))B'(w) = \frac{\partial}{\partial w}\left(\left(\psi_A \circ \chi_B \circ e^{wD} y_0\right)|_{x=1}\right) = \left(\psi_{B^*(t)A'(t)} \circ \chi_B \circ e^{wD} y_0\right)|_{x=1}$.

Proof. All the identities are proved by setting $x = 1$ in (3.2.8), (3.2.9), (3.2.10), (3.2.11), (3.2.12), (3.2.13),(3.2.14) and (3.2.15) and equating the results pairwise as follows. Equations (3.2.8) and (3.2.9) give (1); equations (3.2.10) and (3.2.11) give (2); equations (3.2.12) and (3.2.13) give (3); and equations (3.2.14) and (3.2.15) give (4). □

Each of the identities in Proposition 3.2.2 turns out to be equivalent to the fact that a certain pair of operators are adjoints. In order to see this, our next task will be to put the procedure of setting $x = 1$, used in Proposition 3.2.2, into a context of linear functionals. We shall do this in the next section.

3.3 Umbral connection

We set up a bra-ket notation following [Rm1] so that we may precisely recover umbral results in the formalism there presented.

Notation 3.3.1. Let $f(x) = \sum_{n\geq 0} f_n x^n \in \mathbb{C}[x]$. Then we define

$$\langle A(v)|f(x)\rangle = \sum_{n\geq 0} f_n A_n.$$

So we are now viewing $A(v)$ as a linear functional on $\mathbb{C}[[x]]$. This leads us to the notion of adjoint operators, a key notion in the umbral calculus as presented in [Rm1]. We shall soon show how to recover certain of the same results about adjoints from our point of view.

Definition 3.3.1. We say that a linear operator ϕ on $\mathbb{C}[x]$ and a linear operator ϕ^* on $\mathbb{C}[[v]]$ are adjoints exactly when, for all $A(v)$ and for all $p(x)$, the following identity is satisfied:

$$\langle \phi^*(A(v))|p(x)\rangle = \langle A(v)|\phi(p(x))\rangle.$$

Of course, by linearity, it is equivalent to require that the identity in Definition 3.3.1 be satisfied for $p(x)$ ranging over a basis of $\mathbb{C}[x]$. In addition, we extend the bra-ket notation in the obvious way to handle elements of $\mathbb{C}[x][[w]]$ "coefficient-wise."

Proposition 3.3.1. *If ϕ is a linear operator on $\mathbb{C}[x]$ and ϕ^* is a linear operator on $\mathbb{C}[[v]]$ such that*

$$\langle \phi^*(A(v))|e^{xB(w)}\rangle = \langle A(v)|\phi\left(e^{xB(w)}\right)\rangle,$$

for all $A(v)$ and $B(w)$, then ϕ and ϕ^* are adjoints.

Proof. Equating coefficients of w^n gives us the adjoint equation for a sequence of polynomials $B_n(x)$ of degree exactly n and arbitrary $A(v)$. Since the degree of $B_n(x)$ is n, these polynomials form a basis and so the result follows by linearity. $\square$

The next theorem allows us to translate our "set $x = 1$" procedure from Proposition 3.2.2 into the bra-ket notation.

Theorem 3.3.1. *Let* $u \in \mathbb{C}[y_0, y_1, \cdots, x]$ *be of the form* $u = \sum_{n\geq 0} u_n y_n x^n$ *where* $u_n \in \mathbb{C}$. *Then we have:*

$$\langle A(v)|\psi_{e^t}(u)\rangle = (\psi_A(u))|_{x=1}.$$

Proof. We calculate to get:

$$\langle A(v)|\psi_{e^t}(u)\rangle = \langle A(v)|\sum_{n\geq 0} u_n x^n\rangle = \sum_{n\geq 0} u_n A_n = (\psi_A(u))|_{x=1}.$$

□

Theorem 3.3.2. *We have*

1. $p(x) \in \mathbb{C}[x]$, *viewed as a multiplication operator on* $\mathbb{C}[x]$ *and* $p(\frac{d}{dv})$ *are adjoint operators.*

2. $F(\frac{d}{dx}) \in \mathbb{C}[[\frac{d}{dx}]]$ *and* $F(v)$ *viewed as a multiplication operator on* $\mathbb{C}[[v]]$ *are adjoint operators.*

Proof. We first prove (1). It is obvious that we need only to check the case where $p(x) = x$. By (3.2.1) and Theorem 3.3.1 we have that part (1) of Proposition 3.2.2 is essentially equivalent to

$$A'(B(w)) = \langle A(v)|\psi_{e^t}\left(x\left(\chi_B \circ e^{wD} y_1\right)\right)\rangle = \langle A'(v)|\psi_{e^t} \circ \chi_B \circ e^{wD} y_0\rangle,$$

which in turn, by (3.2.5) and (3.2.6), gives

$$\langle A(v)|xe^{xB(w)}\rangle = \langle \frac{d}{dv} A(v)|e^{xB(w)}\rangle,$$

so that by Proposition 3.3.1 we have established (1).

We now prove (2) in a similar fashion. It is obvious that we need only check the case where $F(t) = t$. By (3.2.1) and Theorem 3.3.1 we have that part (2) of Proposition 3.2.2 is essentially equivalent to

$$A(B(w))B(w) = \langle A(v)|\psi_{e^t} \circ \frac{\partial}{\partial x} \circ \chi_B \circ e^{wD} y_{-1}\rangle = \langle vA(v)|\psi_{e^t} \circ \chi_B \circ e^{wD} y_0\rangle,$$

which in turn, by (3.2.5) and (3.2.6), gives

$$\langle A(v)|B(w)e^{xB(w)}\rangle = \langle vA(v)|e^{x(B(w))}\rangle \quad \Leftrightarrow$$
$$\langle A(v)|\frac{d}{dx}e^{xB(w)}\rangle = \langle vA(v)|e^{x(B(w))}\rangle,$$

so that by Proposition 3.3.1 we get (2). □

Remark 3.3.1. Part (1) of Theorem 3.3.2 appeared as Theorem 2.1.10 in [Rm1] and Part (2) of Theorem 3.3.2 appeared as Theorem 2.2.5 in [Rm1].

In light of the proof of Theorem 3.3.2, we might consider whether parts (3) and (4) of Proposition 3.2.2 also amount to adjoint relationships and, in fact, they do. By (3.2.1) and Theorem 3.3.1, we have that part (3) of Proposition 3.2.2 is essentially equivalent to

$$A(B(w)) = \langle A(v)|\psi_{e^t} \circ \chi_B \circ e^{wD}y_0\rangle = \langle A(B(v))|\psi_{e^t} \circ \chi_t \circ e^{wD}y_0\rangle, \tag{3.3.1}$$

which in turn, by (3.2.5), gives

$$\langle A(v)|e^{xB(w)}\rangle = \langle A(B(v))|e^{xw}\rangle. \tag{3.3.2}$$

We have therefore effectively calculated the adjoint to the substitution map S_B which acts by $S_B(g(v)) = g(B(v))$ for all $g(v) \in \mathbb{C}[[v]]$. We simply need to make a couple of definitions.

Remark 3.3.2. We shall be defining certain linear operators on $\mathbb{C}[x]$ by specifying, for instance, how they act on e^{xw}, which, recall, stands for the formal exponential expansion. Of course, by this we mean that the operator acts only the coefficients of w^n $n \geq 0$. We have already employed similar abuses of notation with the action of $\phi_{f,g}$ in the proof of Proposition 3.2.1 and with the bra-ket notation as mentioned in the comment preceding Proposition 3.3.1.

We now recall the definition of certain "umbral operators"; cf. Section 3.4 in [Rm1] where what we are calling "attached umbral operators" appeared as "umbral operators." More particularly, the umbral operator attached to a sequence $B(w)$ in this work is the same as the umbral operator for $\overline{B}(w)$ in [Rm1].

Definition 3.3.2. We define the umbral operator attached to $B(w)$ to be the unique linear map $\theta_B : \mathbb{C}[x] \to \mathbb{C}[x]$ satisfying:

$$\theta_B e^{xw} = e^{xB(w)}.$$

Theorem 3.3.3. *We have that S_B and θ_B are adjoint operators.*

Proof. The result follows from Proposition 3.3.1 and (3.3.2). $\square$

Remark 3.3.3. Theorem 3.3.3 essentially appeared as Theorem 3.4.1 in [Rm1], although in this work we have chosen some different characterizations of certain objects as definitions as discussed in the introduction. It is not difficult to tie up all the relevant information and we shall indicate the essentials of what is needed along these lines in detail in Section 3.4.

By (3.2.1) and Theorem 3.3.1 we have that part (4) of Proposition 3.2.2 is essentially equivalent to

$$A'(B(w))B'(w) = \frac{\partial}{\partial w}\langle A(v)|\psi_{e^t} \circ \chi_B \circ e^{wD} y_0\rangle = \langle B^*(v)A'(v)|\psi_{e^t} \circ \chi_B \circ e^{wD} y_0\rangle,$$

which in turn, by (3.2.5), gives

$$\begin{aligned} \frac{\partial}{\partial w}\langle A(v)|e^{xB(w)}\rangle = \langle B^*(v)A'(v)|e^{xB(w)}\rangle \quad \Leftrightarrow \\ \langle A(v)|\frac{\partial}{\partial w}e^{xB(w)}\rangle = \langle B^*(v)A'(v)|e^{xB(w)}\rangle. \end{aligned} \tag{3.3.3}$$

We now recall the definition of certain "Sheffer shifts"; cf. Section 3.6 in [Rm1], where what we are calling "attached Sheffer shifts" appeared as "umbral shifts." More particularly, the Sheffer shift attached to a sequence $B(w)$ in this work is the same as the umbral shift for $\overline{B}(w)$ in [Rm1].

Definition 3.3.3. For each $B(w)$, let $D_B : \mathbb{C}[x] \to \mathbb{C}[x]$ be the unique linear map satisfying

$$D_B e^{xB(w)} = \frac{\partial}{\partial w} e^{xB(w)}. \tag{3.3.4}$$

We call D_B the *Sheffer shift* attached to $B(w)$.

Remark 3.3.4. As discussed in the Introduction and Remark 3.1.1, the authors of [FLM2] were concerned with the exponential generating function of the higher derivatives of a composite function, because it roughly resembled "half of" a vertex operator. Following this analogy, we might say that $\psi_{e^t} \circ \chi_B \circ e^{wD} y_0$ is an umbral analogue of ("half of") a vertex operator. Having made this analogy, one can see how using (3.3.4), we have defined attached Sheffer shifts as those operators satisfying an analogue of the $L(-1)$-bracket-derivative property, which is stated as the equality of the first and third expressions of formula (8.7.30) in [FLM2]. See also Remark 3.11.4.

Theorem 3.3.4. *We have that D_B and $B^*(v) \circ \frac{d}{dv}$ are adjoints.*

Proof. The result follows from Proposition 3.3.1 and (3.3.3). □

Remark 3.3.5. Theorem 3.3.4 essentially appeared as part of Theorem 3.6.1 in [Rm1], where the author of that work had already, in addition, shown that the operators $B^*(t) \circ \frac{d}{dt}$ are exactly the surjective derivations on $\mathbb{C}[[t]]$, a routine matter once we note that $B^*(t)$ is an arbitrary element of $\mathbb{C}[[t]]$ having a multiplicative inverse. We also mention a similar caveat for the reader regarding different choices of definitions between the present work and [Rm1] just as discussed in Remark 3.3.3 and in the Introduction.

In closing this section we note obvious characterizations of the attached umbral operators and attached Sheffer shifts in terms of the coefficients of their generating function definitions. For this it is convenient for us to recall the definition of attached Sheffer sequences; cf. Section 2.3 and Theorem 2.3.4 in particular in [Rm1] as well as Proposition 3.4.4 and Remark 3.4.4 in this work. We note that the Sheffer sequence attached to a sequence $B(w)$ in this work is the same as the Sheffer sequence associated to $\overline{B}(w)$ in [Rm1].

Definition 3.3.4. We define the sequence of polynomials $B_n(x)$, the Sheffer sequence attached to $B(w)$, to be the unique sequence satisfying the following:

$$e^{xB(w)} = \sum_{n\geq 0} \frac{x^n B(w)^n}{n!} = \sum_{n\geq 0} \frac{B_n(x) w^n}{n!}.$$

Remark 3.3.6. We recall that the attached Sheffer sequences already appeared explicitly (though, of course not by name) in the proof of Proposition 3.3.1.

Proposition 3.3.2. *We have that $\theta_B : \mathbb{C}[x] \to \mathbb{C}[x]$, the umbral operator attached to $B(w)$, is characterized as the unique linear map satisfying:*

$$\theta_B x^n = B_n(x).$$

□

Proposition 3.3.3. *We have that $D_B : \mathbb{C}[x] \to \mathbb{C}[x]$, the Sheffer shift attached to $B(w)$, is characterized as the unique linear map satisfying:*

$$D_B B_n(x) = B_{n+1}(x).$$

□

3.4 Some historical, contextual results

In this section we shall establish some well-known results of the classical umbral calculus using a mixture of generating function techniques, operator theoretic techniques and some of our previous results. This approach of using a mixture of techniques is intended to be in keeping with the advice given in [Ga]. We include this material for some amount of completeness as well as the convenience of the reader who wishes to more easily see how this work fits into the historical context. I shall continue to use [Rm1], who uses mainly operator theoretic techniques, as my main official reference to the literature, while making occasional informal references to other important sources.

Proposition 3.4.1. *A sequence of polynomials, $B_n(x)$ with $n \geq 0$, is the Sheffer sequence attached to $B(t)$ if and only if*

$$\begin{array}{ll} (i) & B_n(0) = \delta_{n,0} \\ (ii) & \overline{B}\left(\frac{\partial}{\partial x}\right) B_n(x) = nB_{n-1}(x). \end{array}$$

Proof. We first assume that $B_n(x)$ is the Sheffer sequence attached to $B(t)$. Setting $x = 0$ in $e^{xB(t)} = \sum_{n\geq 0} \frac{B_n(x)t^n}{n!}$ gives (i). Since

$$\left(\frac{\partial}{\partial x}\right)^m e^{xB(t)} = B(t)^m e^{xB(t)} \qquad m \geq 0,$$

it is clear by linearity that

$$\overline{B}\left(\frac{\partial}{\partial x}\right)e^{xB(t)} = \overline{B}(B(t))e^{xB(t)} = te^{xB(t)}.$$

The result now follows by equating coefficients of t.

For the converse we assume that

$$\overline{B}\left(\frac{\partial}{\partial x}\right)\sum_{n\geq 0}\frac{B_n(x)t^n}{n!} = t\sum_{n\geq 0}\frac{B_n(x)t^n}{n!}, \tag{3.4.1}$$

which as we saw above is the generating function version of condition (ii). But we may also write

$$\sum_{n\geq 0}\frac{B_n(x)t^n}{n!} = \sum_{n\geq 0}\frac{F_n(t)x^n}{n!},$$

where $F_n(t) \in \mathbb{C}[[t]]$ are yet to be determined. Then by (3.4.1) we have

$$\begin{aligned}
\sum_{n\geq 0}\frac{F_{n+1}(t)x^n}{n!} &= \frac{\partial}{\partial x}\sum_{n\geq 0}\frac{F_n(t)x^n}{n!} \\
&= B\left(\overline{B}\left(\frac{\partial}{\partial x}\right)\right)\sum_{n\geq 0}\frac{F_n(t)x^n}{n!} \\
&= B(t)\sum_{n\geq 0}\frac{F_n(t)x^n}{n!},
\end{aligned}$$

which by equating coefficients of x^n gives

$$F_{n+1}(t) = F_n(t)B(t) \qquad n \geq 0.$$

Since condition (i) says exactly that $F_0(t) = 1$, we have $F_n(t) = B(t)^n$, which is exactly what we need. □

Remark 3.4.1. Proposition 3.4.1 essentially appeared as Theorem 2.4.5 in [Rm1]. Our proofs are quite different, since the approach in [Rm1] is based on operator theoretic language, whereas ours is based on generating functions. Indeed, we are even using different (equivalent) definitions and will only completely show the equivalence once we have proved Proposition 3.4.4. Some readers may recognize these operators $\overline{B}\left(\frac{\partial}{\partial x}\right)$ to be what Rota and Mullin called "delta operators" [MR]. This type of generating function approach was used by Garsia in [Ga] where this result appeared essentially as Theorem

2.5. Two proofs were given in [Ga], the second of which is the generating-function style proof. The first half of our proof is the same as the proof of that implication given in [Ga]. We note finally that this is one of the results which showcases the old-fashioned "umbral" nature of the subject area wherein "subscripts are acting like exponents."

Proposition 3.4.2. *A (nonzero) sequence of polynomials, $B_l(x)$ with $l \geq 0$, is the Sheffer sequence attached to $B(t)$ if and only if*

$$B_l(x+y) = \sum_{m \geq 0} \binom{l}{m} B_{l-m}(x) B_m(y) \qquad \text{for all} \quad l \geq 0. \tag{3.4.2}$$

Proof. We first assume that $B_n(x)$ is attached to $B(t)$. Then

$$\begin{aligned}\sum_{l \geq 0} \frac{B_l(x+y)t^l}{l!} &= e^{(x+y)B(t)} = e^{xB(t)} e^{yB(t)} \\ &= \sum_{n \geq 0} \frac{B_n(x)t^n}{n!} \sum_{m \geq 0} \frac{B_m(y)t^m}{m!} \\ &= \sum_{m,n \geq 0} \frac{B_n(x)B_m(y)}{n!m!} t^{n+m} \\ &= \sum_{l \geq 0} \sum_{m=0}^{l} \frac{B_{l-m}(x)B_m(y)}{(l-m)!m!} t^l,\end{aligned}$$

and equating coefficients in t gives the result.

For the converse, we assume

$$\sum_{l \geq 0} \frac{B_l(x+y)t^l}{l!} = \sum_{n \geq 0} \frac{B_n(x)t^n}{n!} \sum_{m \geq 0} \frac{B_m(y)t^m}{m!}, \tag{3.4.3}$$

which as we have seen is a generating function version of (3.4.2). But, as in the proof of Proposition 3.4.1, we may also write

$$\sum_{n \geq 0} \frac{B_n(x)t^n}{n!} = \sum_{n \geq 0} \frac{F_n(t)x^n}{n!},$$

where $F_n(t) \in \mathbb{C}[[t]]$ are yet to be determined. Then we can rewrite (3.4.3) as

$$\begin{aligned}\sum_{m,n \geq 0} \frac{F_n(t)F_m(t)}{n!m!} x^n y^m &= \sum_{l \geq 0} \frac{F_l(t)(x+y)^l}{l!} \\ &= \sum_{l \geq 0} \sum_{r \geq 0} \frac{x^{l-r} y^r F_l(t)}{(l-r)!r!} \\ &= \sum_{l,s} \frac{x^s y^r}{s!r!} F_{r+s}(t),\end{aligned}$$

which by equating coefficients in x and y gives

$$F_n(t)F_m(t) = F_{n+m}(t).$$

By setting $n = 0$ gives $F_0(t) = 1$ (or else $B_n(x)$ would be 0 for all n) and an easy induction shows that $F_n(t) = F_1(t)^n$, so that by definition the sequence $B_n(x)$ is the Sheffer sequence attached to $F_1(t)$. $\square$

Remark 3.4.2. Proposition 3.4.2 essentially appeared as Theorem 2.4.7 in [Rm1]. One difference is that for us the symbol y is formal (as is typical in the formal calculus used to study vertex operator algebras) instead of being a complex number. Also, we had to state that the sequence was nonzero as an "extra" assumption, because in [Rm1] there was a global assumption made in the opening sentence of Chapter 2, Section 3 that for all polynomial sequences $B_n(x)$ the degree of $B_m(x)$ would be m, an assumption also made by Mullin and Rota in [MR]. This result also appeared as Theorem 4.1 in [Ga] where a generating-function style proof was given. The first half of our proof is the same as the proof of that implication given in [Ga]. It is, of course, essentially because of this result that Rota and Mullin called what we have called attached Sheffer sequences by the name sequences of *binomial type.*

Remark 3.4.3. The reader may have noted how we used the same trick in the proofs of the converse statements in Propositions 3.4.1 and 3.4.2, where we switched which variable we were summing over to study the series we called $F_n(t)$. It is useful to regard our two-variable generating functions as giving the coefficients of (infinite) lower triangular matrices. This trick of focusing on the column generating functions is closely related to the point of view used by Shapiro, Getu, Woan and Woodson [SGWW], about which we shall have more to say in Section 3.7. Of course, this switching to and fro between what we may now regard as rows and columns gives a hint from the generating function point of view as to why these results can be framed in the language of linear functionals and adjoints.

Proposition 3.4.3. *Let $f_n(x)$ $n \geq 0$ be a sequence of elements of $\mathbb{C}[[x]]$ such that the degree of the lowest nonzero term of $f_n(x)$ is n. Polynomials $p(x)$ and $q(x)$ are identical if and only if $\langle f_n(v)|p(x)\rangle = \langle f_n(v)|q(x)\rangle$ for all $n \geq 0$.*

Proof. If $p(x) = \sum_{n\geq 0} p_n x^n$ and $q(x) = \sum_{n\geq 0} q_n x^n$ are identical the conclusion is obvious. For the converse, we shall induct on degree. First of all, we note that the degree of $p(x)$ can be characterized as the highest integer l such that $\langle f_l(v)|p(x)\rangle \neq 0$. Thus we see that $p(x)$ and $q(x)$ must have the same degree. Let $f_n(x) = \sum_{m\geq 0} f_m^n \frac{x^m}{m!}$. If the degree of $p(x)$ and $q(x)$ is l, then we have

$$\langle f_l(z)|p(x)\rangle = f_l^l p_l = \langle f_l(z)|q(x)\rangle = f_l^l q_l,$$

so that the coefficient of the highest degree terms are equal. This equality provides both the basis step and inductive step to reach our result. □

If $f(x) \in \mathbb{C}[[x]]$ is a power series with zero constant term, but nonzero first-degree term, then the sequence of power series $f_n(x) = f(x)^n$ for $n \geq 0$ satisfies the condition needed in the preceding proposition and this is the only case we are interested in.

Proposition 3.4.4. *The sequence of polynomials $p_n(x)$ is attached to $B(t)$ if and only if $\langle \overline{B}(z)^k|p_n(x)\rangle = n!\delta_{n,k}$ for all $k \geq 0$.*

Proof. By (3.3.1) and (3.3.2) we know that

$$\langle A(v)|e^{xB(w)}\rangle = A(B(w)),$$

which specializes to

$$\begin{aligned} w^k = \langle \overline{B}(v)^k|e^{xB(w)}\rangle &= \langle \overline{B}(v)^k| \sum_{n\geq 0} \frac{B_n(x)}{n!} w^n\rangle \\ \Leftrightarrow \langle \overline{B}(v)^k|B_n(x)\rangle &= n!\delta_{n,k} \qquad k \geq 0. \end{aligned} \tag{3.4.4}$$

Thus by Proposition 3.4.3 $B_n(x)$, the polynomials attached to $B(t)$ are the unique solution to (3.4.4), which proves both implications. □

Remark 3.4.4. Essentially it is this characterization of attached Sheffer sequences that Roman uses for his definition (see Theorem 2.3.1 [Rm1]).

Proposition 3.4.5. *If $B_n(x)$ is attached to $B(t)$ then for any $h(t) \in \mathbb{C}[[t]]$ we have*

$$h(t) = \sum_{n\geq 0} \frac{\langle h(v)|B_n(x)\rangle}{n!} \overline{B}(t)^n.$$

Proof. By (3.3.1) and (3.3.2) we know that

$$\langle h(v)|e^{xB(t)}\rangle = h(B(t)),$$

which specializes to

$$\begin{aligned} h(t) &= \langle h(v)|e^{xt}\rangle \\ &= \langle h(v)|e^{xB(\overline{B}(t))}\rangle \\ &= \langle h(v)|\sum_{n\geq 0} \frac{B_n(x)\overline{B}(t)^n}{n!}\rangle. \end{aligned}$$

□

Remark 3.4.5. Essentially, Proposition 3.4.5 appeared as Theorem 2.4.1 in [Rm1].

Proposition 3.4.6. *For any $p(T) \in \mathbb{C}[T]$, we have:*

$$p(t) = \sum_{k\geq 0} \frac{\langle \overline{B}(v)^k|p(x)\rangle}{k!} B_n(t).$$

Proof. It is routine to check that

$$p(t) = \langle e^{tv}|p(x)\rangle$$

by evaluating the case $p(x) = x^m$ and extending by linearity. Noting (as in the proof of Proposition 3.4.5) that

$$e^{tv} = e^{tB(\overline{B}(v))} = \sum_{k\geq 0} \frac{\overline{B}(v)^k B_n(t)}{k!}$$

gives the result. □

Remark 3.4.6. Essentially, Proposition 3.4.6 appeared as Theorem 2.4.2 in [Rm1].

We shall next obtain a relationship between different umbral shifts. We begin by calculating to get:

$$\begin{aligned}
D_B e^{xt} &= D_B e^{xB(\overline{B}(t))} \\
&= \left(D_B e^{xB(t)} \right) |_{t=\overline{B}(t)} \\
&= \left(\frac{\partial}{\partial t} e^{xB(t)} \right) |_{t=\overline{B}(t)} \\
&= \sum_{n\geq 0} \frac{x^n n B^{n-1}(\overline{B}(t)) B'(\overline{B}(t))}{n!} \\
&= B'(\overline{B}(t)) \sum_{n\geq 0} \frac{x^n n t^{n-1}}{n!} \\
&= B'(\overline{B}(t)) \frac{\partial}{\partial t} e^{xt}.
\end{aligned}$$

Thus we have, for any $F(t) \in \mathbb{C}[[t]]$ such that the constant term of $F(t)$ is zero and the first degree term is not:

$$\begin{aligned}
D_B e^{xt} &= \frac{B'(\overline{B}(t))}{F'(\overline{F}(t))} D_F e^{xt} \\
&= D_F \left(\frac{B'(\overline{B}(t))}{F'(\overline{F}(t))} e^{xt} \right) \\
&= D_F \left(\frac{B'(\overline{B}(\frac{d}{dx}))}{F'(\overline{F}(\frac{d}{dx}))} e^{xt} \right).
\end{aligned}$$

The following proposition is now clear.

Proposition 3.4.7. *We have*

$$D_F \circ \frac{B'(\overline{B}(\frac{d}{dx}))}{F'(\overline{F}(\frac{d}{dx}))} = D_B.$$

□

Remark 3.4.7. Essentially, Proposition 3.4.7 appears as Theorem 3.6.5 in [Rm1].

Following more closely to [Rm1] we can obtain the result of Proposition 3.4.7 using adjoints.

Proof. (Second proof of Proposition 3.4.7)

We can calculate using adjoints, using in particular part (2) of Theorem 3.3.2 and Theorem 3.3.4, to get:

$$\begin{aligned}
\langle A(v)|D_B p(x)\rangle &= \langle B^*(v)\frac{d}{dv}A(v)|p(x)\rangle \\
&= \langle \frac{B^*(v)}{F^*(v)}F^*(v)\frac{d}{dv}A(v)|p(x)\rangle \\
&= \langle F^*(v)\frac{d}{dv}A(v)|\frac{B^*(\frac{d}{dx})}{F^*(\frac{d}{dx})}p(x)\rangle \\
&= \langle A(v)|D_F\frac{B^*(\frac{d}{dx})}{F^*(\frac{d}{dx})}p(x)\rangle,
\end{aligned}$$

for all $A(v) \in \mathbb{C}[[v]]$ and for all $p(x) \in \mathbb{C}[x]$, which gives the result. □

3.5 Umbral shifts revisited

In this section we shall show a characterization of the attached Sheffer shifts which will be useful in Section 3.11. We begin by (for temporary convenience) generalizing Definition 3.3.3.

Definition 3.5.1. For each $A(t)$ and $B(t)$, let $D_B^A : \mathbb{C}[x] \to \mathbb{C}[x]$ be the unique linear map satisfying

$$D_B^A e^{xB(t)} = \frac{\partial}{\partial t}A(xB(t)).$$

Recalling the identities (3.2.6) and (3.2.7), we note that

$$\begin{aligned}
\psi_A \circ \chi_B \circ De^{wD}y_0 &= \psi_A \circ \chi_B \circ e^{wD}y_1x_1 \\
&= \psi_A \circ \chi_B \circ \left(e^{wD}y_1\right)\left(e^{wD}x_1\right) \\
&= A'(xB(w))xB'(w),
\end{aligned}$$

so that

$$\begin{aligned}
A'(xB(w))xB'(w) &= \psi_A \circ \chi_B \circ De^{wD}y_0 \\
&= \frac{\partial}{\partial w}\left(\psi_A \circ \chi_B \circ e^{wD}y_0\right) \\
&= D_B^A \circ \psi_{e^t} \circ \chi_B \circ e^{wD}y_0.
\end{aligned}$$

Extracting coefficients in t^n for $n \geq 0$ from the second and fourth terms from the above identity yields:

$$D_B^A \circ \psi_{e^t} \circ \chi_B \circ D^n y_0 = \psi_A \circ \chi_B \circ D^{n+1} y_0.$$

Furthermore, because $\psi_{e^t} \circ \chi_B \circ D^n y_0$ is a polynomial of degree exactly n, this formula characterizes the maps D_B^A.

Although we have briefly generalized the definition for the attached Sheffer shifts (in order to fit more closely with our calculations from Section 3.2), the previous identity shows how it is natural to restrict our attention to the attached Sheffer shifts, and it is this case that will later interest us anyway. We may now state the characterization of the attached Sheffer shifts mentioned in the introduction to this section.

Proposition 3.5.1. *The attached Sheffer shift, $D_B : \mathbb{C}[x] \to \mathbb{C}[x]$ is the unique linear map satisfying*

$$D_B \circ \psi_{e^t} \circ \chi_B \circ D^n y_0 = \psi_{e^t} \circ \chi_B \circ D^{n+1} y_0,$$

for all $n \geq 0$.

□

Remark 3.5.1. Proposition 3.5.1 was announced, together with a more direct proof, as Proposition 6.1 in [R2] and is reproduced in this work as Proposition 2.10.1.

3.6 Sheffer sequences

The classical umbral calculus can be considered to be the study of Sheffer sequences through "umbral" techniques. So far we have only considered a special case, attached Sheffer sequences. The general case may be obtained easily by using the results of the special case and this is probably the shortest route given the efforts we have already made. However, at the risk of being somewhat repetitive we shall describe completely a parallel development since this seems to more explicitly reveal the relationship of the two levels of generality. Recall that our approach began by calculating the higher derivatives of a formal composite function. That is, letting $f(x), g(x) \in \mathbb{C}[[x]]$ such that the constant term of $g(x)$ is zero, we began calculating the higher derivatives of $f(g(x))$. Let $h(x) \in \mathbb{C}[[x]]$. Then the general theory follows by the parallel argument with the starting point of calculating the higher derivatives of the product $h(x)f(g(x))$. By using the automorphism property this is easy. We have

$$e^{w\frac{d}{dx}}h(x)f(g(x)) = \left(\sum_{l\geq 0}\frac{h^{(l)}(x)}{l!}w^l\right)\left(\sum_{n\geq 0}\frac{f^{(n)}(g(x))}{n!}\left(\sum_{m\geq 1}\frac{g^{(m)}(x)}{m!}w^m\right)^n\right).$$

As previously, we abstract the essential information. Consider the algebra

$$\mathbb{C}[\dots y_{-1}, y_0, y_1, y_2, \dots, x_1, x_2, \dots, z_0, z_1, z_2, \dots].$$

Let E be the unique derivation on $\mathbb{C}[\dots y_{-1}, y_0, y_1, y_2, \dots, x_1, x_2, \dots, z_0, z_1, z_2, \dots]$ satisfying the following:

$$\begin{aligned} Ey_i &= y_{i+1}x_1 && i \in \mathbb{Z} \\ Ex_j &= x_{j+1} && j \geq 1 \\ Ez_l &= z_{l+1} && l \geq 0. \end{aligned}$$

Then the question of calculating $e^{w\frac{d}{dx}}h(x)f(g(x))$ is seen to be essentially equivalent to calculating

$$e^{wE}z_0y_0,$$

where we identify E with $\frac{d}{dx}$, $h^{(l)}(x)$ with z_l, $f^{(n)}(g(x))$ with y_n and $g^{(m)}(x)$ with x_m. We could now mimic the proof of Proposition 3.2.1, but it is trivial to invoke it and note that by the automorphism property and (3.2.1) we have

$$e^{wE}z_0y_0 = \sum_{l\geq 0}\frac{z_lw^l}{l!}\sum_{n\geq 0}\frac{y_n\left(\sum_{m\geq 1}\frac{w^mx_m}{m!}\right)^n}{n!}. \tag{3.6.1}$$

In this case we would like to substitute A_n for y_n, xB_n for x_n and C_n for z_n. For flexibility we formalize this with three substitution maps. We already have two of them, so we need only further define ξ_C to be the algebra homomorphism satisfying

$$\xi(C) : \mathbb{C}[\dots, y_{-1}, y_0, y_1, \dots, x_1, x_2, \dots, z_0, z_1, \dots] \to \mathbb{C}[\dots, y_{-1}, y_0, y_1, \dots, x_1, x_2, \dots],$$

with

$$\begin{aligned} \xi(C)(x_j) &= x_j \quad i \geq 1 \\ \xi(C)(y_i) &= y_i \quad i \in \mathbb{Z} \\ \xi(C)(z_l) &= C_l \quad l \geq 0. \end{aligned}$$

Then we have

$$\psi_A \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0 = C(w) A(xB(w)). \tag{3.6.2}$$

Of course, it is easy to calculate $\psi_A \circ \chi_B \circ \xi_C \circ e^{wE}$ on all of

$$\mathbb{C}[\ldots, y_{-1}, y_0, y_1, \ldots, x_1, x_2, \ldots, z_0, z_1,].$$

The following identities are immediate.

$$\psi_{(C)(A \circ B)} \circ \chi_t \circ \xi_1 \circ e^{wE} z_0 y_0 = C(xw) A(B(xw)), \tag{3.6.3}$$

$$\frac{\partial}{\partial w} \circ \psi_A \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0 = C'(w) A(xB(w)) + xC(w) A'(xB(w)) B'(w) \quad \text{and} \tag{3.6.4}$$

$$\begin{aligned} &\psi_{(C'(\overline{B}))(C^{-1}(\overline{B}))(A)+(B^*)(A')} \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0 \\ &= C(w) \left(C'(\overline{B}(xB(w))) C^{-1}(\overline{B}(xB(w))) A(xB(w)) + B^*(xB(w)) A'(xB(w)) \right), \end{aligned} \tag{3.6.5}$$

where the third identity seems too complicated to be useful except that we are next going to set $x = 1$ in the following Proposition.

Proposition 3.6.1. *We have*

1. $C(w)A(B(w)) = \psi_A \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0|_{x=1} = \psi_{(C)(A \circ B)} \circ \chi_t \circ \xi_1 \circ e^{wE} z_0 y_0|_{x=1}$ *and*

2. $C'(w)A(B(w)) + C(w)A'(B(w))B'(w)$
   $= \frac{d}{dw} \left(\left(\psi_A \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0 \right) |_{x=1} \right)$
   $= \psi_{(C'(\overline{B}))(C^{-1}(\overline{B}))(A)+(B^*)(A')} \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0|_{x=1}.$

Proof. Both of the identities are proved by setting $x = 1$ in (3.6.2), (3.6.3), (3.6.4), and (3.6.5), and equating the results pairwise. Equations (3.6.2) and (3.6.3) give (1); equations (3.6.4) and (3.6.5) give (2). $\square$

Each of the identities in Proposition 3.6.1 turns out to be equivalent to the fact that a certain pair of operators are adjoints. By an obvious extension of Theorem 3.3.1, we get that part (1) of Proposition 3.6.1 is essentially equivalent to

$$\begin{aligned} C(w)A(B(w)) &= \langle A(v) | \psi_{e^t} \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0 \rangle \\ &= \langle C(v)A(B(v)) | \psi_{e^t} \circ \chi_t \circ \xi_1 \circ e^{wE} z_0 y_0 \rangle, \end{aligned}$$

which in turn gives

$$\langle A(v) | C(w) e^{xB(w)} \rangle = \langle C(v)A(B(v)) | e^{xw} \rangle. \tag{3.6.6}$$

We can therefore easily calculate the adjoint to $C(v) \circ S_B$ (which acts by $C(v) \circ S_B(A(v)) = C(v)A(B(v))$).

We now recall the definition of certain umbral operators or Sheffer operators, cf. Section 3.5 in [Rm1]. More particularly, the Sheffer operator of $(B(t), C(t))$ in this work is the same as the Sheffer operator for $(1/C(B(t)), \overline{B}(t))$ in [Rm1].

Definition 3.6.1. We define the Sheffer operator of $(B(t), C(t))$ to be the unique linear map $\theta_{C,B} : \mathbb{C}[x] \mapsto \mathbb{C}[x]$ satisfying:

$$\theta_{C,B} e^{xw} = C(w) e^{xB(w)}.$$

Theorem 3.6.1. *We have that $\theta_{C,B}$ and $C(v) \circ S_B$ are adjoint operators.*

Proof. The result follows from Proposition 3.3.1 and (3.6.6). $\square$

Further, by an obvious extension of Theorem 3.3.1 we get that part (2) of Proposition 3.6.1 is essentially equivalent to

$$\begin{aligned} & C'(w)A(B(w)) + C(w)A'(B(w))B'(w) \\ &= \frac{\partial}{\partial w} \langle A(v) | \psi_{e^t} \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0 \rangle \\ &= \langle C'(\overline{B}(v)) C^{-1}(\overline{B}(v)) A(v) + B^*(v) A'(v) | \psi_{e^t} \circ \chi_B \circ \xi_C \circ e^{wE} z_0 y_0 \rangle, \end{aligned}$$

which in turn gives

$$\langle A(v)|\frac{\partial}{\partial w}C(w)e^{xB(w)}\rangle = \langle C'(\overline{B}(v))C^{-1}(\overline{B}(v))A(v) + B^*(v)A'(v)|C(w)e^{xB(w)}\rangle. \tag{3.6.7}$$

We now recall the definition of "Sheffer shifts"; cf. Section 3.7 in [Rm1]. More particularly, the Sheffer shift of $(B(t), C(t))$ in this work is the same as the Sheffer shift for $(1/C(B(t)), \overline{B}(t))$ in [Rm1].

Definition 3.6.2. For each $B(t)$ and $C(t)$, let $D_{B,C} : \mathbb{C}[x] \to \mathbb{C}[x]$ be the unique linear map satisfying

$$D_{B,C}C(w)e^{xB(w)} = \frac{\partial}{\partial w}\left(C(w)e^{xB(w)}\right).$$

We call $D_{B,C}$ the *Sheffer shift* of $(B(t), C(t))$.

Theorem 3.6.2. *We have that $D_{B,C}$ and $C'(\overline{B}(v))C^{-1}(\overline{B}(v)) + B^*(v)\frac{d}{dv}$ are adjoint operators.*

Proof. The result follows from Proposition 3.3.1 and (3.6.7). ☐

In closing this section, we note obvious characterizations of the Sheffer operators and Sheffer shifts in terms of the coefficients of their generating function definitions. For this it is convenient for us to recall the definition of Sheffer sequences, cf. Section 2.3 and Theorem 2.3.4 in particular in [Rm1] (see also Proposition 3.4.4 and Remark 3.4.4 in this work). We note that the Sheffer sequence of $(B(t), C(t))$ in this work is the same as the Sheffer sequence for $(1/C(B(t)), \overline{B}(t))$ in [Rm1].

Definition 3.6.3. We define the sequence of polynomials $S_n(x)$, the Sheffer sequence of $(B(t), C(t))$, to be the unique sequence satisfying the following:

$$C(w)e^{x(B(w))} = C(w)\sum_{n\geq 0}\frac{x^nB(w)^n}{n!} = \sum_{n\geq 0}\frac{S_n(x)w^n}{n!}.$$

Proposition 3.6.2. *We have that $\theta_{C,B}$, the Sheffer operator of $(B(t), C(t))$, is characterized as the unique linear map satisfying:*

$$\theta_{C,B}x^n = S_n(x),$$

where $S_n(x)$ is the Sheffer sequence of $(B(t), C(t))$.

□

Proposition 3.6.3. *We have that $D_{B,C}$, the Sheffer shift of $(B(t), C(t))$, is characterized as the unique linear map satisfying:*

$$D_{B,C}S_n(x) = S_{n+1}(x),$$

where $S_n(x)$ is the Sheffer sequence of $(B(t), C(t))$.

□

3.7 The exponential Riordan group

In this section we recall that the Sheffer operators form a group and we give a matrix representation of this group which, following the authors of [SGWW], we shall call the exponential Riordan group. We have

$$\theta_{G,F} \circ \theta_{C,B} e^{xw} = \theta_{G,F} C(w) e^{xB(w)} = C(w)G(B(w))e^{xF(B(w))},$$

where we treat $B(t)$ as a single, atomic, formal object to get the second equality. Thus

$$\theta_{G,F} \circ \theta_{C,B} = \theta_{(C)(G(B)),F(B)}.$$

Using this, the following theorem is easily verified.

Theorem 3.7.1. *The Sheffer operators form a group whose multiplication is given by:*

$$\theta_{G,F} \circ \theta_{C,B} = \theta_{(C)G(B),F(B)},$$

with identity element $\quad \theta_{1,t},$

and inverses given by $\quad \theta^{-1}_{C,B} = \theta_{(C(\overline{B}))^{-1},\overline{B}}.$

□

Remark 3.7.1. Theorem 3.7.1 appeared as Theorem 3.5.2 in [Rm1].

Since the Sheffer operators are linear maps, it is clear that we should be able to obtain a matrix representation of them. We do this next.

Definition 3.7.1. Let $M_{C,B}$ be the infinite matrix whose entry in the n-th row and m-th column $(m, n \geq 0)$, $M_{C,B}(n, m)$, is the coefficient of the degree m monomial in $\theta_{C,B} x^n$.

Remark 3.7.2. We (perhaps eccentrically) included a 0-th row and column in Definition 3.7.1 for notational convenience in order to naturally match the row and column indices with the corresponding polynomial degrees.

Theorem 3.7.2. *The map sending $\theta_{C,B}$ to $M_{C,B}$ is a group anti-isomorphism where the elements $M_{C,B}$ multiply as matrices. In particular, the set of matrices $M_{C,B}$ for all $C(t)$ and $B(t)$ forms a group under matrix multiplication.*

Proof. It is obvious that we need only show that $M_{C,B} M_{G,F} = M_{(C)(G(B)),F(B)}$. We have

$$\begin{aligned}
\sum_{n,m \geq 0} \frac{M_{(C)(G(B)),F(B)}(n,m) x^m w^n}{n!} &= \theta_{((C)(G(B)),F(B))} e^{xw} \\
&= \theta_{G,F} \circ \theta_{C,B} e^{xw} \\
&= \theta_{G,F} \sum_{n,l \geq 0} \frac{M_{C,B}(n,l) x^l w^n}{n!} \\
&= \sum_{n,l,m \geq 0} \frac{M_{C,B}(n,l) M_{G,F}(l,m) x^m w^n}{n!}.
\end{aligned}$$

so that equating coefficients of x and w we have

$$\sum_{l \geq 0} M_{C,B}(n,l) M_{G,F}(l,m) = M_{(C)(G(B)),F(B)}(n,m),$$

which is what we wanted. □

The group of matrices in Theorem 3.7.2 is the exponential Riordan group and essentially this result (or really a very closely related one) was the starting point for the combinatorial study of [SGWW] (see also Theorem 3.5.5 in [Rm1]).

The following corollary is now routine.

Corollary 3.7.1. *The map sending $\theta_{C,B}$ to $M_{(C(\overline{B}))^{-1},\overline{B}}$ is a group isomorphism where the $M_{C,B}$ multiply as matrices.*

□

The reader may have noted that the proof of Theorem 3.7.2 really did not depend on the full structure of the exponential Riordan group, but rather only on the fact that the Sheffer operators are linear invertible maps. Indeed, we may provide an obvious general statement as follows.

Proposition 3.7.1. *Let L be the set of invertible linear operators on $\mathbb{C}[x]$. For $\theta \in L$, let M_θ be the infinite matrix whose entry in the n-th row and m-th column ($m, n \geq 0$), $M_\theta(n,m)$, is the coefficient of the degree m monomial in $\theta(x^n)$. Then the map ϕ sending θ to M_θ is a group anti-isomorphism where the group product on L is composition and the group product on $\phi(L)$ is matrix multiplication.*

Proof. This is guided substantially by the proof of Theorem 3.7.2. Let $\theta_1, \theta_2 \in L$. Then we may write

$$\theta_1 x^n = \sum_{m\geq 0} \theta_1(n,m)x^m \qquad \theta_1(n,m) \in \mathbb{C} \qquad \text{and}$$

$$\theta_2 x^n = \sum_{m\geq 0} \theta_2(n,m)x^m \qquad \theta_2(n,m) \in \mathbb{C},$$

for all $n \geq 0$. Then we have

$$\begin{aligned}\theta_1 \circ \theta_2 x^n &= \theta_1 \sum_{m\geq 0} \theta_2(n,m)x^m \\ &= \sum_{m,l\geq 0} \theta_2(n,m)\theta_1(m,l)x^l \\ &= \sum_{l\geq 0}\left(\sum_{m\geq 0} \theta_2(n,m)\theta_1(m,l)\right) x^l.\end{aligned}$$

Thus it is clear that $\phi(\theta_1 \circ \theta_2) = \phi(\theta_2)\phi(\theta_1)$ which is what we wanted to show. □

Remark 3.7.3. We could, of course, have left off the assumption of invertibility in the statement of Proposition 3.7.1 and still had a monoid anti-isomorphism.

We close this section with some observations on how we defined the group of Sheffer operators, a subgroup of L. We defined the Sheffer operators by an action on certain types of generating functions, namely, those of the form $C(t)e^{xB(t)}$, which for convenience, let us call Sheffer generating functions. We note the following two propositions.

Proposition 3.7.2. *The representation $C(t)e^{xB(t)}$ of Sheffer generating functions is unique.*

Proof. Let $C_1(t)e^{xB_1(t)}, C_2(t)e^{xB_2(t)}$ be Sheffer generating functions and let $C_1(t)e^{xB_1(t)} = C_2(t)e^{xB_2(t)}$. Then equating the coefficients of the 0-th degree and 1st degree terms in x gives, respectively, that $C_1(t) = C_2(t)$ and $C_1(t)B_1(t) = C_2(t)B_2(t)$ and since $C_1(t), C_2(t)$ have non zero constant term we may cancel them from the second equation. □

Proposition 3.7.3. *The action of the group of Sheffer operators on the Sheffer generating functions is simply transitive.*

Proof. Let $C_1(t)e^{xB_1(t)}, C_2(t)e^{xB_2(t)}$ be Sheffer generating functions. We need a Sheffer operator $\theta_{C,B}$ such that $\theta_{C,B}\left(C_1(w)e^{xB_1(w)}\right) = C_2(w)e^{xB_2(w)}$. We have

$$\theta_{C,B}\left(C_1(w)e^{xB_1(w)}\right) = C_1(w)C(B_1(w))e^{xB(B_1(w))},$$

so that we must have

$$C_1(t)C(B_1(t)) = C_2(t) \quad \text{and} \quad B(B_1(t)) = B_2(t),$$

which in turn is uniquely solved by letting

$$C(t) = C_2(\overline{B_1}(t))C_1^{-1}(\overline{B_1}(t)) \quad \text{and} \quad B(t) = B_2(\overline{B_1}(t)).$$

□

3.8 Matrix multiplication via generating functions

Recall that the "exponential" form of the Sheffer generating functions arose because of Theorem 3.3.1. We never really made use of the exponential structure of the Sheffer generating functions in the previous section. In this section we will show how we may use the structure of more general generating functions to, in a precise sense, perform certain matrix multiplication and see that an ansatz (or two) leads us back to the exponential form so that we may apply the formal Taylor theorem and thereby return to, from a different starting point, the Sheffer generating functions and the exponential Riordan group.

It is a natural question (and one which is independent of all the previous work) to start with countably infinite matrices whose rows and columns are indexed by the nonnegative integers and ask how can we keep track of matrix multiplication by considering certain corresponding generating functions. Consider two 2-variable generating functions

$$\begin{aligned} f(t,x) &= \sum_{n,m\geq 0} f(n,m)\frac{t^n x^m}{a_n b_m} \qquad \text{and} \\ g(t,x) &= \sum_{n,m\geq 0} g(n,m)\frac{t^n x^m}{a_n b_m}, \end{aligned}$$

where $a_n, b_n \in \mathbb{C} \setminus 0$ for all $n \geq 0$. We generically name two such sequences of nonzero complex numbers by a and b, and we say that $f(t,x)$ is a generating function of type (a,b).

Notation 3.8.1. Let $M(\mathbb{C})$ be the space of countably infinite matrices with complex entries so that $(M(n,m))$ where $n,m \geq 0$ is the matrix whose entry in the n-th row and m-th column is $M(n,m)$.

We say that we can multiply two matrices if the sum for each entry in the multiple has only finitely many nonzero terms. That is

Definition 3.8.1. We say two matrices $(M(n,m))$ and $(N(n,m))$ are *left-right multipliable* if for all $m,n \geq 0$ we have that $M(n,l)N(l,m)$ is nonzero for only finitely many $l \geq 0$ and then the matrix multiple is

$$\begin{aligned} (M(n,m))(N(n,m)) &= (P(n,m)), \\ \text{where } P(n,m) &= \sum_{l\geq 0} M(n,l)N(l,m). \end{aligned}$$

Definition 3.8.2. We say two matrices $(M(n,m))$ and $(N(n,m))$ are *strongly left-right multipliable* if for any fixed n,m and $k \geq 0$ we have that $M(n,l)N(l+k,m)$ is nonzero for only finitely many $l \geq 0$.

Of course if the left matrix is lower triangular then the matrices are strongly left-right multipliable and this is the only case in which we are ultimately interested.

We may associate with each generating function which is of type (a,b) a matrix, via the map

$$\alpha_{a,b} : \mathbb{C}[[t,x]] \mapsto M_n(\mathbb{C})$$
$$\alpha_{a,b}(f(t,x)) = (f(n,m)).$$

Now we ask whether there is a nice expression in terms of the generating functions which will give us the generating function of the matrix multiple if that multiple exists. Consider

$$f(t,v)g(u,x) = \sum_{n,m,l,p\geq 0} f(n,m)g(l,p)\frac{t^n v^m u^l x^p}{a_n b_m a_l b_p}.$$

A little thought might indicate replacing v with some u exponent lowering operator like $\frac{\partial}{\partial u}$ and then setting $u = 0$. Assuming all coefficient sums are finitely computable, or equivalently that $\alpha(f(t,x))$ and $\alpha(g(t,x))$ are strongly left-right multipliable, we get

$$\begin{aligned}
&\left(f\left(t, \frac{\partial}{\partial u}\right) g(u,x)\right) |_{u=0} \\
&\quad = \left(\sum_{n,m,l,p\geq 0} f(n,m)g(l,p)\frac{t^n}{a_n}\frac{(l)(l-1)\cdots(l-m+1)u^{l-m}x^p}{b_m a_l b_p}\right) |_{u=0} \\
&\quad = \sum_{n,m,p\geq 0} f(n,m)g(m,p)\frac{m! t^n x^p}{a_n b_m a_m b_p} \\
&\quad = \sum_{n,p\geq 0}\left(\sum_{m\geq 0}\frac{m!}{a_m b_m} f(n,m)g(m,p)\right)\frac{t^n x^p}{a_n b_p}.
\end{aligned}$$

The following result is now clear.

Proposition 3.8.1. *When $a_m b_m = m!$ and when $\alpha_{a,b}(f(t,x))$ and $\alpha_{a,b}(g(t,x))$ are strongly left-right multipliable we have:*

$$\alpha_{a,b}(f(t,x))\alpha_{a,b}(g(t,x)) = \alpha_{a,b}\left(\left(f\left(t,\frac{\partial}{\partial u}\right)g(u,x)\right)|_{u=0}\right).$$

□

So in some sense we know how to multiply certain pairs of matrices with generating functions. This motivates the following definition of a partial multiplication on generating functions.

Definition 3.8.3. Let $f(t,x), g(t,x) \in \mathbb{C}[[t,x]]$. We define the following partial multiplication.

$$f(t,x) * g(t,x) = \left(f\left(t, \frac{\partial}{\partial u}\right) g(u,x) \right) |_{u=0}.$$

Of course the product $f(t,x)*g(t,x)$ is well-defined exactly when for some sequences a, b, we have that $\alpha_{a,b}(f(t,x))$ and $\alpha_{a,b}(g(t,x))$ are strongly left-right multipliable.

In light of the formal Taylor theorem, an ansatz leads us to consider the set of Sheffer generating functions under $*$ multiplication. We may calculate to get the following.

$$\begin{aligned} C(t)e^{xB(t)} * G(t)e^{xF(t)} &= \left(C(t)e^{\frac{\partial}{\partial u}B(t)} \left(G(u)e^{xF(u)} \right) \right) |_{u=0} \qquad (3.8.1) \\ &= \left(C(t)G(u+B(t))e^{xF(u+B(t))} \right) |_{u=0} \\ &= C(t)G(B(t))e^{xF(B(t))}. \qquad (3.8.2) \end{aligned}$$

The following result is now routine.

Theorem 3.8.1. *The set of Sheffer generating functions forms a group under $*$ multiplication, where the identity element is e^{xt} and the inverse of $C(t)e^{xB(t)}$ is given by $C(\overline{B}(t))^{-1}e^{x\overline{B}(t)}$. In addition, the map ψ sending $\theta_{C,B}$ to $\theta_{C,B}^{-1}e^{xt}$ is a group isomorphism.*

□

Further, for any sequences a, b such that $a_m b_m = m!$ we have by Proposition 3.8.1 that $\alpha_{a,b}$ restricted to the Sheffer generating functions is a group isomorphism where the image of $\alpha_{a,b}$ multiply as matrices. The matrix group arising from setting $a_m = m!$ and $b_m = 1$ leads us back to the matrix group encountered in the previous section, the exponential Riordan group. The case arising from setting $a_m = 1$ and $b_m = m!$ gives what the authors of [SGWW] called the Riordan group.

Remark 3.8.1. Of course, it is necessary for $C(t)$ to have a reciprocal and $B(t)$ to have a compositional inverse in order to guarantee inverse elements. However, there are obvious relaxed restrictions yielding perfectly good multiplications. For instance, we need impose no restrictions on the two power series of the right matrix thus giving a module for the various types of Riordan group. Moreover we could further restrict the multiplication yielding submodules and subgroups. For instance, we may consider only those Sheffer generating functions $C(t)e^{xB(t)}$ with $C(t) = 1$. This yields as one example what the authors of [SGWW] called the "associated subgroup" of the Riordan group. For more discussion on this and other subgroups we refer the reader to [Sh].

3.9 Some underlying combinatorics

We shall recall a combinatorial question which will lead us to the matrix groups we have been considering. This section is included for two reasons. The first reason is the convenience of the reader, who will be able to have, briefly presented, a small piece of the underlying combinatorics without needing to find another source. The second is that it was through a combinatorial interpretation of the group multiplication, a description of which shall conclude this section, that I originally realized the connection between umbral calculus and some calculations I was investigating, which themselves were originally motivated by the logarithmic formal calculus as developed in [Mi], and much further developed in [HLZ] for the study of logarithmic tensor product theory for vertex operator algebra modules. This section is expository. There is a vast literature on applying generating functions to counting problems. Rather than try to give any sort of comprehensive list of references, we shall content ourselves to refer the reader to two standard books which discuss the "exponential formula," which is closely related to the subject of our present interest. In Chapter 3 of [W1] the author gives a nice, colorful and useful treatment based on ideas which he writes were suggested to him by Adriano Garsia. A second treatment is given in the first Chapter of [St].

We shall assume little to no experience with generating function techniques in combinatorics and begin with a simple counting problem: What is the number of ways of partitioning L distinct objects into I subsets of sizes 1 and $n+1$, $n \geq 1$, which we may color by choosing from a (pre-)fixed sequence of B_m colors for each subset of size m? That is, we are placing L distinct balls into I bins where each bin may contain either 1 ball or $n+1$ balls and then we color each bin containing 1 ball any of B_1 colors and we color each bin containing $n+1$ balls any of B_{n+1} colors.

Fixing I, it is clear that the answer is 0 unless $L = I + \alpha n$ for some $0 \leq \alpha \leq I$, where we partition the original L objects using exactly α subsets of size $n+1$. Then choosing $n+1$ objects at a time from the original set we get

$$\binom{I+\alpha n}{n+1}\binom{I+(\alpha-1)n-1}{n+1}\cdots\binom{I+n-(\alpha-1)}{n+1}$$

ways of partitioning up the original set where the order we chose the sets of size $n+1$ matter, where we have exactly α subsets of size $n+1$ and where we have ignored the colors. Notice that this product telescopes to give

$$\frac{(I+\alpha n)!}{((n+1)!)^{\alpha}(I-\alpha)!}.$$

The following proposition is now clear.

Proposition 3.9.1. *The number of ways to partition L objects into α subsets of size $n+1$ and $I-\alpha$ subsets of size 1 where we are allowed to color the subsets of size m with B_m colors is*

$$\frac{(I+\alpha n)!}{((n+1)!)^{\alpha}(I-\alpha)!\alpha!}B_1^{I-\alpha}B_{n+1}^{\alpha}.$$

□

We can give an alternate proof that does not use the telescoping trick.

Proof. (Second proof of Proposition 3.9.1) To each ordering of $L = I + \alpha n$ objects, we may associate a partition into α fixed subsets of size $n+1$ and $I-\alpha$ subsets of size 1 by simply grouping the first $n+1$ elements in the ordering and then the next $n+1$, etc., until α groupings have been made. This gives a many-to-one association which partitions the original arrangements into equivalence classes by the obvious pullbacks. More precisely, we may transform the original ordering, while preserving the relevant structure, by 1) rearranging any of the $n+1$ groupings separately, by 2) rearranging the order inside each of the $n+1$ groupings separately and by 3) rearranging the order of the 1 groupings. Thus beginning with a representative original arrangement we may by applying one of the three arrangement procedures transform it into any other element of its pullback partition class. It is easy to see now that each class has the same size which is

$$(n+1)!^{\alpha}(I-\alpha)!\alpha!.$$

Thus we may simply divide $L! = (I+\alpha n)!$ by this class size and take into account the colors to get the desired result since the number of pullback equivalence classes is precisely the number of partitions we are trying to count. □

We notice that our result in Proposition 3.9.1 looks distinctly like a term in a binomial expansion. With a little thought, "hanging" the result on a certain exponential generating function makes this observation precise. We have

$$\begin{aligned}
f(t) &= \sum_{0\le\alpha\le I} \frac{(I+\alpha n)!}{((n+1)!)^{\alpha}(I-\alpha)!\alpha!} B_1^{I-\alpha} B_{n+1}^{\alpha} \frac{t^{I+\alpha n}}{(I+\alpha n)!} \\
&= \sum_{0\le\alpha\le I} \frac{(I+\alpha n)!}{(I-\alpha)!\alpha!} B_1^{I-\alpha} \left(\frac{B_{n+1}}{(n+1)!}\right)^{\alpha} \frac{t^{I+\alpha n}}{(I+\alpha n)!} \\
&= \sum_{0\le\alpha} \frac{1}{I!}\binom{I}{\alpha} B_1^{I-\alpha} \left(\frac{B_{n+1}}{(n+1)!}\right)^{\alpha} t^{I+\alpha n} \\
&= \frac{1}{I!}\sum_{0\le\alpha} \binom{I}{\alpha} (B_1 t)^{I-\alpha} \left(\frac{B_{n+1}t^{n+1}}{(n+1)!}\right)^{\alpha} \\
&= \frac{1}{I!}\left(B_1 t + \frac{B_{n+1}t^{n+1}}{(n+1)!}\right)^{I},
\end{aligned}$$

where $f(t)$ is the exponential generating function whose "coefficient" of $t^L/L!$ is exactly the number of ways of partitioning L objects into I pieces of size 1 and $n+1$ whose subsets of size m may be colored with B_m colors. We now make another algebraic observation that $f(t)$ looks like a term in the Taylor expansion of the exponential function and again hang our object, now on a two-variable generating function, exponential in t and ordinary in x, giving the following.

Proposition 3.9.2. *The number of ways to partition L distinct objects into I subsets of sizes 1 and $n+1$ where we may color the subsets of size m any of B_m colors is the* *coefficient of* $\frac{t^L}{L!}x^I$ *of*

$$e^{\left(B_1 t + \frac{B_{n+1}t^{n+1}}{(n+1)!}\right)x}.$$

$\square$

We can easily generalize this whole procedure. Let us take L objects and count how many ways there are to partition them into I subsets of any sizes where we may color a subset of size m with B_m colors. Given a partition with α_n subsets of size n the second proof of Proposition 3.9.1 obviously generalizes to give that the number of partitions is

$$\frac{\left(\sum_{n\geq 1} n\alpha_n\right)! \prod_{n\geq 1} B_n^{\alpha_n}}{\prod_{n\geq 1}(n!)^{\alpha_n} \prod_{n\geq 1} \alpha_n!}.$$

Next, in parallel to the original special case using the multinomial expansion in place of simply a binomial expansion we get

$$\begin{aligned} f(t) &= \sum_{\sum_{n\geq 1}\alpha_n = I} \frac{\left(\sum_{n\geq 1} n\alpha_n\right)! \prod_{n\geq 1} B_n^{\alpha_n}}{\prod_{n\geq 1}(n!)^{\alpha_n} \prod_{n\geq 1} \alpha_n!} \frac{t^{\sum_{n\geq 1} n\alpha_n}}{\left(\sum_{n\geq 1} n\alpha_n\right)!} \\ &= \frac{1}{I!} \sum_{\sum_{n\geq 1}\alpha_n = I} \frac{I!}{\prod_{n\geq 1}\alpha_n!} \prod_{n\geq 1} \left(\frac{B_n t^n}{n!}\right)^{\alpha_n} \\ &= \frac{1}{I!} \left(\sum_{n\geq 1} \frac{B_n}{n!} t^n\right)^I. \end{aligned}$$

Thus we have the following.

Proposition 3.9.3. *The number of ways to partition L distinct objects into I subsets of any sizes where we may color the subsets of size n any of B_n colors is the coefficient of $\frac{t^L}{L!}x^I$ of $e^{xB(t)}$, where we do not require $B_1 \neq 0$.*

Obviously we can see that attached Sheffer sequences are playing a role here. Indeed, we have interpreted the entries of the matrices of the associated subgroup of the exponential Riordan group which have nonnegative integral entries. Actually, we have done more, since we allow for $B_1 \neq 0$, thus interpreting the matrices with non-negative entries as a module of the associated subgroup of the exponential Riordan group. These matrices are not invertible, of course.

It is natural to ask whether we can interpret any of our algebraic operations concerning Sheffer sequences combinatorially. For instance, how may we interpret the matrix multiplication of the associated subgroup of the exponential Riordan group? A little thought shows that matrix multiplication corresponds to partitioning L objects in two stages. Let us initially ignore the colors. First partition the L objects into J subsets and then partition the J subsets, now considered as objects themselves, into I objects. Of course, this yields a partition of L into I objects. Moreover, by summing over J one gets, at least once, all the ways to partition L objects into I subsets. Thus all we need to show combinatorially is that given two coloring schemes there is indeed a third coloring scheme that arises from this two stage partitioning. Given a subset of size n in the final partition it could have come from a number of intermediate sub-partitions, and these are independent of how any of the other original elements that do not end up in the final subset we are focusing on were arranged. This independence is exactly what we need though to show that the interpretation is, in fact, correct. That is, let the (pre-)fixed lists of numbers of possible colors be $B1_n$ for the left matrix, $B2_n$ for the right matrix and $B3_n$ for the product. Then it is clear that $B3_n$ is equal to the number of ways to partition n distinct objects into any size subsets with those subsets being colored according to the list $B1_n$, with each partition of size m further striped by any of $B2_m$ colors.

Of course, to consider inverses combinatorially we would have to interpret matrices with negative entries. Let us return to thinking of placing balls into bins, except now make the balls into cylinders which can be placed right-side up or upside down. Further, we make our bins such that the cylinders fit "snugly" in the up-and-down direction so that if we turn a bin upside down all the cylinders get turned upside down too, but so that the cylinders can be moved past one another inside each bin. (However, we cannot tell whether the bin itself is right side up or upside down). After we have divided the balls up, but before we color the bins, we line up all the bins next to each other. Then we begin our coloring as usual except that if we are coloring a bin of size m and B_m is negative we turn all of the cylinders in all of the bins upside down and then continue (of course if B_m is positive we only color the bin without any turning and if $B_m = 0$ we "scrap" the whole configuration). If we come to a second negative color we turn all the cylinders right side up again etc. Then we ask for the difference between right side up configurations and upside down configurations of placing L cylinders into I colored bins and this result is what $e^{xB(t)}$ does in fact keep track of as the coefficient of $\frac{t^L}{L!}x^I$. Then, in the two stage partitioning we first place the cylinders into bins coloring and turning them according to the entries of left matrix then we place these bins into bigger bins into which they fit "snugly" coloring and turning the larger bins according to the entries of the right matrix. We then ask for the difference between right side up configurations and upside down configurations of placing L cylinders into I colored, doubly layered bins after this two stage process and this result corresponds to the matrix multiple. The resulting (signed) number of colors allowed for bins of size m for the iterated procedure is the difference of the number of right-side up configurations and the number of upside down configurations of doubly layered bins with m cylinders where there is only one outer bin resulting from the above described two stage partition.

As an example, we note that if $B_n = 1$ for all n then the coefficient of $\frac{t^L}{L!}x^I$ from $e^{xB(t)}$ give the number of ways to partition L distinct objects into I subsets, the ubiquitous *Stirling numbers of the second kind.* If $B_n = (n-1)!$ for all n then the coefficient of $\frac{t^L}{L!}x^I$ from $e^{xB(t)}$ give the number of ways to partition L distinct objects into I cycles, the also ubiquitous *signless Stirling numbers of the first kind.* If we multiplied the matrices corresponding to the Stirling numbers of the second kind and signless Stirling numbers of the first kind together, with the Stirling numbers of the second kind giving the left matrix, then the new matrix would have "colors" B_n determined by the number of ways to divide n balls into an annular bin with up to and including n loose fitting (so that the balls could move past one another) compartments. If we switched the order of the matrices, the product matrix would have "colors" B_n determined by the number of ways to divide n balls into up to and including n cycles.

Now instead of considering the signless Stirling numbers of the first kind, let us consider the case where $B_n = (-1)^{n-1}(n-1)!$, which gives the *Stirling numbers of the first kind.* Now even cycles are upside down and get counted as negative. So if we multiply this on the left by the Stirling numbers of the Second kind the new matrix would have “colors” B_n determined by the difference between the number of ways to divide n balls into an annular bin with an odd number of loose fitting compartments (up to and including n) and into an annular bin with an even number of loose fitting compartments (up to and including n). If we switch the order of the matrices, the new matrix would have “colors” B_n determined by the following. If n is odd we get difference between the number of ways to divide n objects into an odd number of cycles and an even number of cycles. If n is even we get the difference between the number of ways to divide n objects into an even number of cycles and an odd number of cycles. Or in other more algebraic words, we get the n-th row sum of the matrix of Stirling numbers of the first kind. It is well-known that in both cases the answer is the same, and that the new matrix will have $B_1 = 1$ and $B_n = 0$ for $n \geq 1$, so that the Stirling numbers of the first kind and the Stirling numbers of the second kind give matrix inverses. We have thus shown the equivalence of this matrix inverse property with two combinatorial properties and also the algebraic property that the row sums of the Stirling numbers of the first kind are 0 except for the first row which has row sum 1. Of course, to verify the truth of all these statements is algebraically very easy, since it obviously follows from Theorem 3.7.1 which gives

$$\theta_{1,e^x-1}\theta_{1,\log(1+x)} = \theta_{1,e^{\log(1+x)}-1} = \theta_{1,x}.$$

Remark 3.9.1. The equivalent statements of the matrix inverse property of the Stirling numbers of the first and second kinds are, of course, all obtained by identifying the Stirling numbers with elements of the exponential Riordan group and then exploiting the fact that each matrix of this group is fully determined by its leading column.

It is tempting to continue playing around with these sort of combinatorial amusements but perhaps with one final exception we shall leave any further investigations to the reader who will find quite a lot already written on this fascinating subject reaching far beyond our brief observations. The last point that I wish to make is only a remark with a question. It is not difficult to see that the elements of the exponential Riordan group may be represented by infinite products of elements such that the n-th element of the product has 0-s in all entries of the leading column except the first and n-th. The question is what is a natural combinatorial interpretation explaining this representation and more generally are there connections to other work on the subject. I, of course, ask this question merely as an interested non-specialist and may simply be unaware of a standard well-known answer.

3.10 The Virasoro algebra

Our goal in this section will be to show that an operator closely related to the derivations D and E (which appeared in Sections 3.2 and 3.6 respectively) is one of the standard quadratic representations of the $L(-1)$ operator of the Virasoro algebra.

Recall that we began our main investigation by calculating the higher derivatives of the composition of two formal power series $f(x)$ and $g(x)$, where the constant term of $g(x)$ was required to be 0, following a proof given in [FLM2]. In fact, the case that interested the authors in [FLM2] was when $f(x) = e^x$. It is not difficult to specialize our arguments to this case. When we abstract, we get the following set-up: Consider the vector space $y\mathbb{C}[x_1, x_2, x_3, \dots]$ where x_j for $j \geq 1$ are commuting formal variables. Then let $\mathcal{D}$ be the unique derivation on $y\mathbb{C}[x_1, x_2, x_3, \dots]$ satisfying

$$\begin{aligned} \mathcal{D}y &= yx_1 \\ \mathcal{D}x_j &= x_{j+1} \quad j \geq 1. \end{aligned} \tag{3.10.1}$$

The question of calculating $e^{w\frac{d}{dx}}e^{g(x)}$ is seen to be essentially equivalent to calculating

$$e^{w\mathcal{D}}y,$$

where we "secretly" identify $\mathcal{D}$ with $\frac{d}{dx}$, $e^{g(x)}$ with y and $g^{(m)}(x)$ with x_m. It is clear by our identification, and rigorously as an easy corollary of Proposition 3.2.1, that:

$$e^{w\mathcal{D}}y = \sum_{n\geq 0} \frac{y\left(\sum_{m\geq 1} \frac{w^m x_m}{m!}\right)^n}{n!} = ye^{\sum_{m\geq 1} \frac{w^m x_m}{m!}}. \tag{3.10.2}$$

We note that

$$\begin{aligned} \mathcal{D} &= x_1 y\frac{\partial}{\partial y} + x_2\frac{\partial}{\partial x_1} + x_3\frac{\partial}{\partial x_2} + \cdots \\ &= x_1 y\frac{\partial}{\partial y} + \sum_{k\geq 1} x_{k+1}\frac{\partial}{\partial x_k}. \end{aligned}$$

We next shall switch gears in order to recall certain basics about the Virasoro algebra using operators arising from certain Heisenberg Lie algebras. We follow (a variant of) the exposition of this well-known material in [FLM2]. Let $\mathfrak{h}$ be the one-dimensional abelian (complex) Lie algebra with basis element h. We define a nonsingular symmetric bilinear form on $\mathfrak{h}$ by $(ah, bh) = ab$ for all $a, b \in \mathbb{C}$. We recall the (particular) affine Heisenberg Lie algebra $\widehat{\mathfrak{h}}$ which is the vector space

$$\widehat{\mathfrak{h}} = \mathfrak{h} \otimes \mathbb{C}[t, t^{-1}] \oplus \mathbb{C}c,$$

with Lie brackets determined by

$$[ah \otimes t^m, bh \otimes t^n] = (ah, bh)m\delta_{m+n,0}c = abm\delta_{m+n,0}c,$$

where c is central and δ is the Kronecker delta.

We may realize $\widehat{\mathfrak{h}}$ as differential and multiplication operators on a space with infinitely many variables as follows. We consider the space $y\mathbb{C}[x_1, x_2, x_3, \dots]$ and make the following identification:

$$h \otimes t^n = \begin{cases} \alpha(-n)x_{-n} & n < 0 \\ \beta(n)\frac{\partial}{\partial x_n} & n > 0 \\ y\frac{\partial}{\partial y} & n = 0, \end{cases}$$

where $\alpha(n), \beta(n) \in \mathbb{C}$ for $n \geq 1$ and we identify c, the central element, with the multiplication by identity operator. Of course, in this setting $y\frac{\partial}{\partial y}$ is a fancy name for the identity operator, but we wrote it this way so that it appears explicitly as a derivation. It is easy to see that

$$\left[\alpha(n)x_n, \beta(n)\frac{\partial}{\partial x_n}\right] = -\alpha(n)\beta(n),$$

with all other pairs of operators commuting. Thus our identification gives a representation of the Heisenberg Lie algebra exactly when we require that for all $n \geq 1$

$$\alpha(n)\beta(n) = n.$$

For this representation we shall sometimes use the notation $h(n)$ to denote the image of $h \otimes t^n$.

Definition 3.10.1. The Virasoro algebra of central charge 1 is the Lie algebra generated by basis elements 1, a central element, and $L(n)$ for $n \in \mathbb{Z}$ which satisfy the following relations for all $m, n \in \mathbb{Z}$.

$$[L(m), L(n)] = (m-n)L(m+n) + (1/12)(m^3 - m)\delta_{m+n,0}. \tag{3.10.3}$$

Remark 3.10.1. The Virasoro algebra is a central extension of the Witt algebra, the Lie algebra of the derivations of Laurent polynomials in a single formal variable. The correspondence can be seen by identifying $L(n)$ with $-t^{n+1}\frac{d}{dt}$. In fact, (cf. Proposition 1.9.4 in [FLM2]) the Virasoro algebra with a general central element is the unique, up to isomorphism, one dimensional central extension of the Witt algebra.

Theorem 3.10.1. *The operators*

$$L(n) = \frac{1}{2}\sum_{k\in\mathbb{Z}} h(n-k)h(k) \qquad n \neq 0 \qquad \textit{and} \tag{3.10.4}$$

$$L(0) = \frac{1}{2}\sum_{k\in\mathbb{Z}} h(-|k|)h(|k|) \tag{3.10.5}$$

give a representation of the Virasoro algebra of central charge 1*,*

Theorem 3.10.1 appears, for instance, as a special case of Theorem 1.9.6 in [FLM2]. We shall provide a proof in Section 3.13 for the convenience of the reader.

The space $y\mathbb{C}[x_1, x_2, x_3, \ldots]$ is obviously a module for the Virasoro algebra. It is graded by $L(0)$ eigenvalues, which are called *weights*. In the literature, such a module is often called a lowest weight module; this module has y as a lowest weight vector.

We have

$$\begin{aligned} L(0) &= \frac{1}{2}h(0)^2 + h(-1)h(1) + h(-2)h(2) + \cdots \\ &= \frac{1}{2}y\frac{\partial}{\partial y} \circ y\frac{\partial}{\partial y} + \alpha(1)\beta(1)x_1\frac{\partial}{\partial x_1} + \alpha(2)\beta(2)x_2\frac{\partial}{\partial x_2} + \cdots \end{aligned}$$

so that $L(0)y = \frac{1}{2}y$. Thus the lowest weight of the module is $\frac{1}{2}$.

We may now show that by an appropriate (unique) choice of $\alpha(n)$ and $\beta(n)$ we get

$\mathcal{D} = L(-1)$. We have

$$\begin{aligned} L(-1) &= h(-1)h(0) + h(-2)h(1) + h(-3)h(2) + \dots \\ &= \alpha(1)x_1 y\frac{\partial}{\partial y} + \alpha(2)\beta(1)x_2\frac{\partial}{\partial x_1} + \alpha(3)\beta(2)x_3\frac{\partial}{\partial x_2} + \dots. \end{aligned}$$

Therefore it is clear that in order to have $\mathcal{D} = L(-1)$, we need exactly that

$$\begin{aligned} \alpha(1) &= 1 \qquad \text{and} \\ \alpha(n+1)\beta(n) &= 1 \quad n \geq 1, \end{aligned}$$

where we recall that we already have the restriction that $\alpha(n)\beta(n) = n$ for all $n \geq 1$. These two sets of restrictions imply that

$$\begin{aligned} (n+1)\beta(n) &= \beta(n+1) \quad n \geq 1 \\ \beta(1) &= 1, \end{aligned}$$

so that

$$\begin{aligned} \beta(n) &= n! \\ \alpha(n) &= \frac{1}{(n-1)!}, \end{aligned}$$

for all $n \geq 1$, is the unique solution. We record this as a proposition.

Proposition 3.10.1. *The operator $L(-1)$, given by (3.10.4), with (and only with) both $\alpha(n) = \frac{1}{(n-1)!}$ and $\beta(n) = n!$, is identical to the operator $\mathcal{D}$, given by (3.10.1).*

□

For the remainder of this work we shall assume that $\alpha(n) = \frac{1}{(n-1)!}$ and $\beta(n) = n!$.

3.11 Umbral shifts revisited and generalized

We shall continue to consider the space $y\mathbb{C}[x_1, x_2, x_3, \dots]$ as in the previous section, and similarly to some of our previous work, such as in Section 3.2, we shall consider certain substitution maps. Let $\phi_{B(t)}$ denote the following algebra homomorphism.

$$\phi_B : y\mathbb{C}[x_1, x_2, x_3, \dots] \to \mathbb{C}[x]$$

with

$$\phi_{B(t)} x_j = B_j x \qquad j \geq 1$$

$$\text{and} \qquad \phi_{B(t)} y = 1.$$

Then we have

$$\phi_B \circ e^{w\mathcal{D}} y = \phi_B \circ e^{wL(-1)} y = e^{xB(w)}.$$

In light of Proposition 3.5.1, it is routine to show the following.

Theorem 3.11.1. *The attached Sheffer shift, $D_B : \mathbb{C}[x] \to \mathbb{C}[x]$ is the unique linear map satisfying*

$$D_B \circ \phi_B \circ L(-1)^n y = \phi_B \circ L(-1)^{n+1} y,$$

for all $n \geq 0$.

□

With Theorem 3.11.1 as motivation, we make the following definition.

Definition 3.11.1. For $m \geq -1$ we define the operators $\mathcal{D}_B(m) : \mathbb{C}[x] \to \mathbb{C}[x]$ to be the unique linear maps satisfying

$$\mathcal{D}_B(m) \circ \phi_B \circ L(-1)^n y = \phi_B \circ L(m) L(-1)^n y,$$

for all $n \geq 0$.

Of course, $\mathcal{D}_B(-1) = D_B$. These operators are well-defined because $\phi_B \circ L(-1)^n y$ has degree exactly n. In fact, $\phi_B \circ L(-1)^n y = B_n(x)$, the Sheffer polynomial attached to $B(t)$. We call the operators $\mathcal{D}_B(n)$ generalized attached Sheffer shifts.

We would like to use the Virasoro relations to help compute the generalized attached Sheffer shifts. We begin with the following lemma.

Lemma 3.11.1. *There exist rational numbers $f_m(n)$ such that*

$$L(m) L(-1)^n y = f_m(n) L(-1)^{n-m} y,$$

for all $m \geq -1$, $n \geq 0$ such that $n \geq m$.

Proof. We will need that $L(m)L(-1)^n y = 0$ when $m > n$. This is really due to the weights of the vectors, but in this work we shall proceed, in just this special case, with an elementary induction argument. We induct on n. For $n = 0$ this follows essentially because y is a lowest weight vector, but even without considering weights it is easy to directly see given the definition of the operators. By induction (used twice) we have

$$\begin{aligned} L(m)L(-1)^n y &= L(-1)L(m)L(-1)^{n-1}y + [L(m), L(-1)]L(-1)^{n-1}y \\ &= (m+1)L(m-1)L(-1)^{n-1}y \\ &= 0. \end{aligned}$$

We may now focus on the main argument. We establish the boundary cases. Letting $m = -1$ we easily check that $f_{-1}(n) = 1$. The other boundary is $m = n$. We shall use another intermediate induction to establish this case. Our base case then is $m = n = 0$ for which it is easy to check that $f_0(0) = 1/2$. We also have

$$\begin{aligned} L(n)L(-1)^n y &= L(-1)L(n)L(-1)^{n-1}y + [L(n), L(-1)]L(-1)^{n-1}y \\ &= (n+1)L(n-1)L(-1)^{n-1}y, \end{aligned}$$

so that inducting on n we get our result. Moreover we now have the recurrence

$$f_n(n) = (n+1)f_{n-1}(n-1) \qquad n \geq 1, \tag{3.11.1}$$

with, as we have seen, the boundary $f_0(0) = 1/2$.

For our main argument we induct on $m+n$. We have already checked the base case. We then have by induction and using the Virasoro relations that for the remaining cases $m \geq 0$ and $n > m$, we have

$$\begin{aligned} L(m)L(-1)^n y &= L(-1)L(m)L(-1)^{n-1}y + [L(m), L(-1)]L(-1)^{n-1}y \\ &= f_m(n-1)L(-1)^{n-m}y + (m+1)L(m-1)L(-1)^{n-1}y \\ &= \left(f_m(n-1) + (m+1)f_{m-1}(n-1)\right)L(-1)^{n-m}y. \end{aligned}$$

Therefore, not only do the values $f_n(m)$ exist but we have a recurrence for them

$$f_m(n) = f_m(n-1) + (m+1)f_{m-1}(n-1). \tag{3.11.2}$$

□

In the last proposition we found a recurrence for certain values $f_m(n)$. We could extend the range of m and n to include all $m \geq -1$ and $n \in \mathbb{Z}$ and define $f_m(n)$ to be the solution to the recurrence equation found above which coincides when $n \geq 0$ and $n > m$ with the values already defined. In fact, it is easy to find a simpler boundary condition yielding the desired solution other than using the boundary with $m = n$. It is easy to see that instead we may specify that $f_m(0) = 0$ for $m \geq 1$, by considering (3.11.1), which shows that we may specify 0's below the diagonal. Further, it is easy to see from this recurrence equation with given boundary, that $f_m(n)$ is an integer for $m \neq 0$ and that $f_0(n)$ are half integers. It is also easy to see from this recurrence, by induction on n, that we have for $n \geq 0$ that

$$\begin{aligned} f_m(n) &= f_m(0) + (m+1)\sum_{i=1}^{n} f_{m-1}(n-i) \\ &= f_m(0) + (m+1)\sum_{i=0}^{n-1} f_{m-1}(i). \end{aligned} \tag{3.11.3}$$

We now give the natural generalization to Proposition 3.3.3.

Proposition 3.11.1. *We have that $\mathcal{D}_B(m) : \mathbb{C}[x] \to \mathbb{C}[x]$, the generalized Sheffer shift attached to $B(t)$, is characterized as the unique linear map satisfying:*

$$\mathcal{D}_B(m)B_n(x) = f_m(n)B_{n-m}(x), \tag{3.11.4}$$

where by convention $B_n(x) = 0$ for $n \leq -1$.

Proof. By definition 3.11.1 and Lemma 3.11.1 we have

$$\begin{aligned} \mathcal{D}_B(m)B_n(x) &= \phi_B L(m)L(-1)^n y \\ &= f_m(n)\phi_B L(-1)^{n-m} y \\ &= f_m(n)B_{n-m}(x). \end{aligned}$$

□

Remark 3.11.1. The convention in Proposition 3.11.1 that $B_n(x) = 0$ for $n \leq -1$ is only used to ensure that in all cases the right hand side of (3.11.4) is well defined. This condition could have allowed for much more flexibility since we already have that $f_m(n) = 0$ whenever $n - m \leq -1$.

We shall next solve for simple closed (polynomial) expressions for $f_m(n)$ for fixed m. If we are willing to sum over squares, cubes etc. we could compute the answer for nonnegative n for each (fixed) m in turn, using (3.11.3). It is easy to verify that $f_0(n) = n + 1/2$ and $f_1(n) = n^2$ solve the first two cases. To solve for the remaining cases however, for variety, we shall use a heuristic argument making use of the Virasoro algebra relations to derive the answer. We have, for $m + l \leq n$,

$$\begin{aligned}
(l-m)f_{l+m}(n)L(-1)^{n-l-m}y &= (l-m)L(l+m)L(-1)^n y \\
&= [L(l), L(m)]L(-1)^n y \\
&= L(l)L(m)L(-1)^n y - L(m)L(l)L(-1)^n y \\
&= L(l)f_m(n)L(-1)^{n-m} y - L(m)f_l(n)L(-1)^{n-l} y \\
&= f_l(n-m)f_m(n)L(-1)^{n-m-l} y \\
&\quad - f_m(n-l)f_l(n)L(-1)^{n-l-m} y \\
&= \left(f_l(n-m)f_m(n) - f_m(n-l)f_l(n)\right) L(-1)^{n-l-m} y.
\end{aligned}$$

We shall for the time being (unmathematically) ignore the restriction on the indices and get, for whenever all terms are well defined, the identity

$$(l-m)f_{l+m}(n) = f_l(n-m)f_m(n) - f_m(n-l)f_l(n).$$

It is easy to see that the case $l = -1$ recovers 3.11.2. Further, one can check that the case $l = 0$ does not yield any new information. For $l = 1$ we get

$$\begin{aligned}
(1-m)f_{m+1}(n) &= f_1(n-m)f_m(n) - f_m(n-1)f_1(n) \\
&= (n-m)^2 f_m(n) - n^2 f_m(n-1),
\end{aligned}$$

so that it is easy to calculate, by simple substitution, each higher case (in m) starting with $m = 2$. The calculations are not difficult, of course, but I myself "cheated" and used Maple to find and factor the first few answers, which yield an easy and obvious pattern as follows:

$$
\begin{aligned}
f_{-1}(n) &= 1 \\
f_0(n) &= n + 1/2 \\
f_1(n) &= n^2 \\
f_2(n) &= (1/2)n(n-1)(2n-1) \\
f_3(n) &= n(n-1)^2(n-2) \\
f_4(n) &= (1/2)n(n-1)(n-2)(n-3)(2n-3) \\
f_5(n) &= n(n-1)(n-2)^2(n-3)(n-4) \\
f_6(n) &= (1/2)n(n-1)(n-2)(n-3)(n-4)(n-5)(2n-5) \\
f_7(n) &= n(n-1)(n-2)(n-3)^2(n-4)(n-5)(n-6) \\
&\vdots
\end{aligned}
$$

With that as a guide, we may now return to doing rigorous math and state and prove the following theorem.

Theorem 3.11.2. *The unique solution to the recurrence equation (3.11.2) with* $m \geq -1$ *and* $n \in \mathbb{Z}$ *and with boundary given by* $f_{-1}(n) = 1$*,* $f_0(0) = 1/2$ *and* $f_m(0) = 0$ *for* $m \geq 1$ *is given by*

$$
\begin{aligned}
f_{-1}(n) &= 1 \\
f_m(n) &= (1/2)(n(n-1)(n-2)\cdots(n-m+1))(2n-m+1) \quad \textit{for} \quad m \geq 0.
\end{aligned}
$$

Proof. The proof is a straightforward calculation. Let $m \geq 0$ (although admittedly the low m cases are a bit degenerate in this notation). Then we have

$$\begin{aligned}
f_m(n) - f_m(n-1) &= (1/2)(n(n-1)(n-2)\cdots(n-m+1))(2n-m+1) \\
&\quad - (1/2)(n-1)(n-2)\cdots(n-m)(2(n-1)-m+1) \\
&= (1/2)(n-1)(n-2)\cdots(n-m+1)\cdot \\
&\quad \cdot (n(2n-m+1)-(n-m)(2(n-1)-m+1)) \\
&= (1/2)(n-1)(n-2)\cdots(n-m+1)(2mn+2n-m^2-m) \\
&= (n-1)(n-2)\cdots(n-m+1)(m+1)(n-m/2) \\
&= (m+1)f_{m-1}(n-1).
\end{aligned}$$

□

Remark 3.11.2. Recalling (3.11.3), it is easy to see that we could use the last result, perhaps somewhat awkwardly, to solve for the sum of squares and cubes etc., which happens to be related to the Bernoulli numbers, one of the motivating subjects for Blissard [Bli] and is one of the classic problems solved via umbral methods (cf. Chapter 11 [Do] for a nice, succinct old-fashioned umbral style proof and also Chapter 3.11 [Mel]).

Remark 3.11.3. We note that the umbral calculus has long been known to have connections to the Bernoulli numbers and polynomials (see e.g. [Mel]). Bernoulli polynomials have also appeared in the literature of vertex algebra theory (see e.g. [L] and [DLM]). Just as we have been establishing some analogues and connections between umbral calculus and vertex algebra theory, it might be interesting in future work to investigate further possible connections explicitly related to Bernoulli numbers and polynomials.

We shall conclude this chapter by stating and proving the natural generalization to the original formula defining the attached Sheffer shifts in Definition 3.3.3.

Proposition 3.11.2. *For each $m \geq -1$, the map $\mathcal{D}_B(m) : \mathbb{C}[x] \to \mathbb{C}[x]$ is the unique linear map satisfying:*

$$\mathcal{D}_B(m)e^{xB(w)} = \left(w^{m+1}\frac{\partial}{\partial w} + \frac{m+1}{2}w^m\right)e^{xB(w)}.$$

Proof. We calculate, for $m \geq -1$ (although once again the low m cases are a bit degenerate) to get

$$\begin{aligned} w^{m+1}\frac{\partial}{\partial w}e^{xB(w)} &= w^{m+1}\frac{\partial}{\partial w}\sum_{n\geq 0}\frac{B_n(x)w^n}{n!} \\ &= \sum_{n\geq 0}\frac{nB_n(x)w^{n+m}}{n!} \\ &= \sum_{n\geq m}\frac{(n-m)B_{n-m}(x)w^n}{(n-m)!} \\ &= \sum_{n\geq m}\frac{n(n-1)\cdots(n-m)B_{n-m}(x)w^n}{n!} \end{aligned}$$

and

$$\begin{aligned} w^m e^{xB(w)} &= \sum_{n\geq 0}\frac{B_n(x)w^{n+m}}{n!} \\ &= \sum_{n\geq m}\frac{B_{n-m}(x)w^n}{(n-m)!} \\ &= \sum_{n\geq m}\frac{n(n-1)\cdots(n-m+1)B_{n-m}(x)w^n}{n!}, \end{aligned}$$

so that it is easy to check that

$$\begin{aligned} \left(w^{m+1}\frac{\partial}{\partial w}+\frac{m+1}{2}w^m\right)e^{xB(w)} &= \sum_{n\geq m}\frac{f_m(n)B_{n-m}(x)w^n}{n!} \\ &= \mathcal{D}_B(m)e^{xB(w)}. \end{aligned}$$

□

Remark 3.11.4. Building on Remarks 3.1.1 and 3.3.4 we may regard Proposition 3.11.2 as an analogue of formula (8.7.37) in [FLM2] in the cases where the n in (8.7.37) in [FLM2] is restricted so that $n \geq -1$. The h in formula (8.7.37) in [FLM2] should be replaced by the weight of the relevant lowest weight vector, which in our setting seems perhaps to correspond with the lowest weight of the module of the Virasoro algebra which we have been considering, which as we have noted is indeed 1/2.

3.12 Faà di Bruno's formula

We calculate Faà di Bruno's classical formula for the higher derivatives of a composite function, completing a calculation begun in Section 3.1. For $f(x)$ and $g(x) \in \mathbb{C}[[x]]$

such that the constant term of $g(x)$ is 0 we have

$$\begin{aligned}
e^{w\frac{d}{dx}}f(g(x)) &= f(g(x+w)) \\
&= f(g(x)+(g(x+w)-g(x))) \\
&= \left(e^{(g(x+w)-g(x))\frac{d}{dy}}f(y)\right)|_{y=g(x)} \\
&= \sum_{n\geq 0}\frac{f^{(n)}(g(x))}{n!}\left(e^{w\frac{d}{dx}}g(x)-g(x)\right)^n \\
&= \sum_{n\geq 0}\frac{f^{(n)}(g(x))}{n!}\left(\sum_{m\geq 1}\frac{g^{(m)}(x)}{m!}w^m\right)^n \\
&= \sum_{n\geq 0}\frac{f^{(n)}(g(x))}{n!}\sum_{l\geq 0}\sum_{\substack{m_1+\cdots+m_n=l\\ m_i\geq 1}}\frac{g^{m_1}(x)}{m_1!}\cdots\frac{g^{m_n}(x)}{m_n!}w^l, \qquad (3.12.1)
\end{aligned}$$

which by equating coefficients yields:

$$\frac{d}{dx}^l f(g(x)) = l!\sum_{n\geq 0}\frac{f^{(n)}(g(x))}{n!}\sum_{\substack{m_1+\cdots+m_n=l\\ m_i\geq 1}}\frac{g^{m_1}(x)}{m_1!}\cdots\frac{g^{m_n}(x)}{m_n!},$$

which is indeed a formula which satisfies our original requirements. A re-indexing may look more familiar to readers, however. Noticing that the m_i in (3.12.1) give unordered partitions of l in n parts, we may instead index using the number of i's in the relevant partition, which we shall call k_i, so that, in particular, $k_i \geq 0$. Then since we will index by ordered partitions we need to multiply by the number of arrangements of the set $\{1,\cdots,l\}$ with k_i repeats of i, which is obviously $\frac{(k_1+\cdots+k_l)!}{k_1!\cdots k_l!}$. This gives

$$\begin{aligned}
e^{w\frac{d}{dx}}f(g(x)) &= \sum_{n\geq 0}\frac{f^{(n)}(g(x))}{n!}\sum_{\substack{m_1+\cdots+m_n=l\\ m_i\geq 1}}\frac{g^{m_1}(x)}{m_1!}\cdots\frac{g^{m_n}(x)}{m_n!}w^l \\
&= \sum_{l\geq 0}\sum_{k_1+2k_2+\cdots+lk_l=l}\frac{f^{(k_1+\cdots+k_l)}(g(x))}{(k_1+\cdots+k_l)!}\frac{(k_1+\cdots+k_l)!}{k_1!\cdots k_l!}\cdot \\
&\qquad\cdot\left(\frac{g^{(1)}(x)}{1!}\right)^{k_1}\cdots\left(\frac{g^{(l)}(x)}{l!}\right)^{k_l}w^l,
\end{aligned}$$

which by equating coefficients yields:

$$\frac{d}{dx}^{l} f(g(x)) = \sum_{l\geq 0} \sum_{k_1+2k_2+\cdots+lk_l=l} f^{(k_1+\cdots+k_l)}(g(x)) \frac{l!}{k_1!\cdots k_l!} \cdot \\ \cdot \left(\frac{g^{(1)}(x)}{1!}\right)^{k_1} \cdots \left(\frac{g^{(l)}(x)}{l!}\right)^{k_l},$$

which by considering the case $f(x) = e^x$ essentially yields, up to notational differences, formula (12.3.7) in [An] for the Bell polynomials. This formula, in the notes to Chapter 12 of [An], is called Faà di Bruno's formula.

3.13 A standard quadratic representation of the Virasoro algebra of central charge 1

We shall show that the quadratic operators given by (3.10.4) and (3.10.5) do indeed give a representation of the Virasoro algebra of central charge 1. Our verification is a standard elementary approach (see also Section 8.7 of [FLM2] where those authors develop a much deeper proof arising from the theory of vertex operator algebras). All the material in this section is classical and we include it for the sake of completeness. We shall present, for expository purposes, three proofs in increasing order of simplicity.

Theorem 3.13.1. *The operators defined by (3.10.4) and (3.10.5) satisfy the commutation relations (3.10.3).*

Proof. We begin by calculating that for k, l, m and $n \in \mathbb{Z}$ we get

$$\begin{aligned}
[h(k)h(l), h(m)h(n)] &= h(k)h(l)h(m)h(n) - h(m)h(n)h(k)h(l) \\
&= h(k)h(m)h(n)h(l) - h(m)h(n)h(k)h(l) \\
&\quad + [h(l), h(m)]h(k)h(n) + [h(l), h(n)]h(k)h(m) \\
&= [h(l), h(m)]h(k)h(n) + [h(l), h(n)]h(k)h(m) \\
&\quad + [h(k), h(m)]h(n)h(l) + [h(k), h(n)]h(m)h(l).
\end{aligned}$$

Then for m and n nonzero integers such that $m+n \neq 0$ we get

$$\begin{aligned}
[L(m), L(n)] &= (1/4)\left[\sum_{k\in\mathbb{Z}} h(m-k)h(k), \sum_{l\in\mathbb{Z}} h(n-l)h(l)\right] \\
&= (1/4)\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}}[h(m-k)h(k), h(n-l)h(l)] \\
&= (1/4)\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}}[h(m-k), h(n-l)]h(l)h(k) \\
&\qquad\qquad + [h(m-k), h(l)]h(n-l)h(k) \\
&\qquad\qquad + [h(k), h(l)]h(m-k)h(n-l) \\
&\qquad\qquad + [h(k), h(n-l)]h(m-k)h(l) \\
&= (1/4)\sum_{k\in\mathbb{Z}}(2(m-k)h(m+n-k)h(k) + 2kh(m-k)h(n+k)) \\
&= (1/4)\sum_{k\in\mathbb{Z}} 2(m-k)h(m+n-k)h(k) \\
&\quad + (1/4)\sum_{k\in\mathbb{Z}} 2kh(m-k)h(n+k) \\
&= (1/2)\sum_{k\in\mathbb{Z}}(m-k)h(m+n-k)h(k) \\
&\quad + (1/2)\sum_{k\in\mathbb{Z}}(k-n)h(m+n-k)h(k) \\
&= (m-n)/2\sum_{k\in\mathbb{Z}} h(m+n-k)h(k) \\
&= (m-n)L(m+n).
\end{aligned}$$

One must be careful with the case where $m+n=0$. In the previous calculation at one point we ungrouped terms into two separate series, but when $m+n=0$ this leads to undefined sums. Let $m>0$ and we get

$$
\begin{aligned}
[L(m), L(-m)] &= (1/2)\sum_{k\in\mathbb{Z}}((m-k)h(-k)h(k) + kh(m-k)h(-m+k)) \\
&= (1/2)\sum_{k\geq m}((m-k)h(-k)h(k) + kh(m-k)h(-m+k)) \\
&\quad + (1/2)\sum_{k=0}^{m-1}(m-k)h(-k)h(k) + (1/2)\sum_{k=0}^{m-1} kh(m-k)h(-m+k)) \\
&\quad + (1/2)\sum_{k<0}((m-k)h(-k)h(k) + kh(m-k)h(-m+k)) \\
&= (1/2)\sum_{k\geq m}((m-k)h(-k)h(k) + kh(m-k)h(-m+k)) \\
&\quad + (1/2)\sum_{k=0}^{m-1}(m-k)h(-k)h(k) \\
&\quad + (1/2)\sum_{k=0}^{m-1} k(h(-m+k)h(m-k) + [h(m-k), h(-m+k)]) \\
&\quad + (1/2)\sum_{k<0}((m-k)(h(k)h(-k) + [h(-k), h(k)]) \\
&\quad + k(h(-m+k)h(m-k) + [h(m-k), h(-m+k)]) \\
&= (1/2)\sum_{k\geq m}((m-k)h(-k)h(k) + kh(m-k)h(-m+k)) \\
&\quad + (1/2)\sum_{k=0}^{m-1}(m-k)h(-k)h(k) + (1/2)\sum_{k=0}^{m-1} kh(-m+k)h(m-k) \\
&\quad + (1/2)\sum_{k<0}((m-k)h(k)h(-k) + kh(-m+k)h(m-k)) \\
&\quad + (1/2)\sum_{k=0}^{m-1} k[h(m-k), h(-m+k)] \\
&= (1/2)\sum_{k\in\mathbb{Z}}(m-k)h(-|k|)h(|k|) + (1/2)\sum_{k\in\mathbb{Z}} kh(-|m-k|)h(|-m+k|) \\
&\quad + (1/2)\sum_{0\leq k\leq m-1} k(m-k) \\
&= (1/2)\sum_{k\in\mathbb{Z}}((m-k) + (k+m))h(-|k|)h(|k|) + (1/2)\sum_{k=0}^{m-1} k(m-k) \\
&= 2mL(0) + (1/12)m(m-1)(m+1).
\end{aligned}
$$

Finally we calculate to get

$$
\begin{aligned}
[L(0), L(m)] &= (1/4)\left[\sum_{k\in\mathbb{Z}} h(-|k|)h(|k|), \sum_{l\in\mathbb{Z}} h(m-l)h(l)\right] \\
&= (1/4)\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}}[h(-|k|)h(|k|), h(m-l)h(l)] \\
&= (1/4)\sum_{k\in\mathbb{Z}}\sum_{l\in\mathbb{Z}}[h(-|k|), h(l)]h(m-l)h(|k|) + [h(-|k|), h(m-l)]h(l)h(|k|) \\
&\quad + [h(|k|), h(l)]h(-|k|)h(m-l)[h(|k|), h(m-l)]h(-|k|)h(l) \\
&= (1/4)\sum_{k\in\mathbb{Z}}(-2|k|h(m-|k|)h(|k|) + 2|k|h(-|k|)h(m+|k|)) \\
&= (1/2)\sum_{k\geq 0}(-kh(m-k)h(k) + kh(-k)h(m+k)) \\
&\quad + (1/2)\sum_{k<0}(kh(m+k)h(-k) - kh(k)h(m-k)) \\
&= (1/2)\sum_{k\in\mathbb{Z}}(-kh(m-k)h(k) + kh(-k)h(m+k)) \\
&= (1/2)\sum_{k\in\mathbb{Z}}(-kh(m-k)h(k) + (k-m)h(-k+m)h(k)) \\
&= (-m/2)\sum_{k\in\mathbb{Z}} h(m-k)h(k) \\
&= -mL(m).
\end{aligned}
$$

□

We may note that a rough premise for the result of Theorem 3.13.1 and its proof is that commutators of quadratics in Heisenberg operators lead to more quadratic terms or possibly central terms since terms cancel pairwise by reducing to (central) scalars via commutators. This same sort of premise can be used when taking commutators of quadratics of Heisenberg operators with linear terms of Heisenberg operators.

Lemma 3.13.1. *For all $m, k \in \mathbb{Z}$*

$$[L(m), h(k)] = -kh(m+k).$$

Proof. For $k, l, m \in \mathbb{Z}$ we get

$$\begin{aligned}
[h(k)h(l), h(m)] &= h(k)h(l)h(m) - h(m)h(k)h(l) \\
&= [h(l)h(m)]h(k) + h(k)h(m)h(l) - h(m)h(k)h(l) \\
&= [h(l)h(m)]h(k) + [h(k)h(m)]h(l).
\end{aligned}$$

Then for $m \neq 0$ we get

$$\begin{aligned}
[L(m), h(k)] &= (1/2)\left[\sum_{l\in\mathbb{Z}} h(m-l)h(l), h(k)\right] \\
&= (1/2)\sum_{l\in\mathbb{Z}}([h(m-l), h(k)]h(l) + [h(l), h(k)]h(m-l)) \\
&= -kh(m+k)
\end{aligned}$$

and we get

$$\begin{aligned}
[L(0), h(k)] &= (1/2)\left[\sum_{l\in\mathbb{Z}} h(-|l|)h(|l|), h(k)\right] \\
&= (1/2)\sum_{l\in\mathbb{Z}}([h(-|l|), h(k)]h(|l|) + [h(|l|), h(k)]h(-|l|)) \\
&= -kh(k).
\end{aligned}$$

□

Proof. (second proof of Theorem 3.13.1)

For $m, n \in \mathbb{Z}$ such that $n \neq 0$ which is obviously enough to check, by Lemma 3.13.1 we get

$$\begin{aligned}
[L(m), L(n)] &= \frac{1}{2}\left[\sum_{k\in\mathbb{Z}} h(m-k)h(k), L(n)\right] \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}}(h(m-k)h(k)L(n) - L(n)h(m-k)h(k)) \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}}(kh(m-k)h(n+k) + h(m-k)L(n)h(k) - L(n)h(m-k)h(k)) \\
&= \frac{1}{2}\sum_{k\in\mathbb{Z}}(kh(m-k)h(n+k) + (m-k)h(m+n-k)h(k)).
\end{aligned}$$

We may now proceed exactly as in the first proof we gave of Theorem 3.13.1, noting that this time we have merged the first and third cases together. □

Remark 3.13.1. In the proof of Theorem 1.9.6 in [FLM2] the authors computed the central term in a different way than we did, while pointing out various alternative routes including, of course, the one that we took.

We shall next recall a certain well-known notion of normal ordering to give a yet more streamlined proof of Theorem 3.10.1.

For the operators $h(n)$ $n \in \mathbb{Z}$ we recall the following normal ordered product (cf. Section 3.3 [FLM2]).

$$: h(n)h(m) := \begin{cases} h(n)h(m) & \text{if} \quad n \leq m \\ h(m)h(n) & \text{if} \quad m > n \end{cases}$$

and more generally

$$: h(n_1)h(n_2)\cdots h(n_k) := h(n_{\pi(1)})h(n_{\pi(2)})\cdots h(n_{\pi(k)}),$$

where π is the unique permutation such that $n_{\pi(1)} \leq n_{\pi(1)} \leq \cdots \leq n_{\pi(1)}$ where the permutation on the first k positive integers induced from π when restricted to any domain of equal inputs is the identity permutation. We note that normal ordered products force $h(n)$ where n is nonnegative to the right of $h(n)$ where n is negative so that infinite sums of normally ordered products are always well defined.

We shall use the following notational convention, for $1 \leq j \leq k$

$$: h(n_1)h(n_2)\cdots h(n_j + (m))\cdots h(n_k) := h(n_{\pi(1)})h(n_{\pi(2)})\cdots h(n_{\pi(i)} + m)\cdots h(n_{\pi(k)}),$$

where π is the unique permutation such that $n_{\pi(1)} \leq n_{\pi(2)} \leq \cdots \leq n_{\pi(i)} \leq \cdots \leq n_{\pi(1)}$, where the permutation on the first k positive integers induced from π when restricted to any domain of equal inputs is the identity permutation, and where $\pi(i) = j$. That is, for the purposes of the normal ordering any term added in parentheses is "invisible." We have

$$: h(r + (m))h(s) := \begin{cases} : h(r+m)h(s) : +r+m & \text{if } r+m+s = 0, r \leq s \leq 0, m > 0 \\ : h(r+m)h(s) : +s & \text{if } r+m+s = 0, 0 \leq s < r, m < 0 \\ : h(r+m)h(s) : & \text{otherwise.} \end{cases}$$

The definition of the operators $L(n)$ can be conveniently rewritten using normal ordered products as

$$L(n) = (1/2) \sum_{\substack{r,s \in \mathbb{Z} \\ r+s=n}} : h(r)h(s) : \qquad \text{for all } n \in \mathbb{Z}.$$

We conclude with a third proof of Theorem 3.13.1.

Proof. By Lemma 3.13.1, we have for $m, n \in \mathbb{Z}$

$$\begin{aligned}
[L(m), L(n)] &= \left[L(m), (1/2) \sum_{r+s=n} : h(r)h(s) : \right] \\
&= (1/2) \sum_{r+s=n} [L(m), : h(r)h(s) :] \\
&= (1/2) \sum_{r+s=n} (-r : h(r+(m))h(s) : -s : h(r)h(s+(m)) :).
\end{aligned}$$

Now if $m + n \neq 0$, then we get

$$\begin{aligned}
[L(m), L(n)] &= (1/2) \sum_{r+s=n} (-r : h(r+m)h(s) : -s : h(r)h(s+m) :) \\
&= (1/2) \sum_{r+s=n+m} ((m-r) + (m-s)) : h(r)h(s) : \\
&= (m-n)(1/2) \sum_{r+s=n+m} : h(r)h(s) : \\
&= (m-n)L(m+n).
\end{aligned}$$

And if $m + n = 0$ and $n \geq 0$ it is clear that we have

$$\begin{aligned}
[L(-n), L(n)] &= (-n-n)L(-n+n) + (1/2) \sum_{\substack{r > s \geq 0 \\ r+s=n}} - rs + (1/2) \sum_{\substack{s \geq r \geq 0 \\ r+s=n}} - sr \\
&= -2nL(0) - (1/2) \sum_{\substack{r,s \geq 0 \\ r+s=n}} rs \\
&= -2nL(0) - (1/2) \sum_{0 \leq t \leq n} t(n-t) \\
&= -2nL(0) - (n-1)(n)(n+1)/12.
\end{aligned}$$

The result now follows easily. $\square$

Chapter 4

Replacement axioms for the Jacobi identity for vertex algebras and their modules

4.1 Formal calculus summarized

We shall write $x, y, z, x_1, x_2, x_3, \ldots$ for commuting formal variables. In this chapter, formal variables will always commute, and we will not use complex variables. All vector spaces will be over $\mathbb{C}$. Let V be a vector space. We use the following:

$$V[[x, x^{-1}]] = \left\{ \sum_{n \in \mathbb{Z}} v_n x^n | v_n \in V \right\}$$

(formal Laurent series), and some of its subspaces:

$$V((x)) = \left\{ \sum_{n \in \mathbb{Z}} v_n x^n | v_n \in V, v_n = 0 \text{ for sufficiently negative } n \right\}$$

(truncated formal Laurent series),

$$V[[x]] = \left\{ \sum_{n \geq 0} v_n x^n | v_n \in V \right\}$$

(formal power series),

$$V[x, x^{-1}] = \left\{ \sum_{n \in \mathbb{Z}} v_n x^n | v_n \in V, v_n = 0 \text{ for all but finitely many } n \right\}$$

(formal Laurent polynomials), and

$$V[x] = \left\{ \sum_{n \geq 0} v_n x^n | v_n \in V, v_n = 0 \text{ for all but finitely many } n \right\}$$

(formal polynomials). Since some of these spaces are not algebras, we must define multiplication only up to a natural restrictive condition. This condition is the summability condition which is given in Definitions 2.1.4 and 2.1.5 in [LL]. In general, when computing some series we shall say that it "exists" (in keeping with an analogy with analysis) when the coefficient of any monomial has only finitely many contributing terms, that is, that the coefficient of any monomial is finitely computable. If the coefficients are themselves endomorphisms of a vector space then we only require that the coefficient of any monomial be finitely computable when the series is applied to a fixed, but arbitrary, vector (see Remark 2.1.3 in [LL]).

Since some of our spaces, such as $\operatorname{End} V[[x, x^{-1}]]$, are not algebras, but only have a partial multiplication, we are not guaranteed that multiplication is associative even when it is defined. In fact, multiplication is not associative in the usual sense, but there is a replacement property. If $F(x), G(x)$ and $H(x) \in \operatorname{End} V[[x, x^{-1}]]$ and if the three products $F(x)G(x)$, $G(x)H(x)$ and $F(x)G(x)H(x)$ all exist, then

$$(F(x)G(x))H(x) = F(x)(G(x)H(x))$$

(See Remark 2.1.6 [LL] and the preceding discussion). We shall refer to this replacement for associativity as "partial associativity".

Remark 4.1.1. Throughout this chapter, as in [LL], we often extend our spaces to include more than one variable. We state certain properties which have natural extensions in such multivariable settings, which we will also use without further comment.

We define the operator $\operatorname{Res}_x : V[[x, x^{-1}]] \to V$ by the following: For $f(x) = \sum_{n \in \mathbb{Z}} a_n x^n \in V[[x, x^{-1}]]$,

$$\operatorname{Res}_x f(x) = a_{-1}.$$

Further, we shall frequently use the notation e^w to refer to the formal exponential expansion, where w is any formal object for which such expansion makes sense. For instance, we have the linear operator $e^{y\frac{d}{dx}} : \mathbb{C}[[x, x^{-1}]] \to \mathbb{C}[[x, x^{-1}]][[y]]$:

$$e^{y\frac{d}{dx}} = \sum_{n \geq 0} \frac{y^n}{n!} \left(\frac{d}{dx} \right)^n .$$

We have (see (2.2.18) in [LL]), the *automorphism property*:

$$e^{y\frac{d}{dx}}(p(x)q(x)) = \left(e^{y\frac{d}{dx}} p(x) \right) \left(e^{y\frac{d}{dx}} q(x) \right), \tag{4.1.1}$$

for all $p(x) \in \mathbb{C}[x, x^{-1}]$ and $q(x) \in \mathbb{C}[[x, x^{-1}]]$. We use the *binomial expansion convention,* which states that

$$(x+y)^n = \sum_{k \geq 0} \binom{n}{k} x^{n-k} y^k, \tag{4.1.2}$$

where we allow n to be any integer and where we define

$$\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k!};$$

the binomial expression is expanded in nonnegative powers of the second-listed variable. We also have (see Proposition 2.2.2 in [LL]) the *formal Taylor theorem*:

Proposition 4.1.1. *Let* $v(x) \in V[[x, x^{-1}]]$. *Then*

$$e^{y\frac{d}{dx}} v(x) = v(x+y).$$

□

For completeness we include a proof of the following frequently used fact, which equates two different expansions.

Proposition 4.1.2. *For all* $n \in \mathbb{Z}$,

$$(x + (y+z))^n = ((x+y)+z)^n.$$

Proof. If w_1 and w_2 are commuting formal objects, then $e^{w_1+w_2} = e^{w_1}e^{w_2}$. Thus we have

$$(x+(y+z))^n = e^{(y+z)\frac{\partial}{\partial x}} x^n = e^{y\frac{\partial}{\partial x}} \left(e^{z\frac{\partial}{\partial x}} x^n \right) = e^{y\frac{\partial}{\partial x}} (x+z)^n = ((x+y)+z)^n.$$

□

We note as a consequence that for all integers n (and not just nonnegative integers) we have the (non-vacuous) fact that

$$((x+y)-y)^n = (x+(y-y))^n = x^n.$$

We define the formal delta function by

$$\delta(x) = \sum_{n \in \mathbb{Z}} x^n.$$

We have (see Proposition 2.3.21 and Remarks 2.3.24 and 2.3.25 in [LL]) the delta-function substitution property:

Proposition 4.1.3. *For* $f(x,y,z) \in \operatorname{End} V[[x,x^{-1},y,y^{-1},z,z^{-1}]]$ *such that for each fixed* $v \in V$

$$f(x,y,z)v \in \operatorname{End} V[[x,x^{-1},y,y^{-1}]]((z))$$

and such that

$$\lim_{x\to y} f(x,y,z)$$

exists (where the "limit" is the indicated formal substitution), we have

$$\delta\left(\frac{y+z}{x}\right) f(x,y,z) = \delta\left(\frac{y+z}{x}\right) f(y+z,y,z) = \delta\left(\frac{y+z}{x}\right) f(x,x-z,z).$$

□

As in [LL], we use similarly verified substitutions below without comment.

4.2 Formal calculus further developed

Certain elementary identities concerning delta functions are very convenient for dealing with the arithmetic of vertex algebras and, in fact, in some cases, are fundamental to the very notion of vertex algebra. We state and prove some such identities in this section.

The following well-known proposition appears as Proposition 2.3.8 in [LL]. We present an alternate proof which is implicitly exploiting the $\mathcal{S}_3$-symmetry underlying the notion of vertex algebra. We include this alternate proof to emphasize that $\mathcal{S}_3$-symmetry is playing a role in the development of the ideas in this chapter, as we discussed in the introduction. For a precise formulation see Section 2.7 in [FHL] and Section 3.7 in [LL].

Proposition 4.2.1. *We have the following two elementary identities:*

$$x_1^{-1}\delta\left(\frac{x_2+x_0}{x_1}\right) - x_2^{-1}\delta\left(\frac{x_1-x_0}{x_2}\right) = 0 \tag{4.2.1}$$

and

$$x_0^{-1}\delta\left(\frac{x_1-x_2}{x_0}\right) - x_0^{-1}\delta\left(\frac{-x_2+x_1}{x_0}\right) - x_1^{-1}\delta\left(\frac{x_2+x_0}{x_1}\right) = 0. \tag{4.2.2}$$

Proof. First observe that

$$y^{-1}\delta\left(\frac{x}{y}\right) = (x-y)^{-1} + (y-x)^{-1},$$

where we note that a clue as to why this identity holds is that both expressions are annihilated by $x-y$. Then by the formal Taylor theorem

$$\begin{aligned} y^{-1}\delta\left(\frac{x+z}{y}\right) &= e^{z\frac{d}{dx}} y^{-1}\delta\left(\frac{x}{y}\right) \\ &= e^{z\frac{d}{dx}}((x-y)^{-1} + (y-x)^{-1}) \\ &= ((x+z)-y)^{-1} + (y-(x+z))^{-1}. \end{aligned}$$

Being careful with minus signs we may respectively expand all the terms in the left-hand side of (4.2.1) in this manner yielding

$$\begin{aligned} &((x_2+x_0)-x_1)^{-1} + (x_1-(x_2+x_0))^{-1} \\ &-((x_1-x_0)-x_2)^{-1} - (x_2-(x_1-x_0))^{-1}. \end{aligned}$$

Now we get by Proposition 4.1.2 that the first and fourth, and the second and third terms pairwise cancel each other thus giving us (4.2.1). Similarly we may respectively expand all the terms in the left-hand side of (4.2.2) to get

$$\begin{aligned} &((x_1-x_2)-x_0)^{-1} + (x_0-(x_1-x_2))^{-1} \\ &-((-x_2+x_1)-x_0)^{-1} - (x_0-(-x_2+x_1))^{-1} \\ &-((x_2+x_0)-x_1)^{-1} - (x_1-(x_2+x_0))^{-1}. \end{aligned}$$

Now we get by Proposition 4.1.2 that the first and sixth terms, the third and fifth terms, and the second and fourth terms pairwise cancel each other thus giving us (4.2.2). $\square$

A slight variant of the following Proposition appeared in [LL] as Propositions 2.3.26 and 2.3.27:

Proposition 4.2.2. *Let* $g(x_0, x_1, x_2) \in V[[x_0, x_1, x_2]]$*. Next, for* a, b *and* $c \geq 0$*, let*

$$f(x_0, x_1, x_2) = \frac{g(x_0, x_1, x_2)}{x_0^a x_1^b x_2^c}.$$

Then

$$x_1^{-1}\delta\left(\frac{x_2+x_0}{x_1}\right)f(x_0,x_2+x_0,x_2)=x_2^{-1}\delta\left(\frac{x_1-x_0}{x_2}\right)f(x_0,x_1,x_1-x_0)$$

and

$$x_0^{-1}\delta\left(\frac{x_1-x_2}{x_0}\right)f(x_1-x_2,x_1,x_2)-x_0^{-1}\delta\left(\frac{-x_2+x_1}{x_0}\right)f(-x_2+x_1,x_1,x_2)-$$
$$x_1^{-1}\delta\left(\frac{x_2+x_0}{x_1}\right)f(x_0,x_2+x_0,x_2)=0.$$

Proof. We have, for instance, by the (partial) formal delta substitution principle that

$$\begin{aligned}x_0^{-1}\delta\left(\frac{x_1-x_2}{x_0}\right)f(x_1-x_2,x_1,x_2)&=x_0^{-1}\delta\left(\frac{x_1-x_2}{x_0}\right)f((x_0+x_2)-x_2,x_1,x_2)\\&=x_0^{-1}\delta\left(\frac{x_1-x_2}{x_0}\right)f(x_0,x_1,x_2).\end{aligned}\tag{4.2.3}$$

Similar substitutions on the other terms also leave the delta function multiplied by the common factor, $f(x_0,x_1,x_2)$. Therefore (4.2.1) and (4.2.2) yield the result. □

Remark 4.2.1. In the proof of Proposition 4.2.2 we specifically chose the order in which to perform the substitutions, rather than to reverse the equalities in (4.2.3), because it is easier to see that all "existence" type conditions are met.

We also have the following converse proposition, which the author believes has not previously appeared in full, but much of it may be viewed as placing in a more elementary setting the relevant existing arguments used in the axiomatic theory presented in [LL]. We note that a very similar approach was taken in Lemma 2.1 [Li2] where the author had already proved, in a somewhat different manner, some of the implications (related to the Jacobi identity and commutativity and associativity properties, but not skew-associativity properties) of the following proposition.

Proposition 4.2.3. *Let $f(x_1,x_2)$, $g(x_1,x_2)$, and $h(x_1,x_2)\in V((x_1))((x_2))$. We have certain implications among the following statements:*

- *(A) We have*

$$\begin{aligned}x_0^{-1}\delta\left(\frac{x_1-x_2}{x_0}\right)f(x_1,x_2)-&x_0^{-1}\delta\left(\frac{-x_2+x_1}{x_0}\right)g(x_2,x_1)\\&-x_1^{-1}\delta\left(\frac{x_2+x_0}{x_1}\right)h(x_2,x_0)=0.\end{aligned}\tag{4.2.4}$$

- *(B) There exists some* $m_1 \geq 0$ *such that*

$$(x_1 - x_2)^{m_1}(f(x_1, x_2) - g(x_2, x_1)) = 0.$$

- *(C) There exists some* $m_2 \geq 0$ *such that*

$$(x_0 + x_2)^{m_2}(f(x_0 + x_2, x_2) - h(x_2, x_0)) = 0.$$

- *(D) There exists some* $m_3 \geq 0$ *such that*

$$(x_1 - x_0)^{m_3}(g(-x_0 + x_1, x_1) - h(x_1 - x_0, x_0)) = 0.$$

- *(E) There exist* $p_1(x_1, x_2) \in V[[x_1, x_2]]$ *and* $a_1, b_1, c_1 \geq 0$ *such that*

$$f(x_1, x_2) = \frac{p_1(x_1, x_2)}{(x_1 - x_2)^{a_1} x_1^{b_1} x_2^{c_1}}$$

  and

$$g(x_2, x_1) = \frac{p_1(x_1, x_2)}{(-x_2 + x_1)^{a_1} x_1^{b_1} x_2^{c_1}}.$$

- *(F) There exist* $p_2(x_0, x_2) \in V[[x_0, x_2]]$ *and* $a_2, b_2, c_2 \geq 0$ *such that*

$$f(x_0 + x_2, x_2) = \frac{p_2(x_0, x_2)}{x_0^{a_2}(x_0 + x_2)^{b_2} x_2^{c_2}}$$

  and

$$h(x_2, x_0) = \frac{p_2(x_0, x_2)}{x_0^{a_2}(x_2 + x_0)^{b_2} x_2^{c_2}}.$$

- *(G) There exist* $p_3(x_0, x_1) \in V[[x_0, x_1]]$ *and* $a_3, b_3, c_3 \geq 0$ *such that*

$$g(-x_0 + x_1, x_1) = \frac{p_3(x_0, x_1)}{x_0^{a_3} x_1^{b_3}(-x_0 + x_1)^{c_3}}$$

  and

$$h(x_1 - x_0, x_0) = \frac{p_3(x_0, x_1)}{x_0^{a_3} x_1^{b_3}(x_1 - x_0)^{c_3}}.$$

Namely, we have:

(ia)	$(A) \Rightarrow (B)$,	(iia)	$(B) \Rightarrow (E)$,		$(iiia)$	(E) and $(F) \Rightarrow (A)$,
(ib)	$(A) \Rightarrow (C)$,	(iib)	$(C) \Rightarrow (F)$,		$(iiib)$	(E) and $(G) \Rightarrow (A)$,
(ic)	$(A) \Rightarrow (D)$,	(iic)	$(D) \Rightarrow (G)$,	and	$(iiic)$	(F) and $(G) \Rightarrow (A)$.

Remark 4.2.2. In [LL] the authors essentially used only statements (A), (B), (C), (E) and (F). It is an application of the principles of statements (D) and (G) which leads to the new notion of weak skew-associativity.

Proof. The proofs of (ia), (ib), and (ic) are similar. For (ia) we note that (A) trivially implies that the left-hand side of (4.2.4) is lower truncated in powers of x_0. Further, the third term in the left-hand side of (4.2.4) is visibly lower truncated in powers of x_0, and therefore the sum of the remaining two terms must be lower truncated in powers of x_0, which precisely yields (B). The proofs of (ib) and (ic) similarly follow from the lower truncation, in the left-hand side of (4.2.4), of x_1 and x_2 respectively, after the obvious (especially in light of the statements of (C) and (D)) delta function substitutions are made.

The proofs of (iia), (iib) and (iic) are similar. We show only (iia). Since both $f(x_1, x_2) \in V((x_1))((x_2))$ and $g(x_2, x_1) \in V((x_2))((x_1))$ we must have that $(x_1 - x_2)^{m_1} f(x_1, x_2)$ and $(x_1 - x_2)^{m_1} g(x_2, x_1)$ are both $\in V((x_1, x_2))$. So there is some $p_1(x_1, x_2) \in V[[x_1, x_2]]$ and $b, c \geq 0$ such that

$$(x_1 - x_2)^{m_1} f(x_1, x_2) = (x_1 - x_2)^{m_1} g(x_2, x_1) = \frac{p_1(x_1, x_2)}{x_1^b x_2^c}.$$

A careful application of partial associativity allows us to cancel (not simultaneously!) the polynomial terms on the left-hand sides to get the result (cf. Remarks 3.25 and 3.28 in [LL]).

The proofs of (iiia), (iiib) and (iiic) are similar. We show only (iiia). That is, we assume the truth of statements (E) and (F) and prove the truth of statement (A). We have

$$\frac{p_2(x_0,x_2)}{x_0^{a_2}(x_0+x_2)^{b_2}x_2^{c_2}} = f(x_0+x_2,x_2) = e^{x_2\frac{\partial}{\partial x_0}}f(x_0,x_2)$$
$$= e^{x_2\frac{\partial}{\partial x_0}}\frac{p_1(x_0,x_2)}{(x_0-x_2)^{a_1}x_0^{b_1}x_2^{c_1}}$$
$$= \frac{p_1(x_0+x_2,x_2)}{x_0^{a_1}(x_0+x_2)^{b_1}x_2^{c_1}},$$

where we used the consequence of Proposition 4.1.2. We use this consequence without comment below. It is easy to see that one can choose to have $a_1 = a_2$, $b_1 = b_2$ and $c_1 = c_2$. Assuming this, we have

$$p_2(x_0,x_2) = p_1(x_0+x_2,x_2).$$

Then we have

$$h(x_2,x_0) = \frac{p_1(x_0+x_2,x_2)}{x_0^{a_1}(x_2+x_0)^{b_1}x_2^{c_1}}.$$

Considering

$$\frac{p_1(x_1,x_2)}{x_0^{a_1}x_1^{b_1}x_2^{c_1}}$$

in place of $f(x_0,x_1,x_2)$ in Proposition 4.2.2 now gives the result. □

4.3 Vacuum-free vertex algebras

There are many variant definitions of vertex-type algebras. For instance, in [LL] the authors recall Borcherds' notion of vertex algebra [Bor1], but using the formalism of the Jacobi identity (see Definition 3.1.1 in [LL]), which, among other things, lacks a conformal vector, but does include a vacuum vector, the analogue of an identity in a commutative associative algebra. They purposely, for reasons of expository clarity, redundantly state as axioms two defining properties of the vacuum vector analogous to both the right and left identity properties. Indeed, they point out in Proposition 3.6.7 [LL] that the vacuum property is redundant. They further show, in Remark 3.6.8 [LL] that the creation property is not redundant. However, if we require as a separate axiom that the vertex operator map be injective, this asymmetry between the analogues of left and right identity disappears. Indeed, injectivity follows from the creation property, but, as we show (see Proposition 4.5.1), by introducing this further redundancy into the axioms it can be shown that either of the vacuum or creation properties follows from the other when all the other axioms are assumed. Because of this, we shall use injectivity in our statement of the axioms.

In the spirit of studying rings without identity, we may ask what we would have if we removed the vacuum vector altogether. Indeed, many of the various versions of vertex-type algebra already form a hierarchy of specialization as, for instance, vertex algebras specialize to quasi- (or Möbius) vertex algebras (cf. [FHL]), which in turn specialize to vertex operator algebras. So there is already ample precedent for a layered theory which we would be extending. For further justification, we note that the main axiom of any version of vertex-type algebra is some form of the Jacobi identity. It is hoped that by removing the assumption of having a vacuum vector, we avoid any premature distractions from this main axiom in the early development of the theory and that this development also shows more precisely the natural role that the vacuum vector plays vis-à-vis the Jacobi identity when we specialize to that case. With this as motivation, rather than any particular examples (although see [HL] and [BD]) and further, since it turns out that we can recover many results even in this pared-down setting, we proceed to define a vacuum-free vertex algebra:

Definition 4.3.1. A *vacuum-free vertex algebra* is a vector space equipped, first, with an injective linear map (the *vertex operator map)* $V \otimes V \to V[[x, x^{-1}]]$*,* or equivalently, a linear map

$$\begin{aligned} Y(\,\cdot\,,x):\quad V &\to (\mathrm{End}V)[[x,x^{-1}]] \\ v &\mapsto Y(v,x)=\sum_{n\in\mathbb{Z}} v_n x^{-n-1}. \end{aligned}$$

We call $Y(v,x)$ the *vertex operator associated with* v*.* We assume that

$$Y(u,x)v \in V((x))$$

for all $u, v \in V$. Finally, we require that the *Jacobi identity* is satisfied:

$$x_0^{-1}\delta\left(\frac{x_1 - x_2}{x_0}\right) Y(u, x_1)Y(v, x_2) - x_0^{-1}\delta\left(\frac{-x_2 + x_1}{x_0}\right) Y(v, x_2)Y(u, x_1)$$
$$= x_1^{-1}\delta\left(\frac{x_2 + x_0}{x_1}\right) Y(Y(u, x_0)v, x_2).$$

We can immediately apply Proposition 4.2.3 to obtain the next two results.

Proposition 4.3.1. *Let V be a vacuum-free vertex algebra. For all $u, v, w \in V$, we have:*

- *There exists some $m_1 \geq 0$ such that*

$$(x_1 - x_2)^{m_1} \left(Y(u, x_1)Y(v, x_2) - Y(v, x_2)Y(u, x_1)\right) = 0$$

(weak commutativity).

- *There exists some $m_2 \geq 0$ such that*

$$(x_0 + x_2)^{m_2} \left(Y(u, x_0 + x_2)Y(v, x_2)w - Y(Y(u, x_0)v, x_2)w\right) = 0$$

(weak associativity).

- *There exists some $m_3 \geq 0$ such that*

$$(x_1 - x_0)^{m_3} \left(Y(v, -x_0 + x_1)Y(u, x_1)w - Y(Y(u, x_0)v, x_1 - x_0)w\right) = 0$$

(weak skew-associativity).

- *There exist $p_1(x_1, x_2) \in V[[x_1, x_2]]$ and $a_1, b_1, c_1 \geq 0$ such that*

$$Y(u, x_1)Y(v, x_2)w = \frac{p_1(x_1, x_2)}{(x_1 - x_2)^{a_1} x_1^{b_1} x_2^{c_1}}$$

and

$$Y(v, x_2)Y(u, x_1)w = \frac{p_1(x_1, x_2)}{(-x_2 + x_1)^{a_1} x_1^{b_1} x_2^{c_1}}$$

(formal commutativity).

- *There exist* $p_2(x_0, x_2) \in V[[x_0, x_2]]$ *and* $a_2, b_2, c_2 \geq 0$ *such that*

$$Y(u, x_0 + x_2)Y(v, x_2)w = \frac{p_2(x_0, x_2)}{x_0^{a_2}(x_0 + x_2)^{b_2} x_2^{c_2}}$$

and

$$Y(Y(u, x_0)v, x_2)w = \frac{p_2(x_0, x_2)}{x_0^{a_2}(x_2 + x_0)^{b_2} x_2^{c_2}}$$

(formal associativity).

- *There exist* $p_3(x_0, x_1) \in V[[x_0, x_1]]$ *and* $a_3, b_3, c_3 \geq 0$ *such that*

$$Y(v, -x_0 + x_1)Y(u, x_1)w = \frac{p_3(x_0, x_1)}{x_0^{a_3} x_1^{b_3}(-x_0 + x_1)^{c_3}}$$

and

$$Y(Y(u, x_0)v, x_1 - x_0)w = \frac{p_3(x_0, x_1)}{x_0^{a_3} x_1^{b_3}(x_1 - x_0)^{c_3}}$$

(formal skew-associativity).

□

Remark 4.3.1. If we consider certain minimal values, which are "obviously pole clearing," that m_1, m_2 and m_3 may be taken to be in the above proposition, then it is easy to see that they could all be given by the same function on suitable ordered pairs of vectors. For instance, $m_1(u, v)$ could be defined as the negative of the least integer power of x appearing in $Y(u, x)v$, whenever that power is negative, and zero otherwise. As a corollary to this, we see that we could specify which vectors m_1, m_2 and m_3 may depend on, using a more refined statement, each one depending on only two vectors, but we shall not need or want this refinement in this work (see, for instance, Remarks 3.2.2 and 3.4.2 in [LL]).

Proposition 4.3.2. *Any two of weak commutativity, weak associativity and weak skew-associativity can replace the Jacobi identity in the definition of the notion of vacuum-free vertex algebra.*

□

Remark 4.3.2. As we have seen, Proposition 4.3.2 really is a statement of formal calculus and does not need any special information from the (vacuum-free) vertex algebra setting.

Remark 4.3.3. In previous treatments of vertex algebras (with a vacuum vector), as far as the author is aware, weak skew-associativity has been neglected since one can get by perfectly well with the other two weak properties. However, we shall see below that in some ways weak skew-associativity smoothes out some of the theory.

Remark 4.3.4. In the theory of vertex algebras with a vacuum vector, one obtains, without extra assumption, a $\mathcal{D}$ ("derivative") operator which allows one to derive the skew-symmetry relation. Since we do not have a $\mathcal{D}$ operator at this stage we cannot get such a relation, but we still have what we call "vacuum-free skew-symmetry," which is used occasionally without much comment in other treatments.

Proposition 4.3.3. (vacuum-free skew-symmetry) *Let V be a vacuum-free vertex algebra. For all u and $v \in V$, we have*

$$Y(Y(u,x_0)v,x_2) = Y(Y(v,-x_0)u,x_2+x_0).$$

Proof. Notice that the left-hand side of the Jacobi identity is invariant under the substitutions $(x_0,x_1,x_2;u,v) \leftrightarrow (-x_0,x_2,x_1;v,u)$. This means that we get the following relation coming from the right-hand side:

$$\begin{aligned} x_1^{-1}\delta\left(\frac{x_2+x_0}{x_1}\right)Y(Y(u,x_0)v,x_2) &= x_2^{-1}\delta\left(\frac{x_1-x_0}{x_2}\right)Y(Y(v,-x_0)u,x_1) \\ &= x_1^{-1}\delta\left(\frac{x_2+x_0}{x_1}\right)Y(Y(v,-x_0)u,x_2+x_0), \end{aligned}$$

so that taking the residue with respect to x_1 yields the result. □

It turns out that it is enough to know vacuum-free skew-symmetry together with any single one of weak commutativity, weak associativity, or weak skew-associativity in order to recover the entire Jacobi identity. We prove each of these equivalences, arguing in a similar spirit to the proofs given in [LL], where the authors used certain classical guides. Our classical guides use relationships found in commutative associative algebras without identity element, but we shall not have all such relationships. Indeed, the analogues we use are as follows:

$a(bc) = b(ac)$ corresponds to weak commutativity;

$a(bc) = (ab)c$ corresponds to weak associativity;

$a(bc) = (ba)c$ corresponds to weak skew-associativity;

and each of

$a(bc) = (bc)a,$

$(ab)c = (ba)c,$

$a(bc) = a(cb)$ corresponds to vacuum-free skew symmetry.

Proposition 4.3.4. *Weak associativity together with vacuum-free skew-symmetry can replace the Jacobi identity in the definition of the notion of vacuum-free vertex algebra.*

Proof. We follow this analogue: $a(bc) = (ab)c = (ba)c$. Let V be a vacuum-free vertex algebra. We shall show that we get weak skew-associativity, which will be enough. In fact, for all $u, v \in V$, there exists $l \geq 0$ such that

$$\begin{aligned}(x_1 - x_0)^l Y(u, -x_0 + x_1)Y(v, x_1)w &= (x_1 - x_0)^l Y(Y(u, -x_0)v, x_1)w \\ &= (x_1 - x_0)^l Y(Y(v, x_0)u, x_1 - x_0)w.\end{aligned}$$

□

Proposition 4.3.5. *Weak skew-associativity together with vacuum-free skew-symmetry can replace the Jacobi identity in the definition of the notion of vacuum-free vertex algebra.*

Proof. We follow this analogue: $a(bc) = (ba)c = (ab)c$. Let V be a vacuum-free vertex algebra. We shall show that we get weak associativity which will be enough. In fact, for all $u, v, w \in V$, there exists $l \geq 0$ such that

$$\begin{aligned}(x_0 + x_2)^l Y(v, x_0 + x_2)Y(u, x_2)w &= (x_0 + x_2)^l Y(Y(u, -x_0)v, x_2 + x_0)w \\ &= (x_0 + x_2)^l Y(Y(v, x_0)u, x_2)w.\end{aligned}$$

□

Proposition 4.3.6. *Weak commutativity together with vacuum-free skew-symmetry can replace the Jacobi identity in the definition of the notion of vacuum-free vertex algebra.*

Proof. We follow this analogue: $a(bc) = (bc)a = (cb)a = a(cb) = c(ab) = (ab)c$. Let V be a vacuum-free vertex algebra. We shall show that we get weak associativity, which will be enough. For all $u, v, w \in V$, it is easy to see that there exists $l \geq 0$ such that:

$$
\begin{aligned}
&(x_0 + x_2)^l Y(Y(u, x_0 + x_2)Y(v, x_2)w, x_3) \\
&= (x_0 + x_2)^l Y(Y(Y(v, x_2)w, -x_0 - x_2)u, x_3 + (x_0 + x_2)) \\
&= (x_0 + x_2)^l Y(Y(Y(w, -x_2)v, -x_0)u, x_3 + (x_0 + x_2)) \\
&= (x_0 + x_2)^l Y(Y(u, x_0)Y(w, -x_2)v, x_3 + x_2) \\
&= (x_0 + x_2)^l Y(Y(w, -x_2)Y(u, x_0)v, x_3 + x_2) \\
&= (x_0 + x_2)^l Y(Y(Y(u, x_0)v, x_2)w, x_3),
\end{aligned}
$$

so that the result follows from the injectivity of the vertex operators. □

Remark 4.3.5. We did not use the injectivity property of vertex operators in the proofs of Propositions 4.3.4 and 4.3.5, but we did use it in the proof of Proposition 4.3.6.

4.4 Modules

In this section we define the notion of module for a vacuum-free vertex algebra and show a series of results paralleling those of Section 4.3, with one significant exception. We do not have that (along with module skew-symmetry) module weak commutativity can be a replacement axiom, although we do get that module weak associativity and module weak skew-associativity may be used as replacement axioms. A heuristic reason for this may be seen in the fact that in the commutative associative guides, "c" did not remain in the rightmost position for the case of weak commutativity, but it did for the other two cases. We do have a module weak skew-symmetry, which is in a contrast of sorts to the situation with a vertex algebra, where one is tempted to ignore any special skew-symmetric-like property of the module case, since the underlying vertex algebra skew-symmetry is all that one needs. We have already done all the work for this section. We state the results for completeness.

Definition 4.4.1. A *module* for a vacuum-free vertex algebra is a vector space W equipped with a linear map $V \otimes W \to W[[x, x^{-1}]]$, or equivalently, a linear map

$$\begin{aligned} Y_W(\,\cdot\,, x) : V &\to (\mathrm{End} W)[[x, x^{-1}]] \\ v &\mapsto Y_W(v, x) = \sum_{n \in \mathbb{Z}} v_n x^{-n-1}. \end{aligned}$$

We assume that

$$Y_W(u, x)w \in W((x))$$

for all $u \in V$ and $w \in W$. Then finally, we require that the *Jacobi identity* is satisfied:

$$\begin{aligned} x_0^{-1}\delta\left(\frac{x_1 - x_2}{x_0}\right) Y_W(u, x_1)Y_W(v, x_2)w - x_0^{-1}\delta\left(\frac{-x_2 + x_1}{x_0}\right) Y_W(v, x_2)Y_W(u, x_1)w \\ = x_1^{-1}\delta\left(\frac{x_2 + x_0}{x_1}\right) Y_W(Y(u, x_0)v, x_2)w. \end{aligned}$$

Remark 4.4.1. This definition adheres to the principle that a module should satisfy all the defining properties of a vertex algebra that make sense, which is essentially the *a priori* motivation given on page 117 in [LL].

We present the results in parallel order to those in Section 4.3.

Proposition 4.4.1. *Let V be a vacuum-free vertex algebra with module W. For all $u, v \in V$, and $w \in W$, we have:*

- *There exists some $m_1 \geq 0$ such that*

$$(x_1 - x_2)^{m_1} \left(Y_W(u, x_1)Y_W(v, x_2) - Y_W(v, x_2)Y_W(u, x_1)\right) = 0$$

  (weak commutativity).

- *There exists some $m_2 \geq 0$ such that*

$$(x_0 + x_2)^{m_2} \left(Y_W(u, x_0 + x_2)Y_W(v, x_2)w - Y_W(Y(u, x_0)v, x_2)w\right) = 0$$

  (weak associativity).

- *There exists some* $m_3 \geq 0$ *such that*

$$(x_1 - x_0)^{m_3} \left(Y_W(v, -x_0 + x_1)Y_W(u, x_1)w - Y_W(Y(u, x_0)v, x_1 - x_0)w\right) = 0$$

(weak skew-associativity).

- *There exist* $p_1(x_1, x_2) \in W[[x_1, x_2]]$ *and* $a_1, b_1, c_1 \geq 0$ *such that*

$$Y_W(u, x_1)Y_W(v, x_2)w = \frac{p_1(x_1, x_2)}{(x_1 - x_2)^{a_1} x_1^{b_1} x_2^{c_1}}$$

and

$$Y_W(v, x_2)Y_W(u, x_1)w = \frac{p_1(x_1, x_2)}{(-x_2 + x_1)^{a_1} x_1^{b_1} x_2^{c_1}}$$

(formal commutativity).

- *There exist* $p_2(x_0, x_2) \in W[[x_0, x_2]]$ *and* $a_2, b_2, c_2 \geq 0$ *such that*

$$Y_W(u, x_0 + x_2)Y_W(v, x_2)w = \frac{p_2(x_0, x_2)}{x_0^{a_2}(x_0 + x_2)^{b_2} x_2^{c_2}}$$

and

$$Y_W(Y(u, x_0)v, x_2)w = \frac{p_2(x_0, x_2)}{x_0^{a_2}(x_2 + x_0)^{b_2} x_2^{c_2}}$$

(formal associativity).

- *There exist* $p_3(x_0, x_1) \in W[[x_0, x_1]]$ *and* $a_3, b_3, c_3 \geq 0$ *such that*

$$Y_W(v, -x_0 + x_1)Y_W(u, x_1)w = \frac{p_3(x_0, x_1)}{x_0^{a_3} x_1^{b_3} (-x_0 + x_1)^{c_3}}$$

and

$$Y_W(Y(u, x_0)v, x_1 - x_0)w = \frac{p_3(x_0, x_1)}{x_0^{a_3} x_1^{b_3} (x_1 - x_0)^{c_3}}$$

(formal skew-associativity).

□

Remark 4.4.2. Concerning m_1, m_2 and m_3, see Remark 4.3.1.

Proposition 4.4.2. *Any two of weak commutativity, weak associativity and weak skew-associativity (in the sense of Proposition 4.4.1) can replace the Jacobi identity in the definition of the notion of module for a vacuum-free vertex algebra.*

□

Proposition 4.4.3. (vacuum-free skew-symmetry)

For all u and $v \in V$, a vacuum-free vertex algebra with module W, we get the following relation:

$$Y_W(Y(u,x_0)v,x_2) = Y_W(Y(v,-x_0)u,x_2+x_0).$$

Proof. The proof is essentially the same as for Proposition 4.3.3. □

Proposition 4.4.4. *In the definition of the notion of module for a vacuum-free vertex algebra, the Jacobi identity can be replaced by weak associativity (in the sense of Proposition 4.4.1) together with vacuum-free skew-symmetry (in the sense of Proposition 4.4.3).*

Proof. The proof is essentially the same as for Proposition 4.3.4. □

Proposition 4.4.5. *In the definition of the notion of module for a vacuum-free vertex algebra, the Jacobi identity can be replaced by weak skew-associativity (in the sense of Proposition 4.4.1) together with vacuum-free skew-symmetry (in the sense of Proposition 4.4.3).*

Proof. The proof is essentially the same as for Proposition 4.3.5. □

4.5 Vertex algebras with vacuum

We now consider the case of a vacuum-free vertex algebra in which one of the vertex operators acts as the identity. That is, given a vertex algebra V we have a distinguished vector $\mathbf{1} \in V$, which we call the vacuum vector, with the property that $Y(\mathbf{1},x) = 1$, where 1 is the identity endomorphism of V. Continuing the analogy with commutative associative algebras from previous sections, we see that such a vector is analogous to a left identity map. We may then wonder if there is some sort of right identity property. Since vacuum-free skew-symmetry switches the order of the two vectors, we consider specializing to the case where one of the two vectors is the vacuum vector. We get

$$\begin{aligned} Y(Y(u,x)\mathbf{1},z) &= Y(Y(\mathbf{1},-x)u,z+x) \\ &= Y(u,z+x), \end{aligned} \tag{4.5.1}$$

from which it is easy to see that $Y(u,x)\mathbf{1}$ must be a formal power series in x. Therefore we may set $x=0$, to check the constant term, which gives us

$$Y(u_{-1}\mathbf{1},z) = Y(u,z),$$

which by injectivity gives

$$u_{-1}\mathbf{1} = u.$$

With this as motivation, we recall ([Bor1]; cf. [LL]) the definition of vertex algebra (with vacuum) and although the definition contains redundancies we state both the left and right identity properties for purposes of clarity and also since this is traditional.

Definition 4.5.1. A *vertex algebra* is a vacuum-free vertex algebra V together with a distinguished element $\mathbf{1}$ satisfying the following *vacuum property* :

$$Y(\mathbf{1},x) = 1$$

and *creation property* :

$$\begin{gathered} Y(u,x)\mathbf{1} \in V[[x]] \text{ and} \\ Y(u,0)\mathbf{1} = u \text{ for all } u \in V. \end{gathered}$$

We refer to all the parts of the definition of a vertex algebra except for the Jacobi identity as "minor axioms." Consider again the following:

$$\begin{aligned} Y(Y(u,x)\mathbf{1},z) &= Y(Y(\mathbf{1},-x)u,z+x) \\ &= Y(u,z+x) \\ &= e^{x\frac{d}{dz}}Y(u,z). \end{aligned}$$

Checking the first degree term in x gives

$$Y(u_{-2}\mathbf{1}, z) = \frac{d}{dz}Y(u, z).$$

Thus vertex operators are closed under differentiation. Furthermore, we define the $\mathcal{D}$ operator.

Definition 4.5.2. Given a vertex algebra, V, define $\mathcal{D} \in \mathrm{End}(V)$ by

$$\mathcal{D}v = v_{-2}\mathbf{1}.$$

We may now write the $\mathcal{D}$-*derivative property* :

$$Y(\mathcal{D}u, z) = \frac{d}{dz}Y(u, z). \tag{4.5.2}$$

Remark 4.5.1. We note that the philosophy behind this proof of the $\mathcal{D}$-derivative property is again a classical analogue coming from commutative associative algebras with identity, namely, we used as a guide the relation $a \cdot 1 = 1 \cdot a$. In Proposition 3.1.18 [LL], the authors obtain the $\mathcal{D}$-derivative property in a different fashion. Their point of view was to observe that $v_{-2}\mathbf{1}$ is the component of a certain "iterate" and then to look at the "iterate formula", equation 3.1.11 [LL] and "slice down" to get the correct component. We never need to make use of the iterate formula in this work and, in fact, never use the word "iterate" except informally or to say that we won't use it. As we shall see (Propositions 4.5.1, 4.5.2 and 4.5.3 and Remarks 4.5.2 and 4.5.4), the basic theory of the minor properties of a vertex algebra can be handled entirely with vacuum-free skew-symmetry without reference to the Jacobi identity or the iterate formula. However, we also note that the connection between the $\mathcal{D}$-derivative property and properties similar to the "iterate" formula, namely, weak associativity and weak skew-associativity, play an essential role in this treatment, as we see in Propositions 4.5.7, 4.5.10 and 4.6.2, which are used to obtain Theorem 4.6.1.

In the introduction to Section 4.3 we mentioned the equivalence of the vacuum and creation properties provided we separately state injectivity as an axiom. In the introduction to this section we saw how in the presence of vacuum-free skew-symmetry and the other minor axioms, the vacuum property implies the creation property. We now show the converse.

Proposition 4.5.1. *In the presence of vacuum-free skew-symmetry and the other minor axioms of a vertex algebra, the vacuum property and the creation property each imply the other.*

Proof. We have already seen how the vacuum property implies the creation property. We begin the converse statement in a similar fashion by specializing one of the vectors in the formula for vacuum-free skew-symmetry to be $\mathbf{1}$. We get:

$$\begin{aligned} Y(Y(\mathbf{1},x)v,z) &= Y(Y(v,-x)\mathbf{1},z+x) \\ &= e^{x\frac{d}{dz}}Y(Y(v,-x)\mathbf{1},z). \end{aligned} \tag{4.5.3}$$

which by the first part of the creation property gives us that $Y(\mathbf{1},x)v$ is a power series in x. Then extracting the constant term in x we have

$$Y(\mathbf{1}_{-1}v,z) = Y(v_{-1}\mathbf{1},z),$$

which by the second part of the creation property and by injectivity (which follows also from the second part of the creation property as is usually argued) we have

$$\mathbf{1}_{-1}v = v.$$

We now need to show that $\mathbf{1}_{-n}v = 0$ for $n \geq 2$, or in other words that $\frac{d}{dz}Y(\mathbf{1},z) = 0$. It is tempting to try and use the $\mathcal{D}$-derivative property to"peel off" the "outer" Y operator in (4.5.3) but remember that we used the vacuum property to get the $\mathcal{D}$-derivative property so this is not available to us. Instead, we try to imitate the process of getting the $\mathcal{D}$-derivative property by checking the linear term in x. This gives us

$$Y(\mathbf{1}_{-2}v,z) = -Y(v_{-2}\mathbf{1},z) + \frac{d}{dz}Y(v,z).$$

Then further specializing by setting $v = \mathbf{1}$, we have

$$2Y(\mathbf{1}_{-2}\mathbf{1},z) = \frac{d}{dz}Y(\mathbf{1},z). \tag{4.5.4}$$

Acting against $\mathbf{1}$ and extracting the constant term in z gives us, by the second part of the creation property, that

$$2\mathbf{1}_{-2}\mathbf{1} = \mathbf{1}_{-2}\mathbf{1},$$

which gives us that $\mathbf{1}_{-2}\mathbf{1} = 0$. Then substituting back into (4.5.4) gives

$$\frac{d}{dz}Y(\mathbf{1}, z) = 0,$$

which is what we needed to show. □

Remark 4.5.2. As was mentioned in the introduction to Section 4.3, Proposition 3.6.8 [LL] shows that in the presence of the other axioms the vacuum property may be derived. Our proof of that direction in Proposition 4.5.1 shares some common features with the proof in [LL], but one difference that is perhaps worth pointing out is that our assumption was weaker. We needed only to assume vacuum-free skew-symmetry whereas the proof in [LL] explicitly used the Jacobi identity. As was discussed in Remark 4.5.1 much of the theory of the minor properties associated with the vacuum vector can be handled with only the use of vacuum-free skew-symmetry (see Propositions 4.5.2 and 4.5.3 and Remark 4.5.4 below).

We now derive some of the standard "minor properties." Taking the exponential generating function of the higher derivatives and using the $\mathcal{D}$-derivative property (and the formal Taylor theorem) gives

$$Y(e^{x\mathcal{D}}u, z) = e^{x\frac{d}{dz}}Y(u, z) = Y(u, z + x), \tag{4.5.5}$$

which by (4.5.1) gives

$$\begin{aligned} Y(e^{x\mathcal{D}}u, z) &= Y(u, z + x) \\ &= Y(Y(u, x)\mathbf{1}, z), \end{aligned}$$

which, by the injectivity of vertex operators, gives the *strong creation property* :

$$Y(u, x)\mathbf{1} = e^{x\mathcal{D}}u. \tag{4.5.6}$$

We again consider vacuum-free skew-symmetry in light of the $\mathcal{D}$ operator, where we now have:

$$\begin{aligned} Y(Y(u, x)v, z) &= Y(Y(v, -x)u, z + x) \\ &= Y(e^{x\mathcal{D}}Y(v, -x)u, z), \end{aligned}$$

which, by the injectivity of vertex operators, gives us *skew-symmetry*:

$$Y(u,x)v = e^{x\mathcal{D}}Y(v,-x)u. \tag{4.5.7}$$

We may take the derivative of this skew-symmetry formula to get:

$$\begin{aligned}\frac{d}{dx}Y(u,x)v &= \mathcal{D}e^{x\mathcal{D}}Y(v,-x)u + e^{x\mathcal{D}}\frac{d}{dx}Y(v,-x)u\\ &= \mathcal{D}Y(u,x)v + e^{x\mathcal{D}}\frac{d}{dx}Y(v,-x)u\\ &= \mathcal{D}Y(u,x)v - e^{x\mathcal{D}}Y(\mathcal{D}v,-x)u\\ &= \mathcal{D}Y(u,x)v - Y(u,x)\mathcal{D}v, \end{aligned} \tag{4.5.8}$$

where we used skew-symmetry to get the first, second and fourth equalities and the $\mathcal{D}$-derivative property to get the third equality. Observe that the last expression is a commutator, which in fact gives us the *$\mathcal{D}$-bracket derivative property*:

$$[\mathcal{D}, Y(u,x)] = \frac{d}{dx}Y(u,x). \tag{4.5.9}$$

Rearranging the terms of the $\mathcal{D}$-bracket formula makes it resemble a product rule:

$$\mathcal{D}Y(u,x)v = \frac{d}{dx}Y(u,x)v + Y(u,x)\mathcal{D}v. \tag{4.5.10}$$

Of course, because of the $\mathcal{D}$-derivative property, we also have

$$[\mathcal{D}, Y(u,x)] = Y(\mathcal{D}u,x).$$

which, when the terms are rearranged, becomes

$$\mathcal{D}Y(u,x)v = Y(\mathcal{D}u,x)v + Y(u,x)\mathcal{D}v. \tag{4.5.11}$$

Remark 4.5.3. Whereas the $\mathcal{D}$-derivative property may be thought of as an analogue of the power rule for differentiation, the $\mathcal{D}$-bracket derivative property may be thought of as an analogue of the product rule. Indeed, while it is often messy to check what properties look like component-wise, in the case of these two properties the formulas are very familiar. For the $\mathcal{D}$-derivative property we have

$$(\mathcal{D}u)_n = -n(\mathcal{D}u)_{n-1}.$$

For the $\mathcal{D}$-bracket derivative property (as rearranged in the form given in (4.5.11)) we have

$$\mathcal{D}(u_n v) = (\mathcal{D}u)_n v + u_n \mathcal{D}v.$$

Obviously, by the same reasoning as that which we gave for the automorphism property, we have that (4.5.11) gives

$$e^{z\mathcal{D}} Y(u,x)v = Y(e^{z\mathcal{D}}u,x)e^{z\mathcal{D}}v, \tag{4.5.12}$$

which by (4.5.5) gives us

$$e^{z\mathcal{D}} Y(u,x)v = Y(u,x+z)e^{z\mathcal{D}}v, \tag{4.5.13}$$

which formula also follows directly from (4.5.10) and the formal Taylor theorem, again using the same reasoning as the proof of the automorphism property.

We have seen that skew-symmetry, together with the $\mathcal{D}$-derivative property, gives us the $\mathcal{D}$-bracket derivative property. Conversely, we may also derive the $\mathcal{D}$-derivative property assuming only skew-symmetry and the $\mathcal{D}$-bracket derivative property. A careful consideration of (4.5.8), where our assumption is now that the last line is equal to the first line, gives us:

$$\begin{aligned}
Y(u,x)\mathcal{D}v &= -e^{x\mathcal{D}}\frac{d}{dx}Y(v,-x)u \Leftrightarrow \\
e^{-x\mathcal{D}}Y(u,x)\mathcal{D}v &= -\frac{d}{dx}Y(v,-x)u \Leftrightarrow \\
Y(\mathcal{D}v,-x)u &= -\frac{d}{dx}Y(v,-x)u,
\end{aligned}$$

which is the $\mathcal{D}$-derivative property stated for $-x$. Given the $\mathcal{D}$-derivative property and skew-symmetry, we may also recover vacuum-free skew-symmetry as can be seen by the following calculation:

$$\begin{aligned}
Y(u,x)v &= e^{x\mathcal{D}}Y(v,-x)u \Leftrightarrow \\
Y(Y(u,x)v,z) &= Y(e^{x\mathcal{D}}Y(v,-x)u,z) \Leftrightarrow \\
Y(Y(u,x)v,z) &= Y(Y(v,-x)u,z+x).
\end{aligned}$$

We summarize some of our implications in the next two propositions.

Proposition 4.5.2. *In the presence of only the minor axioms of a vertex algebra, but excluding the creation property, vacuum-free skew-symmetry implies the strong creation property, skew-symmetry, the $\mathcal{D}$-derivative property, and the $\mathcal{D}$-bracket derivative property.*

$\square$

Proposition 4.5.3. *In the presence of the minor axioms of a vertex algebra, the following are equivalent:*

- (i) *vacuum-free skew-symmetry*
- (ii) *skew-symmetry together with the $\mathcal{D}$-derivative property*
- (iii) *skew-symmetry together with the $\mathcal{D}$-bracket derivative property.*

$\square$

Remark 4.5.4. Our development of Proposition 4.5.2 largely parallels portions of the material presented in Section 3.1 of [LL]. Perhaps the main difference is that our official proof of the $\mathcal{D}$-derivative property is based on vacuum-free skew-symmetry instead of the iterate formula (see equation 3.1.11 and Proposition 3.1.18 [LL]) and, more generally as well as more roughly, that our point of view is that all of the minor properties are due to vacuum-free skew-symmetry even without the full Jacobi identity.

Recall that we began this section by substituting the vacuum vector into the formula for vacuum-free skew-symmetry. We may pursue a similar analysis with other formulas to further describe the dependencies of weaker axioms on stronger ones.

Proposition 4.5.4. *In the presence of the minor axioms of a vertex algebra, the strong creation property follows from any single one of skew-symmetry, the $\mathcal{D}$-bracket derivative property, or the $\mathcal{D}$-derivative property.*

Proof. We first assume the $\mathcal{D}$-derivative property. By the $\mathcal{D}$-derivative property we have

$$Y(v, z+x)\mathbf{1} = Y(e^{x\mathcal{D}}v, z)\mathbf{1},$$

which by two applications of the creation property allows us to set $z = 0$ and calculate the right-hand side to get

$$Y(v, x)\mathbf{1} = e^{x\mathcal{D}}v,$$

which is the strong creation property.

We now assume the $\mathcal{D}$-bracket derivative property. First, note that by the vacuum property $\mathcal{D}\mathbf{1} = \mathbf{1}_{-2}\mathbf{1} = 0$, so that $e^{x\mathcal{D}}\mathbf{1} = \mathbf{1}$. Then by (4.5.13) we have

$$e^{x\mathcal{D}}Y(v, z)\mathbf{1} = Y(v, z + x)\mathbf{1},$$

which again by two applications of the creation property gives the strong creation property.

Finally, we assume skew-symmetry. We have

$$Y(u, x)\mathbf{1} = e^{x\mathcal{D}}Y(\mathbf{1}, -x)u = e^{x\mathcal{D}}u,$$

which is a third time, the strong creation property. □

We continue to specialize our formulas by substituting in the vacuum vector. In the next proposition, we have the relation $a(b1) = b(a1)$ as a classical guide.

Proposition 4.5.5. *In the presence of the minor axioms of a vertex algebra, skew-symmetry follows from weak commutativity together with the $\mathcal{D}$-bracket derivative property.*

Proof. Let V be a vertex algebra. Let $u, v \in V$. By weak commutativity there exists some $k \geq 0$ such that:

$$\begin{aligned}(x - z)^k Y(u, x)Y(v, z)\mathbf{1} &= (x - z)^k Y(v, z)Y(u, x)\mathbf{1}\\ &= (x - z)^k Y(v, z)e^{x\mathcal{D}}u\\ &= (x - z)^k e^{x\mathcal{D}}Y(v, z - x)u,\end{aligned}$$

and by the creation property we may set $z = 0$ and cancel x^k which gives skew-symmetry. □

In fact, we have more:

Proposition 4.5.6. *In the presence of the minor axioms of a vertex algebra, weak commutativity together with the $\mathcal{D}$-bracket derivative property are equivalent to the Jacobi identity.*

Proof. The result follows from Proposition 4.5.5, Proposition 4.5.3 and Proposition 4.3.6. $\square$

Remark 4.5.5. Proposition 4.5.6 appeared in Theorem 3.5.1 [LL], where Proposition 4.5.5 was obtained during the course of the proof. Our development is similar to, but a variant of, the proof presented in [LL].

We next consider specializing the weak associativity property. We have

Proposition 4.5.7. *In the presence of the minor axioms of a vertex algebra, weak associativity implies the $\mathcal{D}$-derivative property.*

Proof. Let V be a vertex algebra. Let $u, w \in V$. There exists $l \geq 0$ such that

$$(x_0 + x_2)^l Y(u, x_0 + x_2)Y(\mathbf{1}, x_2)w = (x_0 + x_2)^l Y(Y(u, x_0)\mathbf{1}, x_2)w,$$

and also such that the left-hand side of the equation is written in terms of nonnegative powers of $(x_0 + x_2)$. Thus we have:

$$\begin{aligned}(x_0 + x_2)^l Y(u, x_0 + x_2)Y(\mathbf{1}, x_2)w &= (x_0 + x_2)^l Y(u, x_2 + x_0)Y(\mathbf{1}, x_2)w \\ &= (x_0 + x_2)^l Y(u, x_2 + x_0)w.\end{aligned}$$

Then substituting this, we notice that x_0 is appropriately truncated so that we can cancel $(x_0 + x_2)^l$ by multiplying by $(x_2 + x_0)^{-l}$ and applying partial associativity. This gives us:

$$\begin{aligned}Y(u, x_2 + x_0)w = Y(Y(u, x_0)\mathbf{1}, x_2)w \Leftrightarrow \\ e^{x_0 \frac{d}{dx_2}} Y(u, x_2)w = Y(Y(u, x_0)\mathbf{1}, x_2)w.\end{aligned}$$

Looking at the linear term in x_0 gives the result. $\square$

Proposition 4.5.8. *In the presence of the minor axioms of a vertex algebra, weak associativity together with the strong creation property imply the $\mathcal{D}$-bracket derivative formula.*

Proof. Let V be a vertex algebra. Let $u, v \in V$. There exists $l \geq 0$ such that:

$$(x_0 + x_2)^l Y(u, x_0 + x_2) Y(v, x_2)\mathbf{1} = (x_0 + x_2)^l Y(Y(u, x_0)v, x_2)\mathbf{1} \Leftrightarrow$$
$$(x_0 + x_2)^l Y(u, x_0 + x_2) e^{x_2 \mathcal{D}} v = (x_0 + x_2)^l e^{x_2 \mathcal{D}} Y(u, x_0) v,$$

which has only nonnegative powers of x_2 so that we may cancel $(x_0 + x_2)^l$ which gives us the $\mathcal{D}$-bracket derivative formula. □

Remark 4.5.6. Propositions 4.5.7 and 4.5.8 were already essentially obtained as Proposition 2.6 in [Li2]. In fact, Proposition 2.6 in [Li2] was a stronger result, which shows that the assumption of the strong creation property could have been removed from Proposition 4.5.8. Comparing arguments, we note that in the proof of Proposition 4.5.8 we could have canceled $(x_0 + x_2)^l$ in the first line and extracted the coefficient of x_2 using the creation property instead of extracting the coefficients of all powers of x_2 using the strong creation property. We would have arrived at the "unexponentiated" form of the $\mathcal{D}$-bracket derivative formula instead of the "exponentiated" form we did arrive at, in parallel with the fact that the creation property is an unexponentiated form of the strong creation property. Corollary 2.7 in [Li2] recovers the relevant "exponentiated" identities as a consequence.

Proposition 4.5.9. *In the presence of the minor axioms of a vertex algebra, weak associativity together with skew-symmetry are equivalent to the Jacobi identity.*

Proof. By Proposition 4.5.4 we have the strong creation property and so by Proposition 4.5.8 we have the $\mathcal{D}$-bracket derivative formula. Then by Proposition 4.5.3 we have vacuum-free skew-symmetry and so the result follows from Proposition 4.3.4. □

We also have a slight variant proof of the last proposition.

Proof. By Proposition 4.5.7 we have the $\mathcal{D}$-derivative property so that by Proposition 4.5.3 we have vacuum-free skew-symmetry and so the result follows from Proposition 4.3.4. □

Remark 4.5.7. Proposition 4.5.9 appeared in Theorem 3.6.1 in [LL], and Proposition 4.5.8 was essentially obtained in the course of their proof. Our present result generalizes more easily to the module case. In [LL] the authors needed a further argument, which they formulated in Theorem 3.6.3 [LL], in order to obtain the corresponding result for a module. In the course of the proof of Theorem 3.6.3 in [LL], Proposition 4.5.7 was also obtained, though not stated separately. It was our interest in seeking an alternative to Theorem 3.6.3 in [LL] that was the original motivation for the work in this chapter.

And finally, we consider weak skew-associativity.

Proposition 4.5.10. *In the presence of the minor axioms of a vertex algebra, weak skew-associativity implies the $\mathcal{D}$-derivative property.*

Proof. Let V be a vertex algebra. Let $u, w \in V$. There exists $m \geq 0$ such that:

$$\begin{aligned}
(x_1 - x_0)^m Y(u, x_1)w = (x_1 - x_0)^m Y(Y(u, x_0)\mathbf{1}, x_1 - x_0)w &\Leftrightarrow \\
Y(u, x_1)w = Y(Y(u, x_0)\mathbf{1}, x_1 - x_0)w &\Leftrightarrow \\
Y(u, x_1)w = e^{-x_0 \frac{d}{dx_1}} Y(Y(u, x_0)\mathbf{1}, x_1)w &\Leftrightarrow \\
e^{x_0 \frac{d}{dx_1}} Y(u, x_1)w = Y(Y(u, x_0)\mathbf{1}, x_1)w&,
\end{aligned}$$

where the cancellation of $(x_1 - x_0)^m$ was justified because both sides had only non-negative powers of x_0. Then taking coefficient of the first power of x_0 gives us the $\mathcal{D}$-derivative property. $\square$

We can substitute $\mathbf{1}$ for still another vector to get:

Proposition 4.5.11. *In the presence of the minor axioms of a vertex algebra, the following are equivalent:*

(i) weak skew-associativity together with skew-symmetry

(ii) weak skew-associativity together with the $\mathcal{D}$-bracket derivative property.

Proof. Let V be a vector space satisfying the relevant axioms. Let $u, v \in V$. We assume V satisfies weak skew-associativity. By Proposition 4.5.10 we have the $\mathcal{D}$-derivative property. Then by Proposition 4.5.4 we have the strong creation property. Then we have:

$$(x_1 - x_0)^m Y(v, -x_0 + x_1) Y(u, x_1)\mathbf{1} = (x_1 - x_0)^m Y(Y(u, x_0)v, x_1 - x_0)\mathbf{1} \Leftrightarrow$$
$$(x_1 - x_0)^m Y(v, -x_0 + x_1) e^{x_1 \mathcal{D}} u = (x_1 - x_0)^m e^{(x_1 - x_0)\mathcal{D}} Y(u, x_0)v.$$

Observing that both sides are truncated from below in x_1 appropriately we can cancel $(x_1 - x_0)^m$ from both sides to get:

$$Y(v, -x_0 + x_1) e^{x_1 \mathcal{D}} u = e^{(x_1 - x_0)\mathcal{D}} Y(u, x_0)v \Leftrightarrow$$
$$e^{-x_1 \mathcal{D}} Y(v, -x_0 + x_1) e^{x_1 \mathcal{D}} u = e^{-x_0 \mathcal{D}} Y(u, x_0)v,$$

from which it is clear that either the $\mathcal{D}$-bracket derivative property (in exponentiated form) or skew-symmetry each implies the other. □

Remark 4.5.8. We note that the argument in the proof of Proposition 4.5.11 could have been changed to depend on only the creation property instead of the strong creation property in a manner similar to the changes discussed in Remark 4.5.6.

We can now state two more replacement axioms for the Jacobi identity.

Proposition 4.5.12. *In the presence of the minor axioms of a vertex algebra, weak skew-associativity together with either single one of skew-symmetry or the $\mathcal{D}$-bracket derivative property is equivalent to the Jacobi identity.*

Proof. By Proposition 4.5.11 the two statements each follow from the other. Therefore it is enough to show the case where we assume weak skew-associativity together with skew-symmetry. By Proposition 4.5.10 we have the $\mathcal{D}$-derivative property, so that by Proposition 4.5.3 we have vacuum-free skew-symmetry, which in turn gives us the result by Proposition 4.3.5. □

4.6 Modules for a vertex algebras with vacuum

In this section, we give the parallel results to those in Section 4.4, where we now consider modules for a vertex algebra (with vacuum). Since any such module may also be viewed as a module for a vacuum-free vertex algebra, most of the results carry over without comment so we content ourselves with only discussing certain new statements that we get. Most importantly, we show that in the notion of module for a vertex algebra, the Jacobi identity can be replaced by either one (without the other) of weak associativity or weak skew-associativity (in the sense of Proposition 4.4.1).

Definition 4.6.1. A *module* for a vertex algebra V is a vector space W which is a vacuum-free module for V when viewed as a vacuum-free vertex algebra which further satisfies *the vacuum property*

$$Y_W(\mathbf{1}, x) = 1,$$

where 1 is the identity operator on W.

Remark 4.6.1. We do not have an axiom for a module-type of creation property, and this is not merely that we have chosen to remove any redundancy from our axioms. Indeed, our modules are really behaving as left modules and so it does not make sense to have a right identity property, since we cannot act on the vacuum vector.

Following the proof of either Proposition 4.5.7 or Proposition 4.5.10, we have the following *$\mathcal{D}$-derivative property*:

Proposition 4.6.1. *Let W be a module for a vertex algebra V. Then for any $v \in V$, we have*

$$Y_W(\mathcal{D}v, x) = \frac{d}{dx} Y_W(v, x).$$

$\square$

In fact, we have more, since the proofs of Proposition 4.5.7 and Proposition 4.5.10 imply the following:

Proposition 4.6.2. *In the presence of the minor axioms of a module for a vertex algebra, either single one of weak associativity or weak skew-associativity (each in the sense of Proposition 4.4.1) implies the $\mathcal{D}$-derivative property (in the sense of Proposition 4.6.1).*

□

We now conclude with the main result of this chapter. We have already done all the work. The result for weak associativity was obtained in Theorem 4.4.5 in [LL]. It is this result as regards weak associativity (or more precisely, an easy corollary of it, Corollary 4.4.7 [LL]) which entered into the proof in [LL] showing the equivalence of the notion of representation of a vertex algebra with the notion of a vertex algebra module (see Theorem 5.3.15 in [LL]). We have seen that the Jacobi identity may be replaced by weak associativity together with weak skew-associativity (in the sense of Proposition 4.4.1). In fact, by using the algebra skew-symmetry, which we have "for free," we obtain that either one of the two is enough. It is shown in [LL] that the same is not true for weak commutativity (see Remark 4.4.6 in [LL]).

Theorem 4.6.1. *In the presence of the minor axioms of module for a vertex algebra, either single one of weak associativity or weak skew-associativity (each in the sense of Proposition 4.4.1) is equivalent to the Jacobi identity.*

Proof. By Proposition 4.6.2 we have the $\mathcal{D}$-derivative property, and following the proof of Proposition 4.5.3 we have vacuum-free skew-symmetry (in the sense of Proposition 4.4.3). Thus Proposition 4.4.4 and Proposition 4.4.5 give the result. □

References

[A] T.A., Sur la différentiation des fonctions de fonctions, *Nouvelles Annales de Mathématiques* **9** (1850), 119–125.

[An] G. Andrews, The Theory of Partitions, in *Encyclopedia of Mathematics and its Applications*, ed. by G.-C. Rota, Vol. 2, Addison-Wesley, Reading, Massachusetts, (1976), Cambridge University Press, Cambridge (1984), Cambridge Mathematical Library (1998).

[BD] A. Beilinson and V. Drinfeld, manuscript, later incorporated into: *Chiral Algebras*, Colloquium Publications, Vol. 51, American Mathematical Society, 2004.

[BPZ] A. Belavin, A. M. Polyakov and A. B. Zamolodchikov, Infinite conformal symmetries in two-dimensional quantum field theory, *Nucl. Phys.* **B241** (1984), 333–380.

[B] E.T. Bell, The history of Blissard's symbolic calculus, with a sketch of the inventor's life. *Amer. Math. Monthly*, **45** (1938) 414–421.

[Bli] J. Blissard, Theory of generic functions, *Quarterly J. Pure Appl. Math.* **4** (1861), 279–305; **5** (1862), 58–75, 185–208.

[Bl] R. Block, On the Mills-Seligman axioms for Lie algebras of classical type, *Trans. Amer. Math. Soc.* **121** (1966), 378–392.

[Bo] M. Bóna, *A Walk Through Combinatorics*, World Scientific Publishing Co., (2002).

[Bor1] R. E. Borcherds, Vertex algebras, Kac-Moody algebras, and the Monster, *Proc. Natl. Acad. Sci. USA* **83** (1986), 3068–3071.

[Bor2] R. E. Borcherds, Monstrous moonshine and monstrous Lie superalgebras, *Invent. Math.* **109** (1992), 405–444.

[BL] A. Di Bucchiano and D.E. Loeb, A selected survey of umbral calculus, *Elec. J. Combin.*, 3:Dynamical Surveys Section, 1995.

[Ch] W.Y.C. Chen, Context-free grammars, differential operators and formal power series, Conference on Formal Power Series and Algebraic Combinatorics (Bordeaux, 1991); *Theoret. Comput. Sci.* **117** (1993), 113–129.

[CN] J. H. Conway and S. P. Norton, Monstrous moonshine, *Bull. London Math. Soc.* **11** (1979), 308–339.

[DL] C. Dong and J. Lepowsky, *Generalized Vertex Algebras and Relative Vertex Operators*, Progress in Mathematics, Vol. 112, Birkhäuser, Boston, 1993.

[Do] H. Dörrie (transl. D. Antin), *100 Great Problems of Elementary Mathematics*, Dover Publications, Inc., New York, 1965 (orig. German version appeared in 1932).

[DLM] B. Doyon, J. Lepowsky and A. Milas, Twisted vertex operators and Bernoulli polynomials, *Contemp. Math.* **8** (2006) no. 2, 247–307.

[F1] C.F. Faà di Bruno, Sullo sviluppo delle funzioni, *Annali di Scienze Matematiche e Fisiche* **6** (1855), 479–480.

[F2] C.F. Faà di Bruno, Note sur une nouvelle formule de calcul différentiel, *Quarterly J. Pure Appl. Math.* **1** (1857), 359–360.

[FHL] I. Frenkel, Y.-Z. Huang and J. Lepowsky, On axiomatic approaches to vertex operators and modules, *Memoirs Amer. Math. Soc.* **104**, 1993.

[FLM1] I. Frenkel, J. Lepowsky and A. Meurman, A natural representation of the Fischer-Griess Monster with the modular function J as character, *Proc. Natl. Acad. Sci. USA* **81** (1984), 3256–3260.

[FLM2] I. Frenkel, J. Lepowsky and A. Meurman, *Vertex Operator Algebras and the Monster*, Pure and Appl. Math., Vol. 134, Academic Press, New York, 1988.

[Ga] A. Garsia, An exposé of the Mullin-Rota theory of polynomials of binomial type, *Lin. and Multilin. Alg.* **1** (1973), 47–65.

[GF] I.M. Gelfand and D.B. Fuchs, The cohomology of the Lie algebra of vector fields on a circle, *Funk. Anal. i Prilozhen.* **2** (1968), 342–343.

[G] R. L. Griess Jr., The Friendly Giant, *Invent. Math.* **69** (1982), 1–102.

[H] Y.-Z. Huang, *Two-dimensional Conformal Field Theory and Vertex Operator Algebras* Progress in Math. **148**, Birkhäuser, Boston, 1998.

[HL] Y.-Z. Huang and J. Lepowsky, On the $\mathcal{D}$-module and formal-variable approaches to vertex algebras, in: Topics in Geometry: In Memory of Joseph D'Atri, ed. by S. Gindikin, Progress in Nonlinear Differential Equations, Vol. 20, Birkhäuser, Boston, 1996, 175-202.

[HLZ] Y.-Z. Huang, J. Lepowsky and L. Zhang, Logarithmic tensor product theory for generalized modules for a conformal vertex algebra, arXiv:0710.2687v3 [math.QA]

[Jo] W.P. Johnson, The curious history of Faà di Bruno's formula, *Amer. Math. Monthly* **109** (2002), 217–234.

[L] J. Lepowsky, Application of a "Jacobi identity" for vertex operator algebras to zeta values and differential operators, *Lett. Math. Phys.* **53** (2000), no. 2, 551–579.

[LL] J. Lepowsky and H. Li, *Introduction to Vertex Operator Algebras and Their Representations*, Progress in Math., Vol. 227, Birkhäuser, Boston, 2003.

[LW1] J. Lepowsky and R. L. Wilson, Construction of the affine Lie algebra $A_1^{(1)}$ *Commun. Math.Phys.* **62** (1978), 43–53.

[LW2] J. Lepowsky and R. L. Wilson, A new family of algebras underlying the Rogers-Ramanujan identities, *Proc. Nat. Acad. Sci. USA* **78** (1981), 7254–7258.

[Li1] H. Li, Local systems of vertex operators, vertex superalgebras and modules, *J. Pure Applied Algebra* **109** (1996), 143–195.

[Li2] H. Li, Axiomatic G_1-vertex algebras, *Communications in Contemporary Mathematics* **5** (2003), 281–327.

[Lub] D. Lubell, Problem 10992, Problems and Solutions, *American Mathematical Monthly* **110**, 2003, 155.

[Lu] E. Lucas, Théorie nouvelle des nombres de Bernoulli et d'Euler, *Comptes rendus de l'Academie des Sciences (Paris)*, **83** (1876), 539–541; *Annali di Matematica pura ed applicata, Serie 2* **8** (1877), 56–79.

[M] J.M., *Traité de la formule du binôme*, Ward and Lock, 1854; transl. to English by G.W.H. in unpublished manuscript.

[Mel] Z.A. Melzak, *Companion to Concrete Mathematics: Mathematical techniques and various applications*, Pure and Applied Mathematics, John Wiley and Sons, New York-London-Sydney-Toronto, 1973.

[Me] U.H. Meyer, Sur les dériveés d'une fonction de fonction, *Archiv der Mathematik und Physik* **9** (1847), 96–100.

[Mi] A. Milas, Weak modules and logarithmic intertwining operators for vertex operator algebras, in: *Recent Developments in Infinite-Dimensional Lie Algebras and Conformal Field Theory*, ed. S. Berman, P. Fendley, Y.-Z. Huang, K. Misra, and B. Parshall, Contemp. Math., Vol. 297, American Mathematical Society, Providence, RI, 2002, 201–225.

[MR] R. Mullin, and G.-C. Rota, On the foundations of combinatorial theory, in *Graph Theory and its Applications* (B. Harris, ed.), Academic Press, New York-London (1970).

[Ri1] J. Riordan, Derivatives of composite functions, *Bull. Amer. Math. Soc.* **52** (1946), 664–667.

[Ri2] J. Riordan, *Cominatorial Identities*, Wiley, New York, 1968.

[R1] T.J. Robinson, The formal Taylor theorem revisited, to appear.

[R2] T.J. Robinson, New perspectives on exponentiated derivations, the formal Taylor theorem, and Faà di Bruno's formula: Proceedings of the Conference on Vertex Operator Algebras, July 2008, Illinois State University, ed. by M. Bergvelt, G. Yamskulna and W. Zhao, to appear; arXiv:0903.3391.

[R3] T.J. Robinson, Formal calculus and umbral calculus, to appear.

[R4] T. Robinson, On replacement axioms for the Jacobi identity for vertex algebras and their modules, submitted for publication, arXiv:0903.3216.

[Rm1] S. Roman, *The Umbral Calculus*, Pure and Appl. Math., **111**, Academic Press, New York, 1984.

[Rm2] S. Roman, The formula of Faà di Bruno, *Amer. Math. Monthly*, **87** (1980), 805–809.

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

[RKO] G.-C. Rota, D. Kahaner and A. Odlyzko, On the foundations of combinatorial theory VIII: finite operator calculus, *J. Math. Anal. Appl.* **42** (1973), 684–760.

[RR] G.-C. Rota and S. Roman, The umbral calculus, *Adv. in Math.* **27** (1978), 95–188.

[Sc] G. Scott, Formulae of successive differentiation, *Quarterly J. Pure Appl. Math.* **4** (1861), 77–92.

[Sh] L. Shapiro, Bijections and the Riordan group, *Theoret. Comput. Sci.* **307** (2003), 403–413.

[SGWW] L. Shapiro, S. Getu, W.J. Woan and L.C. Woodson, The Riordan group, *Discrete Appl. Math.* **34** (1991), 229–239.

[St] R.P. Stanley, *Enumerative Combinatorics, Vol.2*, Cambridge Studies in Advanced Mathematics **62**, Cambridge University Press, 1999.

[V] M.A. Virasoro, Subsidiary conditions and ghosts in dual-resonance models, *Phys. Rev.* **D1** (1970), 2933–2936.

[W1] H.S. Wilf, *Generatingfunctionology*, Academic Press, Inc., 1990.

[W2] H.S. Wilf, The method of characteristics, and 'problem 89' of Graham, Knuth and Patashnik, arXiv:math/0406620v2[math.CO].

[Wi] R.L. Wilson, Classification of generalized Witt algebras over algebraically closed fields, *Trans. Amer. Math. Soc.* **153** (1971), 191–210.

[Z1] D. Zeilberger, Toward a combinatorial proof of the Jacobian conjecture?, in: Proc. of the "Colloque de combinatoire enumerative," ed. by G. Labelle and P. Leroux; *Lecture Notes in Math.* **1234**, Springer-Verlag, Berlin (1987).

[Z2] D. Zeilberger, Using Rota's umbral calculus to enumerate Stanley's P-Partitions, *Adv. Applied Math.* **41** (2008), 206–217.

Vita

Thomas J. Robinson

2009	Ph. D. in Mathematics, Rutgers University
1998-2002	B. A. in Mathematics and Physics, Carleton College
1998	Graduated from The Ethical Culture Fieldston School.

2002-2009	Teaching assistant, Department of Mathematics, Rutgers University