JOURNAL OF ALGEBRA 64, 54–75 (1980)

A Cyclic Derivative in Noncommutative Algebra

GIAN-CARLO ROTA*

Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

BRUCE SAGAN

Université Louis Pasteur, Paris, France

AND

PAUL R. STEIN†

Los Alamos Scientific Laboratory, Los Alamos, New Mexico

Communicated by D. Buchsbaum

Received April 25, 1979

DEDICATED TO M. P. SCHÜTZENBERGER

... questi che guida in alto gli occhi miei.—Purg. XXI, 124.

Contents. 1. Introduction. 2. Preliminaries. 3. The cyclic derivative. 4. Taylor's formula and the chain rule. 5. Rational functions (Table of cyclic derivatives). 6. Cyclic integration. 7. The Cayley operator. 8. Further work.

1. INTRODUCTION

The problem of extending the notion of derivative to noncommuting polynomials, or to more general noncommutative algebras, has been with us for a long time. Of the various attempts during the past century only one has survived and found notable applications, the *Hausdorff derivative*. F. Hausdorff's prescription for the derivative of a monomial, say $axbx^2cx^3$, is to take the ordinary derivative while preserving the order of the factors, as in the example

$$H\langle axbx^2cx^3\rangle = abx^2cx^2 + 2axbxcx^3 + 3axbx^2cx^2.$$

* Partially supported by NSF Contract MCS 78-02743.

† Work supported by the U.S. Department of Energy, April 1979.

0021-8693/80/050054–22$02.00/0

This rule extends by linearity and continuity to noncommuting formal power series (see, e.g., the discussion in [4]).

A generalization of Taylor's formula for $f(x + a)$, where f is a power series in noncommuting letters, can be obtained using the Hausdorff derivative (Section 4 below). However, computing the Hausdorff derivative of $f(ax)$ does not yield a simple result. This happens because the chain rule for the Hausdorff derivative is more complicated than its commutative analog.

We introduce here an altogether different notion of derivative, which we propose to call the *cyclic derivative*. This derivative applies to a wide class of noncommutative algebras, beginning with the algebra of noncommutative formal power series in a variable x and constants $a, b,..., c$. The Hausdorff derivative can be characterized as the unique linear operator on this algebra satisfying

(1) $H\langle a\rangle = 0$,

(2) $H\langle x\rangle = 1$,

(3) $H\langle fg\rangle = H\langle f\rangle \cdot g + f \cdot H\langle g\rangle$,

where f and g are arbitrary formal power series. We show in Section 3 that there is an analogous characterization of the cyclic derivative when (3) is replaced by the cyclic product rule. A more important feature of the cyclic derivative is the existence of a simple chain rule for the composition of formal power series which reduces to the ordinary chain rule when the elements $a, b,..., x$ commute. As an example, the cyclic derivative of $f(ax)$, where $f(x)$ is a formal power series in x alone, is just $f'(ax) \cdot a$.

The cyclic derivative of a monomial, say $axbx^2cx^3$, is computed using the following three steps. First, one takes all the cyclic permutations of the monomial; so, in the above example $xaxbx^2cx^2$, x^2axbx^2cx, x^3axbx^2c, cx^3axbx^2, xcx^3axbx, x^2cx^2axb, bx^2cx^3ax, xbx^2cx^3a, $axbx^2cx^3$; second, one crosses out all monomials not starting with x, and removes the initial x from the rest; finally one adds the remaining terms, thereby obtaining

$$D\langle axbx^2cx^3\rangle = axbx^2cx^2 + xaxbx^2cx + x^2axbx^2c + cx^3axbx + xcx^3axb + bx^2cx^3a.$$

The extension to formal power series is immediate.

The cyclic derivative is not a derivation (although it is cyclically invariant), and thus the cyclic analog of Taylor's formula is weaker than the Hausdorff version. This is compensated for by the fact that the cyclic derivative enjoys a more elegant chain rule. Thus, the cyclic and Hausdorff derivatives complement each other.

We were led to the definition of the cyclic derivative while reading some work of Turnbull, who was motivated by the properties of the Cayley operator

$\Omega = (\partial/\partial x_{ji})$ of classical invariant theory (Section 7 below). Turnbull computed several properties of the operator $P \to \Omega\,\mathrm{tr}(P)$, where P is a polynomial in constant n by n matrices $A, B,\ldots, C$ and a variable matrix X whose entries x_{ij} are independent indeterminates; he succeeded in obtaining a generalization of Taylor's formula along the lines of Proposition 4.1 below. However, he missed, even in the special case of finite n by n matrices, our main result: the chain rule (see Theorem 4.2). Nonetheless, we take this opportunity to express our indebtedness to Turnbull's pioneering work. We also wish to thank Ira Gessel for his careful reading of the manuscript, and several suggestions.

2. Preliminaries

We are given an alphabet A having a distinguished letter, denoted by x and called the *variable*, and an indefinite supply of other letters $a, b,\ldots$ called *constants*. An element w of the free monoid M generated by A is called a *word*. The *degree* of $w \in M$ is the number of occurrences of the letter x in w while the *length* of w is the total number of letters it contains. The identity element of the monoid is written 1 and is called the *empty word*; it has length zero.

Let K be a field of characteristic zero. The *algebra of noncommutative polynomials* $K\{a, b,\ldots, x\}$ is the set of all linear combinations

$$P = \sum_{i=1}^{n} k_i w_i\,, \qquad k_i \in K,\ w_i \in M$$

with the multiplication inherited from M. We will refer to the elements of K as *scalars* to distinguish them from the constants in A. Note that this algebra is graded by length. We let $p^{(l)}$ denote the homogeneous part of p of length l.

We now define a norm on $K\{a, b,\ldots, x\}$ by

$$\begin{aligned} \| p \| &= 0, && \text{if } p = 0 \\ &= 1/(l+1), && \text{otherwise,} \end{aligned}$$

where l is the smallest integer such that $p^{(l)} \neq 0$. The verification that $\|\ \|$ is indeed a norm is easy and is left as an exercise. The completion of $K\{a, b,\ldots, x\}$ in the topology induced by $\|\ \|$ is the well-known *algebra of noncommutative formal power series* $K\{\{a, b,\ldots, x\}\}$. Consider any $f \in K\{\{a, b,\ldots, x\}\}$, say $f = \lim_{i\to\infty} p_i$. Then for any integer $l \geqslant 0$ there exists an integer $I \geqslant 0$ such that $p^{(i)} = p_I^{(l)}$ for all $i \geqslant I$. Letting $p^{(l)} = p_I^{(l)}$ we obtain in the usual formal expansion,

$$f = \sum_{l=0}^{\infty} p^{(l)}.$$

It can be shown that the polynomials $p^{(l)}$ are independent of the sequence used to define f.

For a given word $w = c_1c_2 \cdots c_n \neq 1$, where the c_i are letters, we define an operator Cw mapping formal power series into formal power series called the *cyclic operator*. The action of Cw on a formal power series f is given by

$$\langle Cw \mid f\rangle = c_1c_2 \cdots c_nf + c_2c_3 \cdots c_nfc_1 + \cdots + c_nfc_1c_2 \cdots c_{n-1}.$$

In addition, we set $\langle C1 \mid f\rangle = 0$ for all formal power series f. An important special case is the action of the operator Cw on the formal power series $f = 1$, which gives

$$\langle Cw \mid 1\rangle = c_1c_2 \cdots c_n + c_2c_3 \cdots c_nc_1 + \cdots + c_nc_1c_2 \cdots c_{n-1}. \tag{2.1}$$

We will abbreviate $\langle Cw \mid 1\rangle$ to $C\langle w\rangle$.

By linearity and continuity we can extend the cyclic operator to polynomials and then to formal power series. If $p = \sum_{i=1}^{r} k_iw_i$ is a polynomial, we have

$$\langle Cp \mid f\rangle = \sum_{i=1}^{r} k_i\langle Cw_i \mid f\rangle;$$

and if $g = \sum_{l=1}^{\infty} p^{(l)}$, then

$$\langle Cg \mid f\rangle = \sum_{l=0}^{\infty} \langle Cp^{(l)} \mid f\rangle.$$

The following lemma is useful:

LEMMA 2.2 (Cyclic Product Rule). *Let f_1, f_2, and f be formal power series. Then*

$$\langle Cf_1f_2 \mid f\rangle = \langle Cf_1 \mid f_2f\rangle + \langle Cf_2 \mid ff_1\rangle.$$

Proof. First let f_1 and f_2 be words, say $f_1 = w_1 = c_1c_2 \cdots c_s$ and $f_2 = w_2 = d_1d_2 \cdots d_t$ so

$$\begin{aligned}\langle Cw_1w_2 \mid f\rangle &= c_1c_2 \cdots c_sd_1d_2 \cdots d_tf + \cdots + c_sd_1d_2 \cdots d_tfc_1c_2 \cdots c_{s-1} \\ &\quad + d_1d_2 \cdots d_tfc_1c_2 \cdots c_s + \cdots + d_tfc_1c_2 \cdots c_sd_1d_2 \cdots d_{t-1} \\ &= c_1c_2 \cdots c_sw_2f + \cdots + c_sw_2fc_1c_2 \cdots c_{s-1} \\ &\quad + d_1d_2 \cdots d_tfw_1 + \cdots + d_tfw_1d_1d_2 \cdots d_{t-1} \\ &= \langle Cw_1 \mid w_2f\rangle + \langle Cw_2 \mid fw_1\rangle.\end{aligned}$$

By linearity and continuity the result for words extends to the case where f_1 and f_2 are formal power series. Q.E.D.

PROPOSITION 2.3. *Let $f_1, f_2, \ldots, f_n$ and f be formal power series. Then*

$$\langle Cf_1f_2 \cdots {}_{an} \mid f_n\rangle = \langle Cf_1 \mid f_2 \cdots f_n f\rangle + \langle Cf_2 \mid f_3 \cdots f_n f f_1\rangle + \cdots + \langle Cf_1 \mid f f_1 f_2 \cdots f_{n-1}\rangle. \tag{2.4}$$

Proof. Letting $g = f_1 f_2 \cdots f_{n-1}$ we have

$$\begin{aligned}\langle Cf_1f_2 \cdots f_n \mid f\rangle &= \langle Cgf_n \mid f\rangle \\ &= \langle Cg \mid f_n f\rangle + \langle Cf_n \mid fg\rangle \\ &= \langle Cf_1f_2 \cdots f_{n-1} \mid f_n f\rangle + \langle Cf_n \mid ff_1f_2 \cdots f_{n-1}\rangle;\end{aligned}$$

Eq. (2.4) now follows by induction. Q.E.D.

As an immediate corollary we infer

COROLLARY 2.5. *If f_1 and f_2 are formal power series then*

$$\langle C(f_1 + f_2)^n \mid f_1 + f_r\rangle = n \cdot \langle C(f_1 + f_2) \mid (f_1 + f_2)^n\rangle.$$

3. THE CYCLIC DERIVATIVE

We now define a basic notion which generalizes the derivative of a commutative formal power series to the noncommutative algebra $K\{\{a, b, \ldots, x\}\}$. Given a word $w = c_1c_2 \cdots c_n$, let the *truncation operator* T act on w by

$$\begin{aligned} T\langle w\rangle = T\langle c_1c_2 \cdots c_n\rangle &= 0 && \text{if } c_1 \neq x \text{ or } w = 1 \\ &= c_2c_3 \cdots c_n && \text{if } c_1 = x. \end{aligned}$$

The operator T extends by linearity to polynomials and then by continuity to formal power series. Now for each formal power series $f \in K\{\{a, b, \ldots, x\}\}$, the *cyclic derivative operator* $Df\colon K\{\{a, b, \ldots, x\}\} \to K\{\{a, b, \ldots, x\}\}$ is the operator mapping formal power series into formal power series by the rule

$$\langle Df \mid g\rangle = T\langle Cf \mid g\rangle.$$

In particular, if g is 1 then the formal power series $\langle Df \mid 1\rangle = T\langle Cf \mid 1\rangle$ is called the *cyclic derivative* of f and is denoted by $D\langle f\rangle$.

EXAMPLE 3.1. For the constant $a \in A$, $a \neq x$, we have $D\langle a\rangle = T\langle Ca \mid 1\rangle = T\langle a\rangle = 0$.

EXAMPLE 3.2. It follows from (2.1) that $Cx^i = ix^i$. Hence if $f = \sum_{i=0}^{\infty} k_i x^i$, $k_i \in K$, then

$$D\langle f\rangle = T\langle C\langle f\rangle\rangle = T\left\langle \sum_{i=0}^{\infty} ik_i x^i \right\rangle = \sum_{i=1}^{\infty} ik_i x^{i-1}.$$

However, if $f = \sum_{i=0}^{\infty} k_i a^i$ then $D\langle f\rangle = T\langle \sum_{i=0}^{\infty} ik_i a^i\rangle = 0$.

EXAMPLE 3.3.

$$\begin{aligned} D\langle ax^n\rangle &= T\langle C\langle x^n\rangle\rangle \\ &= T\langle ax^n + x^n a + x^{n-1}ax + \cdots + xax^{n-1}\rangle \\ &= x^{n-1}a + x^{n-2}ax + \cdots + ax^{n-1} \\ &= C\langle ax^{n-1}\rangle. \end{aligned}$$

We can also write

$$D\langle ax^n\rangle = \langle Da \mid x^n\rangle + \langle Dx \mid x^{n-1}a\rangle + \langle Dx \mid x^{n-2}ax\rangle + \cdots + \langle Dx \mid ax^{n-1}\rangle.$$

This is a special case of

THEOREM 3.4 (Cyclic Derivative Product Rule). *Let $f_1, f_2, \ldots, f_n$ be formal power series. Then*

$$\begin{aligned} D\langle f_1 f_2 \cdots f_n\rangle = \langle Df_1 \mid f_2 f_3 \cdots f_n\rangle + \langle Df_2 \mid f_3 \cdots f_n f_1\rangle \\ + \cdots + \langle Df_n \mid f_1 f_2 \cdots f_{n-1}\rangle. \end{aligned}$$

Proof. By the definition of D and Proposition 2.3

$$\begin{aligned} D\langle f_1 f_2 \cdots f_n\rangle &= T\langle Cf_1 f_2 \cdots f_n \mid 1\rangle \\ &= T\langle Cf_1 \mid f_2 f_3 \cdots f_n 1\rangle + T\langle Cf_2 \mid f_3 \cdots f_n 1 f_1\rangle \\ &\quad + \cdots + T\langle Cf_n \mid 1 f_1 f_2 \cdots f_{n-1}\rangle \\ &= \langle Df_1 \mid f_2 f_3 \cdots f_n\rangle + \langle Df_2 \mid f_3 \cdots f_n f_1\rangle \\ &\quad + \cdots + \langle Df_n \mid f_1 f_2 \cdots f_{n-1}\rangle. \end{aligned}$$ Q.E.D.

The product rule (Theorem 3.4) can be used to characterize the cyclic derivative in the same way that the Hausdorff derivative is characterized as a derivation. Specifically, a map h from $K\{\{a, b,\ldots, x\}\}$ to the algebra of K-linear continuous endomorphisms of $K\{\{a, b,\ldots, x\}\}$

$$h\colon K\{\{a, b,\ldots, x\}\} \to \operatorname{End} K\{\{a, b,\ldots, x\}\}$$

is called a *cyclic derivation* if for all formal power series f_1, f_2, and f we have

$$\langle hf_1f_2 \mid f\rangle = \langle hf_1 \mid f_2f\rangle + \langle hf_2 \mid ff_1\rangle$$

where $\langle hf_1f_2 \mid f\rangle$ denotes the image of f_1f_2 under h evaluated at f.

THEOREM 3.5. *The cyclic derivative operator D is the unique cyclic derivation satisfying*

(i) $Dx = id$ (*the identity endomorphism*),

(ii) *for all constants, a, $Da = 0$ (the zero endomorphism).*

Proof. Conditions (i) and (ii) follow directly from the definition of D; the fact that D is a cyclic derivation can be deduced from Lemma 2.2 by applying T to both sides of the product rule for C.

Now let h be any cyclic derivation satisfying (i) and (ii). Since both D and h are K-linear and continuous it suffices to prove that $hw = Dw$ for any word w. Induct on the length of w. $\langle h1 \mid f\rangle = \langle h1 \mid 1f\rangle + \langle h1 \mid f1\rangle = 2\langle h1 \mid f\rangle$, so $h1 = D1 = 0$. If $w = x$ or $w = a$ we are done by assumption. Now if

$$w = c_1c_2 \cdots c_n = c_1v$$

where the c_i are letters and $v = c_2c_3 \cdots c_n$, we have

$$\begin{aligned} \langle hw \mid f\rangle &= \langle hc_1 \mid vf\rangle + \langle hv \mid fc_1\rangle \\ &= \langle Dc_1 \mid vf\rangle + \langle Dv \mid fc_1\rangle \qquad \text{(by induction)} \\ &= \langle Dw \mid f\rangle. \end{aligned}$$

Q.E.D.

We introduce the *permutation operator*, P. For distinct letters $c_1, c_2, \ldots, c_s$—where c_k may be either a constant or the letter x—we denote by

$$P\langle c_1{}^i c_2{}^j \cdots c_s{}^n\rangle$$

the sum of all distinct words containing i occurrences of c_1, j occurrences of c_2, etc., each word appearing exactly once. For example,

$$\begin{aligned} P\langle a^2x^3\rangle = {} & a^2x^3 + ax^3a + x^3a^2 + x^2a^2x + xa^2x^2 + axax^2 \\ & + xax^2a + ax^2ax + x^2axa + xaxax. \end{aligned}$$

Some properties of the permutation operator are stated in

PROPOSITION 3.6. (i) *For integers $i, j, \ldots, n \geqslant 1$ we have*

$$\begin{aligned} P\langle c_1{}^i c_2{}^j \cdots c_s{}^n\rangle = {} & c_1 \cdot P\langle c_1^{i-1} c_2{}^j \cdots c_s{}^n\rangle + c_2 \cdot P\langle c_1{}^i c_2^{j-1} \cdots c_s{}^n\rangle \\ & + \cdots + c_s \cdot P\langle c_1{}^i c_2{}^j \cdots c_s^{n-1}\rangle. \end{aligned}$$

(ii) (Noncommutative Multinomial Identity)

$$(c_1 + c_2 + \cdots + c_s)^n = \sum_{i+j+\cdots+k=n} P\langle c_1^i c_2^j \cdots c_s^k \rangle.$$

(iii) $C\langle P\langle c_1^i c_2^j \cdots c_s^n \rangle\rangle = (i + j + \cdots + n) \cdot P\langle c_1^i c_2^j \cdots c_s^n \rangle$.

(iv) *If $c_1, c_2, \ldots, c_{s-1}$ are constants, $c_s = x$, and $n \geq 1$ is an integer, then*

$$T\langle P\langle c_1^i c_2^j \cdots x^n \rangle\rangle = P\langle c_1^i c_2^j \cdots x^{n-1} \rangle.$$

Proof. (i) Obvious from the definition of P.

(ii) Inducting on n and using (i) we have

$$(c_1 + c_2 + \cdots + c_s)^n = (c_1 + c_2 + \cdots + c_s) \cdot \sum_{i+j+\cdots+k=n-1} P\langle c_1^i c_2^j \cdots c_s^k \rangle$$
$$= \sum_{i+j+\cdots+k=n} P\langle c_1^i c_2^j \cdots c_s^k \rangle.$$

(iii) If w is any word appearing in $P\langle c_1^i c_2^j \cdots c_s^n \rangle$ then all the cyclic permutations of w also appear in $P\langle c_1^i c_2^j \cdots c_s^n \rangle$. Furthermore, all the cyclic permutations of w appear in $C\langle w \rangle$ with the same multiplicity m; hence $C\langle (1/m) w \rangle$ contains all cyclic permutations of w with multiplicity one. Thus we can find a polynomial

$$p = \frac{1}{m_1} w_1 + \frac{1}{m_2} w_2 + \cdots + \frac{1}{m_k} w_k$$

such that $P\langle c_1^i c_2^j \cdots c_s^n \rangle = C\langle p \rangle$. Now

$$\begin{aligned} C\langle P\langle c_1^i c_2^j \cdots c_s^n \rangle\rangle &= C\langle C\langle p \rangle\rangle \\ &= (i + j + \cdots + n) \cdot C\langle p \rangle \qquad \text{(since } p \text{ is homogeneous of length } i + j + \cdots + n\text{)} \\ &= (i + j + \cdots + n) \cdot P\langle c_1^i c_2^j \cdots c_s^n \rangle. \end{aligned}$$

(iv)
$$\begin{aligned} T\langle P\langle c_1^i c_2^j \cdots x^n \rangle\rangle &= T\langle c_1 \cdot P\langle c_1^{i-1} c_2^j \cdots x^n \rangle + c_2 \cdot P\langle c_1^i c_2^{j-1} \cdots x^n \rangle \\ &\quad + \cdots + x \cdot P\langle c_1^i c_2^j \cdots x^{n-1} \rangle\rangle \\ &= P\langle c_1^i c_2^j \cdots x^{n-1} \rangle. \end{aligned}$$

Q.E.D.

We can now investigate the relationship between the permutation operator P and the cyclic derivative D. For the remainder of this section the letters, a, b, c will denote constants.

COROLLARY 3.7 (Cyclic Exponent Rule).

$$\begin{aligned} D\langle P\langle a^i b^j \cdots x^n \rangle\rangle &= T\langle C\langle P\langle a^i b^j \cdots x^n \rangle\rangle\rangle \\ &= T\langle (i + j + \cdots + n) \cdot P\langle a^i b^j \cdots x^n \rangle\rangle \\ &= (i + j + \cdots + n) \cdot P\langle a^i b^j \cdots x^{n-1} \rangle. \end{aligned}$$

There is also a cyclic analog of the classical polarization operator. For any constant, a, let $\mathbf{a}$ denote the operator mapping the power series f to the power series af.

LEMMA 3.8. *$C\mathbf{a}P\langle a^i b^j c^k \cdots x^n\rangle = (i+1) \cdot P\langle a^{i+1} b^j c^k \cdots x^n\rangle$ where juxtaposition of operators indicates composition.*

Proof. Let $P\langle a^i b^j c^k \cdots x^n\rangle = w_1 + w_2 + \cdots + w_m$. Then

$$\begin{aligned} C\mathbf{a}P\langle a^i b^j c^k \cdots x^n\rangle &= C\langle aw_1 + aw_2 + \cdots + aw_m\rangle \\ &= C\langle aw_1\rangle + C\langle aw_2\rangle + \cdots + C\langle aw_m\rangle. \end{aligned} \tag{3.9}$$

Now each w_r is a summand in $P\langle a^i b^j c^k \cdots x^n\rangle$; so aw_r, and thus each monomial in $C\langle aw_r\rangle$, appears in $P\langle a^{i+1} b^j c^k \cdots x^n\rangle$. In fact, each monomial of $P\langle a^{i+1} b^j c^k \cdots x^n\rangle$ is represented $i+1$ times in (3.9). For if w is a summand in $P\langle a^{i+1} b^j c^k \cdots x^n\rangle$, say

$$w = v_1 a v_2 a \cdots v_{i+1} a v_{i+2}$$

where the v_k are words (possibly empty) not containing a, then w appears in each of $C\mathbf{a}\langle v_2 a v_3 \cdots a v_{i+2} v_1\rangle$, $C\mathbf{a}\langle v_3 a v_4 \cdots a v_{i+2} v_1 a v_2\rangle, \ldots, C\mathbf{a}\langle v_{i+2} v_1 a v_2 a \cdots a v_{i+1}\rangle$ and nowhere else. Q.E.D.

EXAMPLE 3.10 (Cyclic Polarization Operator).

$$\begin{aligned} D\mathbf{a}\langle P\langle a^i b^j c^k \cdots x^n\rangle\rangle &= T\langle C\mathbf{a}P\langle a^i b^j c^k \cdots x^n\rangle\rangle \\ &= T\langle (i+1) \cdot P\langle a^{i+1} b^j c^k \cdots x^n\rangle\rangle \\ &= (i+1) \cdot P\langle a^{i+1} b^j c^k \cdots x^{n-1}\rangle. \end{aligned}$$

It is clear from Example 3.10 that the operators $D\mathbf{a}$ and $D\mathbf{b}$ commute when applied to $P\langle a^i b^j c^k \cdots x^n\rangle$.

The following example will be of use in the sequel.

EXAMPLE 3.11. $(D\mathbf{a})^m\langle x^n\rangle = m! \cdot P\langle a^m x^{n-m}\rangle$, where $m \leqslant n$, and where $(D\mathbf{a})^m$ indicates iteration of the polarization operator m times.

4. TAYLOR'S FORMULA AND THE CHAIN RULE

We recall some elementary facts about the Hausdorff derivative. For any constant a and any word $w = c_1 c_2 \cdots c_n$ the Hausdorff polarization operator H_a is defined as follows. If m of the c_i equal x, then $H_a\langle w\rangle$ is the sum of the m words obtained by replacing each occurrence of x in turn by a. For example, $H_a\langle x^2bxb\rangle = axbxb + xabxb + x^2bab$. The Hausdorff derivative H is obtained by setting the substituted "a" equal to 1 in $H_a\langle w\rangle$, for example, $H\langle x^2bxb\rangle = xbxb + xbxb + x^2b^2$.

In terms of the Hausdorff polarization operator, for every formal power series $f(x)$ in noncommuting letters one derives the following analog of Taylor's formula:

$$f(x + a) = f(x) + H_a\langle f(x)\rangle + \frac{1}{2!}(H_a)^2\langle f(x)\rangle + \cdots = e^{H_a}\langle f(x)\rangle.$$

The analogous formula for the cyclic polarization operator is the following:

PROPOSITION 4.1. *Let $f(x) = \sum_{n=0}^{\infty} k_n x^n$ be a formal power in x having only scalar coefficients $k_n \in K$. Then*

$$f(x + a) = f(x) + \mathbf{D}\mathbf{a}\langle f(x)\rangle + \frac{1}{2!}(D\mathbf{a})^2\langle f(x)\rangle + \cdots = e^{D\mathbf{a}}\langle f(x)\rangle.$$

Proof. This follows easily using the results of Section 3, for

$$\begin{aligned} f(x+a) &= \sum_{n=0}^{\infty} k_n(x+a)^n \\ &= \sum_{n=0}^{\infty} k_n \sum_{i=0}^{n} P\langle a^i x^{n-i}\rangle \qquad \text{(Proposition 3.6)} \\ &= \sum_{i=0}^{\infty}\sum_{n=i}^{\infty} k_n P\langle a^i x^{n-i}\rangle \\ &= \sum_{i=0}^{\infty}\sum_{n=i}^{\infty} \frac{k_n}{i!}(D\mathbf{a})^i\langle x^n\rangle \qquad \text{(Example 3.11)} \\ &= \sum_{i=0}^{\infty} \frac{1}{i!}(D\mathbf{a})^i\left\langle \sum_{n=0}^{\infty} k_n x^n\right\rangle \\ &= \sum_{i=0}^{\infty} \frac{1}{i!}(D\mathbf{a})^i\langle f(x)\rangle. \end{aligned}$$

Q.E.D.

Taylor's formula for the cyclic polarization operator applies to a smaller class of formal power series than does the analogous formula for the Hausdorff polarization operator. This weakness of the cyclic polarization operator is compensated for by the existence of an analog of the chain rule for the cyclic derivative of composite functions.

Let $f = f(x)$ and $g = g(x)$ be noncommutative formal power series in x and assume that g has no term of length 0. Denote by $f(g)$ the power series obtained by replacing every occurrence of x in f by g. Similarly let $D_g\langle f\rangle$ denote the same substitution applied to the power series $D\langle f\rangle$.

THEOREM 4.2 (Chain Rule). *With f and g as above*

$$D\langle f(g)\rangle = \langle Dg \mid D_g\langle f\rangle\rangle.$$

Proof. First consider the case where f is a word $f = v_1xv_2 \cdots xv_m$ where the v_i are words not containing x (possibly empty). Then for any formal power series g:

$$\begin{aligned} D\langle f(g)\rangle &= D\langle v_1gv_2 \cdots gv_m\rangle \\ &= \langle Dv_1 \mid gv_2 \cdots gv_m\rangle + \langle Dg \mid v_2 \cdots gv_mv_1\rangle \\ &\quad + \cdots + \langle Dg \mid v_mv_1gv_2 \cdots v_{m-1}\rangle + \langle Dv_m \mid v_1gv_2 \cdots v_{m-1}g\rangle. \end{aligned}$$

But Dv_i is the zero operator, since v_i contains no x. Hence in this case we have $D\langle f(g)\rangle = \langle Dg \mid D_g\langle f\rangle\rangle$ as desired. The result extends from words to polynomials (by linearity) and then to power series (by continuity). Q.E.D.

Some examples are in order.

EXAMPLE 4.3. Let $f(x) = x^n$, $g(x) = ax$, then

$$\begin{aligned} D\langle (ax)^n\rangle &= \langle Dax \mid D_{ax}\langle x^n\rangle\rangle \\ &= \langle Dax \mid n(ax)^{n-1}\rangle \\ &= T\langle ax \cdot n(ax)^{n-1} + x \cdot n(ax)^{n-1} \cdot a\rangle \\ &= n(ax)^{n-1} \cdot a, \end{aligned}$$

as we might expect from the classical rule of the calculus. Similarly, $D\langle (xa)^n\rangle = a \cdot n(ax)^{n-1}$.

The reader can also verify that the cyclic derivatives of exponentials and logarithms behave much like their classical counterparts, for example, $D\langle e^{ax}\rangle = e^{ax} \cdot a$ and $D\langle \log(1 + ax)\rangle = (1 + ax)^{-1} \cdot a$. Here the functions e^x, $\log(1 + x)$, and $(1 + x)^{-1}$ are defined in terms of their usual power series expansions. It should be noted that the corresponding Hausdorff derivatives of these formal power series do not have any simple form.

The next example shows some of the peculiarities of taking cyclic derivatives:

EXAMPLE 4.4. Let $f(x) = e^x$, $g(x) = axbx$. Then

$$\begin{aligned} D\langle e^{axbx}\rangle &= \langle Daxbx \mid D_{axbx}\langle e^x\rangle\rangle \\ &= \langle Daxbx \mid e^{axbx}\rangle \\ &= bxe^{axbx} \cdot a + e^{axbx} \cdot axb. \end{aligned}$$

In a commutative algebra, this result reduces to $(e^{abx^2})' = 2abxe^{abx^2}$, as expected.

5. Rational Functions

As an application of the results of the previous two sections we show that the derivative of a rational formal power series (rational function) is a rational formal power series. The *rational formal power series* are defined inductively by

(i) monomials are rational formal power series,

(ii) the sums and products of rational formal power series are rational,

(iii) if r is a rational formal power series such that $r^{(0)} = 0$ (r contains no term of length zero) then $(1 - r)^{-1}$ is rational.

THEOREM 5.1. *If r is a rational formal power series then so is its cyclic derivative $D\langle r\rangle$.*

Proof. The proof has three parts.

(i) If w is a monomial and r is any rational formal power series then $\langle Dw \mid r\rangle$ is a rational function. For if $w = c_1c_2 \cdots c_n$ where the c_i are letters, then

$$\langle Dw \mid r\rangle = T\langle c_1c_2 \cdots c_nr + c_2 \cdots c_nrc_1 + \cdots + c_nrc_1c_2 \cdots c_{n-1}\rangle$$

which is the sum of rational functions.

(ii) If r_1 and r_2 are rational functions such that $\langle Dr_1 \mid r\rangle$ and $\langle Dr_2 \mid r\rangle$ are rational for any rational function r then $\langle D(r_1 + r_2) \mid r\rangle$ and $\langle Dr_1r_2 \mid r\rangle$ are also rational functions. This follows because

$$\langle D(r_1 + r_2) \mid r\rangle = \langle Dr_1 \mid r\rangle + \langle Dr_2 \mid r\rangle$$

and

$$\langle Dr_1r_2 \mid r\rangle = \langle Dr_1 \mid r_2r\rangle + \langle Dr_2 \mid rr_1\rangle$$

(iii) If q is a rational function satisfying $q^{(0)} = 0$ and $\langle Dq \mid r\rangle$ is rational for any rational function r, then $\langle D(1 - q)^{-1} \mid r\rangle$ is a rational function. For this case we need the following identity which will be proved later:

$$\left\langle C \frac{1}{1-q} \,\middle|\, r\right\rangle = \left\langle Cq \,\middle|\, \frac{1}{1-q} \cdot r \cdot \frac{1}{1-q}\right\rangle.$$

assuming this formula we have

$$\begin{aligned}\left\langle D \frac{1}{1-q} \,\middle|\, r\right\rangle &= T\left\langle C \frac{1}{1-q} \,\middle|\, r\right\rangle \\ &= T\left\langle Cq \,\middle|\, \frac{1}{1-q} \cdot r \cdot \frac{1}{1-q}\right\rangle \\ &= \left\langle Dq \,\middle|\, \frac{1}{1-q} \cdot r \cdot \frac{1}{1-q}\right\rangle,\end{aligned}$$

which is rational by our assumptions about Dq.

To complete the proof we require:

LEMMA 5.2.

$$\left\langle C \frac{1}{1-q} \,\middle|\, r \right\rangle = \left\langle Cq \,\middle|\, \frac{1}{1-q} \cdot r \cdot \frac{1}{1-q} \right\rangle$$

Proof.

$$\begin{aligned}
\left\langle C \frac{1}{1-q} \,\middle|\, r \right\rangle &= \langle C(1 + q + q^2 + q^3 + \cdots) \mid r \rangle \\
&= \langle C1 \mid r \rangle + \langle Cq \mid r \rangle + \langle Cq^2 \mid r \rangle + \langle Cq^3 \mid r \rangle + \cdots \\
&= 0 + \langle Cq \mid r \rangle + \langle Cq \mid qr \rangle + \langle Cq \mid rq \rangle \\
&\quad + \langle Cq \mid q^2 r \rangle + \langle Cq \mid qrq \rangle + \langle Cq \mid rq^2 \rangle + \cdots
\end{aligned}$$

by Proposition 2.2. Hence we have

$$\begin{aligned}
\left\langle C \frac{1}{1-q} \,\middle|\, r \right\rangle &= \langle Cq \mid r + qr + rq + q^2 r + qrq + rq^2 + \cdots \rangle \\
&= \left\langle Cq \,\middle|\, \frac{1}{1-q} \cdot r \cdot \frac{1}{1-q} \right\rangle. \qquad \text{Q.E.D.}
\end{aligned}$$

As an example, we compute the derivative of the rational power series product

$$\frac{1}{1-ax} \cdot \frac{1}{1-bx}.$$

EXAMPLE 5.3.

$$\begin{aligned}
&D\left\langle \frac{1}{1-ax} \cdot \frac{1}{1-bx} \right\rangle \\
&\quad = \left\langle D \frac{1}{1-ax} \,\middle|\, \frac{1}{1-bx} \right\rangle + \left\langle D \frac{1}{1-bx} \,\middle|\, \frac{1}{1-ax} \right\rangle \\
&\quad = T\left\langle C \frac{1}{1-ax} \,\middle|\, \frac{1}{1-bx} \right\rangle + T\left\langle C \frac{1}{1-bx} \,\middle|\, \frac{1}{1-ax} \right\rangle \\
&\quad = T\left\langle Cax \,\middle|\, \frac{1}{1-ax} \cdot \frac{1}{1-bx} \cdot \frac{1}{1-ax} \right\rangle \\
&\qquad + T\left\langle Cbx \,\middle|\, \frac{1}{1-bx} \cdot \frac{1}{1-ax} \cdot \frac{1}{1-bx} \right\rangle \\
&\quad = \left(\frac{1}{1-ax} \cdot \frac{1}{1-bx} \cdot \frac{1}{1-ax} \right) \cdot a \\
&\qquad + \left(\frac{1}{1-bx} \cdot \frac{1}{1-ax} \cdot \frac{1}{1-bx} \right) \cdot b.
\end{aligned}$$

To conclude this section we present a table of cyclic derivatives.

Table of Cyclic Derivatives

Throughout this table n denotes a positive integer. All transcendental and rational functions of x are defined in terms of their usual Taylor expansions. For the derivatives of Sections I, II, and III, "a" may be replaced by any word containing only constants.

I. Chain rule with $x + a$: $D\langle f(x+a)\rangle = D_{x+a}\langle f(x)\rangle$

1. $D\langle (x+a)^n\rangle = n(x+a)^{n-1}$
2. $D\langle (1-x-a)^{-n}\rangle = n(1-x-a)^{-n-1}$
3. $D\langle e^{x+a}\rangle = e^{x+a}$
4. $D\langle \log(1+x+a)\rangle = (1+x+a)^{-1}$

II. Chain rule with ax: $D\langle f(ax)\rangle = D_{ax}\langle f(x)\rangle \cdot a$

1. $D\langle (ax)^n\rangle = n(ax)^{n-1} \cdot a$
2. $D\langle (1-ax)^{-n}\rangle = n(1-ax)^{-n-1} \cdot a$
3. $D\langle e^{ax}\rangle = e^{ax} \cdot a$
4. $D\langle \log(1+ax)\rangle = (1+ax)^{-1} \cdot a$

III. Chain rule with xa: $D\langle f(xa)\rangle = a \cdot D_{xa}\langle f(x)\rangle$

1. $D\langle (xa)^n\rangle = a \cdot n(xa)^{n-1}$
2. $D\langle (1-xa)^{-n}\rangle = a \cdot n(1-xa)^{-n-1}$
3. $D\langle e^{xa}\rangle = a \cdot e^{xa}$
4. $D\langle \log(1+xa)\rangle = a \cdot (1+xa)^{-1}$

IV. Derivatives of $a \cdot f(x)$

1. $D\langle a \cdot x^n\rangle = \sum_{i=0}^{n-1} x^i \cdot a \cdot x^{n-i-1}$
2. $D\langle a \cdot (1-x)^{-1}\rangle = (1-x)^{-1} \cdot a \cdot (1-x)^{-1}$
3. $D\langle a \cdot (1-x)^{-n}\rangle = \sum_{i=1}^{n} (1-x)^{-i} \cdot a \cdot (1-x)^{-n+i-1}$
4. $D\langle a \cdot e^x\rangle = \sum_{i=0}^{\infty} \frac{1}{(i+1)!} \sum_{j=0}^{i} x^j \cdot a \cdot x^{i-j}$
5. $D\langle a \cdot \log(1+x)\rangle = \sum_{i=0}^{\infty} \frac{(-1)^i}{i+1} \sum_{j=0}^{i} x^j \cdot a \cdot x^{i-j}$

V. Miscellaneous

1. $D\langle e^{axbx}\rangle = bxe^{axbx} \cdot a + e^{axbx} \cdot axb$

2. $$D\left\langle\frac{1}{1-ax}\cdot\frac{1}{1-bx}\right\rangle=\frac{1}{1-ax}\cdot\frac{1}{1-bx}\cdot\frac{1}{1-ax}\cdot a+\frac{1}{1-bx}\cdot\frac{1}{1-ax}\cdot\frac{1}{1-bx}\cdot b$$

3. $$D\langle e^{(1-ax)^{-1}}\rangle=\frac{1}{1-ax}\cdot e^{(1-ax)^{-1}}\cdot\frac{1}{1-ax}\cdot a$$

6. Cyclic Integration

The cyclic integral, namely, the inverse of the cyclic derivative, turns out to be much simpler to compute than the usual commutative integral. Given a formal power series $f \in K\{\{a, b,\dots, x\}\}$ we say that the formal power series g is an *integral* of f,

$$\int \langle f\rangle\, dx = g,$$

whenever $D\langle g\rangle = f$. Clearly, if g_0 is any formal power series such that $D\langle g_0\rangle = 0$ then we also have $\int\langle f\rangle\, dx = g + g_0$. Thus, we begin by describing the kernel of the cyclic derivative D.

For $g_0 \in \ker D$, let $g_0 = \sum_{l=0}^{\infty} p^{(l)}$ be the expansion of g_0 into polynomials homogeneous of length l. Now $D\langle g_0\rangle = \sum_{l=0}^{\infty} D\langle p^{(l)}\rangle = 0$ if and only if $D\langle p^{(l)}\rangle = 0$ for all l, hence it suffices to determine the homogeneous polynomials in the kernel of D.

If $C\langle p^{(l)}\rangle = 0$ then necessarily $D\langle p^{(l)}\rangle = 0$. Conversely, if $D\langle p^{(l)}\rangle = 0$ consider the decomposition

$$p^{(l)} = p_0^{(l)} + p_1^{(l)}$$

where $p_0^{(l)}$ contains all the monomials of degree zero occurring in $p^{(l)}$ while $p_1^{(l)}$ contains all the monomials of positive degree (recall that the degree of a monomial w is the number of occurrences of x in w). Clearly $D\langle p_0^{(l)}\rangle = 0$ and it follows that $D\langle p_1^{(l)}\rangle = 0$. In fact we must have $C\langle p_1^{(l)}\rangle = 0$. For if $C\langle p_1^{(l)}\rangle = \sum_{i=1}^{n} k_i w_i$, where the w_i are distinct words of positive degree and each $k_i \neq 0$, then for some index j we must have $w_j = xv$ for some word v. But now $T\langle w_j\rangle = v \neq T\langle w_i\rangle$ for any $i \neq j$. Thus $k_j v$ is a nonzero summand in $T\langle C\langle p_1^{(l)}\rangle\rangle = D\langle p_1^{(l)}\rangle$ which is a contradiction. Hence $\ker D = \ker C$ for polynomials of positive degree.

The polynomials in $\ker C$ are of the form $p - q$ where q is a cyclic permutation of p. More precisely, linearly order the alphabet A. For any word w let w^* be that summand in $C\langle w\rangle$ which is smallest lexicographically. The word w^* is unique, although it may appear several times in $C\langle w\rangle$.

PROPOSITION 6.1. *Let $V^{(l)}$ be the set of all homogeneous polynomials of length l in the kernel of D. Also let $V_0^{(l)}$ (respectively, $V_1^{(l)}$) be the set of all homogeneous polynomials of length l and degree zero (respectively, positive degree) in the kernel of D. Then:*

(i) *$V^{(l)}$, $V_0^{(l)}$, and $V_1^{(l)}$ are vector spaces for all l with*

$$V^{(l)} = V_0^{(l)} + V_1^{(l)}.$$

(ii) *$V_0^{(l)}$ contains every polynomial of degree 0 and length l.*

(iii) *$V_0^{(l)}$ has as a basis the set*

$$\{w - w^* \mid w \text{ is a word of length } l \text{ and positive degree}, w - w^* \neq 0\}.$$

Proof. (i) Easy, left to the reader.

(ii) Clear, since the derivative of a constant is zero.

(iii) Obviously $w - w^* \in \ker D$ for any w. To show that the given set of expressions span, take $p_1^{(l)} \in V_1^{(l)}$ say

$$p_1^{(l)} = \sum_{i=1}^{I-1} k_i w_i + \sum_{i=I}^{J-1} k_i w_i + \cdots + \sum_{i=M}^{N} k_i w_i \tag{6.2}$$

where the words are grouped so that w_i and w_j occur in the same summation if and only if $C\langle w_i\rangle = C\langle w_j\rangle$.

Now from $\ker D = \ker C$ on $V_1^{(l)}$ we have

$$\begin{aligned}0 = C\langle p_1^{(l)}\rangle &= \sum_{i=1}^{I-1} k_i C\langle w_i\rangle + \sum_{i=I}^{J-1} k_i C\langle w_i\rangle + \cdots + \sum_{i=M}^{N} k_i\langle w_i\rangle \\ &= \sum_{i=1}^{I-1} k_i C\langle w_1^*\rangle + \sum_{i=I}^{J-1} k_i C\langle w_I^*\rangle + \cdots + \sum_{i=M}^{N} k_i C\langle w_M^*\rangle\end{aligned}$$

whence

$$\sum_{i=1}^{I-1} k_i = \sum_{i=I}^{J-1} k_i = \cdots = \sum_{i=M}^{N} k_i = 0.$$

Thus we can write

$$\begin{aligned}p_1^{(l)} &= \sum_{i=1}^{I-1} k_i(w_i - w_1^*) + \sum_{i=I}^{J-1} k_i(w_i - w_I^*) + \cdots + \sum_{i=M}^{N} k_i(w_i - w_M^*) \\ &= \sum_{i=1}^{I-1} k_i(w_i - w_i^*) + \sum_{i=I}^{J-1} k_i(w_i - w_i^*) + \cdots + \sum_{i=M}^{N} k_i(w_i - w_i^*) \\ &= \sum_{i=1}^{N} k_i(w_i - w_i^*), \qquad \text{as desired.}\end{aligned}$$

Linear independence of the $w - w^*$ follows immediately from the fact that each binomial contains a word not found in any other. Q.E.D.

We can now determine which formal power series are integrable:

THEOREM 6.3. *For any formal power series f the following statements are equivalent*:

(i) *f is integrable.*

(ii) *If a word w is a summand in f with coefficient $k \in K$ then every word in $D\langle xw\rangle$ is a summand in f with the same coefficient k.*

Proof. (i) $\to$ (ii) Suppose there exists a formal power series g such that $D\langle g\rangle = f$. Suppose kw is a summand in f, and let the length of w be $l - 1 \geqslant 0$. If we denote by $p_1^{(l)}$ the homogeneous component of g of length l and positive degree then kw is a summand in $D\langle p_1^{(l)}\rangle$. Decomposing $p_1^{(l)}$ as in (6.2) we obtain

$$D\langle p_1^{(l)}\rangle = \sum_{i=1}^{I-1} k_i D\langle w_1^*\rangle + \sum_{i=I}^{J-1} k_i D\langle w_I^*\rangle + \cdots + \sum_{i=M}^{N-1} k_i D\langle w_M^*\rangle.$$

Since the summands in $D\langle w_1^*\rangle, D\langle w_I^*\rangle, \ldots, D\langle w_M^*\rangle$ are all distinct, kw is a summand in $\sum k_i D\langle w_L^*\rangle$ for some index L. So $D\langle xw\rangle = D\langle w_L\rangle^*$ and hence every word in $D\langle xw\rangle$ appears with the same coefficient.

(ii) $\to$ (i). It suffices to show that the lth homogeneous component $p^{(l)}$ of f is integrable. Induct on the number of summands in $p^{(l)}$. If $p^{(l)} = kw$ then $D\langle xw\rangle$ has only one term, so that $xw = (xv)^m$ for some word v not containing x and some integer $m \geqslant 1$. Thus $\int \langle p^{(l)}\rangle\, dx = (1/m)k \cdot xw$.

Now consider $p^{(l)} = \sum_{i=1}^n k_i w_i$ where $k_1 w_1 = kw$. Clearly

$$D\langle xw\rangle = mw + mv_1 + mv_2 + \cdots + mv_r \tag{6.4}$$

where the w and v_i are distinct words and m is an integer greater than 0. By assumption $p^{(l)} - (k/m)\, D\langle xw\rangle$ has $n - r - 1 < n$ summands and by induction we can find g such that

$$\int \left\langle p^{(l)} - \frac{k}{m}\, D\langle xw\rangle\right\rangle dx = g,$$

that is,

$$\int \langle p^{(l)}\rangle\, dx = g + \frac{k}{m} \cdot xw. \qquad \text{Q.E.D.}$$

The second half of the preceding proof gives an algorithm for computing $\int \langle f\rangle\, dx$ whenever the integral exists. Specifically, take any monomial kw in f of minimal length. If $D\langle xw\rangle$ does not satisfy (ii) then the algorithm terminates and f is not integrable. If (ii) is satisfied then $(k/m) \cdot xw$ (m as in (6.4)) is noted as a summand in $\int \langle f \rangle\, dx$ and the procedure is applied again to $f - (k/m) \cdot xw$.

EXAMPLE 6.5. $\int \langle e^{ax} \cdot a \rangle\, dx = e^{ax}$. However Theorem 6.3 shows that $\int \langle e^{ax} \rangle\, dx$ and $\int \langle a \cdot e^{ax} \rangle\, dx$ do not exist. Similar results hold for e^{xa}.

EXAMPLE 6.6.

$$\int \langle P\langle a^i b^j c^k \cdots x^n \rangle \rangle\, dx = \frac{1}{i + j + k + \cdots + n + 1} \cdot P\langle a^i b^j c^k \cdots x^{n+1} \rangle$$

(cf. Corollary 3.7). Also for $i \geqslant 1$

$$\int \langle P\langle a^i b^j c^k \cdots x^n \rangle \rangle\, dx = \frac{1}{i}\, a \cdot P\langle a^{i-1} b^j c^k \cdots x^{n+1} \rangle$$

(cf. Example 3.10). Hence for $i \geqslant 1$

$$D \left\langle \frac{1}{i + j + k + \cdots + n + 1} \cdot P\langle a^i b^j c^k \cdots x^{n+1} \rangle \right.$$
$$\left. - \frac{1}{i}\, a \cdot P\langle a^{i-1} b^j c^k \cdots x^{n+1} \rangle \right\rangle = 0.$$

7. THE CAYLEY OPERATOR

Let A be an algebra which is the quotient of the formal power series algebra $K\{\{a, b,..., x\}\}$ by an ideal I. Suppose that the ideal I is invariant under D, that is, that $\langle Df \mid g \rangle \in I$ whenever either f or g belongs to I. Then D induces an operator, again denoted by D, on the quotient $K\{\{a,..., x\}\}/I$, and such an operator will enjoy the same formal properties as the operator D. Thus, a cyclic derivative can be defined on several algebras of common occurrence; among these, the simplest is the commutative algebra, defined by the ideal generated by the identity $c_1 c_2 - c_2 c_1$, and more generally, the algebras satisfying the standard identities

$$\sum_{\sigma} {}_{\sigma} c_{\sigma 1} c_{\sigma 2} \cdots c_{\sigma n} = 0$$

for some n.

We shall consider one particular case in detail, namely, a finite-dimensional matrix algebra generated by "constant" matrices and a "variable" matrix X, which, as we shall see, can be considered as a matrix whose entries are independent transcendentals x_{ij}. It will turn out that the cyclic derivative on this algebra is $\Omega\, \mathrm{tr}(F)$, where Ω is the Cayley operator of classical invariant theory, and $\operatorname{tr}(F)$ is the trace of the matrix F. This application shows that the cyclic derivative is intimately connected with questions relating to invariant theory.

Consider the noncommutative algebra $K\{\{e_{11}, e_{12}, \ldots, e_{nn}\}\}$ with n^2 constants e_{ij}, $1 \leqslant i, j \leqslant n$, and let $K^1 = K[[x_{11}, x_{12}, \ldots, x_{nn}]]$ be the commutative formal power series algebra in n^2 indeterminates. We denote the algebra of $n \times n$ matrices with entries in K^1 by $\operatorname{Mat}_n(K^1)$. Define a map

$$\phi\colon K\{\{e_{11}, e_{12}, \ldots, e_{nn}, X\}\} \to \operatorname{Mat}_n(K^1)$$

by

$$\phi(e_{ij}) = E_{ij}, \qquad \text{where} \quad E_{ij} = \text{the matrix having 1 in the } (i, j) \text{ position and 0 elsewhere},$$

and

$$\phi(x) = X, \qquad \text{where} \quad X = \sum_{i,j} x_{ij} E_{ij} = [x_{ij}].$$

Obviously ϕ is onto and $I = \ker \phi$ is a two-sided ideal generated by $e_{ij}e_{kl} - \delta_{jk}e_{il}$. Hence,

$$\Phi\colon K\{\{e_{11}, e_{12}, \ldots, e_{nn}, x\}\}/I \to \operatorname{Mat}_n(K^1)$$

is an isomorphism. From the remarks above the cyclic derivative operator extends to $K\{\{e_{11}, e_{12}, \ldots, e_{nn}, x\}\}/I$ and thus to $\operatorname{Mat}_n(K^1)$ with $D\langle X\rangle = I$, $D\langle E_{ij}\rangle = 0$ where I is the identity matrix and 0 is the zero matrix.

Recall that the Cayley operator Ω is defined as the map (see, for example, [6] or [7])

$$\Omega\colon K^1 \to \operatorname{Mat}_n(K^1)$$

where

$$\Omega(f) = \left[\frac{\partial f}{\partial x_{ji}}\right] \qquad \text{for} \quad f \in K^1.$$

The (i, j) entry of this matrix is the partial derivative with respect to x_{ji}, the transpose of what might be expected.

It turns out that the Cayley operator is closely related to the image of the cyclic derivative in the matrix algebra:

THEOREM 7.1. *Let $F = \sum_{l=0}^{\infty} P^{(l)}$ where $P^{(l)}$ is a homogeneous polynomial of length l in the matrix X and matrices with entries in K. Then*

$$D\langle F\rangle = \Omega(\operatorname{trace} F).$$

Proof. By linearity and continuity we can specialize to the case where F is a word, say, of degree d. In fact we can assume that $F = A_1XA_2XA_3 \cdots XA_{d+1}$ where the matrices A_k have scalar entries. Now,

$$
\begin{aligned}
\Omega(\operatorname{trace} F) &= \Omega\left(\sum_{p,q,r,s,t,u,\ldots,v} a^1_{pq}x_{qr}a^2_{rs}x_{st}a^3_{tu}\cdots a^{d+1}_{vp}\right) \\
&= \left[\frac{\partial}{\partial x_{ji}}\left(\sum_{p,q,r,s,t,u,\ldots,v} a^1_{pq}x_{qr}a^2_{rs}x_{st}a^3_{tu}\cdots a^{d+1}_{vp}\right)\right] \\
&= \left[\sum_{p,s,t,u,\ldots,v} a^1_{pj}\hat{x}_{ji}a^2_{is}x_{st}a^3_{tu}\cdots a^{d+1}_{vp}\right. \\
&\qquad \left. + \sum_{p,q,r,u,\ldots,v} a^1_{pq}x_{qr}a^2_{rj}\hat{x}_{ji}a^3_{iu}\cdots a^{d+1}_{vp} + \cdots\right]
\end{aligned}
$$

where $\hat{x}_{ji}$ indicates that x_{ji} is deleted

$$
\begin{aligned}
&= \left[\sum_{p,s,t,u,\ldots,v} a^2_{is}x_{st}a^3_{tu}\cdots a^{d+1}_{vp}a^1_{pj}\right] \\
&\quad + \left[\sum_{p,q,r,u,\ldots,v} a^3_{iu}\cdots a^{d+1}_{vp}a^1_{pq}x_{qr}a^2_{rj}\right] + \cdots \\
&= A_2XA_3\cdots A_{d+1}A_1 + A_3X\cdots A_{d+1}A_1XA_2 + \cdots \\
&= D\langle A_1XA_2XA_3\cdots A_{d+1}\rangle. \qquad \text{Q.E.D.}
\end{aligned}
$$

8. Further Work

Several lines of work are suggested by the preceding considerations.

(1) An investigation of the cyclic derivative in certain specific algebras might lead to more general versions of Taylor's formula. The algebras defined by standard identities mentioned in the preceding section are prime candidates.

(2) Is it possible, by a limiting process, to extend the cyclic derivative to the von Neumann hyperfinite algebra? This might be a foot in the door for an extension of some of the results of the theory of invariants for matrices to von Neumann algebras.

(3) We have not been able to obtain any significant properties of cyclic derivatives in several variables, say for formal series containing x and y. The main difficulty is that the cyclic derivatives in x and y do not commute.

(4) The cyclic chain rule (Theorem 4.2) probably extends to higher order derivatives, but the extension remains to be carried out.

(5) Differential equations in the cyclic derivative seems a promising field of investigation. In this connection, some recent work of Schützenberger is suggestive (see [5]).

(6) The connection between the Cayley operator and the cyclic derivative, developed in Theorem 7.1, suggests that one gets at other operators of classical invariant theory by similar methods. We are thinking of the operators $A \rightarrow \Omega\, \mathrm{tr}(A^{(k)})$, where $A^{(k)}$ is the kth compound matrix of the n by n matrix A, that is, the matrix whose entries are all the k by k minors of A. Is it possible to define operators on the free algebra which reduce to these operators by the method of Section 7? The case $k = n$ is particularly important. We note that for $k > 1$ these operators are not linear.

(7) It may be worthwhile to determine all linear operators which commute with the cyclic derivative. In this connection, it is worth noting that the cyclic derivative is associated to a coalgebra structure on the free algebra in the letters $x, a, b, \dots$. If m is a word, set

$$\Delta m = \sum_{i,j} m_i \otimes m_j ,$$

where m_i and m_j are all words whose product $m_i m_j$ is a cyclic permutation of the word m. Defining a linear functional L as $L(x) = 1$, $L(m) = 0$ for all other words, one sees that

$$Dm = \sum_{i,j} L(m_i)\, m_j .$$

The comultiplication Δ is coassociative, and is in fact associated with a bialgebra structure, where the multiplication of two words is not their juxtaposition but their shuffle product.

(8) By investigating the analogs of Abel's functional equation for the cyclic derivative, one may be led to a cyclic analog of the Lagrange inversion formula. In this connection, some recent work of Gessel points the way [2].

(9) The fact that the cyclic derivatives of elementary transcendental functions are simple generalizations of the commutative case supports the conjecture that a theory of noncommutative hypergeometric functions can be developed in the present context.

(10) The cyclic derivative can be extended to polynomials in the constants $a, a^{-1}, b, b^{-1}, \dots, x, x^{-1}$ by setting $Dx^{-1} = -x^{-1}ax^{-1}$, and one easily verifies that this is the only consistent definition. It is, however, difficult to complete this algebra to an algebra of formal power series.

References

1. S. Eilenberg, "Automata, Languages, and Machines," Vol. A., Academic Press, New York, 1974.
2. I. M. Gessel, A noncommutative generalization and q-analog of the Lagrange inversion formula, preprint.
3. R. C. Lyndon and P. E. Schuff, "Combinatorial Group Theory," Springer, Berlin/New York, 1977.
4. W. Magnus, A. Karrass, and D. Solitar, "Combinatorial Group Theory," Interscience, New York, 1966.
5. M. P. Schützenberger, Solution non-commutative d'une équation différentielle classique, *in* "New Concepts and Technologies in Parallel Information Processing" (E. R. Caianiello Ed.), pp. 381–401, Nordhoff Gröningen, 1975.
6. H. W. Turnbull, On differentiating a matrix, *Proc. Edinburgh Math. Soc. (2)* 1 (1927), 111–128.
7. H. W. Turnbull, A matrix form of Taylor's theorem, *Proc. Edinburgh Math. Soc. (2)* 2 (1929), 33–54.