Notes on the theorem of Baker-Campbell-Hausdorff-Dynkin

Michael Müger
Radboud University, Nijmegen, The Netherlands

April 22, 2020

Abstract

This is a thorough exposition [initially motivated by a course on matrix Lie groups that I am teaching, but things got a bit out of hand] of many proofs for the Baker-Campbell-Hausdorff (BCH) theorem according to which $\log(e^X e^Y)$, where X, Y are non-commuting variables, is a Lie series, and of several for Dynkin's explicit formula for this Lie series. We begin with analytic approaches, extensively discussing the derivative of the exponential function in a Banach algebra, and proving the BCH theorem and Dynkin's formula for small enough X, Y. We then turn to purely algebraic approaches in the framework of formal non-commutative power and Lie series over any field of characteristic zero, giving Eichler's elegant proof of BCH and the classical proof of BCHD based on the criteria of Friedrichs and Dynkin-Specht-Wever, for which we give a slick proof. Apart from the standard Dynkin expansion of the BCH series we prove several others, increasingly combinatorial, due to Reinsch, BCH, Dynkin (not the well-known one!) and Goldberg.

None of the results is new, but I make a point of giving a complete exposition of a sadly neglected pioneering paper by Dynkin (1949), and I have tried hard to make the proofs technically and conceptually accessible.

Contents

1 Introduction

Let $\mathcal{A}$ be an associative algebra over a field $\mathbb{F}$ and $X \in \mathcal{A}$. In order to make sense of $e^X \equiv \exp(X) = \sum_{n=0}^\infty \frac{X^n}{n!}$, we surely want $\mathbb{F}$ to have characteristic zero and $\mathcal{A}$ to be unital (so that we can put $X^0 = 1$). The question of convergence of the series can be studied in a setting of formal power series or of Banach algebras, and we will do this later. For now we ignore all questions of convergence.

If $X, Y \in \mathcal{A}$ commute, then the following well known computation holds:

$$e^{X+Y} = \sum_{n=0}^\infty \frac{(X+Y)^n}{n!} = \sum_{n=0}^\infty \sum_{k=0}^n \binom{n}{k} \frac{X^k Y^{n-k}}{n!} = \sum_{n=0}^\infty \sum_{k=0}^n \frac{X^k Y^{n-k}}{k!\,(n-k)!} = \sum_{r,s=0}^\infty \frac{X^r Y^s}{r!\,s!} = e^X e^Y. \tag{1.1}$$

(For a partial converse see Lemma D.1.) If $XY \neq YX$ then the step $(X+Y)^n = \sum_{k=0}^n \binom{n}{k} X^k Y^{n-k}$ breaks down. But one may wonder whether there still exists a $Z \in \mathcal{A}$ such that $e^X e^Y = e^Z$, and a natural attempt is to put $Z = \log(e^X e^Y)$ using the power series

$$\log X = \sum_{k=1}^\infty \frac{(-1)^{k-1}}{k} (X-1)^k. \tag{1.2}$$

Replacing X by $e^X e^Y$ herein, we obtain the BCH series, denoted $H(X,Y)$ or just H,

$$\begin{aligned} H(X,Y) = \log(e^X e^Y) &= \sum_{k=1}^\infty \frac{(-1)^{k-1}}{k} \left(\sum_{m,n=0}^\infty \frac{X^m Y^n}{m!\,n!} - 1 \right)^k \\ &= \sum_{k=1}^\infty \sum_{m_1+n_1>0} \cdots \sum_{m_k+n_k>0} \frac{(-1)^{k-1}}{k} \frac{X^{m_1} Y^{n_1} \cdots X^{m_k} Y^{n_k}}{m_1!\,n_1! \cdots m_k!\,n_k!} \\ &= \underbrace{\left(X + Y + XY + \frac{X^2+Y^2}{2} + \cdots\right)}_{k=1} \underbrace{- \frac{1}{2}(X^2 + Y^2 + XY + YX + \cdots)}_{k=2} + \cdots \\ &= X + Y + \frac{1}{2}[X,Y] + \cdots, \end{aligned} \tag{1.3}$$

where we have worked out the terms of order one and two (which requires considering $k \in \{1,2\}$) and $[X,Y] := XY - YX$ is the *commutator*. We notice the following: The terms in the infinite sum (1.3) corresponding to some k all have order $\geq k$ in X, Y, due to the condition $m_i + n_i > 0$. And the contribution that a certain k in the sum makes to the order k part of $\log(e^X e^Y)$ is easy to determine: It is just $\frac{(-1)^{k-1}}{k}$ times the sum of all 2^k words in X, Y of length k. But all the $k < \ell$ also contribute terms of order ℓ, and determining those is tedious and error-prone.

1.1 Exercise Compute the contribution of order 3 and show that it equals $\frac{1}{12}([X,[X,Y]] + [Y,[Y,X]])$.

Solution. The order 3 contribution in (1.3) is

$$\underbrace{\left(\frac{X^2Y+XY^2}{2}+\frac{X^3+Y^3}{6}\right)}_{k=1}$$
$$\underbrace{-\frac{1}{2}\left(XYX+XY^2+X^2Y+YXY+\frac{X^3+X^2Y+Y^2X+Y^3+X^3+YX^2+XY^2+Y^3}{2}\right)}_{k=2}$$
$$\underbrace{+\frac{1}{3}(X^3+X^2Y+XYX+XY^2+YX^2+YXY+Y^2X+Y^3)}_{k=3}.$$

Collecting like terms, we obtain

$$\begin{aligned}&\frac{1}{12}\left(X^2Y-2XYX+XY^2+YX^2-2YXY+Y^2X\right)\\&\quad=\frac{1}{12}(X(XY-YX)-(XY-YX)X+Y(YX-XY)-(YX-XY)Y)\\&\quad=\frac{1}{12}([X,[X,Y]]+[Y,[Y,X]]).\end{aligned}$$

■

1.2 Exercise Compute the contribution of order 4 and show that it equals $\frac{1}{24}[Y,[X,[Y,X]]]$.

These results suggest that $\log(e^Xe^Y)$ can be written as a (formal, possibly convergent) series $X+Y+\cdots$, where all order ≥ 2 terms can be expressed as linear combinations of (iterated) commutators $[\cdot,\cdot]$. In this case, one speaks of a Lie series. Indeed, we will give many[1,2] different proofs of the first statement of the following theorem and two for the second, plus several other expansions.

1.3 Theorem (Campbell-Baker-Hausdorff-Dynkin) [3] *Let $\mathcal{A}$ be a unital algebra over a field of characteristic zero and let $X,Y\in\mathcal{A}$. Then*

- *(BCH) $\log(e^Xe^Y)$ is given by a Lie series*
- *(D) with the concrete series representation*

$$H(X,Y)=\log(e^Xe^Y) = \sum_{k=1}^{\infty}\sum_{m_1+n_1>0}\cdots\sum_{m_k+n_k>0}\frac{(-1)^{k-1}}{k\sum_{i=1}^{k}(m_i+n_i)}\frac{1}{m_1!\,n_1!\cdots m_k!\,n_k!}$$
$$\underbrace{[X,[\cdots,[X,}_{m_1}\underbrace{[Y,[\cdots,[Y,}_{n_1}[\ldots\underbrace{[X,[\cdots,[X,}_{m_k}\underbrace{[Y,[\cdots,[Y,}_{n_k}[\cdots]\cdots] \tag{1.4}$$

with the understanding that $[X]:=X$.

[1] An entire book [6] of 550 pages is dedicated to BCHD theorem, its history and at least six proofs and several variations.

[2] Michael Atiyah (1929-2019, British mathematician, Fields medal and Abel prize): "I think it is said that Gauss had ten different proofs for the law of quadratic reciprocity. [By now, more than 200 have been published!] Any good theorem should have several proofs, the more the better. For two reasons: usually, different proofs have different strengths and weaknesses, and they generalize in different directions – they are not just repetitions of each other. And that is certainly the case with the proofs that we came up with [of the Index Theorem]. There are different reasons for the proofs, they have different histories and backgrounds. Some of them are good for this application, some are good for that application. They all shed light on the area. If you cannot look at a problem from different directions, it is probably not very interesting; the more perspectives, the better!" [46]

[3] John Edward Campbell (1862-1924), Irish mathematician. Henry Frederick Baker (1866-1956), British mathematician. Felix Hausdorff (1868-1942), German mathematician who contributed to many areas, particularly point-set topology. Eugene Borisovich Dynkin (1924-2014), Soviet, then American mathematician. Many important contributions to Lie group theory (Dynkin diagrams!) and to probability.

1.4 Remark 1. In the statement of the above theorem we have been deliberately vague. The reason is that it can be formulated in different settings. In the sections that follow we will give precise formulations and proofs in the context of non-commutative formal power and Lie series over a field of characteristic zero, and as an analytic statement in a unital Banach algebra over $\mathbb{R}$ or $\mathbb{C}$.

2. For the history of the theorem see [53, 15, 1, 6]. We limit ourselves to citing the papers [4, 8, 33] that gave rise to the abbreviation BCH (which neglects the contributions of Poincaré[4], Pascal[5] and F. Schur[6]). In any case, Dynkin [18] was the first to give a semi-explicit expansion. Since then there has been a flood of both theoretical approaches of considerable sophistication (e.g. [7, 49]) and computational/algorithmic studies.

3. Evaluating Dynkin's formula (1.4) is unpleasant, even at relatively low order, as already Exercises 1.1, 1.2 demonstrate. (This involved the different formula (1.3), but the combinatorics is the same.) Computational improvements will be provided by Theorem A.1 due Reinsch and by Theorem 9.3, also due to Dynkin.

4. Note that summands in Dynkin's formula with $n_k \geq 2$ vanish since $[Y, Y] = 0$, and the same happens if $n_k = 0$ and $m_k \geq 2$. If $n_k = 1$, then any $m_k \geq 0$ is allowed, and if $n_k = 0$ then we must have $m_k = 1$ due to $m_k + n_k > 0$. Taking this into account, and writing $\mathrm{ad}_X(Y) = [X, Y]$, the formula takes the following form:

$$\begin{aligned} H(X,Y) \;=\; & \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{m_1+n_1>0} \cdots \sum_{m_{k-1}+n_{k-1}>0} \left(\frac{1}{\sum_{i=1}^{k-1}(m_i+n_i)+1} \frac{\mathrm{ad}_X^{m_1}\mathrm{ad}_Y^{n_1}\cdots \mathrm{ad}_X^{m_{k-1}}\mathrm{ad}_Y^{n_{k-1}}(X)}{m_1!\,n_1!\cdots m_{k-1}!\,n_{k-1}!} \right. \\ & \left. + \sum_{m_k\geq 0} \frac{1}{\sum_{i=1}^{k} m_i + \sum_{i=1}^{k-1} n_i + 1} \frac{\mathrm{ad}_X^{m_1}\mathrm{ad}_Y^{n_1}\cdots \mathrm{ad}_X^{m_{k-1}}\mathrm{ad}_Y^{n_{k-1}}\mathrm{ad}_X^{m_k}(Y)}{m_1!\,n_1!\cdots m_{k-1}!\,n_{k-1}!\,m_k!} \right). \end{aligned} \tag{1.5}$$

5. We will not discuss at all how the BCH(D) theorem is applied in the theory of Lie groups (or elsewhere). For this, see e.g. [17, 29, 32, 36, 50]. □

Acknowledgment. I thank Darij Grinberg for countless suggestions and corrections, like exposing a hole in my original proof of Proposition B.7, which led to a much better alternative proof. I am grateful to the authors of [1] for saving [19] from oblivion.

2 Standard analytic proof of BCHD

In this section we give an analytic proof of BCH that was known in essentially this form by 1900. We will then give a straightforward deduction of Dynkin's formula (1.4), which strangely enough was proven only 50 years later, by different methods. Throughout, we will work in a Banach algebra of arbitrary dimension.

2.1 Preparations

2.1 Lemma *Let $\mathcal{A}$ be a unital Banach algebra. (We do not require $\|1\| = 1$.)*

(i) If a sequence $\{X_n\}_{n\in\mathbb{N}_0}$ in $\mathcal{A}$ satisfies $\sum_n \|X_n\| < \infty$ then $\sum_{n=0}^N X_n$ converges to some $X \in \mathcal{A}$ as $N \to \infty$ and $\|X\| \leq \sum_n \|X_n\|$.

(ii) If the power series $f(z) = \sum_{n=0}^\infty a_n z^n$ (where $\{a_n\} \subset \mathbb{C}$) has convergence radius $R > 0$ and $X \in \mathcal{A}$ satisfies $\|X\| < R$ then we can define $f(X) \in \mathcal{A}$ by $\sum_{n=0}^\infty a_n X^n$. (It actually suffices if $r(X) < R$, where $r(X)$ is the spectral radius of X. One always has $r(X) \leq \|X\|$, but the inequality may be strict.)

Proof. (i) This is a well known consequence of completeness.

(ii) The hypothesis $\|X\| < R$ implies that $\sum_n |a_n|\|X\|^n < \infty$. Now apply (i). For the parenthetical remark we observe that for convergence of $\sum_n a_n X^n$ it suffices that $\limsup_n \|X^n\|^{1/n} < R$. Now the claim follows from the Beurling-Gelfand formula $r(X) = \lim_{n\to\infty} \|X^n\|^{1/n}$. ■

If $\mathcal{A}$ is a unital algebra and $A \in \mathcal{A}$ is invertible, we write $\mathrm{Ad}_A(B) = ABA^{-1}$.

[4] Henri Poincaré (1854-1912), French mathematician and mathematical physicist.
[5] Ernesto Pascal (1865-1940). Italian mathematician.
[6] Friedrich Heinrich Schur (1856-1932). German mathematician ($\neq$ Issai Schur).

2.2 PROPOSITION *Let $\mathcal{A}$ be a unital Banach algebra. For all $X \in \mathcal{A}$, define $e^X = \exp(X) = \sum_{n=0}^{\infty} X^n/n!$.*

(i) If $X, Y \in \mathcal{A}$ satisfy $[X, Y] = 0$, i.e. $XY = YX$, then $e^{X+Y} = e^X e^Y = e^Y e^X$.

(ii) For every $X \in \mathcal{A}$ we have that e^X is invertible with $(e^X)^{-1} = e^{-X}$.

(iii) For all $X, Y \in \mathcal{A}$ we have the following connection between ad *and* Ad*:*

$$\mathrm{Ad}_{e^X}(Y) = e^{\mathrm{ad}_X}(Y). \tag{2.1}$$

Proof. (i) This is just the computation of (1.1), which is rigorous since everything converges absolutely.

(ii) Since $[X, -X] = 0$, (i) implies $e^X e^{-X} = e^{X-X} = e^0 = 1$ and similarly $e^{-X} e^X = 1$.

(iii) For $X, Y, Z \in \mathcal{A}$, define $L_X(Z) = XZ$ and $R_Y(Z) = ZY$. Now $L_X, R_Y : \mathcal{A} \to \mathcal{A}$ are bounded linear maps, thus elements of the unital algebra $\mathcal{B}(\mathcal{A})$, which is complete and thus Banach. Quite trivially, $L_X(R_Y(Z)) = XZY = R_Y(L_X(Z))$, so that $[L_X, R_Y] = 0$. Thus by (i) we have the identity $e^{L_X} e^{R_{-X}} = e^{L_X - R_X}$. Furthermore $(L_X - R_X)(Y) = XY - YX = [X, Y] = \mathrm{ad}_X(Y)$, thus $L_X - R_X = \mathrm{ad}_X$. Now

$$\mathrm{Ad}_{e^X}(Y) = e^X Y e^{-X} = L_{e^X} R_{e^{-X}}(Y) = e^{L_X} e^{R_{-X}}(Y) = e^{L_X - R_X}(Y) = e^{\mathrm{ad}_X}(Y),$$

where we have also used the triviality $(L_X)^n = L_{X^n}$, which implies $L_{e^X} = e^{L_X}$, and similarly for R_{e^X}. ■

2.3 REMARK Alternatively, one can prove (2.1) by a differential equation approach, see e.g. [34, p. 66] or [50, p. 14]: Differentiating $U : \mathbb{R} \to \mathcal{A}$, $t \mapsto e^{tX} Y e^{-tX}$ gives

$$U'(t) = Xe^{tX} Y e^{-tX} - e^{tX} Y e^{-tX} X = XU(t) - U(t)X = \mathrm{ad}_X(U(t)).$$

Since ad_X is a constant bounded linear operator on $\mathcal{A}$, this linear differential equation can be solved readily for the initial condition $U(0) = Y$, yielding $U(t) = e^{t\,\mathrm{ad}_X}(Y)$. Putting $t = 1$ gives $\mathrm{Ad}_{e^X}(Y) = U(1) = e^{\mathrm{ad}_X}(Y)$.

The first proof is slightly more elementary in that it does not use a differential equation. □

In Appendix D we collect several further results related to the fact that in general $e^{X+Y} \neq e^X e^Y$ in a non-commutative setting, which however are somewhat tangential to the subject of this paper.

Having discussed the exponential function to some extent, we turn to the logarithm:

2.4 PROPOSITION *Let $\mathcal{A}$ be a unital Banach algebra and $X, Y \in \mathcal{A}$.*

(i) If $X \in \mathcal{A}$ satisfies $\|X - 1\| < 1$ then $W = \log(X)$ defined by (1.2) satisfies $e^W = X$.

(ii) If $\|X\| + \|Y\| < \log 2$ then the series (1.3) for $Z = \log(e^X e^Y)$ converges absolutely and $e^Z = e^X e^Y$.

(iii) If $\|X\| < \log 2$ then $\|e^X - 1\| < 1$ and $\log e^X = X$.

Proof. (i) Since the power series (1.2) (around $X = 1$) has convergence radius $R = 1$, we can use Lemma 2.1(ii) to define $W = \log(X)$ for $X \in \mathcal{A}$ with $\|X - 1\| < 1$. Now e^W makes sense unconditionally. Recall that $\exp(\log z) = z$ if $|z - 1| < 1$. (For a purely algebraic proof of the fact that the composite series $\exp \circ \log$ is the identity cf. Appendix C.) Now the claim $e^W = X$ follows from the equality $\exp(\log(X)) = (\exp \circ \log)(X)$, which is essentially tautological.

(ii) We have

$$\|e^X e^Y - 1\| = \left\| \sum_{\substack{n,m \geq 0 \\ n+m>0}} \frac{X^n Y^m}{n!\, m!} \right\| \leq \sum_{\substack{n,m \geq 0 \\ n+m>0}} \frac{\|X\|^n \|Y\|^m}{n!\, m!} = e^{\|X\| + \|Y\|} - 1.$$

Since the assumption $\|X\| + \|Y\| < \log 2$ is equivalent to $e^{\|X\|+\|Y\|} - 1 < 1$, we can apply (i).

(iii) Arguing as before, $\|e^X - 1\| = \| \sum_{n=1}^{\infty} X^n/n! \| \leq \sum_{n=1}^{\infty} \|X\|^n/n! = e^{\|X\|} - 1$, the hypothesis implies $\|e^X - 1\| < 1$ so that $\log e^X$ is defined, and this equals X by the same argument as in (i). ■

2.5 REMARK 1. Note that the assumption in (iii) cannot be replaced by $\|e^X - 1\| < 1$: If $X = 2\pi i$ then $e^X = 1$ so that $\log e^X = 0 \neq X$.

2. Since we do not know yet that the formulae (1.3) and (1.4) coincide, we haven't proven yet that the r.h.s. of (1.4) converges for $\|X\| + \|Y\| < \log 2$. A similar argument, but using $\|\mathrm{ad}_X\| \leq 2\|X\|$, gives this convergence of (1.4) for $\|X\| + \|Y\| < \frac{\log 2}{2}$. (This condition will turn out to be sufficient for equality of the series (1.3) and (1.4), see Theorem 2.14.) □

2.2 The derivative of the exponential function: Three proofs

Let $\mathcal{A}$ be a unital Banach algebra and $A, B \in \mathcal{A}$. If $[A, B] = 0$ then

$$\frac{d}{dt} e^{A+tB} \bigg|_{t=0} = \frac{d}{dt} e^A e^{tB} \bigg|_{t=0} = e^A B = B e^A. \tag{2.2}$$

But if A and B do not commute, the first step above is not justified. As a first observation one has:

2.6 Lemma *Let $\mathcal{A}$ be a unital Banach algebra and $A, B \in \mathcal{A}$. Then* exp *has all directional derivatives*

$$\frac{d}{dt} e^{A+tB} \bigg|_{t=0} = \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1} = \sum_{r,s=0}^{\infty} \frac{A^r B A^s}{(r+s+1)!}, \tag{2.3}$$

and the map $B \mapsto \frac{d}{dt} e^{A+tB}\big|_{t=0}$ is bounded. This map also is the Fréchet derivative $\frac{de^A}{dA}$.

Proof. By absolute convergence, we may rearrange:

$$e^{A+tB} = \sum_{n=0}^{\infty} \frac{(A+tB)^n}{n!} = \sum_{n=0}^{\infty} \frac{A^n}{n!} + t \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1} + O(t^2),$$

from which the first equality in the lemma is immediate. For the second, one uses that $(n, \ell) \mapsto (r, s) = (n - \ell, \ell - 1)$ is a bijection from $\{1 \le \ell \le n < \infty\}$ to $\mathbb{N}_0^2$. This proves the existence of directional derivatives. We have

$$\left\| \sum_{r,s=0}^{\infty} \frac{A^r B A^s}{(r+s+1)!} \right\| \le \|B\| \sum_{r,s=0}^{\infty} \frac{\|A\|^{r+s}}{(r+s+1)!} = \|B\| \sum_{n=0}^{\infty} (n+1) \frac{\|A\|^n}{(n+1)!} = e^{\|A\|} \|B\|,$$

thus $B \mapsto \frac{d}{dt} e^{A+tB}\big|_{t=0}$ is bounded. For the last claim one must show that

$$\lim_{B \to 0} \frac{\|e^{A+B} - e^A - \frac{de^A}{dA}(B)\|}{\|B\|} = 0,$$

which again is evident. (From the higher perspective of complex analysis in possibly infinite dimensional Banach spaces, the Fréchet differentiability is obvious since exp is an analytic function.) ■

On its own, the preceding result is not very useful. But combined with further algebraic arguments it will actually lead to one (the third) of the proofs of the following important result:

2.7 Proposition *Let $\mathcal{A}$ be a unital Banach algebra.*[7]

(i) If $A, B \in \mathcal{A}$ then

$$\frac{d}{dt} e^{A+tB} \bigg|_{t=0} = e^A \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)!} \operatorname{ad}_A^n(B) = e^A \frac{1 - e^{-\operatorname{ad}_A}}{\operatorname{ad}_A}(B). \tag{2.4}$$

(ii) Let I be some interval and $I \to \mathcal{A}, t \mapsto X(t)$ a differentiable map. Then

$$\frac{d}{dt} e^{X(t)} = e^{X(t)} \sum_{n=0}^{\infty} \frac{(-1)^n}{(n+1)!} \operatorname{ad}_{X(t)}^n \left(\frac{dX(t)}{dt} \right) = e^{X(t)} \frac{1 - e^{-\operatorname{ad}_{X(t)}}}{\operatorname{ad}_{X(t)}} \left(\frac{dX(t)}{dt} \right). \tag{2.5}$$

2.8 Remark 1. The statements are equivalent: (i) is just the special case $X(t) = A + tB$ of (ii), and (ii) is recovered from (i) by combination with the chain rule for Fréchet derivatives, see e.g. [9, Theorem 2.1.1]. (As noted above, exp : $\mathcal{A} \to \mathcal{A}$ is Fréchet differentiable, and for maps $\mathbb{R} \supset I \to \mathcal{A}$ the Fréchet derivative is just the ordinary one.)

2. If $[A, B] = 0$ then all summands $n \ne 0$ in (2.4) vanish and we recover (2.2).

[7] We will not consider the derivative of the exponential map in an abstract Lie group, which would take us too far from our main goals. See e.g. [64, Sect. 2.14].

3. Formal expressions like $\frac{1-e^{-X}}{X}$, which will we encounter frequently, simply mean that we replace z by X in the Taylor series expansion at 0 of a function like $f(z)=\frac{1-e^{-z}}{z}$ that is analytic in a neighborhood of 0. (Of course in a Banach algebra setting this means invoking Lemma 2.1 so that one needs to check that f is analytic on an open disc around zero of radius $R>\|X\|$.) With this convention, the respective second equalities in (2.4) and (2.5) are true by definition and only the first ones need to be proven.

4. The function $f(z)=\frac{1-e^{-z}}{z}$ appearing here is not particularly mysterious: It arises as $\int_0^1 e^{-zt}dt$. Its appearance is one (but not the only) reason why some approaches to the BCH series make contact with the complex of ideas involving difference and umbral calculus and the Bernoulli numbers. But this connection will only become manifest from Section 8 on.

5. Using (2.1), one deduces from (2.4) the equivalent dual formula

$$\frac{d}{dt}e^{A+tB}\bigg|_{t=0}=\mathrm{Ad}_{e^A}\left(\frac{1-e^{-\mathrm{ad}_A}}{\mathrm{ad}_A}(B)\right)e^A=e^{\mathrm{ad}_A}\left(\frac{1-e^{-\mathrm{ad}_A}}{\mathrm{ad}_A}(B)\right)e^A=\frac{e^{\mathrm{ad}_A}-1}{\mathrm{ad}_A}(B)e^A.$$

□

Since the proposition is quite important, we give three proofs!

First proof of Proposition 2.7. With the aim of proving version (ii), we follow Rossmann [50] defining

$$Y(s,t)=e^{-sX(t)}\frac{\partial}{\partial t}e^{sX(t)}.$$

Differentiating w.r.t. s gives

$$\begin{aligned}\frac{\partial Y}{\partial s}&=e^{-sX(t)}(-X(t))\frac{\partial}{\partial t}e^{sX(t)}+e^{-sX(t)}\frac{\partial}{\partial t}\left(X(t)e^{sX(t)}\right)\\&=e^{-sX(t)}(-X(t))\frac{\partial}{\partial t}e^{sX(t)}+e^{-sX(t)}\left(\frac{dX(t)}{dt}e^{sX(t)}+X(t)\frac{\partial}{\partial t}e^{sX(t)}\right)\\&=e^{-sX(t)}\frac{dX}{dt}e^{sX(t)}\quad=\quad\mathrm{Ad}_{e^{-sX(t)}}\left(\frac{dX}{dt}\right)\quad=\quad e^{-s\,\mathrm{ad}_{X(t)}}\left(\frac{dX}{dt}\right),\end{aligned}$$

where we used (2.1) in the last step. With $Y(0,t)=0$ we obtain by integration

$$e^{-X(t)}\frac{d}{dt}e^{X(t)}=Y(1,t)=\int_0^1\frac{\partial}{\partial s}Y(s,t)ds=\int_0^1 e^{-s\,\mathrm{ad}_{X(t)}}\left(\frac{dX}{dt}\right)ds=\left(\int_0^1 e^{-s\,\mathrm{ad}_{X(t)}}ds\right)\left(\frac{dX}{dt}\right),\qquad(2.6)$$

where the last integral is in the Banach algebra $\mathcal{B}(\mathcal{A})$ of bounded linear operators on $\mathcal{A}$. Now the following lemma, applied to $\mathcal{B}=\mathcal{B}(\mathcal{A})$ and $B=\mathrm{ad}_{X(t)}$, immediately gives the proposition in its general form (ii). ■

2.9 Lemma *Let $\mathcal{B}$ be a unital Banach algebra and $B\in\mathcal{B}$. Then*

$$\int_0^1 e^{-sB}ds=\sum_{n=0}^{\infty}\frac{(-B)^n}{(n+1)!}.$$

Proof. Expanding the exponential function, we have

$$\int_0^1 e^{-sB}ds=\int_0^1\sum_{n=0}^{\infty}\frac{(-sB)^n}{n!}ds=\sum_{n=0}^{\infty}\frac{(-B)^n}{n!}\int_0^1 s^n ds=\sum_{n=0}^{\infty}\frac{(-B)^n}{n!\,(n+1)}=\sum_{n=0}^{\infty}\frac{(-B)^n}{(n+1)!}.$$

Here the exchange of integration and summation is allowed due to the uniform norm convergence of the series on the bounded interval $[0,1]$. ■

2.10 Remark 1. From (2.6) one easily derives another frequently encountered, beautifully symmetric formula for the derivative of the exponential function, going back at least to Feynman's [22, (6)] (cf. also [58, (B5)]):

$$\frac{d}{dt}e^{X(t)} = e^{X(t)}\int_0^1 e^{-s\,\mathrm{ad}_{X(t)}}\left(\frac{dX}{dt}\right)ds = e^{X(t)}\int_0^1 \mathrm{Ad}_{e^{-sX(t)}}\left(\frac{dX}{dt}\right)ds = \int_0^1 e^{(1-s)X(t)}\left(\frac{dX}{dt}\right)e^{sX(t)}ds.$$

2. Expanding the exponential functions in the formula just derived into series and using the second identity of (9.12) to evaluate the integral, we again obtain the final expression of (2.3) (or its generalization where $A \leftarrow X(t)$, $B \leftarrow \frac{dX(t)}{dt}$). (Denominators like $(r+s+1)!$ coming from Euler's B-function evaluated at positive integers will reappear later.)

3. If $\mathcal{B}$ is finite dimensional, the operator valued integrals above can be defined coordinate-wise, but if $\mathcal{B}$ is infinite dimensional, they should be understood as a Banach-valued Riemann integrals (or as Bochner integrals). This complication is avoided in the next proof due to Tuynman [63], cf. also [32], which can be regarded as a discretization of the previous proof, the continuous auxiliary variable s being replaced by a discrete one. □

Second proof of Proposition 2.7. We will prove version (i) of the proposition. Define

$$\Delta(A,B) = \left.\frac{d}{dt}e^{A+tB}\right|_{t=0}.$$

We have $\Delta(A,B) = \frac{de^A}{dA}(B)$, thus $B \mapsto \Delta(A,B)$ is linear. (Cf. also Lemma 2.6.) Since the exponential function is infinitely differentiable, $\Delta(A,B)$ is jointly continuous in A,B. For each $m \in \mathbb{N}$ we have

$$e^{A+tB} = \left(\exp\left(\frac{A}{m} + t\frac{B}{m}\right)\right)^m,$$

and differentiating w.r.t. t at $t=0$ gives, using the product rule,

$$\begin{aligned}\left.\frac{d}{dt}e^{A+tB}\right|_{t=0} &= \sum_{k=0}^{m-1} e^{\frac{m-k-1}{m}A}\left.\frac{d}{dt}\exp\left(\frac{A}{m}+t\frac{B}{m}\right)\right|_{t=0} e^{\frac{k}{m}A} \\ &= e^{\frac{m-1}{m}A}\sum_{k=0}^{m-1} e^{-\frac{k}{m}A}\Delta\left(\frac{A}{m},\frac{B}{m}\right)e^{\frac{k}{m}A}.\end{aligned}$$

In view of (2.1) and the linearity of $\Delta(A,B)$ w.r.t. B, this equals

$$e^{\frac{m-1}{m}A}\left[\frac{1}{m}\sum_{k=0}^{m-1} e^{-\frac{k}{m}\mathrm{ad}_A}\right]\left(\Delta\left(\frac{A}{m},B\right)\right).$$

As $m \to \infty$, the factor on the left converges to e^A and $\Delta(A/m,B)$ converges to $\Delta(0,B)=B$ (both w.r.t. $\|\cdot\|$). Now the proof of (2.4) is completed by the following lemma, applied to $\mathcal{B}=\mathcal{B}(\mathcal{A})$ and $B=\mathrm{ad}_A$, which gives convergence of the linear operator in square brackets to $\frac{1-e^{-\mathrm{ad}_A}}{\mathrm{ad}_A}$. ■

2.11 Lemma *Let $\mathcal{B}$ be a unital Banach algebra and $B \in \mathcal{B}$. Then*

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1} e^{-B\frac{k}{m}} = \sum_{\ell=0}^{\infty}\frac{(-B)^\ell}{(\ell+1)!} = \frac{1-e^{-B}}{B}.$$

Proof. We have

$$\frac{1}{m}\sum_{k=0}^{m-1} e^{-B\frac{k}{m}} = \frac{1}{m}\sum_{k=0}^{m-1}\sum_{\ell=0}^{\infty}\frac{(-Bk/m)^\ell}{\ell!} = \sum_{\ell=0}^{\infty}\frac{(-B)^\ell}{\ell!}\frac{1}{m}\sum_{k=0}^{m-1}\left(\frac{k}{m}\right)^\ell. \tag{2.7}$$

Now $\frac{1}{m}\sum_{k=0}^{m-1}\left(\frac{k}{m}\right)^\ell$ is a Riemann sum approximation to the integral $\int_0^1 x^\ell dx$, so that

$$\lim_{m\to\infty}\frac{1}{m}\sum_{k=0}^{m-1}\left(\frac{k}{m}\right)^\ell = \int_0^1 x^\ell dx = \frac{1}{\ell+1}.$$

We may take the limit $m \to \infty$ in (2.7) within the ℓ-sum since $(k/m)^\ell \le 1$ for all ℓ, m and $0 \le k \le m$. (Cf. also Lemma D.5.) We obtain

$$\lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m-1} e^{-B\frac{k}{m}} = \sum_{\ell=0}^{\infty} \frac{(-B)^\ell}{\ell!} \frac{1}{\ell+1} = \sum_{\ell=0}^{\infty} \frac{(-B)^\ell}{(\ell+1)!} = \frac{1-e^{-B}}{B}.$$

■

Third proof of Proposition 2.7. Combine Lemma 2.6 with the following one. ■

2.12 Lemma *Let $\mathcal{A}$ be a unital Banach algebra and $A, B \in \mathcal{A}$. Then*

$$\sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1} = e^A \sum_{s=0}^{\infty} \frac{(-\mathrm{ad}_A)^s}{(s+1)!}(B). \tag{2.8}$$

Proof. We will first prove that the following holds for all $n \in \mathbb{N}$:

$$\sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1} = \sum_{k=1}^{n} \binom{n}{k} A^{n-k} \mathrm{ad}_{-A}^{k-1}(B). \tag{2.9}$$

For $n = 1$ this is just $B = B$. Assume (2.9) holds for n. Since $X_n := \sum_{\ell=1}^n A^{n-\ell} B A^{\ell-1}$ is the sum of all words $w \in \{A, B\}^n$ containing B exactly once, $X_n A$ is the sum of all $w \in \{A, B\}^{n+1}$ containing B exactly once, except for $A^n B$. Thus $X_{n+1} = A^n B + X_n A$, so that

$$\begin{aligned}
\sum_{\ell=1}^{n+1} A^{n+1-\ell} B A^{\ell-1} &= A^n B + \left(\sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1} \right) A \\
&= A^n B + \left(\sum_{k=1}^{n} \binom{n}{k} A^{n-k} \mathrm{ad}_{-A}^{k-1}(B) \right) A \\
&= A^n B + \sum_{k=1}^{n} \binom{n}{k} A^{n-k} \mathrm{ad}_{-A}^{k}(B) + \sum_{k=1}^{n} \binom{n}{k} A^{n+1-k} \mathrm{ad}_{-A}^{k-1}(B) \\
&= \sum_{k=1}^{n+1} \left[\binom{n}{k-1} + \binom{n}{k} \right] A^{n+1-k} \mathrm{ad}_{-A}^{k-1}(B) \\
&= \sum_{k=1}^{n+1} \binom{n+1}{k} A^{n+1-k} \mathrm{ad}_{-A}^{k-1}(B),
\end{aligned}$$

where we used $XA = \mathrm{ad}_{-A}(X) + AX$ with $X = \mathrm{ad}_{-A}^{k-1}(B)$ in the third step, $\binom{n}{n+1} = 0$ in the fourth and $\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$ for $1 \le k \le n+1$ in the fifth, completing the inductive proof of (2.9).

Multiplying (2.9) by $1/n!$ and summing over $n \in \mathbb{N}$ we have

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1} &= \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k=1}^{n} \binom{n}{k} A^{n-k} \mathrm{ad}_{-A}^{k-1}(B) \\
&= \sum_{1 \le k \le n < \infty} \frac{1}{k!\,(n-k)!} A^{n-k} \mathrm{ad}_{-A}^{k-1}(B) \\
&= \sum_{r=0}^{\infty} \frac{A^r}{r!} \sum_{s=0}^{\infty} \frac{\mathrm{ad}_{-A}^s}{(s+1)!}(B) \quad = \quad e^A \sum_{s=0}^{\infty} \frac{\mathrm{ad}_{-A}^s}{(s+1)!}(B),
\end{aligned}$$

where we again used the bijection $\{1 \le k \le n < \infty\} \to \mathbb{N}_0^2,\ (k, n) \mapsto (r, s) = (n-k, k-1)$. ■

2.13 Remark Except for the infinite summation over n, Lemma 2.12 is entirely algebraic, and indeed it will be used again to prove an algebraic version of Proposition 2.7, see Proposition 7.5. □

2.3 Proof(s) of BCHD

2.14 Theorem *Let $\mathcal{A}$ be a Banach algebra over $\mathbb{R}$ or $\mathbb{C}$. Let $X, Y \in \mathcal{A}$ such that $\|X\| + \|Y\| < \frac{\log 2}{2}$. Then Dynkin's Lie series (1.4) converges absolutely to $\log(e^X e^Y)$.*

Proof. Let $X, Y \in \mathcal{A}$ with $\|X\| + \|Y\| < \log 2$ so that (1.3) converges by Proposition 2.4, so that there is $H \in \mathcal{A}$ with $e^H = e^X e^Y$. Then clearly for all $t \in [0,1]$ there is $H(t) \in \mathcal{A}$ with $e^{H(t)} = e^{tX} e^{tY}$. Since the power series are infinitely differentiable on the domain $\|X\| + \|Y\| < \log 2$, the function $H(t)$ is differentiable in t. Differentiating both sides of $e^{H(t)} = e^{tX} e^{tY}$ and applying (2.5) on the l.h.s., we obtain

$$e^{H(t)} \frac{1 - e^{-\mathrm{ad}_{H(t)}}}{\mathrm{ad}_{H(t)}} \left(\frac{dH(t)}{dt} \right) = X e^{tX} e^{tY} + e^{tX} e^{tY} Y = X e^{H(t)} + e^{H(t)} Y.$$

Multiplying with $e^{-H(t)}$ on the right and using $\mathrm{Ad}_{e^{H(t)}} = e^{\mathrm{ad}_{H(t)}}$ (Eq. (2.1)) this becomes

$$\frac{e^{\mathrm{ad}_{H(t)}} - 1}{\mathrm{ad}_{H(t)}} \left(\frac{dH(t)}{dt} \right) = X + e^{\mathrm{ad}_{H(t)}}(Y).$$

(Thus we actually use the alternative formula from Remark 2.8.5.) For $\|X\|, \|Y\|$ and therefore $\|H(t)\|$ small enough, the operator $\frac{e^{\mathrm{ad}_{H(t)}}-1}{\mathrm{ad}_{H(t)}} = \mathrm{id} + O(\mathrm{ad}_{H(t)})$ is invertible, thus

$$\frac{dH(t)}{dt} = \frac{\mathrm{ad}_{H(t)}}{e^{\mathrm{ad}_{H(t)}} - 1} (X + e^{\mathrm{ad}_{H(t)}}(Y)). \tag{2.10}$$

This differential equation looks nasty, but $H(t)$ can be eliminated from the r.h.s. From (1.2) we have $A = \log((e^A - 1) + 1) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (e^A - 1)^k$, convergent (to the correct value) if $\|A\| < \log 2$, and thus

$$\frac{A}{e^A - 1} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (e^A - 1)^{k-1}.$$

This can be applied to $A = \mathrm{ad}_{H(t)}$ since A only appears as e^A on the r.h.s. and since (2.1) gives

$$e^{\mathrm{ad}_{H(t)}} = \mathrm{Ad}_{e^{H(t)}} = \mathrm{Ad}_{e^{tX}} \mathrm{Ad}_{e^{tY}} = e^{\mathrm{ad}_{tX}} e^{\mathrm{ad}_{tY}}.$$

Noting also that $e^{\mathrm{ad}_{tY}}(Y) = Y$, we obtain

$$\frac{dH(t)}{dt} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (e^{\mathrm{ad}_{tX}} e^{\mathrm{ad}_{tY}} - 1)^{k-1} (X + e^{\mathrm{ad}_{tX}}(Y)), \tag{2.11}$$

which has no $H(t)$ on the r.h.s. Before we continue, we discuss the convergence of the series. We have

$$\begin{aligned} \|e^{\mathrm{ad}_{tX}} e^{\mathrm{ad}_{tY}} - 1\| &= \left\| \sum_{m+n>0} \frac{\mathrm{ad}_{tX}^m \mathrm{ad}_{tY}^n}{m!\, n!} \right\| \leq \sum_{m+n>0} \frac{\|\mathrm{ad}_{tX}\|^m \|\mathrm{ad}_{tY}\|^n}{m!\, n!} \\ &= e^{\|\mathrm{ad}_{tX}\| + \|\mathrm{ad}_{tY}\|} - 1 \leq e^{2t(\|X\|+\|Y\|)} - 1, \end{aligned}$$

so that the series in (2.11) converges for all $\|X\| + \|Y\| < \frac{\log 2}{2}$, uniformly in $t \in [0,1]$. Integrating (2.11) over $[0,1]$ and using $H(0) = 0$, we have

$$\log(e^X e^Y) = H(1) = \int_0^1 \left(\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (e^{\mathrm{ad}_{tX}} e^{\mathrm{ad}_{tY}} - 1)^{k-1} (X + e^{\mathrm{ad}_{tX}}(Y)) \right) dt.$$

This already proves the BCH theorem, since the r.h.s. is Lie. To obtain Dynkin's formula, we expand

$$\begin{aligned} \log(e^X e^Y) = & \int_0^1 \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum_{m_1+n_1>0} \cdots \sum_{m_{k-1}+n_{k-1}>0} \\ & \left(\frac{\mathrm{ad}_{tX}^{m_1} \mathrm{ad}_{tY}^{n_1} \cdots \mathrm{ad}_{tX}^{m_{k-1}} \mathrm{ad}_{tY}^{n_{k-1}}(X)}{m_1!\, n_1! \cdots m_{k-1}!\, n_{k-1}!} + \sum_{m_k \geq 0} \frac{\mathrm{ad}_{tX}^{m_1} \mathrm{ad}_{tY}^{n_1} \cdots \mathrm{ad}_{tX}^{m_{k-1}} \mathrm{ad}_{tY}^{n_{k-1}} \mathrm{ad}_{tX}^{m_k}(Y)}{m_1!\, n_1! \cdots m_{k-1}!\, n_{k-1}!\, m_k!} \right) dt. \end{aligned}$$

Using $\mathrm{ad}_{tX}^m = t^m \mathrm{ad}_X^m$ and $\int_0^1 t^n dt = \frac{1}{n+1}$, evaluating the integral simply amounts to counting powers of t, giving exactly formula (1.5), which we have seen to be equivalent to the standard Dynkin formula (1.4). ■

2.15 Remark 1. The proof given above followed [50] (with a few more details), since this argument gives exactly Dynkin's formula (1.4) as opposed to the variations, discussed next, proven by most authors.

2. Other analytic approaches to BCH work with $e^{H(t)} = e^X e^{tY}$ (Hausner/Schwartz [34], Hilgert/Neeb [36], Hall [32]) or with $e^{H(t)} = e^{tX} e^Y$ (Duistermaat/Kolk [17]) instead of $e^{H(t)} = e^{tX} e^{tY}$. We quickly discuss the first, the second being completely analogous. Differentiating $e^{H(t)} = e^X e^{tY}$ and using (2.5) gives

$$e^{H(t)} \frac{1 - e^{-\mathrm{ad}_{H(t)}}}{\mathrm{ad}_{H(t)}} \frac{dH(t)}{dt} = \frac{de^{H(t)}}{dt} = e^X e^{tY} Y = e^{H(t)} Y,$$

from which an argument analogous to the above (but using $H(0) = X$) leads to the integral representation

$$\log(e^X e^Y) = X + \int_0^1 \psi(e^{\mathrm{ad}_X} e^{t\,\mathrm{ad}_Y})(Y) dt, \tag{2.12}$$

where $\psi(t) = \frac{t \log t}{t-1} = t \sum_{n=1}^\infty (-1)^{n-1} \frac{(t-1)^{n-1}}{n}$. While the resulting series expansion

$$\log(e^X e^Y) = X + \sum_{k=1}^\infty \sum_{m_1+n_1>0} \cdots \sum_{m_{k-1}+n_{k-1}>0} \sum_{m_k \geq 0} \frac{(-1)^{k-1}}{k(n_1 + \cdots + n_{k-1} + 1)} \frac{\mathrm{ad}_X^{m_1} \mathrm{ad}_Y^{n_1} \cdots \mathrm{ad}_X^{m_{k-1}} \mathrm{ad}_Y^{n_{k-1}} \mathrm{ad}_X^{m_k}(Y)}{m_1!\, n_1! \cdots m_{k-1}!\, n_{k-1}!\, m_k!} \tag{2.13}$$

is simpler than Dynkin's (1.4),(1.5), it is less symmetric in that $\sum_i n_i$ appears in the denominator instead of $\sum_i (m_i + n_i)$. Another reason to prefer (1.4) is that the algebraic proof of Section 6 naturally leads to (1.4) rather than (2.13). (It is not evident (to this author) how to prove that the right hand sides of (1.4) and of (2.13) are equal without invoking that both are equal to $\log(e^X e^Y)$.)

3. Proofs of the above type have a long history. A version of Proposition 2.7 can be found in the 1891 paper [55] of F. Schur, and integral representations like (2.12) in works of H. Poincaré [45] around 1900. (But it can be difficult to recognize the results in these papers.) One of the earliest books where the results can be found in the above form is [34]. Later ones like [29] (the most thorough one as far as BCHD is concerned, with a computation of the Dynkin series to order 4) and [36, 17, 50, 32] only offer minor variations. Since Dynkin's series (or equivalent ones) can easily be deduced by expansion of integral formulae like (2.12), it seems puzzling that an explicit expansion was obtained only in 1947, by a rather different algebraic approach, discussed in Section 6.

4. In Section 7 we will see that the above proof is not as manifestly analytic as it looks! Essentially the only challenge is to give an algebraic version of Proposition 2.7, whose proof will be similar to the third proof of the latter. And in Section 8 we will give a different proof of the BCH theorem based on (the algebraic version) of Proposition 2.7, much closer (at least in spirit) to the original works [4, 8, 33]. But first we will discuss two algebraic approaches to BCH(D) that have very little in common with what we have done so far. □

3 Basics of non-commutative (Lie) polynomials and series

In this section we will introduce only the modest preliminaries needed for Eichler's algebraic proof of the BCH theorem. Further material on non-commutative polynomials and series will be given in Section 5. **From now on, $\mathbb{F}$ is an arbitrary field of characteristic zero.**

- If A is a set (the *alphabet*), the free monoid A^* generated by A is the set of finite sequences (*words*) with elements in A, with concatenation as (associative) composition and the empty (length zero) sequence ϵ as unit. The length of a word w is denoted $|w|$. Clearly $|w_1 w_2| = |w_1| + |w_2|$. More formally: $A^* = \{(n, f) \mid n \in \mathbb{N}_0, f : \{1, \ldots, n\} \to A\}$, with $|(n, f)| = n$ and $(n, f)(n', f') = (n + n', \{1, \ldots, n + n'\} \mapsto \{f(1), \ldots, f(n), f'(1), \ldots, f'(n')\})$.
  If $A = \emptyset$, one finds $A^* = \{\epsilon\}$. If $|A| = 1$, the map $\{1, \ldots, n\} \to A$ in the above definition is constant, thus carries no information and can be omitted. This shows that $A^* \cong (\mathbb{N}_0, +, 0)$. Note that this does not give an independent construction of $\mathbb{N}_0$ since $\mathbb{N}_0$ was used in our definition of A^*.

- The free monoid has a *universal property*: For every function $A \to M$, where $(M, \cdot, 1)$ is a monoid, there is a unique unital monoid homomorphism $\widehat{f} : A^* \to M$ such that $\widehat{f} \circ \iota = f$, where $\iota : A \hookrightarrow A^*$ is the inclusion map sending letters to one-letter words. Clearly $\widehat{f}$ must send ϵ to 1 and $a_1 \dots a_n$ to $f(a_1) \dots f(a_n)$. This universal property determines A^* up to isomorphism.
- Aside: Applying the above to $M = B^*$, we see that if A, B are sets and $f : A \to B$ a function, one has a unique monoid homomorphism $f^* : A^* \to B^*$ such that $\epsilon \mapsto \epsilon$ and $a_1 \dots a_n \mapsto f(a_1) \dots f(a_n)$. One easily checks that $G(A) = A^*$, $G(f) = f^*$ defines a functor $G : \mathrm{Set} \to \mathrm{Mon}$, the free monoid functor. Now the universal property satisfied by A^* is essentially equivalent to saying that G is a left adjoint of the obvious forgetful functor $F : \mathrm{Mon} \to \mathrm{Set}$.

  If one wants to define A^* without reference to $\mathbb{N}_0$, one may either define it (up to isomorphism) by its universal property or as $G(A)$, where G is a left adjoint of the forgetful functor $F : \mathrm{Mon} \to \mathrm{Set}$. But either way, the proof of existence becomes non-trivial. We won't pursue this.
- If M is a monoid and $\mathbb{F}$ is a field, $\mathbb{F}M$ is the monoid algebra. It consists of the finitely supported functions $M \to \mathbb{F}$ with the obvious structure of $\mathbb{F}$-vector space. Writing an element f of $\mathbb{F}M$ symbolically as $\sum_{m \in M} a_m m$, multiplication is given by $(\sum_{m \in M} a_m m)(\sum_{m' \in M} b_{m'} m') = \sum_{m,m'} a_m b_{m'} mm'$. The function $h : M \to \mathbb{F}$ corresponding to this element is given by the convolution

  $$h_m = \sum_{\substack{m',m'' \in M \\ m'm''=m}} a_{m'} b_{m''}. \tag{3.1}$$

  This is well-defined due to the finite supports of a and b.
- If A is a set and $\mathbb{F}$ a field, the monoid algebra $\mathbb{F}A^*$ for the free monoid A^* is called the free (non-commutative) $\mathbb{F}$-algebra generated by A, denoted $\mathbb{F}\langle A \rangle$. (NB: This is not to be confused with the free commutative $\mathbb{F}$-algebra over A, which is the polynomial ring $\mathbb{F}[A]$.) It has a universal property very similar to the one of the free monoid. (Just replace 'monoid homomorphism' by '$\mathbb{F}$-algebra homomorphism'.) When considered in $\mathbb{F}\langle A \rangle$ we write ϵ as 1 since it is the multiplicative unit.
- Aside: If A is a set and $\mathbb{F}$ a field, we can define the $\mathbb{F}$-vector space $\mathbb{F}A$ generated by A as the set of finitely supported functions $A \to \mathbb{F}$. And if V is a $\mathbb{F}$-vector space, the tensor algebra TV is defined as $\oplus_{n=0}^{\infty} V^{\otimes n}$ with the obvious product operation. Now one can prove a natural (in A) isomorphism $T(\mathbb{F}A) \cong \mathbb{F}(A^*)$. (The two sides correspond to the two possible orders of freely introducing an associative multiplication and a $\mathbb{F}$-vector space structure.) We omit the details since the free algebra formulation is more natural when V comes with a basis.
- If A is a set and $\mathbb{F}$ a field, the algebra $\mathbb{F}\langle\langle A \rangle\rangle$ of formal non-commutative power series over A consists of the set of all functions $A^* \to \mathbb{F}$ with the obvious structure of $\mathbb{F}$-vector space and multiplication given by convolution as in (3.1), with $M = A^*$. Despite the possibly infinite supports of such functions, this is well defined since every word $w \in A^*$ can be factored as $w = w'w''$ only in finitely many ways. (For this reason, the more general 'algebra of formal series' over a monoid M can only be defined if M satisfies such a finite factorization condition. We will not need this generalization.) (Recall that $\mathbb{F}[[A]]$ denotes the usual $\mathbb{F}$-algebra of commutative formal power series over A.) If $A = \{a_1, \dots, a_n\}$, we write $\mathbb{F}\langle\langle a_1, \dots, a_n \rangle\rangle$ instead of $\mathbb{F}\langle\langle \{a_1, \dots, a_n\} \rangle\rangle$. Instead of $f : A^* \to \mathbb{F}$ we write $F = \sum_{w \in A^*} a_w w$ and put $(F, w) = a_w \ \forall w$.

  We will usually denote the formal variables from A by lower case letters and elements of $\mathbb{F}\langle A \rangle$ or $\mathbb{F}\langle\langle A \rangle\rangle$ by capital letters. (This is mildly inconsistent since $A \hookrightarrow \mathbb{F}\langle A \rangle$, but this will not lead to problems.)
- The algebras $\mathbb{F}\langle A \rangle$ and $\mathbb{F}\langle\langle A \rangle\rangle$ have obvious gradations by $\mathbb{N}_0$, where $\mathbb{F}\langle\langle A \rangle\rangle_n$ consists of linear combinations (typically infinite ones in the case of $\mathbb{F}\langle\langle A \rangle\rangle$) of words of length n. For $F = \sum_w a_w w$, we define $F_n = \sum_{w \in A^*, |w|=n} a_w w \in \mathbb{F}\langle\langle A \rangle\rangle_n$.

  Our sets A will always be finite. This implies $\mathbb{F}\langle\langle A \rangle\rangle_n = \mathbb{F}\langle A \rangle_n$ for all $n \in \mathbb{N}_0$.
- We denote by $\mathbb{F}\langle A \rangle_{>0}, \mathbb{F}\langle\langle A \rangle\rangle_{>0}$ the ideals of non-commutative polynomials and series, respectively, with constant term (F, ϵ) equal to zero. We use analogous notations $\mathbb{F}\langle\langle A \rangle\rangle_{\geq k}$, etc., which should explain themselves.
- Given formal series $F = \sum_{n=0}^{\infty} a_n x^n, G = \sum_{n=0}^{\infty} b_n x^n \in \mathbb{F}\langle\langle x \rangle\rangle$, the composite $G \circ F$ is given by

  $$G \circ F = \sum_{n=0}^{\infty} b_n \left(\sum_{m=0}^{\infty} a_m x^m \right)^n = \sum_{k=0}^{\infty} c_k x^k,$$

where

$$c_k = \sum_{\ell=0}^{\infty} b_\ell \sum_{\substack{m_1,\dots,m_\ell \\ m_1+\dots+m_\ell=k}} a_{m_1} \cdots a_{m_\ell}. \tag{3.2}$$

If G is a polynomial or $a_0 = 0$ then each c_k is given by a polynomial in finitely many $b_0, b_1, b_2 \dots$ so that $G \circ F$ is well-defined as an element of $\mathbb{F}\langle\langle x \rangle\rangle$. (Otherwise we have infinite sums like $c_0 = \sum_{\ell=0}^{\infty} b_\ell a_0^\ell$ in $\mathbb{F}$, of which we cannot make sense without a topology on $\mathbb{F}$. A simple example is $F = G = \sum_{n=0}^{\infty} x^n$.) In the more important case $a_0 = 0$, thus $F \in \mathbb{F}\langle\langle x \rangle\rangle_{>0}$, the index ℓ in (3.2) effectively only runs up to k. The above reasoning generalizes to formal series in several 'variables': If $G \in \mathbb{F}\langle\langle x_1, \dots, x_n \rangle\rangle$ and $F_1, \dots, F_n \in \mathbb{F}\langle\langle y_1, \dots, y_m \rangle\rangle_{>0}$ then $G(F_1, \dots, F_n)$ makes sense in $\mathbb{F}\langle\langle y_1, \dots, y_m \rangle\rangle$.

- In particular if $F \in \mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle_{>0}$ then $e^F \equiv \exp(F) \in \mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle$ exists and has constant term 1. And if $F \in 1 + \mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle_{>0}$ then $\log(F)$, where log is as in (1.2), defines an element in $\mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle_{>0}$. (Note that the series defining log has no constant term.) In fact, these maps

$$\exp : \mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle_{>0} \to 1 + \mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle_{>0},$$

$$\log : 1 + \mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle_{>0} \to \mathbb{F}\langle\langle x_1, \dots, x_m \rangle\rangle_{>0}$$

are inverses of each other. To see this, note that the usual functions $\log : (0, 2) \to (-\infty, \log 2)$ and $\exp : (-\infty, \log 2) \to (0, 2)$ are inverses of each other. On these domains, the power series for log and exp converge, so that

$$x = \exp(\log(x)) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_{m=1}^{\infty} \frac{(-1)^{m-1}(x-1)^m}{m} \right)^n \quad \forall x \in (0, 2),$$

$$x = \log(\exp(x)) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} \left(\sum_{n=0}^{\infty} \frac{x^n}{n!} \right)^m \quad \forall x \in (-\infty, \log 2).$$

But this implies that $\exp \circ \log$ and $\log \circ \exp$ are the identity (formal) power series. (A purely algebraic proof of the latter statement is given in Appendix C.)

- If A is an associative algebra, we define a map $[\cdot, \cdot] : A \times A \to A$ by $[x, y] = xy - yx$. One easily verifies that $(A, [\cdot, \cdot])$ is a Lie algebra, i.e. $[\cdot, \cdot]$ is bilinear, antisymmetric ($[y, x] = -[x, y]\ \forall x, y$) and satisfies the Jacobi identity $[x, [y, z]] + [y, [z, x]] + [z, [x, y]] = 0\ \forall x, y, z$.

- If A is a set and $\mathbb{F}$ a field, the intersection of all Lie subalgebras of $(\mathbb{F}\langle A \rangle, [\cdot, \cdot])$ that contain A is a Lie algebra, called the free Lie algebra $\mathcal{L}_{\mathbb{F}}\langle A \rangle$ (over $\mathbb{F}$). Elements of $\mathcal{L}_{\mathbb{F}}\langle A \rangle$ are called Lie polynomials. Note that Lie polynomials have no constant term.

3.1 Lemma *(i) $\mathcal{L}_{\mathbb{F}}\langle A \rangle$ consists of all finite $\mathbb{F}$-linear combinations of elements of A and elements obtained using commutators of finitely elements.*
(ii) If $F \in \mathcal{L}_{\mathbb{F}}\langle A \rangle$ then each homogeneous component F_n is Lie.

Proof. (i) Let $L \subset \mathbb{F}\langle A \rangle$ consist of the finite linear combinations of elements of A and of (correctly bracketed) commutator expressions of elements of A. (E.g. $[[a_1, [a_2, a_3]], a_4]$.) Since $\mathcal{L}_{\mathbb{F}}\langle A \rangle$ is a Lie algebra containing A, we have $L \subseteq \mathcal{L}_{\mathbb{F}}\langle A \rangle$. If $X_1, X_2 \in L$ then $[X_1, X_2] \in L$. Thus L is closed under $[\cdot, \cdot]$ and therefore a Lie algebra. Now the definition of $\mathcal{L}_{\mathbb{F}}\langle A \rangle$ implies $L = \mathcal{L}_{\mathbb{F}}\langle A \rangle$, which was our claim.

(ii) By (i), every $F \in \mathcal{L}_{\mathbb{F}}\langle A \rangle$ is a finite linear combination of commutator expressions involving elements of A. Since each of these is homogeneous and Lie, it follows that F_n is Lie for each n. ■

- If $G \in \mathcal{L}_{\mathbb{F}}\langle\langle x_1, \dots, x_n \rangle\rangle$ and $F_1, \dots, F_n \in \mathcal{L}_{\mathbb{F}}\langle\langle y_1, \dots, y_m \rangle\rangle$ then clearly also the composite $G(F_1, \dots, F_n)$ (see above) is a Lie polynomial.

- Let $\mathbb{F}$ be a field and A a finite set. A formal (non-commutative) series $\sum_{w \in A^*} a_w w \in \mathbb{F}\langle\langle A \rangle\rangle$ is called a Lie series if its constant term vanishes and each homogeneous component $\sum_{|w|=n} a_w w$ is a Lie polynomial, i.e. lies in $\mathcal{L}_{\mathbb{F}}\langle A \rangle$. The set of Lie series is denoted $\mathcal{L}_{\mathbb{F}}\langle\langle A \rangle\rangle$. (The definition of Lie series is slightly more involved if A is infinite, cf. [49]. We will not need this.) Lemma 3.1(ii) implies that $\mathcal{L}_{\mathbb{F}}\langle A \rangle \subset \mathcal{L}_{\mathbb{F}}\langle\langle A \rangle\rangle$.

4 Eichler's algebraic proof of BCH

In this section we will give M. Eichler's[8] [20] little referenced (but see [52, 61, 6]) proof of the following theorem:

4.1 Theorem (BCH) *The formal series for* $\log(e^x e^y) \in \mathbb{F}\langle\langle x, y\rangle\rangle$ *is a Lie series.*

Proof. In what follows, we work in $\mathbb{F}\langle\langle A\rangle\rangle$ throughout, where $A = \{x, y, z\}$. By the observations of the preceding section, $H = \log(e^x e^y)$ is a well-defined element of $\mathbb{F}\langle\langle A\rangle\rangle$ with vanishing constant term and satisfying $e^H = e^x e^y$. Using the $\mathbb{N}_0$-gradation of $\mathbb{F}\langle\langle A\rangle\rangle$ we write $H = \sum_{k=1}^{\infty} H_k$, where $H_k \in \mathbb{F}\langle A\rangle_k$ for each k. Reading (1.4) as a computation in $\mathbb{F}\langle\langle A\rangle\rangle$, we have $H_1(x, y) = x + y$ and $H_2(x, y) = \frac{1}{2}[x, y]$, thus H_1 and H_2 are Lie polynomials. We will prove by induction that each $H_n,\ n \geq 3$ is a Lie polynomial, so that H is a Lie series.

If $X, Y \in \mathbb{F}\langle\langle x, y\rangle\rangle_{>0}$, we obtain composite formal series $H_n(X, Y)$. Replacing $x, y \rightsquigarrow X, Y$ in the identity $e^H = e^x e^y$ shows that the (unique) solution of $e^{H(X,Y)} = e^X e^Y$ is given by $H(X, Y) = \sum_n H_n(X, Y)$. Furthermore:

1. Each H_n is homogeneous of order n, thus $H_n(rX, rY) = r^n H_n(X, Y)$ for all $r \in \mathbb{F}$ for all $n \geq 1$.
2. If $XY = YX$ then $e^X e^Y = e^{X+Y}$, thus $H(X, Y) = \log(e^X e^Y) = X + Y$. Thus $H(X, Y)$ is homogeneous of degree one, implying $H_k(X, Y) = 0$ for all $k \geq 2$. In particular $H_k(rX, sX) = 0$ for all $r, s \in \mathbb{F}$ and $k \geq 2$.
3. Let $X, Y, Z \in \mathbb{F}\langle\langle x, y, z\rangle\rangle_{>0}$. Then the associativity $(e^X e^Y)e^Z = e^X(e^Y e^Z)$ leads to

$$H(H(X, Y), Z) = \log(e^{H(X,Y)} e^Z) = \log(e^X e^Y e^Z) = \log(e^X e^{H(Y,Z)}) = H(X, H(Y, Z)).$$

Expanding the formula $H(H(X, Y), Z) = H(X, H(Y, Z))$ using $H = \sum_{k=1}^{\infty} H_k$ and $H_1(X, Y) = X + Y$, we obtain

$$\begin{aligned} X + Y + Z + \sum_{m=2}^{\infty} H_m(X, Y) + \sum_{\ell=2}^{\infty} H_\ell\left(X + Y + \sum_{m=2}^{\infty} H_m(X, Y), Z\right) \\ = X + Y + Z + \sum_{m=2}^{\infty} H_m(Y, Z) + \sum_{\ell=2}^{\infty} H_\ell\left(X, Y + Z + \sum_{m=2}^{\infty} H_m(Y, Z)\right). \end{aligned} \tag{4.1}$$

While this identity holds for all $X, Y, Z \in \mathbb{F}\langle\langle x, y, z\rangle\rangle_{>0}$, we now restrict it to $X, Y, Z \in V$, where $V = \mathbb{F}\langle x, y, z\rangle_1 = \mathrm{span}_{\mathbb{F}}\{x, y, z\}$, the motivation being that V is a vector space containing $\{x, y, z\}$ and all its elements are Lie.

Let now $n \geq 3$ and assume that $H_m \in \mathbb{F}\langle x, y\rangle$ is a Lie polynomial for each $m < n$. Recall that if K, L, M are Lie polynomials, then so is the composite $K(L, M)$. Thus if $X, Y, Z \in V$ then

$$X + Y + Z + \sum_{m=2}^{n-1} H_m(X, Y) \quad \text{and} \quad X + Y + Z + \sum_{m=2}^{n-1} H_m(Y, Z)$$

are Lie. Since all contributions of degree $\leq n$ to

$$\sum_{\ell=2}^{\infty} H_\ell\left(X + Y + \sum_{m=2}^{\infty} H_m(X, Y), Z\right) \quad \text{and} \quad \sum_{\ell=2}^{\infty} H_\ell\left(X, Y + Z + \sum_{m=2}^{\infty} H_m(Y, Z)\right)$$

involving both H_ℓ and H_m have $\ell, m < n$, they are Lie. Thus the projection of (4.1) to $\mathbb{F}\langle x, y, z\rangle_n$ is

$$H_n(X, Y) + H_n(X + Y, Z) + \mathrm{LIE} = H_n(Y, Z) + H_n(X, Y + Z) + \mathrm{LIE}. \tag{4.2}$$

Put $W = \mathbb{F}\langle x, y, z\rangle/\mathcal{L}_{\mathbb{F}}\langle x, y, z\rangle$ with quotient map $\phi : \mathbb{F}\langle x, y, z\rangle \twoheadrightarrow W$, and define $C : V \times V \to W$ as the map

$$C : V \times V \to W, \quad (K, L) \mapsto \phi(H_n(K, L)).$$

Then (4.2) is equivalent to the 2-cocycle equation

$$C(X, Y) + C(X + Y, Z) = C(X, Y + Z) + C(Y, Z) \qquad \forall X, Y, Z \in V.$$

[8] Martin Eichler (1912-1992), German mathematician who mostly worked on number theory.

The properties 1. and 2. of H_n stated above pass to C. Now the following proposition gives $C = 0$. This means that $H_n(K, L)$ is a Lie polynomial for all $K, L \in V$, in particular for $K = x$, $L = y$. This completes the induction. ■

4.2 Proposition [9] *Let $\mathbb{F}$ be a field of characteristic zero and V, W be $\mathbb{F}$-vector spaces. Let $C : V \times V \to W$ satisfy*

(i) 2-Cocycle identity: $C(X,Y) + C(X+Y,Z) = C(X,Y+Z) + C(Y,Z) \quad \forall X,Y,Z \in V$.

(ii) Homogeneity: There is an $3 \leq n \in \mathbb{N}$ such that $C(rX, rY) = r^n C(X,Y) \quad \forall X, Y \in V$, $r \in \mathbb{F}$.

(iii) $C(rX, sX) = 0 \quad \forall X \in V$, $r, s \in \mathbb{F}$. In particular $C(X,0) = C(0,X) = 0$.

Then $C \equiv 0$.

Proof. Putting $Z = -Y$ in (i) we have $C(X,Y) + C(X+Y,-Y) = C(X,0) + C(Y,-Y)$, thus using (iii):

$$C(X,Y) = -C(X+Y,-Y). \tag{4.3}$$

Similarly with $X = -Y$ in (i) and using (iii) we get $0 = C(-Y, Y+Z) + C(Y,Z)$. Replacing Y, Z by X, Y:

$$C(X,Y) = -C(-X, X+Y). \tag{4.4}$$

Now,

$$C(X,Y) \stackrel{(4.4)}{=} -C(-X,X+Y) \stackrel{(4.3)}{=} C(Y,-X-Y) \stackrel{(4.4)}{=} -C(-Y,-X) \stackrel{(ii)}{=} (-1)^{n+1}C(Y,X). \tag{4.5}$$

Putting $Z = -Y/2$ in (i) gives $C(X,Y) + C(X+Y,-Y/2) = C(X,Y-Y/2) + C(Y,-Y/2)$, thus with (iii):

$$C(X,Y) = C(X,Y/2) - C(X+Y,-Y/2). \tag{4.6}$$

With $X = -Y/2$ in (i) we have $C(-Y/2,Y) + C(-Y/2+Y,Z) = C(-Y/2,Y+Z) + C(Y,Z)$, thus with $Y, Z \to X, Y$:

$$C(X,Y) = C(X/2,Y) - C(-X/2,X+Y). \tag{4.7}$$

Applying (4.6) to both terms on the r.h.s. of (4.7) gives

$$\begin{aligned}
C(X/2,Y) &= C(X/2,Y/2) - C(X/2+Y,-Y/2) \\
&\stackrel{(4.3)}{=} C(X/2,Y/2) + C(X/2+Y/2,Y/2), \\
C(-X/2,X+Y) &= C(-X/2,X/2+Y/2) - C(X/2+Y,-X/2-Y/2) \\
&\stackrel{(4.3)}{=} C(-X/2,X/2+Y/2) + C(Y/2,X/2+Y/2) \\
&\stackrel{(4.4)}{=} -C(X/2,Y/2) + C(Y/2,X/2+Y/2),
\end{aligned}$$

thus

$$\begin{aligned}
C(X,Y) &= C(X/2,Y/2) + C(X/2+Y/2,Y/2) + C(X/2,Y/2) - C(Y/2,X/2+Y/2) \\
&= 2C(X/2,Y/2) + C(X/2+Y/2,Y/2) - C(Y/2,X/2+Y/2) \\
&\stackrel{(ii)}{=} 2^{1-n}C(X,Y) + 2^{-n}C(X+Y,Y) - 2^{-n}C(Y,X+Y) \\
&\stackrel{(4.5)}{=} 2^{1-n}C(X,Y) + 2^{-n}(1+(-1)^n)C(X+Y,Y).
\end{aligned}$$

Collecting, we have

$$(1 - 2^{1-n})C(X,Y) = 2^{-n}(1+(-1)^n)C(X+Y,Y), \tag{4.8}$$

which already implies $C(X,Y) = 0$ if n is odd (since $n \neq 1$). If n is even, replacing X in (4.8) by $X - Y$ we have

$$-(1-2^{1-n})C(X,-Y) \stackrel{(4.3)}{=} (1-2^{1-n})C(X-Y,Y) = 2^{1-n}C(X,Y).$$

[9] This way of stating the proposition is inspired by Eichler's *Autoreferat* in Zentralblatt (Zbl 0157.07601).

Dividing by 2^{1-n} we obtain

$$C(X,Y) = (1-2^{n-1})C(X,-Y),$$

twofold application of which yields

$$C(X,Y) = (1-2^{n-1})^2\, C(X,Y).$$

In view of $n \neq 2$, the numerical factor on the r.h.s. is different from one, so that $C(X,Y) = 0\ \forall X,Y \in V$. ■

4.3 Remark 1. For almost all purposes in Lie theory, the qualitative BCH Theorem 4.1 (plus an easy convergence result, see Remark 2.5) is sufficient. If one only needs this, Eichler's proof probably is the shortest, at least of the algebraic ones.

2. The above proof based on associativity and a resulting 2-cocycle equation is quite interesting and very different from all the others, at least on the surface. It would be good to understand how it relates to the other proofs. □

5 $\mathbb{F}\langle A\rangle$ and $\mathbb{F}\langle\langle A\rangle\rangle$ as Hopf algebras

In the interest of a coherent picture, in this section we give a few more facts than strictly needed later.

5.1 Definition *Let A be a set and $\mathbb{F}$ a field. Denoting by $\otimes$ the algebraic tensor product of $\mathbb{F}$-modules, define*

- *a $\mathbb{F}$-linear map* $\mathrm{conc} : \mathbb{F}\langle A\rangle \otimes \mathbb{F}\langle A\rangle \to \mathbb{F}\langle A\rangle,\ P \otimes Q \mapsto PQ$.
- *a $\mathbb{F}$-algebra homomorphism $\delta : \mathbb{F}\langle A\rangle \to \mathbb{F}\langle A\rangle \otimes \mathbb{F}\langle A\rangle$, uniquely determined by*

$$\delta : 1 \mapsto 1 \otimes 1, \quad a \mapsto a \otimes 1 + 1 \otimes a \qquad \forall a \in A.$$

- *an $\mathbb{F}$-algebra homomorphism $\varepsilon : \mathbb{F}\langle A\rangle \to \mathbb{F},\ \varepsilon(F) = (F,1)$.*
- *a linear map $\alpha : \mathbb{F}\langle A\rangle \to \mathbb{F}\langle A\rangle$ by $\alpha(1) = 1$ and $\alpha(a_1 \cdots a_n) = (-1)^n a_n \cdots a_1$ for $a_1, \ldots, a_n \in A$.*

5.2 Proposition *With the above definitions,*

(i) $(\mathbb{F}\langle A\rangle, \mathrm{conc}, 1)$ *is an (associative) algebra.*

(ii) $(\mathbb{F}\langle A\rangle, \delta, \varepsilon)$ *is a coalgebra, i.e. coassociative* $(\delta\otimes\mathrm{id})\circ\delta = (\mathrm{id}\otimes\delta)\circ\delta$ *with ε as counit* $(\varepsilon\otimes\mathrm{id})\circ\delta = \mathrm{id} = (\mathrm{id}\otimes\varepsilon)\circ\delta$.

(iii) $(\mathbb{F}\langle A\rangle, \mathrm{conc}, 1, \delta, \varepsilon)$ *is a bialgebra, i.e. algebra, coalgebra, and δ, ε are algebra homomorphisms.*

(iv) $(\mathbb{F}\langle A\rangle, \mathrm{conc}, 1, \delta, \varepsilon, \alpha)$ *is a Hopf algebra, i.e. a bialgebra such that α is an anti-homomorphism $\alpha(FG) = \alpha(G)\alpha(F)$ and*

$$\mathrm{conc} \circ (\alpha \otimes \mathrm{id}) \circ \delta = \eta \circ \varepsilon = \mathrm{conc} \circ (\mathrm{id} \otimes \alpha) \circ \delta, \tag{5.1}$$

where $\eta : \mathbb{F} \to \mathbb{F}\langle A\rangle, c \mapsto c1$.

(v) In addition we have $\alpha \circ \alpha = \mathrm{id}$ and cocommutativity $\delta = \sigma \circ \delta$, where $\sigma : \mathbb{F}\langle A\rangle \otimes \mathbb{F}\langle A\rangle \to \mathbb{F}\langle A\rangle \otimes \mathbb{F}\langle A\rangle$ is defined by $\sigma(F \otimes G) = G \otimes F$ for all $F, G \in \mathbb{F}\langle A\rangle$.

(vi) All maps considered above have obvious extensions to $\mathbb{F}\langle\langle A\rangle\rangle$, but note that $\delta(\mathbb{F}\langle\langle A\rangle\rangle) \subseteq \prod_{r,s=0}^{\infty} \mathbb{F}\langle\langle A\rangle\rangle_r \otimes \mathbb{F}\langle\langle A\rangle\rangle_s$, which is bigger than the algebraic tensor product $\mathbb{F}\langle\langle A\rangle\rangle \otimes \mathbb{F}\langle\langle A\rangle\rangle$.

Proof. (i) is known. (ii) Linearity of ε is clear, and multiplicativity $\varepsilon(FG) = \varepsilon(F)\varepsilon(G)$ is immediate by definition of the multiplication of $\mathbb{F}\langle A\rangle$. Since δ is a homomorphism, it suffices to verify $(\delta \otimes \mathrm{id}) \circ \delta(F) = (\mathrm{id} \otimes \delta) \circ \delta(F)$ for $F = 1$ and $F = a \in A$. For $F = 1$, both sides equal $1 \otimes 1 \otimes 1$, and for $F = a \in A$, both sides equal $a \otimes 1 \otimes 1 + 1 \otimes a \otimes 1 + 1 \otimes 1 \otimes a$. Also for the counit property $(\varepsilon \otimes \mathrm{id}) \circ \delta(F) = F = (\mathrm{id} \otimes \varepsilon) \circ \delta(F)$ it suffices to consider $F = 1$ and $F = a \in A$. The first case is trivial, and the second follows readily from $\varepsilon(a) = 0$.

(iii) δ is multiplicative by definition, and for ε this was seen above.

(iv) By linearity, it suffices to verify anti-multiplicativity $\alpha(ww') = \alpha(w')\alpha(w)$ for $w, w' \in A^*$, and this is immediate since α reverses the order of letters in each word. For $F = 1$, the antipode property (5.1) is clearly true since all expressions equal 1. For $F \in A^*$ with $|F| \geq 1$, we have $\varepsilon(F) = 0$ so that it remains to show that the left and right hand sides of (5.1) vanish. We do this for the latter, the other case being analogous.

Since α is an anti-homomorphism, it isn't totally obvious that it suffices to verify (5.1) on $\{1\} \cup A$. Thus let $n \geq 1$ and $a_1, \ldots, a_n \in A$. Noting that $\delta(a) = a \otimes 1 + 1 \otimes a = \sum_{s=0}^{1} a^s \otimes a^{1-s}$ for $a \in A$ we have

$$\delta(a_1 \cdots a_n) = \delta(a_1) \cdots \delta(a_n) = \sum_{s \in \{0,1\}^n} a_1^{s_1} \cdots a_n^{s_n} \otimes a_1^{1-s_1} \cdots a_n^{1-s_n}, \tag{5.2}$$

where we sum over all ways of distributing the factors $a_1, \ldots, a_n$ over the two sides of the tensor product while maintaining their order. Now

$$\begin{aligned} \mathrm{conc} \circ (\mathrm{id} \otimes \alpha) \circ \delta(a_1 \cdots a_n) &= \sum_{s \in \{0,1\}^n} a_1^{s_1} \cdots a_n^{s_n} \alpha(a_1^{1-s_1} \cdots a_n^{1-s_n}) \\ &= \sum_{s \in \{0,1\}^n} (-1)^{\sum_{i=1}^n (1-s_i)} a_1^{s_1} \cdots a_n^{s_n} a_n^{1-s_n} \cdots a_1^{1-s_1}. \end{aligned}$$

Since the middle factor $a_n^{s_n} a_n^{1-s_n} = a_n$ is independent of s_n, the presence of the factor $(-1)^{1-s_n}$ implies that the summation over s_n gives zero.

(v) Involutivity of α and cocommutativity of δ are immediate from the respective definitions.

(vi) The only point that may not be entirely obvious concerns the well-definedness of δ on $\mathbb{F}\langle\langle A \rangle\rangle$. In view of (5.2), wherein the sum of all exponents on the r.h.s. is n, we have

$$\delta(\mathbb{F}\langle A \rangle_n) \subseteq \bigoplus_{k=0}^{n} \mathbb{F}\langle A \rangle_k \otimes \mathbb{F}\langle A \rangle_{n-k} =: (\mathbb{F}\langle\langle A \rangle\rangle \otimes \mathbb{F}\langle\langle A \rangle\rangle)_n,$$

which means that δ preserves degrees so that defining $\delta(F)$ for $F \in \mathbb{F}\langle\langle A \rangle\rangle$ by $\delta(F) = \sum_{n=0}^{\infty} \delta(F_n)$ makes sense. All remaining verifications are straightforward. ■

5.3 Remark For every Lie algebra $\mathcal{L}$, there is a unique associative algebra $U(\mathcal{L})$, the universal enveloping algebra. One can show that $U(\mathcal{L})$ is a Hopf algebra with maps very similar to the above, see e.g. [39]. In fact, one then has an isomorphism $U(\mathcal{L}_{\mathbb{F}}\langle A \rangle) \cong \mathbb{F}\langle A \rangle$ of Hopf algebras, showing that $U(\mathcal{L})$ generalizes $\mathbb{F}\langle A \rangle$. □

5.4 Lemma *Let $\mathbb{F}$ be a field and let A, B be sets.*

(i) For $F, G \in \mathbb{F}\langle\langle A \rangle\rangle$ define

$$d(F,G) = \begin{cases} 0 & \text{if } F = G \\ 2^{-\min\{n \in \mathbb{N} \mid F_n \neq G_n\}} & \text{if } F \neq G \end{cases}$$

Then d is a metric. The topology on $\mathbb{F}\langle\langle A \rangle\rangle$ induced by d is called the A-adic topology.

(ii) A map $\alpha : \mathbb{F}\langle\langle A \rangle\rangle \to \mathbb{F}\langle\langle B \rangle\rangle$ is continuous w.r.t. the A-adic topologies on both sides if and only if for every $r \in \mathbb{N}$ there is $s \in \mathbb{N}$ such that $\phi(F)_{\leq r}$ (the projection of $\phi(F)$ to $\mathbb{F}\langle\langle B \rangle\rangle_{\leq r}$) depends only on $F_{\leq s}$.

(iii) If $\alpha : \mathbb{F}\langle\langle A \rangle\rangle \to \mathbb{F}\langle\langle B \rangle\rangle$ is homogeneous, i.e. $\alpha(\mathbb{F}\langle\langle A \rangle\rangle_n) \subseteq \mathbb{F}\langle\langle B \rangle\rangle_n$ $\forall n \in \mathbb{N}_0$, then it is continuous.

We omit the easy proof. The maps δ, α are homogeneous, thus continuous. Also $\varepsilon : \mathbb{F}\langle\langle A \rangle\rangle \to \mathbb{F}$ is continuous if we identify $\mathbb{F}$ with $\mathbb{F}\langle\langle B \rangle\rangle$ with $B = \emptyset$.

6 Standard algebraic proof of BCHD

This section requires everything from Section 3 and most of Section 5, but not Section 4.

6.1 The Lie-ness criteria of Friedrichs and Dynkin-Specht-Wever

6.1 Definition *Let A be a set and $\mathbb{F}$ a field. We define $\mathbb{F}$-linear maps $D : \mathbb{F}\langle A \rangle \to \mathbb{F}\langle A \rangle$ and $R : \mathbb{F}\langle A \rangle \to \mathcal{L}_{\mathbb{F}}\langle A \rangle \subset \mathbb{F}\langle A \rangle$ by defining them on the words $A^* \subset \mathbb{F}\langle A \rangle$:*

$$\begin{aligned} D: &\quad w \mapsto |w|\, w, \\ R: &\quad 1 = \epsilon \mapsto 0, \quad a \mapsto a, \quad a_1 a_2 \ldots a_n \mapsto [a_1, [a_2, [\cdots, [a_{n-1}, a_n] \cdots] \quad \forall a, a_1, \ldots, a_n \in A. \end{aligned}$$

Note that D and R map the homogeneous components $\mathbb{F}\langle A \rangle_n$ into themselves.

6.2 Lemma *With the above notations, the following holds for all $F \in \mathbb{F}\langle A\rangle$:*

$$\mathrm{conc} \circ (D \otimes \alpha) \circ \delta(F) = R(F). \tag{6.1}$$

Proof. Since this equation is linear, it suffices to prove it on words $F = w \in A^*$. In degree zero, i.e. for $F = 1$, it becomes $D(1)1 = R(1)$, which is true since $D(1) = R(1) = 0$ by definition. Now let $n \in \mathbb{N}, a_1, \ldots, a_n \in A$. Combining (5.2) with $D(a_1^{s_1} \cdots a_n^{s_n}) = (\sum_{i=1}^n s_i) a_1^{s_1} \cdots a_n^{s_n}$, we have

$$\begin{aligned} \mathrm{conc} \circ (D \otimes \alpha)(\delta(a_1 \cdots a_n)) &= \sum_{s \in \{0,1\}^n} (-1)^{\sum_i (1-s_i)} \left(\sum_{i=1}^n s_i \right) a_1^{s_1} \cdots a_n^{s_n} a_n^{1-s_n} \cdots a_1^{1-s_1} \\ &= \sum_{s \in \{0,1\}^n} (-1)^{\sum_i (1-s_i)} \left(\sum_{i=1}^n s_i \right) a_1^{s_1} \cdots a_{n-1}^{s_{n-1}} a_n a_{n-1}^{1-s_{n-1}} \cdots a_1^{1-s_1}. \end{aligned}$$

Writing $\sum_{i=1}^n s_i = (\sum_{i=1}^{n-1} s_i) + s_n$ and noting that $(\sum_{i=1}^{n-1} s_i) a_1^{s_1} \cdots a_{n-1}^{s_{n-1}} a_n a_{n-1}^{1-s_{n-1}} \cdots a_1^{1-s_1}$ is independent of s_n, this term is killed by the s_n-summation due to the presence of the factor $(-1)^{1-s_n}$. The term involving s_n obviously only contributes for $s_n = 1$, for which $(-1)^{1-s_n} = 1$. We thus arrive at

$$\mathrm{conc} \circ (D \otimes \alpha)(\delta(a_1 \cdots a_n)) = \sum_{s_1, \ldots, s_{n-1}} (-1)^{\sum_{i=1}^{n-1}(1-s_i)} a_1^{s_1} \cdots a_{n-1}^{s_{n-1}} a_n a_{n-1}^{1-s_{n-1}} \cdots a_1^{1-s_1}. \tag{6.2}$$

For $n = 1$ the r.h.s. reduces to a_1, which equals $R(a_1)$. For $n \geq 2$ we notice that

$$\sum_{s_{n-1}=0}^{1} (-1)^{1-s_{n-1}} a_{n-1}^{s_{n-1}} a_n a_{n-1}^{1-s_{n-1}} = a_{n-1} a_n - a_n a_{n-1} = [a_{n-1}, a_n].$$

It should be obvious enough without formal inductive argument that working our way outwards in (6.2), we have

$$\mathrm{conc} \circ (D \otimes \alpha)(\delta(a_1 \cdots a_n)) = [a_1, [a_2, [\cdots, [a_{n-1}, a_n] \cdots] = R(a_1 \cdots a_n),$$

completing the proof of (6.1). ■

6.3 Theorem (Friedrichs-Dynkin-Specht-Wever) [10] *If* $\mathrm{char}\, \mathbb{F} = 0$, *the following are equivalent for all* $F \in \mathbb{F}\langle A\rangle$:

(i) F *is a Lie polynomial.*

(ii) $\delta(F) = F \otimes 1 + 1 \otimes F$. *(Friedrichs' criterion)*

(iii) The constant term $(F, 1)$ *vanishes and* $R(F) = D(F)$. *(Criterion of Dynkin-Specht-Wever)*

(The equivalence (i)⇔(ii) is called Friedrichs' theorem, and (i)⇔(iii) the Dynkin-Specht-Wever theorem.)

Proof. (iii)⇒(i) Assume $F \in \mathbb{F}\langle A\rangle$ satisfies $(F, 1) = 0$ and $R(F) = D(F)$. Since $R(F)$ is a Lie polynomial, it follows from $R(F) = D(F)$ that $F_n = \frac{1}{n} R(F)_n$ is a Lie polynomial for all $n \geq 1$. (This is the only place where characteristic zero is used.) Since $F_0 = (F, 1)\epsilon$ is zero by assumption, F is Lie.

(i)⇒(ii) Let $L = \{F \in \mathbb{F}\langle A\rangle \mid \delta(F) = F \otimes 1 + 1 \otimes F\}$, which is a linear subspace. By definition of δ, we have $A \subseteq L$. If $F, G \in L$, the computation

$$\begin{aligned} \delta([F,G]) &= \delta(FG - GF) = (F \otimes 1 + 1 \otimes F)(G \otimes 1 + 1 \otimes G) - (G \otimes 1 + 1 \otimes G)(F \otimes 1 + 1 \otimes F) \\ &= FG \otimes 1 + 1 \otimes FG + F \otimes G + G \otimes F - GF \otimes 1 - 1 \otimes GF - F \otimes G - G \otimes F \\ &= [F,G] \otimes 1 + 1 \otimes [F,G] \end{aligned}$$

shows that $[F, G] \in L$, so that L is closed under commutators. Thus it contains the smallest Lie subalgebra of $\mathbb{F}\langle A\rangle$ that contains A, to wit $\mathcal{L}_{\mathbb{F}}\langle A\rangle$.

[10] Kurt Otto Friedrichs (1901-1982) German American mathematician. Best known for work on differential equations. Wilhelm Otto Ludwig Specht (1907-1985). German mathematician best known for Specht modules in the representation theory of symmetric groups.

(ii)⇒(iii) Assume $F \in \mathbb{F}\langle A\rangle$ satisfies (ii). Applying $\varepsilon \otimes \mathrm{id}$ to both sides of $\delta(F) = F \otimes 1 + 1 \otimes F$ gives $F = \varepsilon(F)1 + \varepsilon(1)F$. With $\varepsilon(1) = 1$ it follows that $(F, 1) = \varepsilon(F) = 0$. Plugging $\delta(F) = F \otimes 1 + 1 \otimes F$ into (6.1) and using $\alpha(1) = 1$, $D(1) = 0$, we get $D(F) = D(F)\alpha(1) + D(1)\alpha(F) = R(F)$. ■

6.4 Remark 1. The easy implications (iii)⇒(i)⇒(ii) are from the classical papers [18, 59, 67] on the criteria of Dynkin-Specht-Wever (1947-49) and Friedrichs (1953) [24]. (Friedrichs gave no general proof, which was supplied soon after by P. M. Cohn, W. Magnus and others.) The beautiful approach to (ii)⇒(iii) based on (6.1) is due to von Waldenfels [65] (1966), who eliminated any reference to the fact that $\mathbb{F}\langle A\rangle$ is the universal enveloping algebra of $\mathcal{L}_{\mathbb{F}}\langle A\rangle$, let alone the Poincaré-Birkhoff-Witt theorem. Cf. also the exposition in [49] and [31, Exercise 1.5.11(c)]. The proof of (6.1) using (5.2) is probably well-known to the experts, but I haven't yet seen it elsewhere.

2. If $(H, m, 1, \delta, \varepsilon, \alpha)$ is a Hopf algebra and $u, v : H \to H$ are linear maps, we define the convolution $u \star v : H \to H$ by $u \star v = m \circ (u \otimes v) \circ \delta$. Associativity of m and coassociativity of δ imply associativity of $\star$, and (5.1) becomes $\alpha \star \mathrm{id} = \mathrm{id} \star \alpha = \eta \circ \varepsilon$. With $H = \mathbb{F}\langle A\rangle$ and $m = \mathrm{conc}$, (6.1) just is $D \star \alpha = R$. Thus $R \star \mathrm{id} = D \star \alpha \star \mathrm{id} = D$, to wit

$$D = \mathrm{conc} \circ (R \otimes \mathrm{id}) \circ \delta. \tag{6.3}$$

(This actually is equivalent to (6.1), the converse working analogously.) Using (6.3) it is also immediate that $\delta(F) = F \otimes 1 + 1 \otimes F$ implies $R(F) = D(F)$. And (6.3) also has an α-free direct proof, see [68], very similar to the proof of (6.1) in [65], but I prefer the essentially non-inductive proof of (6.1) given above.

3. Before the Hopf algebraic view of $\mathbb{F}\langle A\rangle$ and $\mathbb{F}\langle\langle A\rangle\rangle$ became common, Friedrichs' criterion was usually stated differently, see e.g. [24, 43, 47]: For each $a \in A$ introduce a new formal variable a', putting also $A' = \{a' \mid a \in A\}$, in such a way that the variables in A' commute with those in A. Now Friedrichs' criterion is equivalent to $F(a + a') = F(a) + F(a')$ where, e.g., $F(a + a')$ means that each instance $a \in A$ in F is replaced by $a + a'$. To see this equivalence it suffices to realize that the free algebra generated by $A \cup A'$, but with each element of A commuting with each element of A', is nothing other than $\mathbb{F}\langle A\rangle \otimes \mathbb{F}\langle A'\rangle \cong \mathbb{F}\langle A\rangle \otimes \mathbb{F}\langle A\rangle$. Now $F(a + a')$ lives in the latter tensor product, and $F(a + a')$ precisely is $\delta(F)$. This author finds the superiority of the Hopf algebraic formulation quite evident. □

From now on we always assume the alphabet A to be finite. (For our purposes this is sufficient, even though the results hold without this assumption provided one adapts the definition of $\mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle$, cf. [49].)

6.5 Corollary *The analogue of Theorem 6.3 holds for $F \in \mathbb{F}\langle\langle A\rangle\rangle$ if we replace (Lie) 'polynomial' by 'series' in (i) and interpret $\mathbb{F}\langle\langle A\rangle\rangle \otimes \mathbb{F}\langle\langle A\rangle\rangle$ as the completion $\prod_{m,n} \mathbb{F}\langle A\rangle_m \otimes \mathbb{F}\langle A\rangle_n$.*

Proof. We have already seen that δ extends to $\mathbb{F}\langle\langle A\rangle\rangle$, and the observation that R and D preserve degrees of homogeneous elements guarantees that their obvious extensions to $\mathbb{F}\langle\langle A\rangle\rangle$ are well-defined. Now the claim follows since a formal series' being Lie is determined degreewise. ■

6.2 Proof of BCH

6.6 Proposition *Let $F \in \mathbb{F}\langle\langle A\rangle\rangle_{>0}$. Then F is a Lie series if and only if $\delta(e^F) = e^F \otimes e^F$.*

Proof. Recall that $e^F = \exp F$ is well-defined since F has no constant term. By (i)⇔(ii) in Corollary 6.5, F is a Lie series if and only if $\delta(F) = F \otimes 1 + 1 \otimes F$. Both sides of this identity live in $\prod_{n=0}^{\infty} \oplus_{k=0}^{n} \mathbb{F}\langle A\rangle_k \otimes \mathbb{F}\langle A\rangle_{n-k}$ and have zero constant term, so that we can exponentiate them. Since exponentiation of formal series is injective, we find

$$F \in \mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle \quad \Leftrightarrow \quad \delta(F) = F \otimes 1 + 1 \otimes F \quad \Leftrightarrow \quad \exp(\delta(F)) = \exp(F \otimes 1 + 1 \otimes F).$$

As to the r.h.s., since $F \otimes 1$ and $1 \otimes F$ commute, we have

$$\exp(F \otimes 1 + 1 \otimes F) = \exp(F \otimes 1) \exp(1 \otimes F) = (e^F \otimes 1)(1 \otimes e^F) = e^F \otimes e^F,$$

where we also used $\exp(F) \otimes 1 = \exp(F \otimes 1)$ (and $\leftrightarrow$) which follows from 1 being the unit of $\mathbb{F}\langle\langle A\rangle\rangle$. Thus the proof is complete once we prove that $\exp(\delta(F)) = \delta(e^F)$ holds for all $F \in \mathbb{F}\langle\langle A\rangle\rangle_{>0}$.

Since δ is an algebra homomorphism, it satisfies

$$\delta\left(\sum_{n=0}^{N} \frac{F^n}{n!}\right) = \sum_{n=0}^{N} \frac{\delta(F)^n}{n!} \qquad \forall N \in \mathbb{N}. \tag{6.4}$$

Since F has no constant term and δ is degree-preserving, summands $\delta(F^k/k!) = \delta(F)^k/k!$ with $k > N$ do not contribute to the degree $\leq N$ parts of $\delta(\exp(F))$ and $\exp(\delta(F))$, so that (6.4) implies $\delta(\exp(F)) = \exp(\delta(F))$. (This argument is usually, see e.g. [49], couched in topological language using the 'A-adic topology' of $\mathbb{F}\langle\langle A\rangle\rangle$, but this is not really needed here.) ■

(An element X of a Hopf algebra is called *group-like* if $\delta(X) = X \otimes X$ and $\varepsilon(X) = 1$ and is called *primitive* if $\delta(X) = X \otimes 1 + 1 \otimes X$ (although 'Lie-like' would be better). Thus e^X is group-like iff X is primitive.)

6.7 Corollary (BCH) *If A is finite and $X_1, \ldots, X_n \in \mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle$ then $\log(e^{X_1} \cdots e^{X_n}) \in \mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle$. In particular $H = \log(e^x e^y) \in \mathbb{F}\langle\langle x, y\rangle\rangle$ is a Lie series.*

Proof. By assumption, the X_i are Lie, in particular they have no constant terms. By Proposition 6.6, we have $\delta(e^{X_i}) = e^{X_i} \otimes e^{X_i}$ for all i. Since δ is a homomorphism, $\delta(e^{X_1} \cdots e^{X_n}) = e^{X_1} \cdots e^{X_n} \otimes e^{X_1} \cdots e^{X_n}$. With $W = \log(e^{X_1} \cdots e^{X_n})$ we have $e^W = e^{X_1} \cdots e^{X_n}$, thus $\delta(e^W) = e^W \otimes e^W$, and using Proposition 6.6 again, W is Lie. The second statement is a special case (since the generators x, y are Lie). ■

The above approach to BCH is due to P. Cartier[11] [10, 11] and was canonized in the first half of the 1960s by books like [38, 57, 37].

6.3 Proof of Dynkin's formula (1.4)

6.8 Corollary *Define a linear map $P : \mathbb{F}\langle\langle A\rangle\rangle \to \mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle$ by linear extension of*

$$\epsilon \mapsto 0, \qquad a_1 \ldots a_n \mapsto \frac{1}{n} R(a_1 \ldots a_n) = \frac{1}{n}[a_1, [a_2, [\cdots, [a_{n-1}, a_n] \cdots].$$

Then P satisfies $P\restriction_{\mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle} = \mathrm{id}$, thus also $P^2 = P$, so that it is a projection, called the Dynkin idempotent.

Proof. It is clear that the image of P is contained in $\mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle$. Let $F \in \mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle$. Since P respects the $\mathbb{N}_0$-grading, it suffices to prove $P(F_n) = F_n$ for all n. Since Lie series by definition have no constant term, $F_0 = 0 = P(F_0)$. For $n \geq 1$ we have

$$P(F_n) = \frac{1}{n} R(F_n) = \frac{1}{n} D(F_n) = \frac{n}{n} F_n = F_n,$$

where the second identity is due to the Dynkin-Specht-Wever implication (i)⇒(iii) in Corollary 6.5. ■

6.9 Corollary (Dynkin [18]) *Dynkin's formula (1.4) holds as an identity in $\mathcal{L}_{\mathbb{F}}\langle\langle x, y\rangle\rangle$.*

Proof. By Corollary 6.7, the series (1.3) for $H = \log(e^x e^y) \in \mathbb{F}\langle\langle x, y\rangle\rangle$ is a Lie series, thus applying the projection $P : \mathbb{F}\langle\langle A\rangle\rangle \to \mathcal{L}_{\mathbb{F}}\langle\langle A\rangle\rangle$ from Corollary 6.8 does not change it. But since

$$P(x^{m_1} y^{n_1} \cdots x^{m_k} y^{n_k}) = \frac{1}{\sum_{i=1}^{k}(m_i + n_i)} \overbrace{[x, [\cdots, [x,}^{m_1} \overbrace{[y, [\cdots, [y,}^{n_1} [\ldots \overbrace{[x, [\cdots, [x,}^{m_k} \overbrace{[y, [\cdots, [y, [\cdots]}^{n_k} \cdots],$$

the summand-wise application of P to (1.3) gives nothing other than Dynkin's formula (1.4). ■

We note that the above proof of BCH (Subsections 6.1-6.2) is slightly longer than Eichler's, but more conceptual, and a negligible additional effort also proves Dynkin's formula. One wonders whether Eichler's proof in some sense is an infinitesimal version of the above one.

[11] Pierre Cartier (b. 1932). French mathematician. Worked on many subjects, in particular algebraic geometry.

As mentioned before, for most purposes in Lie theory the qualitative BCH theorem is sufficient. But there are other applications, e.g. to differential equations or quantum (field) theory, where it is important to have a managable series expansion. As we have seen, while Dynkin's formula can be proven in different ways, it is hard to work with. For this reason we will now study some alternative approaches to $\log(e^x e^y)$. This will lead us into combinatorial waters of increasing depth.

7 Algebraic version of the proof of BCHD from Section 2.3

The analytic and algebraic proofs of the BCH theorem and of the Dynkin representation (1.4) given in Sections 2 and 6, respectively, look very different and seem to have nothing in common. In this section we will show that this (if it is true at all) is not due to the first proof being analytic and the second algebraic by recasting the first proof in purely algebraic terms. This requires two additions to our algebraic toolkit that will be used again in later sections.

Formal power series in one variable are well known (and briefly discussed in Appendix C), the generalization to several commuting variables being straightforward. And the algebra $\mathbb{F}\langle\langle A\rangle\rangle$ of formal series in non-commuting variables has been discussed extensively above. But one can also consider formal series involving commuting and non-commuting variables:

7.1 Definition *By $\mathbb{F}\langle\langle x_1,\ldots,x_n\rangle\rangle[[t_1,\ldots,t_m]]$, where $n,m\in\mathbb{N}$, we denote the algebra of formal series in n non-commuting variables $x_1,\ldots,x_n$ and m variables $t_1,\ldots,t_m$ that commute with each other and with the x's. As a vector space it is given by*

$$\mathbb{F}\langle\langle x_1,\ldots,x_n\rangle\rangle[[t_1,\ldots,t_m]] = \mathbb{F}^{\mathbb{N}_0^m\times\{x_1,\ldots,x_n\}^*},$$

but we will write its elements as (possibly infinite) linear combinations

$$\sum_{\substack{I=(i_1,\ldots,i_m)\in\mathbb{N}_0^m\\ w\in\{x_1,\ldots,x_n\}^*}} a_{I,w}\, t_1^{i_1}\cdots t_m^{i_m}\, w,$$

whose multiplication is defined by multiplication of basis elements:

$$(t_1^{i_1}\cdots t_m^{i_m}\, w)\,(t_1^{i_1'}\cdots t_m^{i_m'}\, w') = t_1^{i_1+i_1'}\cdots t_m^{i_m+i_m'}\, ww'.$$

7.2 Remark Note that an element of $\mathbb{F}\langle\langle x_1,\ldots,x_n\rangle\rangle[[t_1,\ldots,t_m]]$ can and will be considered as a formal series in commuting variables $t_1,\ldots,t_m$ with coefficients in $\mathbb{F}\langle\langle x_1,\ldots,x_n\rangle\rangle$.

One could also define $\mathbb{F}\langle\langle x_1,\ldots,x_n\rangle\rangle[[t_1,\ldots,t_m]]$ as $\mathbb{F}\langle\langle x_1,\ldots,x_n\rangle\rangle\overline{\otimes}\mathbb{F}[[t_1,\ldots,t_m]]$, but the above approach is more straightforward. (In the same vein, e.g., $\mathbb{F}\langle\langle x,y\rangle\rangle$ is a 'free' tensor product of $\mathbb{F}[[x]]$ and $\mathbb{F}[[y]]$.) □

We now turn to the promised algebraic version of Proposition 2.7, which we state in the greater generality needed later on. All algebras appearing here are unital and all homomorphisms map units to units.

7.3 Definition *If $\mathcal{A}$ is an associative algebra and $\phi:\mathcal{A}\to\mathcal{A}$ an algebra homomorphism, a (ϕ,ϕ)-derivation of $\mathcal{A}$ is a linear map $D:\mathcal{A}\to\mathcal{A}$ such that*

$$D(XY) = D(X)\phi(Y) + \phi(X)D(Y) \qquad \forall X,Y\in\mathcal{A}. \tag{7.1}$$

A derivation of $\mathcal{A}$ is an $(\mathrm{id},\mathrm{id})$-derivation.

7.4 Lemma *Every (ϕ,ϕ)-derivation D satisfies $D(1)=0$.*

Proof. Follows from $D(1) = D(1\cdot 1) = D(1)\phi(1)+\phi(1)D(1) = 2D(1)$. (Recall that $\phi(1)=1$.) ■

For the prototypical example of a derivation see Definition B.2 in Appendix C.

7.5 PROPOSITION *Let $\mathbb{F}$ be a field of characteristic zero, and let $\mathcal{A}$ be an $\mathbb{F}$-algebra of formal series in commuting and/or non-commuting variables. If $\phi : \mathcal{A} \to \mathcal{A}$ is a homomorphism, $D : \mathcal{A} \to \mathcal{A}$ a (ϕ, ϕ)-derivation and $X \in \mathcal{A}$, then*

$$
\begin{aligned}
D(e^X) &= e^{\phi(X)} \sum_{n=0}^{\infty} \frac{(-\mathrm{ad}_{\phi(X)})^n}{(n+1)!}(D(X)) = e^{\phi(X)} \frac{1-e^{-\mathrm{ad}_{\phi(X)}}}{\mathrm{ad}_{\phi(X)}}(D(X)) &\quad (7.2)\\
&= \sum_{n=0}^{\infty} \frac{(\mathrm{ad}_{\phi(X)})^n}{(n+1)!}(D(X))\, e^{\phi(X)} = \frac{e^{\mathrm{ad}_{\phi(X)}}-1}{\mathrm{ad}_{\phi(X)}}(D(X)) e^{\phi(X)}. &\quad (7.3)
\end{aligned}
$$

Proof. We put $A = \phi(X), B = D(X)$ and claim that

$$D(X^n) = \sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1} \qquad \forall n \in \mathbb{N}_0. \tag{7.4}$$

For $n = 0$ this is true since the r.h.s. vanishes and $D(1) = 0$ by Lemma 7.4. Assuming that (7.4) holds for n,

$$
\begin{aligned}
D(X^{n+1}) &= D(X^n X) = D(X^n)\phi(X) + \phi(X^n)D(X)\\
&= \left(\sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1}\right) A + A^n B = \sum_{\ell=1}^{n+1} A^{n+1-\ell} B A^{\ell-1}
\end{aligned}
$$

provides the induction step. (We used the homomorphism and derivation properties of ϕ, D, and the final equality was already shown in the proof of Lemma 2.12.) Dividing (7.4) by $n!$ and summing over $n \in \mathbb{N}_0$, we obtain

$$D(e^X) = \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{\ell=1}^{n} A^{n-\ell} B A^{\ell-1},$$

which equals the l.h.s. of (2.8), whose r.h.s. equals the middle expression in (7.2) if we take the definitions of A, B into account. This proves (7.2) since the proof of (2.8) remains valid in the present formal setting, all sums being locally finite. Since the same holds for (2.1), the dual formula (7.3) follows as in Remark 2.8.5. ■

7.6 THEOREM (BCHD) *The formal series $H = \log(e^x e^y) \in \mathbb{Q}\langle\langle x, y\rangle\rangle$ is a Lie series, and it is given by (1.4), interpreted as element of $\mathbb{Q}\langle\langle x, y\rangle\rangle$.*

Proof. We put $\mathcal{A} = \mathbb{F}\langle x, y\rangle[t]$ and define linear maps $D, I : \mathcal{A} \to \mathcal{A}$ by

$$
\begin{aligned}
D:\ t^n w &\mapsto n t^{n-1} w \qquad \forall n \in \mathbb{N}_0,\ w \in \{x, y\}^*,\\
I:\ t^n w &\mapsto \frac{t^{n+1}}{n+1} w \qquad \forall n \in \mathbb{N}_0,\ w \in \{x, y\}^*.
\end{aligned}
$$

The proof of Lemma B.5 is easily adapted to show that D is a derivation (with $\phi = \mathrm{id}$). It is obvious that $D \circ I = \mathrm{id}_{\mathcal{A}}$ and $I \circ D(F) = F - F_0$, where F_0 is the coefficient of t^0 in the expansion $F = \sum_{n=0}^{\infty} t^n F_n$, where $F_n \in \mathbb{F}\langle\langle x, y\rangle\rangle$. We have

$$D(e^{tx}) = \sum_{n=0}^{\infty} \frac{1}{n!} n t^{n-1} x^n = \sum_{n=0}^{\infty} \frac{t^n x^{n+1}}{n!} = x e^{tx} = e^{tx} x,$$

and similarly $D(e^{ty}) = y e^{ty} = e^{ty} y$. (Deriving these formulae from Proposition 7.5 is possible, but rather roundabout.)

Since $e^{tx} e^{ty} \in \mathcal{A}$ has constant term 1, we can define $\widehat{H} = \sum_{n=0}^{\infty} t^n \widehat{H}_n = \log(e^{tx} e^{ty}) \in \mathcal{A}$. Clearly, $\widehat{H}_0 = 0$. The specialization $\widehat{H}_{t=1} = \sum_{n=0}^{\infty} \widehat{H}_n \in \mathbb{F}\langle\langle x, y\rangle\rangle$ is admissible since it equals $H = \log(e^x e^y)$. Applying the derivation D to both sides of $e^{\widehat{H}} = e^{tx} e^{ty}$ and appealing to (7.3) and the above formulae for $D(e^{tx}), D(e^{ty})$ gives

$$\frac{e^{\mathrm{ad}_{\widehat{H}}}-1}{\mathrm{ad}_{\widehat{H}}}(D(\widehat{H}))\, e^{\widehat{H}} = D(e^{\widehat{H}}) = D(e^{tx} e^{ty}) = x e^{tx} e^{ty} + e^{tx} e^{ty} y = x e^{\widehat{H}} + e^{\widehat{H}} y,$$

whence

$$\frac{e^{\mathrm{ad}_{\widehat{H}}}-1}{\mathrm{ad}_{\widehat{H}}}(D(\widehat{H})) = x + e^{\mathrm{ad}_{\widehat{H}}}(y).$$

The function $f(u) = \frac{e^u-1}{u}$ has $g(u) = \frac{u}{e^u-1}$ as multiplicative inverse: $f \cdot g = g \cdot f = 1$. Thus $f(\mathrm{ad}_{\widehat{H}})$ is invertible with inverse $g(\mathrm{ad}_{\widehat{H}})$, so that

$$D(\widehat{H}) = \frac{\mathrm{ad}_{\widehat{H}}}{e^{\mathrm{ad}_{\widehat{H}}}-1}(x + e^{\mathrm{ad}_{\widehat{H}}}(y)). \tag{7.5}$$

Continuing exactly as in Section 2.3 we arrive at the formal version of (2.11):

$$D(\widehat{H}) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(e^{\mathrm{ad}_{tx}}e^{\mathrm{ad}_{ty}}-1)^{k-1}(x + e^{\mathrm{ad}_{tx}}(y)). \tag{7.6}$$

Applying the anti-derivation operator I from the left and using $I \circ D(\widehat{H}) = \widehat{H} - \widehat{H}_0 = \widehat{H}$, we have

$$\widehat{H} = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} I\left[(e^{\mathrm{ad}_{tx}}e^{\mathrm{ad}_{ty}}-1)^{k-1}(x + e^{\mathrm{ad}_{tx}}(y))\right].$$

Once we expand the exponential functions, as done several times before, the operator I has the same effect of generating the factors like $1+\sum_i(m_i+n_i)$ in the denominator, as in Section 2.3. It also increases the exponents of t, but the specialization $t=1$ makes the t's disappear, yielding for $\widehat{H}_{t=1} = H = \log(e^x e^y)$ the formal series version of Dynkin's formula (1.5). Since the latter clearly is equivalent to (1.4) in each order, the two formulae define the same formal series. ■

7.7 REMARK Up to and including (7.5), the above argument is essentially the same as in Djoković's [16]. But then he contents himself with giving an inductive argument for Lie-ness of the coefficients $H_n \in \mathbb{Q}\langle\langle x,y\rangle\rangle$. (The argument needed to get from (7.5) to (7.6) can in principle be found in [34], but this may well have been the only place in 1975.) □

8 Expansion of $\log(e^x e^y)$ in powers of x. Fourth algebraic proof of BCH

The formal series $H = \log(e^x e^y) \in \mathbb{F}\langle\langle x,y\rangle\rangle$ can be decomposed w.r.t. the usual gradation of $\mathbb{F}\langle\langle x,y\rangle\rangle$ by word length. I.e. $H = \sum_{n=1}^{\infty} H_n$, where $H_n \in \mathbb{F}\langle x,y\rangle_n$ $\forall n$. This was even central in Eichler's proof of BCH.

But there are other gradations of the algebras $\mathbb{F}\langle\langle A\rangle\rangle$: If $a \in A$ and $w \in A^*$, let $|w|_a$ denote the number of occurrences of the letter a in the word w. Clearly $|ww'|_a = |w|_a + |w'|_a$. Thus

$$\mathbb{F}\langle\langle A\rangle\rangle_n^{(a)} := \{F \in \mathbb{F}\langle\langle A\rangle\rangle \mid |w|_a \neq n \ \Rightarrow (F,w) = 0\} \subset \mathbb{F}\langle\langle A\rangle\rangle$$

defines a new gradation of $\mathbb{F}\langle\langle A\rangle\rangle$. Note that if $\{a\} \subsetneq A$, there are arbitrarily long words w with given $|w|_a$, so that $\mathbb{F}\langle\langle A\rangle\rangle_n^{(a)} \not\subseteq \mathbb{F}\langle A\rangle$, even if A is finite.

The aim of this section, where we mostly follow Reutenauer's [49] presentation of very classical results, is to determine $H'_n \in \mathbb{F}\langle\langle x,y\rangle\rangle_n^{(x)}$ (clearly unique) such that $H = \sum_{n=0}^{\infty} H'_n$. (Some would call this a resummation of the original series $\sum_n H_n$.)

Recall that the Bernoulli numbers $B_0, B_1, \ldots \in \mathbb{Q}$ are defined by their exponential generating function:

$$\frac{z}{e^z-1} = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n. \tag{8.1}$$

This is equivalent to

$$\left(\sum_{m=0}^{\infty} \frac{z^m}{(m+1)!}\right)\left(\sum_{n=0}^{\infty} \frac{B_n}{n!} z^n\right) = 1 \qquad \text{and to} \qquad \sum_{n=0}^{N} \frac{B_n}{n!\,(N-n+1)!} = \delta_{N,0} \quad \forall N \in \mathbb{N}_0,$$

from which $B_0 = 1, B_1 = -\frac{1}{2}, B_2 = \frac{1}{6}, \ldots$ can be determined recursively. As a consequence of the fact that $z \mapsto \frac{z}{e^z-1} + \frac{z}{2} = \frac{z}{2}\frac{e^z+1}{e^z-1}$ is even, one has $B_3 = B_5 = \cdots = 0$.

The function $z \mapsto \frac{z}{e^z-1}$ already played a role in the proof of Theorem 2.14, but there it did not lead to an appearance of the Bernoulli numbers. Now it will:

8.1 PROPOSITION *Let $H = \sum_{n=0}^\infty H_n'$ be the decomposition of $H = \log(e^x e^y) \in \mathbb{Q}\langle\langle x, y\rangle\rangle$ w.r.t. the $|\cdot|_x$-grading. Then $H_0' = y$ and*

$$H_1' = \frac{\mathrm{ad}_y}{e^{\mathrm{ad}_y} - 1}(x) = x - \frac{1}{2}\mathrm{ad}_y(x) + \sum_{m=1}^{\infty} \frac{B_{2m}}{(2m)!}\mathrm{ad}_y^{2m}(x). \tag{8.2}$$

Proof. Let $\phi : \mathbb{Q}\langle\langle x, y\rangle\rangle \to \mathbb{Q}\langle\langle x, y\rangle\rangle$ be the unique unital homomorphism such that $x \mapsto 0, y \mapsto y$. Clearly $\phi(F)$ is just the specialization $F_{|x=0}$, which is well-defined. The homomorphism ϕ is continuous and idempotent with image $\mathbb{Q}\langle\langle y\rangle\rangle \subset \mathbb{Q}\langle\langle x, y\rangle\rangle$. Since H_0' by definition is the part of the BCH series $H = \log(e^x e^y)$ involving only powers of y, but no x, we have $H_0' = \phi(H) = \log(e^y) = y$, proving the first claim.

Now let $D : \mathbb{Q}\langle\langle x, y\rangle\rangle \to \mathbb{Q}\langle\langle x, y\rangle\rangle$ be the linear map defined on words by $D(w) = w\delta_{|w|_x,1}$. Thus $D(w) = w$ if w contains x once, and $D(w) = 0$ otherwise. Now D is a (ϕ, ϕ)-derivation, i.e. satisfies $D(ww') = D(w)\phi(w') + \phi(w)D(w')$, since both sides equal ww' if $|ww'|_x = 1$ and both vanish otherwise. Now we are in a position to apply (7.3), obtaining

$$D(e^x)\phi(e^y) + \phi(e^x)D(e^y) = D(e^x e^y) = D(e^H) = \frac{e^{\mathrm{ad}_{\phi(H)}} - 1}{\mathrm{ad}_{\phi(H)}}(D(H))e^{\phi(H)}. \tag{8.3}$$

As shown above, $\phi(H) = y$, and clearly $\phi(e^x) = 1, \phi(e^y) = e^y$. Furthermore, $D(e^x) = x, D(e^y) = 0$, and $D(H) = H_1'$, so that (8.3) reduces to

$$xe^y = \frac{e^{\mathrm{ad}_y} - 1}{\mathrm{ad}_y}(H_1')e^y, \qquad \text{thus} \qquad x = \frac{e^{\mathrm{ad}_y} - 1}{\mathrm{ad}_y}(H_1').$$

Using again that the function $f(z) = \frac{e^z-1}{z}$ has $g(z) = \frac{z}{e^z-1}$ as multiplicative inverse, the above becomes

$$\frac{\mathrm{ad}_y}{e^{\mathrm{ad}_y} - 1}(x) = H_1',$$

which is the first identity in (8.2). The second follows from (8.1) and $B_1 = -\frac{1}{2}$, $B_3 = B_5 = \cdots = 0$. ■

8.2 REMARK 1. Note that H_1', as given by (8.2), is a Lie series, as was to be expected since a linear combination of iterated commutators remains one if some of the Lie summands (those of x-degree different from 1) are removed. (Cf. also Lemma 3.1(ii).)

2. Completely analogously one can consider an expansion $H = \sum_{n=0}^\infty H_n''$ in powers of y, i.e. $H_n'' \in \mathbb{F}\langle\langle x, y\rangle\rangle_n^{(y)}$ for all n. One then proves

$$H_1'' = y + \frac{1}{2}\mathrm{ad}_x(y) + \sum_{m=1}^{\infty} \frac{B_{2m}}{(2m)!}\mathrm{ad}_x^{2m}(y).$$

3. We know from the computations in the Introduction that

$$H = x + y - \frac{1}{2}[y, x] + \frac{1}{12}([y, [y, x]] + [x, [x, y]]) - \frac{1}{24}[y, [x, [y, x]]] + O((x + y)^5), \tag{8.4}$$

from which we immediately get

$$H_1' = x - \frac{1}{2}[y, x] + \frac{1}{12}[y, [y, x]] + O(xy^4). \tag{8.5}$$

In view of $B_2 = \frac{1}{6}, B_3 = 0$, equations (8.2) and (8.5) agree up to order ≤ 3 in y, as they must. Thus the $-\frac{1}{2}$ and $\frac{1}{12}$ in (8.4) really 'are' $\frac{B_1}{1!}$ and $\frac{B_2}{2!}$, respectively, and the vanishing of B_3 'explains' the absence in (8.4) of a term linear in x and cubic in y (or linear in y and cubic in x, by point 2. of the remark). □

We need a few more facts concerning derivations:

8.3 LEMMA *Let D be a derivation (i.e. $\phi = \mathrm{id}$) of an algebra $\mathcal{A}$. Then*

(i) For all $n \in \mathbb{N}_0$ and $X_1, \ldots, X_n \in \mathcal{A}$ we have

$$D(X_1 \cdots X_n) = D(X_1)X_2 \cdots X_n + X_1 D(X_2)X_3 \cdots X_n + \cdots + X_1 \cdots X_{n-1} D(X_n).$$

(ii) For each $n \in \mathbb{N}_0$ and $X, Y \in \mathcal{A}$ we have

$$D^n(XY) = \sum_{k=0}^{n} \binom{n}{k} D^k(X) D^{n-k}(Y).$$

(iii) The following linear map $\mathcal{A} \to \mathcal{A}$ is a continuous homomorphism:

$$\mu : X \mapsto \sum_{n=0}^{\infty} \frac{D^n(X)}{n!} =: e^D(X),$$

provided $\mathcal{A} = \mathbb{F}\langle\langle A \rangle\rangle$ and $D(a)$ has zero constant term $(D(a), 1)$ for all $a \in A$ (or $\mathcal{A}$ is a Banach algebra and D is bounded).

Proof. (i) For $n = 1$ this is trivial, for $n = 0$ it reduces to $D(1) = 0$ (Lemma 7.4), and for $n = 2$ this is just the definition of a derivation. Assume the statement holds for n, and let $X_1, \ldots, X_{n+1} \in \mathcal{A}$. Then

$$\begin{aligned} D(X_1 \cdots X_{n+1}) &= D(X_1 \cdots X_n)X_{n+1} + X_1 \cdots X_n D(X_{n+1}) \\ &= [D(X_1)X_2 \cdots X_n + \cdots + X_1 \cdots X_{n-1} D(X_n)]X_{n+1} + X_1 \cdots X_n D(X_{n+1}) \\ &= D(X_1)X_2 \cdots X_{n+1} + \cdots + X_1 \cdots X_n D(X_{n+1}) \end{aligned}$$

is the inductive step.

(ii) The proof is the same as for the well-known formula for $(fg)^{(n)}$ (which only uses the derivation property).

(iii) With Lemma 7.4 it is immediate that $\mu(1) = e^D(1) = 1$. We compute formally:

$$\begin{aligned} \mu(XY) &= \sum_{n=0}^{\infty} \frac{D^n(XY)}{n!} \overset{(ii)}{=} \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \binom{n}{k} D^k(X) D^{n-k}(Y) \\ &= \sum_{0 \le k \le n < \infty} \frac{D^k(X) D^{n-k}(Y)}{k!\,(n-k)!} = \sum_{k,\ell=0}^{\infty} \frac{D^k(X) D^\ell(Y)}{k!\,\ell!} = \mu(X)\mu(Y). \end{aligned}$$

In the Banach algebra case it is easy to see that $\|\mu(X)\| \le e^{\|D\|}\|X\| < \infty$, so that $\mu(X)$ is well-defined and the above computation is justified by absolute convergence. In the formal algebra case, we note that by (i) we have

$$D(a_1 \cdots a_n) = D(a_1)a_2 \cdots a_n + a_1 D(a_2)a_3 \cdots a_n + \cdots + a_1 \cdots a_{n-1} D(a_n).$$

Now the assumption on D implies that D maps $\mathbb{F}\langle A \rangle_k$ to $\mathbb{F}\langle\langle A \rangle\rangle_{\ge k}$. This implies that the infinite sum in the definition of μ is locally finite, thus well-defined. For the same reason the formal proof of $\mu(XY) = \mu(X)\mu(Y)$ is valid. ■

Given $S \in \mathbb{Q}\langle\langle x, y \rangle\rangle$, we can use the formula in Lemma 8.3(i) to define a unique continuous derivation $D : \mathbb{Q}\langle\langle x, y \rangle\rangle \to \mathbb{Q}\langle\langle x, y \rangle\rangle$ such that $D(x) = 0$ and $D(y) = S$. This derivation will be denoted $S\frac{\partial}{\partial y}$.

8.4 THEOREM *The BCH series $H = \log(e^x e^y)$ is given by*

$$H = \exp\left(H_1' \frac{\partial}{\partial y}\right)(y) = \sum_{n=0}^{\infty} H_n', \quad \text{where} \quad H_n' = \frac{1}{n!}\left(H_1' \frac{\partial}{\partial y}\right)^n (y) \;\in\; \mathbb{Q}\langle\langle x, y \rangle\rangle_n^{(x)}.$$

Proof. Applying the construction given before the theorem to $H'_1 \in \mathbb{Q}\langle\langle x,y\rangle\rangle$ as obtained in Proposition 8.1, we obtain a derivation $D = H'_1 \frac{\partial}{\partial y}$. With this definition, the second statement of the theorem is trivially true for $n=1$, as it must for consistency. Since H'_1 has no constant term, D satisfies the condition in Lemma 8.3(iii), so that we have a well-defined homomomorphism $\mu = e^D : \mathbb{Q}\langle\langle x,y\rangle\rangle \to \mathbb{Q}\langle\langle x,y\rangle\rangle$.

Now by Propositions 7.5 (with $\phi = \mathrm{id}$) and 8.1 we have

$$D(e^y) \stackrel{\text{P. 7.5}}{=} \frac{e^{\mathrm{ad}_y}-1}{\mathrm{ad}_y}(Dy)e^y \stackrel{Dy=H'_1}{=} \frac{e^{\mathrm{ad}_y}-1}{\mathrm{ad}_y}(H'_1)e^y \stackrel{\text{P. 8.1}}{=} \frac{e^{\mathrm{ad}_y}-1}{\mathrm{ad}_y}\left(\frac{\mathrm{ad}_y}{e^{\mathrm{ad}_y}-1}(x)\right)e^y \;=\; xe^y.$$

By induction, taking $D(x)=0$ into account, this gives $D^n(e^y) = x^n e^y$. Dividing by $n!$ and summing over n gives

$$\mu(e^y) = e^D(e^y) = \sum_{n=0}^\infty \frac{D^n}{n!}(e^y) = \sum_{n=0}^\infty \frac{x^n}{n!}e^y = e^x e^y,$$

thus $\mu(e^y) = e^x e^y$. Combining this with 'continuity' of the homomorphism μ, i.e. $\mu(e^y) = e^{\mu(y)}$, we obtain

$$e^{\mu(y)} = e^x e^y = e^H, \qquad \text{thus} \quad \mu(y) = H.$$

In view of the definition of μ, this is the first statement of the theorem. And $H = \sum_{n=0}^\infty H'_n$, where $H'_n = \frac{D^n(y)}{n!}$. Thus to prove the second half of the theorem, we need to show $D^n(y) \in \mathbb{Q}\langle\langle x,y\rangle\rangle_n^{(x)}$.

By definition, $D(y) = H'_1 \in \mathbb{Q}\langle\langle x,y\rangle\rangle_1^{(x)}$. By way of induction, assume $D^n(y) \in \mathbb{Q}\langle\langle x,y\rangle\rangle_n^{(x)}$. Then $D^n(y)$ is a linear combination of words w with $|w|_x = n$. Given such a word, by Lemma 8.3(i) we have $D(w) = \sum_{i=1}^n w_i$, where w_i is obtained from w by replacing the i-th instance of y by H'_1. Since H'_1 has x-degree one, this implies $D(w) \in \mathbb{Q}\langle\langle x,y\rangle\rangle_{n+1}^{(x)}$ and therefore $D^{n+1}(y) \in \mathbb{Q}\langle\langle x,y\rangle\rangle_{n+1}^{(x)}$. This completes the induction. ■

8.5 Remark 1. Also the preceding result has an analogue for the expansion $H = \sum_{n=0}^\infty H''_n$ in powers of y.

2. Using Theorem 8.4 for actual computations is quite painful. But the following is still quite doable: □

8.6 Exercise Use

$$H'_2 = \frac{1}{2!}\left(H'_1\frac{\partial}{\partial y}\right)^2(y) = \frac{1}{2}\left(H'_1\frac{\partial}{\partial y}\right)H'_1$$

to compute H'_2 to order two in (x and) y. Compare with Exercise 1.2.

Nowhere in the above discussion did we assume that $H = \log(e^x e^y)$ is a Lie series. Indeed it can be used to give yet another proof of the following:

8.7 Corollary (BCH) *The BCH series $H = \log(e^x e^y)$ is a Lie series.*

Proof. The claim clearly follows if we show that each H'_n from Theorem 8.4 is Lie. For H'_1 this is clear from Proposition 8.1. With D as in the proof of Theorem 8.4, we by construction have $H'_{n+1} = \frac{D(H'_n)}{n+1}$, so that it suffices to prove that D maps Lie series to Lie series. And indeed, since D is a derivation (of associative algebras), we have

$$D([X,Y]) = D(XY-YX) \;=\; D(X)Y + XD(Y) - D(Y)X - YD(X) = [D(X),Y] + [X,D(Y)],$$

which shows that D (is a Lie algebra derivation, thus) maps linear combinations of commutators to linear combinations of commutators, i.e. Lie polynomials/series. ■

The above algebraic proof of BCH is somewhat longer than those of Eichler and Friedrichs-Cartier(-Dynkin) given before, and certainly less elementary.

9 Dynkin's forgotten paper

As we know, the BCH series is the unique formal series $H \in \mathbb{Q}\langle\langle x,y\rangle\rangle$ such that $e^H = e^x e^y$. In principle, the expansions (1.3) and (1.4) contain all there is to know. We can, e.g., use them to compute to any desired order, or expand in powers of y for fixed order in x as in Section 8. But computationally these series are extremely inconvenient, and the expansion $H = \sum_n H'_n$ in powers of x of the preceding section is no real improvement.

In the remainder of these notes we will focus on expansions than are finer than $H = \sum_n H_n$ or $H = \sum_n H'_n$. The finest possible expansion clearly is $H = \sum_{w\in\{x,y\}^*} (H,w)\, w$. It is in fact trivial to extract a formula for the coefficients $(H,w) \in \mathbb{Q}$ from (1.3):

$$(H,w) = \sum_{k=1}^{|w|} \frac{(-1)^{k-1}}{k} \sum_{m_1+n_1>0} \cdots \sum_{m_k+n_k>0} \frac{\delta_{w,x^{m_1}y^{n_1}\cdots x^{m_k}y^{n_k}}}{m_1!\, n_1! \cdots m_k!\, n_k!} \qquad \forall w \in \{x,y\}^*,$$

where $\delta_{w,x^{m_1}y^{n_1}\cdots x^{m_k}y^{n_k}} = 1$ if the words w and $x^{m_1}y^{n_1}\cdots x^{m_k}y^{n_k}$ are equal and $= 0$ otherwise. Given a word w there can be many $k, m_1, n_1, \ldots, m_k, n_k$ such that $w = x^{m_1}y^{n_1}\cdots x^{m_k}y^{n_k}$, so that the best that can be said about this formula is that all summations are finite, since only $k, m_1, \ldots, n_k \le |w|$ contribute. A more explicit formula, albeit quite complicated, for the coefficients (H,w) has first been given by Goldberg [30] (1955) and elaborated upon or rediscovered by various authors, see in particular [62, 35, 49, 41], and we will discuss it in Section 10.

But since its statement and proof are involved, we will first, by way of introduction, give a complete account of a weaker, but prettier and quite simple, result due to Dynkin [19] (1949) that has been almost totally neglected.[12]

9.1 Dynkin's §1: Finely homogeneous expansion via permutation combinatorics

Following Dynkin we write

$$H = \sum_{p+q>0}^{\infty} H_{p,q}, \qquad \text{where} \qquad H_{p,q} = \sum_{\substack{w\in\{x,y\}^* \\ |w|_x=p,\, |w|_y=q}} (H,w)\, w,$$

i.e. $H_{p,q}$ is the *finely homogeneous* part of H of degree (p,q), consisting of linear combinations of words containing x and y precisely p and q times, respectively, but in any order. Dynkin has given fairly explicit expressions for $H_{p,q}$, see (9.13) and (9.2). This makes his result coarser than Goldberg's, but the ability to compute each coefficient (H,w) individually is of more theoretical than practical interest. (Nothing wrong with that!)

9.1 Remark 1. The BCH theorem implies $H_{n,0} = H_{0,n} = 0$ $\forall n \ge 2$ (since commutator expressions involving only x or y vanish). We already know the first non-vanishing terms:

$$H_{1,0} = x, \quad H_{0,1} = y, \quad H_{1,1} = \frac{1}{2}[x,y], \quad H_{1,2} = \frac{1}{12}[y,[y,x]], \quad H_{2,1} = \frac{1}{12}[x,[x,y]], \quad H_{2,2} = \frac{1}{24}[y,[x,[y,x]]].$$

2. By the BCH theorem, $H_n = \sum_{p=0}^{n} H_{p,n-p}$ is Lie for each n. But the above suggests that each $H_{p,q}$ is Lie on its own. This will follow from the theorem below, whose proof does not use the BCH theorem. But it is also easy to deduce this from BCH, adapting Lemma 3.1(ii): Let F be a homogeneous Lie polynomial of degree n. Thus $F = \sum_{i=1}^{I} c_i F_i$, where $c_i \in \mathbb{F}$ and each F_i is an iterated commutator in x, y. Expanding F_i into a linear combination of words $w \in \{x,y\}^*$, which differ only in the order of their letters, it is clear that each F_i is finely homogeneous. Thus if $p+q=n$ then $F_{p,q}$ is a linear combination of some of the F_i and therefore Lie.

3. Using Proposition 8.1 and Remark 8.2.2, we have very explicit (modulo computing B_n, that is) formulae for $H_{p,q}$ when one of the indices is one:

$$H_{1,q} = \frac{B_q}{q!}\mathrm{ad}_y^q(x), \qquad H_{p,1} = \frac{B_p}{p!}\mathrm{ad}_x^p(y) \qquad \forall p,q \ge 2, \tag{9.1}$$

[12]Perhaps due to the mistaken idea that [19] was a rehash of the much better known [18]. Adding insult to injury, essentially all authors citing [19] actually mean the results of [18]! The only genuine references to [19] of which I am aware are [1, pp. 347-350] and (perhaps, I haven't seen it yet) [66]. Loday [42] mentions some of Dynkin's results without, however, giving a reference to [19]. (The Zentralblatt review (Zbl 0041.16102) of [19] is quite good, but in German. Dynkin's [19] has never been translated from Russian to another language. But, then again, also [18] received a published translation only in 2000.)

(which vanishes for odd $p, q \geq 3$). The rest of these notes is concerned with proving analogous statements for $H_{p,q}$, where $p, q \geq 2$. This turns out to be considerably more involved. $\square$

9.2 Definition *For $n \in \mathbb{N}$, let S_n denote the permutation group of $\{1, \ldots, n\}$. For $\sigma \in S_n$, we define*

$$\begin{aligned}\mathrm{asc}(\sigma) &= \#\{i \in \{1, \ldots, n-1\} \mid \sigma(i) < \sigma(i+1)\}, \\ \mathrm{des}(\sigma) &= \#\{i \in \{1, \ldots, n-1\} \mid \sigma(i) > \sigma(i+1)\},\end{aligned}$$

the numbers of ascents *and* descents *of the permutation $\sigma \in S_n$. (For $n = 1$, we put $\mathrm{asc}(\sigma) = \mathrm{des}(\sigma) = 0$.)*

Note that for every $\sigma \in S_n$ we have $\mathrm{asc}(\sigma) + \mathrm{des}(\sigma) = n - 1$.

9.3 Theorem (Dynkin [19]) *For all p, q with $p + q > 0$ we have*

$$H_{p,q} = \frac{1}{p!\, q!\, (p+q)!\, (p+q)} R_{p+q}(\underbrace{x, \ldots, x}_{p \text{ times}}, \underbrace{y, \ldots, y}_{q \text{ times}}), \tag{9.2}$$

where

$$R_n(x_1, \ldots, x_n) = \sum_{\sigma \in S_n} (-1)^{\mathrm{des}(\sigma)} \mathrm{des}(\sigma)!\, \mathrm{asc}(\sigma)!\ [x_{\sigma(1)}, [x_{\sigma(2)}, [\cdots, [x_{\sigma(n-1)}, x_{\sigma(n)}] \cdots]. \tag{9.3}$$

9.4 Example It is a triviality to check that formulae (9.2-9.3) correctly reproduce $H_{1,0}$ and $H_{0,1}$. For $p = q = 1$, they give $\frac{1}{2!2} R_2(x, y) = \frac{1}{4}([x, y] - [y, x]) = \frac{1}{2}[x, y]$ (since $\mathrm{des}(\mathrm{id}) = 0$ and $\mathrm{des}(\sigma) = 1$ for the only non-trivial permutation $\sigma \in S_2$). To illustrate the theorem on a less trivial term, denote the permutations $\sigma \in S_3$ by $\sigma(1)\sigma(2)\sigma(3)$ and considering them in the order $123, 132, 213, 231, 312, 321$, we correctly find (once again):

$$\begin{aligned} H_{1,2} &= \frac{1}{2!\,3!\,3} (\cancel{2[x, [y, y]]} - \cancel{[x, [y, y]]} - [y, [x, y]] - [y, [y, x]] - [y, [x, y]] + 2[y, [y, x]]) \\ &= \frac{1}{36}([y, [y, x]] - [y, [y, x]] + [y, [y, x]] + 2[y, [y, x]]) = \frac{1}{12}[y, [y, x]]. \end{aligned}$$

$\square$

9.5 Exercise Use Theorem 9.3 to show that $H_{2,2} = \frac{1}{24}[y, [x, [y, x]]]$ (as in Exercises 1.2 and 8.6).

We will first prove Theorem 9.3 assuming the theorems of Baker-Campbell-Hausdorff and Dynkin-Specht-Wever. Then we will remove the dependency on the mentioned theorems, providing yet another proof of BCH.

9.6 Lemma *Define finely homogeneous $Q_I \in \mathbb{F}\langle x_1, \ldots, x_n \rangle$ of degree $I = (i_1, \ldots, i_n)$ by the expansion*

$$\log(e^{t_1 x_1} \cdots e^{t_n x_n}) = \sum_{I = (i_1, \ldots, i_n) \in \mathbb{N}_0^n} t_1^{i_1} \cdots t_n^{i_n}\, Q_I \ \in \mathbb{F}\langle\langle x_1, \ldots, x_n \rangle\rangle[[t_1, \ldots, t_n]]. \tag{9.4}$$

Then for each p, q with $p + q > 0$ we have

$$H_{p,q} = \frac{1}{p!\, q!} Q_{(\underbrace{1, \ldots, 1}_{p+q \text{ times}})}(\underbrace{x, \ldots, x}_{p \text{ times}}, \underbrace{y, \ldots, y}_{q \text{ times}}). \tag{9.5}$$

Proof. Fix $p, q \in \mathbb{N}_0$ and put $n = p + q$ and $x_1 = \cdots = x_p = x$ and $x_{p+1} = \cdots = x_n = y$ in (9.4). We obtain

$$\log(e^{(t_1 + \cdots + t_p)x} e^{(t_{p+1} + \cdots + t_n)y}) = \sum_{I = (i_1, \ldots, i_n) \in \mathbb{N}_0^n} t_1^{i_1} \cdots t_n^{i_n} Q_I(\underbrace{x, \ldots, x}_{p \text{ times}}, \underbrace{y, \ldots, y}_{q \text{ times}}). \tag{9.6}$$

For $p = q = 1$, this formula reduces to

$$\log(e^{t_1 x} e^{t_2 y}) = \sum_{I = (i_1, i_2) \in \mathbb{N}_0^2} t_1^{i_1} t_2^{i_2} Q_I(x, y), \tag{9.7}$$

which is just the finely homogeneous decomposition $\sum_{p,q} H_{p,q}$ of the BCH series (with t_1, t_2 introduced), thus $H_{p,q} = Q_{(p,q)}$. Replacing t_1, t_2 in (9.7) by $t_1+\cdots+t_p$ and $t_{p+1}+\cdots+t_n$, respectively, and calling the summation indices $J = (j_1, j_2)$ for safety, the left hand sides of (9.6) and (9.7) agree, so that we obtain the identity

$$\sum_{j_1,j_2} (t_1 + \cdots + t_p)^{j_1} (t_{p+1} + \cdots + t_n)^{j_2} H_{j_1,j_2} = \sum_{I=(i_1,\ldots,i_n)\in\mathbb{N}_0^n} t_1^{i_1} \cdots t_n^{i_n} \, Q_I(\underbrace{x, \ldots, x}_{p \text{ times}}, \underbrace{y, \ldots, y}_{q \text{ times}}) \tag{9.8}$$

in $\mathbb{F}\langle\langle x, y\rangle\rangle[[t_1, \ldots, t_n]]$. Thus the coefficients (living in $\mathbb{F}\langle x, y\rangle$) of the monomials $t_1^{i_1} \cdots t_n^{i_n}$ appearing on both sides coincide for all $I = (i_1, \ldots, i_n)$.

In particular, this holds for $t_1^1 \cdots t_n^1$. The polynomial $(t_1 + \cdots + t_p)^{j_1} (t_{p+1} + \cdots + t_n)^{j_2}$ on the l.h.s. is homogeneous in the t's with total degree $j_1 + j_2$. Thus it can have a summand $t_1^1 \cdots t_n^1$ (which has degree $n = p+q$) only if $j_1+j_2 = p+q$. So let us assume this. If $j_1 < p$ then none of the monomials in $(t_1+\cdots+t_p)^{j_1}$ can involve each $t_1, \ldots, t_p$. And if $j_1 > p$ then all monomials have a t_i with exponent > 1. Finally, if $(j_1, j_2) = (p, q)$, one checks easily that the coefficient of $t_1^1 \cdots t_n^1$ in $(t_1 + \cdots + t_p)^{j_1} (t_{p+1} + \cdots + t_n)^{j_2}$ is $p!\,q!$. Since the coefficient of $t_1^1 \cdots t_n^1$ on the r.h.s. of (9.8) clearly is given by

$$Q_{\underbrace{(1, \ldots, 1)}_{p+q \text{ times}}}(\underbrace{x, \ldots, x}_{p \text{ times}}, \underbrace{y, \ldots, y}_{q \text{ times}}),$$

we have proven (9.5). ■

From now on we abbreviate $Q_n = Q_{\underbrace{(1, \ldots, 1)}_{n \text{ times}}}$.

9.7 Proposition *For all $n \in \mathbb{N}$ we have*

$$Q_n(x_1, \ldots, x_n) = \frac{1}{n!} \sum_{\sigma\in S_n} (-1)^{\mathrm{des}(\sigma)} \mathrm{des}(\sigma)! \, \mathrm{asc}(\sigma)! \; x_{\sigma(1)} \cdots x_{\sigma(n)}. \tag{9.9}$$

Proof. By definition, $Q_n \in \mathbb{F}\langle x_1, \ldots, x_n\rangle$ is the coefficient of $t_1^1 \cdots t_n^1$ of the formal series $\log(e^{t_1 x_1} \cdots e^{t_n x_n})$. This means that in

$$\begin{aligned} \log(e^{t_1 x_1} \cdots e^{t_n x_n}) &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (e^{t_1 x_1} \cdots e^{t_n x_n} - 1)^k \\ &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \left(\sum_{m_1+\cdots+m_n>0} \frac{t_1^{m_1} \cdots t_n^{m_n} x_1^{m_1} \cdots x_n^{m_n}}{m_1! \cdots m_n!} \right)^k \end{aligned} \tag{9.10}$$

only contributions with $m_i \le 1$ for all i are relevant for the computation of $Q_n(x_1, \ldots, x_n)$. (For such m_i, we have $m_1! \cdots m_n! = 1$.) Furthermore we only need to consider $k \le n$ since summands with $k > n$ have total degree $> n$ in the t's due to the condition $m_1 + \cdots + m_n > 0$.

Collecting the terms in (9.10) in which each t_i appears exactly once, we thus have

$$Q_n(x_1, \ldots, x_n) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \sum_{W_k} w,$$

where $W_k \subset \{x_1, \ldots, x_n\}^*$ consists of those words w that contain each letter $x_1, \ldots, x_n$ precisely once and which can be written as concatenation $w = w_1 \cdots w_k$ of k non-empty words such that the letters in each word w_i appear in their natural order $x_1 < x_2 < \cdots < x_n$. (Such a word is called increasing.)

Since every word w containing each letter $x_1, \ldots, x_n$ exactly once can be written as $x_{\sigma(1)} \cdots x_{\sigma(n)}$ for a unique permutation $\sigma \in S_n$, we have

$$Q_n(x_1, \ldots, x_n) = \sum_{\sigma\in S_n} f(\sigma) \; x_{\sigma(1)} \cdots x_{\sigma(n)}, \qquad \text{where} \qquad f(\sigma) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \chi_{W_k}(w).$$

It remains to show that $f(\sigma) = (-1)^{\mathrm{des}(\sigma)} \frac{\mathrm{asc}(\sigma)!\,\mathrm{des}(\sigma)!}{n!}$.

Every word $w = x_{\sigma(1)} \cdots x_{\sigma(n)}$ can be written as a concatenation $w = w_1 \cdots w_k$ of increasing words, usually not uniquely. (For example, writing only the indices, $1264537 = (126)(45)(37) = (12)(6)(45)(3)(7) = \ldots$) Recall that $\mathrm{des}(\sigma)$ is the number of $i \in \{1, \ldots, n-1\}$ for which $\sigma(i) > \sigma(i+1)$. At these indices the word $w = x_{\sigma(1)} \cdots x_{\sigma(n)}$ fails to be increasing, and it is clear that any decomposition $w = w_1 \cdots w_k$ of w into increasing words must have breaks at these positions, so that $k \geq \mathrm{des}(\sigma) + 1$. There clearly is a unique minimal decomposition with $k = \mathrm{des}(\sigma)+1$. (In the example, this is $(126)(45)(37)$.) All decompositions into $k > \mathrm{des}(\sigma)+1$ increasing words arise from the minimal one by inserting $k - (\mathrm{des}(\sigma) + 1)$ additional breaks. This can be done at the $\mathrm{asc}(\sigma)$ positions that are ascents, which proves the formula

$$\begin{aligned} f(\sigma) &= \sum_{k=\mathrm{des}(\sigma)+1}^{n} \frac{(-1)^{k-1}}{k} \binom{\mathrm{asc}(\sigma)}{k - \mathrm{des}(\sigma) - 1} \\ &= (-1)^{\mathrm{des}(\sigma)} \sum_{m=0}^{\mathrm{asc}(\sigma)} \frac{(-1)^m}{m + \mathrm{des}(\sigma) + 1} \binom{\mathrm{asc}(\sigma)}{m}, \end{aligned} \tag{9.11}$$

where in the second step we substituted $k = \mathrm{des}(\sigma) + 1 + m$, with $m = 0, \ldots, n - 1 - \mathrm{des}(\sigma) = \mathrm{asc}(\sigma)$. Now the following lemma completes the proof. ■

9.8 Lemma *For all $s, t \in \mathbb{N}_0$ the following identities hold:*

$$\sum_{m=0}^{s} \frac{(-1)^m}{m+t+1} \binom{s}{m} = \int_0^1 x^t (1-x)^s \, dx = \frac{s!\, t!}{(s+t+1)!}. \tag{9.12}$$

Proof. With $\int_0^1 x^{m+t}\, dx = \frac{1}{m+t+1}$ and the binomial formula, applied to $(1-x)^s$, we have

$$\sum_{m=0}^{s} \frac{(-1)^m}{m+t+1} \binom{s}{m} = \int_0^1 \sum_{m=0}^{s} (-1)^m x^{m+t} \binom{s}{m} dx = \int_0^1 x^t (1-x)^s \, dx,$$

which is the first identity. For $s = 0$ and all $t \in \mathbb{N}_0$, the second identity is true since it reduces to $\int_0^1 x^t dx = \frac{1}{t+1}$. To do induction over s, assume it holds for a certain s and all t. Now for each $t \in \mathbb{N}_0$ partial integration gives

$$\begin{aligned} \int_0^1 x^t (1-x)^{s+1} \, dx &= \left[\frac{x^{t+1}}{t+1} (1-x)^{s+1} \right]_0^1 + (s+1) \int_0^1 \frac{x^{t+1}}{t+1} (1-x)^s dx \\ &\stackrel{\star}{=} \frac{s+1}{t+1} \frac{s!\,(t+1)!}{(s+(t+1)+1)!} = \frac{(s+1)!\, t!}{((s+1)+t+1)!}, \end{aligned}$$

where the starred equality comes from the induction hypothesis. This completes the proof of the second identity (which just is the evaluation of Euler's function $B(s+1, t+1)$ at positive integers). ■

9.9 Remark As in Section 7, the proof of the Lemma can easily be algebraized using the anti-derivative operator $I : \mathbb{Q}[x] \to \mathbb{Q}[x]$ defined by $x^n \mapsto \frac{x^{n+1}}{n+1}$ and an algebraic version of partial integration. We leave the details to the reader. □

Combining Lemma 9.6 and Proposition 9.7 we now have (with $n = p + q$)

$$H_{p,q} = \frac{1}{p!\,q!} Q_{p+q}(\underbrace{x, \ldots, x}_{p \text{ times}}, \underbrace{y, \ldots, y}_{q \text{ times}}) = \frac{1}{p!\,q!\,(p+q)!} \sum_{\sigma \in S_n} (-1)^{\mathrm{des}(\sigma)} \mathrm{des}(\sigma)!\, \mathrm{asc}(\sigma)!\; X_{\sigma(1)} \cdots X_{\sigma(n)}, \tag{9.13}$$

where $X_1 = \cdots = X_p = x$ and $X_{p+1} = \cdots = X_n = y$. This improves on (1.3). We briefly interrupt our consideration of BCH matters to prove an interesting combinatorial corollary.

9.10 Definition *For $n \in \mathbb{N}$ and $m \in \mathbb{Z}$ we define the* Eulerian number $A(n,m)$ *(also denoted $\left\langle {n \atop m} \right\rangle$) by*

$$A(n,m) = \#\{\sigma \in S_n \mid \mathrm{des}(\sigma) = m\}.$$

Note that, automatically, $A(n,m) = 0$ if $m < 0$ or $m \geq n$. And counting ascents instead would give the same numbers since composition of a permutation σ with the permutation $1 \cdots n \mapsto n \cdots 1$ is a bijection of S_n that exchanges ascents and descents. This observation is equivalent to noting the symmetry $A(n,m) = A(n, n-1-m)$.

9.11 Corollary *For all $n \geq 2$ we have*

$$\sum_{m=0}^{n-1} (-1)^m A(n,m)\, m!\, (n-1-m)! \;=\; 0, \tag{9.14}$$

$$\sum_{m=0}^{n-1} (-1)^m A(n,m)\, m!\, (n-m)! \;=\; (n+1)!\, B_n. \tag{9.15}$$

Proof. Put $x_1 = \cdots = x_n = x$ in Proposition 9.7. (More rigorously, pass to the quotient algebra $\mathbb{F}\langle x_1, \ldots, x_n\rangle / I$, where I is the two-sided ideal generated by all $x_i - x_j$.) Then $e^{t_1 x} \cdots e^{t_n x} = e^{(t_1 + \cdots + t_n)x}$, thus $\log(e^{t_1 x} \cdots e^{t_n x}) = t_1 + \cdots + t_n$. For $n \geq 2$, this has no monomial $t_1 \cdots t_n$, thus $Q_n(x, \ldots, x) = 0$. Thus the r.h.s. in Proposition 9.7 vanishes. But the latter equals

$$\frac{x^n}{n!} \sum_{\sigma \in S_n} (-1)^{\mathrm{des}(\sigma)} \mathrm{des}(\sigma)!\, \mathrm{asc}(\sigma)! = \frac{x^n}{n!} \sum_{m=0}^{n-1} A(n,m)(-1)^m m!\, (n-1-m)!,$$

(since $\mathrm{asc}(\sigma) = n - 1 - \mathrm{des}(\sigma)$), thus the sum on the right vanishes, proving (9.14).

As to the second claim, it will follow from the comparison of two different ways of computing the coefficient (H, xy^n) of the BCH series. On the one hand, since xy^n is of order one in x, it is a summand of H_1' computed in Proposition 8.1. Its coefficient is easily seen to be

$$(H, xy^n) = \frac{(-1)^n B_n}{n!} = \frac{B_n}{n!} \qquad \forall n \geq 2, \tag{9.16}$$

where we took into account that B_n vanishes for odd n.

On the other hand, xy^n is a summand of $H_{1,n}$. Equation (9.13) specializes to

$$H_{1,n} = \frac{1}{n!\,(n+1)!} \sum_{\sigma \in S_{n+1}} (-1)^{\mathrm{des}(\sigma)} \mathrm{des}(\sigma)!\, \mathrm{asc}(\sigma)!\; X_{\sigma(1)} \cdots X_{\sigma(n+1)},$$

where $X_1 = x, X_2 = \cdots = X_{n+1} = y$. The coefficient of xy^n in this sum, thus in H, clearly is

$$(H, xy^n) = \frac{1}{n!\,(n+1)!} \sum_{\substack{\sigma \in S_{n+1} \\ \sigma(1)=1}} (-1)^{\mathrm{des}(\sigma)} \mathrm{des}(\sigma)!\, \mathrm{asc}(\sigma)!.$$

There is an obvious bijection between the $\sigma \in S_{n+1}$ with $\sigma(1) = 1$ and the $\sigma' \in S_n$, given by $\sigma'(i) = \sigma(i+1) - 1\ \forall i = 1, \ldots, n$. For corresponding σ, σ' we have $\mathrm{des}(\sigma) = \mathrm{des}(\sigma')$ and $\mathrm{asc}(\sigma) = \mathrm{asc}(\sigma') + 1$. Thus

$$\begin{aligned}(H, xy^n) &= \frac{1}{n!\,(n+1)!} \sum_{\sigma' \in S_n} (-1)^{\mathrm{des}(\sigma')} \mathrm{des}(\sigma')!\, (\mathrm{asc}(\sigma') + 1)! \\ &= \frac{1}{n!\,(n+1)!} \sum_{m=0}^{n-1} A(n,m)(-1)^m m!\, (n-m)!. \end{aligned} \tag{9.17}$$

Comparing (9.16) and (9.17) gives (9.15). ■

9.12 Remark 1. A direct proof of the corollary, using only the definition of $A(n,m)$ and B_n, will be given in Appendix E, but the above manifestly non-commutative proof perhaps provides more insight.

2. Euler, who introduced the Eulerian numbers to compute power sums $\sum_{k=1}^n a^k k^p$ more general than those covered by the Faulhaber/Bernoulli formula for $\sum_{k=1}^n k^p$, defined them in terms of their generating function, see (E.2). Eq. (9.15) is also very classical, appearing e.g. in [69] (as the first displayed formula on p. 222. I thank Mathoverflow user Efinat-S for providing the reference). Worpitzky defined the Eulerian numbers in yet another manner (via what is now called Worpitzky's identity). The modern definition of the Eulerian numbers in terms of permutations is much more recent. It was discovered in the early 1950's by Carlitz and Riordan. Nowadays there are various proofs, more or less manifestly combinatorial, of the equivalence of the various definitions, see e.g. pages 6-13 in [44].

3. Eq. (9.15) is interesting since it connects the two types of combinatorics, at first sight quite different, involved in our two ways of computing (H, xy^n): On the one hand the commutative world of difference calculus, the Euler-Maclaurin theorem and related matters, on the other the combinatorics of permutations. But such a distinction has been obsolete since the 19th century, cf. Remark E.3, and recent developments, like higher dimensional Euler-Maclaurin theorems, underscore this further. □

Proof of Theorem 9.3, assuming the BCH and Dynkin-Specht-Wever theorems. Since we assume BCH, by Remark 9.1.2 each $H_{p,q}$ is a Lie polynomial, thus also the right hand side of (9.13) is Lie. And since we assume DSW, both sides are invariant under the Dynkin idempotent P from Corollary 6.8, so that

$$\begin{aligned} H_{p,q} &= \frac{1}{p!\,q!\,(p+q)!}\sum_{\sigma\in S_n}(-1)^{\mathrm{des}(\sigma)}\mathrm{des}(\sigma)!\,\mathrm{asc}(\sigma)!\;P(X_{\sigma(1)}\cdots X_{\sigma(n)}) \\ &= \frac{1}{p!\,q!\,(p+q)!\,(p+q)}\sum_{\sigma\in S_n}(-1)^{\mathrm{des}(\sigma)}\mathrm{des}(\sigma)!\,\mathrm{asc}(\sigma)!\;[X_{\sigma(1)},[X_{\sigma(2)},[\cdots,[X_{\sigma(n-1)},X_{\sigma(n)}]\cdots] \\ &= \frac{1}{p!\,q!\,(p+q)!\,(p+q)}R_{p+q}(\underbrace{x,\ldots,x}_{p\text{ times}},\underbrace{y,\ldots,y}_{q\text{ times}}), \end{aligned}$$

where R_n is as in Theorem 9.3, finishing the proof of the latter. ■

9.2 Dynkin's §2. Fifth algebraic proof of BCH

Having proven BCH (many times) and Dynkin-Specht-Wever (just once), we could stop at this point. But Dynkin goes on to finish the proof of Theorem 9.3 *without* assuming any of the mentioned results. We will follow him since this involves nice mathematics (and since having only one proof of Dynkin-Specht-Wever is not satisfactory according to footnote 2).

Since (9.13) was proven without using any deeper results, it suffices to find a self-contained proof of

$$\begin{aligned} &\frac{1}{n}\sum_{\sigma\in S_n}(-1)^{\mathrm{des}(\sigma)}\mathrm{asc}(\sigma)!\,\mathrm{des}(\sigma)!\;[X_{\sigma(1)},[X_{\sigma(2)},[\cdots,[X_{\sigma(n-1)},X_{\sigma(n)}]\cdots] \\ &\qquad = \sum_{\sigma\in S_n}(-1)^{\mathrm{des}(\sigma)}\mathrm{asc}(\sigma)!\,\mathrm{des}(\sigma)!\;X_{\sigma(1)}\cdots X_{\sigma(n)}, \end{aligned} \tag{9.18}$$

which is the content of §2 of [19]. **************************************

Now that we have finished the proof of (9.18), and therefore of Theorem 9.3, without using the BCH and DSW theorems, the fact that $R_n(x,\ldots,x,y,\ldots,y)$ by definition is a Lie polynomial gives yet another proof of:

9.13 Corollary (BCH) *The formal series* $\log(e^x e^y)\in\mathbb{Q}\langle\langle x,y\rangle\rangle$ *is a Lie series.*

9.14 Remark 1. Let us be explicit about what can be found in Dynkin's [19]: From Section 9.1 it contains the material from Definition 9.2 (with the terminology 'correct' and 'wrong' instead of ascents and descents!) through Proposition 9.7. His proofs are short, but sufficient to construct complete ones. He doesn't emphasize that Theorem 9.3 already follows from §1 of his paper together with the BCH and DSW theorems, presumably since it is obvious and he has bigger fish to fry. Nor does he have Corollary 9.11, which is not surprising since the connection between permutations and Eulerian numbers wasn't known yet. As to Section 9.2, ****************

2. One could say that we now have $3 \times 2 + 5 = 11$ proofs of the BCH theorem. Here 3 is the number of proofs of Proposition 2.7, the 2 refers to the analytic proofs of BCH(D) given in Theorem 2.14 and in Remark 2.15.2, whereas the 5 refers to the algebraic proofs in Sections 4, 6.2, 7, 8 and just above. □

10 Goldberg's theorem and its ramifications

Here we will state the generalization of Goldberg's result to $\log(e^{x_1} \cdots e^{x_n})$ and give a complete proof, including the Eulerian connection missing in [30, 41]. (This area is rife with rediscoveries of earlier results: Dynkin's [19] is ignored by everyone, the author of [35] seems unaware of [30], as the authors of [41] are unaware of [35], etc.)

I intend to do this in a way that builds upon the material of Section 9 as much as possible.

****

While Goldberg's result is finer than Dynkin's [19] in that it allows to compute each individual coefficient (H, w), it does *not* lead to a new proof of the Lie nature of H. But since we have proven the latter sufficiently often, we can safely apply Corollary 6.8 to give a Lie representation of the BCH series H as

$$H = \sum_{w \in A^*} \frac{(H, w)}{|w|} R(w).$$

A Reinsch's algorithm for computing the BCH series

In this paper, $\mathbb{N} = \{1, 2, \ldots\}, \mathbb{N}_0 = \{0, 1, 2, \ldots\}$. Expanding the BCH series (1.3) by order of the terms (in powers of X and Y), we obtain $H(X, Y) = \sum_{n=1}^{\infty} H_n$ with

$$H_n = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k} \sum \frac{X^{m_1} Y^{n_1} \cdots X^{m_k} Y^{n_k}}{m_1!\, n_1! \cdots m_k!\, n_k!}, \tag{A.1}$$

where the second summation extends over the $(m_1, n_1, \ldots, m_k, n_k) \in \mathbb{N}_0^{2k}$ satisfying $\sum_i (m_i + n_i) = n$ and $m_i + n_i > 0\ \forall i$. The expression (A.1) makes sense for $X, Y \in \mathcal{A}$ in any associative unital algebra (in characteristic zero), since convergence questions do not arise.

In this section we will briefly discuss a computational approach due to Reinsch [48] that makes the computation of the homogeneous contributions H_n somewhat less painful than using (1.3) directly. The main virtue of his prescription is that it can be implemented very easily on a computer. (Reinsch concedes that his algorithm is not faster than others based on (1.3).)

Let $n \in \mathbb{N}$. Define $(n+1) \times (n+1)$-matrices M, N with entries in $\mathbb{Z}[t_1, \ldots, t_n]$ by

$$M_{ij} = \delta_{i+1,j}, \qquad N_{ij} = \delta_{i+1,j} t_i.$$

These matrices are strictly upper triangular, thus nilpotent with $M^{n+1} = N^{n+1} = 0$. Define $(n+1) \times (n+1)$-matrices F, G with entries in $\mathbb{Q}[t_1, \ldots, t_n]$ by

$$F = e^M = \sum_{k=0}^{n} \frac{M^k}{k!}, \qquad G = e^N = \sum_{k=0}^{n} \frac{N^k}{k!}.$$

As a consequence of $(M^k)_{i,j} = \delta_{i+k,j}$ one immediately has

$$F_{ij} = \frac{1}{(j-i)!}, \qquad G_{ij} = \frac{1}{(j-i)!} \prod_{k=i}^{j-1} t_k$$

with the understanding that $\frac{1}{n!} = 0$ for all $n < 0$.

For $n = 3$ all this amounts to:

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 \\ & 0 & 1 & 0 \\ & & 0 & 1 \\ & & & 0 \end{pmatrix}, \quad N = \begin{pmatrix} 0 & t_1 & 0 & 0 \\ & 0 & t_2 & 0 \\ & & 0 & t_3 \\ & & & 0 \end{pmatrix},$$

$$F = \begin{pmatrix} 1 & 1 & \frac{1}{2!} & \frac{1}{3!} \\ & 1 & 1 & \frac{1}{2!} \\ & & 1 & 1 \\ & & & 1 \end{pmatrix}, \quad G = \begin{pmatrix} 1 & t_1 & \frac{1}{2!}t_1t_2 & \frac{1}{3!}t_1t_2t_3 \\ & 1 & t_2 & \frac{1}{2!}t_2t_3 \\ & & 1 & t_3 \\ & & & 1 \end{pmatrix}.$$

Now, FG is upper triangular, with 1's on the diagonal, so that $(FG - \mathbf{1})^{n+1} = 0$[13]. Therefore

$$\log(FG) = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k}(FG - \mathbf{1})^k \tag{A.2}$$

is a finite sum. ($\mathbf{1}$ is the unit matrix.) Let $A_n = [\log(FG)]_{1,n+1} \in \mathbb{Q}[t_1, \dots, t_n]$ be the top right matrix element.

We call a polynomial $A = \sum_{e=(e_1,\dots,e_n)} c_e\, t_1^{e_1} \cdots t_n^{e_n} \in \mathbb{Q}[t_1, \dots, t_n]$ multiplicity-free if $c_e \neq 0 \;\Rightarrow\; e_i \in \{0, 1\}$ for all $i = 1, \dots, n$. For a multiplicity-free polynomial $A = \sum_{e=(e_1,\dots,e_n)} c_e\, t_1^{e_1} \cdots t_n^{e_n} \in \mathbb{Q}[t_1, \dots, t_n]$, we define

$$Q(A) = \sum_{e \in \{0,1\}^n} c_e\, X_{e_1} \cdots X_{e_n}, \qquad \text{where} \qquad X_0 = X,\ X_1 = Y.$$

A.1 Theorem (Reinsch [48]) *For each $n \in \mathbb{N}$ the following holds:*

(i) All monomials $t_1^{e_1} \cdots t_n^{e_n}$ appearing in A_n are multiplicity-free.

(ii) $Q(A_n) = H_n$.

Proof. (i) By linearity, it is more than sufficient to check that every matrix element of $(FG - \mathbf{1})^k$ is multiplicity-free. One easily checks that

$$(FG - \mathbf{1})_{ij} = \begin{cases} \displaystyle\sum_{i \le r \le j} \frac{1}{(r-i)!} \frac{1}{(j-r)!} \prod_{\ell=r}^{j-1} t_\ell & \text{if } \ i < j \\ 0 & \text{if } i \ge j \end{cases}$$

Thus the non-zero matrix elements $(FG - \mathbf{1})_{ij}$ with $i < j$ are linear combinations of multiplicity-free products of t_ℓ's with $i \le \ell < j$. Any term contributing to $[(FG - \mathbf{1})^k]_{ij}$ is a sum of products

$$(FG - \mathbf{1})_{ii_1}(FG - \mathbf{1})_{i_1 i_2} \cdots (FG - \mathbf{1})_{i_{k-1}j} \qquad \text{with} \qquad i < i_1 < i_2 < \cdots < i_{k-1} < j.$$

Since the different factors involve different t_ℓ's, this product is multiplicity-free.

(ii) We need to prove that $Q(A_n)$ coincides with (A.1). Both expressions are sums over $k = 1, \dots, n$, and we will prove the stronger statement that the respective contributions agree for each $k \le n$ separately, thus

$$Q([(FG - \mathbf{1})^k]_{1,n+1}) = \sum \frac{X^{m_1} Y^{n_1} \cdots X^{m_k} Y^{n_k}}{m_1!\, n_1! \cdots m_k!\, n_k!}. \tag{A.3}$$

Here the sum on the r.h.s. extends over all non-negative indices $m_1, n_1, \dots, m_k, n_k$ satisfying $m_s + n_s > 0$ for all s, as well as the total degree condition $\sum_{s=1}^{k}(m_s + n_s) = n$. By the observations in the proof of (i), the l.h.s. of (A.3) is given by

$$Q\left(\sum_{1=i_0<i_1<i_2<\cdots<i_{k-1}<i_k=n+1} \prod_{s=1}^{k} \left(\sum_{i_{s-1} \le r \le i_s} \frac{1}{(r - i_{s-1})!} \frac{1}{(i_s - r)!} \prod_{\ell=r}^{i_s - 1} t_\ell \right) \right).$$

[13] Thus FG is unipotent.

The variable r is local, thus there is a different one for each value of s. Expanding the product over s gives

$$Q\left(\sum_{1=i_0\le r_1\le i_1\le r_2\le i_2\le\cdots\le i_{k-1}\le r_k\le i_k=n+1}^{\prime}\prod_{s=1}^{k}\frac{1}{(r_s-i_{s-1})!}\frac{1}{(i_s-r_s)!}\prod_{\ell=r_s}^{i_s-1}t_\ell\right). \tag{A.4}$$

Here the accent above the summation symbol represents the additional condition that for each s we must have $i_{s-1}<i_s$. Now we notice that the assignment $(i_\bullet,r_\bullet)\mapsto(m_\bullet,n_\bullet)$ given by

$$m_s=r_s-i_{s-1},\qquad n_s=i_s-r_s\qquad\forall s=1,\ldots,k$$

is a bijection between the summation variables in (A.3) and (A.4), respectively: The condition $m_s+n_s>0$ corresponds to $i_{s-1}<i_s$, and $\sum_s(m_s+n_s)=n$ corresponds to $i_k-i_0=(n+1)-1=n$. With this correspondence it is immediate that the products of factorials appearing in (A.3) and (A.4) coincide. Finally, observing

$$Q\left(\prod_{s=1}^{k}\prod_{\ell=r_s}^{i_s-1}t_\ell\right)=X^{r_1-i_0}Y^{i_1-r_1}\cdots X^{r_k-i_{k-1}}Y^{i_k-r_k}=X^{m_1}Y^{n_1}\cdots X^{m_k}Y^{n_k},$$

since $t_1,\ldots,t_{r_1-1}$ (note that $r_1-i_0=m_1$) are missing from the product, $t_{r_1},\ldots,t_{i_1-1}$ are present (and $i_1-r_1=n_1$), etc., the proof of (A.3) is complete. ∎

A.2 Remark 1. Reinsch also discusses two generalizations: Instead of $\log(e^xe^y)$ one can consider the series expansion of $\log(f(x)f(y))$ for an arbitrary formal power series f with constant term 1. One simply replaces the above F,G by $F=f(M),G=f(N)$. The proof remains essentially the same.

One can also consider more than two formal variables, i.e. compute $\log(e^{X_1}\cdots e^{X_n})$. This requires introducing additional sets of variables besides $t_1,\ldots,t_n$, diminishing somewhat the elegance of the approach. And one can combine both generalizations.

2. Reinsch's result is one of many instances in combinatorics where a problem can be approached by linear algebra methods, as e.g. in the solution of linear difference equations. But from a theoretical point of view, it is a dead end as it adds few new insights, in particular it does not yield the Lie nature of the BCH series. □

B Algebraic proof of $\exp\circ\log=\mathrm{id}$ and $\log\circ\exp=\mathrm{id}$

In this section we exclusively consider $\mathbb{F}[[x]]=\mathbb{F}\langle\langle x\rangle\rangle$, i.e. formal power series in one variable, where $\mathbb{F}$ is a field of characteristic zero.

B.1 Theorem *Define $f,g\in\mathbb{F}[[x]]_{>0}$ by $f(x)=\exp(x)-1$ and $g(x)=\log(1+x)$. Then $f\circ g=\mathrm{id}=g\circ f$.*

This can be proven in many ways, for example using pedestrian direct computation (which is tedious and not very illuminating) or the Lagrange inversion formula. We choose an intermediate route.

B.2 Definition *For $f(x)=\sum_{n=0}^\infty a_nx^n\in\mathbb{F}[[x]]$, we define $f'\in\mathbb{F}[[x]]$ by $f'(x)=\sum_{n=0}^\infty na_nx^{n-1}$.*

B.3 Example For $f(x)=\sum_{n=0}^\infty x^n/n!$ we immediately get $f'=f$. For

$$g(x)=\log(1+x)=\sum_{n=1}^\infty\frac{(-1)^{n-1}}{n}x^n$$

we find

$$g'(x)=\sum_{n=1}^\infty\frac{(-1)^{n-1}}{n}nx^{n-1}=\sum_{n=1}^\infty(-1)^{n-1}x^{n-1}=\sum_{n=0}^\infty(-x)^n=(1+x)^{-1}.$$

□

The following is immediate:

B.4 Lemma *(i) The map $\mathbb{F}[[x]] \to \mathbb{F}[[x]],\ f \mapsto f'$ is linear and continous.*
(ii) If $f(x) = \sum_{n=0}^{\infty} a_n x^n \in \mathbb{F}[[x]]$ satisfies $f' = 0$ then f is constant, i.e. $a_n = 0$ for all $n \geq 1$.

B.5 Lemma (Product rule or derivation property) *Let $f, g \in \mathbb{F}[[x]]$. Then $(fg)' = f'g + fg'$.*

Proof. By linearity of the claim w.r.t. f and g, it suffices to prove it for $f(x) = x^n$, $g(x) = x^m$. Then

$$(fg)' = (x^{n+m})' = (n+m)x^{n+m-1} = (nx^{n-1})x^m + x^n(mx^{m-1}) = f'g + fg'.$$

(This, and similar arguments below, works despite the fact that $\{x^n\}$ is not a Hamel basis of $\mathbb{F}[[x]]$.) ■

B.6 Lemma (Chain rule) *Let $f \in \mathbb{F}[[x]],\ g \in \mathbb{F}[[x]]_{>0}$. Then*

$$(f \circ g)' = (f' \circ g) \cdot g'.$$

Proof. By linearity of the claim w.r.t. f, it suffices to prove it for $f(x) = x^n$, for which the claim reduces to $(g^n)' = ng^{n-1}g'$. For $n = 0, n = 1$ this reduces to $0 = 0$ and $g' = g'$, respectively. Assume the claim holds for n. Then it follows for $n + 1$ by

$$(g^{n+1})' = (g^n g)' = (g^n)'g + g^n g' = ng^{n-1}g'g + g^n g' = ng^n g' + g^n g' = (n+1)g^n g',$$

where we have used the product rule in the second equality and the induction hypothesis in the third. ■

If $f = \sum_{n=0}^{\infty} a_n x^n \in \mathbb{F}[[x]]$ then $[x^n]f$ denotes the coefficient a_n of x^n.

B.7 Proposition (Compositional inverses) *For $f \in \mathbb{F}[[x]]_{>0}$ the following are equivalent:*

(i) $[x^1]f \neq 0$.

(ii) f has a left inverse $g \in \mathbb{F}[[x]]_{>0}$, i.e. $g \circ f = \mathrm{id}$.

(iii) f has a unique inverse $g \in \mathbb{F}[[x]]_{>0}$, i.e. $g \circ f = \mathrm{id} = f \circ g$.

Proof. If $f = \sum_{n=1}^{\infty} a_n x^n$ and $g = \sum_{n=1}^{\infty} b_n x^n$, recall from Section 3 that

$$g \circ f = \sum_{n=1}^{\infty} b_n \left(\sum_{m=1}^{\infty} a_m x^m \right)^n = \sum_{k=1}^{\infty} c_k x^k \qquad \text{with} \qquad c_k = \sum_{\ell=1}^{k} b_\ell \sum_{m_1+\cdots+m_\ell=k} a_{m_1} \cdots a_{m_\ell},$$

in particular $c_1 = a_1 b_1$. Thus $g \circ f = \mathrm{id}$, to wit $g \circ f(x) = x$, is equivalent to $c_k = \delta_{k,1}\ \forall k$.

(ii)⇒(i) If $g \circ f = \mathrm{id}$ then $a_1 b_1 = c_1 = 1$, thus $[x^1]f = a_1 \neq 0$.

(i)⇒(ii) Since $a_1 \neq 0$ by assumption, we take $b_1 = 1/a_1$. For $k \geq 2$, we want

$$b_k a_1^k + \sum_{\ell=1}^{k-1} b_\ell \sum_{m_1+\cdots+m_\ell=k} a_{m_1} \cdots a_{m_\ell} = \sum_{\ell=1}^{k} b_\ell \sum_{m_1+\cdots+m_\ell=k} a_{m_1} \cdots a_{m_\ell} = c_k = 0,$$

which allows to determine the b_k recursively, again using $a_1 \neq 0$.

(iii)⇒(ii) is trivial.

(ii)⇒(iii) As noted in the proof of (ii)⇒(i), the left inverse g of f satisfies $[x^1]g \neq 0$. Thus by (i)⇒(ii), g has a left inverse h. Since clearly f is a right inverse of g, we have $h = h \circ g \circ f = f$. Thus f is a left inverse of g, so that g is a right inverse of f, as was to be shown. This argument also proves uniqueness of inverses. ■

Proof of Theorem B.1. By Example B.3, $f' = f + 1$ and $g'(x) = 1/(1+x)$. Using the chain rule, we compute

$$(g \circ f)' = (g' \circ f)f' = \frac{1}{1+f} f' = \frac{f+1}{1+f} = 1,$$

thus $(g \circ f - x)' = 0$. Combining this with $(g \circ f - x)(0) = 0$, Lemma B.4(ii) gives $g \circ f - x = 0$, thus we have proven $g \circ f = \mathrm{id}$. Now the implication (ii)⇒(iii) of Proposition B.7 gives $f \circ g = \mathrm{id}$. ■

B.8 Remark If we try to prove $(f\circ g)'=1$ as we did for $(g\circ f)'$, we run into

$$(f\circ g)'=(f'\circ g)g'=(f\circ g+1)(x)(1+x)^{-1},$$

which is equal to 1 if and only if $f\circ g=\mathrm{id}$, so that this attempt leads nowhere. □

C Combinatorial proof of $\exp\circ\log=\mathrm{id}$ and $\log\circ\exp=\mathrm{id}$

D More on the exponential function in a Banach algebra

We have seen that in a unital Banach algebra $XY=YX$ implies $e^{X+Y}=e^Xe^Y=e^Ye^X$. There is a partial converse:

D.1 Lemma *Let $\mathcal{A}$ be a unital Banach algebra and $X,Y\in\mathcal{A}$. Then the following are equivalent:*

(i) $XY=YX$.

(ii) There are subsets $S,T\subseteq\mathbb{R}$ both having accumulation points such that $e^{sX+tY}=e^{sX}e^{tY}$ for all $s\in S,t\in T$.

(iii) $e^{sX+tY}=e^{sX}e^{tY}\ \forall s,t\in\mathbb{R}$.

Proof. (i)⇒(iii) Apply Proposition 2.2(i).

(iii)⇒(i) Both sides of the equation in (iii) are invertible, and taking inverses we obtain $e^{-(sX+tY)}=e^{-tY}e^{-sX}$. Replacing s and t by $-s$ and $-t$, respectively, gives $e^{sX+tY}=e^{tY}e^{sX}$, and combining this with (iii) we have $e^{sX}e^{tY}=e^{tY}e^{sX}$. Applying $\frac{\partial^2}{\partial s\partial t}$ to both sides and putting $s=t=0$ gives $XY=YX$.

(iii)⇒(ii) is trivial.

(ii)⇒(iii) By assumption, for every fixed $t\in T$, the map $S\to\mathcal{A}$, $s\mapsto e^{-sX}e^{sX+tY}=e^{tY}$ is constant. Now a well-known argument involving power series, which also works for series with coefficients in a Banach algebra, implies that $s\mapsto e^{-sX}e^{sX+tY}=e^{tY}$ is constant for all $s\in\mathbb{R}$. Thus $e^{sX+tY}=e^{sX}e^{tY}$ holds for all $s\in\mathbb{R}$ and $t\in T$. In the same fashion one extends this to all $t\in\mathbb{R}$. ■

Remarkably, the above result can be strengthened considerably in the case of matrices, cf. [56], in that also the discrete sets $S=T=\mathbb{Z}$ work in (ii)! But $e^{X+Y}=e^Xe^Y=e^Ye^X$ does not imply $XY=YX$:

D.2 Lemma *Let $A=\begin{pmatrix} d_1 & w\\ 0 & d_2\end{pmatrix}$. If $d_1\neq d_2$ but $e^{d_1}=e^{d_2}$ (equivalently, $\frac{d_2-d_1}{2\pi i}\in\mathbb{Z}\backslash\{0\}$) then $e^A=e^{d_1}\mathbf{1}$.*

Proof. The eigenvalues of A evidently are d_1,d_2. If X_1,X_2 are corresponding eigenvectors, they are linearly independent since $d_1\neq d_2$. Thus the 2×2 matrix S having X_1,X_2 as columns is invertible, and $A=S\,\mathrm{diag}(d_1,d_2)S^{-1}$. Now $e^A=S\,\mathrm{diag}(e^{d_1},e^{d_2})S^{-1}$. Since $e^{d_1}=e^{d_2}$, the matrix in the middle is $e^{d_1}\mathbf{1}$, which commutes with S, so that $e^A=e^{d_1}\mathbf{1}$. ■

D.3 Corollary *There are 2×2 matrices X,Y such that $e^{X+Y}=e^Xe^Y=e^Ye^X$, but $XY\neq YX$.*

Proof. Let $X=\begin{pmatrix} i\pi & 0\\ 0 & -i\pi\end{pmatrix}$, $Y=\begin{pmatrix} i\pi & 1\\ 0 & -i\pi\end{pmatrix}$. Now $XY=\begin{pmatrix} -\pi^2 & i\pi\\ 0 & -\pi^2\end{pmatrix}\neq\begin{pmatrix} -\pi^2 & -i\pi\\ 0 & -\pi^2\end{pmatrix}=YX$, yet the preceding lemma gives $e^X=e^Y=-\mathbf{1}$ and $e^{X+Y}=\mathbf{1}$, so that $e^{X+Y}=e^Xe^Y=e^Ye^X$. ■

The importance of multiples of $2\pi i$ in the above construction is underlined by the result, due to Wermuth (1997), that an identity $e^Xe^Y=e^Ye^X$ in a unital Banach algebra does imply $XY=YX$ when the spectra of X and Y are '$2\pi i$-congruence free', i.e. neither contains u,v with $\frac{u-v}{2\pi i}\in\mathbb{Z}\backslash\{0\}$. For a simple proof see [54].

The BCHD theorem, the subject of this paper, is one way of coping with the failure of $e^{X+Y}=e^Xe^Y$ in a non-commutative setting. The Lie-Trotter formula proven below is another such result, often employed together with BCHD in treatments of matrix Lie groups like [32, 50].

D.4 THEOREM *Let $\mathcal{A}$ be a unital Banach algebra.*

(i) If $\{Z_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ such that $Z_n \stackrel{n\to\infty}{\longrightarrow} Z$ then

$$\lim_{n\to\infty}\left(1+\frac{Z_n}{n}\right)^n = e^Z.$$

(ii) For all $X, Y \in \mathcal{A}$, the Lie-Trotter product formula holds:

$$e^{X+Y} = \lim_{n\to\infty}\left(e^{\frac{X}{n}}e^{\frac{Y}{n}}\right)^n.$$

Proof. (ii) Assume (i) holds. If $\|X\| \le 1$ then

$$\|e^X - 1 - X\| = \left\|\sum_{n=2}^\infty \frac{X^n}{n!}\right\| \le \|X\|^2 \sum_{n=2}^\infty 1/n! = \|X\|^2(e-2),$$

thus $e^X = 1 + X + O(\|X\|^2)$ as $\|X\| \to 0$. Expanding the exponential functions, we have

$$e^{\frac{X}{n}}e^{\frac{Y}{n}} = 1 + \frac{X}{n} + \frac{Y}{n} + O\left(\frac{1}{n^2}\right)$$

as $n \to \infty$ (for fixed X, Y). Thus

$$\lim_{n\to\infty}\left(e^{\frac{X}{n}}e^{\frac{Y}{n}}\right)^n = \lim_{n\to\infty}\left(1+\frac{X}{n}+\frac{Y}{n}+O\left(\frac{1}{n^2}\right)\right)^n = \lim_{n\to\infty}\left(1+\frac{X+Y+O\left(\frac{1}{n}\right)}{n}\right)^n = e^{X+Y},$$

where we used (i) in the last step.

(i) We have $Z_n \to Z$, thus $\|Z_n\| \to \|Z\| < \infty$, so that we can pick $C > 0$ such that $\|Z_n\| \le C$ for all n. Now we compute

$$\left(1+\frac{Z_n}{n}\right)^n = \sum_{k=0}^n \binom{n}{k}\left(\frac{Z_n}{n}\right)^k = \sum_{k=0}^n \frac{Z_n^k}{k!}\frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n} = \sum_{k=0}^\infty \frac{Z_n^k}{k!}a_{n,k}, \tag{D.1}$$

where $a_{n,k} = 0$ if $k > n$ and $a_{n,k} = \frac{n}{n}\frac{n-1}{n}\cdots\frac{n-k+1}{n}$ otherwise. We have $0 \le a_{n,k} < 1\ \forall n,k$ and $\lim_{n\to\infty} a_{n,k} = 1$ for each k. Thus in the final sum of (D.1), each summand tends to $Z^k/k!$ as $n \to \infty$. Furthermore, the norm of the kth summand is bounded, uniformly in n, by $C^k/k!$, which sums to $e^C < \infty$. Now the lemma below justifies taking the limit $n \to \infty$ inside the sum, so that

$$\lim_{n\to\infty}\left(1+\frac{Z_n}{n}\right)^n = \lim_{n\to\infty}\sum_{k=0}^\infty \frac{Z_n^k}{k!}a_{n,k} = \sum_{k=0}^\infty \lim_{n\to\infty}\frac{Z_n^k}{k!}a_{n,k} = \sum_{k=0}^\infty \frac{Z^k}{k!} = e^Z,$$

as desired. ∎

The following is a Banach space version of Lebesgue's Dominated Convergence Theorem for the measure space $(\mathbb{N}_0, P(\mathbb{N}_0), \mu)$, where $\mu(B) = \#B$ for each $B \subset \mathbb{N}_0$.

D.5 LEMMA *Let $\mathcal{V}$ be a Banach space, $\{a_{i,n} \in \mathcal{V}\}_{i,n\in\mathbb{N}_0}$ and $\{c_k \ge 0\}_{k\in\mathbb{N}_0}$ and such that*

(a) $\sum_{k=0}^\infty c_k < \infty$.

(b) $\|a_{n,k}\| \le c_k\ \forall n,k$.

(c) $b_k = \lim_{n\to\infty} a_{n,k}$ exists for all $k \in \mathbb{N}_0$.

Then the sums $\sum_{k=0}^\infty a_{n,k}$ and $\sum_{k=0}^\infty b_k$ converge absolutely, and

$$\lim_{n\to\infty}\left(\sum_{k=0}^\infty a_{n,k}\right) = \sum_{k=0}^\infty b_k\ \left(=\sum_{k=0}^\infty \lim_{n\to\infty} a_{n,k}\right). \tag{D.2}$$

Proof. The absolute convergence of $\sum_{k=0}^\infty a_{n,k}$ is immediate by (a), (b), and that of $\sum_{k=0}^\infty b_k$ follows from $\|b_k\| \le c_k$, which is implied by (c).

To prove the limit in (D.2), let $\varepsilon > 0$. In view of $\sum_k c_k < \infty$, there exists $K \in \mathbb{N}$ such that $\sum_{k=K}^\infty c_k < \varepsilon/2$. Since $a_{n,k} \stackrel{n\to\infty}{\longrightarrow} b_k$ for each $k = 0, \ldots, N-1$, there exists an $N \in \mathbb{N}$ such that $n \ge N$ implies $\|a_{n,k} - b_k\| < \frac{\varepsilon}{2K}$ for each $k = 0, \ldots, K-1$. Thus for $n \ge N$ we have $\| \sum_{k=0}^{K-1} a_{n,k} - \sum_{k=0}^{K-1} b_k\| < \varepsilon/2$ and therefore

$$\begin{aligned} \left\| \sum_{k=0}^{K-1} a_{n,k} - \sum_{k=0}^{\infty} b_k \right\| &= \left\| \left(\sum_{k=0}^{K-1} a_{n,k} - \sum_{k=0}^{K-1} b_k \right) - \sum_{k=K}^{\infty} b_k \right\| \\ &\le \left\| \sum_{k=0}^{K-1} a_{n,k} - \sum_{k=0}^{K-1} b_k \right\| + \sum_{k=K}^{\infty} \|b_k\| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

The fact that for each $\varepsilon > 0$ we can find a K means that $\lim_{N\to\infty} \sum_{k=0}^{K-1} a_{n,k} = \sum_{k=0}^\infty b_k$, as claimed. ■

E Some connections between Eulerian and Bernoulli numbers

Our aim here is to give a direct and elementary proof of Corollary 9.11.

E.1 Lemma *(i) The Eulerian numbers satisfy the recursion relation*

$$A(1,0) = 1, \qquad A(n,m) = (n-m)A(n-1,m-1) + (m+1)A(n-1,m) \qquad \forall n \ge 2 \tag{E.1}$$

with the understanding that $A(n,m) = 0$ if $m < 0$ or $m \ge n$.

(ii) For $|x| < \min(1, |t|^{-1})$ we have

$$1 + \sum_{n=1}^{\infty} \frac{x^n}{n!} \sum_{m=0}^{n-1} A(n,m) t^m = \frac{t-1}{t - e^{x(t-1)}}. \tag{E.2}$$

Proof. (i) If $\sigma' \in S_n$ then removing n gives a $\sigma \in S_{n-1}$. Conversely, σ' arises from this (and only this) $\sigma \in S_{n-1}$ by inserting n at the right place. Now, let $\sigma = (\sigma(1), \ldots, \sigma(n-1)) \in S_{n-1}$. To obtain some $\sigma' \in S_n$, we must insert n. If we insert n after $\sigma(n-1)$ or between $\sigma(i)$ and $\sigma(i+1)$ with i a descent, we have $\mathrm{asc}(\sigma') = \mathrm{asc}(\sigma) + 1, \mathrm{des}(\sigma') = \mathrm{des}(\sigma)$. (This gives $m+1$ cases.) If we insert n before $\sigma(1)$ or between $\sigma(i)$ and $\sigma(i+1)$ with i an ascent, we have $\mathrm{asc}(\sigma') = \mathrm{asc}(\sigma), \mathrm{des}(\sigma') = \mathrm{des}(\sigma) + 1$. (This gives $n-1-m+1 = n-m$ cases.) Now (E.1) is immediate.

(ii) In view of $0 \le A(n,m) \le n!$, we have

$$\left| \sum_{n=0}^{\infty} \sum_{m=0}^{n-1} \frac{A(n,m)}{n!} x^n t^m \right| \le \sum_{n=0}^{\infty} |x|^n \sum_{m=0}^{n-1} |t|^m \le \sum_{n=0}^{\infty} |x|^n n \max(1,|t|)^n = \sum_{n=0}^{\infty} n(|x|\max(1,|t|))^n,$$

which converges if $|x|\max(1,|t|) < 1$, defining a smooth function $G(x,t)$ on this domain. Consider the partial differential operator

$$D = 1 + (t - t^2)\frac{\partial}{\partial t} + (tx - 1)\frac{\partial}{\partial x}.$$

A straightforward computation, left to the reader, proves $D\left(\frac{t-1}{t - e^{x(t-1)}}\right) = 0$. Applying the operator D to the l.h.s. of (E.2) gives

$$\begin{aligned} 1 &+ \sum_{n=1}^{\infty} \sum_m \frac{A(n,m)}{n!} \left(x^n t^m + (t-t^2)x^n m t^{m-1} + (tx-1) n x^{n-1} t^m\right) \\ &= 1 + \sum_{n=1}^{\infty} \sum_m \frac{A(n,m)}{n!} \left((m+1)x^n t^m + (n-m)x^n t^{m+1} - n x^{n-1} t^m\right) \\ &= 1 - 1 + \sum_{n=1}^{\infty} \sum_m \frac{x^n t^m}{n!} \left((m+1)A(n,m) + (n+1-m)A(n,m-1) - A(n+1,m)\right) \\ &= 0, \end{aligned}$$

where we used (E.1) with n replaced by $n+1$, thus valid for $n \geq 1$. (Reading this computation backwards perhaps provides some motivation: The point is to find polynomials H, K, L in x, t such that (E.1) implies $HG + KG_t + LG_x \equiv 0$, and $H = 1, K = t - t^2, L = tx - 1$ do the job.)

Thus both sides of (E.2) are annihilated by the linear (in G, G_t, G_x) differential operator D. Since they clearly coincide on the line $x = 0$ (and also on $t = 0$), they coincide everywhere. ■

Direct proof of Corollary 9.11. Let $n \geq 2$. To prove (9.14) it suffices to insert (E.1) into its l.h.s., resulting in

$$\begin{aligned}
\sum_m (-1)^m A(n,m)\, m!\,(n-1-m)! &= \sum_m (-1)^m m!\,(n-1-m)![(n-m)A(n-1,m-1) + (m+1)A(n-1,m)] \\
&= \sum_m (-1)^m m!\,(n-1-m)!\,(n-m)A(n-1,m-1) + \sum_m (-1)^m m!\,(n-1-m)!\,(m+1)A(n-1,m) \\
&= \sum_m (-1)^m m!\,(n-m)!\,A(n-1,m-1) + \sum_m (-1)^m (m+1)!\,(n-1-m)!\,A(n-1,m) \\
&= \sum_m (-1)^{m+1} (m+1)!\,(n-m-1)!\,A(n-1,m) + \sum_m (-1)^m (m+1)!\,(n-1-m)!\,A(n-1,m) \\
&= 0.
\end{aligned}$$

(All the m-summations are over $\mathbb{Z}$.)

The proof of (9.15), which I owe to Ira Gessel [28], is less trivial and requires the generating function. Replacing t in (E.2) by $-u/(1-u)$ and x by $x(1-u)$ (the product of which has absolute value $|ux|$) gives

$$1 + \sum_{n=1}^{\infty} \frac{[x(1-u)]^n}{n!} \sum_{m=0}^{n-1} A(n,m) \left(\frac{-u}{1-u}\right)^m = \frac{-u/(1-u) - 1}{-u/(1-u) - e^{x(1-u)(-u/(1-u)-1)}}.$$

Some rearranging leads to

$$1 + \sum_{n=1}^{\infty} \frac{x^n}{n!} \sum_{m=0}^{n-1} (-1)^m A(n,m)\, u^m (1-u)^{n-m} = \frac{e^x}{1 + u(e^x - 1)}.$$

Integrating this identity over $[0,1]$ with respect to u with the help of (9.12) gives (for $|x| < 1$)

$$1 + \sum_{n=1}^{\infty} \frac{x^n}{n!} \sum_{m=0}^{n-1} (-1)^m A(n,m) \frac{m!\,(n-m)!}{(n+1)!} = \frac{xe^x}{e^x - 1} = x + \frac{x}{e^x - 1} = 1 + \frac{x}{2} + \sum_{n=2}^{\infty} \frac{x^n}{n!} B_n,$$

using $B_1 = -\frac{1}{2}$. Now comparison of coefficients proves (9.15). ■

E.2 Remark An equivalent way of stating the identities of Corollary 9.11, employed by the Wikipedia article on Eulerian numbers, is

$$\sum_{m=0}^{n-1} (-1)^m \frac{A(n,m)}{\binom{n-1}{m}} = 0, \qquad \sum_{m=0}^{n-1} (-1)^m \frac{A(n,m)}{\binom{n}{m}} = (n+1)\, B_n \qquad \forall n \geq 2.$$

But the versions (9.14-9.15) are more useful for our purposes. □

E.3 Remark While the connection (9.15) between Bernoulli numbers and Eulerian numbers is from the 19th century, the permutations did not become involved in this story before 1953. But a different (and better known) connection between permutations and Bernoulli (and Euler) numbers really is from around 1880: Defining the (up-down) *alternating permutations* by

$$\mathcal{T}_n = \{\sigma \in S_n \mid \sigma(1) < \sigma(2) > \sigma(3) < \cdots\},$$

André[14] [2] proved (see [60] for a survey of later developments)

$$\sum_{n=0}^{\infty} \frac{\#\mathcal{T}_n}{n!} x^n = \sec x + \tan x,$$

equivalent to $\#\mathcal{T}_n = T_n$ (tangent numbers) for odd n and $\#\mathcal{T}_n = S_n$ (secant numbers) for even n, so that

$$\#\mathcal{T}_n = (-1)^{(n-1)/2} \frac{2^{n+1}(2^{n+1}-1)}{n+1} B_{n+1} \qquad \text{for all odd } n.$$

(For even n, one has $S_n = (-1)^{n/2} E_n$, where the *Euler numbers*[15] E_n (which vanish for odd n) are defined by $\frac{2}{e^x+e^{-x}} = \sum_{n=0}^{\infty} \frac{E_n}{n!} x^n$.)

Now, as a consequence of the symmetry $A(n,m) = A(n,n-1-m)$, it is clear that $\sum_{m=0}^{n-1} (-1)^m A(n,m) = 0$ for even n. The formula (with Eulerian numbers defined without involvement of permutations)

$$\sum_{m=0}^{n-1} (-1)^m A(n,m) = T_n = \frac{2^{n+1}(2^{n+1}-1)}{n+1} B_{n+1} \qquad \text{for odd } n$$

goes back to Laplace (1777) and Eytelwein (1816)! More recent references are [25, 51]. Not much after the discovery of the meaning of $A(n,m)$ in terms of permutations, purely combinatorial proofs of $\sum_{m=0}^{n-1} (-1)^m A(n,m) = \#\mathcal{T}_n$ for odd n were found, see [23, Théorème 5.6] and [44, Sect. 4.2].

2. When Bernoulli and Euler numbers appear together in alternating fashion as in the formula for $\#\mathcal{T}_n$, e.g. in computation of the volumes of certain polytopes [5] or the computation of $\sum_{k=-\infty}^{\infty} (4k+1)^{-n}$ [21], this usually is due to the identity in question factoring through $\#\mathcal{T}_n$. □

References

[1] R. Achilles, A. Bonfiglioli: The early proofs of the theorem of Campbell, Baker, Hausdorff, and Dynkin. Arch. Hist. Exact Sci. **66**, 295-358 (2012).

[2] D. André: Développement de sec x and tg x. C. R. Math. Acad. Sci. Paris **88**, 965-967 (1879). Sur les permutations alternées. Journ. Math. Pures Appl. (3^e série) **7**, 167-184 (1881).

[3] H. Araki: Expansional in Banach algebras. Ann. Sci. Éc. Norm. Sup. **6**, 67-84 (1973).

[4] H. F. Baker: Alternants and continuous groups. Proc. London Math. Soc. **(2) 3**, 24-47 (1905). (And earlier papers.)

[5] F. Beukers, J. A. C. Kolk, E. Calabi: Sums of generalized harmonic series and volumes. Nieuw Archief v. Wisk. **IV, 11**, 217-224 (1993).

[6] A. Bonfiglioli, R. Fulci: *Topics in noncommutative algebra. The theorem of Campbell-Baker-Hausdorff-Dynkin.* Springer 2012.

[7] N. Bourbaki: *Lie groups and Lie algebras. Part I: Chapters 1-3.* Addison Wesley 1975. (Translation of the French edition. Chapitre I: 1960, 1971. Chapitres II-III: 1972.)

[8] J. E. Campbell: On a law of combination of operators bearing on the theory of continuous transformation groups I+II. Proc. London Math. Soc. **(1) 28**, 381-390 (1897), **(1) 29**, 14-32 (1897).

[9] H. Cartan: *Differential calculus.* Hermann 1971.

[10] P. Cartier: Structure topologique des groupes de Lie. Exposé 22 of Séminaire "Sophus Lie", Éc. Norm. Sup., 1954/55.

[11] P. Cartier: Démonstration algébrique de la formule de Hausdorff. Bull. Soc. Math. France **84**, 241-249 (1956).

[14] Désiré André (1840-1917). French mathematician who mostly worked on combinatorics.

[15] Not to be confused with the Eulerian numbers!

[12] F. Casas: Computational aspects of some exponential identities. *In:* K. Ebrahimi-Fard, M. Barbero Liñán (eds.): *Discrete mechanics, geometric integration and Lie-Butcher series*. Springer 2018.

[13] F. Casas, A. Murua: An efficient algorithm for computing the Baker-Campbell-Hausdorff series and some of its applications. Journ. Math. Phys. **50**, 033513 (2009).

[14] K. T. Chen: Integrations of paths, geometric invariants and a generalized Baker-Hausdorff formula. Ann. Math. **65**, 163-178 (1957).

[15] G. Czichowski: Hausdorff und die Exponentialformel in der Lie-Theorie. Seminar Sophus Lie **2**, 85-93 (1992).

[16] D. Ž. Djoković: An elementary proof of the Baker-Campbell-Hausdorff-Dynkin formula. Math. Zeit. **143**, 209-211 (1975).

[17] J. J. Duistermaat, J. A. C. Kolk: *Lie groups*. Springer 2000.

[18] E. B. Dynkin: Calculation of the coefficients in the Campbell-Hausdorff formula [Russian]. Usp. Mat. Nauk. **57**, 323-326 (1947). English translation in A. A. Yushkevich, G. M. Seitz, A. L. Onishchik (eds.): *Selected papers of E. B. Dynkin with commentary*. American Mathematical Society 2000.

[19] E. B. Dynkin: On the representation by means of commutators of the series $\log(e^x e^y)$ for noncommutative x and y [Russian]. Mat. Sb. (N.S.) **25(67)**, 155-162 (1949).

[20] M. Eichler: A new proof of the Baker-Campbell-Hausdorff formula. Journ. Math. Soc. Japan **20**, 23-25 (1968).

[21] N. Elkies: On the sums $\sum_{k=-\infty}^{\infty}(4k+1)^{-n}$. Amer. Math. Monthly **110**, 561-573 (2003).

[22] R. P. Feynman: An operator calculus having applications in quantum electrodynamics. Phys. Rev. **84**, 108-128 (1951).

[23] D. Foata, M.-P. Schützenberger: *Théorie géométrique des polynômes Eulériens*. Springer 1970. (Retyped in LaTex at arXiv:math/0508232v1)

[24] K. O. Friedrichs: Mathematical aspects of the quantum theory of fields V. Fields modified by linear homogeneous forces. Comm. Pure Appl. Math. **6**, 1-72 (1953). Reproduced in K.O. Friedrichs: *Mathematical aspects of the quantum theory of fields*. Interscience Publishers 1953.

[25] G. Frobenius: Über die Bernoullischen Zahlen und die Eulerschen Polynome. Sitzungsber. Preuss. Akad. Wiss. 809-847 (1910).

[26] I. Fujiwara: Operator calculus of quantized operator. Progr. Theor. Phys. **7**, 433-448 (1952).

[27] I. Fujiwara: On the evaluation of the operator function $\log(e^y e^z)$. Progr. Theor. Phys. **9**, 676 (1953).

[28] I. Gessel: Mathoverflow posting April 20, 2019 (in response to a question of this author). https://mathoverflow.net/questions/328469/eulerian-number-identity-reference-request/328534#328534

[29] R. Godement: *Introduction to the theory of Lie groups*. Springer 2017. (Translation of *Introduction à la théorie des groupes de Lie*, Springer 2004, written in 1973-4.)

[30] K. Goldberg: The formal power series for $\log(e^x e^y)$. Duke Math. Journ. **23**, 13-21 (1956).

[31] D. Grinberg, V. Reiner: Hopf algebras in combinatorics (Version containing solutions). http://www.cip.ifi.lmu.de/~grinberg/algebra/HopfComb-sols.pdf

[32] B. Hall: *Lie groups, Lie algebras and representations. An elementary introduction*. 2nd. ed. Springer 2015.

[33] F. Hausdorff: Die symbolische Exponentialformel in der Gruppentheorie. Ber. Verh. Kgl. Sächs. Ges. Wiss. Leipzig, Math.-phys. Kl. **58**, 19-48 (1906).

[34] M. Hausner, J. T. Schwartz: *Lie groups. Lie algebras*. Nelson and Sons 1968.

[35] J. Helmstetter: Série de Hausdorff d'une algèbre de Lie et projections canoniques de l'algèbre enveloppante. Journ. Alg. **120**, 170-199 (1989).

[36] J. Hilgert, K.-H. Neeb: *Lie-Gruppen und Lie-Algebren*. Vieweg 1991.

[37] G. Hochschild: *The structure of Lie groups*. Holden-Day 1965.

[38] N. Jacobson: *Lie algebras*. Wiley and Sons 1962, Dover 1979.

[39] C. Kassel: *Quantum groups.* Springer 1995.

[40] V. Kathotia: Kontsevich's universal formula for deformation quantization and the Campbell-Baker-Hausdorff formula. Int. Journ. Math. **11**, 523-551 (2000).

[41] H. Kobayashi, N. Hatano, M. Suzuki: Goldberg's theorem and the Baker-Campbell-Hausdorff formula. Physica A **250**, 535-548 (1998).

[42] J.-L. Loday: Série de Hausdorff, idempotents Eulériens et algèbres de Hopf. Expos. Math. **12**, 165-178 (1994).

[43] R. C. Lyndon: A theorem of Friedrichs. Michigan Math. Journ. **3**, 27-29 (1955).

[44] T. K. Petersen: *Eulerian numbers.* Birkhäuser 2015.

[45] H. Poincaré: Sur les groupes continus. Cambr. Phil. Trans. **18**, 220-255 (1899), Quelques remarques sur les groupes continus. Rend. Circ. Mat. Palermo **15**, 321-368 (1901).

[46] M. Raussen, C. Skau: Interview with Michael Atiyah and Isadore Singer. Europ. Math. Soc. Newsletter Sept. 2004.

[47] R. Ree: Lie elements and an algebra associated with shuffles. Ann. Math. **68**, 210-220 (1958).

[48] M. W. Reinsch: A simple expression for the terms in the Baker-Campbell-Hausdorff series. Journ. Math. Phys. **41**, 2434-2442 (2000).

[49] C. Reutenauer: *Free Lie algebras.* Oxford University Press 1993.

[50] W. Rossmann: *Lie groups. An introduction through linear groups.* Oxford University Press 2002.

[51] L. Saalschütz: *Vorlesungen über die Bernoullischen Zahlen, ihren Zusammenhang mit den Secanten-Coefficienten und ihre wichtigeren Anwendungen.* Springer 1893.

[52] A. A. Sagle, R. E. Walde: *Introduction to Lie groups and Lie algebras.* Academic Press 1973.

[53] W. Schmid: Poincaré and Lie groups. Bull. Amer. Math. Soc. **6**, 175-186 (1982).

[54] C. Schmoeger: Remarks on commuting exponentials in Banach algebras. Proc. Amer. Math. Soc. **127**, 1337-1338 (1999).

[55] F. Schur: Zur Theorie der endlichen Transformationsgruppen. Math. Ann. **38**, 263-286 (1891).

[56] C. de Seguins Pazzis: On commuting matrices and exponentials. Proc. Amer. Math. Soc. **141**, 763-774 (2012).

[57] J.-P. Serre: *Lie algebras and Lie groups.* Benjamin 1964, Springer 1992.

[58] R. F. Snider: Perturbation variation methods for a quantum Boltzmann equation. Journ. Math. Phys. **5**, 1580-1587 (1964).

[59] W. Specht: Die linearen Beziehungen zwischen höheren Kommutatoren. Math. Zeit. **51**, 367-376 (1948).

[60] R. P. Stanley: A survey of alternating permutations. *In: Combinatorics and graphs*, Contemp. Math. **531**, 165-196 (2010). See also http://www-math.mit.edu/~rstan/papers/altperm.pdf and the video at https://www.youtube.com/watch?time_continue=13&v=Hmpo3l3fIKE .

[61] J. Stillwell: *Naive Lie theory.* Springer 2008.

[62] R. S. Strichartz: The Campbell-Baker-Hausdorff-Dynkin formula and solutions of differential equations. Journ. Funct. Anal. **72**, 320-345 (1987).

[63] G. M. Tuynman: The derivation of the exponential map of matrices. Amer. Math. Monthly **102**, 818-819 (1995).

[64] V. S. Varadarajan: *Lie groups, Lie algebras, and their representations.* Prentice-Hall 1974, Springer 1984.

[65] W. v. Waldenfels: Zur Charakterisierung Liescher Elemente in freien Algebren. Arch. Math. **17**, 44-48 (1966).

[66] G. E. Wall: A formula in noncommutative differential calculus. II. Austral. Math. Soc. Gaz. **5**, 37-43 (1978).

[67] F. Wever: Über Invarianten in Lieschen Ringen. Math. Ann. **120**, 563-580 (1949).

[68] D. Wigner: An identity in the free Lie algebra. Proc. Amer. Math. Soc. **106**, 639-640 (1989).

[69] J. Worpitzky: Studien über die Bernoullischen und Eulerschen Zahlen. Journ. Reine Angew. Math. **94**, 203-232 (1883).