Available online at www.sciencedirect.com
SCIENCE DIRECT®
LINEAR ALGEBRA AND ITS APPLICATIONS
ELSEVIER Linear Algebra and its Applications 378 (2004) 135–158
www.elsevier.com/locate/laa

On the convergence and optimization of the Baker–Campbell–Hausdorff formula

Sergio Blanes, Fernando Casas *

Departament de Matemàtiques, Universitat Jaume I, 12071 Castellón, Spain

Received 22 May 2003; accepted 18 September 2003

Submitted by R.A. Brualdi

Abstract

In this paper the problem of the convergence of the Baker–Campbell–Hausdorff series for $Z = \log(\mathrm{e}^X \mathrm{e}^Y)$ is revisited. We collect some previous results about the convergence domain and present a new estimate which improves all of them. We also provide a new expression of the truncated Lie presentation of the series up to sixth degree in X and Y requiring the minimum number of commutators. Numerical experiments suggest that a similar accuracy is reached with this approximation at a considerably reduced computational cost.

Keywords: BCH formula; Convergence; Lie algebras; Lie groups

1. Introduction

Let X and Y be non-commutative indeterminates. By introducing the formal series for the exponential function

$$\mathrm{e}^X \mathrm{e}^Y = \sum_{p,q=0}^{\infty} \frac{1}{p!q!} X^p Y^q$$

and substituting this series in the formal series defining the logarithm function

$$\log Z = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} (Z-1)^k$$

* Corresponding author.
E-mail addresses: sblanes@mat.uji.es (S. Blanes), casas@mat.uji.es (F. Casas).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2003.09.010

we obtain

$$\log(\mathrm{e}^X \mathrm{e}^Y) = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} \sum \frac{X^{p_1} Y^{q_1} \cdots X^{p_k} Y^{q_k}}{p_1! q_1! \cdots p_k! q_k!},$$

where the inner summation extends over all non-negative integers $p_1, q_1, \ldots, p_k, q_k$ for which $p_i + q_i > 0$ $(i = 1, 2, \ldots, k)$. Gathering together the terms for which $p_1 + q_1 + p_2 + q_2 + \cdots + p_k + q_k = m$ we can write

$$Z = \log(\mathrm{e}^X \mathrm{e}^Y) = \sum_{m=1}^{\infty} P_m(X, Y), \tag{1}$$

where $P_m(X, Y)$ is a homogeneous polynomial of degree m in the non-commuting variables X and Y.

The Baker–Campbell–Hausdorff (BCH) theorem asserts that every polynomial $P_m(X, Y)$ in (1) is a Lie polynomial, namely it can be expressed in terms of X and Y by addition, multiplication by rational numbers and nested commutators. As usual, the commutator $[X, Y]$ is defined as $XY - YX$. As is well known, this theorem proves to be very useful in various fields of mathematics (theory of linear differential equations [14], Lie group theory [9], numerical analysis [10]) and theoretical physics (perturbation theory, transformation theory, quantum mechanics and statistical mechanics [13,28]). In particular, in the theory of Lie groups, with this theorem one can explicitly write the operation of multiplication in a Lie group in canonical coordinates in terms of the Lie bracket operation in its tangent algebra and also prove the existence of a local Lie group with a given Lie algebra [9].

If $\mathbb{K}$ is any field of characteristic zero, let us denote by $\mathbb{K}\langle X, Y\rangle$ the associative algebra of polynomials in the non-commuting variables X and Y [22]. With the operation $X, Y \longmapsto [X, Y]$ one can introduce a commutator Lie algebra $[\mathbb{K}\langle X, Y\rangle]$ in a natural way. Then, in $[\mathbb{K}\langle X, Y\rangle]$, the set of all Lie polynomials in X and Y (i.e., all possible expressions obtained from X, Y by addition, multiplication by numbers and the Lie operation $[a, b] = ab - ba$) forms a subalgebra $\mathscr{L}(X, Y)$, which in fact is a free Lie algebra with generators X, Y [22]. With this notation, the BCH theorem can be formulated as four statements, each one more stringent than the preceding [27]. Specifically,

(A) The equation $\mathrm{e}^X \mathrm{e}^Y = \mathrm{e}^Z$ has a solution Z in $\mathbb{K}\langle X, Y\rangle$.
(B) The solution Z lies in $\mathscr{L}(X, Y)$.
(C) The exponent Z is an analytic function of X and Y.
(D) The exponent Z can be computed by a series

$$Z(X, Y) = z_1(X, Y) + z_2(X, Y) + \cdots$$

where every polynomial $z_n(X, Y) \in \mathscr{L}(X, Y)$.

In particular Dynkin derived an explicit formula for Z as a series of iterated commutators as

$$Z = \sum_{k=1}^{\infty} \sum_{p_i, q_i} \frac{(-1)^{k-1}}{k} \frac{[X^{p_1} Y^{q_1} \cdots X^{p_k} Y^{q_k}]}{\left(\sum_{i=1}^{k} (p_i + q_i)\right) p_1! q_1! \cdots p_k! q_k!}, \tag{2}$$

where the inner summation is taken over all non-negative integers $p_1, q_1, \ldots, p_k, q_k$ such that $p_1 + q_1 > 0, \ldots, p_k + q_k > 0$ and $[X^{p_1} Y^{q_1} \cdots X^{p_k} Y^{q_k}]$ denotes the right nested commutator based on the word $X^{p_1} Y^{q_1} \cdots X^{p_k} Y^{q_k}$ [7]. This series is called the *BCH formula in the Dynkin form*.

In [27] the question of the global validity of the BCH theorem is analyzed in detail. It is shown that, whereas statement (A) is never in doubt, statements (B)–(D) are only globally valid in a free Lie algebra. Thus, in particular, if X and Y are elements of a normed algebra, the resulting series of normed elements is not guaranteed to converge out of a neighborhood of zero, and therefore cannot be used to compute Z in the large.

There are various statements in the literature concerning the size of this convergence domain depending on the presentation used [23]. It is our purpose to review some of the results published and, by deriving a new recursion procedure for calculating the terms of the series, obtain a larger domain of convergence of the BCH formula.

It is perhaps less known the role the BCH formula plays in the numerical treatment of differential equations on manifolds [11], a branch of numerical analysis which has received considerable attention during the last few years [10]. If $\mathscr{M}$ is a smooth manifold and $\mathfrak{X}(\mathscr{M})$ denotes the linear space of smooth vector fields on $\mathscr{M}$, then a Lie algebra structure can be established in $\mathfrak{X}(\mathscr{M})$ by using the Lie bracket $[X, Y]$ of fields X and Y. In terms of coordinates $x^1, \ldots, x^d$, the components of the field $[X, Y]$ are

$$[X, Y]^i = \sum_{j=1}^{d} \left(X^j \frac{\partial Y^i}{\partial x^j} - Y^j \frac{\partial X^i}{\partial x^j} \right).$$

It follows from the BCH theorem that for any vector fields (linear differential operators) $X, Y \in \mathfrak{X}(\mathscr{M})$ every operator $P_m(X, Y)$ belongs to the Lie algebra $\mathfrak{X}(\mathscr{M})$ and also $\log(\mathrm{e}^X \mathrm{e}^Y)$ if the corresponding series converges.

The flow of a vector field $X \in \mathfrak{X}(\mathscr{M})$ is a mapping $\exp(X)$ defined through the solution of the differential equation

$$u' = X(u), \quad u(0) = p$$

as $\exp(X)(p) = u(1)$. Many numerical methods used to solve differential equations on manifolds are based on compositions of maps that are flows of vector fields. Thus it is important to determine the set of all possible flows that can be obtained by composing a given set of basic flows. In this context the BCH formula provides a useful approach in the construction and analysis of such methods, particularly in the problem of obtaining the order conditions of composition and splitting methods [10]. Typically the vector fields X_i whose flows are to be composed by this class of algorithms depend on a step size h in such a way that $X_i = \mathscr{O}(h^{q_i})$ as $h \to 0$, with

$q_i \geqslant 1$. Then it is clear that $[X_i, X_j] = \mathcal{O}(h^{q_i+q_j})$. When the composition method has order p all terms in the BCH formula which are $\mathcal{O}(h^{p+1})$ can safely be discarded. This has to be done in an efficient way, as well as reducing the total number of Lie brackets involved.

This problem is also addressed in this paper. In particular we get expressions for $z_m(X, Y)$, $m \leqslant 6$, involving the minimum number of commutators and we analyze on some numerical examples the main features of this optimized formulation.

2. Convergence of the BCH formula

In this section we collect various estimates appeared in the literature for the convergent domain of the series $Z = \log(\mathrm{e}^X \mathrm{e}^Y)$. We consider an arbitrary finite-dimensional matrix Lie algebra $\mathfrak{g}$ such that $X, Y \in \mathfrak{g}$. If G is the Lie group corresponding to the Lie algebra $\mathfrak{g}$ then the exponential map $\exp : \mathfrak{g} \longrightarrow G$ is the usual exponential matrix.

2.1. Direct estimates

The first group of results are obtained by considering different explicit presentations of Z and majorizing the terms of the corresponding series.

(a) *The associative presentation.* Z has a unique presentation in the form [8]

$$Z = X + Y + \sum_{n=2}^{\infty} \sum_{w,|w|=n} g_w w \tag{3}$$

in which w denotes a word in the symbols X and Y (i.e., $w = w_1 w_2 \dots w_n$, each w_i being X or Y), g_w is a rational coefficient and the inner sum is taken over all words w with length $|w| = n$. Here the length of w is the number of letters it contains. The coefficients can be computed with a recursive procedure due to Goldberg [8].

If $\mathfrak{g}$ is a complete normed Lie algebra with a norm compatible with associative multiplication, i.e., such that

$$\|XY\| \leqslant \|X\| \|Y\| \tag{4}$$

for all X, Y in $\mathfrak{g}$, then

Theorem 2.1 [24]. *The associative series* (3) *for* $Z = \log(\mathrm{e}^X \mathrm{e}^Y)$ *converges absolutely if* $\|X\| < 1$ *and* $\|Y\| < 1$.

(b) *Lie presentation of Z in the Dynkin form.* When one chooses a norm in $\mathfrak{g}$ and a number $\mu > 0$ such that

$$\|[X, Y]\| \leqslant \mu \|X\| \|Y\| \tag{5}$$

for all X, Y in $\mathfrak{g}$, then the following bound is obtained.

Theorem 2.2 [3, 7]. *The domain of absolute convergence of the infinite series* (2) *contains the open set*

$$C = \left\{ (X, Y) \in \mathfrak{g} \times \mathfrak{g} : \|X\| + \|Y\| < \frac{1}{\mu} \log 2 \right\}.$$

(c) *Lie presentation of Z in the Goldberg form.* In virtue of Dynkin's theorem [22], series (3) can be written as

$$Z = X + Y + \sum_{n=2}^{\infty} \frac{1}{n} \sum_{w, |w|=n} g_w [w], \tag{6}$$

that is, the individual terms are the same as in the associative series (3) except that the word $w = w_1 w_2 \ldots w_n$ is replaced with the right nested commutator $[w] = [w_1, [w_2, \ldots [w_{n-1}, w_n] \ldots]]$ and the coefficient g_w is divided by the word length.

Theorem 2.3 [24, 20]. *With a norm satisfying condition* (5), *the Lie presentation* (6) *for Z based on the Goldberg coefficients converges absolutely provided that* $\|X\| \leqslant \frac{1}{\mu}$ *and* $\|Y\| \leqslant \frac{1}{\mu}$.

Remark. Of course, for a norm in $\mathfrak{g}$ satisfying $\|XY\| \leqslant \|X\|\|Y\|$ it is true that $\|[X, Y]\| \leqslant \mu \|X\|\|Y\|$ with $\mu = 2$. The advantage of considering (5) rests in the fact that for some special classes of Lie algebras and norms μ is actually less than 2. For instance, if $\mathfrak{g} = \mathfrak{so}(n)$, the Lie algebra of $n \times n$ skew-symmetric matrices, and $\| \cdot \|$ is the Frobenius norm, then $\mu = \sqrt{2}$.

2.2. Estimates based on the Magnus expansion

A constructive procedure for obtaining explicitly the fundamental matrix of the differential equation

$$\frac{\mathrm{d}U}{\mathrm{d}t} = A(t)U, \qquad U(0) = I \tag{7}$$

was proposed by Magnus [14] in the form $U(t) = \exp(\Omega(t))$, with the logarithm of $U(t)$ satisfying the equation

$$\Omega' = \mathrm{d}\exp_{\Omega}^{-1} A(t) = \sum_{k=0}^{\infty} \frac{B_k}{k!} \mathrm{ad}_{\Omega}^{k} A(t). \tag{8}$$

Here $\mathrm{ad}_{\Omega}^{0} A = A$, $\mathrm{ad}_{\Omega}^{k} A = [\Omega, \mathrm{ad}_{\Omega}^{k-1} A]$ and B_k are Bernoulli numbers. When the infinite Magnus series $\Omega(t) = \sum_{n=1}^{\infty} \Omega_n(t)$ is substituted in (8) and equal powers of A are collected, one arrives at the recurrence

$$\Omega_1(t) = \int_0^t A(s)\,\mathrm{d}s, \tag{9}$$

$$\Omega_n(t) = \sum_{j=1}^{n-1} \frac{B_j}{j!} \sum_{\substack{k_1+\cdots+k_j=n-1 \\ k_1\geqslant 1,\ldots,k_j\geqslant 1}} \int_0^t \mathrm{ad}_{\Omega_{k_1}} \mathrm{ad}_{\Omega_{k_2}} \cdots \mathrm{ad}_{\Omega_{k_j}} A(s)\,\mathrm{d}s, \quad n \geqslant 2.$$

If the piecewise constant matrix-valued function

$$A(t) = \begin{cases} Y & 0 \leqslant t \leqslant 1, \\ X & 1 < t \leqslant 2 \end{cases} \tag{10}$$

with $X, Y \in \mathfrak{g}$ is considered, then the exact solution of (7) at $t = 2$ is $U(2) = \mathrm{e}^X \mathrm{e}^Y$. By computing $U(2) = \mathrm{e}^{\Omega(2)}$ with recursion (9) one obtains for the first terms

$$\begin{aligned} \Omega_1(2) &= X + Y, \\ \Omega_2(2) &= \tfrac{1}{2}[X, Y], \\ \Omega_3(2) &= \tfrac{1}{12}[X, [X, Y]] + \tfrac{1}{12}[Y, [X, Y]], \\ \Omega_4(2) &= \tfrac{1}{24}[X, [Y, [Y, X]]], \end{aligned} \tag{11}$$

and, in general,

$$\Omega(2) = Z = \log(\mathrm{e}^X \mathrm{e}^Y) = X + Y + \sum_{n=2}^{\infty} G_n(X, Y), \tag{12}$$

where $G_n(X, Y)$ is a homogeneous Lie polynomial in X and Y of degree n. In fact, each $G_n(X, Y)$ is a linear combination of the commutators of the form $[V_1, [V_2, \ldots, [V_{n-1}, V_n] \ldots]]$ with $V_i \in \{X, Y\}$ for $1 \leqslant i \leqslant n$, the coefficients being universal rational constants. It is perhaps for this reason that the Magnus expansion is often referred in the literature as the continuous analogue of the BCH formula.

Now it turns out that the Magnus series is absolutely convergent for t such that

$$\int_0^t \|A(s)\|\,\mathrm{d}s < \frac{2.1737}{\mu}$$

with a norm satisfying (5) [1,17]. In consequence,

Theorem 2.4. *The BCH series in the form* (12) *converges absolutely when* $\|X\| + \|Y\| < 2.1737/\mu$.

The Ω series can also be represented in the form [4,26]

$$\Omega(t) = \sum_{n=1}^{\infty} \frac{1}{n!} \int_0^t \mathrm{d}t_n \cdots \int_0^t \mathrm{d}t_1 (-1)^{n-\Theta_n-1} \Theta_n! (n - \Theta_n - 1)! A(t_n) \cdots A(t_1), \tag{13}$$

where $\Theta_n = \Theta_n(t_n, \ldots, t_1) = v_n + \cdots + v_2$, and the step function $v_k = 1$ if $t_k \geqslant t_{k-1}$ and zero otherwise. It has been shown [18] that the series (13) converges absolutely for t such that

$$t \max_{s\in[0,t]} \|A(s)\| < 2$$

with a norm compatible with associative multiplication. If, on the other hand, one applies Dynkin's theorem then every monomial $A(t_n)\cdots A(t_1)$ in (13) may be replaced with the fully nested commutator $\frac{1}{n}[A(t_n), \ldots, [A(t_2), A(t_1)]]$ and the resulting series is convergent for t such that

$$t \max_{s\in[0,t]} \|A(s)\| < \frac{2}{\mu} \tag{14}$$

with a norm verifying (5). When the explicit formula (13) is applied to the problem defined by (10) then the associative presentation (3) of $\Omega(2) = Z$ is recovered. Therefore this presentation converges if $\max\{\|X\|, \|Y\|\} < 1$, which is just the result of Theorem 2.1. A similar result follows from (14) when a nested commutator structure is considered for $\Omega(t)$. In this case we retrieve the Lie presentation of Z in the Goldberg form (6) and Theorem 2.3.

2.3. *Estimates from associated differential equations*

The third group of convergence proofs are derived by analyzing certain differential equations closely related to the BCH formula.

For sufficiently small $t \in \mathbb{R}$ one has $\exp(tX)\exp(tY) = \exp F(t)$. Here $F(t)$ is an analytic function around $t = 0$ which satisfies the equation [25]

$$\frac{\mathrm{d}F}{\mathrm{d}t} = X + Y + \frac{1}{2}[X - Y, F] + \sum_{p=1}^{\infty} \frac{B_{2p}}{(2p)!}\mathrm{ad}_F^{2p}(X + Y)$$

with initial condition $F(0) = 0$. By writing $F(t) = \sum_{n=1}^{\infty} t^n z_n(X, Y)$, a recursive procedure can be designed for the determination of the coefficients z_n [25]. In fact, $z_1 = X + Y$ and $z_n(X, Y) = G_n(X, Y)$ in (12) for $n \geqslant 2$.

Theorem 2.5 [25]. *The series $\sum_{n=1}^{\infty} z_n(X, Y) = Z(X, Y)$ converges absolutely for all $X, Y \in \mathscr{S} = \{V \in \mathfrak{g} : \|V\| < \frac{\delta}{2\mu}\}$, its sum defines an analytic map of $\mathscr{S} \times \mathscr{S}$ into $\mathfrak{g}$ and $\exp X \exp Y = \exp Z(X, Y)$. Here $\delta > 0$ is a constant such that the equation*

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \frac{1}{2}y + \sum_{p=1}^{\infty} \frac{|B_{2p}|}{(2p)!} y^{2p}, \quad y(0) = 0$$

admits a solution $y(t)$ which is holomorphic in the disc $\{t : |t| < \delta\}$.

In [20] Newman, So and Thompson prove that the explicit value of this number is

$$\delta = \int_0^{2\pi} \frac{\mathrm{d}y}{2 + \frac{1}{2}y - \frac{1}{2}\cot(\frac{1}{2}y)} = 2.17373740\ldots$$

and thus one has

Theorem 2.6 [20]. *Under a norm verifying* (5) *the BCH series* (12) *with all terms of each fixed degree expressed in terms of commutators and grouped as a single term converges absolutely if* $\|X\| < \frac{\delta}{2\mu}$, $\|Y\| < \frac{\delta}{2\mu} = 1.0868/\mu$.

If $Z = \log(\mathrm{e}^X \mathrm{e}^Y)$ is expressed as $Z = \sum_{n=0}^{\infty} U_n$, where now U_n is the homogeneous component of degree n in Y, then by applying similar techniques as in the previous case one obtains

Theorem 2.7 [6]. *The series* $\sum_{n\geqslant 0} U_n$ *converges absolutely for* $\|X\| < \frac{1.2357}{\mu}$, *and* $\|Y\| < \frac{1.2357}{\mu}$.

3. An enlarged convergence domain for the BCH series

In this section we obtain a new convergence theorem for the presentation of $Z = \log(\mathrm{e}^X \mathrm{e}^Y)$ as a series (12) of homogeneous Lie polynomials in X and Y.

For sufficiently small $u, v, t \in \mathbb{R}$ it is true that

$$\exp(utX)\exp(vY) = \exp F_1(u, v, t). \tag{15}$$

If we denote $C(\varepsilon, t) = F_1(\varepsilon, \varepsilon, t)$, then for all sufficiently small $\varepsilon > 0$

$$C(\varepsilon, t) = \sum_{n=1}^{\infty} \varepsilon^n c_n(t), \tag{16}$$

the series being absolutely convergent [25]. Our goal is to provide the terms c_n explicitly and the domain of convergence of the series. This will be done by deriving a differential equation for $C(\varepsilon, t)$ and obtaining its solution as a power series in ε. The coefficients c_n will then be determined by a recursion procedure.

Lemma 3.1. $C(\varepsilon, t)$ *is a solution to the equation*

$$\frac{\partial C}{\partial t} = \mathrm{d}\exp_C^{-1}(\varepsilon X) \tag{17}$$

with the initial condition $C(\varepsilon, 0) = \varepsilon Y$.

Proof. By deriving (15) one gets [25]

$$\frac{\partial}{\partial t}(\exp F_1(u, v, t)) = \mathrm{d}\exp_{F_1}\left(\frac{\partial F_1}{\partial t}\right)\exp F_1(u, v, t),$$

where, formally,

$$\mathrm{d}\exp_{F_1} = \sum_{j=0}^{\infty} \frac{1}{(j+1)!} \mathrm{ad}_{F_1}^j,$$

and

$$\frac{\partial}{\partial t}(\exp(utX)\exp(vY)) = uX\exp(utX)\exp(vY) = uX\exp F_1(u, v, t).$$

Equating differentials,

$$uX = \mathrm{d}\exp_{F_1}\left(\frac{\partial F_1}{\partial t}\right),$$

so that

$$\frac{\partial F_1}{\partial t}(u, v, t) = \mathrm{d}\exp_{F_1}^{-1}(uX).$$

Now we put $u = v = \varepsilon$, and thus

$$\frac{\partial C}{\partial t}(\varepsilon, t) = \mathrm{d}\exp_{C(\varepsilon,t)}^{-1}(\varepsilon X).$$

Finally, by substituting $t = 0$, $u = v = \varepsilon$ in (15), $\exp(\varepsilon Y) = \exp C(\varepsilon, 0)$, so that $C(\varepsilon, 0) = \varepsilon Y$. $\square$

Lemma 3.2. *The coefficients* $c_n(t)$ *in the expansion* (16) *are determined by the recursion formula*

$$\begin{aligned} c_1(t) &= Xt + Y \\ c_n(t) &= \sum_{j=1}^{n-1} \frac{B_j}{j!} \sum_{\substack{k_1+\cdots+k_j=n-1 \\ k_1\geqslant 1,\ldots,k_j\geqslant 1}} \int_0^t \mathrm{ad}_{c_{k_1}(s)} \mathrm{ad}_{c_{k_2}(s)} \cdots \mathrm{ad}_{c_{k_j}(s)} X \, \mathrm{d}s, \quad n \geqslant 2 \end{aligned} \tag{18}$$

Proof. If we denote by $\mathcal{O}(\varepsilon^k)$ any function of the form $\varepsilon \mapsto \varepsilon^k f(\varepsilon)$, where f is analytic around $\varepsilon = 0$, it is clear that

$$\frac{\partial C}{\partial t}(\varepsilon, t) = \sum_{j=1}^{n} \varepsilon^j c_j'(t) + \mathcal{O}(\varepsilon^{n+1})$$

and

$$\mathrm{ad}_{C(\varepsilon,t)} = \varepsilon \mathrm{ad}_{c_1} + \varepsilon^2 \mathrm{ad}_{c_2} + \cdots + \varepsilon^{n-1} \mathrm{ad}_{c_{n-1}} + \mathcal{O}(\varepsilon^n).$$

Therefore

$$\begin{aligned} \mathrm{ad}_{C(\varepsilon,t)}^2 &= \varepsilon^2 \mathrm{ad}_{c_1} \mathrm{ad}_{c_1} + \varepsilon^3 (\mathrm{ad}_{c_1} \mathrm{ad}_{c_2} + \mathrm{ad}_{c_2} \mathrm{ad}_{c_1}) + \cdots \\ &= \sum_{l=2}^{n-1} \varepsilon^l \sum_{\substack{k_1+k_2=l \\ k_1\geqslant 1, k_2\geqslant 1}} \mathrm{ad}_{c_{k_1}} \mathrm{ad}_{c_{k_2}} + \mathcal{O}(\varepsilon^n) \end{aligned}$$

and in general

$$\mathrm{ad}^j_{C(\varepsilon,t)} = \sum_{l=j}^{n-1} \varepsilon^l \sum_{\substack{k_1+\cdots+k_j=l \\ k_1\geqslant 1,\ldots,k_j\geqslant 1}} \mathrm{ad}_{c_{k_1}} \mathrm{ad}_{c_{k_2}} \cdots \mathrm{ad}_{c_{k_j}} + \mathcal{O}(\varepsilon^n).$$

On the other hand, $\mathrm{ad}_{C(\varepsilon,t)} = \mathcal{O}(\varepsilon)$, so that

$$\begin{aligned} \mathrm{d}\exp^{-1}_C(\varepsilon X) &= \varepsilon X + \varepsilon \sum_{j=1}^{n-1} \frac{B_j}{j!} \mathrm{ad}^j_C X + \mathcal{O}(\varepsilon^{n+1}) \\ &= \varepsilon X + \varepsilon \sum_{l=1}^{n-1} \varepsilon^l \sum_{j=1}^{l} \frac{B_j}{j!} \sum_{\substack{k_1+\cdots+k_j=l \\ k_1\geqslant 1,\ldots,k_j\geqslant 1}} \mathrm{ad}_{c_{k_1}} \cdots \mathrm{ad}_{c_{k_j}} X + \mathcal{O}(\varepsilon^{n+1}) \\ &= \varepsilon X + \sum_{l=2}^{n} \varepsilon^l \sum_{j=1}^{l-1} \frac{B_j}{j!} \sum_{\substack{k_1+\cdots+k_j=l-1 \\ k_1\geqslant 1,\ldots,k_j\geqslant 1}} \mathrm{ad}_{c_{k_1}} \cdots \mathrm{ad}_{c_{k_j}} X + \mathcal{O}(\varepsilon^{n+1}). \end{aligned}$$

If we substitute these expressions in (17) and identify the coefficients of ε^l on both sides we get

$$\begin{aligned} c_1'(t) &= X, \\ c_l'(t) &= \sum_{j=1}^{l-1} \frac{B_j}{j!} \sum_{\substack{k_1+\cdots+k_j=l-1 \\ k_1\geqslant 1,\ldots,k_j\geqslant 1}} \mathrm{ad}_{c_{k_1}} \cdots \mathrm{ad}_{c_{k_j}} X, \quad l \geqslant 2. \end{aligned}$$

The initial condition $C(\varepsilon, 0) = \sum_{l\geqslant 1} \varepsilon^l c_l(0) = \varepsilon Y$ implies $c_1(0) = Y$ and $c_l(0) = 0$, $l \geqslant 2$. This proves the lemma. $\square$

The recurrence (18) provides c_n, $n \geqslant 2$, as a linear combination of $n-1$ nested commutators involving n symbols X, Y. Then $C(1, 1) = \sum_{n\geqslant 1} c_n(1)$ is the BCH series in terms of homogeneous Lie polynomials in X and Y, i.e., the presentation (12). Next we analyse its convergence. For that purpose we consider the series

$$v(\varepsilon, t) = \sum_{j=1}^{\infty} \varepsilon^j \|c_j(t)\|. \tag{19}$$

Lemma 3.3. *With a norm in* $\mathfrak{g}$ *satisfying* (5) *then*

$$v(\varepsilon, t) \leqslant \frac{1}{\mu} G^{-1} \left(G(\mu\|Y\|) + \varepsilon\mu\|X\|t \right), \quad 0 \leqslant t \leqslant T \tag{20}$$

holds, *where*

$$G(\tau) = \int_0^\tau \frac{\mathrm{d}x}{2 + \frac{x}{2}\left(1 - \cot\frac{x}{2}\right)} \tag{21}$$

and $T = \sup\{t \geqslant 0 : G(\mu\|Y\|) + \varepsilon\mu\|X\|t < G(2\pi) \equiv \xi = 2.1737738\ldots\}$.

Proof. From (18) it is clear that for $l \geqslant 2$

$$\|c_l(t)\| \leqslant \sum_{p=1}^{l-1} \mu^p \frac{|B_p|}{p!} \sum_{\substack{k_1+\cdots+k_p=l-1 \\ k_1\geqslant 1,\ldots,k_p\geqslant 1}} \int_0^t \|c_{k_1}(s)\|\cdots\|c_{k_p}(s)\|\|X\|\,\mathrm{d}s$$

and thus

$$\begin{aligned}\sum_{l=2}^{N} \varepsilon^l \|c_l(t)\| &\leqslant \varepsilon \sum_{l=2}^{N}\sum_{p=1}^{l-1} \mu^p \frac{|B_p|}{p!} \sum_{\substack{k_1+\cdots+k_p=l-1 \\ k_1\geqslant 1,\ldots,k_p\geqslant 1}} \varepsilon^{l-1} \int_0^t \|c_{k_1}(s)\|\cdots\|c_{k_p}(s)\|\|X\|\,\mathrm{d}s \\ &= \varepsilon \sum_{p=1}^{N-1} \mu^p \frac{|B_p|}{p!} \sum_{l=p}^{N-1} \varepsilon^l \sum_{\substack{k_1+\cdots+k_p=l \\ k_1\geqslant 1,\ldots,k_p\geqslant 1}} \int_0^t \|c_{k_1}(s)\|\cdots\|c_{k_p}(s)\|\|X\|\,\mathrm{d}s.\end{aligned}$$

Let us denote $v_N(\varepsilon, t) = \sum_{l=1}^{N} \varepsilon^l \|c_l(t)\|$. Then it is easy to show that

$$(v_N(\varepsilon, t))^p = \sum_{l=p}^{pN} \varepsilon^l \sum_{\substack{k_1+\cdots+k_p=l \\ k_1\geqslant 1,\ldots,k_p\geqslant 1}} \|c_{k_1}\|\cdots\|c_{k_p}\|,$$

so that, in the last inequality,

$$\sum_{l=p}^{N-1} \varepsilon^l \sum_{\substack{k_1+\cdots+k_p=l \\ k_1\geqslant 1,\ldots,k_p\geqslant 1}} \|c_{k_1}\|\cdots\|c_{k_p}\| \leqslant (v_N(\varepsilon, t))^p$$

and therefore

$$\begin{aligned} v_N(\varepsilon, t) &\leqslant \varepsilon(\|X\|t + \|Y\|) + \varepsilon \int_0^t \sum_{p=1}^{N-1} \frac{|B_p|}{p!} (\mu v_N(\varepsilon, s))^p \|X\|\,\mathrm{d}s \\ &= \varepsilon\|Y\| + \varepsilon \int_0^t \sum_{p=0}^{N-1} \frac{|B_p|}{p!} (\mu v_N(\varepsilon, s))^p \|X\|\,\mathrm{d}s.\end{aligned}$$

Taking the limit $N \to \infty$ we have

$$v(\varepsilon, t) \leqslant \varepsilon\|Y\| + \varepsilon \int_0^t \|X\| g(\mu v(\varepsilon, s))\,\mathrm{d}s, \tag{22}$$

since

$$\sum_{p=0}^{\infty} \frac{|B_p|}{p!} x^p = 2 + \frac{x}{2}\left(1 - \cot \frac{x}{2}\right) \equiv g(x). \tag{23}$$

Now the desired conclusion can be achieved from a result obtained in [16,17]. For completeness we reproduce here the argument.

Let $U(t) \equiv \int_0^t \|X\| g(\mu v(\varepsilon, s))\, \mathrm{d}s$. In consequence,

$$U'(t) = \|X\| g(\mu v(\varepsilon, t)) \leqslant \|X\| g(\mu\varepsilon\|Y\| + \mu\varepsilon U(t)),$$

since g is a non-decreasing function. Thus we have

$$\frac{U'(t)}{g(\mu\varepsilon\|Y\| + \mu\varepsilon U(t))} \leqslant \|X\|$$

because g is positive on the real axis. In fact $g(z)$ is analytic for $|z| < 2\pi$ with positive coefficients in the power series and has no zeros in the ball $|z| < 2\pi$. By integrating this inequality we have

$$\frac{1}{\varepsilon\mu} \int_{\mu\varepsilon\|Y\|}^{\mu\varepsilon\|Y\| + \mu\varepsilon U(t)} \frac{1}{g(x)} \mathrm{d}x \leqslant \|X\| t$$

or

$$G(\mu\varepsilon\|Y\| + \mu\varepsilon U(t)) \leqslant G(\mu\varepsilon\|Y\|) + \mu\varepsilon\|X\| t,$$

where $G(t) \equiv \int_0^t \frac{1}{g(x)} \mathrm{d}x$. Now $G(z)$ is also analytic in $|z| < 2\pi$ and $G'(z) = \frac{1}{g(z)} \neq 0$. Then $y = G(z)$ has an inverse function $z = G^{-1}(y)$ for y in the ball $|y| < G(2\pi)$, which is also analytic there. In consequence, $\mu\varepsilon\|Y\| < 2\pi$ and

$$v(\varepsilon, t) \leqslant \varepsilon\|Y\| + \varepsilon U(t) \leqslant \frac{1}{\mu} G^{-1}(G(\mu\varepsilon\|Y\|) + \mu\varepsilon\|X\| t)$$

for t such that $G(\mu\varepsilon\|Y\|) + \mu\varepsilon\|X\| t$ belongs to the domain of G^{-1}, i.e.,

$$\begin{aligned} &G(\mu\varepsilon\|Y\|) + \mu\varepsilon\|X\| t < G(2\pi) \\ &= \int_0^{2\pi} \frac{1}{g(x)} \mathrm{d}x \equiv \xi = 2.17373738\ldots \qquad \square \end{aligned}$$

If we take $\varepsilon = t = 1$ in (20) then

$$v(\varepsilon = 1, t = 1) = \sum_{n=1}^{\infty} \|c_n(t = 1)\| < \frac{1}{\mu} G^{-1}(\xi) = \frac{2\pi}{\mu}$$

provided $G(\mu\|Y\|) + \mu\|X\| < \xi$. In other words, the BCH series $\sum_{n \geqslant 1} c_n(t = 1)$ is absolutely convergent if

$$\int_0^{\mu\|Y\|} \frac{1}{g(x)} \mathrm{d}x + \mu\|X\| < \xi. \tag{24}$$

On the other hand, it is clear that the roles of X and Y in (15) are interchangeable, i.e., it is equally true that

$$\exp(uX)\exp(vtY) = \exp F_2(u, v, t)$$

and thus for all sufficiently small ε

$$D(\varepsilon, t) \equiv F_2(\varepsilon, \varepsilon, t) = \sum_{n=1}^{\infty} \varepsilon^n d_n(t). \tag{25}$$

By following a similar procedure one has the equivalent results to Lemmas 3.1 and 3.2.

Lemma 3.4. *$D(\varepsilon, t)$ is a solution to the initial value problem*

$$\frac{\partial D}{\partial t} = \mathrm{d}\exp^{-1}_{-D}(\varepsilon Y), \qquad D(\varepsilon, 0) = \varepsilon X.$$

Lemma 3.5. *The coefficients d_n in* (25) *verify the recurrence*

$$d_1(t) = X + Yt,$$
$$d_n(t) = \sum_{j=1}^{n-1} (-1)^j \frac{B_j}{j!} \sum_{\substack{k_1+\cdots+k_j=n-1 \\ k_1 \geqslant 1, \ldots, k_j \geqslant 1}} \int_0^t \mathrm{ad}_{d_{k_1}(s)} \cdots \mathrm{ad}_{d_{k_j}(s)} Y \,\mathrm{d}s, \qquad n \geqslant 2$$

and thus $D(1,1)$ also leads to the BCH series in the presentation (12).

Now, for $n \geqslant 2$

$$\|d_n(t)\| \leqslant \sum_{j=1}^{n-1} \mu^j \frac{|B_j|}{j!} \sum_{\substack{k_1+\cdots+k_j=n-1 \\ k_1 \geqslant 1, \ldots, k_j \geqslant 1}} \int_0^t \|d_{k_1}(s)\| \cdots \|d_{k_p}(s)\| \|Y\| \,\mathrm{d}s$$

and therefore a result similar as Lemma 3.3 holds for the series $\sum_{n \geqslant 1} \varepsilon^n \|d_n(t)\|$ provided that $G(\mu\varepsilon\|X\|) + \mu\varepsilon\|Y\|t < \xi$. We have then proved the following

Theorem 3.6. *The BCH formula in the form* (12), *i.e., expressed as a series of homogeneous Lie polynomials in X and Y, converges absolutely in the domain $D_1 \cup D_2$ of $\mathfrak{g} \times \mathfrak{g}$, where*

$$D_1 = \left\{ (X, Y) : \mu\|X\| < \int_{\mu\|Y\|}^{2\pi} \frac{1}{g(x)} \mathrm{d}x \right\}, \tag{26}$$

$$D_2 = \left\{ (X, Y) : \mu\|Y\| < \int_{\mu\|X\|}^{2\pi} \frac{1}{g(x)} \mathrm{d}x \right\}, \tag{27}$$

and $g(x) = 2 + \frac{x}{2}(1 - \cot \frac{x}{2})$.

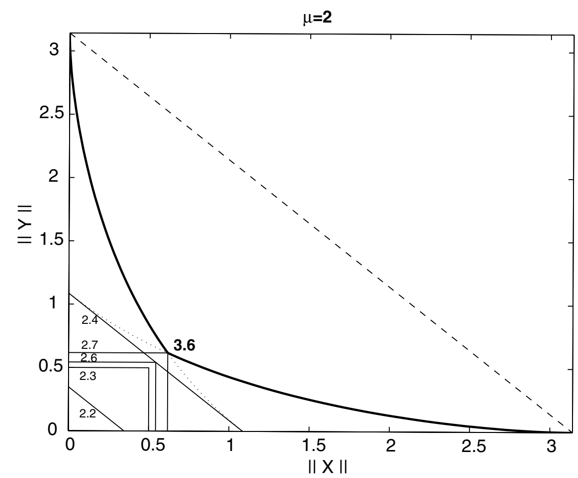

Fig. 1. Bounds for the convergence domain of the BCH series in the Lie presentation when $\mu = 2$ obtained from the different theorems. The number of each theorem is used to label the curves. Except for 2.3, the curves are not part of the convergence domain. The thick curves define the best result and the broken line corresponds to the limit where the exponential representation is guaranteed according to Theorem 4.1.

At this point it is quite illustrative to represent in a figure all the bounds considered here for the convergence domain of the BCH series in the Lie presentation. This is done in Fig. 1 for a value $\mu = 2$. Observe that the domain $D_1 \cup D_2$ given by Theorem 3.6 clearly contains all the previous estimates and that the coordinates of the point where the boundaries of D_1 and D_2 intersect is precisely $\|X\| = \|Y\| = 1.2357/\mu$, the estimate given by Theorem 2.7 for the convergence of the series in terms of homogeneous components in Y.

4. Further discussion on the convergence of the BCH series

Theorem 3.6 provides sufficient conditions for the convergence of the BCH series based on an estimation by norms. In particular it guarantees that the exponent Z can safely be computed with the convergent series (12) in the domain $D_1 \cup D_2$. But, in fact, the domain of analyticity of the function $Z(X, Y)$ is still larger, as the following (unpublished) result shows.

Theorem 4.1 [15]. *Let $X_1, X_2, \ldots, X_k$ $(k \geqslant 2)$ be bounded operators in a Hilbert space $\mathscr{H}$ with* $\dim(\mathscr{H}) > 2$ *and*

$$B_\gamma = \left\{ x = (X_1, \ldots, X_k) : \sum_{i=1}^{k} \|X_i\| < \gamma \right\}.$$

Then the operator function φ, $\varphi(0) = 1_{\mathscr{H}}$, $\varphi(x) = \log(e^{X_k} e^{X_{k-1}} \cdots e^{X_2} e^{X_1})$ (*which is well defined in* B_γ *if* γ *is small enough*) *is in fact an analytic function of* $x = (X_1, \ldots, X_k)$ *in* B_π.

As is well known, given two operators S, T in $\mathscr{H}$, then $\|ST\| \leqslant \|S\|\|T\|$ and $\|[S, T]\| \leqslant 2\|S\|\|T\|$, which corresponds to $\mu = 2$.

The following example shows that the estimate $\gamma = \pi$ provided by Theorem 4.1 is in a certain sense optimal for $\mu = 2$. In other words, the domain of validity of statement (C) of the BCH theorem is, with no further assumptions on X and Y, precisely B_π.

Example 1. Let us consider the matrices

$$A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{28}$$

with the property $[A, B] = 2B$. We take $X = \alpha_1 A$, $Y = \alpha_2 A + \beta B$. Since

$$e^{aA+bB} = \begin{pmatrix} e^a & \frac{b}{a}\sinh a \\ 0 & e^{-a} \end{pmatrix},$$

we have

$$e^X e^Y = \begin{pmatrix} e^{\alpha_1+\alpha_2} & e^{\alpha_1}\frac{\beta}{\alpha_2}\sinh\alpha_2 \\ 0 & e^{-(\alpha_1+\alpha_2)} \end{pmatrix}$$

and therefore

$$Z = \log(e^X e^Y) = (\alpha_1 + \alpha_2)A + \frac{\beta(\alpha_1+\alpha_2)}{\alpha_2} \frac{1 - e^{-2\alpha_2}}{1 - e^{-2(\alpha_1+\alpha_2)}} B, \tag{29}$$

an analytic function if $|\alpha_1 + \alpha_2| < \pi$ with singularities at $\alpha_1 + \alpha_2 = \pm i\pi$. In this case every nested commutator is proportional to B, and thus the BCH series in the presentation (12) gives

$$Z = X + Y + \sum_{n=2}^{\infty} G_n(X, Y) = (\alpha_1 + \alpha_2)A + \beta \sum_{n=1}^{\infty} f_n(\alpha_1, \alpha_2)B. \tag{30}$$

Here $f_1 = 1$ and f_n $(n \geqslant 2)$ is an homogeneous polynomial of degree $n - 1$ in α_1, α_2. By comparing with the exact expression (29) it is clear that the series (30) cannot converge if $|\alpha_1 + \alpha_2| \geqslant \pi$, independently of β.

If we consider the spectral norm, i.e., the matrix norm induced by the Euclidean vector norm, then $\mu = 2$, $\|A\| = \|B\| = 1$,

$$\|X\| = \alpha_1,$$

$$\|Y\| = \alpha_2 \left(1 + \frac{\beta^2}{2\alpha_2^2} + \frac{\beta}{\alpha_2^2}\sqrt{\alpha_2^2 + \frac{\beta^2}{4}}\right)^{1/2} \equiv \alpha_2\sqrt{\Delta(\beta, \alpha_2)}$$

with $\lim_{\beta\to 0} \Delta(\beta, \alpha_2) = 1$. The following results are then immediate.

1. For any $\epsilon > 0$ there exists a $\beta > 0$ such that $\pi < \|X\| + \|Y\| < \pi + \epsilon$ with $\alpha_1 + \alpha_2 = \pi$ and thus the BCH series (30) does not converge. In fact, if $\alpha_1 > 0, \alpha_2 > 0$, $\alpha_1 + \alpha_2 = \pi$ then
$$\pi < \|X\| + \|Y\| = \pi + \alpha_2 \left(-1 + \sqrt{\Delta(\beta, \alpha_2)}\right) < \pi + \epsilon$$
for values of β such that $\Delta(\beta, \alpha_2) < (\epsilon/\alpha_2 + 1)^2$.
2. For any $\epsilon > 0$ there exist values $\alpha_1 > 0$, $\alpha_2 > 0$ with $\alpha_1 + \alpha_2 < \pi$ for which $\|X\| + \|Y\| > \pi + \epsilon$ and the BCH series (30) does converge. This occurs, for instance, for values of β such that $\Delta(\beta, \alpha_2) > ((\pi + \epsilon - \alpha_1)/\alpha_2)^2$.

Let us now assume that $X, Y \in \mathfrak{g}$ are such that $\|[X, Y]\| \ll \mu\|X\|\|Y\|$. Observe that the bounds obtained for the convergence domain of the BCH series depend exclusively on μ and the norm of X and Y, and so the fact that X and Y nearly commute is irrelevant for that purpose. The question is then the following: is it possible to take into account this special feature of the commutator $[X, Y]$ to improve the previous estimates for this particular situation?

In general the answer is negative, as the following example exhibits.

Example 2. Let us consider $X = \alpha A$, $Y = X + \beta B$, with A, B given in (28). Now it is clear that $[X, Y] = 2\alpha\beta B$ and $\|[X, Y]\| = 2|\alpha\beta| \ll 2\|X\|\|Y\|$ if $|\beta| \ll 1$. The expression (29) of Z reads in this case
$$Z = \log(\mathrm{e}^X \mathrm{e}^Y) = 2\alpha A + 2\beta \frac{1 - \mathrm{e}^{-2\alpha}}{1 - \mathrm{e}^{-4\alpha}} B$$
with first singularities at $\alpha = \pm \mathrm{i}\frac{\pi}{2}$ independently of β. Thus one cannot expect the BCH series to converge for $|\alpha| \geqslant \frac{\pi}{2}$ regardless of the value of β.

Nevertheless, as the following lemma shows, even in this case $X + Y$ is a good approximation to Z for sufficiently small values of β.

Lemma 4.2. *The following inequality holds*
$$\|\mathrm{e}^X \mathrm{e}^Y - \mathrm{e}^{X+Y}\| \leqslant K_1 \mathrm{e}^{K_0 + K_1} \tag{31}$$
with
$$\begin{aligned} K_0 &= \|X\| + \|Y\|, \\ K_1 &= \frac{1 - \mathrm{e}^{2K_0}(1 - 2K_0)}{4K_0^2} \|[X, Y]\|. \end{aligned} \tag{32}$$

Proof. Let us consider the general equation (7) and let $\mathrm{e}^{B(t)}$ be an approximate solution, such as that obtained by the Magnus expansion (9). We can define the error of the approximation as
$$E(t) = U(t) - \mathrm{e}^{B(t)}. \tag{33}$$

Our goal is to provide a bound to $E(t)$. To this end we introduce the matrix $U_1(t)$ by $U(t) = \mathrm{e}^{B(t)}U_1(t)$. It is clear that

$$U_1' = A_1(t)U_1, \quad U_1(0) = I \tag{34}$$

with

$$A_1 = \mathrm{e}^{-B}A\mathrm{e}^{B} - \int_0^1 \mathrm{e}^{-xB}B'\,\mathrm{e}^{xB}\,\mathrm{d}x = \mathrm{e}^{-B}\left(\int_0^1 (A - \mathrm{e}^{xB}B'\,\mathrm{e}^{-xB})\,\mathrm{d}x\right)\mathrm{e}^{B}$$
$$= \mathrm{e}^{-B}\int_0^1 \left(A - \mathrm{e}^{-xB}A\,\mathrm{e}^{xB} + \mathrm{e}^{xB}(A - B')\mathrm{e}^{-xB}\right)\mathrm{d}x\,\mathrm{e}^{B}.$$

Taking into account that

$$A - \mathrm{e}^{xB}A\,\mathrm{e}^{-xB} = \int_0^x \mathrm{e}^{uB}[A, B]\,\mathrm{e}^{-uB}\,\mathrm{d}u$$

one gets

$$A_1 = \int_0^1\int_0^x \mathrm{e}^{-(1-u)B}[A, B]\,\mathrm{e}^{(1-u)B}\,\mathrm{d}x\,\mathrm{d}u + \int_0^1 \mathrm{e}^{-(1-x)B}(A - B')\,\mathrm{e}^{(1-x)B}\,\mathrm{d}x. \tag{35}$$

The error can also be written as $E = U(I - U_1^{-1})$ and

$$I - U_1^{-1} = \int_0^t U_1^{-1}(s)A_1(s)\,\mathrm{d}s.$$

If we consider functions $K_0(t)$ and $K_1(t)$ such that

$$\int_0^t \|A(s)\|\,\mathrm{d}s \leqslant K_0(t) \quad \text{and} \quad \int_0^t \|A_1(s)\|\,\mathrm{d}s \leqslant K_1(t)$$

respectively, then

$$\|E\| \leqslant \|U\|\|I - U_1^{-1}\| \leqslant \mathrm{e}^{K_0}K_1\,\mathrm{e}^{K_1}. \tag{36}$$

Now we take $B(t)$ as the approximation obtained with the first term of the Magnus expansion,

$$B(t) = \int_0^t A(s)\,\mathrm{d}s,$$

so that the second term in (35) vanishes, and $A(t)$ is given by (10). In that case

$$\int_0^2 \|A_1(s)\|\,\mathrm{d}s \leqslant \frac{1 - \mathrm{e}^{2K_0}(1 - 2K_0)}{4K_0^2}\int_0^2\int_0^\tau \|[A(\tau), A(s)]\|\,\mathrm{d}\tau\,\mathrm{d}s,$$

so we can choose K_0 and K_1 as in (32) where $[A(\tau), A(s)] = [X, Y]$ if $0 \leqslant s \leqslant 1 < \tau \leqslant 2$ and zero otherwise. □

In other words, if the norm of the commutator $[X, Y]$ is sufficiently small then $Z = \log(\mathrm{e}^X\mathrm{e}^Y)$ can be well approximated by the first term of the BCH formula, $X + Y$, even if the series does not converge. For Example 2, the bound (31) gives

$$\|E\| \leqslant \beta \frac{1 - \mathrm{e}^{4\alpha}(1 - 4\alpha)}{16\alpha^2} \mathrm{e}^{2\alpha} + \mathscr{O}(\beta^2),$$

so that $\|E\| \to 0$ in the limit $\beta \to 0$.

5. Optimization of the BCH series

In most practical applications the BCH series has to be truncated. For instance, a case of some relevance in quantum mechanics is when $[X, Y]$ commutes with both X and Y. Then obviously, $\exp(X)\exp(Y) = \exp(X + Y + \frac{1}{2}[X, Y])$. In the derivation of the order conditions of composition and splitting methods for solving numerically ordinary differential equations, one applies formally the BCH formula to get one exponential of a series in powers of the step size h and compares the truncated series (up to a given order of h) with the flow corresponding to the exact solution [10].

In general, given the presentation (12) in terms of homogeneous Lie polynomials in X and Y of degree n, when the series is truncated one has

$$Z_0^{[N]}(tX, tY) \equiv t(X + Y) + \sum_{n=2}^{N} G_n(tX, tY) = Z(tX, tY) + \mathscr{O}(t^{N+1}) \qquad (37)$$

with $G_n(tX, tY) = t^n G_n(X, Y)$. Our goal in this section is to analyze how the truncated series can be approximated in a computationally efficient way.

In order to proceed, let us consider again the free Lie algebra $\mathscr{L}(X, Y)$ and let $\mathscr{L}_n$ denote the subspace of $\mathscr{L}(X, Y)$ generated by homogeneous elements of degree n ($n \geqslant 1$). Therefore $\mathscr{L} = \bigoplus_{n\geqslant 1} \mathscr{L}_n$, $\mathscr{L}_{n+1} = [\mathscr{L}_n, \mathscr{L}_1]$ and $\mathscr{L}_1 = \mathrm{span}\{X, Y\}$. The dimension c_n of the subspace $\mathscr{L}_n$ is given by Witt's formula [3], its first 10 values being 2, 1, 2, 3, 6, 9, 18, 30, 56, 99. With this notation the BCH theorem establishes that for any $n \geqslant 1$, $G_n \in \mathscr{L}_n$ and thus it can be written as a linear combination of the elements of a basis in $\mathscr{L}_n$. Although there exists a standard procedure to construct a basis in $\mathscr{L}_n$ for all n providing the so-called Hall basis [3], due to the Jacobi identity different selections of independent nested commutators and different basis thus constructed are equally valid.

In this context we can formulate the following two problems concerning the truncated BCH series (37):

Problem 1. Find a basis $\{E_{n,i}\}_{i=1}^{c_n}$ of $\mathscr{L}_n$, $1 \leqslant n \leqslant N$, such that $Z_0^{[N]} = \sum_{n=1}^{N} \sum_{i=1}^{c_n} \alpha_{n,i} E_{n,i}$ has as few non-vanishing coefficients $\alpha_{n,i}$ as possible, i.e., such that $Z_0^{[N]}$ involves the minimum number of independent nested commutators.

Problem 2. Minimize the total number of different commutators required to approximate the expression of $Z_0^{[N]}$ up to order N.

The first problem was afforded in [21,12], obtaining a significant reduction in the number of different nested commutators in the series up to $N = 8$ and 9, respectively. Thus, in particular, $z_1, \dots, z_4$ in the sum $Z_0^{[6]}(tX, tY) = \sum_{n=1}^{6} t^n z_n(X, Y)$ are given by $\Omega_1(2), \dots, \Omega_4(2)$ in (11), respectively and

$$\begin{aligned} z_5(X,Y) &= \frac{1}{6!}(-[X,X,X,X,Y] - 6[X,X,Y,X,Y] - 2[X,Y,Y,X,Y] \\ &\quad + 2[Y,X,X,X,Y] + 6[Y,X,Y,X,Y] + [Y,Y,Y,X,Y]), \\ z_6(X,Y) &= \frac{1}{2\cdot 6!}(-2[X,X,Y,Y,X,Y] + 6[X,Y,X,Y,X,Y] \\ &\quad + [X,Y,Y,Y,X,Y] + [Y,X,X,X,X,Y]), \end{aligned} \tag{38}$$

where $[A_1, A_2, \dots, A_{n-1}, A_n] \equiv [A_1, [A_2, \dots [A_{n-1}, A_n]]]$. Thus, the actual number of terms c'_n ($n \leqslant 6$) appearing in the BCH series is 2, 1, 2, 1, 6, 4. If this basis is adopted, it is easy to verify that an upper bound for the total number of commutators involved in $Z_0^{[N]}$ is $C(N) = \sum_{i=2}^{N-1} c_i + c'_N$.

To reduce this number of commutators and thus solve Problem 2 is quite useful the notion of a graded free Lie algebra. A grading function w is introduced on $\mathscr{L}$ as follows. We assign a grade to the generators, $w(X)$, $w(Y)$, and then the grade is propagated in the basis by additivity: the grade of an element L of the basis of $\mathscr{L}_n$ of the form $L = [L_1, L_2]$ is $w(L) = w(L_1) + w(L_2)$ [19]. In the context of the BCH series, $w(X) = w(Y) = 1$.

Recently, an optimization technique has been proposed which in certain cases allows to write a generic element of a graded free Lie algebra with the minimum number of commutators [2]. The procedure has been applied to optimize numerical integration schemes for linear differential equations based on the Magnus expansion and the family of Runge–Kutta–Munthe-Kaas schemes for differential equations on manifolds [5]. In the particular case of the truncated BCH series $Z_0^{[N]}$ this technique leads to the following optimized expressions.

$N = 3$. Only two commutators are required to obtain $Z_0^{[3]}$. If we denote

$$d_1 = [X, Y], \tag{39}$$

then it is clear that

$$Z_0^{[3]} = X + Y + \tfrac{1}{2}d_1 + \tfrac{1}{12}[X - Y, d_1]. \tag{40}$$

$N = 4$. The expression of $Z_0^{[4]}$ can be written exactly in terms of three commutators as

$$Z_0^{[4]} = X + Y + \tfrac{1}{2}d_2 + \tfrac{1}{4}d_3 \tag{41}$$

with

$$\begin{aligned} d_2 &= [X + \tfrac{1}{6}d_1, Y], \\ d_3 &= [X, -\tfrac{2}{3}d_1 + d_2] \end{aligned} \tag{42}$$

and d_1 given by (39).

$N = 5$. With four commutators we can obtain an expression $\tilde{Z}_0^{[5]}$ which approximates $Z_0^{[5]}$ up to grade 5. In other words, $\tilde{Z}_0^{[5]}(tX, tY) = Z_0^{[5]}(tX, tY) + \mathcal{O}(t^6)$. Specifically, with

$$\begin{aligned} d_{5,1} &= [X, Y], \\ d_{5,2} &= -\frac{1}{168}\left[X + \frac{1}{2}(5 - \sqrt{21})Y, d_{5,1}\right], \\ d_{5,3} &= \left[X + \frac{1}{15}d_{5,1} - \frac{98}{75}(5 + \sqrt{21})d_{5,2}, \frac{75}{196}Y + \frac{5}{196}d_{5,1} + d_{5,2}\right], \\ d_{5,4} &= \left[X - Y + \frac{686}{2775}\sqrt{21}d_{5,3}, \frac{185}{1372}d_{5,1} - d_{5,3}\right], \end{aligned} \tag{43}$$

we have

$$\tilde{Z}_0^{[5]} = X + Y + \frac{23}{196}d_{5,1} + d_{5,3} - \frac{7}{30}d_{5,4}. \tag{44}$$

In (43) irrational numbers appear because a non-linear system of algebraic equations has to be solved. Of course, when all the commutators are combined in (44), we only have rational numbers in the approximated BCH series.

$N = 6$. In principle, the minimum number of commutators required is 5, since z_6 is a linear combination of nested commutators involving six operators. Nevertheless, we have not found solutions of the resulting non-linear system of algebraic equations in the coefficients. With one additional we have the following approximation:

$$\tilde{Z}_0^{[6]} = X + Y + \sum_{i=1}^{5} \beta_i d_i + \beta_6 [d_3, d_4] = Z_0^{[6]} + \mathcal{O}(t^7), \tag{45}$$

where d_1, d_2, d_3 are given by (39) and (42),

$$\begin{aligned} d_4 &= \frac{1}{36}\left[X + \alpha_1 Y + \alpha_2 d_2 + \alpha_3 d_3, 4d_1 - 6d_2 - 3d_3\right], \\ d_5 &= \left[X + x_{5,1}Y + \sum_{i=2}^{4} x_{5,i}d_i, \sum_{i=1}^{4} y_{5,i}d_i\right], \end{aligned} \tag{46}$$

and the numerical values of the coefficients are

$$\begin{array}{lll} \alpha_1 = 2 + \sqrt{3}, & \alpha_2 = -\frac{9(4+5\sqrt{3})}{118}, & \alpha_3 = -\frac{3(110+49\sqrt{3})}{236}, \\ x_{5,1} = 2 - \sqrt{3}, & x_{5,2} = -\frac{3(586+231\sqrt{3})}{1534}, & x_{5,3} = -\frac{3(-17\,972+27\,331\sqrt{3})}{181\,012}, \\ x_{5,4} = -\frac{9(23\,707+4721\sqrt{3}}{90\,506}, & y_{5,1} = \frac{4-\sqrt{3}}{9}, & y_{5,2} = \frac{1+\sqrt{3}}{6}, \\ y_{5,3} = \frac{1+\sqrt{3}}{12}, & y_{5,4} = 1, & \beta_1 = \frac{-9+5\sqrt{3}}{30}, \\ \beta_2 = \frac{4}{5} - \frac{1}{2\sqrt{3}}, & \beta_3 = \frac{3}{20}, & \beta_4 = \frac{-1+\sqrt{3}}{20}, \\ \beta_5 = \frac{1}{20}, & \beta_6 = -\frac{21(-32+19\sqrt{3})}{1180}, & \end{array}$$

In Table 1 we collect the dimension c_n of $\mathscr{L}_n$, the actual number of terms in the BCH series c'_n, the upper bound $C(N)$ and the minimum number of commutators required $F(N)$ up to $N = 6$.

Table 1
Number of commutators in the truncated BCH series

N	c_N	c'_N	$C(N)$	$F(N)$
1	2	2	0	0
2	1	1	1	1
3	2	2	3	2
4	3	1	4	3
5	6	6	12	4
6	9	4	16	6

Next we analyze the quality of the approximations $\widetilde{Z}_0^{[N]}$ in comparison with $Z_0^{[N]}$ for $N = 5, 6$. For this purpose we compute $\|e^Z - e^{Z_0^{[N]}}\|$ and $\|e^Z - e^{\widetilde{Z}_0^{[N]}}\|$ when X and Y are taken as $q \times q$ random matrices with $q = 4, 16$. To consider different situations first we take full random matrices with entries in the interval $(-0.5, 0.5)$ and repeat the computations taking their skew-symmetric projection ($X^S = (X - X^{\mathrm{T}})/2$, $Y^S = (Y - Y^{\mathrm{T}})/2$). In particular we measure

$$\left\| e^{tX} e^{tY} - e^{\widetilde{Z}_0^{[N]}(tX,tY)} \right\|$$

as a function of t. If the accuracy does not improve with N this is a clear indication that the convergence of the BCH formula fails for these values of X, Y and t. In Fig. 2 we show the errors obtained in the BCH approximation using $Z_0^{[N]}$ and $\widetilde{Z}_0^{[N]}$ for $1 \leqslant N \leqslant 6$. For $1 \leqslant N \leqslant 4$ we have that $Z_0^{[N]} = \widetilde{Z}_0^{[N]}$ and both curves coincide.

From the figure we observe that the approximations obtained with $Z_0^{[N]}$ and $\widetilde{Z}_0^{[N]}$ give very similar results. This experiment has been repeated a large number of times with different random matrices and values of q, and similar results were found.

However, the computational cost of $\widetilde{Z}_0^{[N]}$ is considerably smaller than the cost of $Z_0^{[N]}$ for $N > 4$. As mentioned, to compare the performance of different approximations of the same order it is necessary to take into account both the accuracy and the cost. A standard procedure is to define an effective error, E_{f}, which contains both contributions. If $\mathscr{C}$ is the cost and

$$\mathscr{E} = \| \log(e^X e^Y) - Z_0^{[N]} \| \quad \text{or} \quad \mathscr{E} = \| \log(e^X e^Y) - \widetilde{Z}_0^{[N]} \|$$

is the error of the approximation at order N it is usual to take

$$E_{\mathrm{f}} = \mathscr{C}^{N+1} \mathscr{E}.$$

In general, the cost can be considered as proportional to the number of matrix–matrix multiplications, and we can take $\mathscr{C} = x_N/(N+1)$ with $x_N = C(N)$ or $x_N = F(N)$. We have divided this number only to avoid misinterpretations because the large numbers obtained from $\mathscr{C}^{N+1}$. In Fig. 3 we show the results obtained for the effective error when considering full random matrices of dimension 20×20. We observe that the optimal approximations at orders 5 and 6 outperform the effective error obtained from the standard one by several orders of magnitude. As in the previous exam-

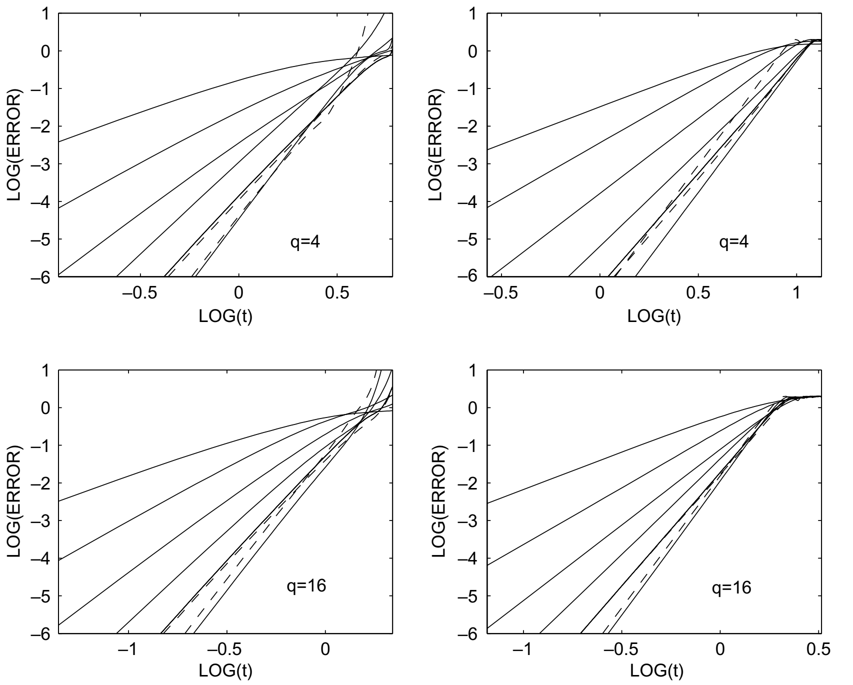

Fig. 2. Error in the BCH approximation using $Z_0^{[N]}$ (—) and $\widetilde{Z}_0^{[N]}$ (- - -) for $1 \leqslant N \leqslant 6$ ($Z_0^{[N]} = \widetilde{Z}_0^{[N]}$ for $1 \leqslant N \leqslant 4$). Full random $q \times q$ matrices are considered in the left figures, and their skew-symmetric projection on the right figures.

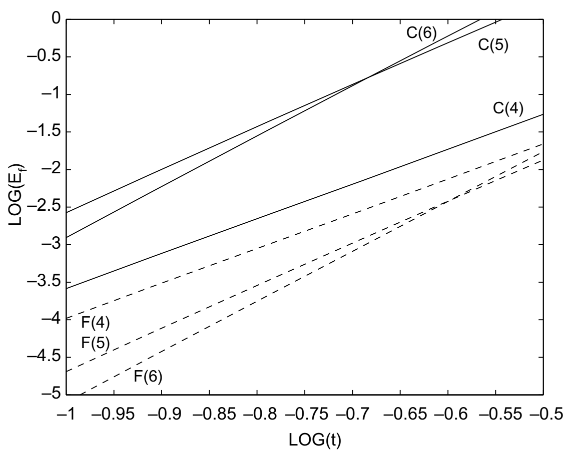

Fig. 3. Effective error in the BCH approximation using $Z_0^{[N]}$ (—) and $\widetilde{Z}_0^{[N]}$ (- - -) for $4 \leqslant N \leqslant 6$ for full random 20×20 matrices.

ple, similar results are obtained when repeated the experiment with other random matrices and dimensions.

Acknowledgements

The authors greatly acknowledge Prof. W. So for providing them Ref. [15] and for some useful remarks. Partial financial support has been given by Fundació Bancaixa. SB has also been supported by Ministerio de Ciencia y Tecnología (Spain) through a contract in the Program Ramón y Cajal 2001.

References

[1] S. Blanes, F. Casas, J.A. Oteo, J. Ros, Magnus and Fer expansions for matrix differential equations: the convergence problem, J. Phys. A: Math. Gen. 31 (1998) 259–268.
[2] S. Blanes, F. Casas, J. Ros, High order optimized geometric integrators for linear differential equations, BIT 42 (2002) 262–284.
[3] N. Bourbaki, Lie Groups and Lie Algebras, Springer, Berlin, 1989 (Chapters 1–3).
[4] I. Bialynicki-Birula, B. Mielnik, J. Plebański, Explicit solution of the continuous Baker–Campbell–Hausdorff problem and a new expression for the phase operator, Ann. Physics 51 (1969) 187–200.
[5] F. Casas, B. Owren, Cost efficient Lie group integrators in the RKMK class, preprint NTNU no. 5/2002, 2002. Accepted for publication in BIT.
[6] J. Day, W. So, R.C. Thompson, Some properties of the Campbell Baker Hausdorff series, Linear and Multilinear Algebra 29 (1991) 207–224.
[7] E.B. Dynkin, Calculation of the coefficients in the Campbell–Hausdorff formula, Dokl. Akad. Nauk SSSR 57 (1947) 323–326, English traslation in: A.A. Yushkevich, G.M. Seitz, A.L. Onishchik (Eds.), Selected Papers of E.B. Dynkin with Commentary, AMS, International Press, 2000.
[8] K. Goldberg, The formal power series for $\log(e^x e^y)$, Duke Math. J. 23 (1956) 13–21.
[9] V.V. Gorbatsevich, A.L. Onishchik, E.B. Vinberg, Foundations of Lie Theory and Lie Transformation Groups, Springer, Berlin, 1997.
[10] E. Hairer, C. Lubich, G. Wanner, Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, Springer, Berlin, 2002.
[11] A. Iserles, H.Z. Munthe-Kaas, S.P. Nørsett, A. Zanna, Lie-group methods, Acta Numer. 9 (2000) 215–365.
[12] M. Kolsrud, Maximal reductions in the Baker–Hausdorff formula, J. Math. Phys. 34 (1993) 270–285.
[13] K. Kumar, On expanding the exponential, J. Math. Phys. 6 (1965) 1928–1934.
[14] W. Magnus, On the exponential solution of differential equations for a linear operator, Comm. Pure Appl. Math. 7 (1954) 649–673.
[15] B.S. Mityagin, unpublished notes, 1990.
[16] P.C. Moan, J.A. Oteo, J. Ros, On the existence of the exponential solution of linear differential systems, J. Phys. A: Math. Gen. 32 (1999) 5133–5139.
[17] P.C. Moan, On backward error analysis and Nekhoroshev stability in the numerical analysis of conservative systems of ODEs, Ph.D. Thesis, University of Cambridge, 2002.
[18] P.C. Moan, J.A. Oteo, Convergence of the exponential Lie series, J. Math. Phys. 42 (2001) 501–508.

[19] H. Munthe-Kaas, B. Owren, Computations in a free Lie algebra, Philos. Trans. Roy. Soc. A 357 (1999) 957–982.
[20] M. Newman, W. So, R.C. Thompson, Convergence domains for the Campbell–Baker–Hausdorff formula, Linear and Multilinear Algebra 24 (1989) 301–310.
[21] J.A. Oteo, The Baker–Campbell–Hausdorff formula and nested commutator identities, J. Math. Phys. 32 (1991) 419–424.
[22] M. Postnikov, Lie Groups and Lie Algebras. Semester V of Lectures in Geometry, URSS Publishers, Moscow, 1994.
[23] W. So, Exponential formulas and spectral indices, Ph.D. Thesis, University of California at Santa Barbara, 1991.
[24] R.C. Thompson, Convergence proof for Goldberg's exponential series, Linear Algebra Appl. 121 (1989) 3–7.
[25] V.S. Varadarajan, Lie Groups, Lie Algebras, and Their Representations, Springer, New York, 1984.
[26] V.A. Vinokurov, The logarithm of the solution of a linear differential equation, Hausdorff's formula, and conservation laws, Soviet Math. Dokl. 44 (1992) 200–205.
[27] J. Wei, Note on global validity of the Baker–Hausdorff and Magnus theorems, J. Math. Phys. 4 (1963) 1337–1341.
[28] R.M. Wilcox, Exponential operators and parameter differentiation in quantum physics, J. Math. Phys. 8 (1967) 962–982.