London Mathematical Society ISSN 1461–1570

ON THE DIMENSION OF CERTAIN GRADED LIE ALGEBRAS ARISING IN GEOMETRIC INTEGRATION OF DIFFERENTIAL EQUATIONS

ARIEH ISERLES AND ANTONELLA ZANNA

Abstract

Many discretization methods for differential equations that evolve in Lie groups and homogeneous spaces advance the solution in the underlying Lie algebra. The main expense of computation is the calculation of commutators, a task that can be made significantly cheaper by the introduction of appropriate bases of function values and exploitation of redundancies inherent in a Lie-algebraic structure by means of graded spaces. In many Lie groups of practical interest a convenient alternative to the exponential map is a Cayley transformation and the subject of this paper is the investigation of graded algebras that occur in this context. To this end we introduce a new concept, a *hierarchical algebra,* a Lie algebra equipped with a countable number of m-nary multilinear operations which display alternating symmetry and a 'hierarchy condition'. We present explicit formulae for the dimension of graded subspaces of free hirarchical algebras and an algorithm for the construction of their basis. The paper is concluded by reviewing a number of applications of our results to numerical methods in a Lie-algebraic setting.

1. *Graded algebras and Lie-group methods*

1.1. *Geometric integration and general Lie-group solvers*

An increasing attention has been paid in recent years to discretization of differential equations that evolve on smooth manifolds. The main reason is that invariants and conservation laws of a differential system can be phrased by restricting the configuration space to a manifold. Discretization methods that respect manifold structure are an important example of *geometric integrators,* computational methods that preserve the underlying geometry and qualitative attributes of the differential system [**1**]. Perhaps the most ubiquitous (and arguably most important) type of a manifold that occurs in practical applications is a *Lie group.* Thus, numerous differential equations of practical interest evolve in Lie groups and the conservation of this feature under discretization is often valuable and sometimes crucial:

- Equations of classical mechanics, robotics and control theory often evolve on the special orthogonal group $\mathrm{SO}_n(\mathbb{R})$. This is also the case with several processes of interest in numerical algebra and computational dynamics;
- Hamiltonian systems are associated with evolution in the symplectic group Sp_n;

Received 31 May 1999.
1991 Mathematics Subject Classification Primary 17B70, Secondary 22E60, 65L99

- Many equations in special relativity evolve in the Lorenz group $\mathrm{SO}_{3,1}(\mathbb{R})$;
- Much of quantum theory can be formulated as evolution in the unitary group $\mathrm{U}_n(\mathbb{C})$;
- Conservation of volume, ubiquitous in equations of incompressible flow, corresponds to evolution in the special linear group $\mathrm{SL}_n(\mathbb{R})$.

Detailed references can be found in [**1**]. The central role of Lie groups in geometric integration is underscored by the fact that, as long as we can discretize a flow evolving in a Lie group G, say, we can also discretize differential systems evolving in an arbitrary homogeneous space which is acted upon by G [**17**]. Thus, for example, a numerical scheme that is guaranteed to respect $\mathrm{O}_n(\mathbb{R})$ structure can be easily adapted to evolve on a sphere, say, a torus, an isospectral manifold, a Stiefel or a Grassmann manifold.

Although exceptions exist [**4, 14, 19**], most Lie-group integrators follow a set pattern. A flow in a finite-dimensional Lie group G can be always written in the form

$$Y' = A(t,Y)Y, \quad t \geqslant 0, \qquad Y(t_0) = Y_0 \in G, \tag{1}$$

where A is an adequately-smooth map from $[t_0,\infty)\times G$ to $\mathfrak{g}$, the Lie algebra corresponding to G. Representing $Y(t) = \exp[\Omega(t)]Y_0$, we can replace (1) with an equation *evolving in the Lie algebra,* namely the so-called 'dexpinv equation'

$$\Omega' = \mathrm{dexp}_{\Omega}^{-1} A(t, \mathrm{e}^{\Omega}Y_0) = \sum_{k=0}^{\infty} \frac{B_k}{k!} \mathrm{ad}\,_{\Omega}^{k} A(t, \mathrm{e}^{\Omega}Y_0), \quad t \geqslant t_0, \qquad \Omega(t_0) = O, \tag{2}$$

where $\{B_k\}_{k\geqslant 0}$ are Bernoulli numbers while $\mathrm{ad}\,_X$ is the adjoint operator in $\mathfrak{g}$,

$$\mathrm{ad}\,_X^m = \overbrace{[X,[X,\ldots,[X,Y]\cdots]]}^{m \text{ times}}, \qquad m \geqslant 0.$$

Instead of solving (1) directly, we apply a numerical scheme to (2) and translate the outcome to G with the exponential map. The reasoning underlying this approach is that, unlike nontrivial Lie groups, Lie algebras are *linear* spaces. Discretization methods that, as most numerical algorithms do, restrict themselves to linear operations and commutation, are guaranteed (within machine accuracy) to evolve in the Lie algebra! Although an implementation of the exponential map calls for much care and algorithmic dexterity [**2**], this procedure nonetheless provides a convenient avenue toward Lie-group solvers.

Perhaps the most versatile example of a Lie-group solver that follows the above pattern is provided by *Runge–Kutta–Munthe-Kaas* methods [**15**]. The main idea is to apply a ν-stage Runge–Kutta method

$$\begin{array}{c|cccc}
c_1 & a_{1,1} & a_{1,2} & \cdots & a_{1,\nu} \\
c_2 & a_{2,1} & a_{2,2} & \cdots & a_{2,\nu} \\
\vdots & \vdots & \vdots & & \vdots \\
c_\nu & a_{\nu,1} & a_{\nu,2} & \cdots & a_{\nu,\nu} \\
\hline
 & b_1 & b_2 & \cdots & b_\nu
\end{array}$$

(cf. [**9**] for Runge–Kutta formalism) to the Lie-algebraic equation (2). Denoting the approximate value of the solution of (1) at time t_N by Y_N and letting $h = t_{N+1} - t_N$, we thus obtain a scheme of the form

$$\left.\begin{aligned} \Phi_m &= h\sum_{l=1}^{\nu} a_{m,k}\Phi_l, \\ K_m &= \mathrm{dexp}^{-1}_{\Phi_m} A(t_N + c_m h, \mathrm{e}^{\Phi_m} Y_N), \end{aligned}\right\} \qquad m = 1, 2, \ldots, \nu,$$
$$\Delta = h\sum_{m=1}^{\nu} b_m K_m, \tag{3}$$
$$Y_{N+1} = \mathrm{e}^{\Delta} Y_N.$$

It is of the same order of precision as the original Runge–Kutta method but it is guaranteed to evolve on G.

Another important example of a Lie-group solver is the *Magnus expansion* which, for linear equations (1) reads

$$\begin{aligned} \Omega(t) = {}& \int_{t_0}^{t} A(\xi)\mathrm{d}\xi - \tfrac{1}{2}\int_{t_0}^{t}\!\!\int_{t_0}^{\xi_1} [A(\xi_2), A(\xi_1)]\mathrm{d}\xi_2\mathrm{d}\xi_1 \\ & + \tfrac{1}{4}\int_{t_0}^{t}\!\!\int_{t_0}^{\xi_1}\!\!\int_{t_0}^{\xi_2} [[A(\xi_3), A(\xi_2)], A(\xi_1)]\mathrm{d}\xi_3\mathrm{d}\xi_2\mathrm{d}\xi_1 \\ & + \tfrac{1}{2}\int_{t_0}^{t}\!\!\int_{t_0}^{\xi_1}\!\!\int_{t_0}^{\xi_1} [A(\xi_3), [A(\xi_2), A(\xi_1)]]\mathrm{d}\xi_3\mathrm{d}\xi_2\mathrm{d}\xi_1 + \cdots, \end{aligned} \tag{4}$$

originally introduced by Wilhelm Magnus [**13**]. Recently it has been cast by Iserles & Nørsett into a form amenable for numerical treatment [**11**]. This has included recursive generation of expansion terms and an effective treatment of multivariate integrals over polytopes by numerical quadrature. A nonlinear version of the Magnus expansion, using collocation, is due to Zanna [**20**].

Yet another example of a Lie-group solver bases on reducing the problem to a Lie algebra is the *Fer expansion*

$$Y(t) = \mathrm{e}^{\Omega^{[0]}(t)}\mathrm{e}^{\Omega^{[1]}(t)} \cdots Y_0$$

[**7**]. In the linear case the Lie-algebra functions $\Omega^{[i]}$ can be obtained recursively [**8**] and also this procedure can be generalized to nonlinear equations [**20**].

1.2. *Quadratic Lie groups and the Cayley transform*

Runge–Kutta–Munthe-Kaas schemes, as well as Magnus and Fer expansions, can be applied to all finite-dimensional Lie groups, regardless of their structure. This is their strength yet it also represents a potential shortcoming. Many Lie groups possess structure that can be exploited to construct more effective discretization methods. A case in point are *quadratic Lie groups* [**10**]: given $J \in \mathrm{GL}_n(\mathbb{R})$, the set of $n \times n$ real nonsingular matrices, we let

$$G = \{Y \in \mathrm{GL}_n(\mathbb{R}) \,:\, YJY^{\mathrm{T}} = J\}. \tag{5}$$

Note that $\mathrm{SO}_n(\mathbb{R})$, Sp_n and $\mathrm{SO}_{3,1}(\mathbb{R})$ are all either quadratic Lie groups or connected components thereof, while $\mathrm{U}_n(\mathbb{C})$ and other complex quadratic groups can be obtained by replacing (5) with $\{Y \in \mathrm{GL}_n(\mathbb{C}) \,:\, YJY^{\mathrm{H}} = J\}$. Thus, quadratic Lie groups account for four out of our five examples of matrix Lie groups that are relevant in applications. This emphasises the importance of this construct and justifies special attention being paid to quadratic Lie-group methods.

The main structural feature of quadratic Lie groups that can be exploited in the context of Lie-group solvers is that we do not need to use (or approximate) the matrix exponential in order to map G to its Lie algebra

$$\mathfrak{g} = \{X \in \mathfrak{gl}_n(\mathbb{R}) \,:\, XJ + JX^{\mathrm{T}} = O\}. \tag{6}$$

An alternative is to employ the *Cayley transform,* letting $Y(t) = \mathrm{cay}[\Omega(t)]Y_0$, $t \geqslant t_0$, where

$$\mathrm{cay}X = (I - \tfrac{1}{2}X)^{-1}(I + \tfrac{1}{2}X). \tag{7}$$

Thus, the evaluation the exponential – a costly procedure – is no longer required and it can be replaced by the much cheaper matrix inversion. Moreover, in place of the dexpinv equation (2) one obtains the considerably simpler *dcayinv equation*

$$\Omega' = \mathrm{dcay}_{\Omega}^{-1} A(t, \mathrm{cay}[\Omega]Y_0) = A - \tfrac{1}{2}[\Omega, A] - \tfrac{1}{4}\Omega A \Omega, \quad t \geqslant t_0, \qquad \Omega(t_0) = O. \tag{8}$$

This has been recognized in [**12**] and further exploited in [**6, 5**]. Recently, Iserles introduced a Cayley-transform equivalent of a Magnus expansion [**10**]. Specifically, for linear equations Ω is expanded in terms of integrals, commutators and *symmetric products* $PQR + RQP$,

$$\begin{aligned}\Omega(t) = {} & \int_{t_0}^{t} A(\xi)\mathrm{d}\xi - \tfrac{1}{2}\int_{t_0}^{t}\int_{t_0}^{\xi_1} [A(\xi_2), A(\xi_1)]\mathrm{d}\xi_2\mathrm{d}\xi_1 \\ & + \tfrac{1}{4}\int_{t_0}^{t}\int_{t_0}^{\xi_1}\int_{t_0}^{\xi_2} [[A(\xi_3), A(\xi_2)], A(\xi_1)]\mathrm{d}\xi_3\mathrm{d}\xi_2\mathrm{d}\xi_1 \\ & - \tfrac{1}{4}\int_{t_0}^{t}\int_{t_0}^{\xi_1}\int_{t_0}^{\xi_1} A(\xi_2)A(\xi_1)A(\xi_3)\mathrm{d}\xi_3\mathrm{d}\xi_2\mathrm{d}\xi_1 + \cdots .\end{aligned} \tag{9}$$

The Cayley expansion (9) has a number of advantages compared to the Magnus expansion (4): the number of terms with any given number of integrals is smaller, the radius of convergence is more generous and, as we have already mentioned, there is no need whatsoever to evaluate a matrix exponential. The general form of (9) has been investigated in great detail in [**10**] by identifying each expanion term with a *bicolour rooted tree,* whereby expansion coefficients can be generated recursively and counted by combinatorial arguments.

1.3. *Quadrature and graded Lie algebras*

Practical implementation of a Magnus or Cayley expansion requires the truncation of the infinite series and an approximation of integrals by quadrature. On the face of it, the latter task is likely to be prohibitively expensive since each expansion term requires an integration in a different multivariate polytope. Using a traditional cubature method would have required a truly enormous number of function evaluations [**3**]. Fortunately, the structure of integrals featuring in either (4) or (9) can be exploited to a very good effect and it leads to a remarkably efficient quadrature [**11**]. Assuming without loss of generality that $t_0 = 0$ and advancing the integration with time step $h > 0$, all such integrals are of the form

$$\mathscr{I}(h) = \int_{h\mathcal{S}} \mathcal{L}(A(\xi_1), A(\xi_2), \ldots, A(\xi_r))\mathrm{d}\xi_r\mathrm{d}\xi_{r-1}\cdots\mathrm{d}\xi_r,$$

where $\mathcal{L}$ is a multilinear form, while $\mathcal{S} \subset \mathbb{R}$ is a polytope,

$$\mathcal{S} = \{\boldsymbol{\xi} \in \mathbb{R}^r \,:\, 0 \leqslant \xi_i \leqslant \xi_{q_i},\ i = 1, 2, \ldots, r\},$$

with $\xi_0 = 1$ and $q_i \leqslant \max\{0, i-1\}$, $i = 1, 2, \ldots, r$. Let $c_1, c_2, \ldots, c_\nu$ be arbitrary distinct *quadrature points* in $[0,1]$ and let $A_k = A(c_k h)$, $k = 1, 2, \ldots, \nu$. We approximate $\mathscr{I}(h)$ by the quadrature formula

$$\mathscr{Q}(h) = h^r \sum_{k_1=1}^{\nu} \sum_{k_2=1}^{\nu} \cdots \sum_{k_r=1}^{\nu} b_{\boldsymbol{k}} \mathcal{L}(A_{k_1}, A_{k_2}, \ldots, A_{k_r}), \tag{10}$$

where the weights $b_{\boldsymbol{k}}$ can be constructed explicitly by integrating cardinal Lagrange's interpolation polynomials [**10, 11**]. The crucial observation is that the order of accuracy of (10) is precisely that of the univariate Gaussian quadrature in $[0,1]$ (with constant weight function) using the nodes $c_1, c_2, \ldots, c_\nu$. In particular, letting the nodes equal the zeros of the νth Legendre polynomial, shifted to $[0,1]$, we obtain

$$\mathscr{Q}(h) = \mathscr{I}(h) + \mathcal{O}\left(h^{2\nu+1}\right)$$

for all sufficiently smooth matrix functions A. Note that we use the very same ν function values in all the many integrals occurring in the expansion!

Quadrature formulae (10) require a very small number of functions evaluations to compute a Magnus or a Cayley expansion, but this is traded off for substantial number of matrix products: the number of combinations increases very fast as a function or r and ν and each such combination requires the computation of a multilinear form $\mathcal{L}(A_{k_1}, A_{k_2}, \ldots, A_{k_r})$. Fortunately, very significant savings in linear algebra can be obtained by exploiting three mechanisms. Firstly, combinations in one multilinear form might be identical (up to know sign change) to different combinations of another multilinear form [**11**]. Secondly, expressed in a suitable basis, many combinations can be disregarded since they are of magnitude $\mathcal{O}\left(h^{q+1}\right)$ where q exceeds the order of the method. Thirdly, a Lie algebra possesses a wide range of redundancies and symmetries which lead to numerous linear dependencies among values of $\mathcal{L}(A_{k_1}, A_{k_2}, \ldots, A_{k_r})$ for different combinations $\boldsymbol{k}$. The last two mechanisms have been analysed in great detail by Munthe-Kaas and Owren in [**16**] in the context of Magnus expansions. The main purpose of the present paper is to extend their analysis to the realm of Cayley expansions and other algorithms specific to quadratic Lie groups.

The first crucial idea in [**16**] is to express the function values $\{A_1, A_2, \ldots, A_\nu\}$ using a different basis, which renders the order of each term more transparent. Specifically, we use the basis $\{B_0, B_1, \ldots, B_{\nu-1}\}$, where

$$\sum_{l=0}^{\nu-1} \frac{1}{l!}(c_k - \tfrac{1}{2})^l B_l = A_k, \qquad k = 1, 2, \ldots, \nu.$$

The quadrature formula (10) assumes the form

$$\widetilde{\mathscr{Q}}(h) = h^r \sum_{l_1=0}^{\nu-1} \sum_{l_2=0}^{\nu-1} \cdots \sum_{l_r=0}^{\nu-1} \tilde{b}_{\boldsymbol{l}} \mathcal{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_r}), \tag{11}$$

where

$$\tilde{b}_{\boldsymbol{l}} = \frac{1}{l_1! l_2! \cdots l_r!} \int_{\mathcal{S}} \prod_{j=1}^{r} (\xi_j - \tfrac{1}{2})^{l_j} \, \mathrm{d}\xi_r \mathrm{d}\xi_{r-1} \cdots \mathrm{d}\xi_1.$$

Assuming $A \in \mathrm{C}^\infty$, it is trivial to verify that $B_l = h^l A^{(l)}(\frac{1}{2}h)$, therefore $B_l = \mathcal{O}\left(h^l\right)$, for $l = 0, 1, \ldots, \nu - 1$. Constructing $\mathcal{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_r})$ from commutators and symmetric products, it follows at once that

$$h^r \mathcal{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_r}) = \mathcal{O}(h^{|\boldsymbol{l}|+r}) \qquad \text{where} \qquad |\boldsymbol{l}| = \sum_{j=1}^{r} l_j.$$

We say that such a term is of *grade* $|\boldsymbol{l}| + r$. Typically, $c_1, c_2, \ldots, c_\nu$ are taken as Gauss–Legendre points, whence (11) is an order-(2ν) quadrature. In that case, we can throw away with impunity all terms of grade exceeding 2ν and replace (11) by

$$\widehat{\mathscr{Q}}(h) = h^r \sum_{|\boldsymbol{l}| \leqslant 2\nu - r} \tilde{b}_{\boldsymbol{l}} \mathcal{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_r}). \tag{12}$$

Inasmuch as this procedure gets rid of a very significant proportion of expansion terms, further savings take place by exploiting linear dependencies in the underlying Lie algebra. In the context of 'plain' Lie algebras, the subject of [**16**], the two dependency-generating mechanisms are the skew-symmetry $[Y, X] + [X, Y] = O$ and the *Jacobi identity* $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = O$, valid for all $X, Y, Z \in \mathfrak{g}$.

Given the generators $B_0, B_1, \ldots, B_{\nu-1}$, we assign to each B_l a grade $\omega(B_l) \in \mathbb{N}$. Grades propagate naturally during the construction of a free Lie algebra: if $\omega(X_i) = w_i$, $i = 1, 2$, $X_1 \neq X_2$, then $\omega([X_1, X_2]) = w_1 + w_2$. Let $\mathscr{K}_m$ be the linear space of all words of length m in the free Lie algebra. In the case $\omega(B_l) \equiv 1$ the dimension of the linear space spanned by $\mathscr{K}_m$ is known and given by the familiar *Witt formula.* Moreover, it is possible to form recursively a basis of this space (e.g. a *Hall* or a *Lyndon* basis). A major result of [**16**] is a generalisation of the Witt formula and of Hall and Lyndon bases to the case when the grades are arbitrary natural numbers. (In the present case we are interested in $w(B_l) = l + 1$, $l = 0, 1, \ldots, \nu - 1$, but other choices of grades are of an independent interest, cf. [**16**].) Further observing that $b_{\boldsymbol{l}} = 0$ for terms of an even grade, Munthe-Kaas and Owren were able to reduce the number of terms in a Magnus expansion by a remarkable factor. For example, an order-10 method requires just 73 terms, instead of 1256567 in a naive implementation!

The main purpose of the present paper is to extend the theory in [**16**] to Lie algebras of the special form (6). The formation of a free algebra for (6) allows the use of both commutators and symmetric products, while its special structure implies the existence of linear dependencies which are absent in other Lie algebras. In Section 2 we define a *hierarchical algebra,* an algebraic construct which encapsulates the special features of the Lie algebra (6), and prove that every hierarchical algebra is an extension of a Lie algebra. This is followed in Section 3 by a determination of the dimension and construction of a basis of a linear space of a given grade within the realm of freely-generated hierarchical algebras. It is interesting to note that not just the method of analysis but the explicit form of our formulae are very different from the Witt and Munthe-Kaas–Owren framework. Finally, in Section 4 we exhibit a number of examples that implement our results, not just in the context of Cayley expansions (9) but also for other computational constructs for quadratic Lie algebras.

2. *Hierarchical algebras*

Let $\mathbb{F}$ be a field of characteristic 0. We recall that an algebra over $\mathbb{F}$ is a triple $(A, +, \circ)$, whereby $(A, +)$ is an abelian group and '$\circ$' is an internal binary operation over A which is linear in both arguments. In particular, a Lie algebra $\mathfrak{g}$ is an algebra whereby the binary operation '$\circ$' is usually represented by a bracket $[\cdot,\cdot]$ and obeys:

a. Skew-symmetry: $\forall X, Y \in \mathfrak{g}$,

$$[X,Y] + [Y,X] = O; \tag{13}$$

b. Jacobi identity: $\forall X, Y, Z \in \mathfrak{g}$,

$$[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = O. \tag{14}$$

For the time being, let $\mathfrak{g}$ represent a set, not necessarily associated with a Lie algebra. Assume that $(\mathfrak{g},+)$ is an abelian group and let us introduce the following family of m-nary operations,

$$[\![\cdot,\ldots,\cdot]\!]_m : \underbrace{\mathfrak{g}\times\cdots\times\mathfrak{g}}_{m\text{ copies}} \to \mathfrak{g}, \qquad m = 1,2,\ldots. \tag{15}$$

Assume that the m-nary operation (15) obeys

a. Alternating symmetry: $\forall X_1,\ldots,X_m \in \mathfrak{g}$,

$$[\![X_1,X_2,\ldots,X_m]\!]_m + (-1)^m[\![X_m,X_{m-1},\ldots,X_1]\!]_m = O. \tag{16}$$

b. Multilinearity: $\forall X_1,\ldots,X_{l-1},X_{l+1},\ldots X_m, Y, Z, \in \mathfrak{g}$ and $\forall \alpha,\beta \in \mathbb{F}$,

$$\begin{aligned}&[\![X_1,\ldots,X_{l-1},\alpha Y+\beta Z,X_{l+1},\ldots,X_m]\!]_m \\ &= \alpha[\![X_1,\ldots,X_{l-1},Y,X_{l+1},\ldots,X_m]\!]_m + \beta[\![X_1,\ldots,X_{l-1},Z,X_{l+1},\ldots,X_m]\!]_m\end{aligned} \tag{17}$$

for all $1 \leqslant l \leqslant m$.

c. Hierarchy condition: For every $n \geqslant 1$, for every index $1 \leqslant l \leqslant m$, and elements $X_1,\ldots,X_{l-1},X_{l+1},\ldots,X_m \in \mathfrak{g}$ and $Y_1,Y_2,\ldots,Y_n \in \mathfrak{g}$, then

$$\begin{aligned}&[\![X_1,\ldots,X_{l-1},[\![Y_1,Y_2,\ldots,Y_n]\!]_n,X_{l+1},\ldots,X_m]\!]_m \\ &= [\![X_1,\ldots,X_{l-1},Y_1,\ldots,Y_n,X_{l+1},\ldots,X_m]\!]_{m+n-1} \\ &\quad - (-1)^n[\![X_1,\ldots,X_{l-1},Y_n,\ldots,Y_1,X_{l+1},\ldots,X_m]\!]_{m+n-1}.\end{aligned} \tag{18}$$

We remark that the hierarchy condition (18) links the m-nary product with similar products of higher order, thus establishing a hierarchy in the family of m-nary operations.

Definition 1. A *hierarchical algebra* is a family $(\mathfrak{g},+,\{[\![\cdots]\!]_m\}_{m=1}^{\infty})$, whereby $(\mathfrak{g},+)$ is an abelian group and each m-nary operation obeys the alternating symmetry condition (16), the multilinearity (17) and the hierarchy condition (18).

The use of the symbol $\mathfrak{g}$, usually reserved for Lie algebras, in a hierarchical algebra setting is justified by the result below.

Theorem 1. *If* $(\mathfrak{g},+,\{[\![\cdots]\!]_m\}_{m=1}^{\infty})$ *is a hierarchical algebra, then* $(\mathfrak{g},+,[\![\cdots]\!]_2)$ *is a Lie algebra.*

Proof. It is sufficient to prove that the binary product $[\![\cdots]\!]_2$ obeys the Jacobi identity since skew-symmetry is an immediate consequence of the alternating symmetry condition (16).

Consider three arbitrary elements $X, Y, Z \in \mathfrak{g}$. Because of the hierarchy condition,

$$\begin{aligned}[\![X,[\![Y,Z]\!]_2]\!]_2 &= [\![X,Y,Z]\!]_3 - [\![X,Z,Y]\!]_3, \\ [\![Y,[\![Z,X]\!]_2]\!]_2 &= [\![Y,Z,X]\!]_3 - [\![Y,X,Z]\!]_3, \\ [\![Z,[\![X,Y]\!]_2]\!]_2 &= [\![Z,X,Y]\!]_3 - [\![Z,Y,X]\!]_3.\end{aligned}$$

Furthermore, the alternating symmetry condition implies that

$$[\![X_1, X_2, X_3]\!]_3 = [\![X_3, X_2, X_1]\!]_3$$

for every $X_1, X_2, X_3 \in \mathfrak{g}$. Hence we deduce that

$$\begin{aligned}
[\![X, [\![Y, Z]\!]_2]\!]_2 + [\![Y, [\![Z, X]\!]_2]\!]_2 &= [\![X, Y, Z]\!]_3 - [\![X, Z, Y]\!]_3 + [\![Y, Z, X]\!]_3 - [\![Y, X, Z]\!]_3 \\
&= [\![X, Y, Z]\!]_3 - [\![Y, X, Z]\!]_3 = -[\![Z, [\![X, Y]\!]_2]\!]_2,
\end{aligned}$$

from which the Jacobi identity follows. □

Theorem 2. *Every quadratic matrix Lie algebra is a hierarchical algebra.*

Proof. Consider the operation

$$[\![X_1, \dots, X_m]\!]_m = X_1 X_2 \cdots X_m - (-1)^m X_m X_{m-1} \cdots X_1$$

with the usual matrix-matrix multiplication. Then, for every $m \geqslant 1$,

$$\begin{aligned}
&[\![X_1, \dots, X_m]\!]_m + (-1)^m [\![X_m, \dots, X_1]\!]_m \\
&= X_1 X_2 \cdots X_m - (-1)^m X_m X_{m-1} \cdots X_1 + (-1)^m (X_m \cdots X_1 - (-1)^m X_1 \cdots X_m) \\
&= X_1 X_2 \cdots X_m - (-1)^m X_m X_{m-1} \cdots X_1 + (-1)^m X_m \cdots X_1 - X_1 \cdots X_m = O,
\end{aligned}$$

hence alternating symmetry. The multilinearity of the above m-nary operation is immediate.

Next, let us prove the hierarchy condition. To this end, given $l \in \{1, 2, \dots, m\}$ and the elements $X_1, \dots, X_{l-1}, X_{l+1}, \dots X_m$ and $Y_1, \dots Y_n \in \mathfrak{g}$, we have

$$\begin{aligned}
&[\![X_1, \dots, X_{l-1}, [\![Y_1, \dots, Y_n]\!]_n, X_{l+1}, \dots, X_m]\!]_m \\
&= X_1 \cdots X_{l-1} Y_1 \cdots Y_n X_{l+1} \cdots X_m - (-1)^n X_1 \cdots X_{l-1} Y_n \cdots Y_1 X_{l+1} \cdots X_m \\
&\quad - (-1)^m X_m \cdots X_{l+1} Y_1 \cdots Y_n X_{l-1} \cdots X_1 \\
&\quad - (-1)^{m+n-1} X_m \cdots X_{l+1} Y_n \cdots Y_1 X_{l-1} \cdots X_1.
\end{aligned}$$

However,

$$\begin{aligned}
&X_1 \cdots X_{l-1} Y_1 \cdots Y_n X_{l+1} \cdots X_m - (-1)^{m+n-1} X_m \cdots X_{l+1} Y_n \cdots Y_1 X_{l-1} \cdots X_1 \\
&= [\![X_1, \cdots, X_{l-1}, Y_1 \cdots Y_n, X_{l+1}, \cdots, X_m]\!]_{m+n-1},
\end{aligned}$$

and, by the same token, we observe that the two middle terms of the right hand side of the above equation compose into the m-nary product

$$-(-1)^n [\![X_1, \cdots, X_{l-1}, Y_n, \cdots, Y_1, X_{l+1}, \cdots, X_m]\!]_{m+n-1}.$$

Hence (18) follows.

It remains to prove that $[\![\cdots]\!]_m$ is internal operation on $\mathfrak{g}$. By induction on m: for $m = 1$, we have $[\![X]\!]_1 = 2X \in \mathfrak{g}$ for any $X \in \mathfrak{g}$. Next, let us assume that for every $k \leqslant m - 1$ the assertion is true. We distinguish two cases: even and odd values of m. Considering first the case when m is even and let J be the same matrix as in (5). We have

$$\begin{aligned}
[\![X_1, \dots, X_{m-1}, X_m]\!]_m J &= (X_1 \cdots X_{m-1} X_m - X_m X_{m-1} \cdots X_1) J \\
&= (-1)^m J (X_1^{\mathrm{T}} \cdots X_m^{\mathrm{T}} - X_m^{\mathrm{T}} \cdots X_1^{\mathrm{T}}) \\
&= J [\![X_1^{\mathrm{T}}, \cdots, X_m^{\mathrm{T}}]\!]_m = -J [\![X_m^{\mathrm{T}}, \dots X_1^{\mathrm{T}}]\!]_m \\
&= -J [\![X_1, \dots X_m]\!]_m^{\mathrm{T}},
\end{aligned}$$

by virtue of the alternating symmetry. Hence

$$[\![X_1, \dots, X_{m-1}, X_m]\!]_m \in \mathfrak{g}.$$

Finally, when m is odd,

$$\begin{aligned}[\![X_1,\ldots,X_m]\!]_m J &= (X_1\cdots X_m + X_m\cdots X_1)J\\ &= -J(X_1^{\mathrm{T}}\cdots X_m^{\mathrm{T}} + X_m^{\mathrm{T}}\cdots X_1^{\mathrm{T}})\\ &= -J[\![X_1,\ldots,X_m]\!]_m^{\mathrm{T}},\end{aligned}$$

hence $[\![X_1,\ldots,X_{m-1},X_m]\!]_m \in \mathfrak{g}$. Similar procedure for complex groups, whereby the transpose operator is replaced by the Hermitian operator. It follows that $[\![X_1,\ldots,X_m]\!]_m \in \mathfrak{g}$ for all values of m. The proof is complete. □

The importance of hierarchical algebras is therefore self-evident: for quadratic matrix Lie algebras, elements of the form $[\![X_1,\ldots,X_m]\!]_m$ are precisely the building blocks of the Cayley transform (9).

It is useful to introduce the following definitions.

Definition 2. If $\mathfrak{g},\mathfrak{h}$ are two hierarchical algebras, we say that $\pi:\mathfrak{g}\to\mathfrak{h}$ is a hierarchical algebra homomorphism if

$$\pi([\![X_1,\ldots,X_m]\!]_m) = [\![\pi(X_1),\ldots,\pi(X_m)]\!]_m \in \mathfrak{h}, \qquad \forall X_1,\ldots,X_m\in\mathfrak{g},$$

for every $m=1,2,\ldots$.

Similarly to the definition of *free Lie algebras* [**16**], we can introduce the definition for a *free hierarchical algebra*.

Definition 3. Let I be a finite or a countable set of indices. We say that the hierarchical algebra $\mathfrak{g}$ is *free* over I if

a. $\forall i\in I$ there corresponds $X_i\in\mathfrak{g}$, and $X_i\neq X_j$ except when $i=j$.
b. For any hierarchical algebra $\mathfrak{h}$ for which there exists a function $i\to Y_i\in\mathfrak{h}$, there exists a unique $\pi:\mathfrak{g}\to\mathfrak{h}$, a hierarchical algebra homomorphism, such that $\pi(X_i)=Y_i$.

Given a set $\{X_i\}_{i\in I}$, denote by $\mathfrak{g}$ the set obtained by the X_is and all the possible combination of those by means of each m-nary operation. By construction, $\mathfrak{g}$ is a hierarchical algebra, and, for any other hierarchical algebra $\mathfrak{h}$ containing the X_is, it is true that $\mathfrak{g}\subseteq\mathfrak{h}$. Hence $\mathfrak{g}$, which is unique up to an isomorphism, is the smallest hierarchical algebra containing the X_is and will be called hierarchical algebra *generated* by the family $\{X_i\}_{i\in I}$ Note that such $\mathfrak{g}$ is free over I.

The homomorphism π is a representation of a free hierarchical algebra $\mathfrak{g}$ in $\mathfrak{h}$. In concrete terms, if $\mathfrak{h}$ is a quadratic Lie algebra, for a given set of indices I the $\pi(X_i)$s might be associated to some function evaluations of $A(t)$ or to some linear combination of those. The X_i can be regarded as *letters* of an alphabet, and the term $[\![X_{i_1},\ldots,X_{i_m}]\!]_m$ is a word of length m generated by $X_{i_1},\ldots,X_{i_m}$. In the next section we shall be interested in counting the number of words of given length as well counting the words of a given weight, whenever a weight is initially associated to each of the generators X_i.

3. *The dimension and a basis of graded linear subspaces*

3.1. *Graded linear spaces*

Let $\mathscr{X}=\{X_1,X_2,\ldots,X_s\}$ be a set of generators of a free hierarchical algebra $\mathfrak{g}$ and suppose that we are given a map $w:\mathscr{X}\to\mathbb{N}$. The natural number $w(X_l)$ is called the *grade* of X_l. The definition of a grade is propagated in the free hierarchical algebra in a natural manner: provided that $w(Y_j) = v_j$, $j = 1, 2, \ldots, m$, we set

$$w([\![Y_1, Y_2, \ldots, Y_m]\!]_m) = \sum_{j=1}^{m} v_j. \tag{19}$$

We assume without loss of generality that $w(X_k) \leqslant w(X_l)$ for $k < l$ and stipulate that $w(X_1) = 1$. The free hierarchical algebra splits into a direct sum of linear subspaces,

$$\mathfrak{g} = \bigoplus_{r \geqslant 1} \mathscr{K}_r^s(\boldsymbol{w}),$$

where $\boldsymbol{w} = [w(X_1), w(X_2), \ldots, w(X_s)]$ and $\mathscr{K}_r^s(\boldsymbol{w})$ is the linear subspace generated by all the elements in $\mathfrak{g}$ of grade equal to $r \geqslant 1$. We pose the following two problems:

1. What is $\dim \mathscr{K}_r^s(\boldsymbol{w})$? and
2. How to construct a basis of $\mathscr{K}_r^s(\boldsymbol{w})$?

The *raison d'etre* for our interest in the dimension and a basis of $\mathscr{K}_r^s(\boldsymbol{w})$ is implicit in the discussion of Section 1. Thus, suppose that the function values $A_1, A_2, \ldots, A_\nu$ generate a free hierarchical algebra, while the multilinear form in (10) has been derived by hierarchical-algebra operations. Letting $w(A_l) \equiv 1$, we deduce that

$$\mathcal{L}(A_{k_1}, A_{k_2}, \ldots, A_{k_r}) \in \mathscr{K}_r^\nu(\mathbf{1})$$

for all $1 \leqslant k_1, k_2, \ldots, k_r \leqslant \nu$. This corresponds to the observation that $A_l = \mathcal{O}(1)$, $1 \leqslant l \leqslant \nu$ implies that

$$h^r \mathcal{L}(A_{k_1}, A_{k_2}, \ldots, A_{k_r}) = \mathcal{O}(h^r)$$

and the quadrature (10) is composed out of $\mathcal{O}(h^r)$ terms. Having evaluated a basis for $\mathscr{K}_r^\nu(\mathbf{1})$, we can construct all the terms in (10) out of basis elements by linear combinations. Therefore, it is the dimension of the basis, rather than the number of different combinations in (10) and of different r-dimensional integrals that determines the cost of the calculation!

Considerably more important is the case when $A_1, A_2, \ldots, A_\nu$ are replaced with the linear combinations $B_0, B_1, \ldots, B_{\nu-1}$. We let $w(B_l) = l + 1$, consistently with $hB_l = \mathcal{O}(h^{l+1})$, $l = 0, 1, \ldots, \nu - 1$. Let $\boldsymbol{w}^\star = [1, 2, \ldots, \nu]$. Since

$$\mathscr{K}_r^\nu(\boldsymbol{w}^\star) \ni h^r \mathcal{L}(B_{l_1}, B_{l_2}, \ldots, B_{l_r}) = \mathcal{O}(h^{|\boldsymbol{l}|+r}),$$

where $|\boldsymbol{l}| = \sum l_i$, we need to retain in (11) only terms of grade $\leqslant 2\nu$. Therefore, the cost of all the quadrature formulae needed in the implementation of the Cayley expansion (9) is reflected in the dimension of $\mathscr{K}_r^\nu(\boldsymbol{w}^\star)$ for $r \leqslant 2\nu$.

We mention in passing that our construction is similar to the concept of graded linear spaces in Lie-algebra theory. The dimension of a grade-r subspace $\mathscr{N}_r^s(\mathbf{1})$, say, of Lie algebra with initial grades $w(X_l) \equiv 1$, $l = 1, 2, \ldots, s$, is given by the familiar *Witt formula*

$$\dim \mathscr{N}_r^s(\mathbf{1}) = \frac{1}{r} \sum_{i|r} \mu(i) s^{r/i}, \tag{20}$$

where μ is the *Möbius function,* while a basis of $\mathscr{N}_r^s(\mathbf{1})$ can be obtained by algorithms due to Hall and to Lyndon [**16, 18**]. The formula (20) and the Hall basis have been generalized to arbitrary sets of weights by Munthe-Kaas and Owren [**16**]. In particular, we note for future reference that

$$\dim \mathscr{N}_r^s(\boldsymbol{w}) = \frac{1}{r}\sum_{i|r}\mu(i)\left(\sum_{k=1}^{n}\lambda_k^{r/i}\right), \tag{21}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$, $n = \max w_k$, are the zeros of

$$p(z) = z^n - \sum_{i=1}^{s} z^{n-w_i}.$$

3.2. *A generating function*

Let $w_l = w(X_l)$, $l = 1, 2, \ldots, s$. In this subsection we construct explicitly a generating function of the sequence $\{d_r^s\}_{r\geqslant 1}$, where

$$d_r^s = \dim \mathscr{K}_r^\nu(\boldsymbol{w}).$$

To this end, however, we first require a technical result about the form of elements in the free hierarchical algebra $\mathfrak{g}$.

Definition 4. We say that $Y \in \mathfrak{g}$, where $\mathfrak{g}$ is a free hierarchical algebra generated by $X_1, X_2, \ldots, X_s$, is *primitive* if there exist $m \geqslant 1$ and $i_1, i_2, \ldots, i_m \in \{1, 2, \ldots, s\}$ such that $Y = [\![X_{i_1}, X_{i_2}, \ldots, X_{i_s}]\!]_m$. Moreover, we say that a basis $\mathscr{B}_r^s$ of $\mathscr{K}_r^s(\boldsymbol{w})$ is *primitive* if all its elements are primitive.

As an example of a non-primitive element, consider

$$[\![X_{i_1}, X_{i_2}, [\![X_{i_3}, X_{i_4}]\!]_2]\!]_3.$$

Lemma 3. *Every $Y \in \mathfrak{g}$ is a linear combination of primitive elements.*

Proof. The lemma is proved by induction on the grade. Clearly, $[\![X_1]\!]_1$ is a primitive element. Suppose thus that the lemma is true for all elements of grade $\leqslant n-1$, $n \geqslant 2$, and suppose that $Y \in \mathscr{K}_n^s(\boldsymbol{w})$ and is non-primitive. Since $\mathfrak{g}$ is the algebra generated by $X_1, \ldots, X_s$, there exist indices $k_1, \ldots k_m$, $m < n$, such that

$$Y = [\![[\![X_{i_{k_1,1}}, \ldots, X_{i_{k_1,k_1}}]\!]_{k_1}, [\![X_{i_{k_2,1}}, \ldots, X_{i_{k_2,k_2}}]\!]_{k_2}, \ldots, [\![X_{i_{k_m,1}}, \ldots, X_{i_{k_m,k_m}}]\!]_{k_m}]\!]_m,$$

where $[\![X_{i_{k_j,1}}, \ldots, X_{i_{k_j,k_j}}]\!]_{k_j}$ has grade $v_j < n$ for all $1 \leqslant j \leqslant m$ and $v_1 + \cdots + v_m = n$. Note that $[\![X_{i_{k_j,1}}, \ldots, X_{i_{k_j,k_j}}]\!]_{k_j}$ is a primitive elements for all $j \leqslant m$. The lemma follows by the multilinearity (17) and the hierarchy condition (18). □

Definition 5. Given $Y \in \mathfrak{g}$ we say that Y is of *length* m if it is primitive and $Y = [\![X_{i_1}, \ldots, X_{i_m}]\!]_m$. If Y is non-primitive its lenght equals the maximum of the lengths of primitive words that compose it.

As an example, the non-primitive word $[\![X_{i_1}, X_{i_2}, [\![X_{i_3}, X_{i_4}]\!]_2]\!]_3$ has length four, since it can be decomposed in

$$[\![X_{i_1}, X_{i_2}, X_{i_3}, X_{i_4}]\!]_4 - [\![X_{i_1}, X_{i_2}, X_{i_4}, X_{i_3}]\!]_4$$

a linear combination of primitive words both having lenght equal to four.

It is a consequence of Lemma 3 that there exists a primitive basis $\mathscr{B}_r^s$ of $\mathscr{K}_r^s(\boldsymbol{w})$. For every $m \geqslant 1$ we denote by $\nu_{r,m}^{[s]} \geqslant 0$ the number of m-nary terms in $\mathscr{B}_r^s$, i.e. of terms of the form $[\![X_{i_1}, X_{i_2}, \ldots, X_{i_m}]\!]_m$. Clearly,

$$d_r^s = \sum_{m=1}^{\infty} \nu_{r,m}^{[s]}.$$

Hence, defining the generating functions

$$\Theta_m(z) := \sum_{r=1}^{\infty} \nu_{r,m}^{[s]} z^r, \quad m \geqslant 1, \qquad W(z) = \sum_{r=1}^{\infty} d_r^s z^r,$$

we obtain

$$W(z) = \sum_{m=1}^{\infty} \Theta_m(z). \tag{22}$$

Our intention is to obtain each function Θ_m, $m \geqslant 1$, in a closed form. To this end we commence by observing that

$$\Theta_1(z) = q(z) := \sum_{i=1}^{s} z^{w_i}. \tag{23}$$

The reason is that Θ_1 contains precisely the contribution of all unary terms, hence of the generators themselves, and each X_i adds z^{w_i} to the sum.

We next obtain Θ_ms recursively, distinguishing between even and odd indices m.

Case 1 $m = 2M$:
Each term $[\![X_{i_1}, X_{i_2}, \ldots, X_{i_{2M}}]\!]_{2M}$ in a primitive basis contributes an expression of the form $z^{w_{i_1}+w_{i_2}+\cdots+w_{i_{2M}}}$ to Θ_{2M}. We need to single out the indices $i_1, i_2, \ldots, i_{2M} \in \{1, 2, \ldots, s\}$ which are allowed in the basis, i.e. which do not interfere with linear independence.

Choose $t \in \{0, 1, \ldots, M-1\}$ and assume that

$$i_1 = i_{2M}, \quad i_2 = i_{2M-1}, \quad \ldots, \quad i_t = i_{2M+1-t}, \quad i_{t+1} < i_{2M-t}. \tag{24}$$

In that case any choice of $1 \leqslant i_{t+2}, \ldots, i_{2M-t-1} \leqslant s$ leads to a linearly independent term. Therefore the contribution to Θ_{2M} is

$$\begin{aligned}
&\sum_{i_1=1}^{s} \cdots \sum_{i_t=1}^{s} \sum_{i_{t+1}=1}^{s-1} \sum_{i_{2M-t}=i_{t+1}+1}^{s} \sum_{i_{t+2}=1}^{s} \cdots \sum_{i_{2M-t-1}=1}^{s} z^{2(w_{i_1}+\cdots+w_{i_t})+w_{t+1}+\cdots+w_{2M-t}}\\
&= \left(\sum_{i=1}^{s} z^{2w_i}\right)^t \left(\sum_{i=1}^{s} z^{y_i}\right)^{2M-2t-2} \sum_{k=1}^{s-1} \sum_{l=k+1}^{s} z^{w_k+w_l}\\
&= [q(z^2)]^t [q(z)]^{2M-2t-2} \sum_{k=1}^{s-1} \sum_{l=k+1}^{s} z^{w_k+w_l}.
\end{aligned}$$

Moreover,

$$\sum_{k=1}^{s-1} \sum_{l=k+1}^{s} z^{w_k+w_l} = \sum_{k=2}^{s} \sum_{l=1}^{k-1} z^{w_k+w_l},$$

therefore

$$\sum_{k=1}^{s-1}\sum_{l=k+1}^{s} z^{w_k+w_l} = \tfrac{1}{2}\left(\sum_{k=1}^{s-1}\sum_{l=k+1}^{s} z^{w_k+w_l} + \sum_{k=2}^{s}\sum_{l=1}^{k-1} z^{w_k+w_l}\right)$$
$$= \tfrac{1}{2}\left(\sum_{k=1}^{s}\sum_{l=1}^{s} z^{w_k+w_l} - \sum_{k=1}^{s} z^{2w_k}\right) = \tfrac{1}{2}\{[q(z)]^2 - q(z^2)\}.$$

We thus deduce that the contributions of all the terms for this value of t is

$$\tfrac{1}{2}[q(z^2)]^t[q(z)]^{2M-2t-2}\{[q(z)]^2 - q(z^2)\}$$
$$= \tfrac{1}{2}\{[q(z^2)]^t[q(z)]^{2M-2t} - [q(z^2)]^{t+1}[q(z)]^{2M-2t-2}\}.$$

All possible candidates for elements in a primitive basis are obtained by allowing t in (24) to range across $\{0,1,\ldots,M-1\}$: note that we can always assume that $i_{t+1} < i_{2M-t}$, otherwise we use alternating symmetry (16) to obtain a linearly-dependent term. Therefore Θ_{2M} can be obtained by summing up in t, series telescope and

$$\begin{aligned}\Theta_{2M}(z) &= \tfrac{1}{2}\sum_{t=0}^{M-1}\{[q(z^2)]^t[q(z)]^{2M-2t} - [q(z^2)]^{t+1}[q(z)]^{2M-2t-2}\}\\ &= \tfrac{1}{2}\{[q(z)]^{2M} - [q(z^2)]^M\}.\end{aligned} \tag{25}$$

Case 2 $m = 2M+1$:
This case is very similar, except for one important detail. We assemble all linearly-independent terms in $[\![X_{i_1}, X_{i_2}, \ldots, X_{i_{2M+1}}]\!]_{2M+1}$. For every $t = 0,1,\ldots,M-1$ we take, similarly to (24),

$$i_1 = i_{2M+1}, \quad i_2 = i_{2M}, \quad \ldots, \quad i_t = i_{2M-t+2}, \quad i_{t+1} < i_{2M-t+1},$$

whence the contribution to Θ_{2M+1} is

$$\tfrac{1}{2}\{[q(z^2)]^t[q(z)]^{2M-2t+1} - [q(z^2)]^{t+1}[q(z)]^{2M-2t-1}\}$$

– the algebra is virtually identical to the previous case. However, if

$$i_1 = i_{2M+1}, \quad i_2 = i_{2M}, \quad \ldots, \quad i_M = i_{M+2}$$

then the middle index i_{M+1} can be allowed to range across all of $\{1,2,\ldots,s\}$, and this contributes

$$[q(z^2)]^M q(z)$$

to Θ_{2M+1}. Summing up in t leads at once to

$$\Theta_{2M+1}(z) = \tfrac{1}{2}\{[q(z)]^{2M+1} + q(z)[q(z^2)]^M\}. \tag{26}$$

Note that this is consistent with (23), therefore valid for all $M \geqslant 0$.

We can now substitute (25) and (26) in (22) and the outcome is the generating function

$$W(z) = \tfrac{1}{2}\sum_{m=1}^{\infty}[q(z)]^m + \tfrac{1}{2}q(z)\sum_{M=0}^{\infty}[q(z^2)]^M - \tfrac{1}{2}\sum_{M=1}^{\infty}[q(z^2)]^M$$
$$= \tfrac{1}{2}\frac{q(z)}{1-q(z)} + \tfrac{1}{2}\frac{q(z)}{1-q(z^2)} - \tfrac{1}{2}\frac{q(z^2)}{1-q(z^2)}$$

$$= \frac{q(z) - \frac{1}{2}q(z^2) - \frac{1}{2}[q(z)]^2}{[1 - q(z)][1 - q(z^2)]}. \tag{27}$$

3.3. *The dimension of graded subspaces*

We intend in this subsection to find explicitly $d_r^s = \dim \mathscr{K}_r^s$. The point of departure is the explicit formula (27) for the generating function. To expand $W(z)$ in Taylor series, we need to invert polynomials $1 - q(z)$ and $1 - q(z^2)$. This issue is explored in the next lemma.

Lemma 4. *Given any distinct $\theta_1, \theta_2, \ldots, \theta_n \in \mathbb{C} \setminus \{0\}$, it is true that*

$$\prod_{l=1}^{n}(1 - \theta_l z)^{-1} = \sum_{k=0}^{\infty}\left[\sum_{l=1}^{n}\frac{\theta_l^{k+n-1}}{\omega_n'(\theta_l)}\right]z^k, \qquad z \in \mathbb{C}, \tag{28}$$

where $\omega_n(z) = \prod_{k=1}^{n}(z - \theta_k)$.

Proof. The expansion (28) is a straightforward consequence of the residuum theorem. The poles of $f(z) = \prod_{l=1}^{n}(1 - \theta_l z)^{-1}$ being simple, it is true that

$$\prod_{l=1}^{n}(1 - \theta_l z)^{-1} = \sum_{l=1}^{n}\frac{f_l}{1 - \theta_l z},$$

where

$$f_l = \lim_{z \to \theta_l^{-1}}(1 - \theta_l z)f(z) = \frac{\theta_l^{n-1}}{\omega_n'(\theta_l)}.$$

The lemma follows by expanding each $1 - \theta_l z$ as geometric series. □

Letting $n = w_s = \max w_k$ be the degree of the polynomial q, we use (28) with $\varphi_n(z) = [1 - q(z)]^{-1}$. Therefore

$$\omega_n(z) = z^n[1 - q(z^{-1})] = z^n - \sum_{k=1}^{s} z^{n-w_k} = p(z),$$

the polynomial p having been already encountered in (21). We again denote its zeros by $\lambda_1, \lambda_2, \ldots, \lambda_n$, whereby, by (28),

$$\frac{1}{1 - q(z)} = \sum_{k=0}^{\infty}\left[\sum_{l=1}^{n}\frac{\lambda_l^{k+1}}{q'(\lambda_l^{-1})}\right]z^k.$$

The expansion of $[1 - q(z^2)]^{-1}$ can be similarly obtained, replacing z with z^2.

Theorem 5. *Suppose that all the zeros $\lambda_1, \lambda_2, \ldots, \lambda_n$ of p away from the origin are distinct. The dimensions of graded subspaces are*

$$d_{2R}^s = \frac{1}{2}\sum_{l=1}^{n}\frac{1}{q'(\lambda_l^{-1})}\{\lambda_l^{2R+1} - \lambda_l^{R+1} + \tfrac{1}{2}\lambda_l^{R+1}[q(\lambda_l^{-1/2}) + q(-\lambda_l^{-1/2})]\}, \tag{29}$$

$$d_{2R+1}^s = \frac{1}{2}\sum_{l=1}^{n}\frac{1}{q'(\lambda_l^{-1})}\{\lambda_l^{2R+2} + \tfrac{1}{2}\lambda_l^{R+3/2}[q(\lambda_l^{-1/2}) - q(-\lambda_l^{-1/2})]\}. \tag{30}$$

Proof. Bearing in mind that $q(z)=1-\prod_{l=1}^n(1-\lambda_l z)$, we replace (27) with

$$W(z)=\tfrac12\left[\frac{1}{1-q(z)}-\frac{1-q(z)}{1-q(z^2)}\right]=\tfrac12\left[\frac{1}{1-q(z)}-\frac{\prod_{l=1}^n(1-\lambda_l z)}{1-q(z^2)}\right]. \tag{31}$$

Let

$$\prod_{l=1}^n(1-\lambda_l z)=\sum_{j=0}^n u_j z^j. \tag{32}$$

Then

$$\begin{aligned}\frac{\prod_{l=1}^n(1-\lambda_l z)}{1-q(z^2)}&=\sum_{k=0}^\infty \varphi_{n,k}\sum_{j=0}^n u_j z^{2k+j}=\sum_{j=0}^\infty\left(\sum_{k=\lfloor (j-n+1)/2\rfloor}^{\lfloor j/2\rfloor}u_{j-2k}\varphi_{n,k}\right)z^j\\&=\sum_{j=0}^\infty\left[\sum_{l=1}^n\frac{1}{q'(\lambda_l^{-1})}\sum_{k=\lfloor (j-n+1)/2\rfloor}^{\lfloor j/2\rfloor}u_{j-2k}\lambda_l^{k+1}\right]z^j,\end{aligned}$$

where we set $u_j=0$ outside the range $0\leqslant j\leqslant n$.

We next consider the sum

$$g_{n,j}(\xi):=\sum_{k=\lfloor (j-n+1)/2\rfloor}^{\lfloor j/2\rfloor}u_{j-2k}\xi^{k+1},\qquad n,j\geqslant 0,\quad \xi\in\mathbb{C}.$$

By straightforward computation

$$\begin{aligned}g_{2N,2J}(\xi)&=\sum_{k=J-N}^{J}u_{2J-2k}\xi^{k+1}=\xi^{J+1}\sum_{k=0}^{N}u_{2k}\xi^{-k},\\ g_{2N,2J+1}(\xi)&=\sum_{k=J-N+1}^{J}u_{2J+1-2k}\xi^{k+1}=z^{J+1}\sum_{k=0}^{N-1}u_{2k+1}\xi^{-k},\\ g_{2N+1,2J}(\xi)&=\sum_{k=J-N}^{J}u_{2J-2k}\xi^{k+1}=\xi^{J+1}\sum_{k=0}^{N}u_{2k}\xi^{-k},\\ g_{2N+1,2J+1}(\xi)&=\sum_{k=J-N}^{J}u_{2J-2k+1}\xi^{k+1}=\xi^{J+1}\sum_{k=0}^{N}u_{2k+1}\xi^{-k}.\end{aligned}$$

We thus deduce from (32) that

$$\begin{aligned}g_{n,2J}(\xi)&=\tfrac12\xi^{J+1}\left[\prod_{i=1}^n\left(1-\frac{\lambda_i}{\xi^{1/2}}\right)+\prod_{i=1}^n\left(1+\frac{\lambda_i}{\xi^{1/2}}\right)\right],\\ g_{n,2J+1}(z)&=\tfrac12 z^{J+3/2}\left[\prod_{i=1}^n\left(1-\frac{\lambda_i}{\xi^{1/2}}\right)-\prod_{i=1}^n\left(1+\frac{\lambda_i}{\xi^{1/2}}\right)\right]\end{aligned}$$

and, letting $\xi=\lambda_l$ and $\prod_{j=1}^n(1-\lambda_j z)=1-q(z)$, the explicit formulae (29) and (30) follow from substitution in (31). □

3.4. *Examples of graded subspaces*

The simplest special case is when the grades of the generators are equal, $w_l\equiv 1$, $l=1,2,\ldots,s$. In that case $q(z)=sz$, therefore $p(z)=z^n-sz^{n-1}$, and we have $n=1$,

$\lambda_1 = s$. Substitution in (29) and (30) affirms that

$$d^s_{2R-1} = \tfrac{1}{2}(s^{2R-1} + s^R), \quad d^s_{2R} = \tfrac{1}{2}(s^{2R} - s^R), \qquad R \geqslant 1. \tag{33}$$

Table 1 displays the dimensions of subspaces of graded Lie and hierarchical algebras with $s = 2$ and $s = 3$. Note that (20) and (33) imply that

$$\dim \mathscr{N}^s_r(\mathbf{1}) \approx \frac{s^r}{r}, \quad \dim \mathscr{K}^s_r(\mathbf{1}) \approx \frac{s^r}{2}, \qquad r \gg 1, \tag{34}$$

consistently with Table 1.

Subspaces	r									
	1	2	3	4	5	6	7	8	9	10
$\dim \mathscr{N}^2_r$	2	1	2	3	6	9	18	30	56	99
$\dim \mathscr{K}^2_r$	2	1	6	6	20	28	72	120	272	496
$\dim \mathscr{N}^3_r$	3	3	8	18	48	116	312	810	2184	5880
$\dim \mathscr{K}^3_r$	3	3	18	36	135	351	1134	3240	9963	29403

Table 1: The dimensions of graded subspaces of Lie and hierarchical algebras: $w_l \equiv 1$.

In our second example we let $s = 2$ and consider $w_1 = 1$, $w_2 = n \geqslant 2$. Therefore $q(z) = z + z^n$, an nth-degree polynomial, and the zeros $\lambda_1, \lambda_2, \ldots, \lambda_n$ obey $\lambda_l^n - \lambda_l^{n-1} = 1$. It is easy to verify that $q'(\lambda_l^{-1}) = 1 - n + n\lambda_l$, all the zeros are simple and

$$\tfrac{1}{2}[q(z) + q(-z)] = \begin{cases} z^n, & n \text{ even}, \\ 0, & n \text{ odd}; \end{cases} \qquad \tfrac{1}{2}[q(z) - q(-z)] = \begin{cases} z, & n \text{ even}, \\ z + z^n, & n \text{ odd}. \end{cases}$$

We deduce that for even n,

$$\dim \mathscr{K}^2_{2R} = \tfrac{1}{2}\sum_{l=1}^{n} \frac{1}{1 - n + n\lambda_l}(\lambda_l^{2R+1} - \lambda_l^{R+1} + \lambda_l^{R+1-n/2})$$

$$\dim \mathscr{K}^2_{2R+1} = \tfrac{1}{2}\sum_{l=1}^{n} \frac{1}{1 - n + n\lambda_l}(\lambda_l^{2R+2} + \lambda_l^{R+1}),$$

while odd n yields

$$\dim \mathscr{K}^2_{2R} = \tfrac{1}{2}\sum_{l=1}^{n} \frac{1}{1 - n + n\lambda_l}(\lambda_l^{2R+1} - \lambda_l^{R+1}),$$

$$\dim \mathscr{K}^2_{2R+1} = \tfrac{1}{2}\sum_{l=1}^{n} \frac{1}{1 - n + n\lambda_l}[\lambda_l^{2R+2} + \lambda_l^{R+1} + \lambda_l^{R+1-(n-1)/2}].$$

It follows from the above formula and (21) that

$$\dim \mathscr{N}^2_r([1,n]) \approx \frac{\lambda^r_{\max}}{r}, \quad \dim \mathscr{K}^2_r([1,n]) \approx \frac{\lambda^{r+1}_{\max}}{1 - n + n\lambda_{\max}}, \qquad r \gg 1, \tag{35}$$

where $\lambda_{\max} = \max_{l=1,2,\ldots,n} |\lambda_l|$.

Our third and last example is probably the most important within the context of geometric integration (cf. the discussion in Subsection 3.1): $w_l = l$, $l = 1, 2, \ldots, s$. Therefore

$$q(z) = \sum_{l=1}^{s} z^l = \frac{z - z^{s+1}}{1 - z}$$

and $n = s$. Moreover, if $q(\lambda_l^{-1}) = 1$ then $\lambda_l^{-s-1} = 2\lambda_l - 1$, therefore

$$q'(\lambda_l^{-1}) = \frac{1 - (s+1)\lambda_l^{-s} + s\lambda_l^{-s-1}}{(1 - \lambda_l^{-1})^2} = \lambda_l \frac{2s - (1+s)\lambda_l}{1 - \lambda_l}.$$

Thus, $q'(\lambda_l^{-1}) = 0$ (hence, a multiple root) may occur only for $\lambda_l = [(s+1)/2]^{1/s} > 0$. This is impossible according to Lemma 6 in the sequel and we deduce that all the λ_ls are distinct. Letting s be even, we have

$$\tfrac{1}{2}[q(z) + q(-z)] = \frac{z^2 - z^{s+2}}{1 - z^2}, \qquad \tfrac{1}{2}[q(z) - q(-z)] = \frac{z - z^{s+1}}{1 - z^2},$$

therefore, letting again $\boldsymbol{w}^\star = [1, 2, \ldots, s]$,

$$\dim \mathscr{K}_{2R}^s(\boldsymbol{w}^\star) = \tfrac{1}{2} \sum_{l=1}^{s} \frac{1 - \lambda_l}{2s - (1+s)\lambda_l} \left(\lambda_l^{2R} - \lambda_l^R + \lambda_l^R \frac{\lambda_l^{-s/2} - 1}{1 - \lambda_l} \right)$$

$$\dim \mathscr{K}_{2R+1}^s(\boldsymbol{w}^\star) = \tfrac{1}{2} \sum_{l=1}^{s} \frac{1 - \lambda_l}{2s - (1+s)\lambda_l} \left(\lambda_l^{2R+1} + \lambda_l^{R+1} \frac{\lambda_l^{-s/2} - 1}{1 - \lambda_l} \right).$$

Likewise, for odd s,

$$\tfrac{1}{2}[q(z) + q(-z)] = \frac{z^2 - z^{s+1}}{1 - z^2}, \qquad \tfrac{1}{2}[q(z) - q(-z)] = \frac{z - z^{s+2}}{1 - z^2}$$

and we obtain

$$\dim \mathscr{K}_{2R}^s(\boldsymbol{w}^\star) = \tfrac{1}{2} \sum_{l=1}^{s} \frac{1 - \lambda_l}{2s - (1+s)\lambda_l} \left(\lambda_l^{2R} - \lambda_l^R + \lambda_l^R \frac{\lambda_l^{-(s-1)/2} - 1}{1 - \lambda_l} \right),$$

$$\dim \mathscr{K}_{2R+1}^s(\boldsymbol{w}^\star) = \tfrac{1}{2} \sum_{l=1}^{s} \frac{1 - \lambda_l}{2s - (1+s)\lambda_l} \left(\lambda_l^{2R} + \lambda_l^{R} \frac{\lambda_l^{-(s-3)/2} - 1}{1 - \lambda_l} \right).$$

We have already observed in (34) and (35) that the asymptotic behaviour of the sequences $\{\mathscr{N}_r^s\}_{r\geqslant 1}$ and $\{\mathscr{K}_r^s\}_{r\geqslant 1}$ for large r is predictable and that the former increases somewhat slower. In the present case, letting $\lambda_{\max}$ be the largest zero, we have

$$\dim \mathscr{N}_r^s(\boldsymbol{w}^\star) \approx \frac{1}{r}\lambda_{\max}^r, \quad \dim \mathscr{K}_r^s(\boldsymbol{w}^\star) \approx \tfrac{1}{2} \frac{\lambda_{\max} - 1}{(1+s)\lambda_{\max} - 2s} \lambda_{\max}^r, \qquad r \gg 1.$$

To estimate $\lambda_{\max}$ we note that $\lim_{s\to\infty} q(z) = z(1-z)^{-1}$. Therefore all the λ_ls except for one tend to zero, while, without loss of generality, $\lambda_{\max} = \lambda_1 \to 2$. More precisely, $\lambda_1^{s+1} - 2\lambda_1^s + 1 = 0$ implies that

$$\lambda_1 = 2 - \frac{1}{s2^s}[1 + o(1)], \qquad s \gg 1.$$

Therefore, approximately

$$\dim \mathscr{N}_r^s(\boldsymbol{w}^\star) \approx \frac{2^r}{r}, \quad \dim \mathscr{K}_r^s(\boldsymbol{w}^\star) \approx 2^{r-2}, \qquad r, s \gg 1. \tag{36}$$

3.5. *Asymptotic behaviour*

We have already observed in (34), (35) and (36) that the dimension of a graded subspace of a Lie algebra is consistently smaller for large r than its hierarchical-algebra counterpart.

Lemma 6. *Let $s \geqslant 2$. The equation*

$$q(z) = 1, \qquad \textit{where} \qquad q(z) = \sum_{l=1}^{s} z^{w_l}, \tag{37}$$

possesses a unique positive root $\eta \in (0,1)$. This root is simple and all other roots of (37) reside in the open disc $|z| > \eta$.

Proof. Since $q(0) = 0$ and $q(1) = s \geqslant 2$, it follows that $q(z) - 1$ changes sign in $(0,1)$. Moreover, $q'(z) > 0$ for $z \geqslant 0$, therefore this root, which we denote by η, is simple and it is the unique positive root of (37).

Let $n = \max w_l$. If $n = 1$ then $q(z) = sz$, $\eta = 1/s$ and there are no other roots. The lemma is certainly true in this case, hence we may assume in the remainder of the proof that $n \geqslant 2$. Let $\xi = |\xi|\mathrm{e}^{\mathrm{i}\theta}$ be another root of (37) and note that we have already proved that $\theta \in (0, 2\pi)$. Since $q(z)$ is a linear combination of at least two different powers of z, it is true that

$$|q(\xi)| = \left| \sum_{l=1}^{s} |\xi|^{w_l} \mathrm{e}^{\mathrm{i}\theta_l} \right| < \sum_{l=1}^{s} |\xi|^{w_l} = q(|\xi|).$$

Since ξ is a root of (37), we deduce that $q(|\xi|) > 1$. Because of the monotonicity of q in $[0,\infty)$, we conclude that $|\xi| > \eta$ and the lemma follows. $\square$

Theorem 7. *Suppose that the roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct. For $r \gg 1$ it is true that*

$$\frac{\dim \mathscr{K}_r^s(\boldsymbol{w})}{\dim \mathscr{N}_r^s(\boldsymbol{w})} = \frac{r}{2\eta q(\eta)}[1 + o(1)], \tag{38}$$

where η is the positive root of (37).

Proof. Since $1/\eta$ dominates the other λ_ls, we deduce from (21), (29) and (30) that, for $r \gg 1$,

$$\dim \mathscr{N}_r^s(\boldsymbol{w}) \approx \frac{\mu(1)\eta^{-r}}{r} = \frac{1}{r\eta^r}$$

$$\dim \mathscr{K}_r^s(\boldsymbol{w}) \approx \frac{\eta^{-r-1}}{2q'(\eta)} = \frac{1}{2q'(\eta)\eta^{r+1}}.$$

The estimate (38) follows easily. $\square$

3.6. *Multiple roots*

Much of the analysis of Subsections 3.3 and 3.5 is based on the assumption that all the roots of the equation $q(z) = 1$ are simple. It is easy to construct as example of a graded hierarchical algebra with multiple roots by taking $s = 9$ and

$$w(X_1) = 1, \quad w(X_2) = \cdots = w(X_6) = 2, \quad w(X_7) = w(X_8) = w(X_9) = 3.$$

Therefore $q(z) = z + 5z^2 + 3z^3$ and

$$q(z) - 1 = ((z+1)^2(3z-1),$$

with a double root at $z = -1$. The generating function (27) remains, of course, valid and we have

$$W(z) = \frac{z + 4z^2 - 2z^3 - 18z^4 - 15z^5 - 6z^6}{(1+z)^2(1-3z)(1+z)^2(1-3z^2)}$$

$$= \tfrac{1}{8}\frac{1}{(1+z)^2} + \tfrac{3}{32}\frac{1}{1+z} + \tfrac{9}{32}\frac{1}{1-3z} + \tfrac{3}{16}\frac{1+3z}{1-3z^2} - \tfrac{1}{16}\frac{5-3z}{1+z^2}$$
$$- \tfrac{1}{8}\frac{3+2z-3z^2}{(1+z^2)^2}$$
$$= z + 5z^2 + 9z^3 + 23z^4 + 72z^5 + 214z^6 + 630z^7 + 1858z^8 + 5579z^9 + \cdots.$$

This particular example is of little independent interest except that it indicates how to generalise our analysis to the case of multiple roots. Explicit expressions (29) and (30) are no longer valid but they can be generalised with moderate effort. Moreover, note that Lemma 6 remains valid and it is easy to confirm that the asymptotic estimate (38) is true also for multiple roots.

3.7. *A basis of* $\mathscr{K}_r^s(\boldsymbol{w})$

Our method for the formation of a basis of the graded linear subspace $\mathscr{K}_r^s(\boldsymbol{w})$ is nothing but an algorithmic rendering of the argument in Subsection 3.2 which has led to the generating function (27). In essence, we choose a basis of primitive elements, removing all elements that are linearly dependent by virtue of the alternating symmetry (16). Note that this is, in a manner of speech, the exact opposite of the standard procedure for the formation of Hall or Lyndon bases of graded subspaces in Lie algebras. The basic building blocks in the latter are iterated commutators of generators, terms of the form

$$[X_{i_1}, [X_{i_2}, [\ldots, [X_{i_{m-1}}, X_{i_m}]\cdots]]],$$

where $i_1, i_2, \ldots, i_m \in \{1, 2, \ldots, s\}$ and linear dependencies occur by virtue of skew-symmetry of the commutator and of the Jacobi identity [**16**].

Henceforth, we denote our basis of $\mathscr{K}_r^s(\boldsymbol{w})$ by $\mathscr{B}_r^s$.

Algorithm 1. *Looping through* r*:*

Step 1: We initially let $\mathscr{B}_{r,1}^s$ consist of all the generators X_l such that $w_l = r$. If there are no such generators, we let $\mathscr{B}_{r,1}^s = \emptyset$.

Step 2: For every even $m = 2M \leqslant r$ we add to $\mathscr{B}_{r,2M}^s$ all the elements

$$[\![X_{i_1}, X_{i_2}, \ldots, X_{i_{2M}}]\!]_{2M}$$

with the $(2M)$-tuples $\boldsymbol{i}$ formed according to the following rule: For every $t = 0, 1, \ldots, M-1$ we take $\boldsymbol{i}$ such that

$$i_j = i_{2M+1-j} = 1, 2, \ldots, s, \qquad j = 1, 2, \ldots, 2M-t-1, \quad j \neq t+1,$$
$$1 \leqslant i_{t+1} < i_{2M-t} \leqslant s,$$

provided that $\sum_{j=1}^{2M} w_{i_j} = r$.

Step 3: For every odd $3 \leqslant m = 2M+1 \leqslant r$ we add to $\mathscr{B}_{r,2M+1}^s$ each element

$$[\![X_{i_1}, X_{i_2}, \ldots, X_{i_{2M+1}}]\!]_{2M+1}$$

for which the $(2M+1)$-tuple $\boldsymbol{i}$ is either of the form

$$i_j = i_{2M+2-j} = 1, 2, \ldots, s, \qquad j = 1, 2, \ldots, 2M-t, \quad j \neq t+1,$$
$$1 \leqslant i_{t+1} < i_{2M+1-t} \leqslant s$$

for some $t \in \{0, 1, \ldots, M-1\}$ or

$$i_j = i_{2M+2-j} = 1, 2, \ldots, s, \quad j = 1, 2, \ldots, M, \qquad i_{M+1} = 1, 2, \ldots, s$$

and, in either case, $\sum_{j=1}^{2M+1} w_{i_j} = r$.

The algorithm above is wasteful, since it forms time and again the same m-tuples and discards the great majority of them, since they do not lead to elements in $\mathscr{K}_r^s(\boldsymbol{w})$. An alternative algorithm, which avoids this pitfall and forms each m-tuple just once, is ideally suited to generate the bases $\mathscr{B}_r^s$ for all $r = 1, 2, \ldots, r^\star$, where $r^\star \geqslant 1$ is given.

Algorithm 2. *Looping through m:*

Step 0: Let $\mathscr{B}_r^s = \emptyset$, $r = 1, 2, \ldots, r^\star$.

Step 1: For every $l = 1, 2, \ldots, s$, provided that $w_l \leqslant r^\star$, add X_l to $\mathscr{B}_{w_l}^s$.

Step 2: For every even $m = 2M \leqslant r^\star$,

$$\begin{aligned} &t = 0, 1, \ldots, M-1, \\ &i_1, i_2, \ldots, i_t, i_{t+2}, \ldots, i_{2M-t-1} = 1, 2, \ldots, s, \\ &1 \leqslant i_{t+1} < i_{2M-t} \leqslant s \end{aligned}$$

evaluate

$$\rho := 2\sum_{j=1}^{t} w_{i_j} + \sum_{j=t+1}^{2M-t} w_{i_j}.$$

If $\rho \leqslant r^\star$, let $i_{2M+1-j} = i_j$, $j = 1, 2, \ldots, t$, and place $[\![X_{i_1}, X_{i_2}, \ldots, X_{i_{2M}}]\!]_{2M}$ in $\mathscr{B}_\rho^s$.

Step 3: For every $m = 2M+1 \leqslant r^\star$, $M \geqslant 1$,

$$\begin{aligned} &t = 0, 1, \ldots, M-1, \\ &i_1, i_2, \ldots, i_t, i_{t+2} \ldots, i_{2M-t} = 1, 2, \ldots, s, \\ &1 \leqslant i_{t+1} < i_{2M+1-t} \leqslant s \end{aligned}$$

evaluate

$$\rho := 2\sum_{j=1}^{t} w_{i_j} + \sum_{j=t+1}^{2m+1-t} w_{i_j}.$$

Provided that $\rho \leqslant r^\star$, let $i_{2M+2-j} = i_j$, $j = 1, 2, \ldots, t$, and place the term $[\![X_{i_1}, X_{i_2}, \ldots, X_{i_{2M+1}}]\!]_{2M+1}$ in $\mathscr{B}_\rho^s$.
Likewise, for every

$$i_1, i_2, \ldots, i_{M+1} = 1, 2, \ldots, s,$$

if

$$\rho := 2\sum_{j=1}^{M} w_{i_j} + w_{i_{M+1}} < r^\star$$

then let $i_{2M+2-j} = i_j$, $j = 1, 2, \ldots, M$, and place $[\![X_{i_1}, X_{i_2}, \ldots, X_{i_{2M+1}}]\!]_{2M+1}$ in the set $\mathscr{B}_\rho^s$.

The output of Algorithm 2 consists of the bases $\mathscr{B}_r^s$ for $r = 1, 2, \ldots, r^\star$. We illustrate it by the example $s = 3$ and $\boldsymbol{w} = [1, 2, 3]$. letting $r^\star = 5$, we commence by setting $\mathscr{B}_r^3 = \emptyset$, $r = 1, 2, \ldots, 5$. Next,

$m = 1$**:** $X_1 \to \mathscr{B}_1^3$, $X_2 \to \mathscr{B}_2^3$, $X_3 \to \mathscr{B}_3^3$.

$m=2$**:** The only possibility is $t=0$, and this corresponds to terms $[\![X_{i_1},X_{i_2}]\!]_2$. Since $1\leqslant i_1<i_2\leqslant 3$, we obtain

$$[\![X_1,X_2]\!]_2\to\mathscr{B}_3^3,\qquad [\![X_1,X_3]\!]_2\to\mathscr{B}_4^3,\qquad [\![X_2,X_3]\!]_2\to\mathscr{B}_5^3.$$

$m=3$**:** Again, the only possible value is $t=0$, corresponding to $[\![X_{i_1},X_{i_2},X_{i_3}]\!]_3$ with $i_2=1,2,3$ and $1\leqslant i_1<i_3\leqslant 3$. Retaining just elements of grade $\leqslant 5$, we have

$$[\![X_1,X_1,X_2]\!]_3\to\mathscr{B}_4^3,\qquad [\![X_1,X_2,X_2]\!]_3\to\mathscr{B}_5^3,\qquad [\![X_1,X_1,X_3]\!]_3\to\mathscr{B}_5^3.$$

The remaining case is $[\![X_{i_1},X_{i_2},X_{i_1}]\!]_3$, whence $i_1,i_2=1,2,3$. We have

$$[\![X_1,X_1,X_1]\!]_3\to\mathscr{B}_3^3,\qquad [\![X_2,X_1,X_2]\!]_3\to\mathscr{B}_5^3,\qquad [\![X_1,X_2,X_1]\!]_3\to\mathscr{B}_4^3,$$
$$[\![X_1,X_3,X_1]\!]_3\to\mathscr{B}_5^3.$$

$m=4$**:** We need to consider $t=0$ and $t=1$. In the first case the terms have the form $[\![X_{i_1},X_{i_2},X_{i_3},X_{i_4}]\!]_4$ with $i_2,i_3=1,2,3$ and $1\leqslant i_1<i_4\leqslant 3$. There exists just one such term of grade $\leqslant 5$, namely

$$[\![X_1,X_1,X_1,X_2]\!]_4\to\mathscr{B}_5^3.$$

For $t=1$ our terms have the form $[\![X_{i_1},X_{i_2},X_{i_3},X_{i_1}]\!]_4$ with $i_1=1,2,3$ and $1\leqslant i_2<i_3\leqslant 3$. Again, just one term survives,

$$[\![X_1,X_1,X_2,X_1]\!]_4\to\mathscr{B}_5^3.$$

$m=5$**:** For $t=0$ we have terms of the form $[\![X_{i_1},X_{i_2},X_{i_3},X_{i_4},X_{i_5}]\!]_5$ for $i_2,i_3,i_4=1,2,3$ and $1\leqslant i_1<i_5\leqslant 3$. No term of this form has grade $\leqslant 5$. Similar situation pertains to the case $t=1$, whereby the terms are $[\![X_{i_1},X_{i_2},X_{i_3},X_{i_4},X_{i_1}]\!]_5$ with $i_1,i_3=1,2,3$ and $1\leqslant i_2<i_4\leqslant 3$. The remaining case results in the term $[\![X_{i_1},X_{i_2},X_{i_3},X_{i_2},X_{i_1}]\!]_5$ with $i_1,i_2,i_3=1,2,3$. One such term is of grade $\leqslant 5$, namely

$$[\![X_1,X_1,X_1,X_1,X_1]\!]_5\to\mathscr{B}_5^3.$$

The outcome are the bases

$\mathscr{B}_1^3=\{X_1\}$;
$\mathscr{B}_2^3=\{X_2\}$;
$\mathscr{B}_3^3=\{X_3,[\![X_1,X_2]\!]_2,[\![X_1,X_1,X_1]\!]_3\}$;
$\mathscr{B}_4^3=\{[\![X_1,X_3]\!]_2,[\![X_1,X_1,X_2]\!]_3,[\![X_1,X_2,X_1]\!]_3\}$;
$\mathscr{B}_5^3=\{[\![X_2,X_3]\!]_2,[\![X_1,X_2,X_2]\!]_3,[\![X_1,X_1,X_3]\!]_3,[\![X_2,X_1,X_2]\!]_3,$
$\quad[\![X_1,X_3,X_1]\!]_3,[\![X_1,X_1,X_1,X_2]\!]_4,[\![X_1,X_1,X_2,X_1]\!]_4,[\![X_1,X_1,X_1,X_1,X_1]\!]_5\}$.

4. *Applications to Cayley-transform methods*

4.1. *The Cayley–BCH formula*

The purpose of this subsection is to present an application of the formalism of hierarchical algebras to the derivation of the Cayley–BCH formula, a relation which in a Cayley-transform setting corresponds to the famous Baker–Hausdorff–Campbell formula.

Let $X, Y \in \mathfrak{g}$, a quadratic algebra of $\mathfrak{gl}_n(\mathbb{R})$. Throughout our analysis, we identify $\mathfrak{g}$ with a free hierarchical algebra. For a given sufficiently small $h > 0$ we wish to derive a function $Z(h)$ so that the equation

$$\operatorname{cay}[Z(h)] = \operatorname{cay}(hX)\operatorname{cay}(hY) \tag{39}$$

holds. We refer to (39) as the *Cayley–BCH formula*.

Introducing the inverse function cay^{-1} of the Cayley transform (7),

$$\operatorname{cay}^{-1} W = 2(W - I)(W + I)^{-1},$$

we deduce that the function

$$Z(h) = \sum_{k=1}^{\infty} Z_k h^k$$

complies with the equation

$$Z(h) = 2\Big[\operatorname{cay}(hX)\operatorname{cay}(hY) - I\Big] \times \Big[\operatorname{cay}(hX)\operatorname{cay}(hY) + I\Big]^{-1}.$$

Consequently,

$$\begin{aligned}
Z(h) &= 2\Big[(I - \tfrac{1}{2}hX)^{-1}(I + \tfrac{1}{2}hX)(I - \tfrac{1}{2}hY)^{-1}(I + \tfrac{1}{2}hY) + I\Big]^{-1} \\
&\qquad \times \Big[(I - \tfrac{1}{2}hX)^{-1}(I + \tfrac{1}{2}hX)(I - \tfrac{1}{2}hY)^{-1}(I + \tfrac{1}{2}hY) - I\Big] \\
&= \Big[(I - \tfrac{1}{2}hX)^{-1}(I + \tfrac{1}{4}h^2 XY)(I - \tfrac{1}{2}hY)^{-1}\Big]^{-1} \\
&\qquad \times \Big[(I - \tfrac{1}{2}hX)^{-1}(X + Y)(I - \tfrac{1}{2}hY)^{-1}\Big] \\
&= h(I - \tfrac{1}{2}hY)(I + \tfrac{1}{4}h^2 XY)^{-1}(X + Y)(I - \tfrac{1}{2}hY)^{-1}.
\end{aligned}$$

Multiplying both sides on the left by $(I - \frac{1}{2}hY)$ and writing $(I + \frac{1}{4}h^2 XY)^{-1}$ as a geometric series, we infer that

$$Z(h)(I - \tfrac{1}{2}hY) = (I - \tfrac{1}{2}hY)\left(\sum_{k=0}^{\infty} \frac{(-1)^k}{2^{2k}}(XY)^k h^{2k+1}\right)(X + Y),$$

hence the recursions

$$\left.\begin{aligned}
Z_{2k+1} &= \tfrac{1}{2} Z_{2k} Y + \frac{(-1)^k}{2^{2k}}(XY)^k (X + Y), \\
Z_{2k+2} &= \tfrac{1}{2} Z_{2k+1} Y - \frac{(-1)^k}{2^{2k+1}} Y (XY)^k (X + Y).
\end{aligned}\right\} \quad k = 0, 1, 2, \ldots, \tag{40}$$

(with starting condition $Z_0 = O$), that determine the Z_ks.

Theorem 8. *With the same notation as above, it is true that*

$$\begin{aligned}
Z_{2k+1} &= \frac{(-1)^k}{2^{2k+1}}(\llbracket X, Y, \ldots, X, Y, X \rrbracket_{2k+1} + \llbracket Y, X, \ldots Y, X, Y \rrbracket_{2k+1}), \\
Z_{2k+2} &= \frac{(-1)^k}{2^{2k+1}}\llbracket X, Y, \ldots, X, Y \rrbracket_{2k+2},
\end{aligned}$$

for $k = 0, 1, 2, \ldots$.

Proof. By induction on k. For $k = 0$ we obtain

$$Z_1 = \frac{1}{2}([\![X]\!]_1 + [\![Y]\!]_1) = X + Y,$$
$$Z_2 = \frac{1}{2}[\![X, Y]\!]_2 = XY - YX,$$

clearly in agreement with the recursions (40). Next, let us assume that the assertion is true for all integers up to k. We deduce from the recurrence relations (40) that

$$\begin{aligned} Z_{2k+1} &= \tfrac{1}{2} Z_{2k} Y + \frac{(-1)^k}{2^{2k}} (XY)^k (X + Y) \\ &= \frac{(-1)^{k-1}}{2^{2k}} ([\![X, Y, \ldots, X, Y]\!]_{2k} Y - (XY)^k (X + Y)) \\ &= \frac{(-1)^k}{2^{2k}} (XY \cdots YX + YX \cdots XY) \\ &= \frac{(-1)^k}{2^{2k+1}} ([\![X, Y, \ldots, Y, X]\!]_{2k+1} + [\![Y, X, \ldots, X, Y]\!]_{2k+1}). \end{aligned}$$

By a similar token,

$$\begin{aligned} Z_{2k+2} &= \tfrac{1}{2} Z_{2k+1} Y - \frac{(-1)^k}{2^{2k+1}} Y (XY)^k (X + Y) \\ &= \frac{(-1)^k}{2^{2k+2}} ([\![X, Y, \ldots, Y, X]\!]_{2k+1} + [\![Y, X, \ldots, X, Y]\!]_{2k+1}) \\ &\quad - \frac{(-1)^k}{2^{2k+1}} Y (XY)^k (X + Y) \\ &= \frac{(-1)^k}{2^{2k+1}} (XYX \cdots YXY + YXY \cdots XYY \\ &\quad - YXY \cdots XYX - YXY \cdots XYY) \\ &= \frac{(-1)^k}{2^{2k+1}} [\![X, Y, X, \ldots, Y, X]\!]_{2k+2}, \end{aligned}$$

hence the assertion is true also for $k + 1$. The proof is complete. ☐

In passing, it is worthwhile to mention that, in variance from the case of the exponential mapping, whereas

$$\exp(hX) \exp(hX) = \exp(2hX),$$

one has

$$\mathrm{cay}(hX)\mathrm{cay}(hX) = \mathrm{cay}(Z(h)),$$

for the Cayley map, where

$$Z(h) = \sum_{k=0}^{\infty} Z_{2k+1} h^{2k+1} = \frac{4}{\mathrm{i}} \sum_{k=0}^{\infty} \left(\frac{\mathrm{i}hX}{2}\right)^{2k+1} = 2hX \left(I + \frac{h^2 X^2}{4}\right)^{-1}.$$

4.2. *Adjoint operators in a Cayley-transform setting*

As above, we consider two generators, say X and Y in $\mathfrak{g}$. We wish to derive a formula for the hierarchical-algebra element

$$\mathrm{cay}(hX) Y \,\mathrm{cay}(-hX)$$

in terms of primitive elements of a hierarchical algebra. To this end, we observe that

$$\mathrm{cay}(V) = I + 2\sum_{k=1}^{\infty} \left(\frac{h}{2}\right)^k V^k, \qquad V \in \mathfrak{g},$$

hence we obtain by direct computation

$$\begin{aligned}\mathrm{cay}(hX)Y\,\mathrm{cay}(-hX) = Y &+ 2\sum_{k=1}^{\infty} \left(\frac{h}{2}\right)^k [X^kY + (-1)^k YX^k] \\ &+ 4\sum_{k=2}^{\infty} \left(\frac{h}{2}\right)^k \sum_{j=1}^{k-1} (-1)^{k-j} X^j Y X^{k-j},\end{aligned}$$

which, using hierarchical-algebra primitive elements, translates to

$$\begin{aligned}\mathrm{cay}(hX)Y\,\mathrm{cay}(-hX) = \frac{1}{2}[\![Y]\!]_1 &+ 2\sum_{k=1}^{\infty} \left(\frac{h}{2}\right)^k [\![X,\ldots,X,Y]\!]_{k+1} \\ &+ 4\sum_{k=2}^{\infty} \left(\frac{h}{2}\right)^k \sum_{j=1}^{\lfloor \frac{k}{2} \rfloor} (-1)^{k-j} \alpha_j [\![\underbrace{X,\ldots,X}_{j}, Y, \underbrace{X,\ldots,X}_{k-j}]\!]_{k+1},\end{aligned}$$

where $\alpha_1 = \cdots = \alpha_{\lfloor (k-1)/2 \rfloor} = 1$ and $\alpha_{k/2} = 1/2$ when k is an even index. Hence, for each k, the terms are precisely the primitive elements of length k with a single occurrence of Y.

Assume that $w(X) = w_X$ and $w(Y) = w_Y$. Then, the grade of a primitive word of length k and $k-1$ occurrences of X is $(k-1)w_X + w_Y$, therefore, identifying grades with order of accuracy, in order to truncate the above formula to an order K we require that the above summation is carried up to the least index $\hat{k}$ such that

$$\frac{K + w_X - w_Y}{w_X} \leqslant \hat{k}$$

holds.

We have mentioned above that the adjoint operator in a Cayley setting is a sum of primitive elements of the hierarchical algebra, with a single occurrence of Y. In greater generality, denote by $\gamma_{R;l}$ the number of primitive words of length R and l occurrences of the letter X. Clearly,

$$\gamma_{R;l} = \gamma_{R;R-l}, \qquad l = 0, 1, \ldots, R,$$

and, furthermore,

$$d_R^2 = \sum_{l=0}^{R} \gamma_{R;l},$$

where d_R^2 is the dimension of the space $\mathscr{K}_R^2(\mathbf{1})$.

Lemma 9. *With the same notation as above, it is true that*

$$\gamma_{2R+1;2l} = \gamma_{2R;2l-1} + \gamma_{2R;2l} + \binom{R}{l}, \qquad l = 0, \ldots, R, \tag{41}$$

$$\gamma_{2R+1;2l+1} = \gamma_{2R;2l} + \gamma_{2R;2l+1} + \binom{R}{l} \qquad l = 0, \ldots, R, \tag{42}$$

$$\gamma_{2R+2;2l} = \gamma_{2R+1;2l-1} + \gamma_{2R+1;2l} - \binom{R+1}{l}, \qquad l = 0, \dots, R+1, \tag{43}$$

$$\gamma_{2R+2;2l+1} = \gamma_{2R+1;2l} + \gamma_{2R+1;2l+1} - \binom{R}{l}, \qquad l = 0, \dots, R. \tag{44}$$

Proof. By induction on the length of the primitive word. Let us consider the index $R = 0$ and words of length one. The only possible primitive words are $[\![X]\!]_1$ and $[\![Y]\!]_1$, hence $\gamma_{1;0} = 1$ and $\gamma_{1;1} = 1$, which is in accordance with (41) and (42) respectively, given that $\gamma_{0;l} = 0$. For words of length 2, we observe that $\gamma_{2;0} = 0 = \gamma_{2;2}$, since there are no $[\![Y,Y]\!]_2$ and $[\![X,X]\!]_2$ words. Instead, $\gamma_{2;1} = 1$, corresponding to $[\![X,Y]\!]_2$, and (43) and (44) are obeyed.

Next, let us assume that the result is true for $k = 1, 2, \dots, 2R$. If $[\![X, Y \dots, Y, X]\!]_{2R}$ is a primitive word of length $2R$ with L occurrences of X, then $[\![X, Y, \dots, Y, X, Y]\!]_{2R+1}$, obtained by appending an Y, is a primitive word of length $2R+1$ with L occurrences of X. Similarly, we can take a primitive word of length $2R$ and $L-1$ occurrences of X and obtain one of length $2R+1$ and L Xs appending an X to it. However, by doing so, we do not obtain all the primitive words, because there are no symmetric words of length $2R$. All the missing words can be obtained choosing R slots (for instance the first R) and letting $\lfloor L/2 \rfloor$ letters X in them. Hence

$$\gamma_{2R+1;L} = \gamma_{2R;L-1} + \gamma_{2R,L} + \binom{R}{\lfloor L/2 \rfloor},$$

which gives (41) and (42). Similar procedure applies for (43) and (44), except that in this case we have to remove symmetric words which do not appear in the set of primitive elements. □

In other words, the number of words with a given number of occurrences of X and Y obeys a generalized version of the binomial coefficient relation. Below is a tabulation of the $\gamma_{R;l}$s.

R	d_R^2															
0	0								0							
1	2							1		1						
2	1						0		1		0					
3	6					1		2		2		1				
4	6				0		2		2		2		0			
5	20			1		3		6		6		3		1		
6	28		0		3		6		10		6		3		0	
7	72	1		4		12		19		19		12		4		1
⋮	⋮															

Each entry can be obtained as the sum of the two entries directly above, plus the corresponding entry of the *alternating Pascal triangle*.

R															
0								0							
1							1		1						
2						-1		-1		-1					
3					1		1		1		1				
4				-1		-1		-2		-1		-1			
5			1		1		2		2		1		1		
6		-1		-1		-3		-2		-3		-1		-1	
7	1		1		3		3		3		3		1		1
⋮															

Theorem 10. *The generating function*

$$\Gamma(x,y)=\sum_{R=0}^{\infty}\sum_{l=0}^{R}\gamma_{R;l}x^{R-l}y^{l}$$

for the coefficients $\gamma_{R;l}$ is

$$\Gamma(x,y)=\frac{x+y-xy-(x^2+y^2)}{(1-x-y)(1-x^2-y^2)}.$$

Proof. Letting $\gamma_{R;l}=0$ for $l>R$, we extend the range of second summation to $l\geqslant 0$. We have

$$\begin{aligned}\Gamma(x,y)&=\sum_{R=0}^{\infty}\gamma_{R;0}x^{R}+\sum_{R=1}^{\infty}\sum_{l=1}^{\infty}\gamma_{R;l}x^{R-l}y^{l}\\&=\frac{x}{1-x^2}+\sum_{R=1}^{\infty}\sum_{l=1}^{\infty}\gamma_{R;l}x^{R-l}y^{l},\end{aligned}$$

since $\gamma_{2k,0}=0$ and $\gamma_{2k+1;0}=1$, $k=0,1,2,\ldots$. Next, we rearrange the second sum separating even and odd values of the first and of the second index and use Lemma 9 to obtain

$$\begin{aligned}\Gamma(x,y)=\frac{x}{1-x^2}&+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\gamma_{2R+1;2l+1}x^{2(R-l)}y^{2l+1}\\&+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\gamma_{2R+1;2l+2}x^{2(R-l)-1}y^{2l+2}\\&+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\gamma_{2R+2;2l+1}x^{2(R-l)+1}y^{2l+1}\\&+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\gamma_{2R+2;2l+2}x^{2(R-l)}y^{2l+2}\\=\frac{x}{1-x^2}&+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\left[\gamma_{2R;2l+1}+\gamma_{2R;2l}+\binom{R}{l}\right]x^{2(R-l)}y^{2l+1}\\&+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\left[\gamma_{2R;2l+2}+\gamma_{2R;2l+1}+\binom{R}{l+1}\right]x^{2(R-l)-1}y^{2l+2}\\&+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\left[\gamma_{2R+1;2l+1}+\gamma_{2R+1;2l}-\binom{R}{l}\right]x^{2(R-l)+1}y^{2l+1}\end{aligned}$$

$$+\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\left[\gamma_{2R+1;2l+2}+\gamma_{2R+1;2l+1}-\binom{R+1}{l+1}\right]x^{2(R-l)}y^{2l+2}.$$

Note that

$$\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\Big[\gamma_{2R;2l}x^{2(R-l)}y^{2l+1}+\gamma_{2R;2l+1}x^{2(R-l)-1}y^{2l+2}$$
$$+\gamma_{2R+1;2l}x^{2(R-l)+1}y^{2l+1}+\gamma_{2R+1;2l+1}x^{2(R-l)}y^{2l+2}\Big]=y\Gamma(x,y),$$

and that

$$\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\Big[\gamma_{2R;2l+1}x^{2(R-l)}y^{2l+1}+\gamma_{2R;2l+2}x^{2(R-l)-1}y^{2l+2}$$
$$+\gamma_{2R+1;2l+1}x^{2(R-l)+1}y^{2l+1}+\gamma_{2R+1;2l+2}x^{2(R-l)}y^{2l+2}\Big]=x\Gamma(x,y)-\frac{x^2}{1-x^2}.$$

Furthermore,

$$\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\binom{R}{l}x^{2(R-l)}y^{2l+1}=y\sum_{R=0}^{\infty}(x^2+y^2)^R=\frac{y}{1-(x^2+y^2)},$$
$$\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\binom{R}{l+1}x^{2(R-l)-1}y^{2l+2}=\frac{x}{1-(x^2+y^2)}-\frac{x}{1-x^2}$$
$$\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\binom{R}{l}x^{2(R-l)+1}y^{2l+1}=\frac{xy}{1-(x^2+y^2)}$$
$$\sum_{R=0}^{\infty}\sum_{l=0}^{\infty}\binom{R+1}{l+1}x^{2(R-l)}y^{2l+2}=\frac{x^2+y^2}{1-(x^2+y^2)}-\frac{x^2}{1-x^2}.$$

Subsituting above and collecting terms, we deduce that

$$(1-x-y)\Gamma(x,y)=\frac{x+y-xy-(x^2+y^2)}{1-(x^2+y^2)},$$

and hence the theorem follows. □

We note that, setting $q(x,y)=x+y$, the above generating function becomes

$$\Gamma(x,y)=\frac{q(x,y)-\frac{1}{2}q(x^2,y^2)-\frac{1}{2}[q(x,y)]^2}{[1-q(x,y)][1-q(x^2,y^2)]},$$

a two-variable counterpart of (27).

Theorem 11. *With the same notation as above, it is true that*

$$\gamma_{2R;2l}=\frac{1}{2}\left[\binom{2R}{2l}-\binom{R}{l}\right],$$
$$\gamma_{2R;2l+1}=\frac{1}{2}\binom{2R}{2l+1},$$
$$\gamma_{2R+1;2l}=\frac{1}{2}\left[\binom{2R+1}{2l}+\binom{R}{l}\right],$$
$$\gamma_{2R+1;2l+1}=\frac{1}{2}\left[\binom{2R+1}{2l+1}+\binom{R}{l}\right],$$

for all $R, l \geqslant 0$.

Proof. The generating function $\Gamma(x,y)$ can be written as

$$\Gamma(x,y) = \tfrac{1}{2}\frac{1}{1-(x+y)} + \tfrac{1}{2}\frac{x+y-1}{1-(x^2+y^2)}.$$

We observe that

$$\frac{1}{1-(x+y)} = \sum_{k=0}^{\infty}\sum_{l=0}^{k}\binom{k}{l}x^{k-l}y^l$$

is the generating function of the binomial coefficients. Similarly,

$$\frac{1}{1-(x^2+y^2)} = \sum_{k=0}^{\infty}\sum_{l=0}^{k}\binom{k}{l}x^{2(k-l)}y^{2l}.$$

Multiplying the latter by $x+y-1$ and matching powers yields the desired result. $\square$

4.3. *Cayley expansions*

The most important application of the counting results presented in Section 3 is to numerical methods based on a Cayley expansion [**10**] for the solution of the differential equation (1). Recall that the solution of the *dcayinv equation* (8) can be written as

$$\Omega(t) = \sum_{m=1}^{\infty}\mathscr{I}_m(t),$$

where functions the $\mathscr{I}_m(t)$ obey the recurrence relation [**10**]

$$\mathscr{I}_m(t) = -\frac{1}{2}\int_0^t [\mathscr{I}_{m-1}(\xi), A(\xi)]\,\mathrm{d}\xi - \frac{1}{4}\sum_{k=1}^{m-2}\int_0^t \mathscr{I}_{m-k-1}(\xi)A(\xi)\mathscr{I}_k(\xi)\,\mathrm{d}\xi, \quad m \geqslant 3, \tag{45}$$

with starting condition

$$\mathscr{I}_1(t) = \int_0^t A(\xi)\,\mathrm{d}\xi.$$

The sum on the right-hand-side of (45) should be considered null when the range of indices is empty. As observed in Section 1, each $\mathscr{I}_m(t)$ features commutators and symmetric products of the form $PQR + RQP = [\![P,Q,R]\!]_3$. We invoke the hierarchy condition of the m-nary operator and claim that

$$\mathscr{I}_m(h) = \sum_{\boldsymbol{k}}\alpha_{\boldsymbol{k}}\int_{h\mathcal{S}}[\![A(\xi_{k_1}),\ldots,A(\xi_{k_m})]\!]_m\,\mathrm{d}\xi_{k_1}\cdots\,\mathrm{d}\xi_{k_m},$$

where the integration is performed over some polytope $\mathcal{S}$. The m-nary product is linear in each of its components! Therefore, given quadrature points $c_1,\ldots,c_\nu$ in $[0,1]$, and denoting $A_k = A(c_k h)$, we can approximate each of the terms

$$\int_{h\mathcal{S}}[\![A(\xi_{k_1}),\ldots,A(\xi_{k_m})]\!]_m\,\mathrm{d}\xi_{k_1}\cdots\,\mathrm{d}\xi_{k_m},$$

with a quadrature formula

$$\mathscr{Q}(h) = h^m\sum_{k_1=1}^{\nu}\cdots\sum_{k_m=1}^{\nu}b_{\boldsymbol{k}}[\![A_{k_1},\ldots,A_{k_m}]\!]_m,$$

the weights $b_{\boldsymbol{k}}$ being obtained by integrating products of cardinal Lagrange's polynomials over the polytope $\mathcal{S}$. In such a case all the terms $A_{k_1}, \ldots, A_{k_m}$ are representations of some free hierarchical algebra symbols $X_1, \ldots, X_m$, and, for each degree m we are required to evaluate a number of terms that is bounded by the dimension of the basis of $\mathscr{K}_m^\nu(\mathbf{1})$. As an example, let us consider two Gaussian nodes $c_1 = \frac{1}{2} - \frac{\sqrt{3}}{6}$ and $c_2 = \frac{1}{2} + \frac{\sqrt{3}}{6}$ in $[0, 1]$, which we can use to generate a method of order four. Taking into account further simplifying assumptions such as time-symmetry, a bound on the number of terms is given by the sum of the dimensions of graded subspaces $\mathscr{K}_m^2(\mathbf{1})$, for $m = 1, \ldots, 3$, i.e. nine (cfr. Table 1). The required primitive terms are

$$
\begin{array}{lllll}
[\![A_1]\!]_1 & [\![A_2]\!]_1 & [\![A_1, A_2]\!]_2 & [\![A_1, A_1, A_1]\!]_3 & [\![A_1, A_1, A_2]\!]_3 \\
[\![A_1, A_2, A_1]\!]_3 & [\![A_2, A_2, A_1]\!]_3 & [\![A_2, A_1, A_2]\!]_3 & [\![A_2, A_2, A_2]\!]_3. &
\end{array}
$$

Considerable savings occur when we choose graded algebras. Assume that $\{B_0, B_1\}$ is a base equivalent to $\{A_1, A_2\}$, but with weights $\boldsymbol{w} = [1, 2]$. Such a basis can be obtained by means of a Vandermonde transformation, as described in Section 1. Since the polynomial

$$p(z) = z^2 - z - 1$$

has two distinct roots, namely $\lambda = \frac{1 \pm \sqrt{5}}{2}$, according to Theorem 5,

$$
\begin{aligned}
\dim \mathscr{K}_{2R}^2[1, 2] &= \tfrac{1}{2} \sum_{l=1}^{2} \frac{1}{-1 + 2\lambda_l} (\lambda_l^{2R+1} - \lambda_l^{R+1} + \lambda_l^R) \\
\dim \mathscr{K}_{2R+1}^2[1, 2] &= \tfrac{1}{2} \sum_{l=1}^{2} \frac{1}{-1 + 2\lambda_l} (\lambda_l^{2R+2} + \lambda_l^{R+1}),
\end{aligned}
$$

and we deduce that $\dim \mathscr{K}_1^2(\boldsymbol{w}) = \dim \mathscr{K}_2^2(\boldsymbol{w}) = 1$ and $\dim \mathscr{K}_3^2(\boldsymbol{w}) = 2$, yielding a total of four terms, corresponding to

$$[\![B_0]\!]_1, \quad [\![B_1]\!]_1, \quad [\![B_0, B_1]\!]_2, \qquad [\![B_0, B_0, B_0]\!]_3.$$

4.4. *Runge–Kutta–Munthe-Kaas schemes in a Cayley-transform setting*

As a last example, we will consider the application of our results to Runge–Kutta–Munthe-Kaas methods when employing the Cayley transform as a map from the Lie algebra to the corresponding Lie group. For the more general setting of homogeneous manifolds, we refer the reader to [**6**].

The numerical methods in question consist of solving the differential equation (1) by obtaining a numerical solution to the *dcayinv equation* (8) in $\mathfrak{g}$ by means of a Runge–Kutta method, and mapping back the result to the Lie group G by means of the Cayley transform (7). Given a ν-stage Runge–Kutta scheme, defined in terms of the tableau

$$
\begin{array}{c|cccc}
c_1 & a_{1,1} & a_{1,2} & \cdots & a_{1,\nu} \\
\vdots & \vdots & \vdots & & \vdots \\
c_\nu & a_{\nu,1} & a_{\nu,2} & \cdots & a_{\nu,\nu} \\
\hline
 & b_1 & b_2 & \cdots & b_\nu
\end{array}
$$

to advance the solution by a single step from t_N to $t_{N+1} = t_N + h$ we compute

$$\begin{array}{rcll} \Psi_i & = & h\sum_{j=1}^{\nu} a_{i,j}\tilde{A}_j, & \\ A_i & = & A(t_N + hc_i), & \qquad i = 1, 2, \ldots, \nu, \\ \tilde{A}_i & = & \mathrm{dcay}_{\Psi_i}^{-1}(A_i), & \end{array}$$

hence,

$$\Omega_{n+1} = h\sum_{i=1}^{\nu} b_i \tilde{A}_i,$$

and, finally,

$$Y_{N+1} = \mathrm{cay}(h\Omega_{N+1})Y_N.$$

The matrices $\tilde{A}_1, \ldots, \tilde{A}_\nu$ are a representation of generators $X_1, \ldots, X_\nu$ of the free hierarchical algebra, with grades $w(X_i) = 1$, $i = 1, 2, \ldots, \nu$. Thus, an expansion in primitive words requires

$$\sum_{m=1}^{p} \mathrm{card}\{W \in \mathscr{B}_m^{\nu}(\mathbf{1}) |\ W \text{ has length} \leqslant 3\}$$

terms, where $p \leqslant 2\nu$ is the order of the numerical method in question. Again, one can use a graded algebra generated by $B_0, \ldots, B_{\nu-1}$ with grades $\boldsymbol{w} = [1, 2, \ldots, \nu]$, and reduce the number of required primitive words to

$$\sum_{m=1}^{p} \mathrm{card}\{W \in \mathscr{B}_m^{\nu}(\boldsymbol{w}) |\ W \text{ has length} \leqslant 3\}.$$

The savings occur by virtue of the fact that usually $\mathscr{B}_m^{\nu}(\boldsymbol{w})$ has an cardinality substantially smaller than $\mathscr{B}_m^{\nu}(\mathbf{1})$.

We should mention that for this particular example, it is not necessarily so that introducing primitive words leads to a significantly cheaper computation, due to the fact that, unlike the dexp^{-1} equation, the dcay^{-1} formula is finite and consists of a very small number of terms.

Acknowledgements

The authors wish to thank members of the Numerical Analysis group at DAMTP, University of Cambridge, for numerous discussions, comments and for allowing us to share our enthusiasm for hierarchical algebras. Particular thanks are due to Malcolm Sabin.

References

1. C. J. BUDD and A. ISERLES, 'Geometric integration: Numerical solution of differential equations on manifolds.' *Philosophical Trans. Royal Soc. A* 357 (1999) 945–956.

2. E. CELLEDONI and A. ISERLES, 'Methods for the approximation of the matrix exponential in a Lie-algebraic setting.' Tech. Rep. 015, Mathematical Sciences Research Institute, Berkeley, California, (1999).

3. R. COOLS, 'Constructing cubature formulae: the science behind the art.' *Acta Numerica* 6 (1997) 1–54.

4. P. Crouch and R. Grossman, 'Numerical integration of ordinary differential equations on manifolds.' *J. Nonlinear Sc.* 3 (1993) 1–33.

5. F. Diele, L. Lopez and R. Peluso, 'The Cayley transform in the numerical solution of unitary differential systems.' *Adv. Comput. Maths* 8 (1998) 317–334.

6. K. Engø, 'On the construction of geometric integrators in the RKMK class.' Tech. Rep. No. 158, Department of Informatics, University of Bergen, Norway, (1998).

7. F. Fer, 'Résolution del l'equation matricielle $\dot{U} = pU$ par produit infini d'exponentielles matricielles.' *Bull. Classe des Sci. Acad. Royal Belg.* 44 (1958) 818–829.

8. A. Iserles, 'Solving linear ordinary differential equations by exponentials of iterated commutators.' *Numerische Mathematik* 45 (1984) 183–199.

9. A. Iserles, *A First Course in the Numerical Analysis of Differential Equations* (Cambridge University Press, Cambridge, 1996).

10. A. Iserles, 'On Cayley-transform methods for the discretization of Lie-group equations.' Tech. Rep. NA1999/4, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, (1999).

11. A. Iserles and S. P. Nørsett, 'On the solution of linear differential equations in Lie groups.' *Philosophical Trans. Royal Soc. A* 357 (1999) 983–1019.

12. D. Lewis and J. C. Simo, 'Conserving algorithms for the dynamics of Hamiltonian systems on Lie groups.' *J. Nonlinear Sci.* 4 (1994) 253–299.

13. W. Magnus, 'On the exponential solution of differential equations for a linear operator.' *Comm. Pure Appl. Maths* VII (1954) 649–673.

14. R. I. McLachlan, G. R. W. Quispel and N. Robidoux, 'Geometric integration using discrete gradients.' *Philosophical Trans. Royal Soc. A* 357 (1999) 1021–1045.

15. H. Munthe-Kaas, 'Runge–Kutta methods on Lie groups.' *BIT* 38 (1998) 92–111.

16. H. Munthe-Kaas and B. Owren, 'Computations in a free Lie algebra.' *Philosophical Trans. Royal Soc. A* 357 (1999) 957–981.

17. H. Munthe-Kaas and A. Zanna, 'Numerical integration of differential equations on homogeneous manifolds.' 'Foundations of Computational Mathematics,' (eds F. Cucker and M. Shub) (Springer Verlag, 1997) pp. 305–315, pp. 305–315.

18. A. L. Onischik (ed.). *Lie Groups and Lie Algebras*, no. 20 in Encyclopaedia of Mathematical Sciences, Berlin, (1993). Springer-Verlag.

19. B. Owren and A. Marthinsen, 'Integration methods based on canonical coordinates of the second kind.' Tech. Rep. Numerics No. 5/1999, Norwegian University of Science and Technology, Trondheim, Norway, (1999).

20. A. Zanna, 'Collocation and relaxed collocation for the Fer and the Magnus expansions.' Tech. Rep. NA1997/17, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, (1997). To appear in *SIAM J. Num. Anal.*

Arieh Iserles `A.Iserles@damtp.cam.ac.uk`
Antonella Zanna `A.Zanna@damtp.cam.ac.uk`

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street
Cambridge
CB3 9EW