Institute of Chemical Technology, Prague

Faculty of Chemical Engineering

Department of Mathematics

Zig+Zag Dynamical Systems

(Dynamické systémy generované dvěma a více vektorovými poli)

PhD Thesis

RNDr. Pavel Pokorný

Supervisor:	Prof. RNDr. Alois Klíč, CSc.
Study program:	Applied Mathematics
Study field:	Applied Mathematics

This thesis was written at the Department of mathematics, Institute of Chemical Technology, Prague, in 2005, summarizing my results of the past ten years. It was defended on May 10, 2006.

For the preparation of this thesis I used exclusively those sources cited in the list of references.

Prague, January 25, 2006,
corrected May 11, 2006. Pavel Pokorný

[74] Pružné kyvadlo a deterministický chaos. Matematika na vysokých školách, Herbertov 1.-3.9.2003.

My software packages published under the GNU General Public License:

[75] `xpplot` graphical data plotting tool for Unix.

[76] `easynum` numerical library in C.

Publications citing my work:

[77] Štěpánek F., Kubíček M., Marek M.: Modeling and Optimization of Continuous Adsorption in a desiccant Rotor. Ind. Eng. Chem. Res. **37**, 1435 (1998) cites [33].

[78] Galan J., Freire E.: Chaos in a Mean Field Model of Coupled Quantum Wells; Bifurcations of Periodic Orbits in a Symmetric Hamiltonian System. Rep. Math. Phys. **44**, 87 (1999) cites [76].

[79] Russo L., Mancusi E., Maffettone P.L., Crescitelli S.: Symmetry Properties and Bifurcation Analysis of a Class of Periodically Forced Chemical Reactors. Chem. Eng. Sci. **57**, 5065 (2002) cites [33].

[80] Almeida J., Peralta-Salas D., Romera M.: Can Two Chaotic Systems Give Rise to Order? Physica **D 200**, 124 (2005) cites [33], [34].

[81] Afraimovich V.: Some Topological Properties of Lattice Dynamical Systems. In Chazottes J. R., Fernandez B. (Eds.): *Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems*. Series: Lecture Notes in Physics, **671**, Springer-Verlag (2005) cites [43].

[82] Greve G. C. F.: Brave GNU World, Issue 5 (1999) cites [75].

This PhD thesis has been announced in:

[83] Brindley J. et. al. (Eds.): UK Nonlinear News, Issue 42 (2006).

Acknowledgment:
I would like to thank to my ancestors for giving education to their children, to my wife Jana for love, patience and understanding, to my supervisor prof. Klíč for friendly support and scientific guidance, to my colleagues for critical comments at the department seminar “Dynamical Systems”, to prof. Kowalski for stimulating discussion and to posters to `sci.math.research` and `sci.nonlinear` newsgroups for inspiring comments.

Typeset by LaTeX under Linux (the author has used no Microsoft software for the preparation of this PhD thesis).

Contents

[61] Non-chaotic Strange Attractors in Quasi-periodically Forced Excitable Systems. 12.-15.6.1994 Woudschotten, Netherlands.

[62] On Dynamical Systems Generated by Two Alternating Vector Fields. Second Int. Conference on Difference Equations and Applications, 7.-11.8.1995, Veszprem, Hungary.

[63] Initiation and Control of Chaos by Blinking - Results. Int. Conference on Nonlinear Dynamics, Chaotic and Complex Systems, 7.-12.11.1995, Zakopane, Poland.

[64] Evaluation of Time Series. Dynamics Days Edinburgh 28.6.-1.7.1998.

[65] Dynamical Systems with Blinking. Dynamics Days Edinburgh 28. 6.-1. 7. 1998.

[66] Zig-Zag Dynamical Systems. Dynamics Days, Como, Italy 20-23.6.1999.

[67] Mathematical Modeling of Chemical and Biological Systems. Bio-mathematical Conference AVČR, Srní, Czech Republic 8.-9.4.2000.

[68] Dynamical Systems Generated by Two Alternating Vector Fields and the Baker-Campbell-Hausdorff formula. Dynamics Days, University of Surrey, England, 25.-29.6.2000.

[69] Excitable Dynamical Systems. XII Czech-Slovak-Polish Optical Conference Wave and Quantum Aspects of Contemporary Optics, Velke Losiny, Czech Republic, 12.-15.9. 2000.

[70] Dynamical Systems and the Baker-Campbell-Hausdorff Formula. Conference Nolta 2000, 17.-21.9.2000 Dresden, Germany.

[71] Mathematical Modeling of Multi-Patch Predator-Prey Model. Bio-mathematical Seminar 30.3.-1.4.2001 Cikhaj, Czech Republic.

[72] On Dynamical Systems Generated by Two Vector Fields. 3rd Scientific Colloquium, Math. Dept. Prague Institute of Chemical Technology, 26.-28.6.2001.

[73] Zig-zag Dynamical Systems and the Baker-Campbell-Hausdorff formula. The Czechoslovak Int. Conference on Differential Equations and Their Applications EQUADIFF 10, Prague, 28.-31.8.2001.

[49] Nonlinear Dynamical Systems and Prediction. Nostradamus Prediction Workshop Zlín, Czech Republic 22-23.9.1998.

[50] Nonlinear Dynamics and Prediction. Nostradamus Prediction Conference 7-8.10.1999 Zlín, Czech Republic.

[51] History and Future of Prediction. Nostradamus Int. Conference, 25.-26. 9. 2001 UTB Zlín, Czech Republic.

[52] Chaos je krásný. 21.5.2002, KMOP MFF UK.

[53] Deterministický chaos a matematické, fyzikální a filozofické souvislosti. JČMF na MFF UK Praha, 19.5.2004.

[54] Úvod do teorie nelineárních dynamických systémů a deterministického chaosu. 6. ročník semináře Matematika na vysokých školách (Determinismus a náhoda) Herbertov 5.-7.9.2005.

Contributed lectures:

[55] Coupled Map Lattices. Summer School Mathematical Modeling in Chemical Engineering. 5.-9.9.1988 Seč, Czech Republic.

[56] Cellular Automata. Summer School Mathematical Modeling in Chemical Engineering. 4.-8.9.1989 Kašperské Hory, Czech Republic.

[57] Excitable Dynamical Systems - Experiment and Modeling. Summer School Mathematical Modeling in Chemical Engineering. 10.-14.9.1990 Harrachov, Czech Republic.

[58] Excitability in Chemical and Biological Systems. Department of Physical Chemistry, Free University of Brussels, Belgium. September 1991.

[59] Reduction of Dimensionality in Excitable Dynamical Systems. Tempus Summer School Mathematical Modeling in Chemical Engineering. 24.-31.8.1992 Prague, Czech Republic.

[60] Origin and Dimension of Strange Non-chaotic Attractors. Euromech Colloquium 20.-25.9.1993 Spala, Poland.

List of symbols	
R	the set of all real numbers
R^n	the set of all n-tuples of real numbers
t	real time
x, x_0, x_1, x_2	points in R^n
$\psi(x,t)$	switching function
p	half of the switching period $T = 2p$
E	identity matrix
G	involution matrix (satisfying $G \cdot G = E$)
u, v, w, U, V, W	vector fields (maps from R^n to R^n)
u'	matrix of partial derivatives $u'_{jk} = \frac{\partial u_j}{\partial x_k}$
I	identity map (i.e. $I(x) = x$ for all $x \in R^n$)
ϕ	smooth map from R^n to R^n
Φ	the set of all smooth maps from R^n to R^n
$\hat u, \hat v, \hat w, \hat U, \hat V, \hat W$	operators (maps from Φ to Φ)
$\hat{\mathcal{I}}$	identity operator (i.e. $\hat{\mathcal{I}}(\phi) = \phi$ for all $\phi \in \Phi$)
exp	exponential $\exp \hat u = \sum_{k=0}^{\infty} \frac{\hat u^k}{k!}$
log	logarithm $\log \hat u = \sum_{k=1}^{\infty} (-1)^{k-1} (\hat u - \hat{\mathcal{I}})^k / k$
$\hat U \vdash\!\dashv \hat V$	BCH series $= \log(\exp \hat U \exp \hat V)$
$\mathcal{O}(p^3)$	terms of order 3 and higher in p
$[\hat U, \hat V]$	commutator of two operators $= \hat U \hat V - \hat V \hat U$
$[\hat A_1, \hat A_2, \ldots, \hat A_r]$	commutator of r operators $= [\hat A_1, [\hat A_2, [\ldots, \hat A_r]]]$
M_r	real $2^r \times 2^r$ matrix
$(\hat w_r)_j$	j-th word of order r, e.g. $(\hat w_3)_6 = \hat V \hat U \hat V$
$(a_r)_j$	real coefficient at $(\hat w_r)_j$
$(\hat c_r)_j$	j-th commutator of order r, e.g. $(\hat c_3)_6 = [\hat V, \hat U, \hat V]$
$(b_r)_j$	real coefficient at $(\hat c_r)_j$
N_{tot}	$= 2^r$, number of all possible r-th order words
N_w	number of r-th order words in BCH series
N_{com}	number of r-th order commutators in BCH series
$\varphi(t,x)$	flow (solution to ODE with initial condition x)

[39] Turzík, Klíč, Dubcová, Pokorný: Stability of Spatially Homogeneous Solutions in Lattice Dynamical Systems. Proceedings of 3rd Scientific Colloquium, Math. Dept. Prague Institute of Chemical Technology, 2001, pp. 38-46.

[40] Klíč, Turzík, Dubcová, Pokorný: A Simple Example of Lattice Dynamical System with Spatial Chaos. Proceedings of 3rd Scientific Colloquium, Math. Dept. Prague Institute of Chemical Technology, 2001, pp. 47-54.

[41] Pokorný, Turzík, Dubcová, Klíč: On Spectra of Certain Class of Linear Operators. Proceedings of 3rd Scientific Colloquium, Math. Dept. Prague Institute of Chemical Technology, 2001, pp. 55-63.

[42] Turzík, Dubcová, Klíč, Pokorný: Application of Bernoulli Shift to Investigation of Stability of Spatially Periodic Solutions in Lattice Dynamical Systems. Dynamical Systems: An Int. Journal Vol.18, No.1 (2003) 23-33. ISSN 1468-9367.

[43] Dubcová, Klíč, Pokorný, Turzík: Stability of Spatially Periodic Solutions in Coupled Map Lattices. Int. J. Bif. Chaos Vol.13 No.2 (2003) 343-356 ISSN 0218-1274.

[44] Proč vržou dveře, aneb Hopfova bifurkace ve vaší ložnici. Ve sborníku Matematika na vysokých školách, Herbertov 1.-3.9.2003 ed. L. Herrmann.

[45] Úvod do teorie nelineárních dynamických systémů a deterministického chaosu. Sborník 6. ročníku semináře Matematika na vysokých školách (Determinismus a náhoda) Herbertov 5.-7.9.2005 ed. L. Herrmann.

[46] Úskalí číslicových počítačů při simulaci nelineárních dynamických systémů. Sborník 6. ročníku semináře Matematika na vysokých školách (Determinismus a náhoda) Herbertov 5.-7.9.2005 ed. L. Herrmann.

[47] Modelování chaotických systémů konečnými automaty. Sborník 6. ročníku semináře Matematika na vysokých školách (Determinismus a náhoda) Herbertov 5.-7.9.2005 ed. L. Herrmann.

Invited lectures:

[48] Blinking Dynamical Systems and Campbell-Baker-Hausdorff formula. 9.Int. Colloquium on Differential Equations, Plovdiv Bulgaria 18-23.8.1998.

[29] Snyder J. D., Subramanian S.: A novel reverse flow strategy for ethylbenzene dehydrogenation in a packed bed reactor. Chem. Eng. Sci. **49**, 5585 (1994).

[30] Varadarajan V. S.: *Lie groups, Lie algebras, and their representations.* Englewood Cliffs; Prentice Hall 1974.

[31] Xiao W. D., Wang H., Yuan W. K.: Unsteady state converters with inter–stage heat removal. Chem. Eng. Sci. **54**, 4629 (1999).

[32] Zufle H., Turek T.: Catalytic combustion in a reactor with periodic flow reversal. Chem. Eng. Proc. **36**, 327 (1997).

Own publications:

[33] Klíč A., Pokorný P.: On Dynamical Systems Generated by Two Alternating Vector Fields. Int. J. Bif. Chaos **6**, 2015 (1996).

[34] Klíč A., Pokorný P., Řeháček J.: Zig-Zag Dynamical Systems and the Baker-Campbell-Hausdorff Formula. Math. Slovaca **52**, 79-97 (2002).

[35] Pokorný P., Marek M.: Origin and Dimension of Strange Non-chaotic Attractors. Series on Nonlinear Science E, Vol. 4 (Eds. Kapitaniak T. and Brindley J.) World Scientific Singapore (1994).

[36] Pokorný P., Schreiber I., Marek M. On the Route to Strangeness without Chaos in the Quasi-periodically Forced van der Pol Oscillator. Chaos, Solitons & Fractals Vol. 7 No. 3 (1996).

[37] Dynamical Systems Generated by Two Alternating Vector Fields and the Baker-Campbell-Hausdorff Formula. In Proceedings Int. Symposium of Nonlinear Theory and its Applications, Nolta 2000, 17.-21.9.2000 Dresden, Germany. IEICE. Eds. S. Oishi and W. Schwarz. ISBN 3-933592-84-4, pp.827-830.

[38] Excitable dynamical systems. In 12th Czech-Slovak-Polish Optical Conference on Wave and Quantum Aspects of Contemporary Optics. Peřina, Hrabovský, Křepelka, Editors. Proceedings of SPIE Vol. 4356, 4356-60 (2001). ISBN 0-8194-4047-7.

1 Summary

Keywords:
dynamical systems,
differential operator,
Baker-Campbell-Hausdorff formula,
sparsest solution.

The original inspiration to this research comes from chemical engineering. Experimental studies have shown that some catalytic processes, which were usually operated in stationary regimes, can be significantly improved (in terms of selectivity, conversion etc.) if they are forced to operate in periodic regimes. A special class of these processes includes reactors with periodic flow reversal. The detailed mathematical description of such a reactor would require a system of partial differential equations with a large number of parameters.

To understand the basic types of dynamical behavior of such systems, we started with models in the form of ordinary differential equations where the flow reversal is modeled by two different vector fields $u, v : R^n \to R^n$ acting one after the other

$$\frac{dx}{dt} = f(x,t)$$

where $f(x,t)$ is periodic in t with period $T = 2p$ satisfying

$$f(x,t) = \begin{cases} u(x) & \text{if} \quad 0 \le t < p \\ v(x) & \text{if} \quad p \le t < 2p. \end{cases}$$

We call this type of dynamical systems “zig+zag dynamical systems”.

The output of our research activity in this field consists of three main parts (besides posters and lectures at international conferences):

- [KP] Klíč A., Pokorný P.: On Dynamical Systems Generated by Two Alternating Vector Fields. Int. J. Bif. Chaos **6**, 2015 (1996).
- [KPR] Klíč A., Pokorný P., Řeháček J.: Zig-Zag Dynamical Systems and the Baker-Campbell-Hausdorff Formula. Math. Slovaca **52**, 79-97 (2002).
- This PhD thesis which we intend to publish in a scientific journal after minor modifications.

In [KP] Dynamical systems generated by two vector fields are studied both analytically and numerically. A special case is considered, namely when the two vector fields are related by an involution. A map G is called involution if its second iterate is equal to identity. A linear involution is represented by a matrix G satisfying

$$G \cdot G = E$$

where E is the identity matrix. Two vector fields u, v are G-related if

$$v(G \cdot x) = G \cdot u(x).$$

This is a good starting point to investigate systems modeling chemical reactors with periodic flow reversal, when the reactant gases flow for a certain time interval from one side and then for another time interval they flow from the opposite side.

Four examples of this type of models were chosen in [KP] for numerical study:

- blinking nodes
- blinking cycles
- blinking Lorenz and
- blinking vortices.

Detailed numerical investigation suggests that for small switching time interval p the resulting system can be approximated by the averaged system

$$\frac{dx}{dt} = w(x)$$

where

$$w(x) = \frac{u(x) + v(x)}{2}.$$

In [KPR] this observation was refined in the following way. Taking the average of the two vector fields is just the first order approximation of an infinite series approximation. The system of two blinking cycles was chosen to test our hypothesis. By blinking cycles we mean the following: a 2-dim system given in polar coordinates (r, ϕ)

$$\frac{dr}{dt} = r(a^2 - r^2)$$

$$\frac{d\phi}{dt} = 1$$

has a stable limit cycle with radius a and with the center in the origin. Writing the system in Cartesian coordinates and shifting the center one unit to the right we get

[15] Matros Y. Sh.: *Unsteady processes in catalytic reactor.* Elsevier, Amsterdam (1985).

[16] Matros Y. Sh.: *Catalytic process under unsteady–state conditions.* Elsevier, Amsterdam (1989).

[17] Nelson E.: *Tensor Analysis.* Princeton University Press, Princeton (1967).

[18] Nieken U., Kolios G., Eigenberger G.: Control of ignited steady state in autothermal fixed bed reactors for catalytic combustion. Chem. Eng. Sci. **49**, 5507 (1994).

[19] Olver P. J.: *Applications of Lie Groups to Differential Equations.* Springer (1993).

[20] Oteo J. A.: The Baker Campbell Hausdorff formula and nested commutator identities. J. Math. Phys. **32**, 419 (1991).

[21] Ottino J. M.: *The kinematics of mixing, stretching, chaos and transportations*. Cambridge University Press (1989).

[22] Pejas W.: Ein Beweis der qualitativen Aussage der Campbell–Hausdorff Formel fur analytische Gruppen. Arch. der Math. **19**, 453-456 (1968).

[23] Řeháček J., Kubíček M., Marek M.: Modeling of a tubular catalytic reactor with flow reversal. Chem. Eng. Sci. **47**, 2897 (1992).

[24] Řeháček J., Kubíček M., Marek M.: Periodic, quasi–periodic and chaotic spatio–temporal patterns in a tubular catalytic reactor with periodic flow reversal. Comp. Chem. Eng. **22**, 283 (1998).

[25] Reinsch M. W.: A simple expression for the terms in the Baker–Campbell–Hausdorff series. J. Math. Phys. **41**, 2434 (2000).

[26] Rossmann W.: *Lie Groups An Introduction Through Linear Groups.* Oxford University Press (2002).

[27] Russo L. et.al.: Symmetry properties and bifurcational analysis of a class of periodically forced chemical reactors. Chem. Eng. Sci. (2002).

[28] Russo L. et.al.: Nonlinear Analysis of a network of three continuous stirred tank reactors with periodic feed switching: symmetry and symmetry breaking. Int. J. Bif. Chaos **14**, 1325 (2004).

References

[1] Baker H. F.: Alternants and continuous groups. Proc. London Math. Soc., II. Ser. **3**, 24-27 (1905).

[2] Bourbaki N.: *Lie Groups and Lie Algebras.* Springer (1989).

[3] Bunimovich G. A. et.al.: SO2 oxidation in reverse flow reactor: influence of a vanadium catalytic dynamic properties. Chem. Eng. Sci. **50**, 565 (1995).

[4] Cartier P.: Demonstration algebrique de la formule de Hausdorff. Bull. Soc. math. France **84**, 241-249 (1956).

[5] Djokovic D. Z.: An Elementary Proof of the Baker–Campbell–Hausdorff–Dynkin Formula. Math. Z. **143**, 209-211 (1975).

[6] Donoho D. L.: For Most Large Under-determined Systems of Linear Equations the Minimal ℓ^1-norm Solution is also the Sparsest Solution. Preprint `http://www-stat.stanford.edu/~donoho`

[7] Dynkin E. B.: On the representation of the series $\log(e^x e^y)$ with non-commuting x and y by commutators. Mat. Sbornik, n. Ser. **25**, 155-162 (1949).

[8] Filippov A. F.: *Differential Equations with Discontinuous Righthand Sides.* Kluwer Academic Publishers Dordrecht (1988).

[9] Hochschild G.: *The structure of Lie groups.*, San Francisco–London–Amsterdam: Holden–Day, 1965.

[10] Howe R.: Very basic Lie theory. Amer. Math. Monthly **90** (1983).

[11] Israel R.: discussion in the `sci.math.research` newsgroup.

[12] Khinast J., Luss D.: Efficient bifurcation analysis of periodically–forced distributed parameters systems. Comp. Chem. Eng. **24**, 139 (2000).

[13] Kolsrud M.: Maximal reductions in the Baker–Hausdorff formula. J. Math. Phys. **34**, 270 (1993).

[14] Lorenz E. N.: Deterministic Nonperiodic Flow. J. Atmos. Sci., Vol. **20**, (1963).

the vector field u while shifting the center one unit to the left we get the vector field v.

Our conjecture in [KPR] was that the vector field describing the behavior of the resulting system is given by the Baker-Campbell-Hausdorff (BCH) formula. Taking the averaged system and also taking only the first two or three terms in the BCH series resulted in a dynamical system with a qualitatively different phase portrait. It was our great joy to see that taking the first four terms gives a system with the phase portrait indistinguishable from the original one.

It took us three years to find a satisfactory proof that the BCH series is indeed the right form for the approximation of the zig+zag dynamical systems. This PhD thesis brings the desired proof along with further results.

The main results presented in this PhD thesis are as follows. In section 3 for a given smooth vector field $u : R^n \to R^n$ we introduce a linear first order differential operator $\hat{u}$ acting on smooth maps $\phi : R^n \to R^n$

$$\hat{u}(\phi) = \phi' \cdot u$$

where both ϕ and u are functions of $x \in R^n$. Thus in full form it reads

$$\hat{u}(\phi)(x) = \phi'(x) \cdot u(x).$$

We show that the solution of ODE

$$\dot{x} = u(x)$$

with the initial condition

$$x(0) = x_0$$

can be formally written (assuming the series converges) as

$$x(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} x^{(k)}(0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \hat{u}^k I x_0 = \exp(t\hat{u}) I x_0.$$

In section 4 we apply this formalism to a dynamical system generated by two alternating vector fields u and v acting one after the other. Our aim is to find a shortcut – a third vector field w that moves the point in the state space from the same initial condition to the same final point in the same time. In section 5 we use the Taylor expansion to find the first few terms of the expansion of the shortcut w. In section 6 we find the full form of the expansion of the shortcut w. To this purpose we first have to replace composition of flows by composition of operators

$$\exp(\hat{V}) I \circ \exp(\hat{U}) I = \exp(\hat{U}) \exp(\hat{V}) I.$$

Then we can use the good old Baker–Campbell–Hausdorff formula that gives (for non-commuting x and y) the solution to

$$\exp z = \exp x \exp y$$

for z in the form of an infinite series of commutators of x and y. For commuting x and y (which is the case for real or complex numbers) we have just

$$z = x + y$$

while for non-commuting x and y (which is the case for square matrices or operators) we have

$$z = x + y + \frac{1}{2}[x, y] + \frac{1}{12}([x, x, y] - [y, x, y]) - \frac{1}{24}[x, y, x, y] +$$
$$+\frac{1}{6!}(-[x, x, x, x, y] - 6[x, x, y, x, y] - 2[x, y, y, x, y] +$$
$$+2[y, x, x, x, y] + 6[y, x, y, x, y] + [y, y, y, x, y]) +$$
$$+\frac{1}{2 \cdot 6!}([x, y, y, y, x, y] + [y, x, x, x, x, y] +$$
$$+6[y, x, x, y, x, y] - 2[y, y, x, x, x, y]) + \dots$$

Here square brackets denote nested commutator defined by (39) and (44).

This form is not unique and in section 7 we derive detailed relation between the form of the BCH series in words and in commutators. This relation can be expressed conveniently by an infinite series of square matrices M_r, where each M_r is a $r \times r$ matrix of integer numbers. Then we apply our results to derive identities between commutators. The simplest of them being

$$[x, y] + [y, x] = 0.$$

Each such identity corresponds to one vector of the kernel of the matrix M_r. These identities can be used to reduce the number of terms of a given order in the BCH series. We formulate the problem to find the sparsest form of the BCH series (i.e. the form that has the least number of terms of a given order). This problem can be solved by finding the sparsest solution to a given linear under-determined system of algebraic equations. We give an explicit algorithm to find the sparsest solution and we present numerical examples for illustration.

In section 13 we generalize our previous results to more than two vector fields.

In section 14 we apply our results to the case of a dynamical system generated by four vector fields $u, v, -u, -v$ and give a useful interpretation of the commutator $[u, v]$.

A.4 Curriculum Vitae

Born: 26.12.1962 in Pardubice, Czech Republic.
Father: Doc. MUDr. Čestmír Pokorný, CSc.
Mother: MUDr. Irena Pokorná.
Married, two children.

1986: RNDr. in Physics, Faculty of Mathematics and Physics, Charles University Prague.
1986 - 1987: Military service.
1987 - 1994: Dept. of Chemical Engineering, Institute of Chemical Technology, Prague.
1.7.-30.9.1991 Dept. of Physical Chemistry, Free University of Brussels, Belgium.
1994 - today: Dept. of Mathematics, Institute of Chemical Technology, Prague.

Teaching experience:
since 1994: Calculus I and Calculus II in Czech.
since 1996: Calculus I and Calculus II in English.
since 1997: Fourier Transform in Czech.

Language experience:
state language examination in English, German and Russian, partially French.
Member of Mensa.

and

$$[\text{Slide}, \text{Wriggle}] = \cos\beta\ (\cos\alpha, \sin\alpha, 0, 0)^T.$$

The last two vector fields are proportional to the vector field

$$(\cos\alpha, \sin\alpha, 0, 0)^T$$

which, in turn, is equal to Drive for $\beta = 0$, thus giving nothing new.

This can be conveniently computed by the following *Mathematica* program

```
Steer = {0,0,0,1};
Drive = {Cos[a+b],Sin[a+b],Sin[b],0};
d[f_]:= Outer[D,f,{x,y,a,b}];
com[f_,g_] := d[g].f - d[f].g;
Wriggle = com[Steer,Drive];
Slide   = com[Wriggle,Drive]//Simplify;
SlideSteer = com[Slide,Steer];
SlideDrive = com[Slide,Drive];
SlideWriggle = com[Slide,Wriggle];

Print["Slide=",Slide];
Print["[Slide,Steer]=",SlideSteer];
Print["[Slide,Drive]=",SlideDrive];
Print["[Slide,Wriggle]=",SlideWriggle];
```


2 Introduction

Considerable attention has been devoted to dynamical systems in the form of a set of n first order ordinary differential equations (ODE)

$$\frac{dx}{dt} = f(x) \tag{1}$$

where $f : R^n \to R^n$ is a continuous vector field.

The present work deals with a generalization of the above dynamical system, namely with dynamical systems with discontinuous right hand side

$$\frac{dx}{dt} = f(x,t) \tag{2}$$

where

$$f(x,t) = \begin{cases} u(x) & \text{if} \quad \psi(x,t) > 0 \\ v(x) & \text{if} \quad \psi(x,t) < 0 \end{cases} \tag{3}$$

where $\psi(x,t)$ (called the switching function) is a continuous function of both the state x and of the time t.

We do not deal with the question how to define $f(x,t)$ for $\psi(x,t) = 0$ here. This question is important when $\psi(x,t)$ depends on x, see [8]. Here we can take either one of the single side limits.

We consider a special case when the switching function $\psi(x,t)$ does not depend on x and is periodic in t with period $T = 2p$ and has the special form

$$\psi(x,t) \begin{cases} > 0 & \text{if} \quad 0 < t < p \\ < 0 & \text{if} \quad p < t < 2p. \end{cases} \tag{4}$$

So we arrive at the **zig+zag dynamical system**

$$\frac{dx}{dt} = f(x,t) \tag{5}$$

where $f(x,t)$ is $2p$-periodic in t satisfying

$$f(x,t) = \begin{cases} u(x) & \text{if} \quad 0 \le t < p \\ v(x) & \text{if} \quad p \le t < 2p. \end{cases} \tag{6}$$

A note on the term “zig+zag”: an alternative name is “zig-zag” where the hyphen (“-”) means simple joining of two words. We use “zig+zag” intentionally to distinguish between plus ("+") and minus ("-") where minus stands for the opposite of plus. This convention is useful when studying systems "zig+zag-zig-zag", i.e. systems generated by four vector fields $u, v, -u, -v$ where the third vector field is equal to the first one when multiplied by minus one and the fourth vector field is equal to the second one when multiplied by minus one, see chapter 14. Still another name could be "dynamical systems generated by two or more alternating vector fields".

2.1 Motivation

The motivation to investigate zig+zag dynamical systems comes from chemical engineering, see [27], [28], [23], [24]. There has been an increasing interest in periodically forced processes in literature recently. Application of forced unsteady operating conditions seems to be beneficial in several heterogeneous catalyzed processes ([3], [29] [31], [12]). Experimental studies ([32], [18]) have shown that some catalytic processes, which were usually operated in stationary regimes, can be significantly improved (in terms of selectivity, conversion etc.) if they are forced to operate in periodic regimes. Theoretical and experimental studies have shown that several processes can be operated auto-thermally avoiding the use of heat exchanger [15], [16].

2.2 Previous results

In [33] we studied zig+zag dynamical systems, where the two vector fields are related by a linear involution G (a map is called involution if its second iteration is identity) i.e. we studied systems of the form

$$v(G \cdot x) = G \cdot u(x)$$

where the matrix G satisfies

$$G \cdot G = E$$

here E is the identity matrix.

For small switching period p the method of averaging was used to approximate the non-autonomous zig+zag system by an autonomous system

$$\frac{dx}{dt} = \frac{u(x) + v(x)}{2}.$$

we have

$$[\text{Steer}, \text{Drive}] = (-\sin(\alpha+\beta), \cos(\alpha+\beta), \cos(\beta), 0)^T \tag{56}$$

meaning

$$\begin{aligned} \dot{x} &= -\sin(\alpha+\beta) \\ \dot{y} &= \cos(\alpha+\beta) \\ \dot{\alpha} &= \cos(\beta) \\ \dot{\beta} &= 0. \end{aligned}$$

Note that the angle α gives the orientation of the car (relative to a chosen direction), the angle β gives the orientation of the driving wheel (relative to the car) and the angle $\alpha+\beta$ gives the direction at which the center of the front axle moves when driving. Then the vector field (56) describes a motion of the car when the car moves in a direction perpendicular to the direction it would drive; rotating with a constant angular velocity $\dot{\alpha} = \cos(\beta)$; with fixed position β of the driving wheel. This motion can be called Wriggle (česky: vrtět se) defined by

$$\text{Wriggle} = [\text{Steer}, \text{Drive}].$$

It is easy to show that

$$[\text{Wriggle}, \text{Steer}] = \text{Drive}$$

thus giving nothing new. But

$$[\text{Wriggle}, \text{Drive}] = (-\sin(\alpha), \cos(\alpha), 0, 0)^T.$$

This motion can be called Slide, thus

$$\text{Slide} = [\text{Wriggle}, \text{Drive}]$$

because the car moves in a direction perpendicular to its axis and does not rotate! This is the very motion needed to park a car to a space that is just a little longer than the length of the car (or to come out of such a difficult position).

To give the complete picture we add that

$$[\text{Slide}, \text{Steer}] = (0, 0, 0, 0)^T,$$

$$[\text{Slide}, \text{Drive}] = \sin\beta\,(\cos\alpha, \sin\alpha, 0, 0)^T,$$

and thus

$$\dot{x} = \dot{\alpha} R_2 \cos(\alpha+\beta) = v\cos(\alpha+\beta)$$
$$\dot{y} = \dot{\alpha} R_2 \sin(\alpha+\beta) = v\sin(\alpha+\beta)$$

where the velocity v is

$$v = \dot{\alpha} R_2.$$

Since

$$\sin\beta = \frac{L}{R_2}$$

we have

$$\dot{\alpha} = \frac{v}{R_2} = \frac{v}{L}\sin\beta.$$

Choosing the length scale so that the length of the car is $L = 1$ and choosing the time scale so that the velocity is $v = 1$ we get the vector field corresponding to driving

$$\text{Drive} = (\cos(\alpha+\beta), \sin(\alpha+\beta), \sin\beta, 0)^T \tag{55}$$

meaning

$$\begin{aligned} \dot{x} &= \cos(\alpha+\beta) \\ \dot{y} &= \sin(\alpha+\beta) \\ \dot{\alpha} &= \sin(\beta) \\ \dot{\beta} &= 0. \end{aligned}$$

Let us compute the commutator

$$[u, v] = v' \cdot u - u' \cdot v$$

of the two vector fields Steer and Drive. As

$$\text{Steer}' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

and

$$\text{Drive}' = \begin{pmatrix} 0 & 0 & -\sin(\alpha+\beta) & -\sin(\alpha+\beta) \\ 0 & 0 & \cos(\alpha+\beta) & \cos(\alpha+\beta) \\ 0 & 0 & 0 & \cos(\beta) \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

In [34] we generalized the method of averaging by including higher order terms leading to the Baker - Campbell - Hausdorff (BCH) series for linear systems and we used the BCH series heuristically for nonlinear system.

The purpose of the present work is to give rigorous reasoning for the usage of the BCH series for nonlinear systems. First, we develop the operator formalism to deal with solution to ODE.

3 Operator formalism

3.1 Introduction

Although we will work only with vector fields on R^n that can be given by ordered n-tuples of smooth functions of n variables, it turns out to be useful and necessary to consider vector fields as differential operators, in agreement with [19]. We use hat to stress the operator nature of a vector field, i.e. $\hat{u}$ will denote the vector field on R^n viewed as a differential operator and u will be its coordinate form in local coordinates. As we work on R^n only, the coordinate map on R^n is the identity map ($Ix = x \; \forall x \in R^n$) and $\hat{u}I = u$ is the coordinate form of the vector field $\hat{u}$ in the coordinate system on R^n.

3.2 Vector field as an operator

Consider ODE

$$\frac{dx}{dt} = u(x) \tag{7}$$

with a smooth vector field $u : R^n \to R^n$ (where n is a fixed positive integer) and with the initial condition

$$x(0) = x_0.$$

By **smooth** we mean in $C^\infty(R^n)$ i.e. having continuous derivatives of any order. We shall write $\dot{x}$ instead of $\frac{dx}{dt}$ where possible. Assuming the solution $x(t) = \varphi(t, x_0)$ can be expressed as a Taylor series

$$x(t) = x(0) + t\dot{x}(0) + \frac{t^2}{2}\ddot{x}(0) + \ldots \tag{8}$$

for t in some neighborhood of zero, we need the time derivatives of x

$$\begin{aligned} \dot{x}(t) &= u(x(t)) \\ \ddot{x}(t) &= u'(x(t)) \cdot u(x(t)) \\ &\vdots \end{aligned}$$

Here u' is the matrix of partial derivatives

$$u'_{jk} = \frac{\partial u_j}{\partial x_k}$$

and the dot stands for multiplication of a matrix and a vector

$$(u' \cdot u)_j = \sum_{k=1}^{n} u'_{jk} u_k.$$

We will use multiple primes to denote higher order derivatives e.g.

$$u''_{jkl} = \frac{\partial^2 u_j}{\partial x_k \partial x_l}.$$

Using the chain rule for the derivative of the composition of maps $s : R^n \to R^n$ and $x : R \to R^n$

$$\frac{d}{dt} s(x(t)) = s'(x(t)) \cdot \dot{x}(t)$$

we can find the time derivative of x of any order:

$$\begin{aligned} \dot{x} &= u \\ \ddot{x} &= u' \cdot u \\ x^{(3)} &= (u' \cdot u)' \cdot u \\ x^{(4)} &= ((u' \cdot u)' \cdot u)' \cdot u \\ &\vdots \end{aligned} \tag{9}$$

(each left hand side is evaluated in t and each right hand side is evaluated in $x(t)$).

Could we introduce multiplication of vector fields $u, v : R^n \to R^n$ by

$$uv = u' \cdot v \tag{10}$$

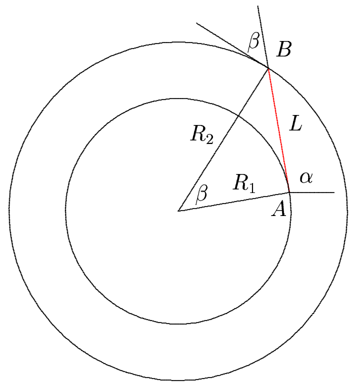

Figure 4: A driving car with a fixed orientation of the driving wheel. A (the center of the rear axle) moves on a circle with radius R_1, B (the center of the front axle) moves on a circle with radius R_2, L is the length of the car, α is the orientation of the car (relative to a chosen direction), β is the orientation of the driving wheel of the car (relative to the axis of the car).

The driver can perform two actions: steering (turning the driving wheel) and driving. Let us denote the vector field describing steering by Steer and the vector field describing driving by Drive. Then

$$\text{Steer} = (0, 0, 0, 1)^T \tag{54}$$

meaning

$$\begin{aligned} \dot{x} &= 0 \\ \dot{y} &= 0 \\ \dot{\alpha} &= 0 \\ \dot{\beta} &= 1. \end{aligned}$$

Using Fig. 4 we find that the center B of the front axle has coordinates

$$x = R_2 \cos(\alpha + \beta - \frac{\pi}{2}) = R_2 \sin(\alpha + \beta)$$

$$y = R_2 \sin(\alpha + \beta - \frac{\pi}{2}) = -R_2 \cos(\alpha + \beta)$$

where T stands for a transformation that puts y at i-th position when σ_i is found (this is the mark left by multiplication by the matrix N) and it puts x at i-th position when σ_i is not found.

A.3 Example 3: Parking problem

To illustrate the Zig+Zag-Zig-Zag system consider the following example from [17] and [26]: how to park your car into a space that is just a little longer than your car?

A car can be driven forward or backward (along an arc depending on the position of the driving wheel). In a situation shown in Fig. 3 where each rectangle represents a car parked in a street, the driver may wish to be able to drive his car to the right instead. Surprising enough, this can be done as a limiting case of a Zig+Zag-Zig-Zag dynamical system.

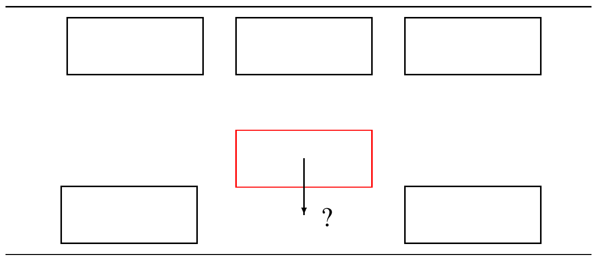

Figure 3: Parking problem: how to park your car into a space that is just a little longer than your car? Each black rectangle represents a parked car in a street, the red rectangle is your car.

The configuration of a car in a plane can be given by four real values (see Fig. 4):

- the position of the center B of the front axle in Cartesian coordinates (x, y),
- the angle α specifying the orientation of the car relative to a chosen direction
- the angle β specifying the orientation of the driving wheel of the car relative to the axis of the car.

then the Taylor series (8) would become an exponential. Unfortunately, this multiplication would not be associative because for $u, v, w : R^n \to R^n$

$$(uv)w = (u' \cdot v)' \cdot w$$

but on the other hand

$$u(vw) = u' \cdot (v' \cdot w)$$

which is different. We still want to express the Taylor series (8) as an exponential.

To this purpose for any given smooth vector field $u : R^n \to R^n$ we introduce a linear first order differential **operator**

$$\hat{u} : \Phi \to \Phi.$$

Here Φ is the set of all smooth maps $\phi : R^n \to R^n$. The action of the operator $\hat{u}$ on any map $\phi \in \Phi$ is given by

$$\hat{u}\phi = \phi' \cdot u \qquad (11)$$

i.e. $\hat{u}\phi$ is a map that assigns to each $x \in R^n$ the point

$$x \mapsto \hat{u}\phi x = (\hat{u}\phi)(x) = (\phi' \cdot u)(x) = \phi'(x) \cdot u(x).$$

The alternative grouping, namely $\hat{u}(\phi(x))$ does not make sense, since the operator can be applied to a map, here ϕ, not to a point $\phi(x)$.

The action of the operator $\hat{u}$ on the identity map I, i.e. a map satisfying

$$I(x) = x \qquad \text{for all } x \in R^n \qquad (12)$$

is

$$\hat{u}I = I' \cdot u = E \cdot u = u \qquad (13)$$

where E is the $n \times n$ identity matrix.

We also define the identity operator $\hat{\mathcal{I}}$ by

$$\hat{\mathcal{I}}\phi = \phi \qquad \text{for all } \phi \in \Phi. \qquad (14)$$

The identity operator $\hat{\mathcal{I}}$ is of order zero (no derivative of ϕ is necessary to evaluate $\hat{\mathcal{I}}\phi = \phi$) and $\hat{\mathcal{I}}$ does not correspond to any vector field, because each operator corresponding to a vector field is of order one.

Note: we will apply the operators only to the identity map I or to a map that results from applying another operator to the identity map I, which is actually applying the composition of operators to the identity map.

Powers of $\hat{u}$ can be used to express time derivatives of $x(t)$ in (9). Namely

$$x^{(k)}(t) = \hat{u}^k I x(t)$$

meaning

$$\begin{aligned} \hat{u}I &= u \\ \hat{u}^2 I &= \hat{u}\hat{u}I = \hat{u}u = u' \cdot u \\ \hat{u}^3 I &= \hat{u}u' \cdot u = (u' \cdot u)' \cdot u \\ \hat{u}^4 I &= \hat{u}(u' \cdot u)' \cdot u = ((u' \cdot u)' \cdot u)' \cdot u \\ &\vdots \end{aligned}$$

And for $t = 0$ we get

$$x^{(k)}(0) = \hat{u}^k I x_0.$$

We define (in accord with [10], [19] etc.) the **exponential** of an operator $\hat{u}$ as

$$\exp \hat{u} = \sum_{k=0}^{\infty} \frac{\hat{u}^k}{k!}. \tag{15}$$

Note that if $\hat{u}$ is a first order differential operator (corresponding to a vector field $u : R^n \to R^n$), then $\hat{u}^2$ is a second order differential operator, $\hat{u}^3$ is a third order differential operator etc. and $\exp \hat{u}$ has not a finite order.

Now we can express the Taylor series (8) by the exponential of the operator $\hat{u}$ corresponding to the vector field u on the right hand side of the ODE (7)

$$x(t) = \varphi(t, x_0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} x^{(k)}(0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \hat{u}^k I x_0 = \exp(t\hat{u}) I x_0. \tag{16}$$

We stress that the exponential in (16) is a formal series, which is equivalent to the Taylor series of the solution of the given ODE. The exponential

$$\exp(t\hat{u})$$

is not an operator in the sense that when applied to a function it gives another function defined on a certain subset of R^n independent of t. As we will see in the next example, (16) gives the solution of the ODE which is defined on a certain interval of t that may depend on the initial condition, which is a common case for ODE.

A.2 Alternative way to find the terms in the BCH series

Reinsch in [25] in order to find the r-th order term in the BCH series uses two $(r+1) \times (r+1)$ matrices

$$M_{ij} = \delta_{i+1,j}$$

and

$$N_{ij} = \delta_{i+1,j}\, \sigma_i$$

i.e.

$$M = \begin{pmatrix} 0 & 1 & 0 & \dots & & & \\ & 0 & 1 & 0 & \dots & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & 0 & 1 \\ & & & & & & 0 \end{pmatrix}$$

and

$$N = \begin{pmatrix} 0 & \sigma_1 & 0 & \dots & & & \\ & 0 & \sigma_2 & 0 & \dots & & \\ & & \cdot & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & 0 & \sigma_r \\ & & & & & & 0 \end{pmatrix}.$$

When these matrices multiply a $(r+1)$-dim vector

$$(0, 0, \dots, 1)^T$$

from the left, then the only nonzero component of the vector moves one position up with each multiplication. Multiplication by the matrix M leaves this nonzero component unchanged, while multiplication by the matrix N multiplies this nonzero component by one more σ_i, thus leaving a mark to be used later by the transformation rule T.

Then the r-th order term of

$$z = \log(\exp(M)\exp(N))$$

is

$$z_r = T(\log(\exp(M)\exp(N)))_{1,r+1}$$

and

$$\mathrm{Ad}\, a\ (x) = [a, x] = ax - xa$$

and D the formal differentiation with respect to t. Then he derives step by step two different results of $(D \exp(h)) \exp(-h)$, namely

$$(D \exp(h)) \exp(-h) = \hat{U} + \exp(\mathrm{Ad}\, h)(\hat{V})$$

and

$$(D \exp(h)) \exp(-h) = f(\mathrm{Ad}\, h)(Dh)$$

where

$$f(x) = \frac{\exp(x) - 1}{x} = \sum_{n=1}^{\infty} \frac{x^{n-1}}{n!} = 1 + \frac{x}{2} + \frac{x^2}{6} + \frac{x^3}{24} + \frac{x^4}{120} + \dots$$

Thus

$$f(\mathrm{Ad}\, h)(Dh) = \hat{U} + \exp(\mathrm{Ad}\, h)(\hat{V}).$$

Introducing

$$g(x) = \frac{1}{f(x)} = \frac{x}{\exp(x) - 1} = 1 - \frac{x}{2} + \frac{x^2}{12} - \frac{x^4}{6!} + \frac{x^6}{6 \cdot 7!} + \dots$$

Dh can be expressed as

$$Dh = g(\mathrm{Ad}\, h)(\hat{U} + \exp(\mathrm{Ad}\, h)(\hat{V})).$$

This is a differential equation for h that provides a recursion formulas for h_n. The coefficient of t^n on the left is

$$(n + 1)h_{n+1}$$

and that on the right is a linear combination of expressions of the form

$$\mathrm{Ad}\, h_{m_1} \dots \mathrm{Ad}\, h_{m_k}(z)$$

where $z = \hat{U}$ or $\hat{V}$ and $m_1 + \ldots + m_k = n$. Thus if h_j are Lie elements (i.e. can be expressed by commutators of $\hat{U}$ and $\hat{V}$ only) for $j \leq n$ so is h_{n+1}. Since $h_0 = 0$ (and also $h_1 = \hat{U} + \hat{V}$), h is a Lie element.

Note: some authors (see e.g. [19]) introduce differential operators acting on real scalar–valued functions alone and only then they let the operator act component–wise on vector–valued functions. We believe that our approach, namely to introduce differential operators acting on vector–valued functions directly is more natural. Also, by using the hat notation (adopted from quantum mechanics) we stress the difference between a map (acting on points in R^n) and an operator (acting on maps from R^n to R^n). This difference is often neglected.

3.3 Example 1: ODE

To illustrate the operator formalism just introduced consider a differential equation

$$\dot{x} = x^2$$

with the initial condition $x(0) = x_0 \in R$.

First, let us solve this ODE by separation of variables:

$$x(t) = \frac{x_0}{1 - t\, x_0}.$$

For

$$|t\, x_0| < 1$$

this solution can be expressed as a Taylor series

$$x(t) = x_0 \sum_{k=0}^{\infty} (t\, x_0)^k.$$

As the second approach we will use the operator formalism introduced above. For the 1-dim vector field $u(x) = x^2$ we have the corresponding operator

$$\hat{u}(\phi) = \phi' \cdot u$$

i.e.

$$\hat{u}(\phi)(x) = \phi'(x) \cdot x^2.$$

This operator when applied to the identity map $I(x) = x$ is

$$(\hat{u} I)(x) = x^2$$

and its k-th power is

$$(\hat{u}^k I)(x) = k!\, x^{k+1}.$$

The exponential of the operator $t\hat{u}$ is

$$\exp(t\hat{u}) = \sum_{k=0}^{\infty} \frac{(t\hat{u})^k}{k!}$$

and this exponential when applied to the identity map I gives in x

$$(\exp(t\hat{u})I)(x) = (\sum_{k=0}^{\infty} \frac{(t\hat{u})^k}{k!} I)(x) = \sum_{k=0}^{\infty} t^k x^{k+1} = x \sum_{k=0}^{\infty} (tx)^k.$$

Then the solution can be expressed as

$$x(t) = (\exp(t\hat{u})I)(x_0) = x_0 \sum_{k=0}^{\infty} (tx_0)^k$$

which is exactly the same result as we obtained by standard methods.

The result obtained by operators is equivalent to the result obtained by Taylor series in both its values and its domain of convergence. This example also demonstrates that we do not have to suppose the existence of a global flow, i.e. a solution defined for all real t. In this example for positive initial condition the solution exists for $t < 1/x_0$ and for negative initial condition the solution exists for $t > 1/x_0$ and for zero initial condition the solution exists for $t \in R$.

The domain of convergence of the Taylor series is a subinterval of the domain of definition of the solution.

4 Two vector fields

Consider a zig+zag dynamical system generated by two vector fields u, v

$$\dot{x} = f(x,t) \tag{17}$$

where $f(x,t)$ is $2p$-periodic in t satisfying

$$f(x,t) = \begin{cases} u(x) & \text{if} \quad 0 \le t < p \\ v(x) & \text{if} \quad p \le t < 2p \end{cases} \tag{18}$$

with the initial condition $x(0) = x_0$, see Fig. 1.

In the first half of the period, i.e. in time p, the vector field u moves the state point from the initial condition $x(0) = x_0$ to the point

$$x_1 = \exp(p\hat{u})Ix_0.$$

14 Zig+Zag-Zig-Zag System

Consider the dynamical system generated by four vector fields $u, v, -u, -v$ acting one after the other. Using the result of the previous chapter for $m = 4$ we have

$$\hat{W} = \sum_{k=1}^{4} \hat{U}_k + \frac{1}{2} \sum_{j=1}^{3} \sum_{k=j+1}^{4} [\hat{U}_j, \hat{U}_k] + \mathcal{O}(p^3) =$$

$$= \hat{U}_1 + \hat{U}_2 + \hat{U}_3 + \hat{U}_4 + \frac{1}{2}([\hat{U}_1, \hat{U}_2] + [\hat{U}_1, \hat{U}_3] +$$

$$+ [\hat{U}_1, \hat{U}_4] + [\hat{U}_2, \hat{U}_3] + [\hat{U}_2, \hat{U}_4] + [\hat{U}_3, \hat{U}_4]) + \mathcal{O}(p^3).$$

As

$$\begin{aligned} \hat{U}_1 &= \hat{U} \\ \hat{U}_2 &= \hat{V} \\ \hat{U}_3 &= -\hat{U} \\ \hat{U}_4 &= -\hat{V} \end{aligned}$$

we have

$$\hat{W} = \hat{U} + \hat{V} - \hat{U} - \hat{V} + [\hat{U}, \hat{V}] + [\hat{U}, -\hat{U}] + [\hat{U}, -\hat{V}] +$$

$$+ [\hat{V}, -\hat{U}] + [\hat{V}, -\hat{V}] + [-\hat{U}, -\hat{V}] + \mathcal{O}(p^3) = [\hat{U}, \hat{V}] + \mathcal{O}(p^3).$$

This gives an important interpretation of the commutator of two vector fields: it is a vector field which is tangent to the curve with parametric equation

$$x_4 = x_4(p)$$

of the "zig+zag-zig-zag" system generated by four vector fields, where the third vector field is the opposite to the first one and the fourth vector field is the opposite to the second one.

A Appendices

A.1 Proof of the BCH formula

The main idea of the proof by Djokovic in [5] that $\hat{W}$ can be expressed by commutators of $\hat{U}$ and $\hat{V}$ only is as follows. He denotes

$$h = \log(\exp(t\hat{U}) \exp(t\hat{V})) = \sum_{n=1}^{\infty} h_n(\hat{U}, \hat{V})\, t^n$$

where $f(x,t)$ is $3p$-periodic in t satisfying

$$f(x,t) = \begin{cases} u_1(x) & \text{if} \quad 0 \le t < p \\ u_2(x) & \text{if} \quad p \le t < 2p \\ u_3(x) & \text{if} \quad 2p \le t < 3p. \end{cases} \tag{52}$$

In analogy with (29) we introduce rescaled vector fields

$$U_k = p\, u_k \qquad \text{for } k = 1, \dots, 3$$

and the corresponding operators $\hat{U}_k$.

We again want to find a shortcut – a single vector field W that moves the point in the state space from the same initial condition $x_0 = x(0)$ to the same final point $x_3 = x(3p)$.

To use our previous results it is convenient to introduce a new symbol $\vDash$ resembling the letter H for Hausdorff (see [2]) by

$$\hat{U} \vDash \hat{V} = \log(\exp \hat{U} \exp \hat{V})$$

or equivalently

$$\exp(\hat{U} \vDash \hat{V}) = \exp \hat{U} \exp \hat{V}. \tag{53}$$

We want to find $\hat{W}$ satisfying

$$\exp \hat{W} = \exp \hat{U}_1 \exp \hat{U}_2 \exp \hat{U}_3.$$

Using (53) twice we get

$$\exp \hat{W} = \exp(\hat{U}_1 \vDash \hat{U}_2) \exp \hat{U}_3 = \exp(\hat{U}_1 \vDash \hat{U}_2 \vDash \hat{U}_3)$$

and thus

$$\hat{W} = \hat{U}_1 \vDash \hat{U}_2 \vDash \hat{U}_3.$$

It is easy to show that our previous result

$$\hat{U} \vDash \hat{V} = \hat{U} + \hat{V} + \frac{1}{2}[\hat{U}, \hat{V}] + \mathcal{O}(p^3)$$

implies

$$\hat{U}_1 \vDash \hat{U}_2 \vDash \hat{U}_3 = \hat{U}_1 + \hat{U}_2 + \hat{U}_3 + \frac{1}{2}([\hat{U}_1, \hat{U}_2] + [\hat{U}_1, \hat{U}_3] + [\hat{U}_2, \hat{U}_3]) + \mathcal{O}(p^3).$$

It is easy to generalize this result from 3 to m vector fields $\hat{U}_1, \dots, \hat{U}_m$ acting one after another. By induction we find

$$\hat{W} = \sum_{k=1}^{m} \hat{U}_k + \frac{1}{2} \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} [\hat{U}_j, \hat{U}_k] + \mathcal{O}(p^3).$$

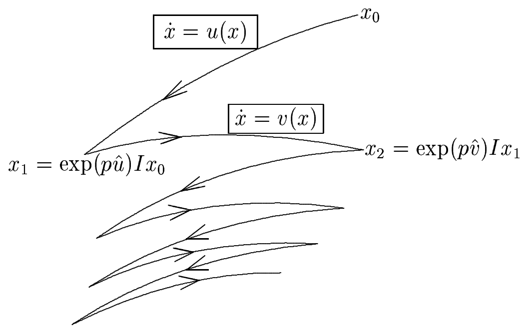

Figure 1: A typical trajectory of a zig+zag dynamical system (17–18) with the initial condition $x(0) = x_0$. In time p the vector field u moves the state point from x_0 to x_1 and then in another time p the vector field v moves the state point from x_1 to x_2. And then these two steps repeat periodically producing a zig+zag shaped trajectory giving the name to this class of dynamical systems.

And then in the second half of the period, i.e. in another time p, the vector field v moves the state point from the point x_1 to the point

$$x_2 = \exp(p\hat{v})I(x_1) = \exp(p\hat{v})I(\exp(p\hat{u})I(x_0)). \tag{19}$$

Our aim is to find a **shortcut** (a short way from x_0 to x_2) – a single vector field, let us call it w, that moves the state point from the same initial condition x_0 to the same final point x_2 in the same time $2p$ (not necessarily going via the point x_1, actually almost surely not). Thus replacing the non-autonomous dynamical system (17) by the autonomous system

$$\dot{x} = w(x).$$

5 Shortcut by Taylor

Assuming the solution to the ODE

$$\dot{x} = u(x)$$

can be written as

$$x(t) = x(0) + tu(x(0)) + \frac{t^2}{2}u'(x(0)) \cdot u(x(0)) + \mathcal{O}(t^3)$$

we have (omitting the argument if equal to $x(0)$)

$$x_1 = x_0 + pu + \frac{p^2}{2}u' \cdot u + \mathcal{O}(p^3)$$

and similarly for the ODE

$$\dot{x} = v(x)$$

assuming we can write the solution in the form of a Taylor series we have

$$x_2 = x_1 + pv(x_1) + \frac{p^2}{2}v'(x_1) \cdot v(x_1) + \mathcal{O}(p^3).$$

Now, we want to express $v(x_1)$ and $v'(x_1)$ in terms of $v(x_0)$ and $v'(x_0)$. Using

$$x_1 - x_0 = pu + \mathcal{O}(p^2)$$

we get

$$v(x_1) = v(x_0) + v'(x_0) \cdot pu(x_0) + \mathcal{O}(p^2)$$

and

$$v'(x_1) = v'(x_0) + \mathcal{O}(p)$$

thus (omitting the argument if equal to x_0)

$$x_2 = x_1 + p(v + v' \cdot pu) + \frac{p^2}{2}v' \cdot v + \mathcal{O}(p^3) =$$

$$= x_0 + pu + \frac{p^2}{2}u' \cdot u + p(v + v' \cdot pu) + \frac{p^2}{2}v' \cdot v + \mathcal{O}(p^3) =$$

$$= x_0 + p(u + v) + \frac{p^2}{2}(u' \cdot u + 2v' \cdot u + v' \cdot v) + \mathcal{O}(p^3). \tag{20}$$

We want (20) to be equal to the solution of the ODE

$$\dot{x} = w(x)$$

for $t = 2p$, namely

$$x_2 = x_0 + 2pw + \frac{1}{2}(2p)^2 w' \cdot w + \mathcal{O}(p^3).$$

Since

$$U'(x) = \mathcal{U}$$

and

$$V'(x) = \mathcal{V}$$

we have

$$W_2(x) = \frac{1}{2}(\mathcal{V} \cdot \mathcal{U} \cdot x - \mathcal{U} \cdot \mathcal{V} \cdot x)$$

so we can write

$$W_2 = \frac{1}{2}(\mathcal{V} \cdot \mathcal{U} - \mathcal{U} \cdot \mathcal{V}) = \frac{1}{2}[\mathcal{V},\mathcal{U}] = -\frac{1}{2}[\mathcal{U},\mathcal{V}].$$

As a result, for linear systems we can use the BCH series directly for matrices, but then we have to multiply by -1 each even order term.

12.1 Ambiguity of matrix commutator

This uncomfortable result leads various authors to define the commutator of two matrices in different ways. While Rossmann in [26] p.14 defines the commutator of two matrices $\mathcal{U}$ and $\mathcal{V}$

$$[\mathcal{U},\mathcal{V}]_{Rossmann} = \mathcal{U} \cdot \mathcal{V} - \mathcal{V} \cdot \mathcal{U},$$

Olver in [19] p.44 defines the same commutator as

$$[\mathcal{U},\mathcal{V}]_{Olver} = \mathcal{V} \cdot \mathcal{U} - \mathcal{U} \cdot \mathcal{V}.$$

We follow Rossmann, because this notation is more common in literature.

13 Zig+Zag+Zug Systems

In the preceding chapters we studied dynamical systems generated by two vector fields $u, v : R^n \to R^n$ acting one after the other in time. In this chapter we investigate a generalization to the case when there are more than two vector fields. As the first step in this generalization is the case with three vector fields, we choose to call them “zig+zag+zug” systems. To be more precise we study dynamical systems of the form

$$\frac{dx}{dt} = f(x,t) \tag{51}$$

Then the solution of the ODE

$$\dot{x} = w(x)$$

with the initial condition

$$x(0) = 1$$

in time $t = 2p$ can be found by Taylor expansion

$$x(t) = x(0) + t\dot{x}(0) + \frac{t^2}{2}\ddot{x}(0) + \frac{t^3}{3!}x^{(3)}(0) + \frac{t^4}{4!}x^{(4)}(0) + \mathcal{O}(t^5)$$

$$x_2 = x(2p) = 1 + 2p + \frac{11p^2}{2} + 17p^3 + \frac{443p^4}{8} + \mathcal{O}\left(p^5\right)$$

which agrees completely with the result obtained without BCH.

12 Application to linear ODE

Suppose the two vector fields U and V are linear, i.e. there are two $n \times n$ matrices $\mathcal{U}$ and $\mathcal{V}$ such that

$$U(x) = \mathcal{U} \cdot x$$

and

$$V(x) = \mathcal{V} \cdot x.$$

Note to symbols: names of matrices are usually ordinary capital letter. We want to distinguish between a vector field and a matrix and still use related symbols, so we choose calligraphic capital letters for matrices.

Then the first order term in the expansion of the operator $\hat{W}$

$$\hat{W}_1 = \hat{U} + \hat{V}$$

gives the first order term in the expansion of the vector field W

$$W_1 = \hat{W}_1 I = (\hat{U} + \hat{V})I = U + V.$$

The second order term in the expansion of the operator

$$\hat{W}_2 = \frac{1}{2}[\hat{U}, \hat{V}]$$

gives

$$W_2 = \hat{W}_2 I = \frac{1}{2}[\hat{U}, \hat{V}]I = \frac{1}{2}(\hat{U}\hat{V} - \hat{V}\hat{U})I = \frac{1}{2}(V' \cdot U - U' \cdot V).$$

We can search for w in the form of a series in p. Denoting w_0 the leading term and denoting pw_1 the first order term and neglecting higher order terms we can write

$$w = w_0 + pw_1 + \mathcal{O}(p^2).$$

Then

$$x_2 = x_0 + 2p(w_0 + pw_1) + \frac{1}{2}(2p)^2(w_0 + pw_1)' \cdot (w_0 + pw_1) + \mathcal{O}(p^3) =$$

$$= x_0 + p(2w_0) + p^2(2w_1 + 2w_0 \cdot w_0) + \mathcal{O}(p^3). \tag{21}$$

Comparing the terms of x_2 from (20) and (21) linear in p we have

$$w_0 = \frac{u + v}{2}.$$

In the method called averaging w is approximated by w_0 only.

Using this w_0 and comparing the terms of x_2 from (20) and (21) of order p^2 we get

$$w_1 = \frac{v' \cdot u - u' \cdot v}{4}.$$

Together we get

$$w = w_0 + pw_1 + \mathcal{O}(p^2) = \frac{u + v}{2} + p\frac{v' \cdot u - u' \cdot v}{4} + \mathcal{O}(p^2). \tag{22}$$

This approach can be used to any order. However, a more elegant way is to use the BCH series, which gives more insight.

6 Shortcut by BCH

6.1 Composition of flows

First, we want to express the composition of maps

$$\exp(p\hat{v})I \circ \exp(p\hat{u})I$$

from (19) by appropriate composition of operators (applied to the identity map).

For this purpose consider a map $s : R^n \to R^n$ applied to the solution of the ODE

$$\dot{x} = u(x)$$

i.e. applied to the map

$$x(t) = \exp(t\hat{u})Ix_0.$$

Let us call the composition S

$$S = s \circ x$$

i.e.

$$S(t) = s(x(t)).$$

Using the chain rule we can find the derivatives of S:

$$\begin{array}{lrcccc} & S(t) & = & s(x(t)) & = & \hat{s}Ix(t) \\ & \dot{S}(t) & = & s'(x(t)) \cdot u(x(t)) & = & \hat{u}\hat{s}Ix(t) \\ \text{in short} & \dot{S} & = & s' \cdot u & = & \hat{u}\hat{s}Ix(t) \\ & \ddot{S} & = & (s' \cdot u)' \cdot u & = & \hat{u}^2\hat{s}Ix(t) \\ & S^{(3)} & = & ((s' \cdot u)' \cdot u)' \cdot u & = & \hat{u}^3\hat{s}Ix(t) \end{array} \tag{23}$$

in general

$$S^{(k)} = \hat{u}^k\hat{s}Ix(t).$$

Assuming we can express $S(t)$ in the form of a Taylor series we have

$$s(x(t)) = S(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} S^{(k)}(0) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \hat{u}^k \hat{s} I x_0 = \exp(t\hat{u})\hat{s}Ix_0.$$

Then consider

$$s(x(t)) = \exp(t\hat{u})\hat{s}Ix_0 \tag{24}$$

for a special operator $\hat{s}$

$$\hat{s} = \exp(p\hat{v}) \tag{25}$$

and its corresponding map (the flow of the ODE $\dot{x} = v(x)$)

$$s = \hat{s}I = \exp(p\hat{v})I. \tag{26}$$

Then after putting (25) and (26) into (24) we have

$$\exp(p\hat{v})I(x(t)) = \exp(t\hat{u})\exp(p\hat{v})Ix_0$$

and

$$\exp(p\hat{v})I \circ \exp(t\hat{u})I = \exp(t\hat{u})\exp(p\hat{v})I. \tag{27}$$

This is a key result that allows to treat zig+zag dynamical systems by operators because we can study composition of flows by investigating "multiplication" of exponentials of operators.

thus

$$x_2 = \frac{x_1}{\sqrt{1 - 2px_1^2}} = \frac{\frac{1}{1-p}}{\sqrt{1 - \frac{2p}{(1-p)^2}}} = \frac{1}{\sqrt{1 - 4p + p^2}}$$

for

$$0 \le p < 2 - \sqrt{3}.$$

The Taylor series of x_2 is

$$x_2 = 1 + 2p + \frac{11p^2}{2} + 17p^3 + \frac{443p^4}{8} + \mathcal{O}\left(p^5\right).$$

Let us now use the BCH series to find x_2 and to compare it with the above result. We first prepare the commutator

$$[u, v] = v' \cdot u - u' \cdot v$$

of two power functions x^m and x^k

$$[x^m, x^k] = kx^{k-1}x^m - mx^{m-1}x^k = (k-m)x^{m+k-1}.$$

Thus

$$[u, v] = [x^2, x^3] = x^4,$$
$$[u, u, v] = [x^2, x^4] = 2x^5,$$
$$[v, u, v] = [x^3, x^4] = x^6,$$
$$[u, v, u, v] = [x^2, x^6] = 4x^7$$

etc.

The shortcut vector field w that moves the same initial condition x_0 to the same final point x_2 (not necessarily, actually almost never, via the point x_1) in the same time $t = 2p$ can be found by the BCH formula, recalling (29)-(31)

$$w = \frac{1}{2p}W =$$

$$= \frac{1}{2p}(U + V + \frac{1}{2}[U, V] + \frac{1}{12}([U, U, V] - [V, U, V]) - \frac{1}{24}[U, V, U, V] + \mathcal{O}(p^5)) =$$

$$= \frac{u+v}{2} + \frac{p}{4}[u, v] + \frac{p^2}{24}([u, u, v] - [v, u, v]) - \frac{p^3}{48}[u, v, u, v] + \mathcal{O}(p^4) =$$

$$= \frac{x^2 + x^3}{2} + \frac{px^4}{4} + \frac{p^2}{24}(2x^5 - x^6) - \frac{p^3x^7}{12} + \mathcal{O}(p^4).$$

As shown in Table 1 we have found the sparsest solution using the algorithm described above for order up to 10 and the minimal number N_{com} of the commutators in the terms of order r in the BCH series is given in the table. For order 11 and 12 we have found a solution using the Gauss elimination without removing columns but this solution is not guaranteed to be the sparsest one because the algorithm described above takes too long for large r. So the numbers in Table 1 give only an upper estimate for N_{com} for r 11 and 12.

11 Example 2: Zig+Zag system

To illustrate the above results consider the following 1-dim example of a zig+zag system with

$$u(x) = x^2$$

acting for time p and

$$v(x) = x^3$$

acting for another time p with the initial condition

$$x(0) = x_0 = 1.$$

We first find x_1 and x_2 by separation of variables and we find the Taylor series of x_2 as a function of the parameter p. Then we compare this result with that obtained by the BCH series.

The ODE

$$\dot{x} = x^2$$

has the solution

$$x(t) = \frac{x(0)}{1 - tx(0)}$$

thus

$$x_1 = \frac{1}{1-p}.$$

The ODE

$$\dot{x} = x^3$$

has the solution

$$x(t) = \frac{x(0)}{\sqrt{1 - 2tx(0)^2}}$$

For $t = p$ we have

$$\exp(p\hat{v})I \circ \exp(p\hat{u})I = \exp(p\hat{u})\exp(p\hat{v})I. \tag{28}$$

For three vector fields $u, v, w : R^n \to R^n$ and for $p > 0$ we can introduce three new vector fields $U, V, W : R^n \to R^n$ by

$$\begin{aligned} U &= pu && (29)\\ V &= pv && (30)\\ W &= 2pw. && (31) \end{aligned}$$

Then the corresponding operators satisfy

$$\begin{aligned} \hat{U} &= p\hat{u} && (32)\\ \hat{V} &= p\hat{v} && (33)\\ \hat{W} &= 2p\hat{w}. && (34) \end{aligned}$$

Then (28) can be written in a simpler form

$$\boxed{\exp(\hat{V})I \circ \exp(\hat{U})I = \exp(\hat{U})\exp(\hat{V})I.} \tag{35}$$

To avoid confusion, if $\hat{U}$ and $\hat{V}$ are operators, so are their exponentials $\exp(\hat{U})$ and $\exp(\hat{V})$. Then $\exp(\hat{U})I$ and $\exp(\hat{V})I$ are these two operators applied to the identity map I from R^n to R^n resulting in another two maps from R^n to R^n. On the left we have the composition of these two maps. On the right the composition of the two exponentials $\exp(\hat{U})$ and $\exp(\hat{V})$ of operators being another operator is applied to the identity map I from R^n to R^n resulting in another map from R^n to R^n.

Note the **reversed order** of $\hat{U}$ and $\hat{V}$ on the left and on the right hand side. Also note that (35) is far from obvious. E.g. it is not true when the identity map I is replaced by another map ϕ, because then even the zero order terms do not agree: on the left we have $\phi\circ\phi$ while on the right we have just ϕ which is not equal unless ϕ is a projection.

Due to the rescaling (32-34), $\hat{U}$ and $\hat{V}$ are proportional to p (for fixed $\hat{u}$ and $\hat{v}$). When p is small, so are $\hat{U}$ and $\hat{V}$ and we can work with series expansions.

As an example we give the second order expansion of both sides of (35). On the left we have

$$\exp(\hat{V})I \circ \exp(\hat{U})I = (\hat{\mathcal{I}} + \hat{V} + \frac{1}{2}\hat{V}\hat{V})I \circ (\hat{\mathcal{I}} + \hat{U} + \frac{1}{2}\hat{U}\hat{U})I + \mathcal{O}(p^3) =$$

$$= (I + V + \frac{1}{2}V' \cdot V) \circ (I + U + \frac{1}{2}U' \cdot U) + \mathcal{O}(p^3) =$$
$$= I + U + V + \frac{1}{2}U' \cdot U + \frac{1}{2}V' \cdot V + V' \cdot U + \mathcal{O}(p^3)$$

and on the right we have

$$\exp(\hat{U})\exp(\hat{V})I = (\hat{\mathcal{I}} + \hat{U} + \frac{1}{2}\hat{U}\hat{U})(\hat{\mathcal{I}} + \hat{V} + \frac{1}{2}\hat{V}\hat{V})I + \mathcal{O}(p^3) =$$
$$= (\hat{\mathcal{I}} + \hat{U} + \hat{V} + \frac{1}{2}\hat{U}\hat{U} + \frac{1}{2}\hat{V}\hat{V} + \hat{U}\hat{V})I + \mathcal{O}(p^3) =$$
$$= I + U + V + \frac{1}{2}U' \cdot U + \frac{1}{2}V' \cdot V + V' \cdot U + \mathcal{O}(p^3).$$

The presence of the term $V' \cdot U$ in the second order expansion shows that it is not possible to exchange $\hat{U}$ and $\hat{V}$.

Let us return to our original goal. The first vector field u moves the point in the state space from the initial condition x_0 to the point

$$x_1 = \exp(p\hat{u})Ix_0$$

in time p and then the second vector field v moves the point from x_1 to

$$x_2 = \exp(p\hat{v})Ix_1 = \exp(p\hat{v})I \circ \exp(p\hat{u})Ix_0$$

in another time p. We want to find the shortcut – the vector field w that brings the point in the state space from the initial condition x_0 to the final point x_2 in the same time i.e. $2p$

$$x_2 = \exp(2p\hat{w})Ix_0.$$

So we have the equation for $\hat{w}$

$$\exp(p\hat{v})I \circ \exp(p\hat{u})I = \exp(2p\hat{w})I.$$

Due to the rescaling (32-34) we can write

$$\exp(\hat{V})I \circ \exp(\hat{U})I = \exp(\hat{W})I.$$

Using (35) to the left hand side we get

$$\exp\hat{U}\exp\hat{V}I = \exp\hat{W}I.$$

Assuming the operators are applied to the identity map I only we want to solve

$$\exp\hat{U}\exp\hat{V} = \exp\hat{W} \qquad (36)$$

for $\hat{W}$.

coefficient $(b_r)_k$ in front of a commutator ending in $\ldots, \hat{V}, \hat{U}$. In this way we reduce the number of unknowns to one half once more to get $2^r/4$ unknowns.

We can look at the system of linear algebraic equations as of finding the sparsest linear combination of columns of the matrix to get the right hand side vector. Thus we can remove any column that is a multiple of another column and we put the unknown corresponding to the removed column to zero. Unfortunately we cannot remove a column that is a linear combination of other columns because then the solution to the new system might have worse sparsity (more nonzero components). But we can remove any row (any equation) that is a linear combination of other rows.

Performing these reductions along with the Gauss elimination the system can be reduced considerably ending with m rows (equations) and n unknowns. After this reduction we try to find the sparsest solution.

First, we test whether there is a solution containing only one nonzero component. To do this we remove all but one columns of the matrix M_r^T. This can be done in n ways. We try to solve this system in all these n cases. If the rank of the matrix of coefficients is equal to the rank of the augmented matrix (the matrix with the right hand side column added) then the system has solution (according to the Frobenius theorem), if the ranks do not agree, the system has no solution. If we find a solution we are done.

If not, we test whether there is a solution containing two nonzero components. To do this we remove all but two columns of the matrix M_r^T. This can be done in

$$\binom{n}{2}$$

ways. In all these cases we again test whether there is a solution using the Frobenius theorem. If so, we are done.

If not, we test whether there is a solution containing three nonzero components. To do this we remove all but three columns of the matrix M_r^T. This can be done in

$$\binom{n}{3}$$

ways. In all these cases we again test whether there is a solution using the Frobenius theorem. If so, we are done. If not we test whether there is a solution containing 4, then 5, then 6 and so on solutions. This algorithm stops after a finite (though possibly large) number of steps.

This proves the first statement. And we can also write

$$\exp \hat{V} \exp \hat{U} \exp[\hat{U}, \hat{V}] = \exp(\hat{U} + \hat{V}) \exp(-\frac{1}{2}[\hat{U}, \hat{V}]) \exp[\hat{U}, \hat{V}] =$$

$$= \exp(\hat{U} + \hat{V}) \exp(\frac{1}{2}[\hat{U}, \hat{V}]).$$

This proves the second statement.

10.2 Disprove of Kolsrud conjecture

Kolsrud in [13] computes the sparsest form of the BCH series up to the order 9 by intuitive applications of simplification rules. He observes that for the order r between 1 and 9 if the order r is a prime number then the minimal number of commutators to express the terms of the BCH series of order r is equal to the rank of the matrix M_r (in our symbols). Unfortunately, we have found that for $r = 11$ which is prime, there is a solution having 181 terms (see Table 1). This number may not be the least possible value, but it is definitely less than the rank of the matrix M_r which is 186 for $r = 11$. This result disproves the Kolsrud conjecture.

In the next subsection we describe our systematic algorithm to find the sparsest form of the BCH series.

10.3 Our algorithm to find the sparsest solution

We use a systematic approach to find the sparsest solution to (49), when the right hand side $\vec{a}_r$ is already known. First, we reduce the size of the problem. For the order r there are originally 2^r equations for 2^r unknowns. Since

$$[\hat{U}, \hat{U}] = 0$$

any commutator ending in two $\hat{U}$'s is equal to zero (and the same is true for $\hat{V}$) and we can put to zero each coefficient $(b_r)_k$ in front of a commutator ending in two $\hat{U}$'s or in two $\hat{V}$'s. In this way we reduce the number of unknowns to one half to get $2^r/2$ unknowns.

Further, since

$$[\hat{U}, \hat{V}] = -[\hat{V}, \hat{U}]$$

any commutator ending in $\ldots, \hat{V}, \hat{U}$ is equal to minus one times the same commutator with the last two symbols exchanged. Thus we can put to zero each

6.2 Non-identity map

Note however, that

$$\hat{a}I = \hat{b}I$$

does not imply

$$\hat{a}\phi = \hat{b}\phi \qquad \text{for all } \phi \in \Phi$$

or equivalently that

$$\hat{a}I = 0$$

does not imply

$$\hat{a}\phi = 0 \qquad \text{for all } \phi \in \Phi.$$

To see this, consider the second order operator

$$\hat{a} = \hat{u}\hat{k}$$

where $\hat{k}$ is a first order operator corresponding to a constant vector field k

$$k(x) = \text{const.}$$

and $\hat{u}$ is another first order operator corresponding to a vector field u. Then

$$\hat{a}\phi = \hat{u}\hat{k}\phi = (\phi' \cdot k)' \cdot u = \phi'' \cdot k \cdot u.$$

Here $\phi'' \cdot k \cdot u$ is a map which in the point $x \in R^n$ has the value

$$\hat{a}\phi(x) = \phi''(x) \cdot k(x) \cdot u(x) = \phi''(x) \cdot \text{ const. } \cdot u(x).$$

Then $\hat{a}\phi$ is nonzero for quadratic ϕ while zero for linear ϕ and thus for identity. So we have

$$\hat{a}I = 0$$

while nonzero result for nonlinear ϕ

$$\hat{a}\phi = \phi'' \cdot k \cdot u.$$

7 BCH series

Using the series (15) for the exponential of an operator and using the series

$$\log \hat{u} = \sum_{k=1}^{\infty} (-1)^{k-1} (\hat{u} - \hat{\mathcal{I}})^k / k \tag{37}$$

for the logarithm (assuming the series converges in accord with [19], [10] etc.) we get the solution of (36) for W in the form of the Hausdorff series [2]

$$\hat{W} = \log(\exp \hat{U} \exp \hat{V}). \tag{38}$$

In the next section we present a detailed derivation of a few low order terms of this series.

Reinsch in [25] uses an interesting alternative approach to get the terms in the Baker–Campbell–Hausdorff series, see Appendix A.2.

7.1 Second order expansion

The second order expansion in p gives

$$\exp \hat{U} = \hat{\mathcal{I}} + \hat{U} + \frac{1}{2}\hat{U}^2 + \mathcal{O}(p^3)$$

$$\exp \hat{V} = \hat{\mathcal{I}} + \hat{V} + \frac{1}{2}\hat{V}^2 + \mathcal{O}(p^3).$$

Denoting

$$\hat{Z} = \exp \hat{U} \exp \hat{V} - \hat{\mathcal{I}} =$$
$$= (\hat{\mathcal{I}} + \hat{U} + \frac{1}{2}\hat{U}^2)(\hat{\mathcal{I}} + \hat{V} + \frac{1}{2}\hat{V}^2) - \hat{\mathcal{I}} + \mathcal{O}(p^3) =$$
$$= \hat{U} + \hat{V} + \hat{U}\hat{V} + \frac{1}{2}\hat{U}^2 + \frac{1}{2}\hat{V}^2 + \mathcal{O}(p^3)$$

we have

$$\hat{Z}^2 = (\hat{U} + \hat{V} + \hat{U}\hat{V} + \frac{1}{2}\hat{U}^2 + \frac{1}{2}\hat{V}^2)^2 + \mathcal{O}(p^3) =$$
$$= \hat{U}^2 + \hat{U}\hat{V} + \hat{V}\hat{U} + \hat{V}^2 + \mathcal{O}(p^3)$$

and

$$\hat{W} = \log(\exp \hat{U} \exp \hat{V}) = \hat{Z} - \frac{1}{2}\hat{Z}^2 + \mathcal{O}(p^3) =$$

infinitely many solutions. We can search for the sparsest solution, i.e. the solution that has the least possible number of nonzero elements.

The sparsest solution of an under-determined system is not only elegant, it is also practical, because it contains the information in the most compressed way. This is desirable when we want to store or to transmit the information. This problem has been discussed intensively recently, see [6] and the references there.

10.1 Finite forms of BCH series

In this section we present two special cases when the BCH series has only a finite number of terms.

If $\hat{U}$ and $\hat{V}$ commute, i.e. if

$$[\hat{U}, \hat{V}] = 0$$

then

$$\hat{W} = \log(\exp \hat{U} \exp \hat{V})$$

has finitely many terms, namely

$$\hat{W} = \hat{U} + \hat{V}.$$

This is also the case when $\hat{U}$ and $\hat{V}$ are replaced by (complex) numbers.

Another special case of finite form of the BCH series is this one: If

$$[\hat{U}, \hat{V}] \neq 0$$
$$[\hat{U}, \hat{U}, \hat{V}] = 0$$
$$[\hat{V}, \hat{U}, \hat{V}] = 0$$

then

$$\hat{W} = \hat{U} + \hat{V} + \frac{1}{2}[\hat{U}, \hat{V}].$$

As a consequence we can write

$$\exp \hat{U} \exp \hat{V} = \exp(\hat{U} + \hat{V}) \exp(\frac{1}{2}[\hat{U}, \hat{V}]) = \exp \hat{V} \exp \hat{U} \exp[\hat{U}, \hat{V}].$$

Proof: since $[\hat{U}, \hat{V}]$ commutes with both $\hat{U}$ and $\hat{V}$ and thus also with $(\hat{U} + \hat{V})$ (even though $\hat{U}$ and $\hat{V}$ do not commute) we can write

$$\exp \hat{U} \exp \hat{V} = \exp(\hat{U} + \hat{V} + \frac{1}{2}[\hat{U}, \hat{V}]) = \exp(\hat{U} + \hat{V}) \exp(\frac{1}{2}[\hat{U}, \hat{V}]).$$

of order r in the BCH series is

$$N_w(r) = N_w(r-1).$$

- Conjecture 2:
  If the order r is even and more than 2 then the number $N_w(r)$ of words of order r in the BCH series is

$$N_w(r) = N_w(r-1).$$

- Observation 3:
  If the order r is between 2 and 12 then the minimal number $N_{com}(r)$ of commutators of order r is strictly less than then the number $N_w(r)$ of words of order r in the BCH series.

  Thus it is more convenient to write the BCH series in commutators than in words, because then the results are shorter. There is one even more important reason to write the terms using commutators: a commutator of operators corresponding to vector fields corresponds to a vector field as well.

- Conjecture 3:
  If the order r is more than 1 then the minimal number $N_{com}(r)$ of commutators of order r is strictly less than then the number $N_w(r)$ of words of order r in the BCH series.

- Observation 4:
  For certain r (e.g. 2, 3, 5, 7, 11) the number N_w of words reaches a remarkably high value. Considering that two words, namely the word consisting of $\hat{U}$'s only and the word consisting of $\hat{V}$'s only cannot be generated as a linear combination of commutators, the value $N_w = N_{tot} - 2$ is the highest possible result consistent with the theory.

10 Sparsest form of BCH series

Using (38) with exp given by (15) and log given by (37) we get the series to solve (36) in words. Then we can solve (49) to get the series in commutators. As mentioned earlier, the system of linear algebraic equations (49) is under-determined, it has more unknowns (the coefficients $(b_r)_k$) than independent equations. It has

$$= \hat{U} + \hat{V} + \hat{U}\hat{V} + \frac{1}{2}\hat{U}^2 + \frac{1}{2}\hat{V}^2 - \frac{1}{2}(\hat{U}^2 + \hat{U}\hat{V} + \hat{V}\hat{U} + \hat{V}^2) + \mathcal{O}(p^3) =$$

$$= \hat{U} + \hat{V} + \frac{1}{2}(\hat{U}\hat{V} - \hat{V}\hat{U}) + \mathcal{O}(p^3).$$

7.2 Commutator

Denoting the **commutator** of two operators

$$[\hat{U}, \hat{V}] = \hat{U}\hat{V} - \hat{V}\hat{U} \tag{39}$$

we get

$$\hat{W} = \hat{U} + \hat{V} + \frac{1}{2}[\hat{U}, \hat{V}] + \mathcal{O}(p^3). \tag{40}$$

We say that $\hat{U}$ and $\hat{V}$ commute if $[\hat{U}, \hat{V}] = 0$.

Note that

$$[\hat{U}, \hat{V}] + [\hat{V}, \hat{U}] = 0 \tag{41}$$

for any two operators. This property is called **skew symmetry**.

Going back from $\hat{U}$, $\hat{V}$ and $\hat{W}$ to $\hat{u}$, $\hat{v}$ and $\hat{w}$ using (29-31) we get

$$\hat{w} = \frac{1}{2p}\hat{W} = \frac{1}{2}(\hat{u} + \hat{v} + \frac{p}{2}[\hat{u}, \hat{v}] + \mathcal{O}(p^2)).$$

Our ultimate goal is not the operator $\hat{w}$ but rather the vector field

$$w = \hat{w}I,$$

i.e. the shortcut from the initial point x_0 to the final point x_2 of the zig+zag system. To this purpose we must apply the commutator $[\hat{u}, \hat{v}]$ to the identity map I

$$[\hat{u}, \hat{v}]I = \hat{u}\hat{v}I - \hat{v}\hat{u}I = v' \cdot u - u' \cdot v.$$

Thus a commutator of two vector fields is defined as

$$[u, v] = v' \cdot u - u' \cdot v$$

and in coordinates

$$([u, v](x))_j = (([\hat{u}, \hat{v}]I)(x))_j = \sum_{k=1}^{n} \frac{\partial v_j(x)}{\partial x_k} u_k(x) - \frac{\partial u_j(x)}{\partial x_k} v_k(x).$$

To avoid confusion, if $\hat{u}$ and $\hat{v}$ are operators, so is their commutator $[\hat{u}, \hat{v}]$. Then $[\hat{u}, \hat{v}]I$ is this operator applied to the identity map from R^n to R^n, the result being another map from R^n to R^n. Then $([\hat{u}, \hat{v}]I)(x)$ is this map evaluated in the point x, the result being a point in R^n. And finally $(([\hat{u}, \hat{v}]I)(x))_j$ is the j-th coordinate of this point.

Then

$$w = \hat{w}I = \frac{u+v}{2} + p\frac{v' \cdot u - u' \cdot v}{4} + \mathcal{O}(p^2). \qquad (42)$$

This agrees with (22) completely. It is always a pleasure to get the same result by two different ways.

7.3 Commuting vector fields

We say that two vector fields u, v commute if

$$[u, v] = 0.$$

This occurs if and only if their corresponding operators $\hat{u}$, $\hat{v}$ commute. In one dimension it is easy to find all the vector fields that commute with a given vector field $u : R \to R$. The condition

$$v'(x)u(x) - u'(x)v(x) = 0$$

is a linear differential equation for the function $v(x)$ that can be solved by separation of variables to give

$$v(x) = k\, u(x) \qquad k \in R.$$

Thus a one dimensional vector field commutes with its multiples only. Alternatively, this can be shown by considering, for nonzero f

$$(\frac{g}{f})' = \frac{g'f - f'g}{f^2}$$

and thus

$$[f, g] = f^2(\frac{g}{f})'.$$

Then $[f, g] = 0$ if

$$(\frac{g}{f})' = 0$$

and this is for

$$\frac{g}{f} = k.$$

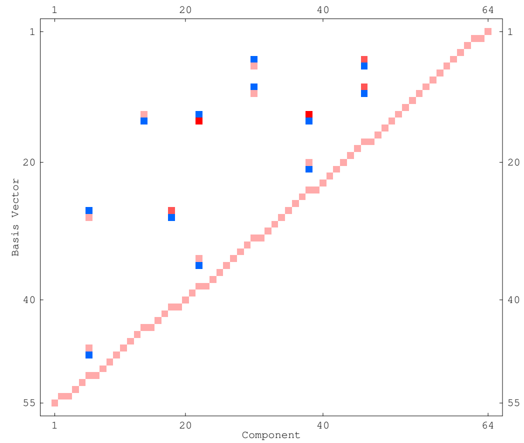

Figure 2: One possible basis of the kernel of the matrix M_6^T. The matrix M_6^T is a 64×64 matrix. Its rank is 9, so the dimension of the kernel is 55. Each vector in the kernel has 64 components. Thus there are 55 rows each consisting of 64 squares in this picture. Each row depicts one vector in the basis of the kernel. Red squares represent positive components while blue squares represent negative ones. The saturation of the color shows the absolute value of the corresponding element. Zero components are shown as blank squares (most of the components are zero).

- Observation 1:
  If the order r is odd and between 1 and 11 then the number $N_w(r)$ of words of order r in the BCH series as a function of r forms a strictly increasing sequence 2, 6, 30, 126, 390, 2046.

- Conjecture 1:
  If the order r is odd then the number $N_w(r)$ of words of order r in the BCH series as a function of r forms a strictly increasing sequence.

- Observation 2:
  If the order r is even and between 4 and 12 then the number $N_w(r)$ of words

$$= -\hat{U}\hat{U}\hat{V}\hat{V} + 2\hat{U}\hat{V}\hat{U}\hat{V} - 2\hat{V}\hat{U}\hat{V}\hat{U} + \hat{V}\hat{V}\hat{U}\hat{U}$$

and similarly for

$$[\hat{V}, \hat{U}, \hat{U}, \hat{V}] = \ldots = -\hat{U}\hat{U}\hat{V}\hat{V} + 2\hat{U}\hat{V}\hat{U}\hat{V} - 2\hat{V}\hat{U}\hat{V}\hat{U} + \hat{V}\hat{V}\hat{U}\hat{U}.$$

As a consequence both

$$\hat{W}_4 = -\frac{1}{24}([\hat{U}, \hat{V}, \hat{U}, \hat{V}])$$

and

$$\hat{W}_4 = -\frac{1}{24}([\hat{V}, \hat{U}, \hat{U}, \hat{V}])$$

are correct. This changes the form of the result but not its sparsity. Each form has exactly one term.

As an illustration in Fig. 2 we present a basis of the kernel of M_6^T shown as a 55×64 matrix. Here positive elements are shown as red squares, negative elements are shown as blue squares (we adopt this convention from electronics) and zero elements as blank squares. The higher the absolute value of an element, the higher the saturation of the color. As an example, the 13-th vector in this basis is

$$(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, -3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,$$

$$0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)^T$$

corresponding to the identity

$$[\hat{U}, \hat{U}, \hat{V}, \hat{V}, \hat{U}, \hat{V}] - 3[\hat{U}, \hat{V}, \hat{U}, \hat{V}, \hat{U}, \hat{V}] +$$

$$+3[\hat{V}, \hat{U}, \hat{U}, \hat{V}, \hat{U}, \hat{V}] + [\hat{V}, \hat{V}, \hat{U}, \hat{U}, \hat{V}, \hat{U}] = 0.$$

Note that there is just one minus sign and three plus signs in this identity, as seen also in Fig. 2 where the 13-th row contains one blue square and 3 red squares.

9 Our observations and conjectures about BCH series

Based on results given in Tab. 1 we formulate the following observations and conjectures:

7.4 Higher order commutators

We can generalize the notation used for a commutator of two operators $\hat{A}_1$, $\hat{A}_2$

$$[\hat{A}_1, \hat{A}_2] = \hat{A}_1\hat{A}_2 - \hat{A}_2\hat{A}_1$$

to commutators of three or more operators. For three operators we have

$$[\hat{A}_1, \hat{A}_2, \hat{A}_3] \stackrel{def}{=} [\hat{A}_1, [\hat{A}_2, \hat{A}_3]] = \hat{A}_1[\hat{A}_2, \hat{A}_3] - [\hat{A}_2, \hat{A}_3]\hat{A}_1 =$$
$$= \hat{A}_1(\hat{A}_2\hat{A}_3 - \hat{A}_3\hat{A}_2) - (\hat{A}_2\hat{A}_3 - \hat{A}_3\hat{A}_2)\hat{A}_1 =$$
$$= \hat{A}_1\hat{A}_2\hat{A}_3 - \hat{A}_1\hat{A}_3\hat{A}_2 - \hat{A}_2\hat{A}_3\hat{A}_1 + \hat{A}_3\hat{A}_2\hat{A}_1$$

Note that

$$[\hat{A}_1, \hat{A}_2, \hat{A}_3] + [\hat{A}_2, \hat{A}_3, \hat{A}_1] + [\hat{A}_3, \hat{A}_1, \hat{A}_2] = 0 \qquad (43)$$

for any three operators. This property is called **Jacobi identity**.

Similarly, we can define a commutator of order r (i.e. a commutator of r operators) for any positive integer r by

$$[\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_r] = [\hat{A}_1, [\hat{A}_2, [\ldots, \hat{A}_r]]] \qquad (44)$$

when we add for the sake of completeness

$$[\hat{A}] = \hat{A}.$$

Note that a commutator of order r is linear in each one of its arguments.

When each $\hat{A}_i$ is either $\hat{U}$ or $\hat{V}$ (the two letters in a two-letter alphabet) then we can express each commutator $[\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_r]$ of order r as a certain linear combination of 2^r words (ordered n-tuples of elements that are either the symbol $\hat{U}$ or the symbol $\hat{V}$) consisting of r letters. These linear combinations can be conveniently express using a square $2^r \times 2^r$ matrix M_r. We will need these matrices later, so we give examples of M_1, M_2 and M_3 and then we derive a recurrent relation for them.

For $r = 1$ we have

$$\begin{array}{rcl} [\hat{U}] & = & \hat{U} \\ [\hat{V}] & = & \hat{V} \end{array}$$

and in the matrix form

$$\begin{pmatrix} [\hat{U}] \\ [\hat{V}] \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \hat{U} \\ \hat{V} \end{pmatrix} = M_1 \cdot \begin{pmatrix} \hat{U} \\ \hat{V} \end{pmatrix}.$$

For $r = 2$ we have

$$\begin{aligned}
[\hat{U},\hat{U}] &= 0 \\
[\hat{U},\hat{V}] &= \hat{U}\hat{V} - \hat{V}\hat{U} \\
[\hat{V},\hat{U}] &= \hat{V}\hat{U} - \hat{U}\hat{V} \\
[\hat{V},\hat{V}] &= 0
\end{aligned}$$

and in the matrix form

$$\begin{pmatrix} [\hat{U},\hat{U}] \\ [\hat{U},\hat{V}] \\ [\hat{V},\hat{U}] \\ [\hat{V},\hat{V}] \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \hat{U}\hat{U} \\ \hat{U}\hat{V} \\ \hat{V}\hat{U} \\ \hat{V}\hat{V} \end{pmatrix} = M_2 \cdot \begin{pmatrix} \hat{U}\hat{U} \\ \hat{U}\hat{V} \\ \hat{V}\hat{U} \\ \hat{V}\hat{V} \end{pmatrix}.$$

For $r = 3$ we have

$$\begin{aligned}
[\hat{U},\hat{U},\hat{U}] &= 0 \\
[\hat{U},\hat{U},\hat{V}] &= \hat{U}\hat{U}\hat{V} - 2\hat{U}\hat{V}\hat{U} + \hat{V}\hat{U}\hat{U} \\
[\hat{U},\hat{V},\hat{U}] &= -\hat{U}\hat{U}\hat{V} + 2\hat{U}\hat{V}\hat{U} - \hat{V}\hat{U}\hat{U} \\
[\hat{U},\hat{V},\hat{V}] &= 0 \\
[\hat{V},\hat{U},\hat{U}] &= 0 \\
[\hat{V},\hat{U},\hat{V}] &= -\hat{U}\hat{V}\hat{V} + 2\hat{V}\hat{U}\hat{V} - \hat{V}\hat{V}\hat{U} \\
[\hat{V},\hat{V},\hat{U}] &= \hat{U}\hat{V}\hat{V} - 2\hat{V}\hat{U}\hat{V} + \hat{V}\hat{V}\hat{U} \\
[\hat{V},\hat{V},\hat{V}] &= 0
\end{aligned}$$

and in the matrix form

$$\begin{pmatrix} [\hat{U},\hat{U},\hat{U}] \\ [\hat{U},\hat{U},\hat{V}] \\ [\hat{U},\hat{V},\hat{U}] \\ [\hat{U},\hat{V},\hat{V}] \\ [\hat{V},\hat{U},\hat{U}] \\ [\hat{V},\hat{U},\hat{V}] \\ [\hat{V},\hat{V},\hat{U}] \\ [\hat{V},\hat{V},\hat{V}] \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 2 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} \hat{U}\hat{U}\hat{U} \\ \hat{U}\hat{U}\hat{V} \\ \hat{U}\hat{V}\hat{U} \\ \hat{U}\hat{V}\hat{V} \\ \hat{V}\hat{U}\hat{U} \\ \hat{V}\hat{U}\hat{V} \\ \hat{V}\hat{V}\hat{U} \\ \hat{V}\hat{V}\hat{V} \end{pmatrix} =$$

r	N_{rules}	L_{max}
1	0	
2	3	2
3	6	2
4	13	2
5	26	2
6	55	4
7	110	4
8	226	8
9	456	15
10	925	26
11	1862	43
12	3761	76
⋮	⋮	⋮

Table 3: This table shows for each order r between 1 and 12 the number N_{rules} of independent rules (identities) of order r (which is the dimension of the kernel of the matrix M_r^T, cf. Table 1) and the maximal length L_{max} of a non-decomposable rule of order r. The field of L_{max} for $r = 1$ is blank because there are no rules for $r = 1$, the kernel of M_1 is just a zero vector.

which corresponds to the identity

$$(\hat{c}_4)_6 - (\hat{c}_4)_{10} = 0$$

i.e.

$$[\hat{U},\hat{V},\hat{U},\hat{V}] - [\hat{V},\hat{U},\hat{U},\hat{V}] = 0.$$

I expected the plus sign instead of the minus sign, in analogy with

$$[\hat{U},\hat{V}] + [\hat{V},\hat{U}] = 0.$$

But the above identity is indeed correct, as can be shown by direct evaluation of

$$[\hat{U},\hat{V},\hat{U},\hat{V}] = [\hat{U},[\hat{V},[\hat{U},\hat{V}]]] = [\hat{U},[\hat{V},\hat{U}\hat{V} - \hat{V}\hat{U}]] =$$

$$= [\hat{U},\hat{V}\hat{U}\hat{V} - \hat{V}\hat{V}\hat{U} - \hat{U}\hat{V}\hat{V} + \hat{V}\hat{U}\hat{V}] =$$

$$= \hat{U}\hat{V}\hat{U}\hat{V} - \hat{U}\hat{V}\hat{V}\hat{U} - \hat{U}\hat{U}\hat{V}\hat{V} + \hat{U}\hat{V}\hat{U}\hat{V} - \hat{V}\hat{U}\hat{V}\hat{U} + \hat{V}\hat{V}\hat{U}\hat{U} + \hat{U}\hat{V}\hat{V}\hat{U} - \hat{V}\hat{U}\hat{V}\hat{U} =$$

8 Identities

For any order r each vector in the kernel of the matrix M_r^T corresponds to a certain identity. Three vectors in the kernel of M_2^T given in (50) correspond to three identities

$$(\hat{c}_2)_1 = [\hat{U}, \hat{U}] = 0$$
$$(\hat{c}_2)_2 + (\hat{c}_2)_3 = [\hat{U}, \hat{V}] + [\hat{V}, \hat{U}] = 0$$

and

$$(\hat{c}_2)_4 = [\hat{V}, \hat{V}] = 0.$$

This is a direct consequence of the skew symmetry of the commutator.

Oteo in [20] finds some of these identities by comparison of the BCH series computed by various methods. Our approach is more systematic. Based on the kernel of the matrices M_r we are able to give both the total number and the form of these identities.

Each such identity says that a certain linear combination of a certain number (we call this number the length L of the identity) of commutators is equal to zero. Table 3 shows the number of independent identities and the maximal length of an non-decomposable identity for order r between 1 and 12. We say that an identity is non-decomposable if none of its proper parts is an identity itself. E.g. the identity

$$[\hat{U}, \hat{V}] + [\hat{V}, \hat{U}] = 0$$

is non-decomposable, while the identity

$$[\hat{U}, \hat{U}] + [\hat{V}, \hat{V}] = 0$$

is decomposable, because it can be decomposed into two identities, namely

$$[\hat{U}, \hat{U}] = 0$$

and

$$[\hat{V}, \hat{V}] = 0.$$

For higher order r it may be less obvious whether two forms of the result are equal or not.

It was a surprise for me that the kernel of the matrix M_4^T contains the vector

$$(0, 0, 0, 0, 0, 1, 0, 0, 0, -1, 0, 0, 0, 0, 0, 0)$$

$$= M_3 \cdot \begin{pmatrix} \hat{U}\hat{U}\hat{U} \\ \hat{U}\hat{U}\hat{V} \\ \hat{U}\hat{V}\hat{U} \\ \hat{U}\hat{V}\hat{V} \\ \hat{V}\hat{U}\hat{U} \\ \hat{V}\hat{U}\hat{V} \\ \hat{V}\hat{V}\hat{U} \\ \hat{V}\hat{V}\hat{V} \end{pmatrix}.$$

In general

$$(\hat{c}_r)_k = \sum_{j=1}^{2^r} (M_r)_{k,j} \, (\hat{w}_r)_j \qquad (45)$$

where $(\hat{c}_r)_k$ is the k-th commutator of order r, and $(\hat{w}_r)_j$ is the j-th word consisting of r letters.

To work with the matrices M_r it is convenient to derive recurrent relations for them. The rule

$$[\hat{A}, \hat{X}] = \hat{A}\hat{X} - \hat{X}\hat{A}$$

where $\hat{A}$ is either $\hat{U}$ or $\hat{V}$ and $\hat{X}$ is a commutator corresponding to a row of the matrix M_r, implies that each row in the upper half of the matrix M_{r+1} (i.e. a row corresponding to a commutator beginning with $\hat{U}$) is a difference of two rows:

- a row of matrix M_r corresponding to the operator $\hat{X}$ padded by zeroes on the right and
- this row with zeroes inserted to even positions

and that each row in the bottom half of the matrix M_{r+1} (i.e. a row corresponding to a commutator beginning with $\hat{V}$) is a difference of two rows:

- a row of matrix M_r corresponding to the operator $\hat{X}$ padded by zeroes on the left and
- this row with zeroes inserted to odd positions

namely,

$$
(M_{r+1})_{i,j} = \begin{cases}
(M_r)_{i,j} - (M_r)_{i,\frac{j+1}{2}} & \text{for } i \leq 2^r,\ j \leq 2^r, j \text{ odd} \\
(M_r)_{i,j} & \text{for } i \leq 2^r,\ j \leq 2^r, j \text{ even} \\
- (M_r)_{i,\frac{j+1}{2}} & \text{for } i \leq 2^r,\ j > 2^r, j \text{ odd} \\
0 & \text{for } i \leq 2^r,\ j > 2^r, j \text{ even} \\
0 & \text{for } i > 2^r,\ j \leq 2^r, j \text{ odd} \\
- (M_r)_{i-2^r,\frac{j-2^r}{2}} & \text{for } i > 2^r,\ j \leq 2^r, j \text{ even} \\
(M_r)_{i-2^r,j-2^r} & \text{for } i > 2^r,\ j > 2^r, j \text{ odd} \\
(M_r)_{i-2^r,j-2^r} - (M_r)_{i-2^r,\frac{j-2^r}{2}} & \text{for } i > 2^r,\ j > 2^r, j \text{ even}
\end{cases} \tag{46}
$$

This can be conveniently computed by the following *Mathematica* code:

```
m={{1,0},{0,1}};
For[r=2,r<=3,r++,
 n=2^(r-1);
 m=
 Table[PadRight[m[[i]],2n]-Insert[m[[i]],0,Table[{j+1},{j,n}]],{i,n}]
 ~Join~
 Table[PadLeft[m[[i]],2n]-Insert[m[[i]],0,Table[{j},{j,n}]],{i,n}];
 Print["r=",r," m=",MatrixForm[m]];
];
```

7.5 Higher order expansion in words

The number N_w of words in

$$\hat{W} = \log(\exp \hat{U} \exp \hat{V})$$

of order r is too large to be listed here even for moderately large r, see Tab. 1. To give a picture of the series we present the expansion up to order 6 only

$$\hat{W} = \hat{W}_1 + \hat{W}_2 + \hat{W}_3 + \hat{W}_4 + \hat{W}_5 + \hat{W}_6 + \mathcal{O}(p^7)$$

where

$$
\begin{aligned}
\hat{W}_1 &= \hat{U} + \hat{V} \\
\hat{W}_2 &= \frac{1}{2}(\hat{U}\hat{V} - \hat{V}\hat{U}) \\
\hat{W}_3 &= \frac{1}{12}(\hat{U}\hat{U}\hat{V} - 2\hat{U}\hat{V}\hat{U} + \hat{U}\hat{V}\hat{V} + \hat{V}\hat{U}\hat{U} - 2\hat{V}\hat{U}\hat{V} + \hat{V}\hat{V}\hat{U}) \\
\hat{W}_4 &= \frac{1}{24}(\hat{U}\hat{U}\hat{V}\hat{V} - 2\hat{U}\hat{V}\hat{U}\hat{V} + 2\hat{V}\hat{U}\hat{V}\hat{U} - \hat{V}\hat{V}\hat{U}\hat{U})
\end{aligned}
$$

plus the general solution of the corresponding homogeneous equation. In other words we can add any vector from the kernel of the matrix M_r^T (i.e. a vector that is mapped to zero) and we get another solution of the linear system. But this another solution corresponds to the same term in the BCH series.

To illustrate this, let us consider $r = 2$. One possible basis of the kernel of M_r^T is

$$(1,0,0,0)^T, \quad (0,1,1,0)^T, \quad (0,0,0,1)^T. \tag{50}$$

Then the general solution to (49) is

$$\vec{b}_2 = \begin{pmatrix} 0 \\ \frac{1}{2} \\ 0 \\ 0 \end{pmatrix} + k_1 \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} + k_2 \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} + k_3 \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix},$$

with $k_1, k_2, k_3 \in R$. The corresponding second order term in the BCH series is

$$
\begin{aligned}
\hat{W}_2 &= \frac{1}{2}(\hat{c}_2)_2 + k_1(\hat{c}_2)_1 + k_2((\hat{c}_2)_2 + (\hat{c}_2)_3) + k_3(\hat{c}_2)_4 = \\
&= \frac{1}{2}[\hat{U},\hat{V}] + k_1[\hat{U},\hat{U}] + k_2([\hat{U},\hat{V}] + [\hat{V},\hat{U}]) + k_3[\hat{V},\hat{V}].
\end{aligned}
$$

In this case (for $r = 2$) it is trivial to see that the new terms add nothing new but they may change the form of the result. As an example consider

$$
\begin{aligned}
k_1 &= 0, \\
k_2 &= -\frac{1}{2}, \\
k_3 &= 0.
\end{aligned}
$$

Then we get

$$\hat{W}_2 = -\frac{1}{2}[\hat{V},\hat{U}]$$

which is equal to the result given above, just in a different form. The form of the result may be changed in two ways:

- one term may be replaced by another, e.g. $[\hat{U},\hat{V}]$ may be replaced by $-[\hat{V},\hat{U}]$, or
- the number of terms may be changed.

We want to find the **sparsest** solution to (49), i.e. the solution that has the least possible number of nonzero terms.

this linear equation has not a unique solution. The matrix M_r is singular, its rank is less than its size 2^r, see Tab. 1. But the solution exists. As an illustration we present the results of order up to eight here

$$\begin{aligned}
\hat{W}_5 &= \frac{1}{6!}(-[\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}] - 6[\hat{U},\hat{U},\hat{V},\hat{U},\hat{V}] - 2[\hat{U},\hat{V},\hat{V},\hat{U},\hat{V}] + \\
&\quad +2[\hat{V},\hat{U},\hat{U},\hat{U},\hat{V}] + 6[\hat{V},\hat{U},\hat{V},\hat{U},\hat{V}] + [\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}]) \\
\hat{W}_6 &= \frac{1}{2\cdot 6!}([\hat{U},\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}] + [\hat{V},\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}] + \\
&\quad +6[\hat{V},\hat{U},\hat{U},\hat{V},\hat{U},\hat{V}] - 2[\hat{V},\hat{V},\hat{U},\hat{U},\hat{U},\hat{V}]) \\
\hat{W}_7 &= \frac{1}{6\cdot 7!}([\hat{U},\hat{U},\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}] - 3[\hat{U},\hat{U},\hat{U},\hat{U},\hat{V},\hat{U},\hat{V}] - \\
&\quad -4[\hat{U},\hat{U},\hat{U},\hat{V},\hat{V},\hat{U},\hat{V}] + 12[\hat{U},\hat{U},\hat{V},\hat{U},\hat{U},\hat{U},\hat{V}] + \\
&\quad +9[\hat{U},\hat{U},\hat{V},\hat{U},\hat{V},\hat{U},\hat{V}] - 9[\hat{U},\hat{U},\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}] + \\
&\quad +18[\hat{U},\hat{V},\hat{U},\hat{U},\hat{V},\hat{U},\hat{V}] + 24[\hat{U},\hat{V},\hat{U},\hat{V},\hat{V},\hat{U},\hat{V}] + \\
&\quad +3[\hat{U},\hat{V},\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}] - 3[\hat{V},\hat{U},\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}] - \\
&\quad -9[\hat{V},\hat{U},\hat{U},\hat{U},\hat{V},\hat{U},\hat{V}] + 4[\hat{V},\hat{U},\hat{U},\hat{V},\hat{V},\hat{U},\hat{V}] - \\
&\quad -6[\hat{V},\hat{U},\hat{V},\hat{U},\hat{U},\hat{U},\hat{V}] - 30[\hat{V},\hat{U},\hat{V},\hat{U},\hat{V},\hat{U},\hat{V}] - \\
&\quad -3[\hat{V},\hat{U},\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}] + 3[\hat{V},\hat{V},\hat{U},\hat{U},\hat{V},\hat{U},\hat{V}] - \\
&\quad -6[\hat{V},\hat{V},\hat{U},\hat{V},\hat{V},\hat{U},\hat{V}] - [\hat{V},\hat{V},\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}]) \\
\hat{W}_8 &= \frac{1}{3\cdot 8!}(6[\hat{U},\hat{U},\hat{V},\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}] - 24[\hat{U},\hat{V},\hat{V},\hat{U},\hat{V},\hat{V},\hat{U},\hat{V}] + \\
&\quad +6[\hat{U},\hat{V},\hat{V},\hat{V},\hat{V},\hat{U},\hat{U},\hat{V}] - 2[\hat{U},\hat{V},\hat{V},\hat{V},\hat{V},\hat{V},\hat{U},\hat{V}] - \\
&\quad -2[\hat{V},\hat{U},\hat{U},\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}] - 12[\hat{V},\hat{U},\hat{U},\hat{V},\hat{U},\hat{U},\hat{U},\hat{V}] + \\
&\quad +6[\hat{V},\hat{U},\hat{U},\hat{V},\hat{U},\hat{V},\hat{U},\hat{V}] - 6[\hat{V},\hat{U},\hat{V},\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}] - \\
&\quad -60[\hat{V},\hat{U},\hat{V},\hat{U},\hat{U},\hat{V},\hat{U},\hat{V}] + 8[\hat{V},\hat{U},\hat{V},\hat{V},\hat{U},\hat{U},\hat{U},\hat{V}] + \\
&\quad +6[\hat{V},\hat{V},\hat{U},\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}] + 48[\hat{V},\hat{V},\hat{U},\hat{V},\hat{U},\hat{U},\hat{U},\hat{V}] - \\
&\quad -25[\hat{V},\hat{V},\hat{V},\hat{U},\hat{U},\hat{U},\hat{U},\hat{V}]).
\end{aligned}$$

7.7 Non-uniqueness of results in commutators

Unlike the results for $\hat{W}_r$ in words, the results for $\hat{W}_r$ in commutators are not unique with the only exception of $r = 1$. This is the consequence of the fact that the rank of the matrix M_r that appears in (49) is less than its size 2^r for $r > 1$.

The general solution to (49) can be written as a sum of a particular solution

$$\begin{aligned}
\hat{W}_5 &= \frac{1}{6!}(-\hat{U}\hat{U}\hat{U}\hat{U}\hat{V} + 4\hat{U}\hat{U}\hat{U}\hat{V}\hat{U} + 4\hat{U}\hat{U}\hat{U}\hat{V}\hat{V} - 6\hat{U}\hat{U}\hat{V}\hat{U}\hat{U} - \\
&\quad -6\hat{U}\hat{U}\hat{V}\hat{U}\hat{V} - 6\hat{U}\hat{U}\hat{V}\hat{V}\hat{U} + 4\hat{U}\hat{U}\hat{V}\hat{V}\hat{V} + 4\hat{U}\hat{V}\hat{U}\hat{U}\hat{U} - \\
&\quad -6\hat{U}\hat{V}\hat{U}\hat{U}\hat{V} + 24\hat{U}\hat{V}\hat{U}\hat{V}\hat{U} - 6\hat{U}\hat{V}\hat{U}\hat{V}\hat{V} - 6\hat{U}\hat{V}\hat{V}\hat{U}\hat{U} - \\
&\quad -6\hat{U}\hat{V}\hat{V}\hat{U}\hat{V} + 4\hat{U}\hat{V}\hat{V}\hat{V}\hat{U} - \hat{U}\hat{V}\hat{V}\hat{V}\hat{V} - \hat{V}\hat{U}\hat{U}\hat{U}\hat{U} + \\
&\quad +4\hat{V}\hat{U}\hat{U}\hat{U}\hat{V} - 6\hat{V}\hat{U}\hat{U}\hat{V}\hat{U} - 6\hat{V}\hat{U}\hat{U}\hat{V}\hat{V} - 6\hat{V}\hat{U}\hat{V}\hat{U}\hat{U} + \\
&\quad +24\hat{V}\hat{U}\hat{V}\hat{U}\hat{V} - 6\hat{V}\hat{U}\hat{V}\hat{V}\hat{U} + 4\hat{V}\hat{U}\hat{V}\hat{V}\hat{V} + 4\hat{V}\hat{V}\hat{U}\hat{U}\hat{U} - \\
&\quad -6\hat{V}\hat{V}\hat{U}\hat{U}\hat{V} - 6\hat{V}\hat{V}\hat{U}\hat{V}\hat{U} - 6\hat{V}\hat{V}\hat{U}\hat{V}\hat{V} + 4\hat{V}\hat{V}\hat{V}\hat{U}\hat{U} + \\
&\quad +4\hat{V}\hat{V}\hat{V}\hat{U}\hat{V} - \hat{V}\hat{V}\hat{V}\hat{V}\hat{U}) \\
\hat{W}_6 &= \frac{1}{2\cdot 6!}(-\hat{U}\hat{U}\hat{U}\hat{U}\hat{V}\hat{V} + 4\hat{U}\hat{U}\hat{U}\hat{V}\hat{U}\hat{V} + 4\hat{U}\hat{U}\hat{U}\hat{V}\hat{V}\hat{V} - 6\hat{U}\hat{U}\hat{V}\hat{U}\hat{U}\hat{V} - \\
&\quad -6\hat{U}\hat{U}\hat{V}\hat{U}\hat{V}\hat{V} - 6\hat{U}\hat{U}\hat{V}\hat{V}\hat{U}\hat{V} - \hat{U}\hat{U}\hat{V}\hat{V}\hat{V}\hat{V} + 4\hat{U}\hat{V}\hat{U}\hat{U}\hat{U}\hat{V} - \\
&\quad -6\hat{U}\hat{V}\hat{U}\hat{U}\hat{V}\hat{V} + 24\hat{U}\hat{V}\hat{U}\hat{V}\hat{U}\hat{V} + 4\hat{U}\hat{V}\hat{U}\hat{V}\hat{V}\hat{V} - 6\hat{U}\hat{V}\hat{V}\hat{U}\hat{U}\hat{V} - \\
&\quad -6\hat{U}\hat{V}\hat{V}\hat{U}\hat{V}\hat{V} + 4\hat{U}\hat{V}\hat{V}\hat{V}\hat{U}\hat{V} - 4\hat{V}\hat{U}\hat{U}\hat{U}\hat{V}\hat{U} + 6\hat{V}\hat{U}\hat{U}\hat{V}\hat{U}\hat{U} + \\
&\quad +6\hat{V}\hat{U}\hat{U}\hat{V}\hat{V}\hat{U} - 4\hat{V}\hat{U}\hat{V}\hat{U}\hat{U}\hat{U} - 24\hat{V}\hat{U}\hat{V}\hat{U}\hat{V}\hat{U} + 6\hat{V}\hat{U}\hat{V}\hat{V}\hat{U}\hat{U} - \\
&\quad -4\hat{V}\hat{U}\hat{V}\hat{V}\hat{V}\hat{U} + \hat{V}\hat{V}\hat{U}\hat{U}\hat{U}\hat{U} + 6\hat{V}\hat{V}\hat{U}\hat{U}\hat{V}\hat{U} + 6\hat{V}\hat{V}\hat{U}\hat{V}\hat{U}\hat{U} + \\
&\quad +6\hat{V}\hat{V}\hat{U}\hat{V}\hat{V}\hat{U} - 4\hat{V}\hat{V}\hat{V}\hat{U}\hat{U}\hat{U} - 4\hat{V}\hat{V}\hat{V}\hat{U}\hat{V}\hat{U} + \hat{V}\hat{V}\hat{V}\hat{V}\hat{U}\hat{U})
\end{aligned}$$

7.6 Higher order expansion in commutators

It is obvious that the second order term can be written using a commutator

$$\hat{W}_2 = \frac{1}{2}[\hat{U}, \hat{V}].$$

It is still quite easy to find that the third order and the fourth order terms can be written using commutators as well

$$\hat{W}_3 = \frac{1}{12}([\hat{U}, \hat{U}, \hat{V}] - [\hat{V}, \hat{U}, \hat{V}])$$

$$\hat{W}_4 = -\frac{1}{24}[\hat{U}, \hat{V}, \hat{U}, \hat{V}].$$

It is a nontrivial fact that all the terms of $\hat{W} = \log(\exp \hat{U} \exp \hat{V})$ can be expressed as linear combinations of commutators only. Seven different proofs of this fact are given in [1], [4], [5], [7], [9], [22], [30]. The main idea of the proof by Djokovic

r	N_{tot}	N_w	rank(M_r)	N_{com}	$d/r!$
1	2	2	2	2	1
2	4	2	1	1	1
3	8	6	2	2	2
4	16	4	3	1	1
5	32	30	6	6	6
6	64	28	9	4	2
7	128	126	18	18	6
8	256	124	30	13	3
9	512	390	56	38	10
10	1024	388	99	31	2
11	2048	2046	186	≤ 181	6
12	4096	2044	335	≤ 180	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 1: This table shows for each order r between 1 and 12 the total number $N_{tot} = 2^r$ of all possible words consisting of r letters ($\hat{U}$ or $\hat{V}$), the number N_w of words appearing in the r-th order terms of BCH series, the rank of the matrix M_r relating the commutators and the words consisting of r letters, the least possible number N_{com} of commutators in the r-th order terms of BCH, and the common denominator d (divided by factorial of r) of the terms of order r of the BCH series.

[5] is given in Appendix. The consequence of this is that $\log(\exp \hat{U} \exp \hat{V}) = \hat{U} + \hat{V}$ if and only if $\hat{U}$ and $\hat{V}$ commute. For a short introduction to the Lie theory see [10].

The formula giving

$$\hat{W} = \log(\exp \hat{U} \exp \hat{V})$$

in commutators of $\hat{U}$ and $\hat{V}$ is called the Baker – Campbell – Hausdorff formula. It is named after the British mathematician Henry Frederick Baker (1866-1956), after the Irish mathematician John Edward Campbell (1862-1924) and after the German mathematician Felix Hausdorff (1868-1942).

To find the explicit form of the r-th order term in the BCH series as a linear combination of commutators $(\hat{c}_r)_k$ with coefficients $(b_r)_k$ (to be found), namely

$$\hat{W}_r = \sum_{k=1}^{2^r} (b_r)_k \ (\hat{c}_r)_k \qquad (47)$$

words	commutators
$(\hat{w}_1)_1 = \hat{U}$	$(\hat{c}_1)_1 = [\hat{U}]$
$(\hat{w}_1)_2 = \hat{V}$	$(\hat{c}_1)_2 = [\hat{V}]$
$(\hat{w}_2)_1 = \hat{U}\hat{U}$	$(\hat{c}_2)_1 = [\hat{U}, \hat{U}]$
$(\hat{w}_2)_2 = \hat{U}\hat{V}$	$(\hat{c}_2)_2 = [\hat{U}, \hat{V}]$
$(\hat{w}_2)_3 = \hat{V}\hat{U}$	$(\hat{c}_2)_3 = [\hat{V}, \hat{U}]$
$(\hat{w}_2)_4 = \hat{V}\hat{V}$	$(\hat{c}_2)_4 = [\hat{V}, \hat{V}]$
$(\hat{w}_3)_1 = \hat{U}\hat{U}\hat{U}$	$(\hat{c}_3)_1 = [\hat{U}, \hat{U}, \hat{U}]$
$(\hat{w}_3)_2 = \hat{U}\hat{U}\hat{V}$	$(\hat{c}_3)_2 = [\hat{U}, \hat{U}, \hat{V}]$
$(\hat{w}_3)_3 = \hat{U}\hat{V}\hat{U}$	$(\hat{c}_3)_3 = [\hat{U}, \hat{V}, \hat{U}]$
$(\hat{w}_3)_4 = \hat{U}\hat{V}\hat{V}$	$(\hat{c}_3)_4 = [\hat{U}, \hat{V}, \hat{V}]$
$(\hat{w}_3)_5 = \hat{V}\hat{U}\hat{U}$	$(\hat{c}_3)_5 = [\hat{V}, \hat{U}, \hat{U}]$
$(\hat{w}_3)_6 = \hat{V}\hat{U}\hat{V}$	$(\hat{c}_3)_6 = [\hat{V}, \hat{U}, \hat{V}]$
$(\hat{w}_3)_7 = \hat{V}\hat{V}\hat{U}$	$(\hat{c}_3)_7 = [\hat{V}, \hat{V}, \hat{U}]$
$(\hat{w}_3)_8 = \hat{V}\hat{V}\hat{V}$	$(\hat{c}_3)_8 = [\hat{V}, \hat{V}, \hat{V}]$
$\vdots$	$\vdots$

Table 2: This table explains the symbols for words and commutators used in the text for order r between 1 and 3. To avoid confusion, note that $\hat{w}$ is a word, while $\hat{W}$ is the BCH series.

supposing we know the coefficients $(a_r)_j$ in the linear combinations of the words $(\hat{w}_r)_j$ giving $\hat{W}_r$, namely

$$\hat{W}_r = \sum_{j=1}^{2^r} (a_r)_j \ (\hat{w}_r)_j \qquad (48)$$

all we need to do is to put (45) into (47) and to compare it with (48). By comparing coefficients in front of individual words $(\hat{w}_r)_j$ on the left and on the right we get

$$\sum_{k=1}^{2^r} (b_r)_k \ (M_r)_{k,j} = (a_r)_j$$

which in matrix form is

$$M_r^T \cdot \vec{b}_r = \vec{a}_r \qquad (49)$$

(here T means transpose). This is the linear equation to give the coefficients $(b_r)_k$ in front of commutators in the BCH series. With the exception of order $r = 1$