TECHNISCHE
UNIVERSITÄT
WIEN

DIPLOMARBEIT

Local error analysis for generalised splitting methods

ausgeführt am

Institut für
Analysis und Scientific Computing
TU Wien

unter der Anleitung von

Ao. Univ. Prof. Dr. Winfried Auzinger

durch

Maximilian Brunner
Matrikelnummer: 01528246
Schumanngasse 28/7
1180 Wien

Wien, am 14. August 2020

Verfasser

Betreuer

Abstract

This thesis introduces the basic theory of splitting methods for evolution equations. A symmetrised version of the defect is discussed and the defect is established as an asymptotically correct local error estimator. The general background for high order splitting such as the *Baker-Campbell-Hausdorff* formula and symmetrised versions thereof are treated and order conditions for high order splittings are extracted. In particular, we take a closer look at skew-hermitian matrices. In addition, we cover a 'dual' approach - the Zassenhaus splitting - and discuss the main ingredients Magnus provided for the analysis of the Zassenhaus splitting. A symmetrisation of Magnus' approach is made. Next, we introduce *inner symmetrised defects* and elaborate on its Taylor expansion. This is the key component to the more basic approach. We focus on the error expansion of the Strang splitting - our basic case of the more general Zassenhaus type setting. The systematic treatment of the general case offers ideas for further generalisations and provides a basis for a good understanding of the high level theory results. It is based on the Faà di Bruno identity and Bell polynomials play a key role when generalising the Lie expansion formula. We use Feynman diagrams for a compact and clear picture of the derivatives we will encounter. In the end, we have successfully recovered the order condition previously seen in the BCH formula by using the Taylor approach. We will conclude the thesis with an application of the order conditions to a physical problem.

Keywords: evolution equations, splitting methods, symmetrised defect, order conditions, local error estimators, Zassenhaus splitting, Bell polynomials, Faà di Bruno formula

Danksagung

Ich möchte mich bei Prof. Winfried Auzinger für die konstruktive und hilfreiche Kritik und das gute, respektvolle Arbeitsumfeld bedanken. Ebenso fand ich die dabei gewonnenen Einblicke in die wissenschaftliche Arbeitsweise durch die Arbeitsgruppe mit Othmar Koch und Harald Hofstätter lehrreich.

Mein Dank gebührt Victor Wenin, zum einen für das Korrekturlesen der Arbeit, zum anderen aber auch für den freundschaftlichen Rat und die vielen, unerschöpflichen Diskussionen während des ganzen Studiums über Gott, die Welt und Mathematik.

Dem österreichschen Staat danke ich für die erhaltene Studienbeihilfe, die das Studium finanziell leistbar gemacht hat.

Bei Martina Schwarz möchte ich mich für dies und das und einfach für alles bedanken.

Abschließend möchte ich mich bei meiner ganzen Familie bedanken, aber ganz besonders bei meinen Eltern. Ihre bedingungslose Unterstützung und die Freiheit meinen eigenen Weg zu gehen, ermöglichte mir nicht nur das Studium, sondern auch viel persönliches Wachstum. Ich danke von ganzem Herzen.

Maximilian Brunner

Wien, 14. August 2020

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Diplomarbeit selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 20. August 2020

Name des Autors

Contents

1 Introduction

In quantum mechanics, the Schrödinger equation encodes the dynamic of a quantum system. The energy of the system is given by the Hamiltonian operator H and the evolution equation in generic form reads

$$-\mathrm{i}\partial_t u(x,t) = H(u(x,t))$$

where we call u a wave function. Most of the time in physics, the Hamiltonian H consists of more independent summands. For example, the most common splitting

$$H = -\tfrac{1}{2}\,\partial_x^2 + V(x)$$

consists of a kinetic term $2^{-1}\,\partial_x^2$ and a potential $V(x)$. After linearisation and semidiscretisation in time, we may ask for efficient numerical schemes for calculating the dynamics given by the abstract problem

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}u(t) = H(u(t)) = (A+B)(u(t)). \tag{1.1}$$

The exact solution for given initial vector $u(0) = u_0$ reads

$$u(t) = \mathrm{e}^{t\,H}\,u_0 = \mathrm{e}^{t\,(A+B)}\,u_0.$$

For commutative matrices A and B, we have

$$\mathrm{e}^A\mathrm{e}^B = \mathrm{e}^{A+B} = \mathrm{e}^B\mathrm{e}^A,$$

which does not hold true in general. Oftentimes, it is true that

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}u(t) = A(u(t)) \quad \text{resp.} \quad \tfrac{\mathrm{d}}{\mathrm{d}\,t}u(t) = B(u(t))$$

are easier to solve than the original problem. The easiest splitting methods are approximations of the form

$$\mathrm{e}^{t\,(A+B)} \approx \mathrm{e}^{tA}\,\mathrm{e}^{tB}.$$

Section 3 is based on an observation seen in [Auzinger and Koch, 2018]: For a fundamental solution $\mathcal{E}(t)$ of (1.1), the following equations hold true:

$$\mathcal{E}' = H\mathcal{E} \quad \text{resp.} \quad \mathcal{E}' = \frac{1}{2}\{\mathcal{E}, H\}. \tag{1.2}$$

For a numerical scheme $\mathcal{S}(t)$, the equations (1.2) cannot hold true, but we can use it to define two possible defects - which are closely related to the local error:

$$\mathcal{S}'(t) - H\mathcal{S} \quad \text{and} \quad \mathcal{S}' - \frac{1}{2}\{\mathcal{S}, H\}.$$

Using an integral representation for the local error, we will arrive at local error estimators, which will be proportional to the defects. The main result in Section 3 is a theorem which states that the symmetrised defect is an approximation of higher order for the local error.

In Section 4, we compare the Lie-Trotter splitting - the prototype problem for the classical approach - with the Strang splitting - the easiest case for a symmetrised splitting method. During the course of that, we will introduce the famous Baker-Campbell-Hausdorff formula as well as its symmetrised version. The main source for this section is the thorough treatment of high order splitting methods done by [Singh, 2018]. A paper written by [Van Loan, 1977] offers an error bound for perturbations. Taking a look at the high order splitting

$$e^{t\,(A+B)} \approx e^{\frac{1}{2}tA}\, e^{\frac{1}{2}tB}\, e^{t^3C}\, e^{\frac{1}{2}tB}\, e^{\frac{1}{2}tA}, \tag{1.3}$$

we can get an explicit condition on how to choose C to increase the order of the numerical scheme, which we aim to reconstruct with a more basic approach in the course of this thesis.

In Section 5, we consider a closely related problem called the Zassenhaus splitting. We will take a look at an algebraic approach offered by [Magnus, 1954] and apply the same reasoning to the symmetrised splitting approach.

The main aim of this thesis is to reconstruct the order conditions for C with a more basic approach - a Taylor expansion. In Section 6, we proceed in the same way as [Auzinger, 2018]: We introduce the so-called inner symmetrised defect and calculate its Taylor expansion explicitly for Strang splitting.

[Auzinger, 2018] partially offered ideas on how to approach the more general setting as in (1.3). In Section 7, we start with a calculation of the inner symmetrised defect for the more general setting. The steps needed for the Taylor expansion are generalised - with the help of Bell polynomials and the Faà di Bruno identity. We provide a systematic approach on how to reconstruct the order conditions from the inner symmetrised defect. In the course of that, a formalism based on graphs - Feynman diagrams - as seen in [Johnson-Freyd, 2009] is developed and applied to our problem. Further generalisations are easily anticipated, especially on how to keep an overview on inflating recurrence relations. At the end, we elaborate the connection between the defect and the inner symmetrised defect.

Section 8 provides insights to a possible application - the Schrödinger equation mentioned at the very beginning of the introduction.

2 Preliminaries

Marianne: – *Pierrot.*
Ferdinand: *Je m'appelle Ferdinand!*
— *from* Pierrot le Fou

As naming things is arbitrary, it is convenient to mention the most important definitions right at the beginning to be familiar with the notation encountered in this thesis.

2.1 General setting

The problem setting and terminologies are extracted from [Auzinger et al., 2014, Section 2].

We start with a Banach space Ω. The *linear evolution equation* is stated as follows

$$u'(t) := \tfrac{\mathrm{d}}{\mathrm{d}\,t}u(t) = H\,u(t), \quad t \geq 0, \tag{LEE}$$

resp. the *nonlinear evolution equation*

$$u'(t) := \tfrac{\mathrm{d}}{\mathrm{d}\,t}u(t) = H(u(t)), \quad t \geq 0, \tag{NLEE}$$

with given initial value

$$u_0 := u(t_0) \in \Omega. \tag{IV}$$

Notice that the given operators A, B and H do not depend on time, both in (LEE) and (NLEE).

Over the course of this thesis, we will focus on a special form of the *time evolution operator* H, namely that can be split up additively into two operators A and B as

$$H = A + B. \tag{SPLIT}$$

This property (SPLIT) is key for the development of numerical schemes: The problem can then be split up such that the subproblems

$$u' = Au \quad \text{and} \quad u' = Bu$$

have known and easier computable solutions. Glueing them together by certain schemes will yield a sufficiently good approximation of the more difficult problem we initially wanted to solve - a linear or nonlinear evolution equation without any assumption on splitting.

We define an arbitrary operator $X \in \{A, B, H\}$ as

$$X \colon D(X) \subseteq \Omega \to \Omega,$$

where $D(X)$ is the domain of the operator X. We assume that the domains of H, A and B are related by

$$D(A) \cap D(B) = D(H) \subseteq \Omega.$$

Further treatment of these problem using operator theory can be found in [Evans, 2010, Section 7.4.1]for the linear case as well as for nonlinear operators in [Evans, 2010, Section 9.6.2].

As we assume semidiscretisation in time of problem (LEE), the involved operators A, B and H are linear as well as $X \colon D(X) \subseteq \Omega = \mathbb{R}^n \to \mathbb{R}^n$ for $X \in \{A, B, H\}$ and therefore matrices.

We start with common notions, see for instance [Schöberl, 2016].

Definition 1 (Exponential linear of operator). Let $X\colon D(X) \subseteq \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator. We define the *exponential of an operator* e^X as

$$\mathrm{e}^X := \sum_{k=0}^{\infty} \frac{X^k}{k!}.$$

This definition applies for matrices, too. ◇

In the case of commuting operators A and B, $\mathrm{e}^A\mathrm{e}^B = \mathrm{e}^{A+B} = \mathrm{e}^B\mathrm{e}^A$ holds true. Notice that a different matrix we will call the generator can be found (namely $A+B$) to express (and thus simplify) $\mathrm{e}^A\mathrm{e}^B$. Numerically, this will be indeed a simplification as matrix addition is cheap in comparison to multiplication and exponentiation.

Definition 2 (Commutator, anticommutator and Lie product). Let A and B be operators defined on the same Banach space as given above. The *commutator* resp. *anticommutator* of A and B are defined as follows:

$$[A, B] := A\,B - B\,A, \quad \{A, B\} := A\,B - B\,A.$$

We call the commutator $[A, B]$ the *Lie product of A and B*. ◇

Remark 1. The commutator can be seen as a measure for curvature in the space of operators. If $[A, B] = 0$, the order of application does not matter. With other words, the application in the space of operators is isotropic.

Example 1. Using the definition of the exponential series, one can immediately see that for time-independent H, the fundamental solution $\mathcal{E}(t) = e^{t\,H}$ of the evolution equation (LEE) satisfies

$$\mathcal{E}' = H\mathcal{E} = \mathcal{E}H = \frac{1}{2}(H\mathcal{E} + \mathcal{E}H) = \frac{1}{2}\{\mathcal{E}, H\}, \tag{2.1}$$

since $[H, e^{t\,H}] = 0$, i. e. they commute. This example will be important in Section 3.

Let X and Y be matrices. We consider the ring of noncommutative power series in variables X and Y - see [Auzinger and Herfort, 2014, p. 247] - denoted by

$$\mathbb{C}\langle X, Y\rangle$$

and we call X and Y generators of the ring.

Definition 3 (Lie element). We define the generators X and Y of the ring $\mathbb{C}\langle X, Y\rangle$ to be a *Lie element* of degree one. Recursively, we define all linear combinations of Lie elements of degree one with coefficients in $\mathbb{C}$ to be Lie elements and any Lie product of Lie elements to be Lie element too, where the degree indicates the number of recurrences used - see [Magnus, 1954, p.652]. ◇

We introduce a convenient notation for capturing repeated application of Lie products.

Definition 4 (Nested commutators). The *nested commutators of U and X* are defined as follows:

$$\begin{aligned} [U,X]_0 &:= X, \\ [U,X]_1 &:= [U,X], \\ \text{and generally,} \quad [U,X]_j &:= [U,[U,X]_{j-1}], \quad j \in \mathbb{N}. \end{aligned}$$

◇

Definition 5 (Matrix Lie group). We call G a *matrix Lie group*, if G is closed subgroup of $\mathrm{GL}_n(\mathbb{C}) = \{A \in \mathbb{C}^{n\times n} \mid A \text{ is invertible}\}$. ◇

This definition is found in [Hall, 2003, the equivalence mentioned on p. 4].

Example 2 (Unitary Group $U(n)$). We call the set

$$U(n) := \{A \in \mathrm{GL}_n(\mathbb{C}) \mid A^* A = \mathbb{1}\}$$

unitary group where $(A^*)_{jk} = \overline{A_{kj}}$. With other words, the adjoint coincides with the inverse, i. e. $A^* = A^{-1}$. This definition is found in [Hall, 2003, p.6]. $U(n)$ is an example of a matrix Lie group.

Definition 6 (Matrix Lie algebra). Let G be a matrix Lie group. The *matrix Lie algebra* $\mathfrak{g}$ *of G* is defined by

$$\mathfrak{g} := \{X \in \mathbb{C}^{n\times n} \mid \mathrm{e}^{tX} \in G \text{ for all real } t\}.$$ ◇

In this thesis, we will frequently deal with unitary or skew-hermitian matrices. We will now see how they are related:

Example 3 (Skew hermitian matrices). We follow [Hall, 2003, Section 2.5.4, p. 40] and denote the Lie algebra $\mathfrak{u}(n)$ of $U(n)$ by

$$\mathfrak{u}(n) := \{X \in \mathrm{GL}_n(\mathbb{C}) \mid X^* + X = 0\}.$$

By definition, $\mathrm{e}^{tX} \in U(n)$ for all real t. This yields

$$\mathrm{e}^{-tX} = \left(\mathrm{e}^{tX}\right)^{-1} \stackrel{!}{=} \left(\mathrm{e}^{tX}\right)^* = \mathrm{e}^{tX^*}.$$

A sufficient condition is $X^* = -X$. Differentiating at time $t = 0$ yields necessity.

2.2 Flows

We want to model the dynamics of problem (LEE). The natural concept appearing is that of a flow. Let us take a look at the slightly more general problem and introduce the concept of flow - see [Deuflhard and Bornemann, 2013, p. 131]: Let the autonomous equation

$$u'(t) = f(t,u), \quad u(t_0) = u_0$$

be given with a sufficiently smooth solution $u \in C^1([t_0,T];\mathbb{R}^n)$.

Definition 7 (Exact flow). Let $I \subseteq \mathbb{R}$ be the maximal existence interval[1]. A function Φ_s^t defined on $I \times \Omega$ is called *exact flow*, if it satisfies

$$\Phi_t^t u = u \qquad (\textit{consistency}) \tag{2.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}\tau} \Phi_t^{t+\tau} u \Big|_{\tau=0} = f(t, u) \qquad (\textit{local conformity}) \tag{2.3}$$

$$\Phi_r^t \Phi_s^r u = \Phi_s^t u \qquad (\textit{semigroup property}) \tag{2.4}$$

for all $(t, u) \in I \times \Omega$. ◇

Remark 2. The map $\Phi_s^t(x)$ encodes the dynamics of the system. For a given particle $x \in \Omega$ in the state space Ω at time s, the flow $\Phi_s^t(x)$ returns the element of the state space at time t.

For numerical schemes, we define:

Definition 8 (Consistent one-step integrator of order p). Let $\mathcal{E}(t)$ be the exact flow of (NLEE). We call a (numerical) flow $\mathcal{S} = \mathcal{S}(t)$ a *consistent one-step integrator of order* p if it fulfils

$$\mathcal{E}(t) = \mathcal{S}(t) + \mathcal{O}(t^{p+1}), \quad t \to 0$$

and agree at starting time t_0. ◇

Remark 3. For numerical schemes, [Deuflhard and Bornemann, 2013, p. 131] points out that the only possible relaxation lies in suspending (2.4) as 2.2 is elementary and 2.3 is the only connection to the right-hand side.

We also introduce the concept of symmetry for flows. This is captured by the adjoint method, which needs the definition of the inverse method.

Let $\mathcal{E}(t)$ be the fundamental solution of the problem (LEE). Due to the well-known *Theorem of Picard-Lindelöf* - assuming sufficient smoothness, see [Prüss and Wilke, 2010, Satz 2.2.2, p.31] - we get a maximal interval of existence for the solution $u(t)$. A Lemma about the *Wronski determinant* - see [Prüss and Wilke, 2010, Lemma 3.1.2, p.44] - guarantees the invertibility for the maximal time interval given by the Theorem of Picard-Lindelöf, if we assume invertibility at time $t = 0$.

As the existence of an inverse method is now clear, we follow the traces of [Hairer et al., 2006].

Definition 9 (Adjoint method). Let $\Psi_t := \Psi_0^t$ be a consistent one-step integrator for the exact flow $\Phi_t := \Phi_0^t$. We call Ψ_t^* the *adjoint method of* Ψ_t, if it fulfils

$$\Psi_{t,*} := \Psi_t^* := (\Psi_{-t})^{-1}. \tag{2.5}$$

This definition is found in [Hairer et al., 2006, Definition 3.1, p.42]. Sometimes, it will be convenient to write the asterisk in the subscript as we will deal with derivatives. ◇

[1] For further details, cf. Theorem of Picard-Lindelöf.

Definition 10 (Symmetric method). A method is called *symmetric* or *selfadjoint*[2], if

$$\Psi_t^* = \Psi_t \tag{2.6}$$

holds - cf. [Hairer et al., 2006, Definition 3.1, p.42]. ⋄

The next lemma shows that adjoint methods have a desirable property: The leading term of the method is of the same order as the leading term of the original method. In addition, the leading term of the error differs only by sign.

Lemma 1 (Error term adjoint method). Let the exact flow Φ_t of (LEE) be given. We consider a one-step method Ψ_t of order p which fulfils

$$\Psi_t(u_0) = \Phi_t(u_0) + C(u_0)t^{p+1} + \mathcal{O}(t^{p+2}).$$

Then the adjoint method Ψ_t^* is of order p as well and it can be written as

$$\Psi_t^*(u_0) = \Phi_t(u_0) + (-1)^p\, C(u_0)t^{p+1} + \mathcal{O}(t^{p+2}).$$

Proof. This proof can be found in [Hairer et al., 2006, Theorem 3.2, p.42f]. □

An immediate consequence is the following corollary:

Corollary 1 (Error of symmetric methods). The expected order of symmetric schemes is even.

2.3 Sylvester-type equations

When taking a look at problem (LEE), we will naturally encounter a so-called *Sylvester-type equation*. This can be seen as a generalisation of first order differential equations to a noncommutative setting. [Auzinger et al., 2014, Section 2] examine this type of equation in further detail.

The following Lemma states the most prominent feature of a *Sylvester-type equation*.

Lemma 2. Let A and B be time-independent operators, the time-dependent operator $\mathcal{X}(t)$ and the (possibly time-dependent) inhomogeneity $\mathcal{C}(t)$ be given[3]. The *inhomogeneous Syvester-type equation* reads

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{X} = \mathcal{X}A + B\mathcal{X} + \mathcal{C}, \quad \mathcal{X}(0) = \mathcal{X}_0.$$

The solution of this initial value problem can be represented as

$$\mathcal{X}(t) = \mathrm{e}^{t\,B}\,\mathcal{X}(0)\,\mathrm{e}^{t\,A} + \int_0^t \mathrm{e}^{(t-\tau)\,B}\,\mathcal{C}(\tau)\,\mathrm{e}^{(t-\tau)\,A}\,\mathrm{d}\,\tau. \tag{2.7}$$

Proof. Plugging in $t = 0$, the equation holds true.

[2]We consider bounded operators only and therefore, we need not care about subtleties.
[3]From now on, we will stick to the convention that time-dependent letters will be written caligraphically.

Taking the derivate and using the Leibnitz rule, we arrive at

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{X}(t) &= \tfrac{\mathrm{d}}{\mathrm{d}\,t}\left(\mathrm{e}^{t\,B}\,\mathcal{X}(0)\,\mathrm{e}^{t\,A} + \int_0^t \mathrm{e}^{(t-\tau)\,B}\,\mathcal{C}(\tau)\,\mathrm{e}^{(t-\tau)\,A}\;\mathrm{d}\,\tau\right) \\
&= B\,\mathrm{e}^{t\,B}\,\mathcal{X}(0)\,\mathrm{e}^{t\,A} + \mathrm{e}^{t\,B}\,\mathcal{X}(0)\,\mathrm{e}^{t\,A}\,A + \mathrm{e}^{(t-\tau)\,B}\,\mathcal{C}(\tau)\,\mathrm{e}^{(t-\tau)\,A}\Big|_{\tau=t} \\
&\quad + \int_0^t B\,\mathrm{e}^{(t-\tau)\,B}\,\mathcal{C}(\tau)\,\mathrm{e}^{(t-\tau)\,A} + \mathrm{e}^{(t-\tau)\,B}\,\mathcal{C}(\tau)\,\mathrm{e}^{(t-\tau)\,A}\,A\;\mathrm{d}\,\tau \\
&= B\,\mathcal{X} + \mathcal{X}\,A + \mathcal{C}.
\end{aligned}$$

The claim follows.

This is a well-known results of the theory for non-commutative linear differential equations of first order (e.g. found in [Behr et al., 2018, Thm. 2.2] about Sylvester-type differential equations). □

This concludes the section about preliminaries and we are ready to proceed with numerics now.

3 Local error estimation using defects

"Although this may seem a paradox, all exact science is dominated by the idea of approximation."
— Bertrand Russell

When examining an integrator, it is useful being able to estimate local errors of numerical integrators that quantify the deviation of the integrator from the exact flow it estimates. A measure for the local error made is the defect. In our context, we compare a numerical flow associated by a numerical scheme with the theoretical, exact flow of the underlying evolution equation.

The backbone of this section is seen in [Auzinger et al., 2019b, Section 2 and 3] - where all statements can be found for the nonlinear case, too. A compressed version of the ideas presented is found in [Auzinger et al., 2019a, p.11-15]. The linear case is more approachable and thus quite insightful - for the nonlinear case we refer to literature. We will take a look at the classical approach for defects first and then introduce the symmetrised version of the defect. The symmetrised defect will yield a higher order in convergence in comparison to the classical approach, which will be formulated as a Theorem, which is found in [Auzinger et al., 2019b, Theorem 4, p.8].

We start by introducing the following definition:

Definition 11 (Local error). Let $\mathcal{E}(t)$ be the exact flow of (NLEE) and let $\mathcal{S} = \mathcal{S}(t)$ be a consistent one-step integrator of order $p \in \mathbb{N}$. We call the difference between the methods $\mathcal{L}(t)$ the *local error*

$$\mathcal{L}(t) := \mathcal{S}(t) - \mathcal{E}(t). \tag{3.1}$$

$\diamond$

Since $\mathcal{S}(0) = \mathcal{E}(0)$, we can easily see that the initial value is given by

$$\mathcal{L}(0) = 0. \tag{3.2}$$

3.1 Classical and symmetrised defect

We start with some vocabulary, which follows [Auzinger and Koch, 2018].
Recalling and restating parts of equation (2.1), the exact flow fulfils

$$\mathcal{E}' = H\mathcal{E} \quad \text{resp.} \quad \mathcal{E}' = \frac{1}{2}\{\mathcal{E}, H\}. \tag{3.3}$$

We know that these equations cannot be fulfilled by a numerical scheme $\mathcal{S}$. Replacing $\mathcal{E}$ by $\mathcal{S}$, we obtain a local measure for its 'discrepancy'. Formally, we can investigate

$$\mathcal{S}'(t) - H\mathcal{S} \quad \text{and} \quad \mathcal{S}' - \frac{1}{2}\{\mathcal{S}, H\}.$$

These two possible approaches[4] will be covered now.

Clarifying the term we are talking about for the classical approach, we define the

[4] Actually, any convex combination is a possible way for defining the defect, see [Auzinger et al., 2019b, Remark p.5]

Definition 12 (Classical version of defect). For the numerical flow $\mathcal{S}(t)$, the expression

$$\mathcal{D}_c(t) := S'(t) - HS(t)$$

is called the *classical defect.* ⋄

In the classical case, the defect satisfies - by definition - $\mathcal{S}' = H\mathcal{S} + \mathcal{D}_c$. Taking the derivative of the local error, we obtain using the left-hand equality in(3.3)

$$\mathcal{L}'(t) = \mathcal{S}'(t) - \mathcal{E}'(t) = H\mathcal{S}(t) + \mathcal{D}_c(t) - H\mathcal{E}(t) = H\mathcal{L}(t) + \mathcal{D}_c(t). \tag{3.4}$$

Equation (3.4) is of generic form for a linear ordinary differential equation of order 1, where the defect plays the role of an inhomogeneity with initial value $\mathcal{L}(0) = 0$ as observed earlier in (3.2). Equation (3.4) is an ordinary differential equation where $A = 0$, $B = H$ with known initial value. According to Lemma 2, the solution can be represented as

$$\begin{aligned}\mathcal{L}(t) &= \cancel{\mathrm{e}^{t\,H}\,\mathcal{L}(0)\,\mathrm{e}^{t\,0}} + \int_0^t \mathrm{e}^{(t-\tau)\,H}\,\mathcal{D}_c(\tau)\,\underbrace{\mathrm{e}^{(t-\tau)\,0}}_{=\mathbb{1}}\,\mathrm{d}\,\tau \\ &= \int_0^t \underbrace{\mathcal{E}((t-\tau))\,\mathcal{D}_c(\tau)}_{=:\,\Theta_c(\tau)}\,\mathrm{d}\,\tau. \end{aligned} \tag{3.5}$$

We call the integrand Θ_c.

For our purposes, the symmetrised version of the defect is more important than the classical. Motivated by the right-hand of (3.3), we define analogously the

Definition 13 (Symmetrised version of defect). For a numerical flow $\mathcal{S}(t)$, the expression

$$\mathcal{D}_s(t) := \mathcal{S}'(t) - \tfrac{1}{2}\{\mathcal{S}(t), H\}. \tag{3.6}$$

is called *symmetrised* or *symmetric defect.* ⋄

Again, taking the derivative of

$$\mathcal{L}(t) := \mathcal{S}(t) - \mathcal{E}(t), \quad \mathcal{L}(0) = 0,$$

using $\mathcal{S}' = \mathcal{D}_s(t) + \frac{1}{2}\{\mathcal{S}, H\}$ as well as the second option in (3.3), we end up with

$$\begin{aligned}\mathcal{L}'(t) &= \mathcal{S}'(t) - \mathcal{E}'(t) = \mathcal{D}_s(t) + \tfrac{1}{2}\{\mathcal{S}, H\} - \tfrac{1}{2}\{\mathcal{E}, H\} = \tfrac{1}{2}\{\mathcal{L}, H\} + \mathcal{D}_s(t) \\ &= \tfrac{1}{2}\mathcal{L}H + \tfrac{1}{2}H\mathcal{L} + \mathcal{D}_s(t). \end{aligned} \tag{3.7}$$

Applying Lemma 2 with $A = B = 2^{-1}H$ and inhomogeneity $\mathcal{C}(t) = \mathcal{D}_s(t)$, we obtain following representation and define the integrand Θ_s as

$$\begin{aligned}\mathcal{L}(t) &= \cancel{\mathrm{e}^{\frac{1}{2}\,t\,H}\,\mathcal{L}(0)\,\mathrm{e}^{\frac{1}{2}\,t\,H}} + \int_0^t \mathrm{e}^{\frac{1}{2}\,(t-\tau)\,H}\,\mathcal{D}_c(\tau)\,\mathrm{e}^{\frac{1}{2}\,((t-\tau)\,H}\,\mathrm{d}\,\tau \\ &= \int_0^t \underbrace{\mathcal{E}(\tfrac{1}{2}(t-\tau))\,\mathcal{D}_s(\tau)\;\mathcal{E}(\tfrac{1}{2}(t-\tau))}_{=:\,\Theta_s}\,\mathrm{d}\,\tau. \end{aligned} \tag{3.8}$$

3.2 The generic barrier and local error estimators

The main idea of local error estimators is - refer to [Auzinger et al., 2019b, Section 3] - to replace the integral representation of the local error by an a-posteriori estimator $\tilde{\mathcal{L}}_c(t)$ resp. $\tilde{\mathcal{L}}_s(t)$, which are both asymptotically correct. Formally, we approximate the integral for $\Theta \in \{\Theta_c, \Theta_s\}$

$$\mathcal{L}(t) = \int_0^t \Theta(\tau)\ \mathrm{d}\,\tau \approx \Theta(t) + \mathcal{O}(t^{p+2}), \quad t \to 0, \tag{3.9}$$

where p stands for the assumed order of the one-step scheme Θ. We refer to 3.9 as the *generic barrier*.

For classical and symmetrised defect, the fact that we consider one-step schemes of order p yields that the local error is of order p, i. e.

$$\mathcal{L}(t) = \mathcal{O}(t^{p+1}) \tag{3.10}$$

and hence, we have

$$\mathcal{D}_c(t) = \mathcal{D}_s(t) = \mathcal{O}(t^p), \quad \text{in particular} \quad \tfrac{\mathrm{d}^k}{\mathrm{d}\,t^k}\mathcal{D}_c(0) = \tfrac{\mathrm{d}^k}{\mathrm{d}\,t^k}\mathcal{D}_s(0) = 0, \ \text{for all } k = 1, \dots, p-1.$$

Now, approximating $\Theta(\tau)$ for $\Theta \in \{\Theta_c, \Theta_s\}$ by a Taylor expansion around $t = 0$, we see that

$$\Theta(\tau) = \frac{\tau^p}{p!}\Theta^{(p)}(0) + \mathcal{O}(\tau^{p+1}). \tag{3.11}$$

Coming back to (3.5) resp. (3.8), we find

$$\begin{aligned}\mathcal{L}(t) = \int_0^t \Theta(\tau)\ \mathrm{d}\,\tau = \int_0^t \left(\frac{\tau^p}{p!}\Theta^{(p)}(0) + \mathcal{O}(\tau^{p+1})\right)\mathrm{d}\,\tau = \frac{t^{p+1}}{(p+1)!}\Theta^{(p)}(0) + \mathcal{O}(t^{p+2})\\ = \tfrac{1}{p+1}\, t\,\Theta(t) + \mathcal{O}(t^{p+2}) =: \tilde{\mathcal{L}}(t) + \mathcal{O}(t^{p+2}),\end{aligned} \tag{3.12}$$

where we define $\tilde{\mathcal{L}}(t) = t(p+1)^{-1}\Theta(t)$ for both the classical and symmetrised defect. Notice that we used (3.11) for $\tau = t$ in the last step.

Definition 14 (Local error estimator $\tilde{\mathcal{L}}(t)$). We call expressions $\tilde{\mathcal{L}}(t)$ asymptotically correct *local error estimators* of $\mathcal{L}(t)$. ◇

In summary, we have found an asymptotically correct approximation for the local error estimators in terms of the defect, namely

$$\tilde{\mathcal{L}}_c(t) = \tfrac{1}{p+1}\, t\,\Theta_c(t) = \tfrac{1}{p+1}\, t\,\mathcal{D}_c(t) \quad \text{resp.} \quad \tilde{\mathcal{L}}_s(t) = \tfrac{1}{p+1}\, t\,\Theta_s(t) = \tfrac{1}{p+1}\, t\,\mathcal{D}_s(t).$$

3.3 An improved barrier for symmetrised defect

The next definition introduces a corrected scheme for the symmetrised defect $\mathcal{D}_s(t)$. The corrected scheme turns out to be a local error estimator of higher order than those seen previously in (3.12).

Definition 15 (Corrected scheme $\mathcal{R}$). We define the *corrected scheme* $\mathcal{R}$

$$\mathcal{R}(t) := \mathcal{S}(t) - \tfrac{1}{p+1}\, t\, \mathcal{D}_s(t) = \mathcal{S}(t) - \tilde{\mathcal{L}}_s(t).$$

◇

The definition is motivated by the observation that adding and subtracting (3.12) for the symmetrised case yields

$$\mathcal{L}(t) = \mathcal{S}(t) - \mathcal{E}(t) = \tilde{\mathcal{L}}_s(t) + \mathcal{S}(t) - \tilde{\mathcal{L}}_s(t) - \mathcal{E}(t) =: \tilde{\mathcal{L}}_s(t) + \mathcal{R}(t) - \mathcal{E}(t). \tag{3.13}$$

Importantly and surprisingly, the local error of the corrected scheme is increased by one order of magnitude, i. e. order $p+2$ instead of $p+1$[5], at least for self-adjoint schemes $\mathcal{S}(t)$. This is the result of the following theorem, which can be found in [Auzinger and Koch, 2018, Theorem 1] for the linear problem (LEE) or in [Auzinger et al., 2019b, Theorem 4] for the nonlinear case (NLEE). For the symmetrised defect, we therefore get

$$\mathcal{L}(t) = \tfrac{1}{p+1} t\, \mathcal{D}_s(t) + \mathcal{O}(t^{p+3}). \tag{3.14}$$

Theorem 1 (Symmetrised defect - corrected error estimator). Let $\mathcal{S}(t)$ denote a given one-step scheme of order $p \in \mathbb{N}$ with $p \geq 2$, for (LEE). Assume that $\mathcal{S}(t)$ is self-adjoint[6], i. e.

$$\mathcal{S}(-t)\mathcal{S}(t) = \mathbb{1}.$$

Let $\mathcal{R}(t)$ be the corrected scheme as in Definition 15.
Then the local error $\mathcal{L}_{\mathcal{R}}(t) = \mathcal{R}(t) - \mathcal{E}(t)$ - see (3.13) - satisfies

$$\mathcal{L}_{\mathcal{R}}(t) = \mathcal{O}(t^{p+3}), \quad t \to 0. \tag{3.15}$$

The corrected scheme is of order $p+2$ and is therefore *even* too.

Remark 4. The local error of the corrected scheme is of higher order in comparison to the generic barrier seen in (3.12).

Proof. Corollary 1 about self-adjoint schemes points out a plausible way of where the increase of order can possibly stem from, so let us check how close $\mathcal{R}(t)$ is to being selfadjoint:

$$\begin{aligned}\mathcal{R}(-t)\,\mathcal{R}(t) &= \Big(\mathcal{S}(-t) + \tfrac{1}{p+1} t\, \mathcal{D}_s(-t)\Big)\ \Big(\mathcal{S}(t) - \tfrac{1}{p+1} t\, \mathcal{D}_s(t)\Big) \\ &= \mathcal{S}(-t)\,\mathcal{S}(t) + \tfrac{1}{p+1} t\,(\mathcal{D}_s(-t)\,\mathcal{S}(t) - \mathcal{S}(-t)\,\mathcal{D}_s(t)) - \tfrac{1}{(p+1)^2}\, t^2\, \mathcal{D}_s(-t)\,\mathcal{D}_s(t).\end{aligned}$$

As $\mathcal{D}_s(t)$ is of order p by construction, the terms are already ordered by increasing order. Notice that the leading term $\mathcal{S}(-t)\mathcal{S}(t) = \mathbb{1}$ by assumption. Taking the derivative of this identity, we obtain

$$\begin{aligned} 0 &= \tfrac{\mathrm{d}}{\mathrm{d}\,t}\,(\mathcal{S}(-t)\mathcal{S}(t)) = -\mathcal{S}'(-t)\,\mathcal{S}(t) + \mathcal{S}(-t)\,\mathcal{S}'(t) \\ &= -\left(\mathcal{D}_s(-t) + \tfrac{1}{2}\{\mathcal{S}(-t), H\}\right)\ \mathcal{S}(t) + \mathcal{S}(-t)\left(\mathcal{D}_s(t) + \tfrac{1}{2}\{\mathcal{S}(t), H\}\right) \\ &= \underbrace{-\tfrac{1}{2}\{\mathcal{S}(-t), H\}\,\mathcal{S}(t) + \mathcal{S}(-t)\,\tfrac{1}{2}\{\mathcal{S}(t), H\}}_{=:\,2\,\mathcal{X}(t) \overset{?}{=} 0} + \underbrace{(\mathcal{S}(-t)\,\mathcal{D}_s(t) - \mathcal{D}_s(-t)\,\mathcal{S}(t))}_{\text{critical term for order improvement}}.\end{aligned}$$

[5] cf. (3.10).

[6] Notice that the self-adjointness assumption implies even order, cf. Corollary 1.

If we can prove that $\mathcal{X}$ vanishes, then the critical term for the selfadjointness of $\mathcal{R}(t)$ ceases to be an order barrier. Therefore, examing $\mathcal{X}(t)$ more closely by expanding the anticommutator, we arrive at

$$\begin{aligned}\mathcal{X}(t) &= -\left(\mathcal{S}(-t)\,H + H\,\mathcal{S}(-t)\right)\,\mathcal{S}(t) + \mathcal{S}(-t)\,\left(\mathcal{S}(t)\,H + H\,\mathcal{S}(t)\right)\\ &= -\mathcal{S}(-t)\,H\,\mathcal{S}(t) - H\,\underbrace{\mathcal{S}(-t)\,\mathcal{S}(t)}_{=\mathbb{1}} + H\,\underbrace{\mathcal{S}(-t)\,\mathcal{S}(t)}_{=\mathbb{1}} + \mathcal{S}(-t)\,H\,\mathcal{S}(t) = 0.\end{aligned}$$

Summing up, we have shown that $\mathcal{R}(t)$ is what [Auzinger et al., 2019b, (3.3a)] call almost self-adjoint, i. e.

$$\mathcal{R}(-t)\,\mathcal{R}(t) = \mathbb{1} - \tfrac{1}{(p+1)^2}\,t^2\,\mathcal{D}_s(-t)\,\mathcal{D}_s(t).$$

As $\mathcal{D}_s(t)$ is of order p, we see that

$$\mathcal{R}(-t)\,\mathcal{R}(t) = \mathbb{1} + \mathcal{O}(t^{2\,p+2}).$$

On the other side, we know that $\mathcal{R}(t)$ is at least of order $p+1$, so we can expand the corrected local error $\mathcal{R}(t)$ as follows

$$\mathcal{R}(t) = \mathcal{E}(t) + t^{p+2}\,C + \mathcal{O}(t^{p+3}), \quad t \to 0.$$

Now, as $2\,p+2 \geq p+3$ for all $p \geq 1$, we arrive at an equation for C, namely

$$\begin{aligned}\mathcal{R}(-t)\,\mathcal{R}(t) &= \left(\mathcal{E}(-t) + (-t)^{p+2}\,C + \mathcal{O}(t^{p+3})\right)\,\left(\mathcal{E}(t) + t^{p+2}\,C + \mathcal{O}(t^{p+3})\right)\\ &= \mathcal{E}(-t)\,\mathcal{E}(t) + t^{p+2}\,(\mathcal{E}(-t)\,C + C\,\mathcal{E}(t)) + \mathcal{O}(t^{p+3}).\end{aligned}$$

Notice that $(-t)^{p+2} = t^{p+2}$ for even p as assumed. Since $\mathcal{E}(-t)\,\mathcal{E}(t) = \mathbb{1}$, we extract the equation (as all other terms lower than $2\,p+2$ must vanish, especially the term of order $p+2$), hence

$$\begin{aligned}0 &= \mathcal{E}(-t)\,C + C\,\mathcal{E}(t) = \left(\mathbb{1} - t\,H + \mathcal{O}(t^2)\right)\,C + C\,\left(\mathbb{1} + t\,H + \mathcal{O}(t^2)\right)\\ &= 2\,C + \mathcal{O}(t), \quad t \to 0.\end{aligned}$$

For odd p, the reasoning boils down to the even case, as the equation above would yield a trivial equation for $-C + C = 0$ due to the missing switch of sign in $(-t)^{p+2}$.

We therefore conclude that $C = 0$ and the result is proven. □

4 Splitting methods and high level theory

"[...] the best thing since wrestling
Infesting in your kids' ears and nesting [...]"
— Eminem, *Without me*

We will see in this section that the error of splitting methodscan be expressed in terms graded nested commutations.

4.1 Lie-Trotter Splitting

The easiest method one can think of when considering splitting as in (LEE) is to apply the operators consecutively, that is

$$\mathrm{e}^{H} = \mathrm{e}^{A+B} \approx \mathrm{e}^{A}\mathrm{e}^{B}.$$

Let the flow Φ_h^A associated with A respectively the flow Φ_h^B associated with B be given as well as the initial value u_0. The Lie-Trotter numerical flow is given by

$$\Psi_h = \Phi_h^A \circ \Phi_h^B.$$

This means, at first we apply B and use this value $\Phi_h^B(u_0)$ as a starting point for the flow of A. The end point is defined to be $u_1 := \Psi_h$. This consecutive application of individual flows is illustrated in Figure 4.1.

A visualisation is the following:

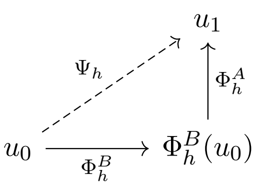


Figure 1: Schematic visualisation of the Lie-Trotter Splitting.

For further visualisations, we will stick to the convention laid out in Figure 4.1 that solid arrows symbolise given maps and dashed arrows stand for resulting items.

4.2 Baker-Campbell-Hausdorff formula

Assuming noncommutativity of A and B, the formula for the generator C for $\mathrm{e}^{A}\mathrm{e}^{B} = \mathrm{e}^{C}$ gets more involved. It is described by the so-called Baker-Campbell-Hausdorff formula as the next theorem states. The theorem and its proof can be found in Section 7.6 (and an exercise thereof) and Section 7.7 in [Stillwell, 2008]. We need the following

Definition 16 (Degree[7] of nested Lie bracket). Let $F(A, B)$ be a nested commutator. We say that the nested commutator is of *degree* k, if

$$F(tA, tB) = t^k F(A, B).$$

◇

[7] Compare with Definition 3.

This definition is motivated due to the fact that for instance [Singh, 2018, Section 2.2] uses tA and tB instead of A and B for the Baker-Campbell-Hausdorff formula. Casually speaking, it quantifies *how often* operators are commutated with each other.

Example 4. For instance, we call $F(A, B) = [tA, tB]$ a nested commutator of degree $k = 2$ as

$$[tA, tB] = t^2\,[A, B].$$

If we take $F(A, B) = [[B, A], A]$, we see

$$F(tA, tB) = [[tB, tA], tA] = t^3\,[[B, A], A]$$

is a nested commutator of degree $k = 3$.

Theorem 2 (Baker-Campbell-Hausdorff formula). Consider two quadratic matrices A and B of the same dimension $n \in \mathbb{N}$ over $\mathbb{C}$. Then the first terms of the generator D, i. e. the matrix D such that $\mathrm{e}^D \approx \mathrm{e}^A\,\mathrm{e}^B$, are given by

$$\begin{aligned} D &= A + B + \frac{1}{2}[A, B] - \frac{1}{24}\left([[B, A], A] + 2\,[[B, A], B]\right) + \dots \\ &= \sum_{k=1}^{\infty} F_k(A, B) =: \mathrm{BCH}(A, B), \end{aligned}$$

where the $F_k(A, B)$ represent the sum of all occurring terms of degree k in the Lie-algebra and are thus homogeneous polynomials of degree k in A and B. We call the appearing series *Baker-Campbell-Hausdorff formula* of A and B. Moreover, for each $k \geq 1$, the polynomial is a linear combination of A and B and (possibly nested) commutators of A and B.

Remark 5. [Magnus, 1954, p. 652] emphasises that the statement of the Baker-Campbell-Hausdorff theorem is that $\mathrm{BCH}(A, B)$ is a Lie element.

Proof. A purely algebraic proof without bells and whistles is given in [Stillwell, 2008, p.154ff] which credits Eichler to have found this comparably short proof. □

In numerics, we are interested in approximating the occurring series by partial sums, i. e.

$$\sum_{k=1}^{\infty} F_k(tA, tB) = \sum_{k=1}^{\infty} t^k\,F_k(A, B) = \sum_{k=1}^{K} t^k\,F_k(A, B) + \mathcal{O}(t^{K+1}), \quad t \to 0.$$

The error of the approximation of the exponentiation is conveyed when exponentiation is performed. This is captured by the next results.

Lemma 3 (Error bound for perturbation). Let $A \in \mathbb{C}^{n\times n}$ be given. For a perturbation ΔA, the following upper bound holds true:

$$\|\mathrm{e}^{t\,(A+\Delta A)}\|_2 \leq t\,\|\mathrm{e}^{t\,A}\|_2 \|\Delta A\|_2\,\mathrm{e}^{t\,(\mu(A)-\eta(A)+\|\Delta A\|_2)}, \tag{4.1}$$

with the maximal eigenvalue of $\frac{A+A^*}{2}$ and the maximal real part of eigenvalues of A, i. e.,

$$\mu(A) := \mu_2(A) = \max\left\{\lambda \,\middle|\, \tfrac{A+A^*}{2}v = \lambda v, v \neq 0\right\} \quad \text{and} \quad \eta(A) = \max\left\{\mathrm{Re}(\lambda) \mid Av = \lambda v, v \neq 0\right\}.$$

The expression $\mu(A)$ is called *logarithmic norm of A w. r. t. the 2-norm.*

Proof. The idea was given in [Singh, 2018, p.32f], refering to [Van Loan, 1977, Lemma 1 and Theorem 3]. A quick sketch of the main ingredients goes as follows:

Claim 1 (Dahlquist's bound). The following inequality is valid for all $0 \le s \le t$:

$$\|e^{As}\|_2 \le e^{\mu(A)s}. \tag{4.2}$$

Proof of claim. This estimate about the logarithmic 2-norm $\mu(A)$ is found in [Dahlquist, 1959, Equation (1.3.8) on page 14] for the special case of the 2-norm. Dahlquist's publication covers the time-dependent case $A(t)$, too. In Kato's terms, it gives an upper bound for the unperturbed operator A, which is in our case simply a matrix. □

Roughly speaking, the next claim gives a bound for the effect a perturbation ΔA has on Dahlquist's bound (4.2).

Claim 2 (Kato's bound). Let the inequality (4.2) hold true. Then, we have for all $0 \le s \le t$:

$$\|e^{(A+\Delta A)s}\|_2 \le e^{(\mu(A)+\|\Delta A\|_2)s}. \tag{4.3}$$

Proof of claim. This identity is found in [Kato, 1995, Theorem 2.1, particularly Equation (2.7) on page 497ff]. Notice the assumptions of Theorem 2.1 in Kato's book require a bound for the unperturbed operator A, which we have established with the previous proven claim. □

Claim 3.

$$e^{(A+\Delta A)t} - e^{At} = \int_0^t e^{A(t-s)}\, \Delta A\, e^{(A+\Delta A)s}\, \mathrm{d}\, s \tag{4.4}$$

Proof of claim. The proof is found in [Higham, 2008, Problem 10.1 with solution provided on page 371]. Define $X(t) := e^{(A+\Delta A)t}$ and we differentiate:

$$X'(s) = (A+\Delta A)X(s), \quad X(0) = \mathbb{1}.$$

We notice that this is a homogeneous - i. e. $C = 0$ - Sylvester-type equation with $A = 0$ and $B = A + \Delta A$. Hence, equation (2.7) in Lemma 2 establishes the claim. □

Applying the 2-norm on equation (4.4), using the triangle inequality for integrals, we get

$$\begin{aligned}\|e^{(A+\Delta A)t} - e^{At}\|_2 &\le \int_0^t \|e^{A(t-s)}\, \Delta A\, e^{(A+\Delta A)s}\|_2\, \mathrm{d}\, s \\ &\le \|\Delta A\|_2 \int_0^t \|e^{A(t-s)}\|_2\, \|e^{(A+\Delta A)s}\|_2\, \mathrm{d}\, s \end{aligned} \tag{4.5}$$

The second inequality exploits the submultiplicativity of the norm, as well as the fact that real numbers do commute. Now, applying Dahlquist's estimate (4.2) as well as Kato's bound (4.3), we arrive at

$$\begin{aligned} &\le \|\Delta A\|_2 \int_0^t e^{\mu(A)(t-s)}\, e^{(\mu(A)+\|\Delta A\|_2)s}\, \mathrm{d}\, s \\ &= \|\Delta A\|_2\, e^{\mu(A)t} \int_0^t e^{\|\Delta A\|_2 s}\, \mathrm{d}\, s \le \|\Delta A\|_2\, e^{\mu(A)t}\, t\, e^{\|\Delta A\|_2 t}, \end{aligned}$$

where we have used the monotonicity of the integrand and integral involved. To consider the relative error on the left-hand side, $1 = \frac{\|e^{At}\|_2}{\|e^{At}\|_2}$ is inserted on the right-hand side. This opens up options for the use of the Schur decomposition of the matrix $e^{At} = QSQ^{-1}$ with Q being unitary and an upper triangular matrix S with the same spectrum as A. A well-known fact is that a unitary matrix U preserves the 2-norm of an arbitrary and complex matrix X, i. e.

$$\|UX\|_2 = \|X\|_2.$$

Therefore, we get for all λ in the spectrum of A that

$$\|e^{At}\|_2 = \|Q\,S\,Q^{-1}\|_2 = \|S\|_2 \geq |e^{\lambda t}| = |e^{\mathrm{Re}(\lambda)}|\,|e^{\mathrm{i}\,\mathrm{Im}(\lambda)}| \tag{4.6}$$

holds true. Due to Euler, we have $\|e^{\mathrm{i}\,x}\|_2 = 1$ for real x. The inequality is valid for the eigenvalue with maximal real part, too.

Summarising, we have

$$\|e^{(A+\Delta A)t} - e^{At}\|_2 \leq t\,\|\Delta A\|_2\,\|e^{At}\|_2 e^{(\mu(A)-\eta(A)+\|\Delta A\|_2)t}$$

as claimed. □

Remark 6. In particular, equation (4.6) establishes that

$$\mu(A) - \eta(A) \geq 0.$$

Remark 7 (skew-hermitian matrices). An important class for physical problems are the skew-hermitian matrices. Let A be a skew-hermitian matrix, we have

$$\mu(A) = 0, \quad \text{as} \quad \tfrac{A+A^*}{2} = 0.$$

The fact that all eigenvalues of skew-hermitian matrices are purely imaginary implies

$$\eta(A) = 0.$$

In addition, we know that the matrix exponential of a skew-hermitian matrix A is unitary - cf. Example 7. Therefore, the expression in (4.5) reduces to

$$\|e^{t\,(A+\Delta A)}\|_2 \leq t\,\|\Delta A\|_2. \tag{4.7}$$

Corollary 2. Let $p \in \mathbb{N}$. Let the skew-hermitian matrices M and N satisfy the bound

$$\|M - N\|_2 \leq C\,t^p. \tag{4.8}$$

Given that M and N are close enough - they satisfy 4.8 -, the bound (4.7) simplifies to

$$\|e^M - e^N\|_2 \leq \|M - N\|_2 \leq C\,t^p. \tag{4.9}$$

Proof. This Lemma is mentioned in [Singh, 2018, Lemma 2.2.1 on page 34]. As the difference of two skew-hermitian matrices is skew-hermitian, so is $\Delta A = M - N$. The claim then follows immediately from Remark 7. □

Corollary 3. Let A and B be skew-hermitian. We see that the Lie-Trotter splitting is of order 1 as

$$\mathrm{e}^{tH} = \mathrm{e}^{t(A+B)} = \mathrm{e}^{tA}\mathrm{e}^{tB} + \mathcal{O}(t^2), \quad t \to 0.$$

Proof. Following [Singh, 2018, p.33], we define $M := tA + tB$ and $N := \mathrm{BCH}(tA, tB)$, hence $\Delta A = \mathrm{BCH}(tA, tB) - M = t^2[A, B] + \mathcal{O}(t^3)$ for $t \to 0$. Using Corollary 2, Remark 7 as well as the skew-hermiticity of $tA + tB$, we arrive at

$$\|\mathrm{e}^{tA}\,\mathrm{e}^{tB} - \mathrm{e}^{tA+tB}\|_2 = \|\mathrm{e}^{\mathrm{BCH}(tA,tB)} - \mathrm{e}^{tA+tB}\|_2 \leq \|\mathrm{BCH}(tA, tB) - (tA + tB)\|_2 = Ct^2.$$

In the first step, we remembered that $\mathrm{BCH}(tA, tB)$ is the generator of $\mathrm{e}^{tA}\,\mathrm{e}^{tB}$. □

Remark 8. Later, we consider $\Delta A = t^2\,C$. Stressing that - though we never assumed that ΔA is time-dependent - we can cover this case too. This is due to the fact that we are interested in the asymptotic behaviour of $t \to 0$, so that the reasoning as in Lemma 3 will be valid for small enough t, as t will be an upper bound for t^3 - at least in the skew-hermitian case.

4.3 Strang Splitting

There are more advanced schemes available than the Lie Trotter splitting, which can exhibit desirable properties. For instance, symmetrising the Lie-Trotter splitting leads to the so-called Strang splitting, i. e.

$$\mathrm{e}^{H} = \mathrm{e}^{A+B} \approx \mathrm{e}^{\frac{1}{2}A}\mathrm{e}^{B}\mathrm{e}^{\frac{1}{2}A}.$$

The numerical flow Ψ^h is then defined as

$$\Psi_h := \Phi^A_{h/2} \circ \Phi^B_h \circ \Phi^A_{h/2}.$$

Figure 2 depicts the symmetry of the situation in comparison to the Lie-Trotter splitting.

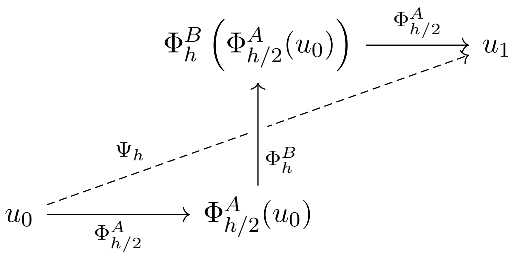


Figure 2: Schematic visualisation of the Strang Splitting.

4.4 Symmetric Baker-Campbell-Hausdorff formula

Analogously, we may look for an expression E as an exponent of the Strang-splitting. We ask for E such that

$$\mathrm{e}^{\frac{1}{2}A}\mathrm{e}^{B}\mathrm{e}^{\frac{1}{2}A} = \mathrm{e}^{E}.$$

The answer will be the symmetrised version of the Baker-Campbell-Hausdorff formula, which is of the same flavour as Theorem 2:

Theorem 3 (Symmetric version of Baker-Campbell-Hausdorff formula). The expression E is called the *symmetric Baker-Campbell-Hausdorff formula* and its first terms are given by

$$E = (A + B) - \frac{1}{24}\left([[B, A], A] + 2\,[[B, A], B]\right) + \dots .$$

Proof. The first few terms are found in [Singh, 2018, Equation (2.33) on page 35]. □

Remark 9. Taking a look at the terms involved, we see (not only in the first terms) that there are only *odd* degrees of nested commutators involved.

Remark 10. Using the already known BCH-formula, we have, using associativity

$$\begin{aligned} \mathrm{e}^{E} = \mathrm{e}^{\mathrm{sBCH}(A,B)} = \mathrm{e}^{\frac{1}{2}A}\mathrm{e}^{B}\mathrm{e}^{\frac{1}{2}A} = \mathrm{e}^{\mathrm{BCH}(\frac{1}{2}A,B)}\mathrm{e}^{\frac{1}{2}A} = \mathrm{e}^{\mathrm{BCH}(\mathrm{BCH}(\frac{1}{2}A,B),\frac{1}{2}A)} \\ = \mathrm{e}^{\frac{1}{2}A}\mathrm{e}^{\mathrm{BCH}(B,\frac{1}{2}A)} = \mathrm{e}^{\mathrm{BCH}(\frac{1}{2}A,\mathrm{BCH}(B,\frac{1}{2}A))}, \end{aligned}$$

hence we have extracted the identity

$$\mathrm{sBCH}(A, B) = \mathrm{BCH}(\mathrm{BCH}(\tfrac{1}{2}A, B), \tfrac{1}{2}A) = \mathrm{BCH}(\tfrac{1}{2}A, \mathrm{BCH}(B, \tfrac{1}{2}A)).$$

Corollary 4. Let A and B be skew-hermitian matrices. Then,

$$\mathrm{e}^{tH} - \mathrm{e}^{\frac{1}{2}tA}\mathrm{e}^{tB}\mathrm{e}^{\frac{1}{2}tA} = \mathrm{e}^{t(A+B)} - \mathrm{e}^{\frac{1}{2}tA}\mathrm{e}^{tB}\mathrm{e}^{\frac{1}{2}tA} = \mathcal{O}(t^3), \quad t \to 0.$$

Proof. As in the previous Corollary 3, little adaptions for the formula of the symmetric BCH formula yield:

$$\|\mathrm{e}^{\frac{1}{2}tA}\mathrm{e}^{tB}\mathrm{e}^{\frac{1}{2}tA} - \mathrm{e}^{tA+tB}\|_2 = \|\mathrm{e}^{\mathrm{sBCH}(tA,tB)} - \mathrm{e}^{tA+tB}\|_2 \le \|\mathrm{sBCH}(tA, tB) - (tA + tB)\|_2 = Ct^3.$$

Further details can be found in [Singh, 2018, p. 35]. □

4.5 Higher order splitting

As we have seen in the previous sections, a vital aspect for higher convergence order of a numerical method is the fact that the exponents need to be approximated with higher order.

Generally, the action of matrix exponentiation is expensive to compute. One can ask if there is a tradeoff possible when using suiting numerical methods. Calculating nested matrix commutators beforehand and saving the matrix exponentials will yield higher convergence order of the numerical schemes - the drawback is that matrix exponentials are not sparse in general. Therefore, memory consumption increases and more exponentials need to be calculated.

A generic approach might be to ask which matrix C should be inserted to enhance convergence order (doing this in a symmetric way due to the fact that only odd terms occur, which will result in a higher increase of order):

$$\mathrm{e}^{\frac{1}{2}tA}\mathrm{e}^{tB}\mathrm{e}^{\frac{1}{2}tA} = \mathrm{e}^{\frac{1}{2}tA}\mathrm{e}^{\frac{1}{2}tB}\mathrm{e}^{\frac{1}{2}tB}\mathrm{e}^{\frac{1}{2}tA} \approx \mathrm{e}^{\frac{1}{2}tA}\mathrm{e}^{\frac{1}{2}tB}\mathrm{e}^{t^3C}\mathrm{e}^{\frac{1}{2}tB}\mathrm{e}^{\frac{1}{2}tA}.$$

Knowing the sBCH formula, we see for skew-hermitian matrices that the choice

$$\begin{aligned} C &= \frac{1}{24}\left([[B,A],A] + 2\,[[B,A],B]\right) \\ &= \frac{1}{24}\left([A,[A,B]] + 2\,[B,[A,B]]\right) \end{aligned} \tag{4.10}$$

with a similar reasoning as in the previous corollaries yields a scheme of order 3 at least. We will reconstruct C up to certain orders with a more practicable and more basic approach in Chapter 7.

5 The 'dual' analogon of BCH - the Zassenhaus formula

Give your son Marvin two sport shirts for a present.
The first time he wears one of them, look at him
sadly and say in your Basic Tone of Voice,
"The other one you didn't like?"
— Dan Greenburg, *How to be a Jewish mother*

Up to now, the focus was on an approximation of $\mathrm{e}^{tH} = \mathrm{e}^{tA}\mathrm{e}^{tB} + \mathcal{O}(t^2)$ for $t \to 0$ for given $H = A + B$. A different, yet 'dual' choice is the *Zassenhaus* splitting. It takes a look at *what is missing* to get an equality in terms of adding further exponentials. A good summary of the results about Zassenhaus splitting using BCH and sBCH is found in [Singh, 2018, Chapter 4].

5.1 Zassenhaus formula in general form

The idea is to start with $\mathrm{e}^{t(A+B)}$ and develop a chain of generators whose Lie degree gradually increases, i. e. with D_k including Lie-elements of degree k. Formally,

$$\mathrm{e}^{tH} = \mathrm{e}^{t(A+B)} = \mathrm{e}^{tA}\,\mathrm{e}^{tB}\,\mathrm{e}^{t^2 D_2}\,\mathrm{e}^{t^3 D_3}\,\cdots\,. \tag{5.1}$$

The order of the approximation is the criterion for the number of exponentials we consider. Notice that its validity is an immediate consequence of the BCH or sBCH formula. Using the BCH formula, we arrive at

$$D_2 = -\tfrac{1}{2}[A,B], \quad D_3 = -\tfrac{1}{6}(2[[A,B],B] + [[A,B],A]), \quad \text{etc.}$$

[Magnus, 1954] offers an algebraic approach on how to derive the equations that lead to the formulas of D_k. It is worth seeing an outline of the main ingredients.

Magnus starts with defining a linear operator performing bracketing from the left, which is called

Definition 17 (Dynkin operator). Let us consider the ring of power series $\mathbb{C}\langle X,Y\rangle$. We define for every element $F(X,Y) \in \mathbb{C}\langle X,Y\rangle$ the Lie element F by the *Dynkin operator* δ. For the *Dynkin operator* δ, the following properties hold true for $F_1, F_2 \in \mathbb{C}\langle X,Y\rangle$ and $c \in \mathbb{C}$:

$$\delta(F_1 + cF_2) := \delta(F_1) + c\delta(F_2) \qquad (\textit{linearity}) \tag{5.2}$$

For any monomial $W_1 \cdots W_n$, where $W_i \in \{A,B\}$ is a generator, we define

$$\delta(W_1 \cdots W_n) := [[\ldots[[W_1,W_2],W_3]\ldots],W_n] \qquad (\textit{left bracketing}) \tag{5.3}$$

$$\delta(\mathbb{1}) := 0 \qquad (\textit{identity}) \tag{5.4}$$

◇

The definition in [Magnus, 1954, p. 662] can be found in [Khukhro and Khoukhro, 1998, p. 70] from a different angle in more modern terms. Notice that the elements of the free Lie algebra $\mathbb{C}\langle X,Y\rangle$ are naturally graded - cf. Definition 16 - and a possible choice for the basis are Lyndon words, which is nontrivial due to the Jacobi identity. For more details, see [Auzinger and Herfort, 2014, p. 247].

We quote - from [Magnus, 1954, p. 662] - the following Lemma.

Lemma 4. Let Y be a homogenous Lie element of degree n[8] and let $Z \in \mathbb{C}\langle X, Y\rangle$ be arbitary. Then

1. $$\delta(Y) = nY,$$

2. $$\delta(Y^2 Z) = 0$$

hold true.

Magnus applies the Dynkin operator δ to equation (5.1), so the left-hand side - using the definition as well as Lemma 4 - yields

$$\delta(\mathrm{e}^{t(A+B)}) = \delta(\mathbb{1}) + t\delta(A+B) + \frac{t^2}{2!}\delta((A+B)^2) + \ldots = t(A+B).$$

The right-hand side looks as follows

$$\begin{aligned}\delta(\mathrm{e}^{tA}\mathrm{e}^{tB}\mathrm{e}^{t^2D_2}\mathrm{e}^{t^3D_3}\cdots) &= \delta\left(\sum_{k=0}^{\infty}\frac{(tA)^k}{k!}\sum_{k=0}^{\infty}\frac{(tB)^k}{k!}\sum_{k=0}^{\infty}\frac{(tD_2)^k}{k!}\cdots\right)\\ &= \underbrace{0}_{\text{degree 0 terms}} + \underbrace{t(A+B)}_{\text{degree 1 terms}} + \underbrace{t^2(\delta(AB)+\delta(D_2))}_{\text{degree 2 terms}} + \ldots.\end{aligned}$$

Comparing coefficients and using the lemma, we arrive at

$$\delta(D_2) = -\delta(AB) = -\frac{1}{2}[A, B].$$

When comparing with the BCH formula, the same procedure yields for D_3

$$D_3 = -\tfrac{1}{6}(2[[A, B], B] + [[A, B], A]).$$

5.2 Symmetric Zassenhaus formula

The same idea can be used for a symmetrised form. The symmetrised version of equation (5.1) is

$$\mathrm{e}^{tH} = \mathrm{e}^{t(A+B)} = \cdots\, \mathrm{e}^{t^5E_5}\, \mathrm{e}^{t^3E_3}\, \mathrm{e}^{\frac{1}{2}tA}\, \mathrm{e}^{tB}\, \mathrm{e}^{\frac{1}{2}tA}\, \mathrm{e}^{t^3E_3}\, \mathrm{e}^{t^5E_5} \cdots, \tag{5.5}$$

where E_{2k+1} are Lie-elements of degree $2k+1$ for $k \in \mathbb{N}$.

Following Magnus' approach, i. e. applying the Dynkin operator δ, we derive an equation for E_3. The left-hand side result is the same as in the nonsymmetric case, so due to Lemma 4 above, it holds

$$\delta(\mathrm{e}^{t(A+B)}) = \delta(\mathbb{1}) + t\delta(A+B) + \frac{t^2}{2!}\delta((A+B)^2) + \ldots = t(A+B).$$

[8] possibly a linear combination of Lie-elements with the same degree n

The right-hand side using the linearity of δ yields

$$\begin{aligned}
&\delta(\cdots \mathrm{e}^{t^5E_5}\,\mathrm{e}^{t^3E_3}\,\mathrm{e}^{\frac{1}{2}tA}\,\mathrm{e}^{tB}\,\mathrm{e}^{\frac{1}{2}tA}\,\mathrm{e}^{t^3E_3}\,\mathrm{e}^{t^5E_5}\,\cdots)\\
&\qquad = \delta\left(\cdots \sum_{k=0}^{\infty}\frac{(t^3E_3)^k}{k!}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}tA\right)^k}{k!}\sum_{k=0}^{\infty}\frac{(tB)^k}{k!}\sum_{k=0}^{\infty}\frac{\left(\frac{1}{2}tA\right)^k}{k!}\sum_{k=0}^{\infty}\frac{(t^3E_3)^k}{k!}\cdots\right)\\
&\qquad = \delta\left(\underbrace{\mathbb{1}}_{\text{degree 0 terms}} + \underbrace{t(\tfrac{1}{2}A + B + \tfrac{1}{2}A)}_{\text{degree 1 terms}} + \underbrace{t^2(\tfrac{1}{2}(AB+BA) + \tfrac{1}{4}(A^2+B^2)}_{\text{degree 2 terms}}\right)\\
&\qquad\quad + \delta\left(\underbrace{t^3\left(2E_3 + \tfrac{1}{4}(ABA + \tfrac{1}{2}A^2B + AB^2 + B^2A + \tfrac{1}{2}BA^2) + \tfrac{1}{6}(A^3+B^3)\right)}_{\text{degree 3 terms}} + \ldots\right),
\end{aligned}$$

where $\ldots$ signifies higher order terms.

As previously seen, the degree 0 term vanishes. The first order term can be found on the left-hand side as well - using the points in Lemma 4, so therefore, the remaining degrees need to vanish too. Taking a look at degree 2, we have

$$\begin{aligned}
\delta(\tfrac{1}{2}(AB+BA) + \tfrac{1}{4}(A^2+B^2)) &= \tfrac{1}{2}(\delta(AB) + \delta(BA)) + \tfrac{1}{4}\delta((A^2\cdot 1 + B^2\cdot 1)\\
&= \underbrace{\tfrac{1}{2}([A,B] + [B,A])}_{=0} + 0 = 0.
\end{aligned}$$

For degree 3, we use the first point of the Lemma 4 for the E_3-term, yielding

$$\delta(2E_3) = 6E_3. \tag{5.6}$$

The second point of Lemma 4 simplifies a great deal, namely

$$\begin{aligned}
&\delta\big(\tfrac{1}{4}(ABA + \tfrac{1}{2}A^2B + AB^2 + B^2A + \tfrac{1}{2}BA^2) + \tfrac{1}{6}(A^3+B^3)\big) = \tfrac{1}{4}\delta\left(ABA + AB^2 + \tfrac{1}{2}BA^2\right)\\
&\qquad = \tfrac{1}{4}\left([[A,B],A] + [A,B],B] + \tfrac{1}{2}[[B,A],A]\right)\\
&\qquad = \tfrac{1}{4}\left([A,B],B] - \tfrac{1}{2}[[B,A],A]\right).
\end{aligned} \tag{5.7}$$

In the last step, we used $[[A,B],A] = -[[B,A],A]$. As (5.6) + (5.7) = 0 due to comparison with the left-hand side, we arrive at

$$E_3 = \tfrac{1}{48}\left(2[[B,A],A] + \tfrac{1}{2}[[B,A],B]\right).$$

This is exactly what the symmetric BCH formula modulo division by 6 (due to the step in equation (5.6)) returns as results in [Singh, 2018, Section 4.1.1, p. 74] reveal.

6 Splitting methods and a basic Taylor approach

"In theory, there is no difference between theory and practice. In practice, there is."

disputed

We start with a special defect representation, which will be the key to the more general approach we need for the reconstruction of (4.10). First, we will consider the Strang splitting as a special case. Many of the ideas originate in an unpublished draft and were extended in private communication with [Auzinger, 2020] and his research group.

6.1 Special representation of defect for symmetric flow

We start with quoting an auxiliary result that shows that symmetric methods can be represented as a composition of a numerical scheme with its adjoint.

Lemma 5. Assume that a symmetric method Ψ_t can be expanded as a power series in t. Then there exists another method $\widehat{\Psi}_t$ such that

$$\Psi_t = \left(\widehat{\Psi}_{t/2} \circ \widehat{\Psi}^*_{t/2}\right)$$

Proof. This Lemma as well as the proof thereof can be found in [Hairer et al., 2006, Lemma V.3.2]. □

The following Lemma allows us to get the inner symmetrised defect $\widehat{\mathcal{D}}(t)$ starting from a symmetric method represented by composition.

Lemma 6. Consider a self-adjoint one-step operator $\mathcal{S} = \mathcal{S}(t)$ of the form [9]

$$\mathcal{S}(t) = \mathcal{T}_*(t)\,\mathcal{T}(t),$$

where $\mathcal{T}_*(t) = \mathcal{T}^{-1}(-t)$ is the adjoint of $\mathcal{T}(t)$.
Then its symmetrised defect $\mathcal{D}_s(t)$ as defined in equation (3.6) can be expressed as

$$\mathcal{D}_s(t) = \mathcal{T}_*(t)\,\widehat{\mathcal{D}}_s(t)\,\mathcal{T}(t), \tag{6.1}$$

with the *inner symmetrised defect* defined as

$$\widehat{\mathcal{D}}_s(t) := \mathcal{T}(-t)\,\mathcal{T}_*'(t) + \mathcal{T}'(t)\,\mathcal{T}_*(-t) - \tfrac{1}{2}\mathcal{T}(-t)\,H\,\mathcal{T}_*(t) - \tfrac{1}{2}\mathcal{T}(t)\,H\,\mathcal{T}_*(-t). \tag{6.2}$$

Moreover, $\widehat{\mathcal{D}}(t)$ is an even function, i. e.

$$\widehat{\mathcal{D}}_s(t) = \widehat{\mathcal{D}}_s(-t).$$

Proof. First, we notice that the method is essentially a composition as stated in Lemma 5.

[9] Notice that we use t instead of $2^{-1}t$, but we allow us this notational inconsistency.

Expanding $\mathcal{D}_s(t)$, using the product rule and proper bracketing, we arrive at

$$\begin{aligned}\mathcal{D}_s(t) &= \mathcal{S}'(t) - \tfrac{1}{2}H\mathcal{S}(t) - \tfrac{1}{2}\mathcal{S}(t)H \\ &= \mathcal{T}_*'(t)\,\mathcal{T}(t) + \mathcal{T}_*(t)\,\mathcal{T}'(t) - \tfrac{1}{2}H\mathcal{T}_*(t)\,\mathcal{T}(t) - \tfrac{1}{2}\mathcal{T}_*(t)\,\mathcal{T}(t)H \\ &= \mathcal{T}_*(t)\left(\mathcal{T}_*^{-1}(t)\,\mathcal{T}_*'(t) + \mathcal{T}'(t)\,\mathcal{T}^{-1}(t) - \tfrac{1}{2}\mathcal{T}_*^{-1}(t)\,H\,\mathcal{T}_*(t) - \tfrac{1}{2}\mathcal{T}(t)\,H\,\mathcal{T}^{-1}(t)\right)\mathcal{T}(t) \\ &=: \mathcal{T}_*(t)\,\widehat{\mathcal{D}}_2(t)\,\mathcal{T}(t).\end{aligned}$$

By self-adjointness assumption, we have

$$\mathcal{T}^{-1}(t) = \mathcal{T}_*(-t) \quad \text{and} \quad \mathcal{T}_*^{-1}(t) = \mathcal{T}(-t). \tag{6.3}$$

So, we see

$$\widehat{\mathcal{D}}_s(t) = \mathcal{T}(-t)\,\mathcal{T}_*'(t) + \mathcal{T}'(t)\,\mathcal{T}_*(-t) - \tfrac{1}{2}\mathcal{T}(-t)\,H\,\mathcal{T}_*(t) - \tfrac{1}{2}\mathcal{T}(t)\,H\,\mathcal{T}_*(-t),$$

meaning the inner symmetrised defect $\widehat{\mathcal{D}}_s(t)$ is indeed of the form (6.2) as claimed.

Furthermore, differentiating the self-adjointness assumption (6.3), we end up with

$$0 = \tfrac{\mathrm{d}}{\mathrm{d}\,t}\big(\mathcal{T}(-t)\,\mathcal{T}_*(t)\big) = -\mathcal{T}'(-t)\,\mathcal{T}_*(t) + \mathcal{T}(-t)\,\mathcal{T}_*'(t),$$

which implies

$$\widehat{\mathcal{D}}_s(t) = \mathcal{T}'(-t)\,\mathcal{T}_*(t) + \mathcal{T}'(t)\,\mathcal{T}_*(-t) - \tfrac{1}{2}\mathcal{T}(-t)\,H\,\mathcal{T}_*(t) - \tfrac{1}{2}\mathcal{T}(t)\,H\,\mathcal{T}_*(-t),$$

whence $\widehat{\mathcal{D}}_s(t) = \widehat{\mathcal{D}}_s(-t)$, as asserted. This concludes the proof. □

Remark 11. Remarkably, there is a connex between Lemma 6 and Theorem 1 for the linear case. Namely, it implies

$$\begin{aligned}\mathcal{S}(-t)\,\mathcal{D}_s(t) &= \mathcal{T}_*(-t)\,\underbrace{\mathcal{T}(-t)\,\mathcal{T}_*(t)}_{=\mathbb{1}}\,\widehat{\mathcal{D}}(t)\,\mathcal{T}(t), \\ \mathcal{D}_s(-t)\,\mathcal{S}(t) &= \mathcal{T}_*(-t)\,\underbrace{\widehat{\mathcal{D}}(-t)}_{=\widehat{\mathcal{D}}(t)}\,\underbrace{\mathcal{T}(-t)\,\mathcal{T}_*(t)}_{=\mathbb{1}}\,\mathcal{T}(t).\end{aligned}$$

We easily arrive at the vital identity used in the proof of Theorem 1, i. e.

$$\mathcal{S}(-t)\,\mathcal{D}_s(t) = \mathcal{D}_s(-t)\,\mathcal{S}(t).$$

6.2 Symmetrised defect for Strang splitting

In this section, we will develop a defect representation for Strang splitting, which is so-to-say the basic case of the more general approach taken later on. The Strang splitting reads

$$\mathcal{S} = \mathcal{S}(t) = \mathrm{e}^{\frac{1}{2}tA}\,\mathrm{e}^{tB}\,\mathrm{e}^{\frac{1}{2}tA}, \quad \mathcal{S}(0) = \mathbb{1},$$

and[10]

$$\begin{aligned}\mathcal{T} &= \mathcal{T}(t) = \mathrm{e}^{\frac{1}{2}tB}\,\mathrm{e}^{\frac{1}{2}tA}, \quad \mathcal{T}(0) = \mathbb{1}, \\ \mathcal{T}_* &= \mathcal{T}_*(t) = \mathrm{e}^{\frac{1}{2}tA}\,\mathrm{e}^{\frac{1}{2}tB}, \quad \mathcal{T}_*(0) = \mathbb{1}.\end{aligned}$$

[10] We see that we could use $2^{-1}t$ as the argument too.

First, we observe that

$$\mathcal{S} = \mathcal{T}_*\,\mathcal{T}.$$

Next, we notice that the derivatives of $\mathcal{T}$ and $\mathcal{T}_*$ satisfy homogeneous Sylvester-type equations as a simple calculation reveals:

$$\mathcal{T}' = \mathcal{T}\,(\tfrac{1}{2}A) + (\tfrac{1}{2}B)\,\mathcal{T} \tag{6.4}$$

and

$$\mathcal{T}_*' = (\tfrac{1}{2}A)\,\mathcal{T}_* + \mathcal{T}_*\,(\tfrac{1}{2}B). \tag{6.5}$$

Hence, we get the following expression for $\mathcal{S}'(t)$ using (6.4) and (6.5)

$$\begin{aligned}
\mathcal{S}' &= (\mathcal{T}_*\,\mathcal{T})' = \mathcal{T}_*'\,\mathcal{T} + \mathcal{T}_*\,\mathcal{T}' \\
&= \big(\tfrac{1}{2}\,A\,\mathcal{T}_* + \mathcal{T}_*\,\tfrac{1}{2}\,A\big)\,\mathcal{T} + \mathcal{T}_*\,\big(\mathcal{T}\,\tfrac{1}{2}\,A + \tfrac{1}{2}\,A\,\mathcal{T}\big) \\
&= \tfrac{1}{2}\,A\,\mathcal{T}_*\,\mathcal{T} + \mathcal{T}_*\,\tfrac{1}{2}\,A\,\mathcal{T} + \mathcal{T}_*\,\mathcal{T}\,\tfrac{1}{2}\,A + \mathcal{T}_*\,\tfrac{1}{2}\,A\,\mathcal{T} \\
&= \mathcal{T}_*\,B\,\mathcal{T} + \tfrac{1}{2}\{\mathcal{S}, A\}.
\end{aligned}$$

For the symmetrised defect $\mathcal{D}_s$ as in Definition 13, we use $\mathcal{T}_*\,B\,\mathcal{T} = (\tfrac{1}{2} + \tfrac{1}{2})\,\mathcal{T}_*\,B\,\mathcal{T}$ and the definition of the commutator resp. anticommutator to get

$$\begin{aligned}
\mathcal{D}_s &= \mathcal{S}' - \tfrac{1}{2}\{\mathcal{S}, H\} \\
&= \cancel{(\tfrac{1}{2}A)\,\mathcal{T}_*\,\mathcal{T}} + (\tfrac{1}{2} + \tfrac{1}{2})\,\mathcal{T}_*\,B\,\mathcal{T} + \cancel{\mathcal{T}_*\,\mathcal{T}\,(\tfrac{1}{2}A)} - \tfrac{1}{2}\mathcal{T}_*\,\mathcal{T}\,(\cancel{A} + B) - \tfrac{1}{2}(\cancel{A} + B)\,\mathcal{T}_*\,\mathcal{T} \\
&= \tfrac{1}{2}[\mathcal{T}_*, B]\,\mathcal{T} - \tfrac{1}{2}\,\mathcal{T}_*\,[\mathcal{T}, B] \\
&= \tfrac{1}{2}\,(\mathcal{T}_*\,B\,\mathcal{T} - B\,\mathcal{T}_*\,\mathcal{T} - \mathcal{T}_*\,\mathcal{T}\,B + \mathcal{T}_*\,B\,\mathcal{T}) \\
&= \mathcal{T}_*\,\big(B - \tfrac{1}{2}\mathcal{T}_*^{-1}\,B\,\mathcal{T}_* - \tfrac{1}{2}\mathcal{T}\,B\,\mathcal{T}^{-1} + B\big)\mathcal{T} \\
&=: \mathcal{T}_*\,\widehat{\mathcal{D}}_s\,\mathcal{T}.
\end{aligned} \tag{6.6}$$

We may call $\widehat{\mathcal{D}}_s(t)$ the *inner symmetrised defect* for Strang splitting. Connecting all loose ends with the self-adjointness assumption (6.3), we arrive at

$$\begin{aligned}
\widehat{\mathcal{D}}_s(t) :&= B - \tfrac{1}{2}\big(\mathcal{T}_*^{-1}(t)\,B\,\mathcal{T}_*(t) + \mathcal{T}(t)\,B\,\mathcal{T}^{-1}(t)\big) \\
&= B - \tfrac{1}{2}\big(\mathcal{T}(-t)\,B\,\mathcal{T}_*(t) + \mathcal{T}(t)\,B\,\mathcal{T}_*(-t)\big) \\
&= \widehat{\mathcal{D}}_s(-t)
\end{aligned} \tag{6.7}$$

and notice that the inner symmetrised defect $\widehat{\mathcal{D}}_s(t)$ is indeed even.
Now we introduce a convenient new variable $\Xi(t)$ defined as

$$\Xi(t) := \mathrm{e}^{\frac{1}{2}tB}\,\big(\mathrm{e}^{\frac{1}{2}tA}B\,\mathrm{e}^{-\frac{1}{2}tA}\big)\,\mathrm{e}^{-\frac{1}{2}tB} = \mathcal{T}(t)\,B\,\mathcal{T}_*(-t), \tag{6.8}$$

and thus for the inner symmetrised defect $\widehat{\mathcal{D}}_s$, we get

$$\widehat{\mathcal{D}}_s(t) = B - \tfrac{1}{2}\big(\Xi(t) + \Xi(-t)\big), \quad \widehat{\mathcal{D}}_s(0) = 0. \tag{6.9}$$

6.3 Derivations, ad, Ad and the Lie expansion formula

This section deals with new vocabulary which turns out to be helpful. The following definition is found in [Schuster, 2018, p. 5].

Definition 18 (Derivation). Let A be an algebra over $\mathbb{C}$, i. e. a vectorspace over $\mathbb{C}$ endowed with a bilinear product $\cdot\colon A \times A \to A, (M, N) \mapsto M \cdot N =: MN$. A derivation D is a linear map $D\colon A \to A$ such that for all $M, N \in A$

$$D(MN) = MD(N) + D(M)N \tag{6.10}$$

holds true. ◇

In the literature, nested commutators are sometimes denoted with ad.

Definition 19 (ad-notation). Let us consider the nested commutators of U and X as in Definition 4. We define

$$\begin{aligned} &\mathrm{ad}_U^0(X) := [U, X]_0 = X,\\ &\mathrm{ad}_U^1(X) := [U, X]_1 = [U, X],\\ \text{and generally,}\quad &\mathrm{ad}_U^j(X) := [U, X]_j = [U, [U, X]_{j-1}] = [U, \mathrm{ad}_U^{j-1}(X)], \quad j \in \mathbb{N}. \end{aligned}$$

◇

Remark 12. The ad-notation stems from Lie theory. Let $\mathfrak{g}$ be a Lie algebra over $\mathbb{C}$. For every $Y \in \mathfrak{g}$, the map

$$\mathrm{ad}_Y(X)\colon \mathfrak{g} \to \mathfrak{g}, X \mapsto \mathrm{ad}_Y(X) := [Y, X], \quad X \in \mathfrak{g}$$

defines a linear map. Thanks to the Jacobi identity, we can easily see that ad_Y is a derivation as

$$\mathrm{ad}_{[U,V]}(X) = [[U, V], X] \overset{\text{Jacobi}}{=} [U, [V, X]] - [V, [U, X]] = [U, \mathrm{ad}_V(X)] + [\mathrm{ad}_U(X), V].$$

In this case, we notice that $D(\cdot)(X) := \mathrm{ad}_\cdot(X)$ is a derivation using linearity. This remark is provided by [Schuster, 2018, p. 6].

Remark 13 (Derivation and inductive structure yields Leibniz). As we know that ad_Y is a derivation due to Remark 12, we see that the differential operator $\frac{\mathrm{d}}{\mathrm{d}t}$ is a derivation too in the setting given in equation (6.11a). As [Schuster, 2018, p. 9] points out, the Leibniz rule is an immediate consequence of an inductive argument.

The next definition is found in [Hall, 2003, Definition 2.22].

Definition 20 (Ad). Let G be a matrix Lie group with Lie Algebra $\mathfrak{g}$. Then, we define, for fixed $A \in G$, the linear map $\mathrm{Ad}_A\colon \mathfrak{g} \to \mathfrak{g}$ by

$$\mathrm{Ad}_A(X) := A\, X\, A^{-1}.$$

◇

Remark 14. In general, Lie theory - see [Hall, 2003, Theorem 2.21] shows in an abstract manner that the aforementioned map ad exists as a restriction of Ad once we are sure that Ad_A is a Lie homomorphism. Notice that Ad - the conjugation - plays the same role in the Lie group as the commutator ad does in the Lie algebra.

The following Lie expansion formula is established by [Hall, 2003, Theorem 2.21, (3)]. The next finding is key to the Taylor expansion of $\Xi(t)$.

Lemma 7 (Lie expansion formula). The t-dependent exponential

$$\mathrm{e}^{\mathrm{ad}_{tU}}(X) = \mathrm{e}^{t\,U} X\, \mathrm{e}^{-t\,U}$$

satisfies

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{\mathrm{ad}_{tU}}(X) = \mathrm{ad}_U\big(\mathrm{e}^{\mathrm{ad}_{tU}}(X)\big). \tag{6.11a}$$

More generally, for all $q \in \mathbb{N}_0$,

$$\tfrac{\mathrm{d}^q}{\mathrm{d}\,t^q}\,\mathrm{e}^{\mathrm{ad}_{tU}}(X) = \mathrm{ad}_U^q\big(\mathrm{e}^{\mathrm{ad}_{tU}}(X)\big), \tag{6.11b}$$

$$\tfrac{\mathrm{d}^q}{\mathrm{d}\,t^q}\,\mathrm{e}^{\mathrm{ad}_{tU}}(X)\big|_{t=0} = \mathrm{ad}_U^q(X).$$

Those time derivatives are the main ingredients for the Taylor expansion of $\mathrm{e}^{\mathrm{ad}_{tU}}(X)$.

Proof. This proof is found in [Hall, 2003, Theorem 2.21], but in much clearer form in the second edition [Hall, 2015, Theorem 3.34]. The first identity (6.11a) is obtained by differentiation:

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{\mathrm{ad}_{tU}}(X) = U\mathrm{e}^{\mathrm{ad}_{tU}}(X) - \mathrm{e}^{\mathrm{ad}_{tU}}(X)\,U = [U, \mathrm{e}^{\mathrm{ad}_{tU}}(X)] = \mathrm{ad}_U\big(\mathrm{e}^{\mathrm{ad}_{tU}}(X)\big).$$

The proof of equation (6.11b) works by straightforward induction after noticing that ad_U is time-independent. □

6.4 Taylor expansion of symmetrised defect $\widehat{\mathcal{D}}_s(t)$

We are ready to address the Taylor expansion of $\Xi(t)$ as defined in equation (6.8). First, we rewrite (6.8) using ad-notation and find

$$\Xi(t) = \mathrm{e}^{\frac{1}{2}tB}\big(\mathrm{e}^{\frac{1}{2}tA} B\, \mathrm{e}^{-\frac{1}{2}tA}\big)\,\mathrm{e}^{-\frac{1}{2}tB} = \Big(\mathrm{e}^{\mathrm{ad}_{\frac{1}{2}tB}} \circ \mathrm{e}^{\mathrm{ad}_{\frac{1}{2}tA}}\Big)(B). \tag{6.12}$$

[Auzinger, 2018, Appendix B] introduced the following notation in order to see if Remark 13 applies to the product of exponentials in (6.12). Notice that Remark 13 does not state that $\mathrm{e}^{\mathrm{ad}_{t\,U}}(X)$ is a derivation. We will prove this directly as an statement on the inheritance of derivations in series is not at hand.

Definition 21. Define

$$f_{p,U}(t)(X) := (\mathrm{ad}_U^p \circ \mathrm{e}^{\mathrm{ad}_{tU}})(X) = [U, \mathrm{e}^{tU} X \mathrm{e}^{-tU}]_p.$$

Furthermore, let Δ be the differential operator acting on objects of the type $f_{p,U}$ by

$$(\Delta f_{p,U})(t)(X) = f_{p+1,U}(t)(X) = (\mathrm{ad}_U \circ f_{p,U})(t)(X) = \tfrac{\mathrm{d}}{\mathrm{d}\,t} f_{p,U}(t)(X).$$

◇

We observe that $f_{p,U}$ is motivated by (6.11a). The next Lemma shows that Δ is indeed a derivation.

Lemma 8 (Δ is a derivation). This proof was written by [Auzinger, 2018]. According to Definition 21, Δ is a *derivation*, i.e.,

$$\begin{aligned}\Delta\big(f_{q,V}\circ f_{p,U}\big) &= f_{q+1,V}\circ f_{p,U} + f_{q,V}\circ f_{p+1,U}\\ &= \Delta f_{q,V}\circ f_{p,U} + f_{q,V}\circ \Delta f_{p,U}.\end{aligned} \tag{6.13}$$

Proof. We start with the left-hand-side to see

$$\begin{aligned}\Delta\big(f_{q,V}\circ f_{p,U}\big)(X)(t) &= \tfrac{\mathrm{d}}{\mathrm{d}\,t}\big(\big(f_{q,V}\circ f_{p,U}\big)(X)(t)\big)\\ &= \tfrac{\mathrm{d}}{\mathrm{d}\,t}\,[V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}]_q\\ &= \big[V,\tfrac{\mathrm{d}}{\mathrm{d}\,t}\underbrace{\big(\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}\big)}_{=:\mathcal{W}(t)}\big]_q.\end{aligned}$$

Further calculations reveal

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}\,t}\,\mathcal{W}(t) &= V\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV} + \mathrm{e}^{tV}\tfrac{\mathrm{d}}{\mathrm{d}\,t}\big([U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\big)\mathrm{e}^{-tV} - \mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}V\\ &= [V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}] + \mathrm{e}^{tV}\big[U,\tfrac{\mathrm{d}}{\mathrm{d}\,t}(\mathrm{e}^{tU}X\,\mathrm{e}^{-tU})]_p\big]\mathrm{e}^{-tV}\\ &= [V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}] + \mathrm{e}^{tV}[U,[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p]\mathrm{e}^{-tV}\\ &= [V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}] + \mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_{p+1}\,\mathrm{e}^{-tV}.\end{aligned}$$

In total, we see that

$$\begin{aligned}\Delta\big(f_{q,V}\circ f_{p,U}\big)(X)(t) &= [V,[V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}]]_q + [V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_{p+1}\,\mathrm{e}^{-tV}]_q\\ &= [V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_p\,\mathrm{e}^{-tV}]_{q+1} + [V,\mathrm{e}^{tV}[U,\mathrm{e}^{tU}X\,\mathrm{e}^{-tU}]_{p+1}\,\mathrm{e}^{-tV}]_q,\end{aligned}$$

whence the claim (6.13) is proven. ☐

Δ is indeed a derivation and the next corollary is immediate.

Corollary 5 (Leibniz rule for Δ). Let $q\in\mathbb{N}_0$. The following holds true:

$$\tfrac{\mathrm{d}^q}{\mathrm{d}\,t^q}\big(\mathrm{e}^{\mathrm{ad}_{tV}}\circ\mathrm{e}^{\mathrm{ad}_{tU}}\big)(X) = \sum_{j=0}^{q}\binom{q}{j}\,\big(\mathrm{ad}_V^j\circ\mathrm{e}^{\mathrm{ad}_{tV}}\big)\circ\,\big(\mathrm{ad}_U^{q-j}\circ\mathrm{e}^{\mathrm{ad}_{tU}}\big)(X). \tag{6.14a}$$

$$\begin{aligned}\tfrac{\mathrm{d}^q}{\mathrm{d}\,t^q}\big(\mathrm{e}^{\mathrm{ad}_{tV}}\circ\mathrm{e}^{\mathrm{ad}_{tU}}\big)(X)\big|_{t=0} &= \sum_{j=0}^{q}\binom{q}{j}\,\big(\mathrm{ad}_V^j\circ\mathrm{ad}_U^{q-j}\big)(X)\\ &= \sum_{j=0}^{q}\binom{q}{j}\,[V,[U,X]_{q-j}]_j.\end{aligned} \tag{6.14b}$$

Proof. This corollary was found in [Auzinger, 2018]. We get the Leibniz rule by simple induction as noticed in Remark 13. The last equation switches to the nested commutator notation and the simplification due to the evaluation at $t=0$ is clear as

$$\mathrm{e}^{\mathrm{ad}_{t\,U}}\Big|_{t=0} = \mathbb{1},$$

☐

Now we are ready for calculating the Taylor expansion of $\Xi(t)$ as seen in (6.12).

Lemma 9 (Taylor expansion $\Xi(t)$). The Taylor expansion of $\Xi(t)$ is given by

$$\Xi(t) = \sum_{q=0}^{7} \tfrac{1}{q!}\, t^q\, 2^{-q} \sum_{k=0}^{q-1} \binom{q}{k} [B,[A,B]_{q-k}]_k + \mathcal{O}(t^8). \tag{6.15}$$

Proof. This Lemma is also found in [Auzinger, 2018]. We begin by remarking that $\mathrm{e}^{\mathrm{ad}_{1/2\,t\,U}}$ changes the recurrence (6.11a) by a factor 2^{-1}, namely

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{\mathrm{ad}_{1/2\,t\,U}}(X) = 2^{-1}\,\mathrm{ad}_U\big(\mathrm{e}^{\mathrm{ad}_{1/2\,t\,U}}(X)\big).$$

By induction, we get the factor 2^{-q}, as we take q derivatives. Notice that the factor 2 does not affect the core of the argument of both Lemma 6.13 and Lemma 5.
The Taylor expansion of $\Xi(t)$ around $t = 0$ looks as follows

$$\Xi(t) = \sum_{q=0}^{7} \tfrac{1}{q!}\, t^q\, \Xi^{(q)}(0) + \mathcal{O}(t^8).$$

The statement about evaluation in (6.14b) for $U = A$ and $V = X = B$ introduces the sum and the binomial coefficient with the additional factor 2^{-k} for $k = 0, \ldots, 7$.

We notice that the q-th summand vanishes as $[B,[A,B]_0] = [B,B] = 0$ and hence, the claim is proven. □

What is still missing is the error expansion of $\widehat{\mathcal{D}}_s(t)$ in case of the Strang splitting, see (6.9). This is the content of the following Lemma.

Lemma 10 (Error expansion for Strang). Let $\widehat{\mathcal{D}}_s(t)$ be given as in (6.8), namely

$$\widehat{\mathcal{D}}_s(t) = B - \tfrac{1}{2}\big(\Xi(t) + \Xi(-t)\big).$$

Then the error expansion for the symmetrised defect $\widehat{\mathcal{D}}_s(t)$ in case of the Strang splitting is given as

$$\widehat{\mathcal{D}}_s(t) = -\sum_{q=2,\ q\ \text{even}}^{6} \tfrac{1}{q!}\, t^q\, 2^{-q} \sum_{k=0}^{q-1} \binom{q}{k} [B,[A,B]_{q-k}]_k + \mathcal{O}(t^8), \tag{6.16}$$

where the $\mathcal{O}(t^8)$ remainder depends on integrals involving terms of the form $\big[B, \Xi_{[A,B]_{8-k}}(t)\big]_k$ for $k = 0 \ldots 8$.

Proof. This is a simple corollary of Lemma 9. The odd terms cancel out as they are antisymmetric and the even terms remain. □

Let us recapitulate: First, we calculated a special representation for the symmetrised defect $\mathcal{D}_s(t)$ where we got a general formula for the inner symmetrised defect $\widehat{\mathcal{D}}(t)$. Next, we turned toward the Strang splitting and calculated its inner symmetrised defect $\widehat{\mathcal{D}}_s(t)$ as a function of $\Xi(t)$. After introducing derivations, we have seen that the differential operator $\frac{\mathrm{d}}{\mathrm{d}\,t}$ is a derivation and hence by induction, we extracted the Leibniz rule. The Leibniz rule allowed us to write down the Taylor expansion of $\Xi(t)$. Finally, we were able to explicitly write down the Taylor expansion of the inner symmetrised defect $\widehat{\mathcal{D}}_s(t)$ of the Strang splitting. This Taylor expansion will appear in the next section, but a more general approach needs to be considered.

7 Taylor expansion for the generalised approach

ALPHA: *Adequate for what?*
SIGMA: *Adequate for the solution of the problem we want to solve.*
THETA*[aside]: Pragmatism! Has everybody lost interest in* truth?
— Imre Lakatos, *Proofs and Refutations*

We have already dealt with the Taylor expansion for the Strang splitting. Now, we are interested in deriving the Taylor expansion of a more general approach where the factor $t^2\,C$. So, formally we take a look at the problem

$$\mathcal{S}(t) := \mathrm{e}^{\frac{1}{2}\,t\,A}\mathrm{e}^{\frac{1}{2}\,t\,B}\mathrm{e}^{t^3\,C}\mathrm{e}^{\frac{1}{2}\,t\,B}\mathrm{e}^{\frac{1}{2}\,t\,A}, \quad \mathcal{S}(t) = \mathbb{1}.$$

We will reproduce the order condition in Section 4.5

$$C = \tfrac{1}{24}\,([A,[A,B] + 2\,[B,[A,B])$$

by means of an explicit formula for the Taylor expansion of the problem considered above.

7.1 Bell polynomials and Faà di Bruno's formula

Starting with preliminaries, we introduce the Bell polynomials. These occur when considering higher derivatives of $f(g(t))$.

Theorem 4 (Partial Bell polynomials). The partial Bell polynomials are defined as

$$B_{q,k}(x_1, x_2, \ldots, x_{q-k-1}) := \sum \frac{q!}{c_1!c_2!\cdots(1!)^{c_1}(2!)^{c_2}\cdots}\,(x_1)^{c_1}(x_2)^{c_2}\cdots(x_{n-k+1})^{c_{n-k+1}}, \tag{7.1a}$$

with summation over all nonnegative integers $c_i \geq 0$ such that

$$\begin{aligned} c_1 + 2\,c_2 + 3\,c_3 + \ldots + (n-k+1)\,c_{n-k+1} &= q, \\ c_1 + c_2 + c_3 + \ldots + c_{n-k+1} &= k. \end{aligned} \tag{7.1b}$$

The Bell polynomials have integer coefficients, are homogeneous of degree $k \in \mathbb{N}_0$, and are of weight $q \in \mathbb{N}_0$.

Proof. [Comtet, 1974, p.133f] defines the Bell polynomials differently but shows in [Comtet, 1974, Theorem A on p. 134] that their explicit expressions are of the form as given in (7.1).[11] The claims are covered by Theorem A, too. □

Let us take a look at higher order derivatives of $f(g(t))$. The first derivative is well known as the chain rule. The formula for the generalised chain rule is named

[11] Homogeneity and weight are defined by (7.1b). Since we are only interested in the explicit formula for the Bell polynomials, we use the explicit formula as a defintion. The statement of the Theorem appears sound in the context Comtet introduces the Bell polynomials.

Theorem 5 (Faà di Bruno identity). For sufficiently smooth real-valued functions, we have

$$\frac{\mathrm{d}^q}{\mathrm{d}t^q}\, f(g(t)) = \sum_{k=1}^{q} f^{(k)}(g(t)) \cdot B_{q,k}\left(g'(t), g''(t), \ldots, g^{(q-k+1)}(t)\right), \tag{7.2}$$

with the partial Bell polynomials defined by (7.1).

Proof. The theorem is stated and proven for instance in [Comtet, 1974, p.139] or in a more modern fashion in [Krantz and Parks, 2002, p.17]. □

Remark 15 (Complete Bell polynomials). In literature - see [Comtet, 1974, Section 3.3, p. 134], the complete Bell polynomials are defined as

$$B_q(x_1, \ldots, x_q) := \sum_{k=0}^{q} B_{q,k}(x_1, x_2, \ldots, x_{q-k-1}).$$

The Faà di Bruno Theorem is closely related where the derivatives $f^{(k)}(g(t))$ can be understood as weights.

7.2 Symmetrised defect for generalised approach

We would like to follow the procedure already used for Strang splitting. The first thing to do is to check out the assumptions of Lemma 6 to establish well-definedness of the symmetrised defect.

7.2.1 Well-definedness of symmetrised defect

As in the Strang case, where

$$\mathcal{T} = \mathcal{T}(t) = \mathrm{e}^{\frac{1}{2}tB}\,\mathrm{e}^{\frac{1}{2}tA}, \quad \mathcal{T}(0) = \mathbb{1},$$
$$\mathcal{T}_* = \mathcal{T}_*(t) = \mathrm{e}^{\frac{1}{2}tA}\,\mathrm{e}^{\frac{1}{2}tB}, \quad \mathcal{T}_*(0) = \mathbb{1},$$

we introduce a new variable

$$\mathcal{Z} := \mathcal{Z}(t) := \mathrm{e}^{t^3\,C}, \quad \mathcal{Z}(0) = \mathbb{1}, \tag{7.3}$$

to arrive at

$$\mathcal{S}(t) = \mathcal{T}_*(t)\,\mathcal{Z}(t)\,\mathcal{T}(t) = \mathcal{T}_*(t)\mathcal{Z}^{\frac{1}{2}}(t)\mathcal{Z}^{\frac{1}{2}}(t)\mathcal{T}(t) =: \mathcal{R}_*(t)\,\mathcal{R}(t). \tag{7.4}$$

A short calculation reveals that $\mathcal{R}_*(t)$ is no abuse of notation. Using the definition of $\mathcal{Z}(t)$ and $\mathcal{T}_*(t) = \mathcal{T}(-t)^{-1}$, we see

$$\mathcal{R}_*(t) = \mathcal{T}_*(t)\,\mathcal{Z}(t) = \left(\mathcal{Z}^{-\frac{1}{2}}(t)\,\mathcal{T}_*(t)^{-1}\right)^{-1} = \left(\mathcal{Z}^{\frac{1}{2}}(-t)\,\mathcal{T}(-t)^{-1}\right)^{-1} = \mathcal{R}(-t)^{-1},$$

whence Lemma 6 guarantees the well-definedness of $\widehat{\mathcal{D}}_s(t)$ as in equation (6.2)

$$\mathcal{D}_s(t) = \mathcal{R}_*(t)\,\widehat{\mathcal{D}}_s(t)\,\mathcal{R}(t), \quad \mathcal{S}(t) = \mathbb{1}.$$

We now aim at constructing $\widehat{\mathcal{D}}_s$ explicitly.

7.2.2 Explicit representation of symmetric defect

It is convenient to start by looking at the derivative of $\mathcal{Z} = \mathcal{Z}(t)$,

$$\mathcal{Z}' = (3\,t^2\,C)\,\mathcal{Z} = \mathcal{Z}\,(3\,t^2\,C) = \tfrac{3}{2}t^2\,\{\mathcal{Z}, C\}. \tag{7.5}$$

Recalling the symmetric defect $\mathcal{D}_s$ according to Definition 13, we have

$$\begin{aligned}\mathcal{S}' &= \mathcal{T}_*'\,\mathcal{Z}\,\mathcal{T} + \mathcal{T}_*\,\mathcal{Z}\,\mathcal{T}' + \mathcal{T}_*\,\mathcal{Z}'\,\mathcal{T}\\ &= \big((\tfrac{1}{2}A)\,\mathcal{T}_* + \mathcal{T}_*\,(\tfrac{1}{2}B)\big)\mathcal{Z}\mathcal{T} + \tfrac{3}{2}t^2\,\mathcal{T}_*\{\mathcal{Z}, C\}\mathcal{T} + \mathcal{T}_*\mathcal{Z}\big(\mathcal{T}\,(\tfrac{1}{2}A) + (\tfrac{1}{2}B)\,\mathcal{T}\big),\end{aligned}$$

and

$$\tfrac{1}{2}\{\mathcal{S}, H\} = \tfrac{1}{2}\mathcal{T}_*\mathcal{Z}\mathcal{T}(A+B) + \tfrac{1}{2}(A+B)\mathcal{T}_*\mathcal{Z}\mathcal{T}.$$

Together with equation (3.6), we obtain

$$\begin{aligned}\mathcal{D}_s &= \mathcal{S}'(t) - \tfrac{1}{2}\{\mathcal{S}, H\}\\ &= \big(\cancel{(\tfrac{1}{2}A)\,\mathcal{T}_*} + \mathcal{T}_*\,(\tfrac{1}{2}B)\big)\mathcal{Z}\mathcal{T} + \tfrac{3}{2}t^2\,\mathcal{T}_*\{\mathcal{Z}, C\}\mathcal{T} + \mathcal{T}_*\mathcal{Z}\big(\cancel{\mathcal{T}\,(\tfrac{1}{2}A)} + (\tfrac{1}{2}B)\,\mathcal{T}\big)\\ &\quad - \tfrac{1}{2}\mathcal{T}_*\mathcal{Z}\mathcal{T}(\cancel{A} + B) + \tfrac{1}{2}(\cancel{A} + B)\mathcal{T}_*\mathcal{Z}\mathcal{T}\\ &= \tfrac{1}{2}\big([\mathcal{T}_*, B]\,\mathcal{Z}\,\mathcal{T} - \mathcal{T}_*\,\mathcal{Z}\,[\mathcal{T}, B]\big) + \tfrac{3}{2}t^2\,\mathcal{T}_*\{\mathcal{Z}, C\}\mathcal{T}.\end{aligned} \tag{7.6}$$

As we attempt to extract $\widehat{\mathcal{D}}_s$ from the equation

$$\mathcal{D}_s = \mathcal{T}_*\,\mathcal{Z}^{\frac{1}{2}}\widehat{\mathcal{D}}_s\,\mathcal{Z}^{\frac{1}{2}}\mathcal{T},$$

we proceed from equation (7.6) to see that

$$\begin{aligned}2\,\mathcal{D}_s &= [\mathcal{T}_*, B]\,\mathcal{Z}\,\mathcal{T} - \mathcal{T}_*\,\mathcal{Z}\,[\mathcal{T}, B] + 3\,t^2\,\mathcal{T}_*\{\mathcal{Z}, C\}\mathcal{T}\\ &= \mathcal{T}_*\,B\,\mathcal{Z}\,\mathcal{T} - B\,\mathcal{T}_*\,\mathcal{Z}\,\mathcal{T} - \mathcal{T}_*\,\mathcal{Z}\,\mathcal{T}\,B + \mathcal{T}_*\,\mathcal{Z}\,B\,\mathcal{T} + 3\,t^2\mathcal{T}_*\{\mathcal{Z}, C\}\mathcal{T}\\ &= \mathcal{T}_*\,\big(B\,\mathcal{Z} - \mathcal{T}_*^{-1}B\,\mathcal{T}_*\,\mathcal{Z} - \mathcal{Z}\,\mathcal{T}\,B\,\mathcal{T}^{-1} + \mathcal{Z}\,B + 3\,t^2\{\mathcal{Z}, C\}\big)\mathcal{T}\\ &= \mathcal{T}_*\mathcal{Z}^{\frac{1}{2}}\Big(\mathcal{Z}^{-\frac{1}{2}}B\,\mathcal{Z}^{\frac{1}{2}} - \mathcal{Z}^{-\frac{1}{2}}\mathcal{T}_*^{-1}B\,\mathcal{T}_*\,\mathcal{Z}^{\frac{1}{2}} - \mathcal{Z}^{\frac{1}{2}}\,\mathcal{T}\,B\,\mathcal{T}^{-1}\mathcal{Z}^{-\frac{1}{2}}\\ &\qquad + \mathcal{Z}^{\frac{1}{2}}B\,\mathcal{Z}^{-\frac{1}{2}} + 3\,t^2\,\underbrace{\mathcal{Z}^{-\frac{1}{2}}\{\mathcal{Z}, C\}\mathcal{Z}^{-\frac{1}{2}}}_{=\,2\,C}\Big)\mathcal{Z}^{\frac{1}{2}}\mathcal{T}.\end{aligned}$$

In the last step, we used that $\mathcal{Z}^{-\frac{1}{2}}\{\mathcal{Z}, C\}\mathcal{Z}^{-\frac{1}{2}} = 2\,C$ as $\mathcal{Z}$ is a function of C and hence commutes freely - whence the anticommutator yields the factor 2.

By the self-adjointness assumption as in $\mathcal{T}^{-1}(t) = \mathcal{T}_*(-t)$ as well as in $\mathcal{Z}(t)^{-\frac{1}{2}} = \mathcal{Z}(-t)^{\frac{1}{2}}$, it is easily seen that $\widehat{\mathcal{D}}_s(t)$ is an even function as stated by Lemma 6:

$$\begin{aligned}2\,\widehat{\mathcal{D}}_s(t) &= \big(\mathcal{Z}(t)^{-\frac{1}{2}}B\,\mathcal{Z}(t)^{\frac{1}{2}} + \mathcal{Z}(t)^{\frac{1}{2}}B\,\mathcal{Z}(t)^{-\frac{1}{2}}\big)\\ &\quad - \big(\mathcal{Z}(t)^{-\frac{1}{2}}\mathcal{T}_*^{-1}(t)\,B\,\mathcal{T}_*(t)\,\mathcal{Z}(t)^{\frac{1}{2}} + \mathcal{Z}(t)^{\frac{1}{2}}\,\mathcal{T}(t)\,B\,\mathcal{T}^{-1}(t)\,\mathcal{Z}(t)^{-\frac{1}{2}}\big) + 6\,t^2\,C\\ &= \big(\mathcal{Z}(-t)^{\frac{1}{2}}B\,\mathcal{Z}(t)^{\frac{1}{2}} + \mathcal{Z}(t)^{\frac{1}{2}}B\,\mathcal{Z}(-t)^{\frac{1}{2}}\big)\\ &\quad - \big(\mathcal{Z}(-t)^{\frac{1}{2}}\mathcal{T}(-t)\,B\,\mathcal{T}_*(t)\,\mathcal{Z}(t)^{\frac{1}{2}} + \mathcal{Z}(t)^{\frac{1}{2}}\,\mathcal{T}(t)\,B\,\mathcal{T}_*(-t)\,\mathcal{Z}(-t)^{\frac{1}{2}}\big) + 6\,t^2\,C\\ &= 2\,\widehat{\mathcal{D}}_s(-t).\end{aligned} \tag{7.7}$$

Inserting the expressions for $\mathcal{Z}(t)$, $\mathcal{T}(t)$ and $\mathcal{T}_*(t)$ in equation (7.7) and noticing that $[\mathrm{e}^{t\,B}, B] = 0$, we conclude that the expression for $\widehat{\mathcal{D}}_s(t)$ is of the form

$$\begin{aligned}
\widehat{\mathcal{D}}_s(t) &= \tfrac{1}{2}\mathrm{e}^{\frac{1}{2}t^3\,C}\,\underline{\mathrm{e}^{\frac{1}{2}tB}\left(B - \mathrm{e}^{\frac{1}{2}tA}\,B\,\mathrm{e}^{-\frac{1}{2}tA}\right)\mathrm{e}^{-\frac{1}{2}tB}}\,\mathrm{e}^{-\frac{1}{2}t^3\,C} \\
&\quad + \tfrac{1}{2}\mathrm{e}^{-\frac{1}{2}t^3\,C}\,\underline{\mathrm{e}^{-\frac{1}{2}tB}\left(B - \mathrm{e}^{-\frac{1}{2}tA}\,B\,\mathrm{e}^{\frac{1}{2}tA}\right)\mathrm{e}^{\frac{1}{2}tB}}\,\mathrm{e}^{\frac{1}{2}t^3\,C} + 3\,t^2\,C \\
&= \tfrac{1}{2}\mathrm{e}^{\frac{1}{2}t^3\,C}\left(B - \Xi(t)\right)\mathrm{e}^{-\frac{1}{2}t^3\,C} \\
&\quad + \tfrac{1}{2}\mathrm{e}^{-\frac{1}{2}t^3\,C}\left(B - \Xi(-t)\right)\mathrm{e}^{\frac{1}{2}t^3\,C} + 3\,t^2\,C, \quad \widehat{\mathcal{D}}(0) = 0,
\end{aligned} \tag{7.8}$$

with $\Xi(t) = \mathrm{e}^{\frac{1}{2}tB}\left(\mathrm{e}^{\frac{1}{2}tA}B\,\mathrm{e}^{-\frac{1}{2}tA}\right)\mathrm{e}^{-\frac{1}{2}tB}$ as defined in section 6.2, more precisely in equation (6.9) with expansion given by Lemma 9. The underlined terms correspond to the expansion given in Lemma 10 for the case of vanishing $C = 0$ - which recovers the Strang splitting.

7.3 The Lie expansion formula extended

Shifting to ad-notation, we see that equation (7.8) looks like

$$\begin{aligned}
\widehat{\mathcal{D}}_s(t) &= \tfrac{1}{2}\mathrm{e}^{\mathrm{ad}_{1/2\,t^3\,C}}(B - \Xi(t)) + \tfrac{1}{2}\mathrm{e}^{-\mathrm{ad}_{1/2\,t^3\,C}}(B - \Xi(-t)) + 3\,t^2\,C \\
&= \tfrac{1}{2}\mathrm{Ad}_{1/2\,t^3\,C}(B - \Xi(t)) + \tfrac{1}{2}\mathrm{Ad}_{-1/2\,t^3\,C}(B - \Xi(-t)) + 3\,t^2\,C
\end{aligned} \tag{7.9}$$

In Remark 12, we introduced $D(\cdot) = \mathrm{ad}_{\cdot}(X)$ and now define $f(\cdot) := \mathrm{Ad}_{\cdot\,C}$ and $g(t) = \frac{1}{2}\,t^3$. The main ingredient for the error expansion of equation (7.9) will be of the form

$$\mathrm{e}^{\mathrm{ad}_{t^3 C}}(X) = \mathrm{e}^{t^3\,C}\,X\,\mathrm{e}^{-t^3\,C} = \mathrm{e}^{D(g(t))}(X) = \left((f \circ g)\,(t)\right)(X),$$

where we assume time-indenpendency of X.

This clamours for a generalisation of Lemma 7, the Lie expansion formula which will become a main building block for calculating the defect directly.

Lemma 11 (Lie expansion formula extended). Let X be fixed and time-independent. Then, the t-dependent exponential

$$\mathrm{e}^{\mathrm{ad}_{t^3\,C}}(X) = \mathrm{e}^{t^3\,C}\,X\,\mathrm{e}^{-t^3\,C} \tag{7.10a}$$

satisfies the recurrence relation

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{\mathrm{ad}_{t^3\,C}}(X) = (3t^2)\,\mathrm{ad}_C\big(\mathrm{e}^{\mathrm{ad}_{t^3\,C}}(X)\big). \tag{7.10b}$$

More generally, for all $q \in \mathbb{N}_0$, we have

$$\tfrac{\mathrm{d}^q}{\mathrm{d}\,t^q}\,\mathrm{e}^{\mathrm{ad}_{t^3\,C}}(X) = \sum_{k=0}^{q}[C, \mathrm{e}^{\mathrm{ad}_{t^3\,C}}(X)]_k \cdot B_{q,k}\left(3\,t^2, 6\,t, 6, 0, \ldots, 0\right). \tag{7.10c}$$

Evaluating the derivatives at $t = 0$ yields

$$\tfrac{\mathrm{d}^q}{\mathrm{d}\,t^q}\,\mathrm{e}^{\mathrm{ad}_{t^3\,C}}(X)\big|_{t=0} = \sum_{k=1}^{q}[C, X]_k \cdot B_{q,k}\,(0, 0, 6, 0, \ldots, 0)\,, \tag{7.10d}$$

which gives the terms needed for a Taylor expansion at $t = 0$.

Remark 16. Lemma 11 can be generalised for arbitrary $g(t)$. The recurrence as in (7.10b) will read as

$$\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{\mathrm{ad}_{g(t)\,C}}(X) = g'(t)\,\mathrm{ad}_C\big(\mathrm{e}^{\mathrm{ad}_{g(t)\,C}}(X)\big). \tag{7.11a}$$

and for any $q \in \mathbb{N}_0$, the derivatives take the form (7.10c)

$$\tfrac{\mathrm{d}^q}{\mathrm{d}\,t^q}\,\mathrm{e}^{\mathrm{ad}_{g(t)\,C}}(X) = \sum_{k=0}^{q}[C, \mathrm{e}^{\mathrm{ad}_{g(t)\,C}}(X)]_k \cdot B_{q,k}\left(g^{(1)}(t), g^{(2)}(t), \dots, g^{(q-k+1)}(t)\right). \tag{7.11b}$$

Remark 17 (Connection to (7.9)). For the problem posed in (7.9), the building blocks are of the form

$$\begin{aligned}
\mathrm{e}^{\pm\frac{1}{2}t^3\,C}\big(B - \Xi(\pm t)\big)\,\mathrm{e}^{\mp\frac{1}{2}t^3\,C} \quad &\rightsquigarrow \quad \mathrm{e}^{g(t)\,C}\big(X\big)\,\mathrm{e}^{-g(t)\,C}, \quad g(t) = \tfrac{1}{2}t^3 \\
&\rightsquigarrow \quad \mathrm{e}^{\mathrm{ad}_{g(t)\,C}}\big(X\big), \qquad\quad g(t) = t^3.
\end{aligned}$$

A sensible approach for the moment is to consider $g(t) = t^3$ and assume time-independence of X - which is what Lemma 11 covers. The derivatives are $g'(t) = 3\,t^2$, $g''(t) = 6\,t$, $g^{(3)}(t) = 6$ and $g^{(k)}(t) = 0$, for $k \geq 4$, and thus simplify the appearing Bell polynomials to

$$B_{q,k}(3\,t^2, 6\,t, 6, \underbrace{0,\dots,0}_{q-k-2 \text{ times}}) = \sum \frac{q!}{c_1!\,c_2!\,c_3!}\left(\frac{3\,t^2}{1!}\right)^{c_1}\left(\frac{6\,t}{2!}\right)^{c_2}\left(\frac{6}{3!}\right)^{c_3}, \tag{7.12a}$$

with summation over all nonnegative integers $c_1, c_2, c_3 \geq 0$, such that

$$\begin{aligned} c_1 + 2\,c_2 + 3\,c_3 &= q, \\ c_1 + c_2 + c_3 &= k. \end{aligned} \tag{7.12b}$$

Remark 18. Notice that $g(0) = 0$ plays an important role when getting from (7.10c) to (7.10d). It guarantees that $\mathrm{e}^{\mathrm{ad}.}(X)$ reduces to the identity.

Remark 19. For a polynomial $g(t)$, the maximum number of c_i's that may appear corresponds to the degree of $g(t)$.

Proof. The general formula for the coefficients $B_{q,k}(3\,t^2, 6\,t, 6, 0, \dots, 0)$ in (7.10b) was found by calculating the first few derivatives using Maple by applying the recurrence relation (7.10b) repeatedly. The sequence of the coefficients was looked up on [OEIS]. This was proposed by [Hofstätter, 2019] in private communication. Then, the factors in the Bell polynomials were associated with the derivatives of $g(t)$ and the theory found in literature afterwards.

Now, let us start with the case $q = 0$, which is obvious as $[U, V]_0 = V$ holds true. The case $q = 1$ follows by applying the product rule

$$\begin{aligned}
\tfrac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{t^3\,C}\,X\,\mathrm{e}^{-t^3\,C} &= \left(\tfrac{\mathrm{d}}{\mathrm{d}\,t}\,\mathrm{e}^{t^3\,C}\right)\,X\,\mathrm{e}^{-t^3\,C} + \mathrm{e}^{t^3\,C}\,X\,\left(\tfrac{\mathrm{d}}{\mathrm{d}\,t}\,\mathrm{e}^{-t^3\,C}\right) \\
&= (3\,t^2)\,\left(C\,\mathrm{e}^{t^3\,C}\,X\,\mathrm{e}^{-t^3\,C} - \mathrm{e}^{t^3\,C}\,X\,\mathrm{e}^{-t^3\,C}\,C\right) = (3\,t^2)\,[C, \mathrm{e}^{t^3\,C}\,X\,\mathrm{e}^{-t^3\,C}] \\
&= (3t^2)\,\mathrm{ad}_C\big(\mathrm{e}^{\mathrm{ad}_{t^3\,C}}(X)\big),
\end{aligned}$$

which gives the recurrence as claimed.
For the more general case, with

$$f(\cdot) = \mathrm{e}^{\mathrm{ad}_{\cdot\, U}}(X) \quad \text{and} \quad g(t) = t^3$$

we notice that

$$\mathrm{e}^{\mathrm{ad}_{t^3\, C}}(X) = (f \circ g)(X). \tag{7.13}$$

For the derivatives of $f(y)$, we see - using Leibniz rule again - that

$$\begin{aligned} f'(y) = \tfrac{\mathrm{d}}{\mathrm{d}\, y}\mathrm{e}^{\mathrm{ad}_{y\, C}}(X) &= \tfrac{\mathrm{d}}{\mathrm{d}\, t}\left(\mathrm{e}^{y\, C}\, X\, \mathrm{e}^{-y\, C}\right) = C\,(\mathrm{e}^{y\, C}\, X\, \mathrm{e}^{-y\, C}) - (\mathrm{e}^{y\, C}\, X\, \mathrm{e}^{-y\, C})\, C \\ &= \mathrm{ad}_C(\mathrm{e}^{y\, C}\, X\, \mathrm{e}^{-y\, C}) = \mathrm{ad}_C(\mathrm{e}^{\mathrm{ad}_{y\, C}}(X)). \end{aligned}$$

As C is time-indepedent, the induction step simplifies as follows

$$\tfrac{\mathrm{d}^n}{\mathrm{d}\, y^n} f(y) = \tfrac{\mathrm{d}^{n-1}}{\mathrm{d}\, y^{n-1}}\, \mathrm{ad}_C(\mathrm{e}^{\mathrm{ad}_{y\, C}}(X)) = \mathrm{ad}_C(\tfrac{\mathrm{d}^{n-1}}{\mathrm{d}\, y^{n-1}}\mathrm{e}^{\mathrm{ad}_{y\, C}}(X)) \tag{7.14}$$

$$= \mathrm{ad}_C\, \mathrm{ad}_C^{n-1}(\mathrm{e}^{\mathrm{ad}_{y\, C}}(X))) = \mathrm{ad}_C^n(\mathrm{e}^{\mathrm{ad}_{y\, C}}(X))). \tag{7.15}$$

The 'Faà di Bruno' identity (7.2) applied to equation (7.13) yields

$$\begin{aligned} \tfrac{\mathrm{d}^q}{\mathrm{d}\, t^q}\, \mathrm{e}^{\mathrm{ad}_{t^3\, C}}(X) &= \sum_{k=1}^{q} f^{(k)}(g(t)) \cdot B_{q,k}\left(g'(t), g''(t), \ldots, g^{(q-k+1)}(t)\right) \\ &\overset{(7.14)}{=} \sum_{k=1}^{q} \mathrm{ad}_C^k(\mathrm{e}^{\mathrm{ad}_{t^3\, C}}(X)) \cdot B_{q,k}\left(g'(t), g''(t), \ldots, g^{(q-k+1)}(t)\right), \end{aligned}$$

whence the conjecture follows when switching to ad-notation as in $\mathrm{ad}_U^k(Z) = [U, Z]_k$.
Evaluation at $t = 0$ and remarking that $\mathrm{e}^{\mathrm{ad}_{0\, C}}(X) = \mathbb{1}(X)$ shows the remaining claim. □

Remark 20. As we are concerned with evaluation at $t = 0$, we notice that barely any Bell-polynomials are non-vanishing. The procedure can be depicted by using trees as in Figure 3.

Looking at the base case - the recurrence relation (7.10b) - one can either decrease the order by one (differentiating the polynomial part, which corresponds going left in the tree) or getting a multiplicative inner derivate $g'(t) = 3\, t^2$ (corresponding to a differentiation of the always existing factor $\mathrm{e}^{\mathrm{ad}_{t^3\, C}}(X)$), which means to increase the order by two in each step (i. e. going right in the tree).

The level of the tree corresponds to the number of derivatives executed - given by q, starting with the 0-th derivative at the very top. As the Bell polynomials $B_{n,k}(x_1 := g'(t), x_2 := g''(t), \ldots)$ are a multinomial in $x_1, x_2, \ldots$, we notice that k indicates the number of x_i scaled by the exponents involved in a given level according to the formula $c_1 + c_2 + c_3 = k$. For instance, in the third level, we have the term 6, which is $g'''(t)$ and hence $k = 1$. The term $54t^3$ is a multiplication of $3t^2\, 6t$ where $g'(t)$ and $g''(t)$ are present, hence $k = 2$. The last term present in this level, $(3t^2)^3$, yields $k = 3$ due to the exponent.
A Maple routine for Bell polynomials is provided in [OEIS] by Vladimir Kruchinin. It generates the following coefficients

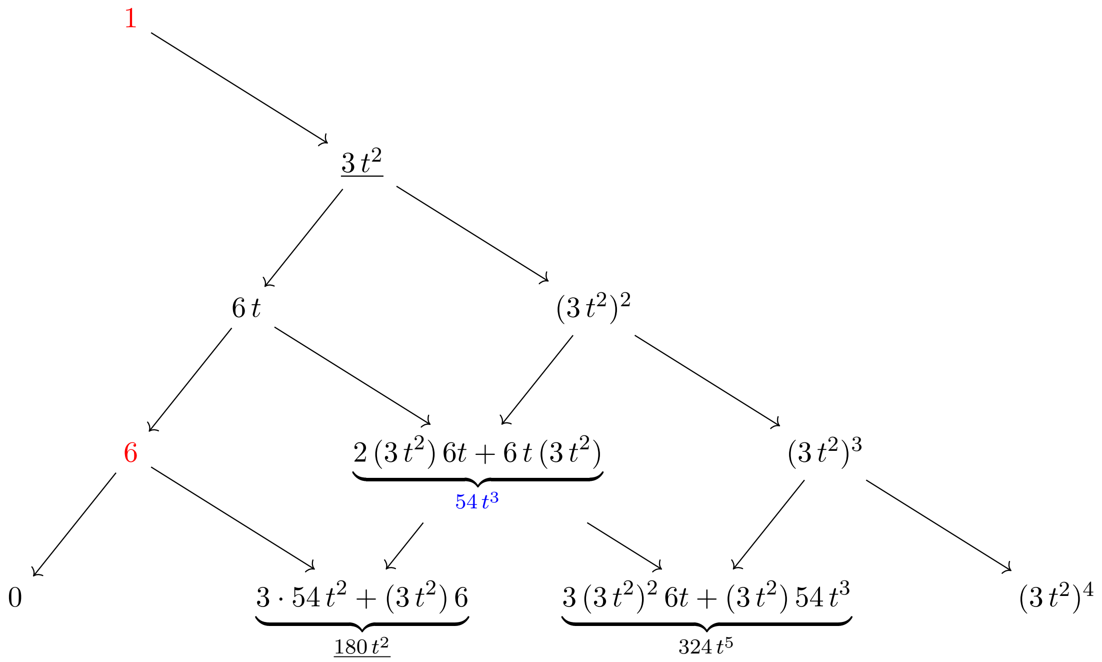


Figure 3: Depiction for higher derivatives of the recurrence given in (7.10b) ($q = 0, \dots, 3$). The level of the tree correspond to the number of derivatives taken.

$$B_{q,k}(3\,t^2, 6\,t, 6, 0, \dots, 0) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3t^2 & 6t & 6 \\ 0 & 0 & 9t^4 & 54\,t^3 \\ 0 & 0 & 0 & 27t^6 \end{pmatrix}.$$

Noticing that $54\,t^3$ corresponds to the addition of the subtrees in Figure 3. The rows of the matrix are the levels in Figure 3. The red constant entries do not vanish in a Taylor expansion when evaluated at time $t = 0$.

7.4 Graphs - a different approach to the combinatorics involved

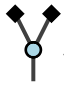
This section is devoted to introducing a different way of formalising the occurring derivatives. The idea was first seen in [Johnson-Freyd, 2009] and was adapted to the problem encountered in (7.8).

We want to build up trees where derivatives are thought of as being sources and the sink is being located at the very bottom of the diagram. We interpret incoming edges from above to nodes as a derivative with respect to the previous node. Summing up this idea, yields the

Definition 22 (Feynman diagrams). Let $f(t)$ be a $\mathbb{R}$-valued function with $f(0) = 0$. We define the *Feynman diagram of* $f(t)$ to be

$$:= f(t) \quad \text{and derivatives w. r. t. } t \text{ as} \quad \blacklozenge.$$

The first and the second derivatives of $f(t)$ are depicted as follows:

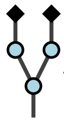
$f'(t) =$  and $f''(t) =$  .

A straight-forward generalisation defines higher derivatives.

Let $g(t)$ be $\mathbb{R}$-valued function with $g(0) = 0$. When considering derivatives of composed functions $(f \circ g)(t)$, the Feynman diagram looks as follows:

$\frac{\mathrm{d}}{\mathrm{d}t}(f \circ g)(t) = f'(g(t))\, g'(t) = \frac{\mathrm{d}f}{\mathrm{d}t}\frac{\mathrm{d}g}{\mathrm{d}x} =$  .

The Feynman diagram of $(f'' \circ g)(g')^2$ is

$(f'' \circ g)(g')^2 =$  ,

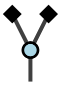
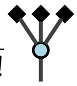
as there are two incoming edges from the top for f and two times one incoming edge for g. These are the building blocks for higher order derivatives and its Feynman diagrams. ⋄

Remark 21. Even though we will take a look at matrix-valued functions, the derivatives of $g(t) = t^3$ are scalar-valued and therefore can be commuted to arbitrary places.

Remark 22. We assume that $f(t)$ has no constant term in its Taylor expansion, i. e. $f(0) = 0$. Note that this is not a big issue due to the fact that if there is a nonvanishing constant term, we consider the centred function $\tilde{f}(t) = f(t) - f(0)$ instead. The derivatives $\tilde{f}^{(k)}(t) = f^{(k)}(t)$ coincide for all $k \geq 1$. The reason for mentioning this technicality is that Feynman diagrams do not depict the evaluation point.

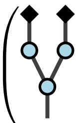
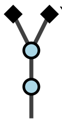
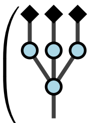
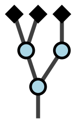
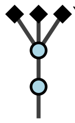
The Feynman diagrams allow us to write the Taylor expansion of a function $f(t)$ as

$$f(t) - f(0) = f'(0)\, t + \tfrac{1}{2!} f''(0)\, t^2 + \tfrac{1}{3!} f'''(0)\, t^3$$

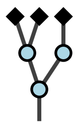
$=$  $t + \frac{1}{2!}$  $t^2 + \frac{1}{3!}$  $t^3 + \mathcal{O}(t^4), \quad t \to 0.$

In addition, they offer a very insightful way of interpreting Faà di Bruno as stated in [Johnson-Freyd, 2009]. We see that - omitting the evaluations at $t = 0$ in the Taylor expansion- that the trees depict the situation in an organised way:

$$f(g(t)) - f(g(0)) = (f' \circ g)g'\, t + \tfrac{1}{2!}\left((f'' \circ g)(g')^2 + (f' \circ g)g''\right)\, t^2 + \tfrac{1}{3!}\left((f''' \circ g)(g')^3 + 3\,(f'' \circ g)g'g'' + (f' \circ g)g'''\right)\, t^3 + \mathcal{O}(t^4)$$

$=$  $t + \frac{1}{2!}\Big($  $+$  $\Big)\, t^2 + \frac{1}{3!}\Big($  $+ 3$  $+$  $\Big)\, t^3 + \mathcal{O}(t^4),$ (7.16) for $t \to 0$.

Imagining an enumeration of the derivatives (for instance $d_1, d_2, \ldots, d_n$), the factor $\frac{1}{n!}$ can be seen as the order of the symmetric group S_n - the permutation group for n factors. The additional prefactor of the term $3\,(f'' \circ g)g'g''$ in the Taylor expansion stems from the fact that there are three 2-cycles in S_3 (i. e. $(d_1, d_2), (d_2, d_3), (d_1, d_3)$) - with other words three ways to get to the tree



We notice that the levels in Figure 3 correspond to the graphs in the Taylor expansion given in equation (7.16).

Now, we are familiar with Feynman diagrams, which was the aim of this section.

7.5 Time dependency - how things are put together

As seen in Remark 17, Lemma 11 dealt with the simplifying assumption of a time-independent X. The next question to ask is what changes when we assume time dependency of $\mathcal{X}(t)$ for the extended Lie expansion formula as given in the defect $\widehat{\mathcal{D}}_s(t)$ in equation (7.8). For our case, $\mathcal{X}(t) = B - \Xi(\pm t)$.

Remark 23. It is helpful mentioning that the notation for $\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))$ is pointing toward another use of the chain rule. However, expanding the definitions, we need to use the product rule, as

$$\frac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t)) = \frac{\mathrm{d}}{\mathrm{d}\,t}\left(\mathrm{e}^{t^3 C}\,\mathcal{X}(t)\,\mathrm{e}^{-t^3 C}\right) = \left(\frac{\mathrm{d}}{\mathrm{d}\,t}\mathrm{e}^{\mathrm{ad}_{t^3 C}}\right)(\mathcal{X}(t)) + \mathrm{e}^{\mathrm{ad}_{t^3 C}}\left(\frac{\mathrm{d}}{\mathrm{d}\,t}\mathcal{X}(t)\right). \tag{7.17}$$

In more abstract terms, it reads

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\,t} f(g(t), \mathcal{X}(t)) &= \left(\frac{\mathrm{d}\,f}{\mathrm{d}\,g}\right)(g(t), \mathcal{X}(t))\,g'(t) + \left(\frac{\mathrm{d}\,f}{\mathrm{d}\,\mathcal{X}}\right)(g(t), \mathcal{X}(t))\,\mathcal{X}'(t) \\ &= g'(t)\,\mathrm{ad}_C\left(f\left(g(t), \mathcal{X}(t)\right)\right) + f(g, \mathcal{X}'(t)). \end{aligned} \tag{7.18}$$

The last equation follows from the fact that

$$\frac{\mathrm{d}}{\mathrm{d}\,y(t)}\mathrm{e}^{\pm y(t)\,C} = \pm y'(t) C\,\mathrm{e}^{\pm y(t)\,C} = \pm \mathrm{e}^{\pm y(t)\,C}\,C y'(t),$$

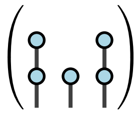
as $\mathrm{e}^{\pm y(t)\,C}$ is a function of C and hence naturally commutes as well as the fact that $[C, \mathcal{X}(t)] \neq 0$ in the general case.

We notice that the recurrence relation (7.18) generalises equation (7.10b).

7.5.1 Feynman diagram representation

When considering higher order derivatives of $\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))$, we will apply the recurrence in equation (7.18) repeatedly. Several applications of the recurrence relation tend to inflate the terms to keep track of. We introduce Feynman diagrams to simplify this task:

$$\mathrm{e}^{t^3 C}\,\mathcal{X}(t)\,\mathrm{e}^{-t^3 C} =: \tag{7.19}$$



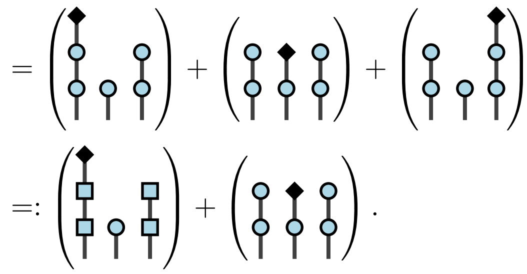
A straight-forward evaluation at $t = 0$ yields

$$\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\Big|_{t=0} = \mathcal{X}(0). \tag{7.20}$$

7.5.2 First derivative

The first derivative looks as follows:

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t)) = \tag{7.21}$$



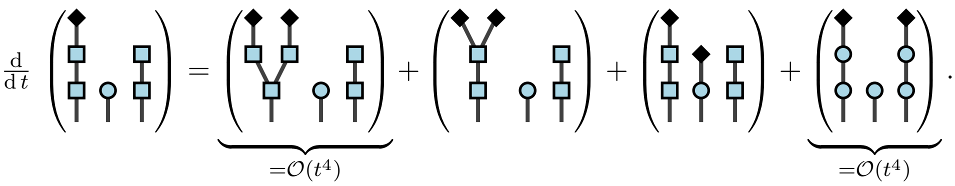
Notice that the squares are used to represent the ad - as seen in the recurrence relation (7.18). We can immediately evaluate at $t = 0$ to see

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\,\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\Big|_{t=0} = \mathcal{X}^{(1)}(0), \tag{7.22}$$

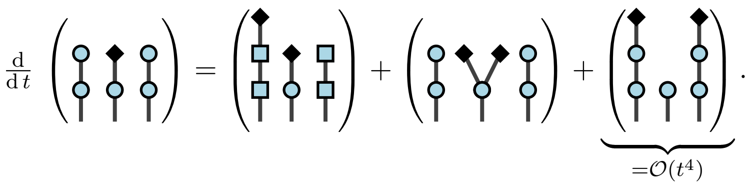
as the first subgraph in the first summand indicates the presence of a factor $g'(t) = 3\,t^2$.

7.5.3 Second derivative

Continuing the derivation process, we arrive at

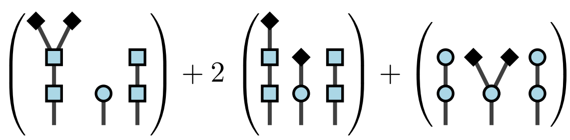


Moreover, we also get



The aim is to reproduce the scheme up to order 4 and hence we omit the terms of order 4 in the second derivative. Therefore, the equation holds true for $t \to 0$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\,\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t)) = \;\ldots\; + \mathcal{O}(t^4).$$



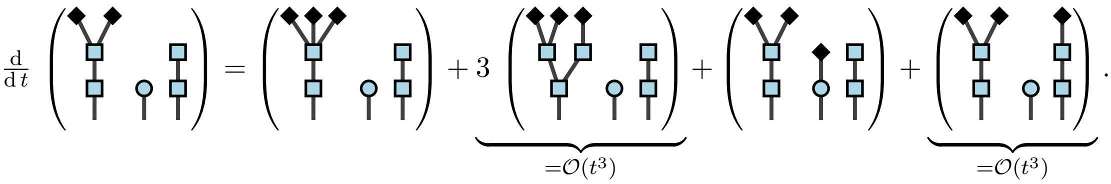
The evaluation results in

$$\frac{\mathrm{d}^2}{\mathrm{d}\,t^2}\,\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\Big|_{t=0} = \mathcal{X}^{(2)}(0), \tag{7.23}$$

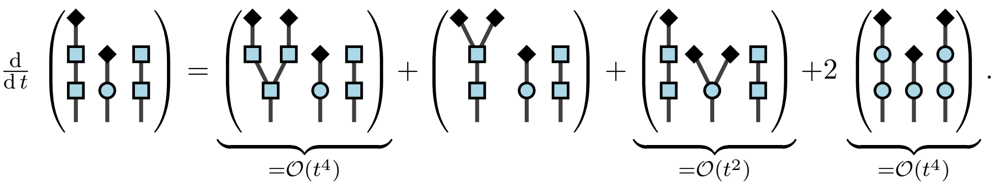
as the first term is proportional to $g''(t) = 6\,t$ and the second summand is proportional to $g'(t) = 3\,t^2$.

7.5.4 Third derivative

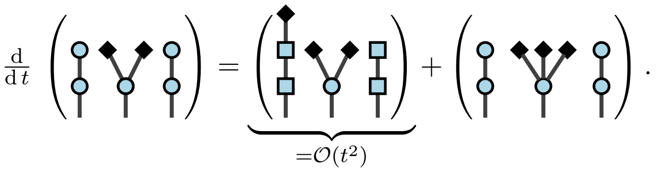
For the third derivative, we need to calculate three derivatives.
Firstly, we observe



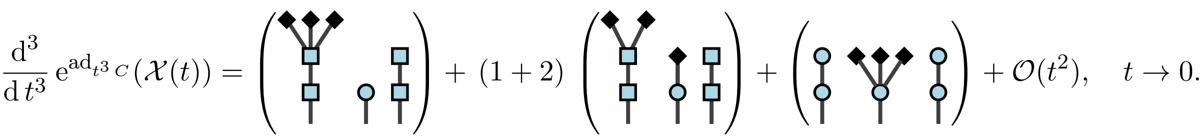
Secondly, the calculation reveals that



Thirdly, we get



Remembering that the second term is scaled by a factor 2, we arrive at



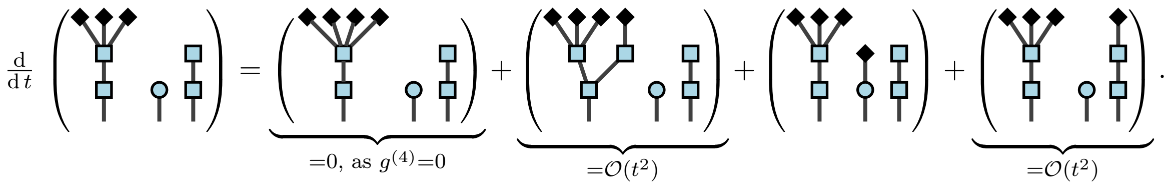
Immediately, we have the evaluation at hand:

$$\frac{\mathrm{d}^3}{\mathrm{d}\,t^3}\,\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\Big|_{t=0} = 6\,[C, \mathcal{X}(0)] + \mathcal{X}^{(3)}(0). \tag{7.24}$$

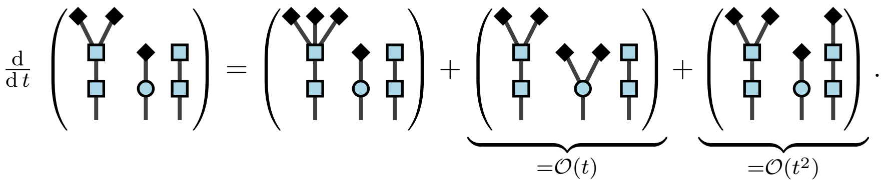
The prefactor 6 stems from the third derivative of $g(t) = t^3$ and the second term vanishes since $g''(0) = 0$.

7.5.5 Fourth derivative

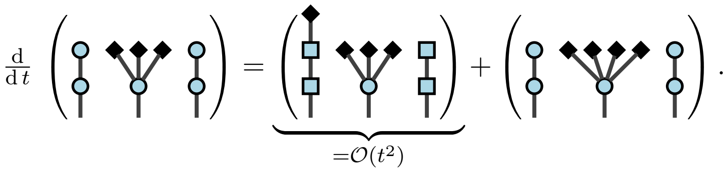
Deriving the first term of the third derivative, we get



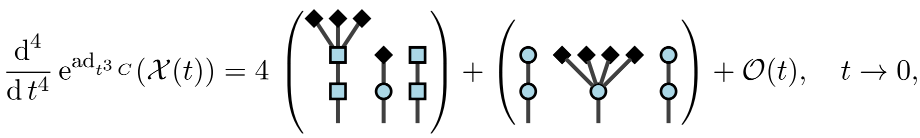
The second summand yields



Furthermore, we see that the third term gives



In total, it looks as follows



and without effort, we can evaluate at $t = 0$ to see that

$$\frac{\mathrm{d}^4}{\mathrm{d}\,t^4}\,\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\Big|_{t=0} = 24\,[C, \mathcal{X}^{(1)}(0)] + \mathcal{X}^{(4)}(0). \tag{7.25}$$

The factor 24 is a product of $g^{(3)}(0) = 6$ and the prefactor 4 in the first term.

7.5.6 Taylor expansion of the time-dependent case

Putting all pieces together, we are able to write down the Taylor expansion around $t = 0$ of the time-dependent Lie expansion formula. This is the content of the next lemma.

Lemma 12 (Lie Formula Extended with time dependent $\mathcal{X}(t)$). Let $g(t) = t^3$ be given. For time-dependent $X = X(t)$, the Taylor expansion around $t = 0$ reads

$$\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t)) = \mathcal{X}(0) + \mathcal{X}^{(1)}(0)\,t + \frac{1}{2}\mathcal{X}^{(2)}(0)\,t^2 + \left([C, \mathcal{X}(0)] + \frac{1}{6}\,\mathcal{X}^{(3)}(0)\right)\,t^3$$
$$+ \left([C, \mathcal{X}^{(1)}(0)] + \frac{1}{4!}\,\mathcal{X}^{(4)}(0)\right)\,t^4 + \mathcal{O}(t^5).$$

Proof. We write out the Taylor expansion around $t = 0$:

$$\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t)) = \mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\Big|_{t=0} + \frac{1}{1!}\left(\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\right)^{(1)}\Big|_{t=0} t + \frac{1}{2!}\left(\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\right)^{(2)}\Big|_{t=0} t^2 + \frac{1}{3!}\left(\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\right)^{(3)}\Big|_{t=0} t^3 + \frac{1}{4!}\left(\mathrm{e}^{\mathrm{ad}_{t^3 C}}(\mathcal{X}(t))\right)^{(4)}\Big|_{t=0} t^4 + \mathcal{O}(t^5).$$

The claim is an immediate consequence of inserting the expressions for the derivatives that we obtained in the previous subsections in (7.20), (7.22), (7.23), (7.24) and (7.25) as well as possible cancellation of prefactors. □

We are able to calculate the Taylor expansion of the symmetrised defect given by equation (7.9), which is what the following lemma is about.

Lemma 13 (Taylor expansion of inner symmetrised defect). Let $g(t) = \frac{1}{2}t^3$ and let the expression of the symmetrised defect be given by equation (7.9), namely

$$\widehat{\mathcal{D}}_s(t) = \tfrac{1}{2}\mathrm{e}^{\mathrm{ad}_{1/2\,t^3 C}}(B - \Xi(t)) + \tfrac{1}{2}\mathrm{e}^{-\mathrm{ad}_{1/2\,t^3 C}}(B - \Xi(-t)) + 3\,t^2\,C.$$

Then, the Taylor expansion of $\widehat{\mathcal{D}}_s(t)$ for $t \to 0$ is given by

$$\widehat{\mathcal{D}}_s(t) = \left(-\frac{1}{8}\left([A,B]_2 + 2\,[B,[A,B]_1]\right) - 3\,C\right) t^2 + \frac{1}{4}\,[C,[A,B]]\,t^4 - \frac{1}{384}\left([A,B]_4 + 4\,[B,[A,B]_3] + 6\,[B,[A,B]_2]_2 + 4\,[B,[A,B]]_3\right) t^4 + \mathcal{O}(t^5) \tag{7.26}$$

Proof. We recall the definition of $\Xi(t)$ as seen in equation (6.8):

$$\Xi(t) = \mathrm{e}^{\frac{1}{2}tB}\left(\mathrm{e}^{\frac{1}{2}tA}B\,\mathrm{e}^{-\frac{1}{2}tA}\right)\mathrm{e}^{-\frac{1}{2}tB},$$

and due to the fact that $\Xi(0) = B$, we see

$$t^0: \quad \underbrace{\mathrm{e}^{\mathrm{ad}_0}}_{=\mathbb{1}}(B - \Xi(0)) = 0 \quad \text{resp.} \quad \underbrace{\mathrm{e}^{\mathrm{ad}_0}}_{=\mathbb{1}}(B - \Xi(0)) = 0 \quad \text{and hence} \quad \widehat{\mathcal{D}}_s(0) = 0,$$

which coincides with (7.26) for $t = 0$. Notice that strictly speaking, the 0 in the subindex of ad is meant to represent the 0-matrix.

Next, we take a look at the coefficient of order 1. We recall that in Lemma 9, the term of first order of Ξ is given by:

$$\begin{aligned}(B - \Xi(t))^{(1)}\Big|_{t=0} &= -2^{-1}\sum_{k=0}^{1-1}\binom{1}{k}[B,[A,B]_{1-k}]_k \\ &= -2^{-1}[A,B].\end{aligned}$$

If we consider $B - \Xi(-t)$ instead, we get an additional minus sign as the first order term is always antisymmetric. Keeping track of the minus due to $B - \Xi(\pm t)$ and the prefactors given in (11), we conclude

$$t^1: \quad \widehat{\mathcal{D}}_s^{(1)}(0) = -\tfrac{1}{2}\,2^{-1}[A,B] + \tfrac{1}{2}\,2^{-1}[A,B] = 0.$$

This is expected due to the symmetry of the scheme.

With a similar argument about parity - this time noticing that the second order term of $\Xi(t)$ is necessarily even - and recalling $\Xi(t)$ as in Lemma 9, we have

$$t^2 : \quad \frac{1}{2!}\left((\tfrac{1}{2}+\tfrac{1}{2}) \left(-2^{-2}\sum_{k=0}^{2-1}\binom{2}{k}[B,[A,B]_{2-k}]_k \right) + 6\,C \right) = -\frac{[A,B]_2 + 2\,[B,[A,B]_1]}{8} + 3\,C.$$

The last term is due to the second derivative of the $3\,t^2\,C$ term on the left hand side.
In total, if we set

$$C = \tfrac{1}{24}\left([A,[A,B]] + 2\,[B,[A,B]]\right)$$

we have a scheme of order 3 at least. This is what is stated in (4.10) and what we intended to reproduce.

The remaining claim is about higher order errors.

For order 3, we need to consider two terms as in Lemma 12. We will consider $\mathcal{X}^{(3)}(0)$ first. Lemma 9 states that

$$(B-\Xi(t))^{(3)}\Big|_{t=0} = -2^{-3}\sum_{k=0}^{3-1}\binom{3}{k}[B,[A,B]_{3-k}]_k.$$

Notice that there is an additional minus sign in the case of $\Xi(-t)$ owing to the fact that the third order term is antisymmetric. Thus, for the first term, we have $\mathcal{X}^{(3)}(0) = 0$.
The second term of order 3, $[C,\mathcal{X}(0)]$ vanishes due to the fact that $\mathcal{X}(0) = 0$ in both cases.

The fourth order term consists of two terms. Firstly, we turn towards $\mathcal{X}^{(4)}(0)$. With the aid of Lemma 9, we get

$$\begin{aligned}(B-\Xi(\pm t))^{(4)}\Big|_{t=0} &= -2^{-4}\sum_{k=0}^{4-1}\binom{4}{k}[B,[A,B]_{4-k}]_k \\ &= -\tfrac{1}{16}\left([A,B]_4 + \binom{4}{1}[B,[A,B]_3] + \binom{4}{2}[B,[A,B]_2]_2 + \binom{4}{3}[B,[A,B]]_3\right) \\ &= -\tfrac{1}{16}\left([A,B]_4 + 4\,[B,[A,B]_3] + 6\,[B,[A,B]_2]_2 + 4\,[B,[A,B]]_3\right) =: \textcircled{A}.\end{aligned}$$

This is due to the fact that the t^4 term is symmetric and hence

$$\tfrac{1}{2}(B-\Xi(t))^{(4)}\big|_{t=0} + \tfrac{1}{2}(B-\Xi(-t))^{(4)}\big|_{t=0} = \textcircled{A}.$$

Last but not least, it holds that $[C,\mathcal{X}^{(1)}(0)]$ is even as $\Xi(-t)$ is odd but due to the chain rule - as $g(t) = -\frac{1}{2}\,t^3$ - the sign of $\mathrm{e}^{\mathrm{ad}_{-1/2\,t^3\,C}}(B-\Xi(-t))$ changes again. Notice that the additional factor 2^{-1} is also a consequence of the scaling of $g(t)$. In total, we have

$$\tfrac{1}{2}[C,(B-\Xi(t))^{(1)}\big|_{t=0}] + (-1)\tfrac{1}{2}[C,(B-\Xi(-t))^{(1)}\big|_{t=0}] = -\tfrac{1}{4}\,[C,[A,B]].$$

All in all, with the correct prefactors, the fourth order term reads as

$$t^4 : \quad -\tfrac{1}{384}\left([A,B]_4 + 4\,[B,[A,B]_3] + 6\,[B,[A,B]_2]_2 + 4\,[B,[A,B]]_3\right) - \tfrac{1}{4}\,[C,[A,B]]$$

and this concludes the proof. □

Remark 24. We expect the terms of order 5 to vanish due to the symmetry of the method. For the order 6 error terms, the same techniques developed in this section can be applied.

We recall equation (6.6):

$$\mathcal{D}_s(t) = \mathcal{T}_*(t)\,\widehat{\mathcal{D}}(t)\,\mathcal{T}(t).$$

As we are finite dimensional, all norms are equivalent, hence we choose the 2-norm and use its submultiplicativity:

$$\|\mathcal{D}_s(t)\|_2 \leq \|\mathcal{T}_*(t)\|_2\,\|\widehat{\mathcal{D}}(t)\|_2\,\|\mathcal{T}(t)\|_2.$$

The matrix exponential of skew-hermitian matrices is unitary - see Example 7 - and the product of unitary matrices is unitary and therefore, we get

$$\|\mathcal{T}(t)\|_2 \leq \|\mathrm{e}^{\frac{1}{2}tB}\,\mathrm{e}^{\frac{1}{2}tA}\|_2 = 1$$

as well as

$$\|\mathcal{T}_*(t)\|_2 \leq \|\mathrm{e}^{\frac{1}{2}tA}\,\mathrm{e}^{\frac{1}{2}tB}\|_2 = 1.$$

In total, we get

$$\|\mathcal{D}_s(t)\|_2 \leq \|\widehat{\mathcal{D}}_s(t)\|_2. \tag{7.27}$$

Finally, we restate equation (3.14) for the global order $p = 4$ and replace the expression for $\widehat{\mathcal{D}}_s(t)$ by the Taylor expansion in Lemma 13 for $t \to 0$:

$$\begin{aligned}
\|\mathcal{L}(t)\|_2 &\overset{(3.14)}{=} \|\tfrac{1}{p+1}t\,\mathcal{D}_s(t) + \mathcal{O}(t^{p+3})\|_2 \leq \|\tfrac{1}{p+1}t\,\mathcal{D}_s(t)\|_2 + \mathcal{O}(t^7)\\
&\overset{(7.27)}{\leq} \tfrac{1}{p+1}t\,\|\widehat{\mathcal{D}}_s(t)\|_2 + \mathcal{O}(t^7)\\
&\overset{(7.26)}{=} \tfrac{t}{p+1}\underbrace{\left(-\frac{1}{8}\left([A,B]_2 + 2\,[B,[A,B]_1]\right) - 3\,C\right)}_{\text{second order condition}} t^2 + \mathcal{O}(t^5).
\end{aligned}$$

Solving the second order condition for C yields

$$C = \tfrac{1}{24}[A,[A,B]] + \tfrac{1}{12}[B,[A,B]], \tag{7.28}$$

which is what we desired to reproduce - see (4.10).

We observe that the choice of C according to the second order condition (7.28) results in a fourth order scheme due to the symmetry of the scheme.

All in all, we have managed to reproduce the result (4.10) by the symmetric BCH formula as previously seen. The approach using Feynman diagrams helps to keep an overview on what is essential due to the increasing amount of terms being present. It also offers a way for further generalisations without using much high-level theory.

8 Application to the Schrödinger Equation

So. Jetzt läßt das Zucken nach. Jetzt liegst du still.
Wir sind am Ende von Physiologie und Theologie,
die Physik beginnt.
— Alfred Döblin, *Berlin Alexanderplatz*

In this section, we consider the semiclassical Schrödinger equation as our test problem. A thorough treatment of the semiclassical Schrödinger equations is found in [Jin et al., 2011], from which the problem setting in the introductory section will be briefly restated.

Let $0 < \varepsilon \ll 1$ be the dimensionless *semiclassical parameter*. The posed problem looks as follows: For a sufficiently smooth potential $V(x)$, we have

$$\mathrm{i}\,\varepsilon\,\partial_t u^\varepsilon(x,t) = -\frac{\varepsilon^2}{2}\,\partial_x^2\,u^\varepsilon(x,t) + V(x)\,u^\varepsilon(x,t), \quad u^\varepsilon(0,x) = u_0^\varepsilon(x) \tag{8.1}$$

where $(x,t) \in \mathbb{R}^d \times \mathbb{R}$ and $d \in \mathbb{N}$ indicating the space dimension of the problem and the wave function $u^\varepsilon(t,x)$ is sufficiently smooth, complex valued and unknown. The parameter ε displays the ratio between the microscopic and macroscopic scale being present in the problem - for small ε we notice that the potential is macroscopic, the kinetic term microscopic.

Dividing equation (8.1) by $\mathrm{i}\,\varepsilon$, we get

$$\partial_t u^\varepsilon(x,t) = \mathrm{i}\left(\frac{\varepsilon}{2}\,\partial_x^2 - \frac{1}{\varepsilon}\,V(x)\right)u^\varepsilon(x,t), \quad u^\varepsilon(0,x) = u_0^\varepsilon(x). \tag{8.2}$$

In order to see the factors that arise for the commutators better, we replace $\varepsilon\,2^{-1}$ with ε and absorb the factor 2 in $V(x)$ without distinguishing this notationally.

Hence, exact time evolution operator $\mathcal{E}(t)$ is given by

$$\mathcal{E}(t) = \mathrm{e}^{\mathrm{i}\,t\left(\varepsilon\,\partial_x^2 - \varepsilon^{-1}V\right)},$$

with a potential term $V = V(x)$ being only dependent on x.

We approximate the exact evolution operator $\mathcal{E}(t)$ with a numerical scheme $\mathcal{S}(t)$ using the usual notation, i. e.

$$\mathcal{E}(t) \approx \mathcal{S}(t) = \mathrm{e}^{\frac{1}{2}tA}\,\mathrm{e}^{\frac{1}{2}tB}\,\mathrm{e}^{t^3\,C}\,\mathrm{e}^{\frac{1}{2}tB}\,\mathrm{e}^{\frac{1}{2}tA}, \quad t \to 0 \tag{8.3}$$

where

$$A = \mathrm{i}\,\varepsilon\,\partial_x^2,$$
$$B = -\mathrm{i}\,\varepsilon^{-1}V.$$

We assume A and B to be finite-dimensional approximations without distinguishing this notationally. Inserting a sufficiently smooth test function u, we have

$$Au = \mathrm{i}\,\varepsilon\,u_{xx},$$
$$Bu = -\mathrm{i}\,\varepsilon^{-1}V \cdot u.$$

We will now calculate C with the commutators that appear in the second order condition (7.28). First, we notice that the commutator $[A, B]$ can be written as

$$\begin{aligned}
[A,B]\,u &= A\,B\,u - B\,A\,u \\
&= A\,(-\mathrm{i}\,\varepsilon^{-1} V \cdot u) - B\,(\,\mathrm{i}\,\varepsilon\, u_{xx}) \\
&= \mathrm{i}\,\varepsilon\,(-\mathrm{i}\,\varepsilon^{-1} V \cdot u)_{xx} + \mathrm{i}\,\varepsilon^{-1} V \cdot (\,\mathrm{i}\,\varepsilon\, u_{xx}) \\
&= (V \cdot u)_{xx} - V \cdot u_{xx} \\
&= V_{xx} \cdot u + 2\,V_x \cdot u_x + \cancel{V \cdot u_{xx}} - \cancel{V \cdot u_{xx}} \\
&= V_{xx} \cdot u + 2\,V_x \cdot u_x.
\end{aligned}$$

We are now able to calculate $[A, [A, B]]$:

$$\begin{aligned}
[A,[A,B]]\,u &= A\,[A,B]\,u - [A,B]\,A\,u \\
&= A\,(V_{xx} \cdot u + 2\,V_x \cdot u_x) - [A,B]\,(\mathrm{i}\,\varepsilon\, u_{xx}) \\
&= \mathrm{i}\,\varepsilon\,(V_{xx} \cdot u + 2\,V_x \cdot u_x)_{xx} - V_{xx} \cdot (\mathrm{i}\,\varepsilon\, u_{xx}) - 2\,V_x \cdot (\mathrm{i}\,\varepsilon\, u_{xx})_x \\
&= \mathrm{i}\,\varepsilon\,\Big(V_{xxxx} \cdot u + 2\,V_{xxx} \cdot u_x + V_{xx} \cdot u_{xx} \\
&\qquad + 2\,V_{xxx} \cdot u_x + 4\,V_{xx} \cdot u_{xx} + \cancel{2\,V_x \cdot u_{xxx}} \\
&\qquad - V_{xx} \cdot u_{xx} - \cancel{2\,V_x \cdot u_{xxx}}\Big) \\
&= \mathrm{i}\,\varepsilon\,\big(V_{xxxx} \cdot u + 4\,V_{xxx} \cdot u_x + 4\,V_{xx} \cdot u_{xx}\big).
\end{aligned}$$

Another calculation yields an expression for $[B, [A, B]]$:

$$\begin{aligned}
[B,[A,B]]\,u &= B\,[A,B]\,u - [A,B]\,B\,u \\
&= B\,(V_{xx} \cdot u + 2\,V_x \cdot u_x) - [A,B]\,(-\mathrm{i}\,\varepsilon^{-1} V \cdot u) \\
&= -\mathrm{i}\,\varepsilon^{-1} V \cdot (V_{xx} \cdot u + 2\,V_x \cdot u_x) \\
&\quad - V_{xx} \cdot (-\mathrm{i}\,\varepsilon^{-1} V \cdot u) - 2\,V_x \cdot (-\mathrm{i}\,\varepsilon^{-1} V \cdot u)_x \\
&= \mathrm{i}\,\varepsilon^{-1}\,\big(-\cancel{V \cdot V_{xx} \cdot u} - \cancel{2\,V \cdot V_x \cdot u_x} + \cancel{V_{xx} \cdot V \cdot u} + 2\,V_x \cdot V_x \cdot u + \cancel{2\,V_x \cdot V \cdot u_x}\big) \\
&= 2\,\mathrm{i}\,\varepsilon^{-1}\, V_x^2 \cdot u.
\end{aligned} \tag{8.4}$$

Putting things together as in equation (7.26), we can write Cu as

$$Cu = \tfrac{1}{24}\,\mathrm{i}\,\varepsilon\big(V_{xxxx} \cdot u + 4\,V_{xxx} \cdot u_x + 4\,V_{xx} \cdot u_{xx}\big) + \tfrac{1}{6}\,\mathrm{i}\,\varepsilon^{-1}\, V_x^2 \cdot u$$

and hence C to be expressed as

$$\begin{aligned}
C &= \tfrac{1}{24}\,\mathrm{i}\,\varepsilon\big(V_{xxxx} + 4\,V_{xxx} \circ \partial_x + 4\,V_{xx} \circ \partial_x^2\big) + \tfrac{1}{6}\,\mathrm{i}\,\varepsilon^{-1}\, V_x^2 \\
&= \tfrac{1}{24}\,\mathrm{i}\,\varepsilon\,\partial_x^4 V + \tfrac{1}{6}\,\partial_x^3 V \circ \partial_x + \tfrac{1}{6}\,\partial_x^2 V \circ \partial_x^2 + \tfrac{1}{6}\,\mathrm{i}\,\varepsilon^{-1}\,(\partial_x V)^2.
\end{aligned} \tag{8.5}$$

The key aspect of this application is that (8.5) shows that C comparably cheap to calculate. Often, this is due to the fact that derivatives of the potential $V(x)$ are themselves relatively cheap to calculate.

Notice that equation (8.4) implies that

$$[B,[A,B]]_3 = [B,[B,[B,[A,B]]]] = [B,[B,[A,B]]] = 0$$

due to the fact that $V(x)$ is a scalar. Therefore, we get an expression for the inner symmetrised defect $\widehat{\mathcal{D}}_s(t)$:

$$\widehat{\mathcal{D}}_s(t) = \tfrac{1}{384}\,(96\,[C,[A,B]] - [A,B]_4 + 4\,[B,[A,B]_3] + 6\,[B,[A,B]_2]_2)\; t^4 + \mathcal{O}(t^5). \tag{8.6}$$

A numerical experiment for the harmonic potential

$$V(x) = \tfrac{1}{2}\,\omega^2\,x^2$$

with exact solution

$$u = \left(\tfrac{\omega}{\pi}\right)^{1/4}\,\mathrm{e}^{-\omega/2(\mathrm{i}t+x^2)}$$

is performed in [Auzinger and Koch, 2018, Section 4], verifying order 4 of the generalised Strang splitting in (8.3). Notice that the roles of A and B are interchanged in comparsion to Auzinger.

In total, we see that physics offer a wide variety of possible applications for high order splitting methods. The condition for a fourth order scheme is experimentally verified and shifting the derivatives to the potential might be an inexpensive way to calculate computionally challenging problems. The general approach seen in Section 7 might be applied to different scalings in the future.

References

[OEIS] The on-line encyclopedia of integer sequences. `http://oeis.org/A187082`. Accessed: 2020-07-25.

[Auzinger, 2018] Auzinger, W. (2018). Local error representation via symmetrized defect, for some symmetric splittings. unpublished.

[Auzinger, 2020] Auzinger, W. (2018-2020). private communication.

[Auzinger and Herfort, 2014] Auzinger, W. and Herfort, W. (2014). Local error structures and order conditions in terms of lie elements for exponential splitting schemes. *Opuscula Mathematica*, 34(2).

[Auzinger et al., 2019a] Auzinger, W., Hofstätter, H., and Koch, O. (2019a). Precise local error control for symmetric one-step schemes applied to nonlinear evolution equations. Conference presentation.

[Auzinger et al., 2019b] Auzinger, W., Hofstätter, H., and Koch, O. (2019b). Symmetrized local error estimators for time-reversible one-step methods in nonlinear evolution equations. *Journal of Computational and Applied Mathematics*, 356:339–357.

[Auzinger and Koch, 2018] Auzinger, W. and Koch, O. (2018). An improved local error estimator for symmetric time-stepping schemes. *Applied Mathematics Letters*, 82:106–110.

[Auzinger et al., 2014] Auzinger, W., Koch, O., and Thalhammer, M. (2014). Defect-based local error estimators for splitting methods, with application to Schrödinger equations, Part II. Higher-order methods for linear problems. *Journal of Computational and Applied Mathematics*, 255:384–403.

[Behr et al., 2018] Behr, M., Benner, P., and Heiland, J. (2018). Solution formulas for differential Sylvester and Lyapunov equations. *arXiv preprint arXiv:1811.08327*.

[Comtet, 1974] Comtet, L. (1974). *Advanced combinatorics : the art of finite and infinite expansions*. Reidel, Dordrecht [u.a.], rev. and enl. ed. edition.

[Dahlquist, 1959] Dahlquist, G. (1959). *Stability and error bounds in the numerical integration of ordinary differential equations*. Kungliga Tekniska Högskolans handlingar. Stockholm.

[Deuflhard and Bornemann, 2013] Deuflhard, P. and Bornemann, F. (2013). *Numerische Mathematik 2. Gewöhnliche Differentialgleichungen*. De Gruyter, Berlin [u.a.], 4. edition.

[Evans, 2010] Evans, L. C. (2010). *Partial differential equations*. Graduate studies in mathematics ; 19. American Math. Soc., Providence, RI, 2. ed. edition.

[Hairer et al., 2006] Hairer, E., Lubich, C. V., and Wanner, G. V. (2006). *Geometric numerical integration*. Springer series in computational mathematics ; 31. Springer, Berlin [u.a.], 2. ed. edition.

[Hall, 2003] Hall, B. C. (2003). *Lie groups, Lie algebras, and representations*. Graduate texts in mathematics ; 222. Springer, Cham [u.a.], 2. ed. edition.

[Hall, 2015] Hall, B. C. (2015). *Lie groups, Lie algebras, and representations : an elementary introduction*. Graduate texts in mathematics. Springer, Cham [u.a.], 2. edition.

[Higham, 2008] Higham, N. J. (2008). *Functions of matrices : theory and computation*. SIAM, Philadelphia, Pa.

[Hofstätter, 2019] Hofstätter, H. (2019). private communication.

[Jin et al., 2011] Jin, S., Markowich, P. A., and Sparber, C. (2011). Mathematical and computational methods for semiclassical schrödinger equations. *Acta Numerica*, 20:121–209.

[Johnson-Freyd, 2009] Johnson-Freyd, T. (2009). Is there a faa di bruno-like formula for composition of three functions? MathOverflow. URL:https://mathoverflow.net/q/5888 (version: 2009-11-18).

[Kato, 1995] Kato, T. (1995). *Perturbation Theory for Linear Operators*. Classics in mathematics. Springer Berlin Heidelberg, second edition.

[Khukhro and Khoukhro, 1998] Khukhro, E. I. and Khoukhro, E. I. (1998). *p-Automorphisms of Finite p-Groups*, volume 246. Cambridge University Press.

[Krantz and Parks, 2002] Krantz, S. G. and Parks, H. R. (2002). *A primer of real analytic functions*. Birkhäuser advanced texts : Basler Lehrbücher. Birkhäuser, Boston, Mass. [u.a.], 2. ed. edition.

[Magnus, 1954] Magnus, W. (1954). On the exponential solution of differential equations for a linear operator. *Communications on pure and applied mathematics*, 7(4):649–673.

[Prüss and Wilke, 2010] Prüss, J. and Wilke, M. V. (2010). *Gewöhnliche Differentialgleichungen und dynamische Systeme*. Grundstudium Mathematik. Birkhäuser, Basel.

[Schuster, 2018] Schuster, F. (2017-2018). Lie Algebren und Darstellungstheorie. University Lecture. `https://dmg.tuwien.ac.at/schuster/SkriptLieAlg.pdf`.

[Schöberl, 2016] Schöberl, J. (2016). Splitting methods. `http://www.asc.tuwien.ac.at/~schoeberl/wiki/lva/numpde15/splitting.pdf`.

[Singh, 2018] Singh, P. (2018). *High accuracy computational methods for the semiclassical Schrödinger equation*. PhD thesis, University of Cambridge.

[Stillwell, 2008] Stillwell, J. (2008). *Naive Lie theory*. Undergraduate texts in mathematics. Springer, New York, NY.

[Van Loan, 1977] Van Loan, C. (1977). The sensitivity of the matrix exponential. *SIAM Journal on Numerical Analysis*, 14(6):971–981.