Effective approximation for linear time-dependent Schrödinger equation

Arieh Iserles* & Karolina Kropielnicka†

November 30, 2011

Abstract

The computation of the linear Schrödinger equation presents substantive challenges because of the presence of a large parameter. Assuming periodic boundary conditions, the standard approach to its solution consists of semi-discretization with a spectral method, followed by an exponential splitting. Our contention in this paper is that this is sub-optimal. Using other means of semi-discretization, in particular finite differences or spectral collocation, one obtains matrices whose commutators are small. This opens up the possibility of using high-order splittings with arguments made up of nested commutators. In particular, we introduce a variation on the theme of the classical Zassenhaus splitting which is time-symmetric and separates powers of the large parameter, rather than separating powers of the time step.

1 Introduction

Linear Schrödinger equation plays central role in a wide range of applications and is the fundamental model of quantum mechanics (Griffiths 2004). Its computation presents many enduring challenges (Jin, Markowich & Sparber 2011) which form the centrepiece of this paper.

We consider the standard *linear Schrödinger equation* in one space variable,

$$\mathrm{i}\varepsilon u_t = -\frac{\varepsilon^2}{2m}u_{xx} - V(x)u, \qquad t \geq 0, \quad x \in [-1,1], \tag{1.1}$$

where $u = u(x,t)$, given with an initial condition and periodic boundary conditions, where the potential V is a periodic function of period 2 and $\|V\|_\infty = 1$. The important case of $\varepsilon = \hbar \approx 1,05457168 \cdot 10^{-34}$, the *reduced Planck constant*, is probably outside the reach of accurate computation, here we focus on larger, but still exceedingly small, values of ε. The minute size of ε is a major source of difficulties in the numerical

*Department of Applied Mathematics and Mathematical Physics, University of Cambridge, Wilberforce Rd, Cambridge CB3 0WA, UK.

†Department of Applied Mathematics and Mathematical Physics, University of Cambridge, Wilberforce Rd, Cambridge CB3 0WA, UK, and Institute of Mathematics, University of Gdańsk, Wit Stwosz Str. 57, 90-952 Gdańsk, Poland.

discretization of (1.1) because, using naive approach, the very rapid oscillations require spatial resolution of $\mathcal{O}(\hbar)$ which, needless to say, is often impractical or, at best, exceedingly expensive. This is the motivation to pursue alternative approaches, based in the main on the concept of *splittings* (Jin et al. 2011, McLachlan & Quispel 2002).

It is important to discuss the assumption on the periodicity of the boundary conditions and of the function V: in many applications of interest the initial condition is a *wave packet* – a function which is, to all intents and purposes, zero outside a fairly small subinterval. The equation being hyperbolic, signals propagate at finite speed, therefore nothing of much interest happens near the boundaries in long time intervals. We might impose there either zero Dirichlet or periodic conditions and our choice of the latter is motivated by the considerably simpler mathematical framework and by the availability of faster numerical algorithms, due in the main to the *fast Fourier transform (FFT)*.

In greater generality, the body of ideas in this paper can be extended with little extra effort to a multivariate linear Schrödinger equation

$$\mathrm{i}\varepsilon u_t(\boldsymbol{x},t) = -\frac{\varepsilon^2}{2m}\nabla u(\boldsymbol{x},t) - V(\boldsymbol{x})u(\boldsymbol{x},t), \qquad t \geq 0, \quad \boldsymbol{x} \in [-1,1]^d,$$

with appropriate periodic boundary conditions. This generalisation is fairly transparent and, for the sake of simplicity, we consider here just the univariate case.

In the sequel we let $\omega = 1/\varepsilon$, thereby rewriting (1.1) in the form

$$u_t(x,t) = \frac{\mathrm{i}}{2m\omega}u_{xx}(x,t) + \mathrm{i}\omega V(x)u(x,t). \tag{1.2}$$

We complement (1.2) with a smooth initial condition at $t = 0$,

$$u(x,0) = \phi(x), \qquad x \in [-1,1], \tag{1.3}$$

where $\phi(-1) = \phi(1)$, and periodic boundary condition at $x = \pm 1$,

$$u(-1,t) = u(1,t), \qquad t \geq 0. \tag{1.4}$$

The large size of ω causes high oscillations in the solution, rendering standard numerical methods – whether finite differences, spectral methods or finite elements – inefficient when solving (1.2). The standard approach which actually works combines spectral discretization of the space variables with exponential splitting in time (Jin et al. 2011).

In general, numerical analysis of (1.2) commences with *semi-discretization:* discretizing the space variable while retaining the time derivative intact. In other words, we seek an approximate solution of the equation in a finite-dimensional Hilbert space $\mathcal{H}$. The new variables are a representation of the solution in this space with respect to some basis: in the case of finite differences these are the function values on a spatial grid, for spectral methods these are expansion coefficients with respect to a suitable orthogonal basis while for finite elements, depending on precise implementation details, these can be either expansion coefficients in finite-element functions or nodal values of the solution. In all cases we approximate (1.2) by the homogeneous system of *ordinary differential equations (ODEs)*

$$\boldsymbol{u}' = (\mathrm{i}cA + \mathrm{i}\omega B)\boldsymbol{u}, \quad t \geq 0, \qquad \boldsymbol{u}(0) = \boldsymbol{\phi}, \tag{1.5}$$

where $c = (N+\frac{1}{2})^2/(2m\omega)$, $\boldsymbol{u}(t)$ is a vector, $\boldsymbol{\phi}$ is a representation of $u(x,0)$ with respect to the same basis, while A and B are $(2N+1)\times(2N+1)$ matrices. The reasons for the definition of the scaling factor c will become apparent in the sequel.

The exact solution to that problem is of course known,

$$\boldsymbol{u}(t) = \mathrm{e}^{\mathrm{i}t(cA+\omega B)}\boldsymbol{\phi},$$

except that direct calculation of the exponential is ineffective for all but minute step-sizes $\Delta t > 0$, due to the large size of ω.

The main idea in designing effective means to compute the above matrix exponential are based on the concept of *splitting* (McLachlan & Quispel 2002). Let $\tau = \mathrm{i}(\Delta t)$. The most ubiquitous is the *Strang splitting*

$$\mathrm{e}^{\tau(cA+\omega B)} \approx \mathrm{e}^{\frac{1}{2}\tau cA}\mathrm{e}^{\tau\omega B}\mathrm{e}^{\frac{1}{2}\tau cA}.$$

We commit above an error of $\mathcal{O}(\tau^3)$: such a splitting is said to be of *order* 2. However, the order of spatial discretization typically is much larger than 2 – in particular, for spectral methods the order of convergence is "spectral": the error decays faster than a reciprocal of any power of the number of degrees of freedom. It makes absolutely no sense to couple an exceedingly precise discretization in space with a crude discretization in time, and this motivates the quest for higher-order splittings. Typically, such splittings are of the form

$$\mathrm{e}^{\alpha_1\tau cA}\mathrm{e}^{\beta_1\tau\omega B}\mathrm{e}^{\alpha_2\tau cA}\mathrm{e}^{\beta_2\tau\omega B}\cdots\mathrm{e}^{\alpha_r\tau cA}\mathrm{e}^{\beta_r\tau\omega B}\mathrm{e}^{\alpha_r\tau cA}\cdots\mathrm{e}^{\beta_2\tau\omega B}\mathrm{e}^{\alpha_2\tau cA}\mathrm{e}^{\beta_1\tau\omega B}\mathrm{e}^{\alpha_1\tau cA}. \tag{1.6}$$

Note a number of features of (1.6). Firstly, it is *symmetric:* it reads the same from left or right. This has two important advantages. Such splittings are always of an even order and they preserve useful structural features of the underlying solution (Hairer, Lubich & Wanner 2006). Secondly, it *separates scales:* the argument of each exponential scales either like ω or like ω^{-1}. This is critical once we wish to evaluate the exponentials to good accuracy using, for example, Krylov subspace methods (Hochbruck & Lubich 1997).

There are a number of well-known means to obtain splittings (1.6) of higher order. The most popular is the *Yošida device* (Yošida 1990). Denote the Strang splitting by $S_1(\tau;A,B)$ and set

$$S_2(\tau;A,B) = S_1(\alpha_1\tau;A,B)S_1((1-2\alpha_1)\tau;A,B)S_1(\alpha_1\tau;A,B),$$

where $\alpha_1 = (1-2^{1/3})^{-1}$. Then the order of S_2 is 4 and it is also symmetric. In general, letting

$$S_{r+1}(\tau;A,B) = S_r(\alpha_r\tau;A,B)S_r((1-2\alpha_r)\tau;A,B)S_r(\alpha_r\tau;A,B)$$

with $\alpha_r = (2-2^{1/r})^{-1}$, $r\in\mathbb{N}$, results in a method of order $2(r+1)$. In principle, this allows methods of arbitrarily high order, except that this requires inordinately large number of terms. Denote the number of individual exponentials in S_r by s_r, hence $s_1 = 3$. It is easy to see that $s_{r+1} = 3s_r - 2$, because two exponentials with a scaled A argument can be amalgamated. Therefore $s_r = 2\cdot 3^{r-1}+1$ exponentials for order

$2r$: a fairly steep cost, clearly unacceptable once we want to match the high order of spatial discretization.

Although it is possible to improve the 'yield' of (1.6) somewhat, the number of exponentials imposes significant restriction on order (McLachlan & Quispel 2002). The reason is another feature of (1.6), namely that the arguments of all exponentials are scaled matrices A and B. The alternative, using the Baker–Campbell–Hausdorff formula, leads to arguments which are linear combinations of commutators. The latter course of action is considered flawed, because commutators tend to be large and so, in the context of the Schrödiner equation, should be avoided – having large ω is bad enough! A major point that we strive to make in this paper is that *this assertion is false!* Once we choose correctly the space discretization, commutators become small and we are free to use them in our quest to decrease the number of exponentials in a splitting.

In Section 2 we present a simple example of a finite difference method which leads to small commutators. This is followed by a general conditions on finite difference methods that ensure the consistency of (1.5), while producing small commutators.We then consider the highest-order finite difference methods supported on the grid: pseudospectral methods which exhibit up to order $2N$ with $2N+1$ grid points.

Section 3 is devoted to spectral collocation. While the standard spectral semi-discretization does not generate small commutators, the less familiar spectral collocation is up to the task.

Having established a framework for splittings that allow for nested commutators of A and B, in Section 4 we consider a highly effective new splitting algorithm, the *symmetric Zassenhaus formula.* This requires fairly technical and tedious algebra, which we relegate to Appendices A and B.

A word about stability. In all our discussion we assume that $c = \mathcal{O}(1)$ for $\omega \gg 1$, i.e. that $N \sim \omega^{1/2}$. In other words, our asymptotic analysis is in the frequency ω, requiring the dimension of the semi-discretization to grow as ω increases. This is at odds with classical stability analysis: fixed ω, $N \to \infty$. However, our purpose here is accuracy, rather than stability. Stability, formally, is assured: both τcA and τB are, in all our semi-discretizations, skew-Hermitian matrices. Hence so are all their commutators and their linear combinations. Therefore each component in any of the splittings considered here is an exponential of a normal matrix (hence, eigenvalue analysis suffices for stability (Iserles, Munthe-Kaas, Nørsett & Zanna 2000)) which, in addition, is skew-Hermitian (hence all exponentials are unitary, thus of unit ℓ_2 matrix norm).

2 Finite difference methods

2.1 The BCH formula and splittings

The core of the problem in splitting the exponential function is the fact that, unless X and Y commute, $\mathrm{e}^{tX}\mathrm{e}^{tY} \neq \mathrm{e}^{t(X+Y)}$. More specifically

$$\mathrm{e}^{tX}\mathrm{e}^{tY} = \mathrm{e}^{\mathrm{BCH}(tX,tY)}, \qquad \mathrm{BCH}(tX,tY) = \sum_{m=1}^{\infty} Z_m(X,Y)t^m, \tag{2.1}$$

where

$$
\begin{aligned}
&Z_1 = X+Y, \qquad Z_2 = \tfrac{1}{2}[X,Y], \qquad Z_3 = \tfrac{1}{12}[X,[X,Y]] - \tfrac{1}{12}[Y,[X,Y]],\\
&Z_4 = -\tfrac{1}{24}[X,[Y,[X,Y]]], \qquad Z_5 = \tfrac{1}{720}([Y,[Y,[Y,[X,Y]]]] - [X,[X,[X,[X,Y]]]])\\
&\qquad + \tfrac{1}{360}([Y,[X,[X,[X,Y]]]] - [X,[Y,[Y,[X,Y]]]]) + \tfrac{1}{120}([Y,[Y,[X,[X,Y]]]]\\
&\qquad - [X,[X,[Y,[X,Y]]]])
\end{aligned}
$$

and so on – this is the famous *Baker–Campbell–Hausdorff (BCH) formula* (Iserles et al. 2000, Suzuki 1977).

For our purposes, however, it is useful to replace the BCH formula (2.1) by *symmetric BCH*,

$$
\mathrm{e}^{tX}\mathrm{e}^{tY}\mathrm{e}^{tX} = \mathrm{e}^{\mathrm{sBCH}(tX,tY)}, \qquad \mathrm{sBCH}(tX,tY) = \sum_{m=0}^{\infty} W_m(X,Y)t^{2m+1}, \tag{2.2}
$$

where

$$
\begin{aligned}
&W_0 = 2X+Y,\\
&W_1 = -\tfrac{1}{6}[Y,[X,Y]] - \tfrac{1}{6}[X,[X,Y]],\\
&W_2 = \tfrac{7}{360}[X,[X,[X,[X,Y]]]] + \tfrac{1}{360}[Y,[Y,[Y,[X,Y]]]] - \tfrac{1}{90}[X,[Y,[Y,[X,Y]]]]\\
&\qquad + \tfrac{1}{45}[Y,[X,[X,[X,Y]]]] + \tfrac{1}{60}[X,[X,[Y,[X,Y]]]] + \tfrac{1}{30}[Y,[Y,[X,[X,Y]]]]
\end{aligned}
$$

and so on.

Using either (2.1) or (2.2), it is possible to obtain splittings which, for the same order, require less exponentials than (1.6), at the cost of allowing commutators to feature in the argument: a familiar splitting of the kind is *Yošida's second method* (Yošida 1990), often used in symplectic approximation of Hamiltonian problems with partitioned Hamiltonian function. More examples can be found in (McLachlan & Quispel 2002). Yet, as we have already mentioned, methods for the solution of (1.2) avoid commutators. The reason is as follows. The matrix A scales like $\mathcal{O}(1)$ for $\omega \gg 1$, while $\omega B = \mathcal{O}(\omega)$. Therefore, a naive estimate is

$$
\|\overbrace{[\omega B,[\omega B,[\omega B,\ldots,[\omega B}^{r\ \text{times}},A]\cdots]]]\| \le (2\omega)^r\|B\|^r\|A\| = \mathcal{O}(\omega^r) \tag{2.3}
$$

Assuming that this upper bound is realistic, this seems to indicate that al least some commutators grow very rapidly with r. The only means at our disposal to counteract this rapid growth in the size of the commutators is using suitably small time step $\tau = \Delta t$ and this analysis seems to indicate the need for $\tau = \mathcal{O}(\omega^{-1})$. This stability restriction is clearly unacceptable in realistic algorithms, hence the imperative to avoid commutators.

The main message of this paper is that the above reasoning is flawed, at least for some semi-discretization methods.

2.2 A simple example

Let N be a sufficiently large natural number. We impose in $[-1,1]$ an equidistant mesh on $[-1,1]$ with spacing $\Delta x = 1/(N+\frac{1}{2})$ and mesh points $x_k = k/(N+\frac{1}{2})$,

$k=-N,-N+1,\ldots,N$. We semi-discretize (1.2)–(1.4) on this grid and this yields the ODEs (1.5), where

$$\boldsymbol{u}(t)=\begin{bmatrix} u_{-N}(t)\\ u_{-N+1}(t)\\ \vdots\\ u_{N-1}(t)\\ u_N(t)\end{bmatrix}\approx\begin{bmatrix} u(x_{-N},t)\\ u(x_{-N+1},t)\\ \vdots\\ u(x_{N-1},t)\\ u(x_N,t)\end{bmatrix},\qquad \boldsymbol{\phi}=\begin{bmatrix} \phi(x_{-N})\\ \phi(x_{-N})\\ \vdots\\ \phi(x_{N-1})\\ \phi(x_N)\end{bmatrix}.$$

Both A and B are $(2N+1)\times(2N+1)$ matrices. A approximates the second derivative operator, while B is diagonal, with the values of the periodic function ϕ at the grid points along its diagonal,

$$B=\begin{bmatrix} V_{-N} & 0 & \cdots & \cdots & 0\\ 0 & V_{-N+1} & 0 & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & 0 & V_{N-1} & 0\\ 0 & \cdots & \cdots & 0 & V_N \end{bmatrix} \tag{2.4}$$

where $V_k=V(x_k)$.

The simplest example of this state of affairs is the familiar symmetric discretization of the second derivative,

$$u_{xx}(x,t)=\frac{1}{(\Delta x)^2}[u(x-\Delta x,t)-2u(x,t)+u(x+\Delta x,t)]+\mathcal{O}\big((\Delta x)^2\big).$$

Since $(\Delta x)^{-2}=(N+\frac{1}{2})^2$, This results in the *circulant*

$$A=\begin{bmatrix} -2 & 1 & 0 & \cdots & 1\\ 1 & -2 & 1 & \ddots & \vdots\\ 0 & \ddots & \ddots & \ddots & 0\\ \vdots & \ddots & \ddots & \ddots & 1\\ 1 & \cdots & 0 & 1 & -2 \end{bmatrix}. \tag{2.5}$$

Note in passing that $\mathrm{i}A,\mathrm{i}B\in\mathfrak{su}(2N+1)$, the set of $(2N+1)\times(2N+1)$ skew-Hermitian matrices. Since $\mathfrak{su}(2N+1)$ is a Lie algebra, a linear space closed under commutation, we deduce that all their commutators (and their linear combinations) are also skew-Hermitian. Consequently, as we have already commented by the end of Section 1, $\mathrm{e}^{\tau(cA+\omega B)}$, all the exponentials in our splittings and all their products are all unitary matrices.

Note further that, as we have already mentioned, A scales like $\mathcal{O}(1)$ for $N,\omega\gg 1$. Of greater interest is the constant $c=1/(2m\omega(\Delta x)^2)$. We choose N so that $c=\mathcal{O}(1)$ for $\omega\gg 1$ – this compels us to select $N=\mathcal{O}\big(\omega^{1/2}\big)$ (i.e., $\Delta x=\mathcal{O}\big(\omega^{-1/2}\big)$), a choice which is maintained in the sequel.

According to (2.3), the commutator $[\omega B,A]$ should be large since so is ωB. In our case, however, the upper bound in (2.3) is a gross overestimate: direct computation confirms that

$$([B,A])_{k,\ell} = \begin{cases} V_k - V_{k+1}, & \ell = k+1 \pmod{2N+1}, \\ V_{k+1} - V_k, & k = \ell+1 \pmod{2N+1}, \\ 0, & \text{otherwise} \end{cases}$$

$$\approx \begin{cases} -(\Delta x)V'(x_k), & \ell = k+1 \pmod{2N+1}, \\ (\Delta x)V'(x_k), & k = \ell+1 \pmod{2N+1}, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, bearing in mind that $\Delta x = \mathcal{O}(\omega^{-1/2})$, $\|[\omega B, A]\| = \mathcal{O}(\omega^{1/2})$. Moreover, each additional commutation with ωB scales like $\mathcal{O}(\omega^{1/2})$ and, in place of the upper bound (2.3), we have the estimate

$$\|[\overbrace{\omega B, [\omega B, [\omega B, \ldots, [\omega B}^{r\text{ times}}, A] \cdots]]]\| = \mathcal{O}\left(\omega^{r/2}\right) \tag{2.6}$$

– the order of magnitude is roughly the square root of what would have been the size of nested commutators in a general case. This is a crucial observation because each nested commutator is ultimately multiplied by a power of the time step τ, itself a small parameter, and this renders commutators small and manageable in practical computations.

2.3 General finite difference schemes

Is the estimate (2.6) specific to the simple matrix (2.5) or does it extend to other finite difference schemes?

Let us consider the ODE (1.5), where the matrix B is given by (2.4), while the elements of A obey

$$\sum_{\ell=-N}^{N} A_{k,\ell} = \sum_{\ell=-N}^{N} (\ell-k)A_{k,\ell} = \sum_{\ell=-N}^{N} (\ell-k)^3 A_{k,\ell} = 0, \qquad \sum_{\ell=-N}^{N} (\ell-k)^2 A_{k,\ell} = 2 \tag{2.7}$$

for $k = -N, \ldots, N$. (Note that all these conditions are trivially satisfied for A given by (2.5).) Let $v \in \mathrm{C}^3[-1,1]$ be periodic and $v_k = v(x_k)$, $k = -N, \ldots, N$. We thus have

$$\begin{aligned} \sum_{\ell=0}^{N-1} A_{k,\ell} v_\ell = \sum_{\ell=0}^{N-1} & A_{k,\ell}[v(x_k) + (\ell-k)\Delta x v'(x_k) + \tfrac{1}{2}(\ell-k)^2(\Delta x)^2 v''(x_k) \\ & + \tfrac{1}{6}(\ell-k)^3(\Delta x)^3 v'''(x_k) + \mathcal{O}\big((\Delta x)^4\big)] = v''(x_k) + \mathcal{O}\big((\Delta x)^2\big)\,, \end{aligned}$$

hence $A\boldsymbol{v}$ approximates v'' on the grid. Likewise,

$$\sum_{\ell=-N}^{N} (\ell-k)\eta A_{k,\ell} v_\ell = 2(\Delta x) v'(x_k) + \mathcal{O}(\Delta x)\,,$$

$$\sum_{\ell=-N}^{N} (\ell-k)^2 \eta A_{k,\ell} v_\ell = 2v(x_k) + \mathcal{O}(\Delta x)\,.$$

Sacrificing spurious generality, we observe that, with periodic boundary conditions, it makes sense to use the same discretization scheme at every point, 'wrapping round' the boundary. In other words, a sensible choice of the matrix A is as a *symmetric circulant,*

$$A_{k,\ell} = a_{N,k-\ell}, \quad (k-\ell) \bmod N, \qquad k,\ell = -N,\ldots,N$$

where $a_{N,-k} = a_{N,k}$. Thus, for example, for $N = 2$,

$$A = \begin{bmatrix} a_{2,0} & a_{2,1} & a_{2,2} & a_{2,2} & a_{2,1} \\ a_{2,1} & a_{2,0} & a_{2,1} & a_{2,2} & a_{2,2} \\ a_{2,2} & a_{2,1} & a_{2,0} & a_{2,1} & a_{2,2} \\ a_{2,2} & a_{2,2} & a_{2,1} & a_{2,0} & a_{2,1} \\ a_{2,1} & a_{2,2} & a_{2,2} & a_{2,1} & a_{2,0} \end{bmatrix}.$$

Note that, bearing in mind the periodicity of v, each product $A\boldsymbol{v}$ is thus a discrete convolution and can be computed by FFT in $\mathcal{O}(N\log N)$ operations. This is important because, ultimately, the computational expense in forming Krylov subspace approximations of the matrix exponential reduces to products of the form $A\boldsymbol{v}$ and $B\boldsymbol{v}$ and the latter costs just $\mathcal{O}(N)$ operations.

The *symbol* of the Toeplitz matrix A is

$$a_N(z) = \sum_{\ell=-N}^{N} a_{N,\ell} z^\ell, \qquad z \in \mathbb{C}.$$

We have

$$\sum_{\ell=-N}^{N} (\ell-k)^s A_{k,\ell} = \sum_{\ell=-N}^{N} \ell^s a_{N,\ell}, \qquad s \in \mathbb{Z}_+,$$

and it is a simple to confirm that

$$\sum_{\ell=-N}^{N} a_{N,\ell} = a_N(1), \qquad \sum_{\ell=-N}^{N} \ell a_{N,\ell} = a_N'(1), \qquad \sum_{\ell=-N}^{N} \ell^2 a_{N,\ell} = a_N'(1) + a_N''(1),$$
$$\sum_{\ell=-N}^{N} \ell^3 a_{N,\ell} = 2a_N'(1) + 3a_N''(1) + a_N'''(1).$$

Therefore, we can rewrite (2.7) in the form

$$a_N(1) = a_N'(1) = 0, \qquad a_N''(1) = 2, \qquad a_N'''(1) = -6. \tag{2.8}$$

Lemma 1 *Let A and B be $(2N+1)\times(2N+1)$ matrices, such that A is a symmetric circulant, $B = \operatorname{diag}\boldsymbol{V}$ and conditions (2.7) (alternatively, (2.8)) are satisfied for every $N \gg 1$. Then*

$$\lim_{\Delta x\to 0} [cA, B]\boldsymbol{v} = \frac{1}{2m\omega}[\partial_x^2, V\cdot]v, \tag{2.9}$$

where $V\cdot$ is the operation of multiplying a function by V.

Proof By straightforward differentiation, the action of the exact Lie bracket is

$$[\partial_x^2, V\cdot]v = \partial_x^2(Vv) - V\partial_x^2 v = V''v + 2V'v'.$$

On the other hand, given $k = -N, \ldots, N$ and bearing in mind our assumption that $c = \mathcal{O}(1)$,

$$[A,B]_{k,\ell} = (V_\ell - V_k)a_{N,k-\ell}, \qquad k,\ell = -N,\ldots,N. \tag{2.10}$$

Therefore, expanding functions about x_k, letting $V_k^{(i)} = V^{(i)}(x_k)$ etc. and using conditions (2.7),

$$\begin{aligned}
\sum_{\ell=-N}^{N} [(\Delta x)^{-2}A, B]_{k,\ell}v_\ell &= \frac{1}{(\Delta x)^2}\sum_{\ell=-N}^{N} a_{N,\ell}V_{k-\ell}v_{k-\ell} - \frac{1}{(\Delta x)^2}V_k \sum_{\ell=-N}^{N} a_{N,\ell}v_{k-\ell}\\
&= \frac{1}{(\Delta x)^2}\sum_{\ell=-N}^{N} a_{N,\ell}[V_k - (\Delta x)\ell V_k' + \tfrac{1}{2}(\Delta x)^2\ell^2 V_k'' - \tfrac{1}{6}(\Delta x)^3\ell^3 V_k''']\\
&\qquad \times [v_k - (\Delta x)\ell v_k' + \tfrac{1}{2}(\Delta x)^2\ell^2 v_k'' - \tfrac{1}{6}(\Delta x)^3\ell^3 V_k''']\\
&\quad - \frac{1}{(\Delta x)^2}V_k \sum_{\ell=-N}^{N} a_{N,\ell}[v_k - (\Delta x)\ell v_k' + \tfrac{1}{2}(\Delta x)^2\ell^2 v_k'' - \tfrac{1}{6}(\Delta x)^3\ell^3 V_k''']\\
&\qquad + \mathcal{O}\big((\Delta x)^2\big)\\
&= V_k''v_k + 2V_k'v_k' + \mathcal{O}\big((\Delta x)^2\big) = [\partial_x^2, V\cdot]v\Big|_{x=x_k} + \mathcal{O}\big((\Delta x)^2\big)\,.
\end{aligned}$$

This completes the proof. □

How large is the commutator $[B, A]$? Here we must distinguish between two strategies, a thread that will run through the sequel of this paper. The first, which we call *semi-finite (SF),* is that there exists some $s \in \mathbb{N}$ so that $a_{N,k} = 0$ for $|k| \geq s+1$ and every N – in other words, A is a banded circulant and the bandwidth is bounded as ω and N become infinite. The second, termed by us the *full matrix (FM) strategy,* is when no such finite bandwidth is imposed: typically in that case, the matrix A is full.

In the SF case it follows at once from (2.10) that

$$|[A,B]_{k,\ell}| \leq \tilde{c}(\Delta x), \tag{2.11}$$

where $\tilde{c} = s(\Delta x)\|V'\|_\infty \max_{|k|\leq s}|a_{N,k}|/2$.

In the FM case, e.g. for *pseudospectral methods* (Fornberg 1998), we assume additionally that for every fixed $k \in \mathbb{N}$

$$\lim_{N\to\infty} a_{N,k} = a_k, \qquad \text{such that} \qquad \lim_{k\to\infty} a_k = 0. \tag{2.12}$$

(Recall that $a_{N,-k} = a_{N,k}$.) Given $\delta > 0$, can can just choose $K(\delta)$ so that

$$|a_{N,k}| \leq \frac{\delta}{2}, \qquad |k| \geq K(\delta), \qquad N \geq |k|.$$

In that case, (2.10) and $\|V\|_\infty = 1$ imply that

$$|[A,B]_{k,\ell}| \le \begin{cases} \frac{1}{2}K(\delta)|a_{N,k-\ell}|\|V'\|_\infty \Delta x, & |k-\ell| \le K(\delta)-1, \\ \delta, & |k-\ell| \ge K(\delta), \end{cases} \qquad k,\ell = -N,\ldots,N. \tag{2.13}$$

We will demonstrate in the sequel that the highest-order finite difference choice of $a_{N,k}$, as well as the spectral collocation approach of Section 3, result in $a_k = \mathcal{O}\big(k^{-2}\big)$ in (2.12), therefore $K(\delta) = \lceil \delta^{-1/2} \rceil$. Thus, taking $\delta = (\Delta x)^{2/3}$ results in (2.13) in the upper bound

$$|[A,B]|_{k,\ell}| \le \tilde{c}(\Delta x)^{2/3}, \qquad k,\ell = -N,\ldots,N, \tag{2.14}$$

for some $\tilde{c} > 0$. Note the difference between (2.11) and (2.14): the price of considerably higher spatial accuracy is that the commutator is larger – but still considerably smaller than the naive bound of $\mathcal{O}(1)$.

The bounds (2.11) and (2.14) generalize immediately to any nested commutator.

Lemma 2 *Let C be a nested commutator in the free Lie algebra generated by A and B, with $r-q$ occurrences of A and q of ωB, $q,r \ge 1$. Then, assuming $\Delta x = \mathcal{O}\big(\omega^{-1/2}\big)$, it is true that*

$$\|C\| \le \hat{c}\omega^{\mu q}, \qquad \omega \gg 1, \tag{2.15}$$

for some $\hat{c} > 0$, where $\mu = 1/2$ and $\mu = 2/3$ in for SF and FM strategies, respectively.

Proof By a trivial extension of our analysis, each A 'contributes' $\mathcal{O}(1)$, while each commutation with B 'contributes' $\mathcal{O}((\Delta x)^\mu)$. (2.15) follows at once. □

2.4 The highest-order scheme

Given any $s \in \mathbb{N}$, we wish to approximate the second derivative of f at x by a linear combination of $f(x+k\Delta x)$, $k = -s,\ldots,s$.

Proposition 3 *The highest-order equispaced approximation to the second derivative using $2s+1$ function values is*

$$f''(x) = \frac{1}{(\Delta x)^2} \sum_{k=-s}^{s} a_{s,k} f(x+k\Delta x) + \mathcal{O}\big((\Delta x)^{2s}\big),$$

where

$$a_{s,k} = \begin{cases} \dfrac{2(-1)^{k-1}s!^2}{k^2(s-k)!(s+k)!}, & k \ne 0, \\ -\displaystyle\sum_{\ell \ne 0} a_{s,\ell}, & k = 0. \end{cases} \tag{2.16}$$

Proof Using the terminology of (Iserles 2008), we wish to find $a_{s,k}$, $k = -s,\ldots,s$, such that

$$\frac{1}{(\Delta x)^2} \sum_{k=-s}^{s} a_{s,k} \mathrm{E}^k = \mathrm{D}^2 + \mathcal{O}\big((\Delta x)^{2s}\big) = \frac{(\log \mathrm{E})^2}{(\Delta x)^2} + \mathcal{O}\big((\Delta x)^{2s}\big),$$

where D and E are the *differential* and *shift operators*, $\mathrm{E} = \mathrm{e}^{(\Delta x)\mathrm{D}}$. Since $\mathrm{E} = \mathrm{I} + \mathcal{O}(\Delta x)$, this is equivalent to

$$\sum_{k=-s}^{s} a_{s,k} z^k = (\log z)^2 + \mathcal{O}\big(|z-1|^{2s+2}\big), \qquad z \to 1.$$

Our first conclusion is that $\sum_{k=-s}^{s} a_{s,k} = 0$, confirming the formula for $a_{s,0}$ in (2.16). Differentiation and multiplication by z yield

$$\sum_{k=-s}^{s} k a_{s,k} z^k = 2\log z + \mathcal{O}\big(|z-1|^{2s+1}\big).$$

Differentiating again and multiplying by z^{s+1} results in

$$\sum_{k=0}^{2s} (k-s)^2 a_{s,k-s} z^k - 2z^s = \mathcal{O}\big(|z-1|^{2s}\big).$$

We have an $(2s)$-degree polynomial in z on the left and it can be $\mathcal{O}\big(|z-1|^{2s}\big)$ only if it is a constant multiple of $(1-z)^{2s}$,

$$\sum_{k=0}^{2s} (k-s)^2 a_{s,k-s} z^k = 2z^s + d(z-1)^{2s}.$$

Using $a_{s,-k} = a_{s,k}$, it is easy, though, to rewrite the left-hand side as

$$\sum_{k=1}^{s} k^2 a_{s,k} (z^{s-k} + z^{s+k}),$$

while the right-hand side is

$$\begin{aligned} & 2z^s + d(-1)^s \binom{2s}{s} z^s + d\sum_{k=0}^{2s} (-1)^k \binom{2s}{k} z^k \\ = {} & 2z^s + d(-1)^s \binom{2s}{s} z^s + d\sum_{k=1}^{s} (-1)^{s-k} \binom{2s}{s-k} (z^{s-k} + z^{s+k}). \end{aligned}$$

Comparing coefficients, we have $d = (-1)^{s-1} s!^2/(2s)!$ and (2.16) follows for $k \neq 0$. □

Using the Sterling formula, it is trivial to prove that for any fixed $k \in \mathbb{Z} \setminus \{0\}$

$$\lim_{s\to\infty} \frac{2(-1)^{k-1} s!^2}{k^2 (s-k)!(s+k)!} = \frac{2(-1)^{k-1}}{k^2}.$$

Therefore the bound (2.14) holds in this case.

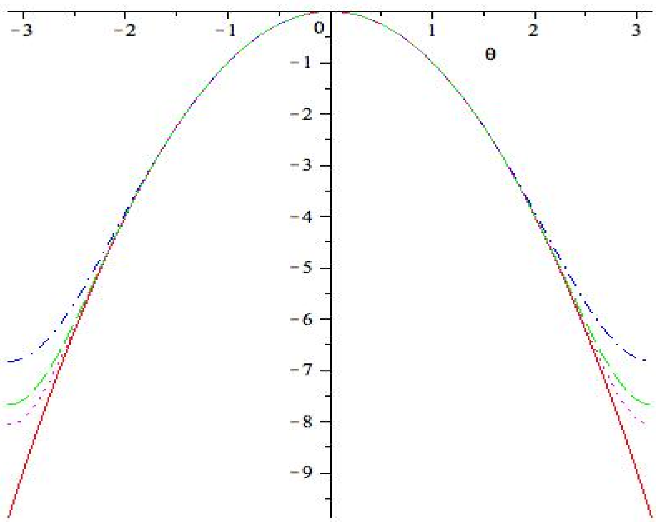

Figure 2.1: Finite differences: $-\theta^2$ (thick solid line) and $a_s(\mathrm{e}^{\mathrm{i}\theta})$ for $s = 5, 10, 15$ (dash-dot, dash and dot lines, respectively) for $\theta \in [-\pi, \pi]$.

Employing (2.16), we obtain a matrix A with the symbol

$$a_s(z) = \sum_{\substack{k=-s\\k\neq 0}} \frac{2(-1)^{k-1} s!^2}{k^2 (s-k)!(s+k)!} z^k - \sum_{\substack{k=-s\\k\neq 0}} \frac{2(-1)^{k-1} s!^2}{k^2 (s-k)!(s+k)!},$$

while it follows at once from the order conditions that

$$\sum_{k=-s}^{s} a_{s,k} \cos k\theta = -\theta^2 + \mathcal{O}\big(\theta^{2s+2}\big), \qquad \theta \to 0.$$

Fig. 2.1 displays how well $a_s(\mathrm{e}^{\mathrm{i}\theta})$ approximates $-\theta^2$ for different values of $s \geq 1$, demonstrating how the increasing order at the origin translates into approximation in the entire interval $[-\pi, \pi]$.

Consequently,

$$a_s(\mathrm{e}^{\mathrm{i}\theta}) = \sum_{k=-s}^{s} a_{s,k} [1 - \tfrac{1}{2} k^2 \theta^2 + \mathcal{O}\big(\theta^4\big)] = -\theta^2 + \mathcal{O}\big(\theta^{2s+2}\big), \qquad \theta \to 0$$

and it is easy to confirm that (2.8) holds.

In Fig. 2.2 we display the error committed once we approximate the second derivative of a periodic function by the Nth finite difference scheme (2.16) (i.e., taking $s = N$) for three different values of N and for two functions: $u(x) = (2 + \sin \pi x)^{-1}$ and $u(x) = \mathrm{e}^{\cos \pi x}$. Note that the second function is entire, while the first is meromorphic: this explains why the precision is so much higher in the bottom row. However,

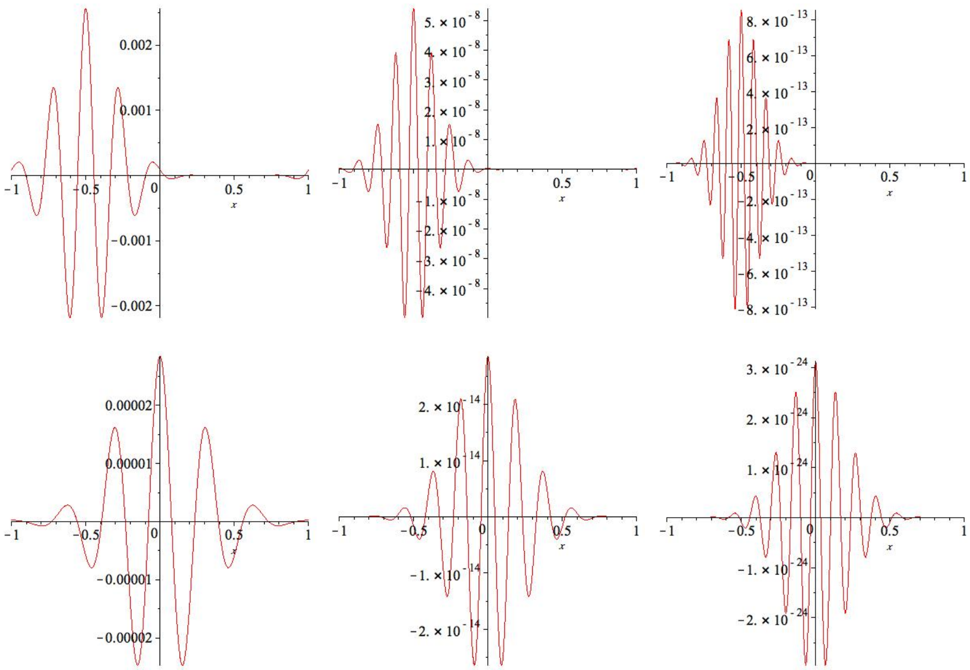

Figure 2.2: The error committed in approximating the second derivative of $u(x) = (2+\sin\pi x)^{-1}$ (top row) and of $u(x) = \mathrm{e}^{\cos\pi x}$ by finite differences, using $s = N = 10$ (left column), $s = N = 20$ (central column) and $s = N = 30$ (right column).

the remarkable observation is that in both cases, using just a small number of grid points, we obtain very good accuracy.

Fig. 2.1 seems to indicate that $\lim_{s\to\infty} a_s(\mathrm{e}^{\mathrm{i}\theta}) = -\theta^2$. To prove this is indeed true, we recall that $a_{s,k} \stackrel{s\to\infty}{\longrightarrow} (-1)^{k-1}/k^2$ for $k \neq 0$, therefore

$$
\begin{aligned}
a_s(\mathrm{e}^{\mathrm{i}\theta}) \stackrel{s\to\infty}{\longrightarrow} F(\theta) &:= 4\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^2}\cos k\theta - 4\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k^2} \\
&= \ 4\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^2}\cos k\theta - \frac{\pi^2}{3}.
\end{aligned}
$$

Note that $F(0) = 0$. Moreover,

$$
F'(\theta) = 4\sum_{k=1}^{\infty} \frac{(-1)^k}{k}\sin k\theta = -4\arctan\frac{\sin\theta}{1+\cos\theta} = -2\theta.
$$

Therefore indeed $F(\theta) = -\theta^2$, confirming the observation from Fig. 2.1. Moreover, since the Euclidean norm of an infinite Toeplitz matrix is the maximum of the modulus of its symbol over the unit circle, we deduce that for $s \gg 1$ it is true that $\|A\| \approx \pi^2$.

3 Spectral collocation method

3.1 The method

As we have already mentioned, a spectral method is the most popular means of semi-discretizing (1.2) (Jin et al. 2011). This results in a diagonal matrix A with $A_{k,k} = -k^2\pi^2/(N+\frac{1}{2})^2$ and a *dense* matrix B. Consequently, there is absolutely no reason for commutators to be small. Fortunately, this is not the case with *spectral collocation* (Fornberg 1998). Like in Section 2, we impose in $[-1,1]$ the grid $\{x_k\}_{k=-N}^{N}$ and interpolate periodic functions there by trigonometric polynomials $1, \cos\pi x, \ldots, \cos\pi N x$. It is elementary that the scaled *Dirichlet kernel*

$$K_N(x) = \frac{1}{2N+1}\sum_{k=-N}^{N}\cos k\pi x = \frac{\sin((N+\frac{1}{2})\pi x)}{(2N+1)\sin(\frac{1}{2}\pi x)} \tag{3.1}$$

obeys

$$K_N(x_\ell) = \begin{cases} 1, & \ell = 0, \\ 0, & \ell \neq 0, \end{cases}$$

therefore $K_N(x - x_r)$ is the cardinal function of interpolation at x_r and the requisite interpolating trigonometric polynomial to a periodic function f is the discrete convolution $\sum_{k=-N}^{N} f(x_k)K_N(x-x_k)$. It follows, differentiating (3.1) twice, that $A_{k,\ell} = a_{N,\ell-k}$, where

$$a_{N,m} = \frac{K_N''(x_m)}{(N+\frac{1}{2})^2} = \begin{cases} -\dfrac{\pi^2 N(N+1)}{2(N+\frac{1}{2})^2}, & m = 0, \\ (-1)^{m+1}\dfrac{\pi^2\cos\frac{\pi m}{2N+1}}{2(N+\frac{1}{2})^2\sin^2\frac{\pi m}{2N+1}}, & 1 \le |m| \le N. \end{cases} \tag{3.2}$$

It is trivial to verify that, similarly to the finite-difference case $\lim_{m\to\infty} a_{N,m} = 2(-1)^{m+1}/(m^2)$ and $\|A\| = \mathcal{O}(1)$.

We again define $a_N(z) = \sum_{k=-N}^{N} a_{N,k}z^k$, the symbol of A. A number of symbols, acting on the unit disc, are displayed in Fig. 3.2 and it can be seen that they also seem to approximate $-\theta^2$. Comparison with Fig. 2.1 is instructive: it seems that, while for finite differences, $a_N(\mathrm{e}^{\mathrm{i}\theta})$ always lives between the real axis and $-\theta^2$, for spectral collocation it is always underneath $-\theta^2$. We do not pursue further this observation in the current paper, since it is of no relevance to its argument.

It is easy to see that $\sum_{\ell=-N}^{N}\ell^{2s+1}a_{N,\ell} = 0$ for every $s \in \mathbb{Z}_+$ and it is possible to prove, at considerably greater difficulty, that $\sum_{\ell=-N}^{N} a_{N,\ell} = 0$. However, computer simulation indicates that $\sum_{\ell=-N}^{N}\ell^2 a_{N,\ell} \neq 2$, although the difference decays rapidly with growing N. Therefore we cannot Lemma 1 to argue that, as $N \to \infty$, we obtain the correct Lie bracket. Fortunately, a direct proof of this fact follows from properties of Dirichlet kernels.

Lemma 4 *Let matrices $A = (a_{N,\ell-k})$ and $B = \mathrm{diag}(V(x_k))$ originate in spectral collocation, while $v_k = v(x_k)$ for a periodic $v \in \mathrm{C}^2[-1,1]$. Then*

$$\lim_{\Delta x\to 0}[(\Delta x)^{-2}A, B]\boldsymbol{v} = [\partial_x^2, V\,\cdot]v. \tag{3.3}$$

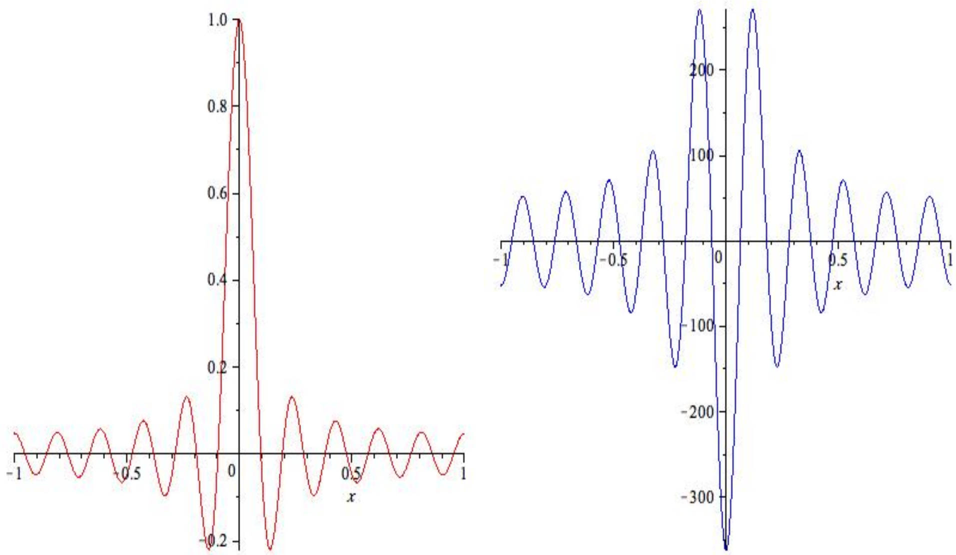

Figure 3.1: The cardinal function K_{10} and its second derivative K_{10}'', respectively.

Proof Since $x_{\ell-k} = x_\ell - x_k$ and K_N is an even function, by direct computation, for every $k = -N, \ldots, N$,

$$\begin{aligned}\frac{[(AB-BA)\boldsymbol{v}]_k}{(\Delta x)^2} &= -\sum_{\ell=-N}^{N} [V(x_\ell) - V(x_k)] K_N''(x_{k-\ell}) v(x_k) \\ &= -\frac{1}{N+\frac{1}{2}} \int_{-1}^{1} K_N''(x_k - x) V(x) v(x)\,\mathrm{d}x \\ &\quad + \frac{V(x_k)}{N+\frac{1}{2}} \int_{-1}^{1} K_N''(x_k - x) v(x)\,\mathrm{d}x + \mathcal{O}(\Delta x) \\ &= -\int_{-1}^{1} \mathcal{D}_N''(x_k - x) V(x) v(x)\,\mathrm{d}x + V(x_k) \int_{-1}^{1} \mathcal{D}_N''(x_k - x) v(x)\,\mathrm{d}x \\ &\quad + \mathcal{O}(\Delta x)\,,\end{aligned}$$

where $\mathcal{D}_N = K_N/(N+\frac{1}{2})$ is the Dirichlet kernel. Integrating by parts, it is immediate that for any periodic $g \in \mathrm{C}^2[-1,1]$

$$\int_{-1}^{1} \mathcal{D}_N''(y - x) g(x)\,\mathrm{d}x = \int_{-1}^{1} \mathcal{D}_N(y - x) g''(x)\,\mathrm{d}x, \qquad y \in [-1,1],$$

while it is an elementary feature of Dirichlet kernels that

$$\lim_{N\to\infty} \int_{-1}^{1} \mathcal{D}_N(y - x) g(x)\,\mathrm{d}x = g(y).$$

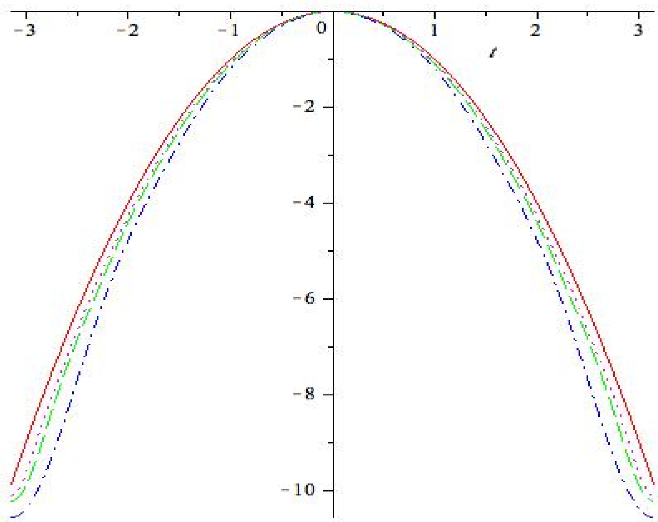

Figure 3.2: Spectral collocation: $-\theta^2$ (thick solid line) and $a_s(\mathrm{e}^{\mathrm{i}\theta})$ for $s = 5, 10, 15$ (dash-dot, dash and dot lines, respectively) for $\theta \in [-\pi, \pi]$.

Therefore

$$\lim_{\Delta x \to 0} (\Delta x)^{-2}[A, B]\boldsymbol{v} = -(Vv)'' + Vv'' = [\partial_x^2, V\cdot]v.$$

$\square$

In Fig. 3.3 we display the error of spectral collocation for the same two cases (and with the same number of grid points) as in Fig. 2.2. The degree of improvement once we replace finite differences by spectral collocation is startling but should not be surprising. Large as the order of the finite-difference method (2.16) with $s = N$ might be, ultimately it cannot compete with spectral convergence!

But how does spectral collocation compare with the 'real deal' spectral method? The latter is nothing but standard Fourier expansion, truncated for modes $\geq |N| + 1$ and differentiated twice, and its error is reported in Fig. 3.4. Remarkably, both errors are very similar – spectral collocation is somewhat smaller, but this is a fluke. This is not surprising. The loose phrase 'accuracy' refers to $\mathrm{L}_\infty[-1, 1]$ error. A truncated Fourier expansion is an $\mathrm{L}_2[-1, 1]$ projection. It is quite good L_∞ projection: according to the *Lozinski–Khakskiladze Theorem* (Lozinski 1948), a truncated Fourier operator is the best L_∞ linear projection. Specifically, the *Lebesgue constant* of truncated Fourier expansion is $(2/\pi)\log N + c_1$, where $c_1 \in \mathbb{R}$ (Fejér 1910). However, the Lebesgue constant of trigonometric interpolation at equally-spaced points (i.e., of spectral collocation) is almost as good: it is $(2/\pi)^2 \log N + c_2$ for some c_2 (Galkin 1971). To all intents and purposes, spectral collocation is just as good as a conventional spectral method.

However, once we consider the size of commutators, spectral collocation enjoys great advantage. Sinc the matrix B is diagonal, with values of V along its diagonal,

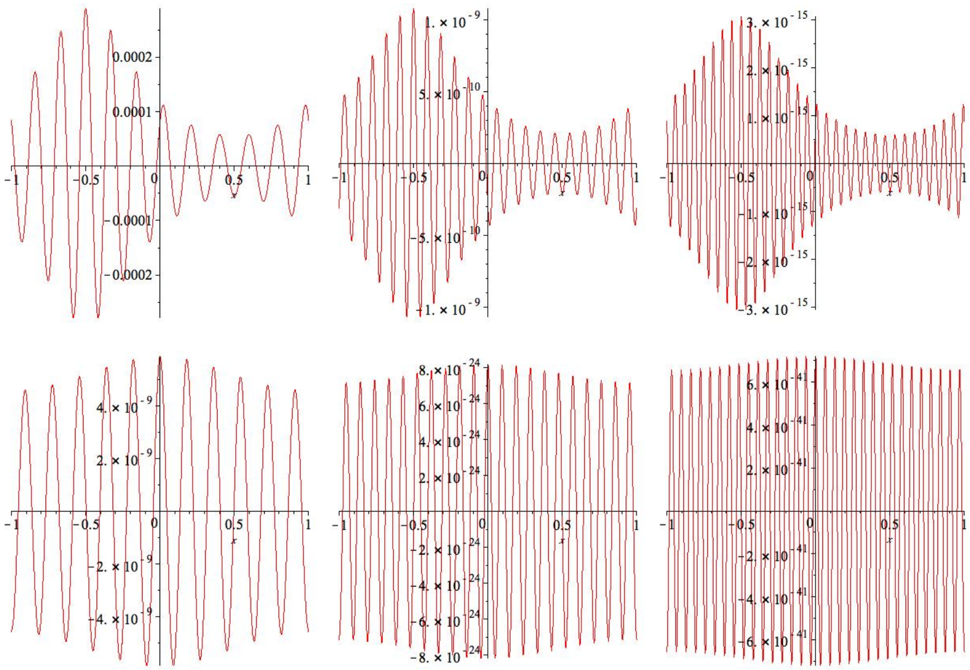

Figure 3.3: The error committed in approximating the second derivative of $u(x) = (2+\sin \pi x)^{-1}$ (top row) and of $u(x) = \mathrm{e}^{\cos \pi x}$ by spectral collocation using $s = N = 10$ (left column), $s = N = 20$ (central column) and $s = N = 30$ (right column).

Lemma 2 remains valid and the bound (2.15) holds with $\mu = \frac{2}{3}$.

4 Multiscale exponential splittings

4.1 Symmetric Zassenhaus splitting

Our point of departure is the linear ODE (1.5), where we assume that $\Delta x = \mathcal{O}\big(\omega^{-1/2}\big)$ and wish to express its solution by approximating $\mathrm{e}^{\tau(cA+\omega B)}$ as a product of simpler exponentials. However, our definition of 'simple' is at odds with the splitting (1.6): we are perfectly willing to accept arguments which are linear combinations of commutators but insist that each argument scales like a non-negative power of ω. In other words, we approximate

$$\mathrm{e}^{\tau(cA+\omega B)} \approx \mathrm{e}^{\omega^{s_1} U_1(\tau;cA,B)} \mathrm{e}^{\omega^{s_2} U_2(\tau;cA,B)} \cdots \mathrm{e}^{\omega^{s_M} U_M(\tau;cA,B)} \tag{4.1}$$

where $\tau = \mathrm{i}\Delta t$ is of small magnitude, for some $M \geq 3$ and $s_1, \ldots, s_M \in \mathbb{Z}_+$. We further stipulate that (4.1) is symmetric: $U_{M+1-i} = U_i$ for $i = 1, 2, \ldots, M$.

We commence from the *Zassenhaus splitting*

$$\mathrm{e}^{\tau(X+Y)} = \mathrm{e}^{\tau X} \mathrm{e}^{\tau Y} \mathrm{e}^{\tau^2 V_2(X,Y)} \mathrm{e}^{\tau^3 V_3(X,Y)} \mathrm{e}^{\tau^4 V_4(X,Y)} \cdots, \tag{4.2}$$

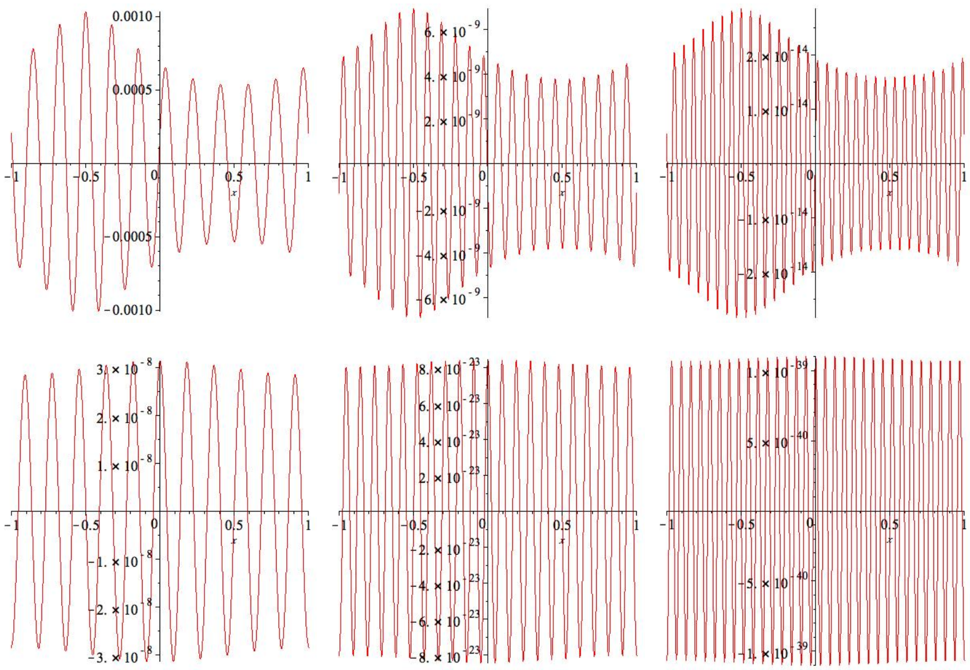

Figure 3.4: The error committed in approximating the second derivative of $u(x) = (2+\sin\pi x)^{-1}$ (top row) and of $u(x) = \mathrm{e}^{\cos\pi x}$ by Fourier series using $s = N = 10$ (left column), $s = N = 20$ (central column) and $s = N = 30$ (right column).

where

$$
\begin{aligned}
V_2(X,Y) &= -\tfrac{1}{2}[X,Y],\\
V_3(X,Y) &= \tfrac{1}{3}[Y,[X,Y]] + \tfrac{1}{6}[X,[X,Y]],\\
V_4(X,Y) &= -\tfrac{1}{24}[X,[X,[X,Y]]] - \tfrac{1}{8}[Y,[X,[X,Y]]] - \tfrac{1}{8}[Y,[Y,[X,Y]]]
\end{aligned}
$$

(Oteo 1991). However, we make two crucial changes to the splitting (4.2): firstly, we wish to expand in powers of ω, rather than τ. Secondly, we wish the splitting to be symmetric. Specifically, given $s \in \mathbb{N}$, we seek functions $R_i(\tau; cA, B)$, $i = 0, \ldots, s-1$ and $Q_s(\tau; cA, B)$ such that

$$
\begin{aligned}
\mathrm{e}^{\tau(cA+\omega B)} = {}& \mathrm{e}^{R_0(\tau;cA,B)}\mathrm{e}^{\omega R_1(\tau;cA,B)} \cdots \mathrm{e}^{\omega^{s-1}R_{s-1}(\tau;cA,B)}\mathrm{e}^{\omega^s Q_s(\tau;cA,B)}\\
&\times \mathrm{e}^{\omega^{s-1}R_{s-1}(\tau;cA,B)} \cdots \mathrm{e}^{\omega R_1(\tau;cA,B)}\mathrm{e}^{R_0(\tau;cA,B)}.
\end{aligned} \tag{4.3}
$$

This results in the approximate *sth symmetric Zassenhaus splitting*

$$
\begin{aligned}
&\mathrm{e}^{R_0(\tau;cA,B)}\mathrm{e}^{\omega R_1(\tau;cA,B)} \cdots \mathrm{e}^{\omega^{s-2}R_{s-2}(\tau;cA,B)}\mathrm{e}^{2\omega^{s-1}R_{s-1}(\tau;cA,B)}\mathrm{e}^{\omega^{s-2}R_{s-2}(\tau;cA,B)}\\
&\quad\times \cdots \mathrm{e}^{\omega R_1(\tau;cA,B)}\mathrm{e}^{R_0(\tau;cA,B)}.
\end{aligned} \tag{4.4}
$$

Our next step is an explicit construction of the splitting (4.3).

4.2 A recursive algorithm

We set

$$Q_0(\tau; A, B) = \tau c A + \tau\omega B.$$

Given $\ell \in \mathbb{Z}_+$, suppose that we have already obtained $R_i(\tau; A, B)$ for $i = 0, \ldots, \ell - 1$, as well as

$$Q_\ell(\tau; A, B) = \sum_{m=\ell}^{\infty} \omega^m \mathcal{W}_m^{[\ell]}(\tau; A, B)$$

– thus, $\mathcal{W}_0^{[0]} = \tau c A$, $\mathcal{W}_1^{[0]} = \tau B$ and $\mathcal{W}_m^{[0]} = O$ for $m \geq 2$. We set

$$X = -\tfrac{1}{2}\omega^\ell \mathcal{W}_\ell^{[\ell]}, \qquad Y = Q_\ell$$

– note that $2X + Y = \sum_{m=\ell+1}^{\infty} \omega^m \mathcal{W}_m^{[\ell]} = \mathcal{O}\big(\omega^{\ell+1}\big)$.

Recalling the symmetric BCH operator (2.2), we set

$$R_\ell(\tau, A, B) = \tfrac{1}{2}\omega^\ell \mathcal{W}_\ell^{[\ell]}, \qquad Q_{\ell+1}(\tau; A, B) = \mathrm{sBCH}(X, Y). \tag{4.5}$$

Note that $R_\ell = \mathcal{O}\big(\omega^\ell\big)$, as required. Since (cf. (2.2))

$$Q_{\ell+1}(\tau; A, B) = \mathrm{sBCH}(-\tfrac{1}{2}\omega^\ell \mathcal{W}_\ell^{[\ell]}, \omega^\ell \mathcal{W}_\ell^{[\ell]} + \mathcal{O}\big(\omega^{\ell+1}\big)) = \mathcal{O}\big(\omega^{\ell+1}\big)\,,$$

it follows that there exist $\mathcal{W}_j^{[\ell+1]}$, $j \geq \ell + 1$, such that

$$Q_{\ell+1}(\tau; A, B) = \sum_{j=\ell+1}^{\infty} \omega^j \mathcal{W}_j^{[\ell+1]}$$

and the definition of the recursive step is complete.

In a computation of this kind the devil is in the detail – and there is plenty of detail! Fortunately, this computation needs be done just once, tabulated and subsequently used in different implementations of our algorithm (with either finite differences or spectral collocation). Our main resource is a tabulation of all the commutator terms in symmetric BCH splitting up to order 19, written in the Hall basis (Murua 2010). Recall that each term W_m in (2.2) is a linear combination of nested commutators in the free Lie algebra $\mathcal{F}$ over the alphabet $\{A, B\}$.[1]

It is convenient to equip each nested commutator in $\mathcal{F}$ with a *grade* $\kappa \in \mathbb{N}$, namely the number of 'letters' A and B therein – once we replace A and B by $\tau c A$ and $\tau\omega B$, this is the same as saying that the term in question is $\mathcal{O}(\tau^\kappa)$. Note that grade is inherited through commutation: if Z_1 and Z_2 are of grades κ_1 and κ_2 respectively, $[Z_1, Z_2]$ is of grade $\kappa_1 + \kappa_2$. Likewise, we say that the *girth* $\sigma \in \mathbb{Z}_+$ of a term is the number of times B occurs in the term in question. The girth propagates like a grade, consistently with commutation, but there is a crucial difference. An important observation, at the heart of many calculations in free Lie algebras (cf. for example (Iserles et al. 2000)) is that the dimension d_κ of the linear space of all grade-κ terms

[1] The reader will be well advised to bear in mind that, while in (2.2) we expand in powers of t, our current concern is in an expansion in powers of ω.

in $\mathcal{F}$ is quite small, for example $d_0=0$, $d_1=2$, $d_2=1$, $d_3=2$, $d_4=3$, $d_5=6$, $d_6=9$ and $d_7=18$. Such terms can be conveniently expressed in terms of a Hall basis, cf. Table 1 in Appendix A. However, the number of distinct terms of girth ≥ 1 in the basis is infinite: observe that

$$[\overbrace{A,[A,[A,\ldots,A}^{r\text{ times}},B]]],$$

of grade $r+1$, is of girth 1 for every $r\in\mathbb{N}$.

The assembly of the $\mathcal{W}_m^{[\ell+1]}$ is painstaking, yet routine. Essentially, we need to identify all nested commutators of terms of the form $\omega^{j_i}\mathcal{W}_{j_i}^{[\ell]}$ so that the sum of the j_is is m. Thus, for example, for $\ell=0$ we use Table 1 to calculate

$$\begin{aligned}
\mathcal{W}_1^{[1]} &= \mathcal{W}_1^{[0]} + \tfrac{1}{24}[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]] + \tfrac{1}{1920}[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]] + \mathcal{O}(\tau^7)\,,\\
\mathcal{W}_2^{[1]} &= \tfrac{1}{12}[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]] - \tfrac{1}{1440}[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]]\\
&\quad - \tfrac{1}{1440}[\mathcal{W}_0^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]] - \tfrac{1}{720}[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]]\\
&\quad + \mathcal{O}(\tau^7)\,,\\
\mathcal{W}_3^{[1]} &= \tfrac{1}{240}[\mathcal{W}_1^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]] - \tfrac{1}{720}[\mathcal{W}_0^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]]\\
&\quad - \tfrac{1}{720}[\mathcal{W}_0^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]] - \tfrac{1}{360}[[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}],[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]\\
&\quad + \mathcal{O}(\tau^7)\,,\\
\mathcal{W}_4^{[1]} &= -\tfrac{1}{720}[\mathcal{W}_1^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_1^{[0]},[\mathcal{W}_0^{[0]},\mathcal{W}_1^{[0]}]]]] + \mathcal{O}(\tau^7)\,,\\
\mathcal{W}_j^{[1]} &= \mathcal{O}(\tau^7)\,, \qquad j\geq 5,
\end{aligned}$$

while for $\ell=1$,

$$\begin{aligned}
\mathcal{W}_2^{[2]} &= \mathcal{W}_2^{[1]},\\
\mathcal{W}_3^{[2]} &= \mathcal{W}_3^{[1]},\\
\mathcal{W}_4^{[2]} &= \mathcal{W}_4^{[1]} + \tfrac{1}{24}[[\mathcal{W}_2^{[1]},\mathcal{W}_1^{[1]}],\mathcal{W}_1^{[1]}],\\
\mathcal{W}_5^{[2]} &= \mathcal{W}_5^{[1]} + \tfrac{1}{24}[[\mathcal{W}_3^{[1]},\mathcal{W}_1^{[1]}],\mathcal{W}_1^{[1]}] + \tfrac{1}{12}[[\mathcal{W}_2^{[1]},\mathcal{W}_1^{[1]}],\mathcal{W}_2^{[1]}],\\
\mathcal{W}_6^{[2]} &= \mathcal{W}_5^{[1]} + \tfrac{1}{24}[[\mathcal{W}_3^{[1]},\mathcal{W}_1^{[1]}],\mathcal{W}_1^{[1]}] + \tfrac{1}{12}[[\mathcal{W}_2^{[1]},\mathcal{W}_1^{[1]}],\mathcal{W}_2^{[1]}]
\end{aligned} \tag{4.6}$$

and so on. Unfortunately, this is not all, because this procedure generates terms which do not belong to the Hall basis. In that case we express them as linear combinations of terms in the basis, a fairly straightforward procedure, and we are done.

We present in Appendix B the formulæ for $\mathcal{W}_j^{[\ell]}$, $\ell\in\mathbb{Z}_+$, given up to $\mathcal{O}(\tau^9)$. Note that, although we have derived the formulæ there by hand, it is perfectly possible to do so, to an arbitrary odd power of τ, using symbolic algebra.

4.3 The structure of the $\mathcal{W}_m^{[\ell]}$s

As can be observed from Appendix B, the leading term of each $\mathcal{W}_m^{[\ell]}$, $m\geq\ell$, scales like $\zeta_m^{[\ell]} := \omega^m c^{r-m}\tau^r$, where m is the number of Bs in this term and r is its girth –

the latter must be odd, because each term in (2.2), due to symmetry, always consist of an odd number of letters.

Since the leading term, having the least power of τ, is the one of least girth, it is easy to see that, for $\ell \geq 2$, it is the same as the girth of the leading term in $\mathcal{W}_m^{[\ell-1]}$: this can be seen at once from (4.6) and trivially generalised to all $\ell \geq 2$, because the least-girth term is inherited from $\mathcal{W}_m^{[\ell-1]}$. This implies that $\zeta_m^{[\ell]} = \zeta_m^{[\ell-1]}$ and it is enough to determine $\zeta_m^{[1]}$.

The leading term in $\mathcal{W}_m^{[1]}$ is the least-girth term consisting of exactly m Bs. However, the only 'building bricks' are $\mathcal{W}_0^{[0]} = \tau c A$ and $\mathcal{W}_1^{[0]} = \tau B$. Therefore, least-girth term contain m occurrences of B and either a single or two occurrences of A, so that r is odd. In principle, there can be several such terms, but for each we thus have $\zeta_m^{[1]} = \omega^m c\tau^{m+1}$ when m is even and $\zeta_m^{[1]} = \omega^m c^2 \tau^{m+2}$ when it is odd. Since $\zeta_\ell^{[\ell]} = \zeta_\ell^{[1]}$, we deduce that for every $\ell \in \mathbb{N}$ it is true that

$$R_\ell(\tau; A, B) = \mathcal{O}\Big(\omega^\ell c^{\frac{1}{2}[3-(-1)^\ell]} \tau^{2\lfloor(\ell+1)/2\rfloor+1}\Big). \tag{4.7}$$

The estimate (4.7) is valid regardless of what are the matrices A and B, as long as they are $\mathcal{O}(1)$ in the underlying parameters. However, each term in Q_ℓ has at least ℓ commutations with B, and we know from Subsection 2.3 that, within our setting (either finite differences or spectral collocation) each such commutation scales the term in question by $\omega^{-1/2}$ in the SF case and as $\Delta x \to 0$, $\omega^{-1/3}$ with the FM strategy. Therefore, we may replace (4.7) by

$$R_\ell(\tau; cA, B) = \begin{cases} \mathcal{O}\Big(\omega^{\ell/2} c^{\frac{1}{2}[3-(-1)^\ell]} \tau^{2\lfloor(\ell+1)/2\rfloor+1}\Big), & \text{Semi-finite strategy,} \\ \mathcal{O}\Big(\omega^{2\ell/3} c^{\frac{1}{2}[3-(-1)^\ell]} \tau^{2\lfloor(\ell+1)/2\rfloor+1}\Big), & \text{Full matrix strategy.} \end{cases} \tag{4.8}$$

In the SF case $R_\ell = \mathcal{O}\big(\omega^{\ell/2}\tau^{\ell+1}\big)$ when ℓ is even, $\mathcal{O}\big(\omega^{\ell/2}\tau^{\ell+2}\big)$ otherwise. Therefore, we may use $\Delta t = \mathcal{O}\big(\omega^{-1/2}\big)$, of the same order of magnitude as Δx. In the FM case we have $R_\ell = \mathcal{O}\big(\omega^{2\ell/3}\tau^{\ell+1}\big)$ and $R_\ell = \mathcal{O}\big(\omega^{2\ell/3}\tau^{\ell+2}\big)$ for even and odd ℓ, respectively. Therefore, the restriction on the time step is somewhat more stringent, $\Delta t = \mathcal{O}\big(\omega^{-2/3}\big) = \mathcal{O}\big((\Delta x)^{4/3}\big)$.

5 Conclusions

We have considered in this paper two families of spatial discretizations of the linear Schrödinger equation which share the feature that the commutators of the matrices corresponding to the 'derivative' and 'function multiplication' operators are small. Such schemes can be applied in either the SF strategy, which keeps the spatial order bounded as $\Delta x \to 0$ or the FM strategy, whereby the order may become infinite or even the convergence might proceed at spectral speed.

The relatively modest size of the commutators allows for the use of splittings that incorporate commutators, and this motivated our introduction of the symmetric Zassenhaus splitting. It allowed us to separate scales, since each argument in the splitting scales like a power of ω. Practical implementation of this splitting requires $\Delta t \sim \Delta x$ in the SF case and $\Delta t \sim (\Delta x)^{4/3}$ in the FM case.

This is a preliminary study into this subject matter and the authors are fully aware that this paper opens just as many questions as it purports to answer. In particular, the following issues are of an interest:

1. *SF or FM?* Most numerical analysts, other things being equal, would prefer the FM strategy, in particular in the case of spectral collocation, since it leads to spectral convergence. (Note that the cost of underlying linear algebra, as long as we compute exponentials with Krylov subspace methods, is essentially the same, once we use FFT.) However, things are not equal! The SF strategy allowed for larger time step and, given that the time discretization is invariably of restricted order, arguably this offsets the improved spatial accuracy of FM. This, however, is an issue that can be completely resolved only upon extensive numerical experimentation.

2. *How to implement the Zassenhaus splitting?* The standard means to compute $\mathrm{e}^C\boldsymbol{v}$, where $\boldsymbol{v} \in \mathbb{C}^{2N+1}$ and $C \in \mathfrak{su}(2N+1)$ is a Krylov subspace method (Hochbruck & Lubich 1997). To this end we need to compute recursively $\boldsymbol{w}_1 := C\boldsymbol{v}$, $\boldsymbol{w}_2 := C^2\boldsymbol{v} = C\boldsymbol{w}_1$ etc. In our case each C is a linear combination of nested commutators of A and B, hence each product can be reduced to FFTs (multiplying the circulant A by a vector), at the expense of $\mathcal{O}(N\log N)$ operations, and to products of a diagonal matrix B by a vector, at the cost of $\mathcal{O}(N)$ operations. It is easy to observe, though, that some operations are repeated and the task here is to group them together so as to minimize the cost.

   As an example, consider the first few terms of $\mathcal{W}_1^{[1]}$ and $\mathcal{W}_2^{[1]}$. On the face of it, we need 33 matrix-vector products to evaluate

   $$\left(\tau B + \frac{1}{24}\tau^3 c^2[[B,A],A]\right)\boldsymbol{v} \qquad \text{and} \qquad \frac{1}{12}\tau^3 c[[B,A],B]\boldsymbol{v}.$$

   However, let

   $$\boldsymbol{z}_A = A\boldsymbol{v}, \quad \boldsymbol{z}_B = B\boldsymbol{v}, \quad \boldsymbol{z}_{AA} = A\boldsymbol{z}_A, \quad \boldsymbol{z}_{BA} = B\boldsymbol{z}_A, \quad \boldsymbol{z}_{AB} = A\boldsymbol{z}_B$$

   etc. Then

   $$\left(\tau B + \frac{1}{24}\tau^3 c^2[[B,A],A]\right)\boldsymbol{v} = \boldsymbol{z}_B + \frac{\tau^3 c^2}{24}(\boldsymbol{z}_{BAA} - 2\boldsymbol{z}_{ABA} + \boldsymbol{z}_{AAB}),$$
   $$\frac{1}{12}\tau^3 c[[B,A],B]\boldsymbol{v} = -\frac{\tau^3 c}{12}(\boldsymbol{z}_{ABB} - 2\boldsymbol{z}_{BAB} + \boldsymbol{z}_{BBA}),$$

   at the cost of just 12 matrix-vector products. Further economies might be possible by aggregating terms, similarly to the design of optimal Magnus integrators in (Blanes, Casas & Ros 2002).

   Another issue for future exploration is an efficient choice of dimensions for Krylov subspace methods. Such methods approximate a product of a matrix exponential and a vector by projecting it into a finite-dimensional subspace of $\mathbb{C}^{2N+1}$. In the case of exponentials originating in the symmetric Zassenhaus splitting (4.4), different exponentials are multiplied by different powers of the large parameter ω. Given that the arguments are skew-Hermitian, this means that the exponentials rotate at different speeds: essentially, (4.4) separates scales. This has clear implications to the choice of dimension of the finite-dimensional space onto which we should project these exponentials. Clearly, good choice of dimension is ω-dependent but precise dependency is unknown.

3. *Nonlinear Schrödinger and other equations.* The phenomenon of small commutators persists once (1.1) is replaced by

$$\mathrm{i}\hbar u_t = -\frac{\hbar^2}{2m} u_{xx} - V(x,u), \qquad t \geq 0, \quad x \in [-1,1], \tag{5.1}$$

with appropriate initial and periodic boundary conditions and with *potential* $V(\,\cdot\,,u)$ of period 2 – the *nonlinear Schrödinger equation (NLS).* On the face of it, our methodology can be extended to this setting with nonlinear algebraic equations solved in the usual way, by iteration. However, although much remains to be done to understand the analytic structure of NLS, enough is known to impose a raft of additional requirements on the numerical solution (Jin et al. 2011). Such requirements depend on the potential, but also on different possible applications of (5.1). Serious investigation of our methodology in the NLS context must address these highly nontrivial issues.

With greater generality, our approach makes sense for a raft of equations of the form

$$u_t = \mathcal{L}u + \omega f(x,t,u) \qquad \text{or} \qquad u_{tt} = \mathcal{L}u + \omega f(x,t,u),$$

where $\mathcal{L}$ is a spatial linear differential operator and $\omega \gg 1$, e.g. the *nonlinear Klein–Gordon equation*

$$u_{tt} = \nabla^2 u + \omega f(u).$$

However, again, much further analysis is necessary.

Acknowledgements

We wish to thank Alexei Shadrin (Cambridge) for teaching us the requisite approximation theory, Philipp Bader (Valencia) and Hans Munthe-Kaas (Bergen) for presenting us with two completely different, yet equally insightful, proofs that $\sum_{-N}^{N} a_{N,\ell} = 0$ for the spectral collocation method. We are also grateful to Philipp Bader, Peter Markowich (Cambridge) and Jennifer Ryan (Delft) for comments on an early draft of the paper.

References

Blanes, S., Casas, F. & Ros, J. (2002), 'High order optimized geometric integrators for linear differential equations', *BIT* **42**, 262–284.

Fejér, L. (1910), 'Lebesguesche Konstanten und divergente Fourierreihen', *J. f. d. Reine u. Angew. Math.* **138**, 22–53.

Fornberg, B. (1998), *A Practical Guide to Pseudospectral Mathods*, Cambridge University Press, Cambridge.

Galkin, P. V. (1971), 'Estimate for Lebesque constants', *Trudy Mat. Inst. Steklov.* **109**, 3–5.

Griffiths, D. J. (2004), *Introduction to Quantum Mechanics*, 2nd edn, Prentice Hall, Upper Saddle River, NJ.

Hairer, E., Lubich, C. & Wanner, G. (2006), *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, 2nd edn, Springer Verlag, Berlin.

Hochbruck, M. & Lubich, C. (1997), 'On Krylov subspace approximations to the matrix exponential operator', *SIAM J. Numer. Anal.* **34**, 1911–1925.

Iserles, A. (2008), *A First Course in the Numerical Analysis of Differential Equations*, 2nd edn, Cambridge University Press, Cambridge.

Iserles, A., Munthe-Kaas, H. Z., Nørsett, S. P. & Zanna, A. (2000), 'Lie-group methods', *Acta Numerica* **9**, 215–365.

Jin, S., Markowich, P. & Sparber, C. (2011), 'Mathematical and computational methods for semiclassical Schrödinger equations', *Acta Numerica* **20**, 121–210.

Lozinski, S. M. (1948), 'On a class of linear operators', *Doklady Akad. Nauk SSSR* **61**, 19–196.

McLachlan, R. I. & Quispel, G. R. W. (2002), 'Splitting methods', *Acta Numerica* **11**, 341–434.

Murua, A. (2010), 'The symmetric Baker–Campbell–Hausdorff formula up to terms of degree 20 written in a Hall basis or a Lyndon basis', `http://www.ehu.es/ccwmuura/research/sbchHall19.dat`.

Oteo, J. A. (1991), 'The Baker–Campbell–Hausdorff formula and nested commutator identities', *J. Math. Phys.* **32**, 419–424.

Suzuki, M. (1977), 'On the convergence of exponential operators—the Zassenhaus formula, BCH formula and systematic approximations', *Comm. Math. Phys.* **57**, 193–200.

Yošida, H. (1990), 'Construction of higher order symplectic integrators', *Phys. Lett.* **150**, 262–268.

Table 1:

j	Nested commutator	χ_j	girth	grade
1	A	1	0	1
2	B	1	1	1
3	$[B,A]$	0	1	2
4	$[[B,A],A]$	$-\frac{1}{24}$	1	3
5	$[[B,A],B]$	$-\frac{1}{12}$	2	3
6	$[[[B,A],A],A]$	0	1	4
7	$[[[B,A],A],B]$	0	2	4
8	$[[[B,A],B],B]$	0	3	4
9	$[[[[B,A],A],A],A]$	$\frac{7}{5760}$	1	5
10	$[[[[B,A],A],A],B]$	$\frac{7}{1440}$	2	5
11	$[[[[B,A],A],B],B]$	$\frac{1}{180}$	3	5
12	$[[[[B,A],B],B],B]$	$\frac{1}{720}$	4	5
13	$[[[B,A],A],[B,A]]$	$\frac{1}{480}$	2	5
14	$[[[B,A],B],[B,A]]$	$-\frac{1}{360}$	3	5
15	$[[[[[B,A],A],A],A],A]$	0	1	6
16	$[[[[[B,A],A],A],A],B]$	0	2	6
17	$[[[[[B,A],A],A],B],B]$	0	3	6
18	$[[[[[B,A],A],B],B],B]$	0	4	6
19	$[[[[[B,A],B],B],B],B]$	0	5	6
20	$[[[[B,A],A],A],[B,A]]$	0	2	6
21	$[[[[B,A],A],B],[B,A]]$	0	3	6
22	$[[[[B,A],B],B],[B,A]]$	0	4	6
23	$[[[B,A],B],[[B,A],A]]$	0	3	6
24	$[[[[[[B,A],A],A],A],A],A]$	$-\frac{31}{967680}$	1	7
25	$[[[[[[B,A],A],A],A],A],B]$	$-\frac{31}{161280}$	2	7
26	$[[[[[[B,A],A],A],A],B],B]$	$-\frac{13}{30240}$	3	7
27	$[[[[[[B,A],A],A],B],B],B]$	$-\frac{53}{120960}$	4	7
28	$[[[[[[B,A],A],B],B],B],B]$	$-\frac{1}{5040}$	5	7
29	$[[[[[[B,A],B],B],B],B],B]$	$-\frac{1}{30240}$	6	7
30	$[[[[[B,A],A],A],A],[B,A]]$	$-\frac{53}{161280}$	2	7
31	$[[[[[B,A],A],A],B],[B,A]]$	$-\frac{11}{12096}$	3	7
32	$[[[[[B,A],A],B],B],[B,A]]$	$-\frac{3}{4480}$	4	7

A The Hall basis of $\mathcal{F}$

We display in Table 1 the terms in the Hall basis of the free Lie algebra $\mathcal{F}$ up to grade 7, grouped by grade.

Table 1 (contd)

j	Nested commutator	χ_j	girth	grade
33	$[[[[[B,A],B],B],B],[B,A]]$	$-\frac{1}{10080}$	5	7
34	$[[[[B,A],A],[B,A]],[B,A]]$	$-\frac{1}{4032}$	3	7
35	$[[[[B,A],B],[B,A]],[B,A]]$	$-\frac{1}{6720}$	4	7
36	$[[[[B,A],A],A],[[B,A],A]]$	$-\frac{19}{80640}$	2	7
37	$[[[[B,A],A],B],[[B,A],A]]$	$-\frac{1}{10080}$	3	7
38	$[[[[B,A],B],B],[[B,A],A]]$	$\frac{17}{40320}$	4	7
39	$[[[[B,A],A],A],[[B,A],B]]$	$-\frac{53}{60480}$	3	7
40	$[[[[B,A],A],B],[[B,A],B]]$	$-\frac{19}{13440}$	4	7
41	$[[[[B,A],B],B],[[B,A],B]]$	$-\frac{1}{5040}$	5	7

Here χ_j is the coefficient of the jth term in the symmetric BCH expansion (2.2) – note that the coefficients of even-grade terms are nil, consistently with the expansion being an odd function of τ.

B The coefficients $\mathcal{W}_m^{[\ell]}$

We tabulate the coefficients $\mathcal{W}_m^{[\ell]}(\tau, cA, B)$ which are required in the construction of the symmetric Zassenhaus splitting (4.4). All the terms are presented up to an error of $\mathcal{O}(\tau^9)$. Note that, thanks to the symmetric BCH expansion (2.2) being odd in t, only odd powers of t are present in the $\mathcal{W}_m^{[\ell]}$s.

B.1 $\ell = 0$

$$\begin{aligned}\mathcal{W}_0^{[0]} &= \tau c A,\\ \mathcal{W}_1^{[0]} &= \tau B.\end{aligned}$$

B.2 $\ell = 1$

$$\begin{aligned}\mathcal{W}_1^{[1]} = \tau B &+ \frac{1}{24}\tau^3 c^2 [[B,A],A] + \frac{1}{1920}\tau^5 c^4 [[[[B,A],A],A],A]\\ &+ \frac{1}{322560}\tau^7 c^6 [[[[[[B,A],A],A],A],A],A] + \mathcal{O}(\tau^9),\end{aligned}$$

$$
\begin{aligned}
\mathcal{W}_2^{[1]} = {} & \frac{1}{12}\tau^3 c[[B,A],B] - \tau^5 c^3 \left(\frac{1}{1440}[[[[B,A],A],A],B] + \frac{1}{1440}[[[[B,A],A],B],A] \right. \\
& \left. + \frac{1}{720}[[[[B,A],B],A],A] \right) + \tau^7 c^5 \left(\frac{17}{483840}[[[[[[B,A],A],A],A],A],B] \right. \\
& - \frac{19}{120960}[[[[[[B,A],A],A],A],B],A] + \frac{11}{40320}[[[[[[B,A],A],A],B,A],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],A],A],A] + \frac{1}{30240}[[[[[[B,A],B],A],A],A],A] \\
& \left. - \frac{1}{96768}[[[[[B,A],A],A],A],[B,A]] + \frac{13}{241920}[[[[B,A],A],A],[[B,A],A]] \right) \\
& + \mathcal{O}(\tau^9)\,,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}_3^{[1]} = {} & \tau^5 c^2 \left(\frac{1}{240}[[[[B,A],A],B],B] - \frac{1}{720}[[[[B,A],A],A],B] - \frac{1}{720}[[[[B,A],B],B],A] \right. \\
& \left. - \frac{1}{360}[[[B,A],B],[B,A]] \right) + \tau^7 c^4 \left(-\frac{19}{120960}[[[[[[B,A],A],A],A],B],B] \right. \\
& + \frac{11}{40320}[[[[[[B,A],A],A],B],A],B] + \frac{11}{40320}[[[[[[B,A],A],A],B],B],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],A],A],B] - \frac{1}{6048}[[[[[[B,A],A],B],A],B],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],B],A],A] + \frac{1}{30240}[[[[[[B,A],B],A],A],A],B] \\
& + \frac{1}{30240}[[[[[[B,A],B],A],A],B],A] + \frac{1}{30240}[[[[[[B,A],B],A],B],A],A] \\
& - \frac{41}{120960}[[[[[B,A],A],A],B],[B,A]] + \frac{23}{40320}[[[[[B,A],A],A],A],[B,A]] \\
& - \frac{1}{10080}[[[[[B,A],A],A],A],[B,A]] + \frac{1}{6720}[[[[B,A],A],[B,A]],[B,A]] \\
& + \frac{1}{1440}[[[[B,A],A],B],[[B,A],A]] - \frac{5}{8064}[[[[B,A],A],A],[[B,A],A]] \\
& \left. + \frac{41}{120960}[[[[B,A],A],A],[[B,A],A]] \right) + \mathcal{O}(\tau^9)\,,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}_4^{[1]} = &-\frac{1}{720}\tau^5 c[[[[B,A],B],B],B] + \tau^7 c^3 \left(\frac{11}{40320}[[[[[[B,A],A],A],B],B],B]\right.\\
&-\frac{1}{6048}[[[[[[B,A],A],B],A],B],B] - \frac{1}{6048}[[[[[[B,A],A],B],B],A],A]\\
&-\frac{1}{6048}[[[[[[B,A],A],B],B],B],A] + \frac{1}{30240}[[[[[[B,A],B],A],A],B],A]\\
&+\frac{1}{30240}[[[[[[B,A],B],A],B],A],B] + \frac{1}{30240}[[[[[[B,A],B],A],B],B],A]\\
&+\frac{1}{30240}[[[[[[B,A],B],B],A],A],B] + \frac{1}{30240}[[[[[[B,A],B],B],A],B],A]\\
&+\frac{1}{30240}[[[[[[B,A],B],B],B],A],A] + \frac{23}{40320}[[[[[B,A],A],B],B],[B,A]]\\
&-\frac{1}{10080}[[[[[B,A],B],A],B],[B,A]] - \frac{1}{10080}[[[[[B,A],B],B],A],[B,A]]\\
&+\frac{1}{6720}[[[[B,A],B],[B,A]],[B,A]] - \frac{17}{40320}[[[[B,A],B],B],[[B,A],A]]\\
&\left.+\frac{7}{5760}[[[[B,A],A],B],[[B,A],B]] - \frac{1}{5040}[[[[B,A],B],A],[[B,A],B]]\right) + \mathcal{O}(\tau^9)\,,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}_5^{[1]} = \tau^7 c^2 &\left(-\frac{1}{6048}[[[[[[B,A],A],B],B],B],B] + \frac{1}{30240}[[[[[[B,A],B],A],B],B],B]\right.\\
&+\frac{1}{30240}[[[[[[B,A],B],B],A],B],B] + \frac{1}{30240}[[[[[[B,A],B],B],B],A],B]\\
&+\frac{1}{30240}[[[[[[B,A],B],B],B],B],A] - \frac{1}{10080}[[[[[B,A],B],B],B],[B,A]]\\
&\left.-\frac{1}{5040}[[[[B,A],B],B],[[B,A],B]]\right) + \mathcal{O}(\tau^9)\,,
\end{aligned}
$$

$$
\mathcal{W}_6^{[1]} = \frac{1}{30240}\tau^7 c[[[[[[B,A],B],B],B],B],B] + \mathcal{O}(\tau^9)\,,
$$

$$
\mathcal{W}_m^{[1]} = \mathcal{O}(\tau^9)\,, \qquad m \ge 7.
$$

B.3 $\ell = 2$

$$
\begin{aligned}
\mathcal{W}_2^{[2]} = {} & \frac{1}{2}\tau^3 c[[B,A],B] + \tau^5 c^3 \Big(-\frac{1}{1440}[[[[B,A],A],A],B] - \frac{1}{1440}[[[[B,A],A],B],A] \\
& - \frac{1}{720}[[[[B,A],B],A],A] \Big) + \tau^7 c^5 \Big(\frac{17}{483840}[[[[[[B,A],A],A],A],A],B] \\
& - \frac{19}{120960}[[[[[[B,A],A],A],A],B],A] + \frac{11}{40320}[[[[[[B,A],A],A],B],A],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],A],A],A] + \frac{1}{30240}[[[[[[B,A],B],A],A],A],A] \\
& - \frac{1}{96768}[[[[[B,A],A],A],A],[B,A]] + \frac{13}{241920}[[[[B,A],A],A],[[B,A],A]] \Big) + \mathcal{O}(\tau^9)\,;
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}_3^{[2]} = {} & \tau^5 c^2 \Big(\frac{1}{240}[[[[B,A],A],B],B] - \frac{1}{720}[[[[B,A],A],A],B] - \frac{1}{720}[[[[B,A],B],B],A] \\
& - \frac{1}{360}[[[B,A],B],[B,A]] \Big) + \tau^7 c^4 \Big(-\frac{19}{120960}[[[[[[B,A],A],A],A],B],B] \\
& + \frac{11}{40320}[[[[[[B,A],A],A],B],A],B] + \frac{11}{40320}[[[[[[B,A],A],A],B],B],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],A],A],B] - \frac{1}{6048}[[[[[[B,A],A],B],A],B],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],B],A],A] + \frac{1}{30240}[[[[[[B,A],B],A],A],A],B] \\
& + \frac{1}{30240}[[[[[[B,A],B],A],A],B],A] + \frac{1}{30240}[[[[[[B,A],B],A],B],A],A] \\
& - \frac{41}{120960}[[[[[B,A],A],A],B],[B,A]] + \frac{23}{40320}[[[[[B,A],A],A],A],[B,A]] \\
& - \frac{1}{10080}[[[[[B,A],A],A],A],[B,A]] + \frac{1}{6720}[[[[B,A],A],[B,A]],[B,A]] \\
& + \frac{1}{1440}[[[[B,A],A],B],[[B,A],A]] - \frac{5}{8064}[[[[B,A],A],A],[[B,A],A]] \\
& + \frac{41}{120960}[[[[B,A],A],A],[[B,A],A]] \Big) + \mathcal{O}(\tau^9)\,,
\end{aligned}
$$

$$\begin{aligned}\mathcal{W}_4^{[2]} = {}& \frac{1}{480}\tau^5 c[[[[B,A],B],B],B] + \tau^7 c^3 \left(\frac{59}{241920}[[[[[B,A],A],A],B],B],B] \right.\\
&- \frac{47}{241920}[[[[[B,A],A],B],A],B],B] - \frac{1}{6048}[[[[[B,A],A]B],B],A],B]\\
&- \frac{1}{6048}[[[[[B,A],A],B],B],B],A] - \frac{1}{40320}[[[[[B,A],B],A],A],B],B]\\
&+ \frac{1}{30240}[[[[[B,A],B],A],B],A],B] + \frac{1}{30240}[[[[[B,A],B],B],A],B],A]\\
&+ \frac{1}{30240}[[[[[B,A],B],B],A],A],B] + \frac{1}{30240}[[[[[B,A],B],B],A],B],A]\\
&+ \frac{1}{30240}[[[[[B,A],B],B],B],A],A] + \frac{23}{40320}[[[[B,A],A],B],B],[B,A]]\\
&- \frac{1}{10080}[[[[B,A],B],A],B],[B,A]] - \frac{1}{10080}[[[[B,A],B],B],A],[B,A]]\\
&+ \frac{1}{6720}[[[B,A],B],[B,A]],[B,A]] - \frac{67}{241920}[[[B,A],B],B],[[B,A],A]]\\
&\left. - \frac{13}{5760}[[[B,A],A],B],[[B,A],B]] - \frac{1}{5040}[[[B,A],B],A],[[B,A],B]] \right) + \mathcal{O}(\tau^9)\,;\end{aligned}$$

$$\begin{aligned}\mathcal{W}_5^{[2]} = {}& \tau^7 c^2 \left(\frac{1}{120960}[[[[[B,A],A],B],B],B],B] - \frac{1}{40320}[[[[[B,A],B],A],B],B],B] \right.\\
&- \frac{1}{40320}[[[[[B,A],B],B],A],B],B] + \frac{1}{30240}[[[[[B,A],B],B],B],A],B]\\
&+ \frac{1}{30240}[[[[[B,A],B],B],B],B],A] - \frac{1}{10080}[[[[B,A],B],B],B],[B,A]]\\
&\left. + \frac{23}{60480}[[[B,A],B],B],[[B,A],B]] - \frac{1}{8640}[[[[B,A],B],[B,A]],B],B] \right) + \mathcal{O}(\tau^9)\,;\end{aligned}$$

$$\mathcal{W}_6^{[2]} = \frac{1}{53760}\tau^7 c[[[[[B,A],B],B],B],B],B] + \mathcal{O}(\tau^9)\,,$$

$$\mathcal{W}_m^{[2]} = \mathcal{O}(\tau^9)\,, \qquad m \ge 7.$$

B.4 $\ell = 3$

$$
\begin{aligned}
\mathcal{W}_3^{[3]} = \tau^5 c^2 \Bigg(& \frac{1}{240}[[[[B,A],A],B],B] - \frac{1}{720}[[[[B,A],B],A],B] - \frac{1}{720}[[[[B,A],B],B],A] \\
& - \frac{1}{360}[[[[B,A],B]],[B,A]] \Bigg) + \tau^7 c^4 \Bigg(-\frac{19}{120960}[[[[[[B,A],A],A],A],B],B] \\
& + \frac{11}{40320}[[[[[[B,A],A],A],B],A],B] + \frac{11}{40320}[[[[[[B,A],A],A],B],B],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],A],A],B] - \frac{1}{6048}[[[[[[B,A],A],B],A],B],A] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],B],A],A] + \frac{1}{30240}[[[[[[B,A],B],A],A],A],B] \\
& + \frac{1}{30240}[[[[[[B,A],B],A],A],B],A] + \frac{1}{30240}[[[[[[B,A],A],A],B],A],A] \\
& + \frac{1}{30240}[[[[[[B,A],B],B],A],A],A] - \frac{41}{120960}[[[[[B,A],A],A],B],[B,A]] \\
& + \frac{23}{40320}[[[[[B,A],A],B],A],[B,A]] - \frac{1}{10080}[[[[[B,A],B],A],A],[B,A]] \\
& + \frac{1}{6720}[[[[B,A],A],[B,A]],[B,A]] + \frac{1}{1440}[[[[B,A],A],B],[[B,A],A]] \\
& - \frac{5}{8064}[[[[B,A],B],A],[[B,A],A]] + \frac{41}{120960}[[[[B,A],A],A],[[B,A],B]] \Bigg) \\
& + \mathcal{O}(\tau^9)\,,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}_4^{[3]} = \frac{\tau^5 c}{480}[[[[B,A],B],B],B] + \tau^7 c^3 \Bigg(& \frac{59}{241920}[[[[[[B,A],A],A],B],B],B] \\
& - \frac{47}{241920}[[[[[[B,A],A],B],A],B],B] - \frac{1}{6048}[[[[[[B,A],A],B],B],A],B] \\
& - \frac{1}{6048}[[[[[[B,A],A],B],B],B],A] - \frac{1}{40320}[[[[[[B,A],B],A],A],B],B] \\
& + \frac{1}{30240}[[[[[[B,A],B],A],B],A],B] + \frac{1}{30240}[[[[[[B,A],B],A],B],B],A] \\
& + \frac{1}{30240}[[[[[[B,A],B],B],A],A],B] + \frac{1}{30240}[[[[[[B,A],B],B],A],B],A] \\
& + \frac{1}{30240}[[[[[[B,A],B],B],B],A],A] + \frac{23}{40320}[[[[[B,A],A],B],B],[B,A]] \\
& - \frac{1}{10080}[[[[[B,A],B],A],B],[B,A]] - \frac{1}{10080}[[[[[B,A],B],B],A],[B,A]] \\
& + \frac{1}{6720}[[[[B,A],B],[B,A]],[B,A]] - \frac{67}{241920}[[[[B,A],B],B],[[B,A],A]] \\
& - \frac{13}{5760}[[[[B,A],A],B],[[B,A],B]] - \frac{1}{5040}[[[[B,A],B],A],[[B,A],B]] \Bigg) \\
& + \mathcal{O}(\tau^9)\,,
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{W}_5^{[3]} = \tau^7 c^2 \biggl(& \frac{1}{120960}[[[[[[B,A],A],B],B],B],B] - \frac{1}{40320}[[[[[B,A],B],A],B],B],B] \\
& - \frac{1}{40320}[[[[[B,A],B],B],A],B],B] + \frac{1}{30240}[[[[[B,A],B],B],B],A],B] \\
& + \frac{1}{30240}[[[[[B,A],B],B],B],B],A] - \frac{13}{60480}[[[[B,A],B],B],B],[B,A]] \\
& + \frac{1}{3780}[[[B,A],B],B],[[B,A],B]] \biggr) + \mathcal{O}\bigl(\tau^9\bigr)\,.
\end{aligned}
$$

$$
\mathcal{W}_6^{[3]} = \frac{1}{53760}\tau^7 c[[[[[B,A],B],B],B],B],B] + \mathcal{O}\bigl(\tau^9\bigr)\,,
$$

$$
\mathcal{W}_m^{[3]} = \mathcal{O}\bigl(\tau^9\bigr)\,, \qquad m \geq 7.
$$

B.5 $\ell \geq 4$

We now have

$$
\begin{aligned}
& \mathcal{W}_m^{[4]} = \mathcal{W}_m^{[3]} + \mathcal{O}\bigl(\tau^9\bigr)\,, \quad m = 4,5,6, \qquad \mathcal{W}_m^{[4]} = \mathcal{O}\bigl(\tau^9\bigr)\,, \quad m \geq 7, \\
& \mathcal{W}_m^{[5]} = \mathcal{W}_m^{[4]} + \mathcal{O}\bigl(\tau^9\bigr)\,, \quad m = 5,6, \qquad \mathcal{W}_m^{[5]} = \mathcal{O}\bigl(\tau^9\bigr)\,, \quad m \geq 7, \\
& \mathcal{W}_6^{[6]} = \mathcal{W}_6^{[5]} + \mathcal{O}\bigl(\tau^9\bigr)\,, \qquad \mathcal{W}_m^{[6]} = \mathcal{O}\bigl(\tau^9\bigr)\,, \quad m \geq 7 \\
& \mathcal{W}_m^{[\ell]} = \mathcal{O}\bigl(\tau^9\bigr)\,, \qquad \ell, m \geq 7.
\end{aligned}
$$