Acta Numerica (2024), pp. 1–161 Printed in the United Kingdom
doi:10.1017/S0962492923000077

Splitting methods for differential equations

Sergio Blanes
Instituto de Matemática Multidisciplinar,
Universitat Politècnica de València, 46022-Valencia, Spain
E-mail: serblaza@imm.upv.es

Fernando Casas
Departament de Matemàtiques and IMAC, Universitat Jaume I,
12071-Castellón, Spain
E-mail: Fernando.Casas@uji.es

Ander Murua
Konputazio Zientziak eta A.A. Saila, Informatika Fakultatea, EHU/UPV,
Donostia/San Sebastián, Spain
E-mail: Ander.Murua@ehu.es

This overview is devoted to splitting methods, a class of numerical integrators intended for differential equations that can be subdivided into different problems easier to solve than the original system. Closely connected with this class of integrators are composition methods, in which one or several low-order schemes are composed to construct higher-order numerical approximations to the exact solution. We analyse in detail the order conditions that have to be satisfied by these classes of methods to achieve a given order, and provide some insight about their qualitative properties in connection with geometric numerical integration and the treatment of highly oscillatory problems. Since splitting methods have received considerable attention in the realm of partial differential equations, we also cover this subject in the present survey, with special attention to parabolic equations and their problems. An exhaustive list of methods of different orders is collected and tested on simple examples. Finally, some applications of splitting methods in different areas, ranging from celestial mechanics to statistics, are also provided.

2020 Mathematics Subject Classification: Primary 65L05, 65L20, 65P10
Secondary 65M12, 65M70

CONTENTS

1. Introduction

1.1. Lie–Trotter and Strang methods

If, as has sometimes been argued, there are only ten big ideas in numerical analysis and all the rest are merely variations on those themes, *splitting* is undoubtedly one of them (MacNamara and Strang 2016). In fact, we could say that, at least since Descartes stated his four rules of logic in the *Discours de la méthode*,[1] the notion of subdividing a complicated problem into its simpler constituent parts, solving each one of them separately and combining those separated solutions in a controlled way to get a solution to the original overall problem, constitutes a guiding principle in all areas of science and philosophy.

In the realm of numerical analysis of differential equations, this basic principle can be stated as follows. Suppose we have the abstract initial value problem

$$x' \equiv \frac{\mathrm{d}x}{\mathrm{d}t} = f(x), \quad x(0) = x_0 \tag{1.1}$$

associated to an ordinary differential equation (ODE) or a partial differential equation (PDE), in which case f is a certain spatial partial differential operator. Furthermore, suppose that f does not depend explicitly on time and can be decomposed as

$$f = f_1 + \cdots + f_m, \quad m \geq 2, \tag{1.2}$$

so that each initial value problem $x' = f_j(x)$, $x(0) = x_0$ is easier to solve than (1.1).

[1] 'Le second, de diviser chacune des difficultés que j'examinarais, en autant de parcelles qu'il se pourrait, et qu'il serait requis pour les mieux résoudre. Le troisième, de conduire par ordre mes pensées, en commençant par les objets les plus simples et les plus aisés à connaître, pour monter peu à peu, comme par degrés, jusques à la connaissance des plus composés; et supposant même de l'ordre entre ceux qui ne se précèdent point naturellement les uns les autres.' (René Descartes, *Discours de la méthode*, Seconde partie).

Most commonly, they can be integrated exactly in closed form. Then it is possible to take advantage of the decomposition (1.2) to get accurate approximations to the solution of (1.1) by means of *splitting methods*.

The best example to start with is perhaps the case of a linear differential equation defined in $\mathbb{R}^D$ and $m = 2$, namely

$$x' = f_1(x) + f_2(x) = F_1\, x + F_2\, x, \quad x(0) = x_0, \tag{1.3}$$

where F_1 and F_2 are $D \times D$ matrices and $x \in \mathbb{R}^D$. The solution reads

$$x(t) = \mathrm{e}^{t(F_1+F_2)} x_0,$$

so that by computing this matrix exponential directly we would have solved (1.3) without requiring splitting methods. Associated with (1.3), we have the matrix differential equation

$$\frac{\mathrm{d}X}{\mathrm{d}t} = (F_1 + F_2)X, \quad X(0) = I, \tag{1.4}$$

in the sense that

$$x(t) = X(t)\, x_0 \quad \text{and} \quad X(t) = \mathrm{e}^{t(F_1+F_2)}.$$

This is useful (mainly) for theoretical purposes, since we usually try to compute $\mathrm{e}^{t(F_1+F_2)} x_0$ directly instead of first computing the matrix exponential and then multiplying it by x_0.

It often happens that evaluating the action of $X(t)$ on x_0 is difficult or computationally expensive. If, however, this is not the case for each e^{tF_j} separately, then one may use the well-known Lie product formula (Reed and Simon 1980, p. 295)

$$\mathrm{e}^{t(F_1+F_2)} = \lim_{n\to\infty} \left(\mathrm{e}^{\frac{t}{n}F_2}\, \mathrm{e}^{\frac{t}{n}F_1}\right)^n. \tag{1.5}$$

To get an approximate solution of (1.3) at the final time $t = t_f$, we subdivide the interval $[0, t_f]$ into N steps of length h, with $Nh = t_f$, and compute the sequence

$$x_{n+1} = \mathrm{e}^{hF_2}\, \mathrm{e}^{hF_1}\, x_n, \quad n \geq 0, \tag{1.6}$$

so that $x_{n+1} \approx x(t_{n+1} = (n+1)h)$. This is the so-called *Lie–Trotter scheme*. When the matrices commute, the sequence produces the exact solution. To put it another way, if the commutator $[F_1, F_2] \equiv F_1F_2 - F_2F_1 = 0$, then $\exp(h(F_1 + F_2)) = \exp(hF_2)\exp(hF_1)$. Otherwise, a direct calculation shows that

$$\mathrm{e}^{h(F_1+F_2)} - \mathrm{e}^{hF_2}\mathrm{e}^{hF_1} = \frac{1}{2}h^2[F_1, F_2] + O(h^3)$$

as $h \to 0$, and hence the previous approximation is only of first order of accuracy. Another version of the method is possible, of course, by reversing the order of F_1 and F_2, namely

$$x_{n+1} = \mathrm{e}^{hF_1}\, \mathrm{e}^{hF_2}\, x_n, \quad n \geq 0 \tag{1.7}$$

has the same order of accuracy and properties as (1.6). Needless to say, for any

$m > 2$, it results in

$$x_{n+1} = \mathrm{e}^{hF_m} \cdots \mathrm{e}^{hF_2}\, \mathrm{e}^{hF_1}\, x_n$$

(or any other permutation of the matrices F_j).

A higher-order approximation can be achieved by considering a symmetrized version of (1.6),

$$\mathrm{e}^{h(F_1+F_2)} - \mathrm{e}^{\frac{1}{2}hF_1}\mathrm{e}^{hF_2}\mathrm{e}^{\frac{1}{2}hF_1} = Ch^3 + O(h^4), \tag{1.8}$$

where the constant C can be obtained either by comparing Taylor series or by applying the Baker–Campbell–Hausdorff (BCH) formula as (Varadarajan 1984) $C = \frac{1}{24}([F_1, [F_1, F_2]] + 2[F_2, [F_1, F_2]])$. Therefore the sequence

$$x_{n+1} = \mathrm{e}^{\frac{1}{2}hF_1}\mathrm{e}^{hF_2}\mathrm{e}^{\frac{1}{2}hF_1}\, x_n, \quad n \geq 0 \tag{1.9}$$

produces a second-order approximation for the solution of (1.3). This corresponds to the *Strang* (splitting) *scheme*. Again, if the role of F_1 and F_2 is interchanged, we have another version of the Strang splitting scheme.

Simple generalizations to the case $m > 2$ include in particular the product

$$\mathrm{e}^{\frac{1}{2}hF_1}\, \mathrm{e}^{\frac{1}{2}hF_2} \cdots \mathrm{e}^{hF_m} \cdots \mathrm{e}^{\frac{1}{2}hF_2}\, \mathrm{e}^{\frac{1}{2}hF_1}.$$

Higher-order splitting methods could in principle be constructed by including more exponentials with their corresponding coefficients in a time step, namely

$$\Psi(h) = \mathrm{e}^{a_{s+1}hF_1}\, \mathrm{e}^{b_s hF_2}\, \mathrm{e}^{a_s hF_1} \cdots \mathrm{e}^{b_1 hF_2}\, \mathrm{e}^{a_1 hF_1}. \tag{1.10}$$

The number s as well as the coefficients a_j, b_j are chosen so that

$$\Psi(h) = \mathrm{e}^{h(F_1+F_2)} + O(h^{r+1})$$

as $h \to 0$ for a given order r. In (1.10), the first and last exponentials correspond to F_1. This format is convenient for implementation, since the last exponential in one step can be concatenated with the first one at the next step, thus reducing the number of evaluations by one. This corresponds to the well-known FSAL (first same as last) property. The situation when we have an exponential of F_2 as the first and last term is recovered by taking $a_1 = a_{s+1} = 0$.

1.2. Flows and differential operators

The Lie–Trotter scheme can be easily generalized to any system (1.1)–(1.2) when the solution is no longer given by exponentials, as in the linear case. If $m = 2$, it is equivalent to the following.

Algorithm 1.1 (Lie–Trotter). Starting from $x_0 = x(0)$, for $n \geq 0$,

- solve $y_1' = f_1(y_1)$, $y_1(t_n) = x_n$, in $[t_n, t_{n+1}]$;
- set $y_{n+1/2} = y_1(t_{n+1})$;
- solve $y_2' = f_2(y_2)$, $y_2(t_n) = y_{n+1/2}$, in $[t_n, t_{n+1}]$;

- finally, set $x_{n+1} = y_2(t_{n+1})$.

Alternatively, if we denote the solution of equation (1.1) for each $t \in \mathbb{R}$ as

$$x(t) = \varphi_t^{[f]}(x_0),$$

then Algorithm 1.1 can be formally expressed as

$$x_{n+1} = \chi_h(x_n) \equiv \varphi_h^{[2]}\left(\varphi_h^{[1]}(x_n)\right) = \left(\varphi_h^{[2]} \circ \varphi_h^{[1]}\right)(x_n), \tag{1.11}$$

where we have used the simplified notation $\varphi_t^{[j]}(x_0)$ for the solutions $y(t) = \varphi_t^{[f_j]}(x_0)$ of the subproblems $y' = f_j(y)$, $y(0) = x_0$. Analogously, the Strang splitting (1.9) is generalized as follows.

Algorithm 1.2 (Strang). From $x_0 = x(0)$, for $n \geq 0$,

- solve $y_1' = f_1(y_1)$, $y_1(t_n) = x_n$, in $[t_n, t_{n+1/2}]$, with $t_{n+1/2} = (n + \frac{1}{2})h$;
- set $y_{n+1/2} = y_1(t_{n+1/2})$;
- solve $y_2' = f_2(y_2)$, $y_2(t_n) = y_{n+1/2}$, in $[t_n, t_{n+1}]$;
- set $\hat{y}_{n+1/2} = y_2(t_{n+1})$;
- solve $y_1' = f_1(y_1)$, $y_1(t_{n+1/2}) = \hat{y}_{n+1/2}$, in $[t_{n+1/2}, t_{n+1}]$;
- finally, set $x_{n+1} = y_1(t_{n+1})$

or, in short,

$$x_{n+1} = S_h^{[2]}(x_n) \equiv \left(\varphi_{h/2}^{[1]} \circ \varphi_h^{[2]} \circ \varphi_{h/2}^{[1]}\right)(x_n). \tag{1.12}$$

If equation (1.1) corresponds to a (nonlinear) ordinary differential equation evolving in $\mathbb{R}^D$,

$$x' = f(x), \quad x(0) = x_0 \in \mathbb{R}^D, \tag{1.13}$$

f is called the *vector field*. If (1.13) admits for each $x_0 \in \mathbb{R}^D$ a unique solution $x(t)$ defined for all $t \in \mathbb{R}$, the map

$$\begin{aligned} \varphi_t^{[f]}: \quad & \mathbb{R}^D \longrightarrow \mathbb{R}^D \\ & x_0 \longmapsto x(t) = \varphi_t^{[f]}(x_0) \end{aligned}$$

is referred to as the *t-flow* (Arnold 1989). Thus, for each value of the real parameter t, $\varphi_t^{[f]}$ maps $\mathbb{R}^D$ in $\mathbb{R}^D$ in such a way that $\varphi_t^{[f]}(z)$ is the value at time t of the solution of the system with initial value z at time 0, whereas, for fixed x_0 and varying t, $\varphi_t^{[f]}(x_0)$ is the solution of the initial value problem (1.13).

It is worth mentioning that the solution $x(t)$ of (1.1) is in general defined for a *maximal time interval* $(t_{\min}(x_0), t_{\max}(x_0))$ (with $-\infty \leq t_{\min}(x_0) \leq 0 \leq t_{\max}(x_0) \leq +\infty$). Furthermore, the vector field f of many systems of ordinary differential equations is singular (or undefined) for some $x \in \mathbb{R}^D$. Thus, in general, f is defined for some open set $\mathcal{U} \subset \mathbb{R}^D$. Hence, for a given $t \in \mathbb{R}$, the t-flow $\varphi_t^{[f]}$ is a

map from

$$\mathcal{D}_t = \{x_0 \in \mathcal{U} : t \in (t_{\min}(x_0), t_{\max}(x_0))\}$$

to $\mathcal{U}$. In general, one may have different domains of definition $\mathcal{U}$ (resp. $\mathcal{D}_t$) for each f_j (resp. $\varphi_t^{[f_j]}$). In this general situation the compositions in (1.11) and (1.12) are not well-defined for all x_0, h and n. In order to avoid these technicalities, we will assume in what follows that $\mathcal{U} = \mathbb{R}^D$, $t_{\min} = -\infty$ and $t_{\max} = +\infty$, for each vector field f_j.

Associated with the vector field f is the Lie derivative or *Lie operator* F (Arnold 1989), mapping smooth functions $g : \mathbb{R}^D \to \mathbb{R}$ into the real-valued function $F\,g : \mathbb{R}^D \to \mathbb{R}$ such that, for $x \in \mathbb{R}^D$,

$$(F\,g)(x) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} g\big(\varphi_t^{[f]}(x)\big),$$

that is,

$$(F\,g)(x) = f(x) \cdot \nabla g(x).$$

Then the flow of (1.13) verifies (Sanz-Serna and Calvo 1994, Hairer, Lubich and Wanner 2006)

$$g\big(\varphi_t^{[f]}(x)\big) = \big(\mathrm{e}^{tF} g\big)(x) \equiv \left(\sum_{k\geq 0} \frac{t^k}{k!} F^k g\right)(x).$$

The operator $X(t) \equiv \mathrm{e}^{tF}$ is called *Lie transformation*, and can be seen as the formal solution of the operator equation

$$\frac{\mathrm{d}X}{\mathrm{d}t} = X\,F, \quad X(0) = I. \tag{1.14}$$

This can be seen as follows: on the one hand,

$$\frac{\mathrm{d}}{\mathrm{d}t} g\big(\varphi_t^{[f]}(x)\big) = \frac{\mathrm{d}}{\mathrm{d}t}(X(t)g)(x) = \left(\frac{\mathrm{d}X(t)}{\mathrm{d}t} g\right)(x),$$

and on the other hand

$$\frac{\mathrm{d}}{\mathrm{d}t} g\big(\varphi_t^{[f]}(x)\big) = (Fg)\big(\varphi_t^{[f]}(x)\big) = X(t)(Fg)(x).$$

Lie operators satisfy some remarkable properties (Arnold 1989). In particular, although they do not commute, their commutator is nevertheless a first-order linear differential operator. Specifically, let F and G be the Lie operators associated with f and g, respectively, and $u : \mathbb{R}^D \to \mathbb{R}$ a given smooth function. Then

$$[F, G]u = (FG - GF)\,u = \sum_{i,j=1}^{D} \left(f_j \frac{\partial g_i}{\partial x_j} - g_j \frac{\partial f_i}{\partial x_j}\right) \frac{\partial u}{\partial x_i}$$

and it is possible to associate a new vector field to this differential operator, $w = (f, g)$, with components

$$w_i = (f, g)_i = \sum_{j=1}^{D} \left(f_j \frac{\partial g_i}{\partial x_j} - g_j \frac{\partial f_i}{\partial x_j} \right).$$

It is called the *Lie–Poisson bracket* of f and g, and its Lie operator W satisfies $W = [F, G]$.

Now suppose that $f(x) = f_1(x) + f_2(x)$, so that each part $x' = f_j(x)$ is exactly solvable (or can be numerically solved up to round-off accuracy) with flow $x(t) = \varphi_t^{[j]}(x_0)$. Letting F_1 and F_2 denote the Lie operators associated with f_1 and f_2, respectively, it holds that

$$g\left(\varphi_t^{[1]}(x)\right) = \left(\mathrm{e}^{tF_1} g\right)(x), \quad g\left(\varphi_t^{[2]}(x)\right) = \left(\mathrm{e}^{tF_2} g\right)(x).$$

Then, for the first-order approximation $\chi_h = \varphi_h^{[2]} \circ \varphi_h^{[1]}$ furnished by Algorithm 1.1, we have $g(\chi_h(x)) = (\Psi(h)g)(x)$, where $\Psi(h)$ is a series of linear differential operators defined as

$$\Psi(h) = \mathrm{e}^{hF_1}\, \mathrm{e}^{hF_2}. \tag{1.15}$$

Notice that the exponentials of Lie derivatives in (1.15) appear in reverse order with respect to the maps in the integrator (Hairer *et al.* 2006, p. 88). Of course, the same procedure can be applied to the Strang splitting, resulting in the product

$$\Psi(h) = \mathrm{e}^{\frac{h}{2}F_1}\, \mathrm{e}^{hF_2}\, \mathrm{e}^{\frac{h}{2}F_1}. \tag{1.16}$$

These considerations show that: (i) splitting methods for the problem (1.1)–(1.2) defined in a certain function space (e.g. with partial differential equations) can be formulated in terms of the solution of each subproblem (either exact or approximate) by means of Algorithms 1.1 and 1.2, and (ii) splitting methods applied to nonlinear ODEs evolving in $\mathbb{R}^D$ can also be formally expressed as products of exponentials of differential operators, since it is possible to transform the original nonlinear problem into a linear one with the Lie formalism. This observation is very useful when analysing the order conditions for a method to be of a given order. In particular, we have the same order conditions for linear and nonlinear ODEs (see Section 2).

The previous integrators are sometimes called *multiplicative operator-splitting methods*, especially in the literature concerning the numerical treatment of partial differential equations. In that area, we still have to specify how to solve each initial value sub-problem in Algorithms 1.1–1.2 as well as the boundary conditions. Moreover, we should take into account that, for a given differential equation, different ways to carry out the splitting in fact lead to different integrators.

1.3. Adjoint method, conjugate method

The flow $\varphi_t^{[f]}$ of (1.13) verifies $\left(\varphi_{-t}^{[f]}\right)^{-1} = \varphi_t^{[f]}$, but this property is not shared by many numerical integrators, and in particular by the map χ_h corresponding to the Lie–Trotter scheme.

In general, if $\psi_h(x)$ represents a numerical method of order at least one, i.e. $\psi_h(x) = x + h\,f(x) + O(h^2)$, then $(\psi_{-h})^{-1}(x) = x + h\,f(x) + O(h^2)$, so that

$$\psi_h^* \equiv (\psi_{-h})^{-1}$$

is also a numerical method of order at least one. It is called the *adjoint method* of ψ_h (Sanz-Serna and Calvo 1994). In words, stepping forwards with the given method ψ_h is the same as stepping backwards with the inverse of its adjoint ψ_h^*. If $\psi_h \equiv \chi_h = \varphi_h^{[2]} \circ \varphi_h^{[1]}$, then clearly $\chi_h^* = \varphi_h^{[1]} \circ \varphi_h^{[2]}$. Additional examples are the explicit and implicit Euler methods

$$x_{n+1} = \psi_h^e(x_n) = x_n + hf(x_n), \quad x_{n+1} = \psi_h^i(x_n) = x_n + hf(x_{n+1}),$$

since $\psi_h^i = (\psi_h^e)^*$.

Whenever an integrator satisfies

$$\psi_h = \psi_h^* = (\psi_{-h})^{-1},$$

it is called a *time-symmetric* or *self-adjoint* method. Alternatively, $x_{n+1} = \psi_h(x_n)$ is time-symmetric if and only if, exchanging $h \leftrightarrow -h$ and $x_n \leftrightarrow x_{n+1}$, we get the same expression, i.e. $\psi_{-h}(x_{n+1}) = x_n$. The Strang scheme (1.12) is an example of a time-symmetric method.

It is in fact straightforward to construct time-symmetric methods using the adjoint: given an arbitrary method ψ_h of order $r \geq 1$, then the compositions

$$\psi_{h/2} \circ \psi_{h/2}^* \quad \text{and} \quad \psi_{h/2}^* \circ \psi_{h/2} \tag{1.17}$$

are time-symmetric methods of order $r \geq 2$ (Sanz-Serna and Calvo 1994). Further, symmetric methods are necessarily of even order, as we will show in Section 2. Notice that the Strang method (1.12) is simply

$$S_h^{[2]} = \chi_{h/2}^* \circ \chi_{h/2},$$

where χ_h is given by (1.11). Additional examples are the trapezoidal rule $\psi_h^t = \psi_{h/2}^i \circ \psi_{h/2}^e$ and the midpoint rule $\psi_h^m = \psi_{h/2}^e \circ \psi_{h/2}^i$.

The Strang scheme can also be expressed as

$$\begin{aligned} S_h^{[2]} &= \varphi_{-h/2}^{[1]} \circ \left(\varphi_h^{[1]} \circ \varphi_h^{[2]}\right) \circ \varphi_{h/2}^{[1]} \\ &= \pi_h^{-1} \circ \chi_h^* \circ \pi_h, \end{aligned} \tag{1.18}$$

with $\pi_h = \varphi_{h/2}^{[1]}$. In the terminology of dynamical systems, the Strang and Lie–Trotter schemes are said to be *conjugate* to each other by the ($O(h)$-near to identity) map $\pi_h = \varphi_{h/2}^{[1]}$, which can be considered as a change of coordinates. Furthermore,

the result of n applications of the Strang scheme can be recovered from n applications of Lie–Trotter by carrying out just an initial transformation at the initial step and its *inverse* at the final step.

Since many dynamical properties are invariant under changes of coordinates, conjugate methods provide the same characterization of these properties. Other examples of conjugate methods are the trapezoidal and midpoint rules, the map π_h being in this case the implicit Euler method (Hairer *et al.* 2006). We will treat conjugate methods in detail in Sections 4 and 5.

1.4. The mathematical pendulum

The simple mathematical pendulum constitutes a standard example of a nonlinear Hamiltonian system. In appropriate units, the corresponding Hamiltonian function reads

$$H(q,p) = \frac{1}{2}p^2 + (1-\cos q), \tag{1.19}$$

where q denotes the angle from the vertical suspension point and p is the associated momentum.

As is well known, the equations of motion of a generic Hamiltonian system with Hamiltonian $H(q,p)$, and $q, p \in \mathbb{R}^d$, are given by (Goldstein 1980)

$$\frac{\mathrm{d}q}{\mathrm{d}t} = \nabla_p H, \quad \frac{\mathrm{d}p}{\mathrm{d}t} = -\nabla_q H, \tag{1.20}$$

the function $H(q,p)$ remains constant along the evolution, and the corresponding t-flow, denoted $\varphi_t^{[H]}$, is a *symplectic* transformation (Arnold 1989): its Jacobian matrix $\varphi_t'^{[H]}$ verifies the identity

$$(\varphi_t'^{[H]})^\top J\, \varphi_t'^{[H]} = J \quad \text{for } t \geq 0,$$

where J is the basic canonical matrix

$$J = \begin{pmatrix} 0_d & I_d \\ -I_d & 0_d \end{pmatrix}. \tag{1.21}$$

In the particular case of (1.19), $d = 1$, and the equations of motion are

$$\frac{\mathrm{d}q}{\mathrm{d}t} = p, \quad \frac{\mathrm{d}p}{\mathrm{d}t} = -\sin q.$$

Given a Hamiltonian $H(q,p)$ that can be decomposed as

$$H(q,p) = H_1(q,p) + H_2(q,p), \tag{1.22}$$

it makes sense to split the equations of motion (1.20) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} q \\ p \end{pmatrix} = \begin{pmatrix} \nabla_p H_1(q,p) \\ -\nabla_q H_1(q,p) \end{pmatrix} + \begin{pmatrix} \nabla_p H_2(q,p) \\ -\nabla_q H_2(q,p) \end{pmatrix},$$

so that each subsystem is itself Hamiltonian. In that case, we can then apply

Algorithm 1.1 by composing both maps and form the first-order scheme

$$x_{n+1} = \chi_h(x_n) \equiv \left(\varphi_h^{[H_2]} \circ \varphi_h^{[H_1]}\right)(x_n), \quad n = 0, 1, 2, \ldots, \tag{1.23}$$

where $x_n = (q_n, p_n)^\top$. Similarly, Algorithm 1.2 gives the second-order scheme

$$x_{n+1} = S_h^{[2]}(x_n) \equiv \left(\varphi_{h/2}^{[H_1]} \circ \varphi_h^{[H_2]} \circ \varphi_{h/2}^{[H_1]}\right)(x_n), \quad n = 0, 1, 2, \ldots. \tag{1.24}$$

Notice that, since both χ_h and $S_h^{[2]}$ are defined as compositions of flows of Hamiltonian systems and the composition of symplectic maps is also symplectic (Arnold 1989), then both (1.23) and (1.24) are *symplectic integrators* (Sanz-Serna and Calvo 1994).

The fact that schemes (1.23) and (1.24) share the symplectic property with the exact flow has remarkable consequences in practice concerning the preservation of properties and the error propagation for long-time integrations, as we will shortly illustrate.

For the Hamiltonian (1.19) describing the pendulum (and in fact for many other mechanical systems), one can separate the contributions of the kinetic energy $T(p) = \frac{1}{2}p^2$ and the potential energy $V(q) = 1 - \cos q$, so that a natural splitting of the form (1.22) is then

$$H(q, p) = T(p) + V(q). \tag{1.25}$$

This corresponds to splitting the equations of motion (1.20) into the subsystems

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} \nabla_p T(p) \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} 0 \\ -\nabla_q V(q) \end{pmatrix}, \tag{1.26}$$

which in turn implies that

$$\varphi_t^{[T]}: \begin{pmatrix} q_0 \\ p_0 \end{pmatrix} \longmapsto \begin{pmatrix} q_0 + t\,\nabla T_p(p_0) \\ p_0 \end{pmatrix} \tag{1.27}$$

and

$$\varphi_t^{[V]}: \begin{pmatrix} q_0 \\ p_0 \end{pmatrix} \longmapsto \begin{pmatrix} q_0 \\ p_0 - t\,\nabla V_q(q_0) \end{pmatrix}. \tag{1.28}$$

Then the first-order scheme (1.23) reduces to

$$q_{n+1} = q_n + h\,\nabla_p T(p_n), \quad p_{n+1} = p_n - h\,\nabla_q V(q_{n+1}), \quad n = 0, 1, 2 \ldots \tag{1.29}$$

Compared to the explicit Euler method

$$q_{n+1} = q_n + h\,\nabla_p T(p_n), \quad p_{n+1} = p_n - h\,\nabla_q V(q_n),$$

it only differs in that $\nabla_q V$ is evaluated at the updated value q_{n+1} instead of q_n. It makes sense, then, to call scheme (1.29) the symplectic Euler–VT method: we first compute the gradient of the kinetic energy T and then compute the gradient of the potential energy V.

In accordance with our treatment in Section 1.3, the adjoint of (1.29) corresponds to composing the maps $\varphi_t^{[T]}$ and $\varphi_t^{[V]}$ in reverse order,

$$p_{n+1} = p_n - h\,\nabla_q V(q_n), \quad q_{n+1} = q_n + h\,\nabla T_p(p_{n+1}), \tag{1.30}$$

so we call it the symplectic Euler–TV method.

Obviously, our discussion of schemes (1.29) and (1.30) above applies for any Hamiltonian system whose Hamiltonian function can be written in the (so-called separable) form (1.25). That is, the Lie–Trotter scheme leads to the two variants of the symplectic Euler method when it is applied to separable Hamiltonian systems.

As for the Strang splitting scheme, described in Algorithm 1.2 in general, and in (1.24) for Hamiltonian problems, when $H(q,p) = T(p) + V(q)$ it reduces to the much celebrated *Störmer–Verlet method* (Hairer, Lubich and Wanner 2003). Specifically, depending on the order in which the parts are evaluated, we have the following two variants.

Algorithm 1.3 (Störmer–Verlet–VTV). From $(q_0, p_0) = (q(0), p(0))$, for $n \geq 0$,

- $p_{n+1/2} = p_n - \frac{h}{2}\nabla_q V(q_n)$;
- $q_{n+1} = q_n + h\nabla_p T(p_{n+1/2})$;
- $p_{n+1} = p_{n+1/2} - \frac{h}{2}\nabla_q V(q_{n+1})$.

Algorithm 1.4 (Störmer–Verlet–TVT). From $(q_0, p_0) = (q(0), p(0))$, for $n \geq 0$,

- $q_{n+1/2} = q_n + \frac{h}{2}\nabla_p T(p_n)$;
- $p_{n+1} = p_n - h\nabla_q V(q_{n+1/2})$;
- $q_{n+1} = q_{n+1/2} + \frac{h}{2}\nabla_p T(p_{n+1})$.

Clearly, Algorithms 1.3 and 1.4 correspond to *time-symmetric* methods and can be obtained by composing the Euler–TV method and its adjoint. Specifically, if χ_h corresponds to method (1.30), then

$$S_h^{[2]} \equiv \chi_{h/2}^* \circ \chi_{h/2} = \varphi_{h/2}^{[V]} \circ \varphi_h^{[T]} \circ \varphi_{h/2}^{[V]} \tag{1.31}$$

recovers Störmer–Verlet–VTV, whereas the TVT version corresponds to $\varphi_{h/2}^{[T]} \circ \varphi_h^{[V]} \circ \varphi_{h/2}^{[T]}$.

Figure 1.1(a) shows trajectories of the pendulum (1.19), starting from three different initial conditions $(q_0, p_0) = (-5, \frac{5}{2}), (1, 1), (\frac{1}{10}, 0)$, corresponding to different regions of the phase space. For $(q_0, p_0) = (-5, \frac{5}{2})$ it holds that $T(p(t)) > V(q(t))$, whereas for $(q_0, p_0) = (1, 1)$ we have $T(p(t)) \simeq V(q(t))$ on average. Finally, for $(q_0, p_0) = (\frac{1}{10}, 0)$, the system can be seen as a slightly perturbed harmonic oscillator. In this case we could consider the following decomposition:

$$H = \frac{1}{2}(p^2 + q^2) + \left(1 - \frac{1}{2}q^2 - \cos q\right) \equiv H_1(q,p) + H_2(q), \tag{1.32}$$

where $H_1(q,p) = \frac{1}{2}(p^2+q^2)$ corresponds to the harmonic oscillator, whose exact solution is known (a rotation in phase space). Moreover, $|H_2(q)| = \varepsilon|H_1(q,p)|$, with $\varepsilon \approx 10^{-3}$ along the orbit originated in $(\frac{1}{10},0)$. With splitting (1.32), the map $\varphi_h^{[H_1]} \circ \varphi_h^{[H_2]}$ reads

$$\begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = R(h) \begin{pmatrix} q_n \\ p_n + h(q_n - \sin q_n) \end{pmatrix}, \quad \text{with } R(h) = \begin{pmatrix} \cos h & \sin h \\ -\sin h & \cos h \end{pmatrix}, \tag{1.33}$$

whereas the second-order scheme $\varphi_{h/2}^{[H_1]} \circ \varphi_h^{[H_2]} \circ \varphi_{h/2}^{[H_1]}$ can be formulated as

$$\begin{aligned} \begin{pmatrix} q_{n+1/2} \\ p_{n+1/2} \end{pmatrix} &= R(h/2) \begin{pmatrix} q_n \\ p_n \end{pmatrix}, \\ p^*_{n+1/2} &= p_{n+1/2} + h(q_{n+1/2} - \sin q_{n+1/2}), \\ \begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} &= R(h/2) \begin{pmatrix} q_{n+1/2} \\ p^*_{n+1/2} \end{pmatrix}. \end{aligned} \tag{1.34}$$

To illustrate the different splittings, we take $(\frac{1}{10},0)$ as the initial condition, integrate until the final time $t_f = 500$ and measure the relative error in energy, $|H(q_n,p_n) - H(q_0,p_0)|/|H(q_0,p_0)|$, along the trajectory. The step size is taken so that all the methods tested require the same number of evaluations of the potential (and thus essentially involve the same computational cost): 1200 evaluations (Figure 1.1(b)) and 2400 evaluations (Figure 1.1(c)). The schemes we test are as follows: on the one hand, the Störmer–Verlet algorithm (1.31) (denoted by S_2 in the graphs) and the fourth-order Runge–Kutta–Nyström splitting method proposed in Blanes and Moan (2002) ($\mathrm{RKN}_6 4$), as representatives of the $T+V$ splitting; on the other hand, the specially adapted schemes (1.34), denoted $(2,2)$, and the $(10,6,4)$ integrator presented in Blanes *et al.* (2013*b*), both for the case where $H = H_1 + H_2$, with H_2 small compared with H_1. The notation $(10,6,4)$ refers to the fact that the local error of the method is of order $O(\varepsilon h^{11} + \varepsilon^2 h^7 + \varepsilon^3 h^5)$ if H_2 is ε times smaller than H_1. For analogy, we label the Strang method applied to the perturbed harmonic oscillator (1.32) as $(2,2)$. Notice that the splitting (1.32) is more advantageous for this initial condition (the simple method (1.34) behaves better than the fourth-order scheme), and that the improvement with respect to the $T+V$ splitting is approximately of the size of ε. We therefore see that for this type of problem it is possible to construct integrators, providing much more accurate results with the same computational effort.

Since all the schemes are symplectic integrators and the evolution is taking place in a compact domain, the error in energy is bounded (in contrast to standard non-symplectic methods, whose error in energy usually grows linearly with t), whereas the error in phase space (q,p) grows linearly when applied to near-integrable Hamiltonian systems (Hairer *et al.* 2006). Notice, however, that if the scheme is conjugate to another more accurate integrator, then the global error in phase space will remain bounded for some time interval before it starts growing linearly; see

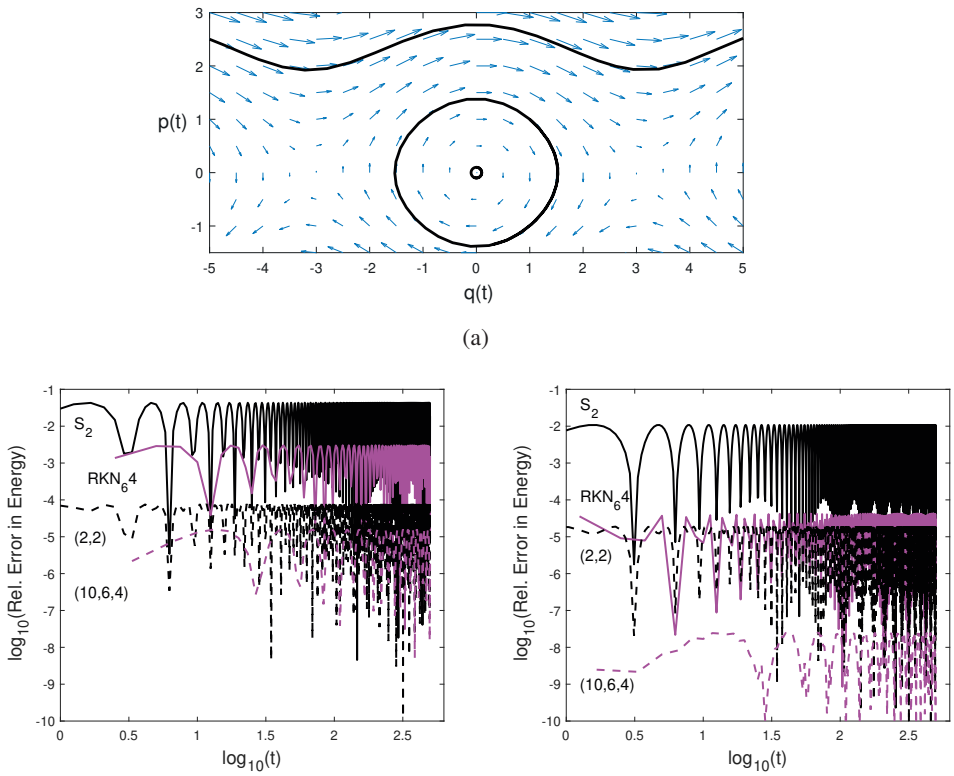


Figure 1.1. Simple pendulum. (a) Phase space and three trajectories with initial conditions $(q_0, p_0) = (-5, \frac{5}{2}), (1, 1), (\frac{1}{10}, 0)$. (b,c) Relative error in energy committed by different splitting methods along the solution with initial condition $(q_0, p_0) = (\frac{1}{10}, 0)$ in the interval $t \in [0, 500]$ with (b) 1200 evaluations and (c) 2400 evaluations of the potential.

Section 4.5 for more details. This feature is illustrated in Figure 1.2: we integrate the system starting with the same initial condition and final time $t_f = 500$ with $h = \frac{5}{12}$ and compute the relative error

$$\|(q(t_n), p(t_n)) - (q_n, p_n)\| / \|(q(t_n), p(t_n))\|,$$

with different schemes. The reference solution $(q(t_n), p(t_n))$ is computed numerically with very high accuracy. Specifically, we test the following integrators: the Lie–Trotter method, equation (1.23) (denoted LT in the graph) and the Störmer–Verlet (S_2) method for the splitting $H = T + V$, and scheme (2, 2) and (1.33) ($\mathrm{LT}_{\mathrm{pert}}$), which corresponds to the Lie–Trotter method applied to the perturbed harmonic

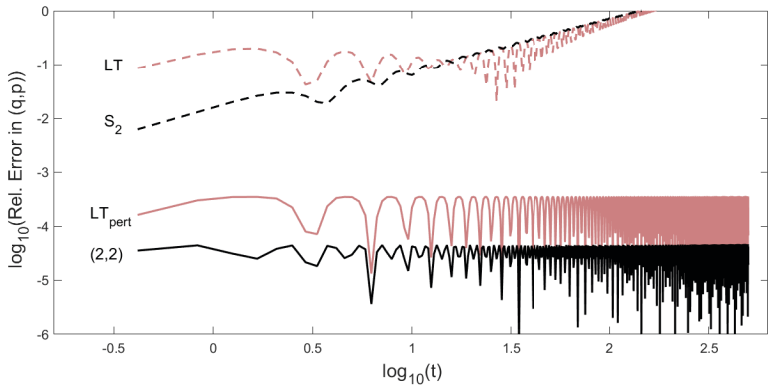


Figure 1.2. Pendulum. Relative error in phase space for different splitting methods along the solution with initial condition $(q_0, p_0) = (\frac{1}{10}, 0)$ in the interval $t \in [0, 500]$ and time step $h = \frac{5}{12}$.

oscillator (1.32). We observe that, since LT and S_2 are conjugate to each other (see (1.18)), their errors are quite similar after some time interval. On the other hand, $(2, 2)$ and $\mathrm{LT}_{\mathrm{pert}}$ (which are also conjugate to each other) show a different behaviour: no linear growth is visible, and the error of $\mathrm{LT}_{\mathrm{pert}}$ is larger than that of (2,2) by approximately the same factor for the whole time interval considered in Figure 1.2. This is related to the fact that $(2, 2)$ and $\mathrm{LT}_{\mathrm{pert}}$ are conjugate to another method with a local error essentially $O(h^3\varepsilon^2)$; see Sections 4.5 and 5.7. More comments on these observations along with additional explanations will be given in Section 4.5.

1.5. The gravitational N-body problem

Another popular example to illustrate the behaviour and performance of splitting methods corresponds to the important problem in classical mechanics of a planetary system modelled as N bodies (a massive star and $N - 1$ planets) under mutual gravitational Newtonian interaction. This is also a Hamiltonian system with

$$H(q, p) = \sum_{i=0}^{N-1} \frac{1}{2m_i} p_i^\top p_i - G \sum_{i=1}^{N-1} \sum_{j=0}^{i-1} \frac{m_i m_j}{\|q_i - q_j\|}. \tag{1.35}$$

Here (q, p) denote the 'supervectors' composed by the positions $q_i \in \mathbb{R}^3$ and momenta $p_i \in \mathbb{R}^3$ of the 'Sun' $(i = 0)$ and the $N-1$ planets $(i = 1, \ldots, N-1)$ in some Cartesian coordinate system: $q = (q_0, q_1, \ldots, q_{N-1})^\top$, $p = (p_0, p_1, \ldots, p_{N-1})^\top$. In (1.35), m_i is the mass of the ith body, and G is the universal gravitational constant.

Now the equations of motion (1.20) read

$$\begin{aligned} \frac{\mathrm{d}q_i}{\mathrm{d}t} &= \nabla_{p_i} H = \frac{1}{m_i} p_i, \quad i = 0, \dots, N-1, \\ \frac{\mathrm{d}p_i}{\mathrm{d}t} &= -\nabla_{q_i} H = -G \sum_{j \neq i} \frac{m_i m_j}{\|q_i - q_j\|^3} (q_i - q_j), \quad i, j = 0, \dots, N-1. \end{aligned} \tag{1.36}$$

Since the kinetic energy $T(p)$ and the potential energy $V(q)$ are in this case

$$T(p) = \sum_{i=0}^{N-1} \frac{1}{2m_i} p_i^\top p_i, \quad V(q) = -G \sum_{i=1}^{N-1} \sum_{j=0}^{i-1} \frac{m_i m_j}{\|q_i - q_j\|}, \tag{1.37}$$

it also makes sense to separate the Hamiltonian (1.35) as $H(q, p) = T(p) + V(q)$, so that the symplectic Euler and the Störmer–Verlet schemes can be applied to arbitrary configurations of the bodies. However, this choice is suboptimal for planetary systems, where planets describe near-Keplerian orbits around the central star. An alternative procedure taking advantage of the hierarchical nature of the motion of the planets around the central massive body was first proposed in Wisdom and Holman (1991), and is known in the literature as the Wisdom–Holman integration map. It essentially consists in changing coordinates so that the transformed H can be written as an integrable part H_1 (corresponding to the Keplerian motion of the planets) and a small perturbation H_2 (that accounts for the gravitational interaction of the planets among themselves), and then applying the second-order scheme (1.24) to this new Hamiltonian.

Specifically, Wisdom and Holman (1991) consider a linear canonical change of variables to rewrite (1.35) in the so-called Jacobi coordinates $(\hat{q}_i, \hat{p}_i)$, $i = 0, 1, \dots, N-1$. Here $\hat{q}_0$ is the position of the centre of mass of the system, $\hat{q}_1$ is the relative position of the first planet with respect to the central star, and for $i = 2, \dots, N-1$, $\hat{q}_i$ is the position of the ith planet relative to the centre of mass of the central star and the planets with lower indices. That is,

$$\hat{q}_0 = \frac{1}{M_{N-1}} \sum_{j=0}^{N-1} m_j q_j, \quad \hat{q}_i = q_i - \frac{1}{M_i} \sum_{j=0}^{i-1} m_j q_j, \quad i = 1, \dots, N-1, \tag{1.38}$$

where $M_i = \sum_{j=0}^{i} m_j$ for $i = 0, \dots, N-1$. This can be written in a more compact way as $\hat{q} = Aq$, where A is an $N \times N$ invertible matrix with mass-dependent entries, $\hat{q} = (\hat{q}_0, \dots, \hat{q}_{N-1})^\top$ and $q = (q_0, \dots, q_{N-1})^\top$.

The conjugate momenta $\hat{p}_i$, $i = 0, 1, \dots, N-1$, are uniquely determined by the requirement that the change of variables be canonical (i.e. $p = A^\top \hat{p}$), so that the transformed Hamiltonian function is obtained by rewriting (1.35) in the new variables: $\hat{H}(\hat{q}, \hat{p}) \equiv H(A^{-1}\hat{q}, A^\top \hat{p})$.

It is straightforward to check that the kinetic energy, expressed in terms of $\hat{p}$, has the same diagonal structure as in p:

$$T = \sum_{i=0}^{N-1} \frac{1}{2m_i} p_i^\top p_i = \sum_{i=0}^{N-1} \frac{1}{2\hat{m}_i} \hat{p}_i^\top \hat{p}_i,$$

where

$$\hat{m}_0 = M_{N-1}, \quad \hat{m}_i = \frac{M_{i-1}}{M_i} m_i, \quad i = 1, \ldots, N-1.$$

The Hamiltonian in the new variables $\hat{H}(\hat{q}, \hat{p})$ can be split as $\hat{H}(\hat{q}, \hat{p}) = H_1(\hat{q}, \hat{p}) + H_2(\hat{q})$, where

$$\begin{aligned} H_1(\hat{q}, \hat{p}) &= \frac{1}{2\hat{m}_0} \hat{p}_0^\top \hat{p}_0 + \sum_{i=1}^{N-1} \left(\frac{1}{2\hat{m}_i} \hat{p}_i^\top \hat{p}_i - \frac{m_0 m_i}{\|\hat{q}_i\|} \right), \\ H_2(\hat{q}) &= V(q) + G \sum_{i=1}^{N-1} \frac{m_0 m_i}{\|\hat{q}_i\|} \\ &= G m_0 \sum_{i=1}^{N-1} m_i \left(\frac{1}{\|\hat{q}_i\|} - \frac{1}{\|q_i - q_0\|} \right) - G \sum_{i=2}^{N-1} \sum_{j=1}^{i-1} \frac{m_i m_j}{\|q_i - q_j\|}, \end{aligned} \tag{1.39}$$

and q has to be expressed in terms of $\hat{q}$ according to $q = A^{-1}\hat{q}$.

Observe that the potential energy does not depend on $\hat{q}_0$, so $\hat{p}_0$ is constant (in fact it is the linear momentum of the system) and therefore we can remove it from $\hat{H}$ if we assume that the centre of the mass is at rest.

Clearly, for fixed $\hat{q}$ and varying mass ratios,

$$\hat{q}_i = q_i - q_0 + O(\varepsilon), \quad H_2(\hat{q}) = O(\varepsilon) \quad \text{as } \varepsilon \equiv \frac{1}{m_0} \max_{1 \le i \le N-1} m_i \to 0.$$

Hamiltonian H_1 can then be considered as a collection of $N-1$ two-body problems, and H_2 as a perturbation. It turns out that the flow $\varphi_h^{[H_1]}$ can be computed with the algorithm proposed in Danby (1988, p. 165), for example, whilst H_2 only depends on $\hat{q}$, and thus its flow $\varphi_h^{[H_2]}$ can be explicitly evaluated in an efficient way. Notice that the number of terms in H_1 grows linearly with the number of bodies N, whereas the number of terms in H_2 grows quadratically.

As an illustrative example, we next consider the outer Solar System modelled as a six-body system with the inner Solar System ($i = 0$), the four giant planets ($i = 1, 2, 3, 4$) and Pluto ($i = 5$), all considered as point masses. The initial conditions for each planet are taken at Julian time (TDB) 2440400.5 (28 June 1969), obtained from the DE430 ephemerides (Folkner *et al.* 2014) and normalized so that the centre of mass of the system is at rest. A schematic diagram of the trajectories is shown in Figure 1.3(a) with the initial (circles) and final (stars) positions of each object after 200 000 days. For this problem we test the same methods as for the pendulum in Figure 1.1: on the one hand, Störmer–Verlet (1.31), S_2,

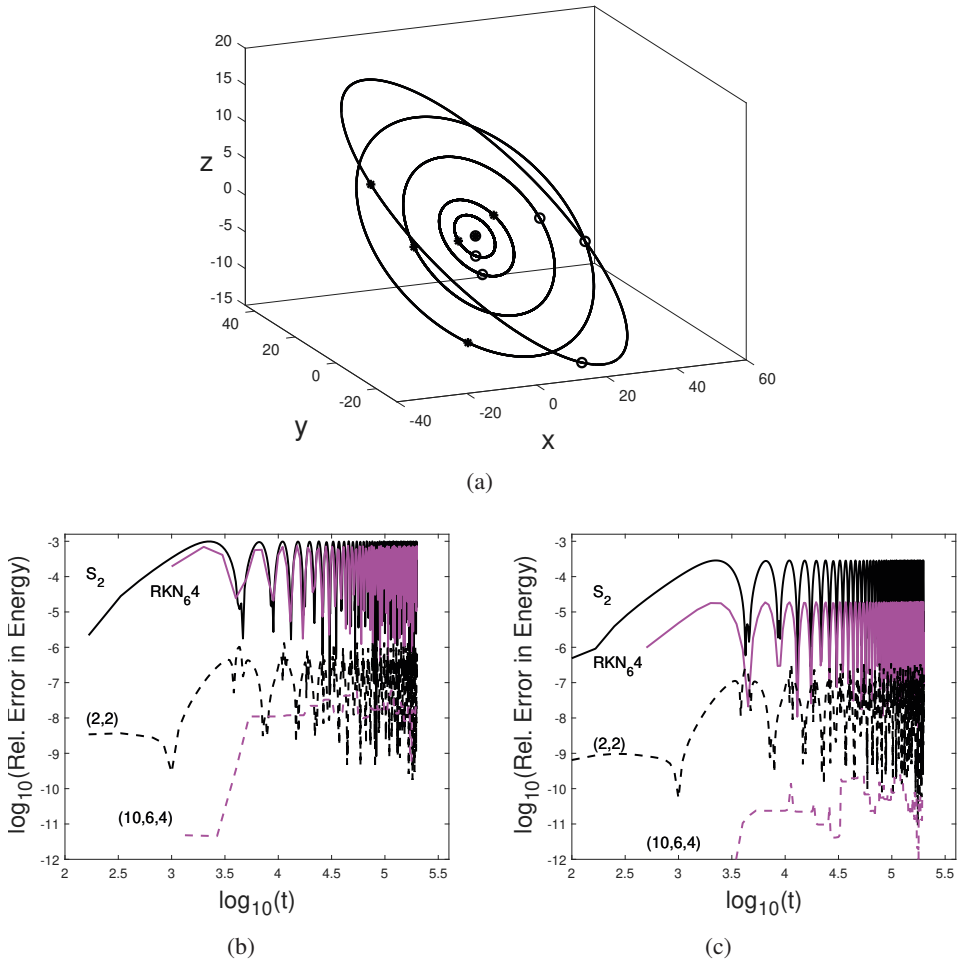


Figure 1.3. (a) Trajectories of the six-body system modelling the outer Solar System. (b,c) Relative error in energy as a function of time for an interval of 200 000 days obtained with different splitting methods with (b) 1200 evaluations and (c) 2400 evaluations of the force.

and the fourth-order Runge–Kutta–Nyström splitting method of Blanes and Moan (2002), $\mathrm{RKN}_6 4$ when the Hamiltonian (1.35) is separated into kinetic and potential energy (1.37); on the other hand, the specially adapted schemes (1.34) (called in this setting the Wisdom–Holman integrator, and denoted (2,2) as before) and the $(10, 6, 4)$ integrator presented in Blanes *et al.* (2013*b*) when H is expressed as $H(\hat{q}, \hat{p}) = H_1(\hat{q}, \hat{p}) + H_2(\hat{q})$, with (1.39) in Jacobi coordinates. We integrate for a relatively short time interval, $t_f = 200\,000$ days (or approximately 46 periods of Jupiter and two periods of Pluto), and compute the relative error in the energy with each integrator with a step size so that all of them require the same number

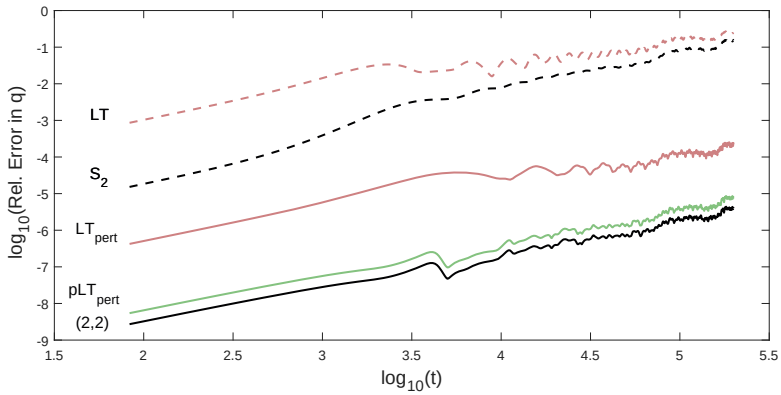


Figure 1.4. Outer Solar System. Relative error in position as a function of t for a time interval $[0, t_f = 200\,000]$ days with $h = t_f/1200$ obtained with different splitting methods.

of force evaluations. The results are displayed in Figure 1.3 with (b) 1200 and (c) 2400 evaluations. As in the previous example of the pendulum, the error in energy remains bounded in all cases, and scheme $(2, 2)$ provides an error energy almost 1000 times smaller than S_2. This illustrates the fact that, by taking a splitting adapted to the structure of the problem and designing integrators taking these specific features into account, it is possible to greatly improve the efficiency.

In Figure 1.4 we display the relative errors in positions $q = (q_0, \ldots, q_5) \in \mathbb{R}^{18}$ along the time integration for the same methods as in Figure 1.2. As in the pendulum problem, the error for LT is larger than for S_2 at the beginning of the integration interval, but they become similar after some time. Here $(2, 2)$ is also more accurate than S_2 by a factor significantly smaller than ε. Compared to the pendulum problem, the error grows linearly right from the beginning for schemes $\mathrm{LT}_{\mathrm{pert}}$ and $(2, 2)$ (although with a smaller slope than LT and S_2). The curve labelled by $\mathrm{pLT}_{\mathrm{pert}}$ corresponds to the relative error obtained by applying $\mathrm{LT}_{\mathrm{pert}}$ with initial conditions $(\bar{q}_0, \bar{p}_0) = \varphi_{h/2}^{[H_1]}(q_0, p_0)$. Notice that the phase errors of $\mathrm{pLT}_{\mathrm{pert}}$ are very similar to those of $(2, 2)$. All these observations will be accounted for in Section 4.5.

1.6. The time-dependent Schrödinger equation

The basic object to study the time evolution of a system in quantum mechanics is the (time-dependent) Schrödinger equation. In the case of one particle of unit mass in a potential $V(x)$, it reads (Messiah 1999)

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\psi(x, t) = -\frac{1}{2}\Delta\psi(x, t) + V(x)\psi(x, t), \quad \psi(x, 0) = \psi_0(x). \tag{1.40}$$

Here $\psi\colon \Omega \subset \mathbb{R}^3 \times \mathbb{R} \longrightarrow \mathbb{C}$ is the wave function, representing the state of the system, and $\hbar$ is the reduced Planck constant. The quantity $|\psi(x,t)|^2$ represents a probability density for the position of the particle, in the sense that the probability of the particle to be located in $S \subset \Omega$ at time t is $\int_S |\psi(x,t)|^2 \,\mathrm{d}x$. The equation is then defined in the Hilbert space $L^2(\Omega, \mathbb{C})$.

If we introduce the self-adjoint operators $\hat{T}$, $\hat{V}$ acting on $\psi \in L^2(\Omega, \mathbb{C})$ as

$$\hat{T}\psi = -\frac{1}{2}\Delta\psi, \quad \hat{V}\psi = V(x)\psi,$$

then a straightforward calculation shows that $[\hat{V}, [\hat{T}, \hat{V}]]\psi = |\nabla V|^2\psi$, and therefore

$$[\hat{V}, [\hat{V}, [\hat{T}, \hat{V}]]]\psi = 0. \tag{1.41}$$

A standard approach for applying splitting methods in this setting consists in first discretizing the equation in space. If we consider the one-dimensional case for simplicity and if the wave function is negligible outside a space interval $[a, b]$ on the time interval of interest, then we can limit ourselves to the study of the equation on that finite interval with periodic boundary conditions (Lubich 2008). After rescaling, the periodic interval can always be restricted to $[-\pi, \pi]$. In this way, the original problem is transformed into ($\hbar = 1$)

$$\mathrm{i}\frac{\partial}{\partial t}\psi(x,t) = -\frac{1}{2}\frac{\partial^2\psi}{\partial x^2}(x,t) + V(x)\psi(x,t), \quad x \in [-\pi, \pi], \tag{1.42}$$

with $\psi(-\pi, t) = \psi(\pi, t)$ for all t.

The wave function is then approximated by a trigonometric polynomial $u(x,t)$ whose coefficients are obtained by requiring that the approximation satisfies (1.42) in a grid of M equispaced points $x_j = -\pi + j \cdot 2\pi/(M-1)$ on the interval $[-\pi, \pi]$. The vector $u = (u_0, \ldots, u_{M-1})^\top \in \mathbb{C}^M$ formed by the grid values $u_j \approx \psi(x_j, t)$, $j = 0, 1, \ldots, M-1$, then verifies the M-dimensional linear ODE

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}u(t) = H\,u(t) = (T+V)u(t), \quad u(0) = u_0 \in \mathbb{C}^M. \tag{1.43}$$

Here $V = \mathrm{diag}(V(x_j))$ and $T = -\frac{1}{2}D$, where D is the second-order periodic spectral differentiation matrix (Trefethen 2000). As is well known, $Tu = \mathcal{F}^{-1}D_T\mathcal{F}u$, where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the forward and backward discrete Fourier transform, and D_T is also diagonal. The transformations $\mathcal{F}$ and $\mathcal{F}^{-1}$ are computed with the fast Fourier transform (FFT) algorithm, requiring $O(M \log M)$ operations.

Notice that solving equations $\mathrm{i}u' = Tu$ and $\mathrm{i}u' = Vu$ is done trivially by using exponentials of diagonal matrices and FFTs, namely,

$$\left(\mathrm{e}^{\tau V}u\right)_j = \mathrm{e}^{\tau V(x_j)}u_j, \quad \mathrm{e}^{\tau T}u = \mathcal{F}^{-1}\mathrm{e}^{\tau D_T}\mathcal{F}\,u,$$

for a time step h, with $\tau = -\mathrm{i}h$. Therefore, splitting methods constitute a valid alternative way to approximate the solution $u(t) = \mathrm{e}^{\tau H}u_0 = \mathrm{e}^{\tau(T+V)}u_0$, which may be prohibitively expensive to evaluate for large values of M. Thus the Lie–Trotter

scheme reads

$$\mathrm{e}^{\tau(T+V)} = \mathrm{e}^{\tau T}\,\mathrm{e}^{\tau V} + O(\tau^2), \tag{1.44}$$

whereas the second-order Strang splitting constructs the numerical approximation u_{n+1} at time $t_{n+1} = t_n + \Delta t$ by

$$u_{n+1} = \mathrm{e}^{\tau/2V}\,\mathrm{e}^{\tau T}\,\mathrm{e}^{\tau/2V}\,u_n. \tag{1.45}$$

The resulting scheme is called the *split-step Fourier* method in the chemical literature (Feit, Fleck Jr and Steiger 1982) and has some remarkable properties. In particular, it is unitary and time-symmetric (Lubich 2008), as is the exact solution $\mathrm{e}^{\tau H}$.

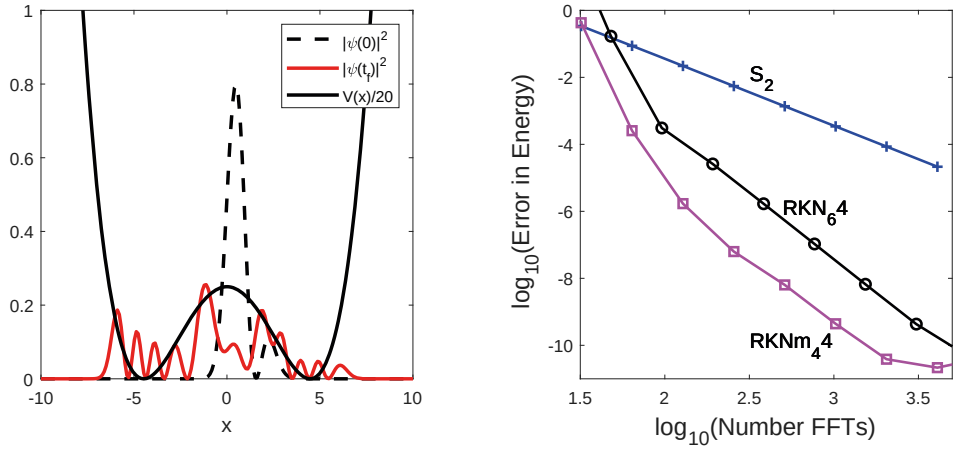
Relation (1.41) still holds for the matrices T and V if the number of points M in the space discretization is sufficiently large, and in fact $[V,[T,V]]$ is diagonal if the derivatives of the potential are computed first and then evaluated at the corresponding space grid.

We next illustrate the procedure with the one-dimensional double-well potential

$$V(x) = \frac{1}{80}(x^2-20)^2, \tag{1.46}$$

and the initial wave function $\psi(x,0) = \psi_0(x) = \sigma\cos^2(x)\,\mathrm{e}^{-\frac{1}{2}(x-1)^2}$, where σ is an appropriate normalizing constant. We take $M = 256$ discretization points on the interval $x \in [-13, 13]$ and integrate the resulting linear ODE (1.43) in the interval $0 \le t \le t_f = Nh = 10$. Figure 1.5(a) shows $|\psi_0(x)|^2$, $|\psi(x,t_f)|^2$ and the potential $V(x)$, whereas Figure 1.5(b) shows an efficiency diagram. Specifically, we display the error in energy measured at the final time, $|u_N^\top H u_N - u_0^\top H u_0|$, as a function of the number of FFT calls (and its inverse) as an estimate of the computational effort of each method. The lines correspond to the Strang splitting (1.45), S_2, with time steps $h = 10/2^k$, $k = 1, 2, \ldots, 12$, the fourth-order RKN splitting method $\mathrm{RKN}_6 4$ from Blanes and Moan (2002), already illustrated in Figures 1.1 and 1.3, and another fourth-order scheme including the double commutator $[V,[T,V]]$ into its formulation (denoted $\mathrm{RKNm}_4 4$). In this diagram the slope of each line for sufficiently small h indicates the order of the scheme. As in the previous examples, by taking into account the specific features of the problem at hand it is possible to construct more accurate and efficient numerical approximations.

The error in energy also remains bounded for these unitary integrators, as in the previous examples involving classical Hamiltonian problems, whereas the error in the wave function grows linearly with t, unless the scheme is conjugate to another more accurate one, in which case it is bounded for some time before linear growth takes place. To illustrate this feature, in Figure 1.6 we depict how the error in the solution $\|u(t_n) - u_n\|$ evolves with time for a longer integration interval $t \in [0, 1000]$, with step size $h = \frac{1}{20}$. As usual, the reference solution is computed numerically with sufficiently high accuracy, and the tested schemes are as follows: Lie–Trotter, equation (1.44) (LT), Strang, equation (1.45), and a variant of Strang



(a) (b)

Figure 1.5. Time-dependent Schrödinger equation with a double-well potential. (a) $V(x)$, initial and final wave function with $M = 256$ discretization points. (b) Relative error in energy at the final time vs. number of FFTs for different values of the time step obtained with the Strang method S_2, and two fourth-order splitting schemes: one involving six evaluations of V ($\mathrm{RKN}_6 4$) and another with four evaluations of V, and incorporating in addition the double commutator $[V, [T, V]]$ ($\mathrm{RKNm}_4 4$).

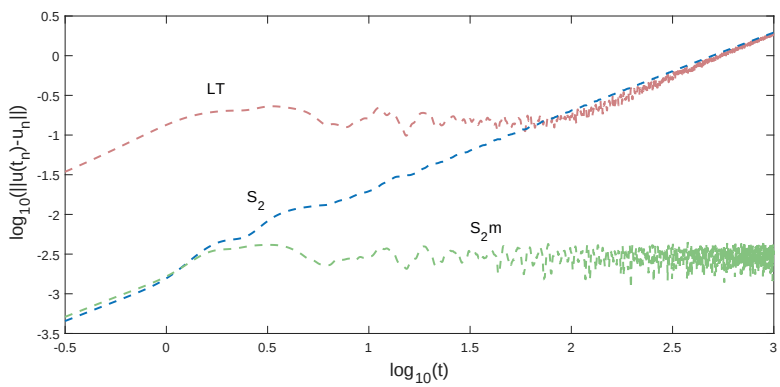


Figure 1.6. Time-dependent Schrödinger equation with a double-well potential: relative error in the wave function as a function of t with $h = 1/20$ for Strang (S_2), Lie–Trotter (LT) and scheme S_2m involving a double commutator and conjugate to a method of order four.

involving a double commutator, namely

$$S_2m: \quad u_{n+1} = \mathrm{e}^{\tau/2V+(\tau^3/48)[V,[T,V]]}\,\mathrm{e}^{\tau T}\,\mathrm{e}^{\tau/2V+(\tau^3/48)[V,[T,V]]}\,u_n. \tag{1.47}$$

This scheme does not require any additional FFTs, and in addition it is actually conjugate to a method of order four. We observe that LT approaches S_2 after a transition time, whereas the error for S_2m remains bounded for the whole interval. In fact, its linear error growth only appears when the time interval is doubled.

1.7. Splitting methods as geometric numerical integrators

As the examples of Sections 1.2–1.6 illustrate, splitting schemes, even of low order of consistency such as Lie–Trotter and Strang methods, preserve by construction structural properties of the exact solution, such as symplecticity (in classical Hamiltonian dynamics) and unitarity (in quantum evolution problems). This feature gives them qualitative superiority with respect to other standard integrators in practice, especially when long time intervals are concerned. In this sense, splitting methods constitute an important class of *geometric numerical integrators*.

Although the idea that numerical integrators applied to an ordinary differential equation should preserve as many properties of the system as possible has been implicitly assumed since the early days of numerical analysis, it is fair to say that in the classical consistency/stability approach the emphasis has been on other issues. In particular, the goal has mainly been to compute the solution of (1.1) at time $t_f = Nh$ with a global error $\|x_N - x(t_f)\|$ smaller than a prescribed tolerance and as efficiently as possible. To do that, we choose the class of method (one-step, multistep, extrapolation, etc.), the order (fixed or adaptive) and the time step (constant or variable). This approach has proved to be very fruitful, giving rise to highly tuned and thoroughly tested software packages generally available to solve a great variety of problems.

On the other hand, there are special types of problems arising in many fields of science and applied mathematics that possess an underlying geometric structure which influences the qualitative character of their solutions, so we naturally aim to construct numerical approximations that preserve this structure. Classical Hamiltonian systems such as those illustrated previously constitute a case in point. It turns out, however, that many numerical integrators included in standard software packages do not take into account these distinctive features of the equations to be solved, and the question is whether it is possible to design, analyse and apply new schemes providing approximate solutions that share one or several geometric properties with the exact solution. This is precisely the realm of geometric numerical integration, a terminology introduced in Sanz-Serna (1997).

According to McLachlan and Quispel (2006):

> 'Geometric integration' is the term used to describe numerical methods for computing the solution of differential equations, while preserving one or more physical/mathematical properties of the system *exactly* (i.e., up to round-off error).

Thus, rather than primarily taking into account prerequisites such as consistency and stability, the aim is to reproduce the qualitative features of the solution of the differential equation being discretized, in particular its geometric properties, such as the symplectic character (for Hamiltonian systems) and unitarity (quantum mechanics), but also the phase-space volume (for divergence-free vector fields), time-reversal symmetries, first integrals of motion (energy, linear and angular momentum), Casimirs, Lyapunov functions, etc. In these structure-preserving methods we try to incorporate as many of these properties as possible and, as a result, they exhibit improved qualitative behaviour. In addition, they typically allow for a significantly more accurate integration for long-time intervals than with general-purpose methods (Hairer *et al.* 2006, Blanes and Casas 2016).

Although splitting methods have a long history in numerical mathematics and have been applied, sometimes with different names, in many different contexts (e.g. partial differential equations, quantum statistical mechanics, chemical physics and molecular dynamics), it is fair to say that the interest in splitting has revived with the advent of geometric numerical integration, and new and very efficient schemes have been put to use to solve a wide variety of problems. The reason is clear: if the problem (1.3) has some property that is deemed to preserve (symplectic, unitary, volume-preserving, etc.) and each subproblem $x' = f_j(x)$, $x(0) = x_0$ can be integrated exactly or by means of a numerical method preserving these properties, then the splitting method constructed by composing the solution of the subproblems is also symplectic, unitary, volume-preserving, etc. In other words, splitting methods provide by construction approximations lying in the same group of diffeomorphisms as the system $x' = f(x)$ (McLachlan and Quispel 2002). Here we assume of course that each subproblem $x' = f_j(x)$ possesses the same characteristic feature as the total problem considered.

1.8. Relevance of splitting methods

Given a certain differential equation $x' = f(x)$, the application of splitting methods to solve the corresponding initial value problem involves three main steps (McLachlan and Quispel 2002).

(1) Choosing the set of terms f_j such that $f = f_1 + \cdots + f_m$. Different ways of decomposing f may give rise to integrators with different qualitative behaviour and efficiency, as we have seen in the previous examples.

(2) Solving each subproblem $x' = f_j(x)$ either exactly or approximately.

(3) Combining these solutions to get an approximation for the original overall problem.

Being such a simple idea, it is hardly surprising that the splitting principle can be used in so many different settings. In particular, one may

- split the differential equation into linear and nonlinear parts;

- in an ODE describing a Hamiltonian system with an additional small dissipation, separate the Hamiltonian part and the dissipation;
- decompose into parts describing different physical processes, for example diffusion and reaction in partial differential equations;
- get approximations to the original problem by solving one space direction at a time (dimensional splitting in PDEs).

In addition, splitting methods possess some advantages concerning their implementation, in particular the following.

- They are typically explicit.
- Their storage requirements are quite small. The algorithms are sequential and the solution at intermediate stages can be stored in the solution vectors.
- Programming higher-order schemes is no more difficult than Lie–Trotter and Strang splitting methods, at least in the context of ODEs. Usually, a few more lines of code is all that is required to deal with the additional stages.
- As stated earlier, they can preserve a wide variety of structures possessed by the differential equation.

They also present some disadvantages, of course. Among them, we can mention the following.

- Splitting schemes of order three or higher necessarily involve negative coefficients. In other words, they require substeps that go backwards in time, and this has severe repercussions when applying them to, for instance, reaction–diffusion equations (see Section 6).
- Although it is possible to construct accurate high-order splitting methods, stability can be an issue, in the sense that their stability interval might be reduced to render them useless in practice. This aspect has to be taken seriously when designing new methods.
- Ordinary splitting does not capture the correct steady-state solutions (where these solutions exist) (MacNamara and Strang 2016), in the sense that the numerical solutions obtained converge to limits that are not steady-state solutions but just approximations of them. This can lead to unacceptable errors, for instance in the simulation of combustion. In this context, balanced splitting techniques have been introduced to correct this flaw (Speth, Green, MacNamara and Strang 2013).

Splitting methods constitute an important tool in different areas of science where the evolution of systems is governed by differential equations. In addition to Hamiltonian systems, they can be successfully applied in the numerical study of Poisson systems, systems possessing integrals of motion (such as energy and angular momentum) and systems with (continuous, discrete and time-reversal)

symmetries. In fact, splitting methods have been designed (often independently) and extensively used in fields as distant as molecular dynamics, simulation of storage rings in particle accelerators, celestial mechanics, astronomy, quantum (statistical) mechanics, plasma physics, hydrodynamics and Markov chain Monte Carlo methods.

Operator-splitting methods also appear outside the realm of differential equations, and in particular in optimization, in a variety of different special forms and different denominations (gradient-projection, proximal-gradient, alternating direction method of multipliers or ADMM, split Bregman, etc.). All of them are related to special types of splitting methods, such as Douglas–Rachford and Peaceman–Rachford schemes. More details can be found in several contributions collected in the comprehensive book by Glowinski, Osher and Yin (2016*a*).

1.9. Some historical remarks

There is ample consensus that the beginning of splitting is related to the product formula (1.5). What is not so clear is the origin of the formula itself. Thus Reed and Simon (1980, p. 295) establish it as 'the classical theorem of Lie', but give no exact source, whereas Chorin, Huges, Marsden and McCracken (1978) call it 'the 1875 formula of S. Lie', citing the classical treatise of Lie (1888), and, based on this reference, Glowinski *et al.* (2016*a*) even ascribe to Lie himself 'the first operator-splitting scheme recorded in history'. The problem is that the reference Lie (1888) is clearly not from 1875, and it is not evident (at least to us) that this formula appears there explicitly.

On the other hand, as pointed out in Cohen, Friedland, Kato and Kelly (1982), the result (1.5) can be found in several references published during the 1950s, namely Butler and Friedman (1955) and Golden (1957), whereas it was Trotter (1959) who generalized it to self-adjoint linear operators, without mentioning Lie or these previous references. Subsequently, formula (1.5), even in the matrix case, has been attributed to Trotter (Bellman 1970, p. 181). In view of the situation, we believe we are not committing an act of historical injustice by referring to the approximation (1.6) and Algorithm 1.1 as the *Lie–Trotter* scheme.

With respect to the splitting method (1.9), it first appeared in print in Strang (1968) as an alternative way to solve multidimensional problems with one-dimensional operators. We have already seen that, when applied to Hamiltonian systems of the form $H(q, p) = T(p) + V(q)$, it leads to Algorithm 1.4 when composing the flows associated to $T(p)$ and $V(q)$. It is called the *Störmer–Verlet method* since it was used by the astronomer Carl Störmer (1907) in his computations of the motion of ionized particles in the Earth's magnetic field, and by Loup Verlet (1967) in molecular dynamics. It is also referred to as the *leapfrog method* in the context of PDEs describing wave propagation, and as the *Wisdom–Holman method* when applied to the splitting (1.39) (Wisdom and Holman 1991). In fact, it can be found

in several classical references, the oldest one being perhaps Newton's *Principia*.[2] For a detailed account the reader is referred to the enlightening review by Hairer *et al.* (2003).

As well as symplecticity when applied to Hamiltonian systems, the Störmer–Verlet method preserves many other geometric properties of the exact flow associated with an ordinary differential equation. This includes the preservation of all linear first integrals (such as the linear momentum), and quadratic first integrals of the form $I(q,p) = p^\top C q$ for Hamiltonian systems, where C is a symmetric matrix. In other words, $I(q,p)$, computed along the numerical trajectory, is constant. A classical example is the angular momentum in N-body problems if the forces only depend on the distances of the particles.

All these favourable properties, in addition to its optimal stability property and reversibility, help us to understand why this method is probably the most widely used splitting scheme and geometric integrator, especially in molecular dynamics (Schlick 2010), condensed matter simulations (Ceperley 1995) and sampling with the hybrid Monte Carlo method (Neal 2011).

The convenience of designing numerical integration methods that, by construction, preserve the symplectic structure when applied to Hamiltonian systems was duly recognized during the 1950s in the field of accelerator physics. Thus, in the words of an early pioneer, 'if one wishes to examine solutions to differential equations, adoption of a "Hamiltonian" or "canonical" integration algorithm would be reassuring' (Laslett 1986). This was the point of view adopted in a pioneering paper by de Vogelaere (1956),[3] where he devoted himself to the task of designing 'a method of integration which, if there was no round-off error, would give a solution with the contact transformation property'. Here *contact transformation* has to be understood as *symplectic transformation*. The first-order schemes proposed by de Vogelaere (1956), although implicit in general, turn out to be explicit when $H(q,p) = T(p) + V(q)$, in which case they reproduce the symplectic Euler schemes (1.29) and (1.30).

It was another accelerator physicist, Ronald Ruth (1983), who presented what is probably the first splitting method of order three. This paper can be considered as the actual starting point in the systematic exploration of symplectic integrators along several parallel avenues: (i) the use of generating functions in the context of Hamiltonian mechanics to produce appropriate canonical transformations approximating the exact flow in each integration step (Feng and Qin 1987, Channell and Scovel 1990); (ii) the conditions that Runge–Kutta methods have to satisfy to be symplectic (Sanz-Serna 1988, Lasagni 1988, Suris 1988); (iii) the design of explicit symplectic methods of order four and higher for Hamiltonian systems that can be split into two pieces which can be solved exactly when considered as

[2] *Philosophiae Naturalis Principia Mathematica*, Book 1, Section 2, Proposition 1.

[3] See also Skeel and Cieśliński (2020) for the context of the work, and the preprint itself, typeset in LaTeX.

independent systems (Neri 1988, Forest 1992), with the help of the Lie formalism. This approach was further elaborated in Yoshida (1990). Working in the context of the hybrid Monte Carlo algorithm for dynamical fermions, a splitting method of order four was also independently proposed around the same time in Campostrini and Rossi (1990). In parallel developments, what is now called the Suzuki–Yoshida composition technique for increasing the order of numerical integrators appeared in Creutz and Gocksch (1989) for Monte Carlo simulations and in Suzuki (1990) and Yoshida (1990).

We should also mention the papers by de Raedt and de Raedt (1983) and Takahashi and Imada (1984), who pioneered the use of double commutators to get approximations of higher order than those obtained by the Störmer–Verlet method in path-integral Monte Carlo simulations: in the first case by constructing a fourth-order splitting scheme, and in the second, a method that it is also of order four by conjugation. In fact, scheme (1.47) is conjugate to the one proposed in Takahashi and Imada (1984). Ruth (1983) also presents a third-order method using double commutators.

That splitting and composition methods could be used to construct integrators for problems evolving in groups other than the symplectic group was emphasized in Forest and Ruth (1990) and further developed in Feng (1992), with the aim of constructing schemes able to preserve different structures.

The 1990s saw a dramatic increase in the interest and applications of splitting integrators in several fields, often with spectacular results. We should mention in particular those achieved in Wisdom and Holman (1991), revealing the existence of chaotic phenomena in the Solar System by numerically integrating the planetary equations of motion over very large time intervals.

The state of the art of splitting methods in the context of geometric numerical integration was masterfully summarized in a review paper by McLachlan and Quispel (2002), which has greatly influenced subsequent investigation in the field, as testified by its growing number of citations over the years, in many different areas.

Among the huge number of published works on splitting methods, the following surveys are worth highlighting.

- The monograph by Yanenko (1971) (an English translation of the Russian edition published in 1967) was perhaps the first to be devoted to the method of splitting (or *method of fractional steps*) for solving 'complicated problems of mathematical physics in several variables'. Those include the numerical treatment of parabolic and hyperbolic equations, as well as boundary value problems for the Laplace and Poisson equations, and several applications in elasticity theory and hydrodynamics. It is based on the early contributions of Peaceman, Rachford, Douglas and several authors of the Soviet school (Dyakonov, Marchuk, Samarskii, Yanenko and others).

- The exhaustive review article by Marchuk (1990), included in Volume I of the *Handbook of Numerical Analysis* (Ciarlet and Lions 1990), can be seen as an update of the previous work, with a systematic study of operator splitting and alternating direction methods for solving linear and nonlinear partial differential equations. It includes convergence analyses and new applications to problems in hydrodynamics, meteorology and oceanography.
- The review paper by McLachlan and Quispel (2002) mainly focused on the application of splitting methods as geometric numerical integrators for various classes of ordinary differential equations. In that context, they carried out a classification of ODEs and their integration methods into different categories, and also examined the question of how to decompose a given vector field into much simpler vector fields, as well as the composition of these elementary flows.
- Books and monographs dealing with geometric numerical integration and structure-preserving algorithms contain plenty of material on splitting methods. Among them, we can cite the influential work of Sanz-Serna and Calvo (1994), the canonical reference by Hairer *et al.* (2006), as well as Leimkuhler and Reich (2004), Leimkuhler and Matthews (2015) and Blanes and Casas (2016).
- The multi-author book by Glowinski *et al.* (2016*a*) constitutes an excellent illustration of the ample scope and wide range of applications that today's operator-splitting methods are able to deal with. These include the numerical solution of problems modelled by linear and nonlinear partial differential equations and inequalities, problems in information sciences and image processing, and large-scale optimization problems, among others.

1.10. Plan of the paper

In this paper we will focus on splitting methods applied to evolutionary problems, mostly described by ordinary differential equations. These can directly model the problem we are interested in, or they can result from evolutionary PDEs previously discretized in space. Particular attention will be addressed to problems possessing special properties, very often from a geometric origin, that are worth preserving via the numerical methods. In so doing, we will follow a strategy similar to that in Blanes, Casas and Murua (2008*b*), trying to avoid any duplication of the material already gathered in the classic references cited above, and including new results, schemes and applications which have appeared in the literature during the last few years.

In particular, no general rule is provided here on how to split the defining operator in (1.1). As mentioned earlier, this issue is further analysed in McLachlan and Quispel (2002), and in fact some of the open problems listed there are related to it. We have already seen in the examples provided in this section that several splittings of the same problem are possible, often leading to methods with very

different performances. Moreover, in certain cases, the original system has several geometric properties that are simultaneously preserved along the evolution, whereas different splittings preserve different properties, and it is generally difficult to find one splitting that preserves most of them.

With these considerations in mind, the rest of the paper is organized as follows. In Section 2 we first review the general composition technique and then provide a detailed analysis of the order conditions required by splitting and composition methods to achieve a given order of accuracy. There are some relevant problems, however, whose particular structure allows us to design specially adapted methods, and some of them are reviewed in Section 3, where we also show how to adapt existing splitting methods to non-autonomous problems.

In Section 4 we summarize some of the qualitative properties possessed by splitting methods in the context of geometric numerical integration of ordinary differential equations, with special attention to the idea of processing, whereas Section 5 is devoted to the treatment of highly oscillatory problems.

Splitting methods are particularly well adapted to deal with partial differential equations whose defining operator contains contributions coming from very different physical sources, so they have a long history in this area. Section 6 contains a brief survey, with special emphasis on Schrödinger equations and general parabolic evolution equations. The existence of negative coefficients in the methods, however, leads to an order barrier for parabolic equations, and Section 7 reviews splitting methods with complex coefficients as a possible way to overcome this order barrier.

In Section 8 we present an extended list of existing methods, classifying them into different families and giving the appropriate references. Their corresponding coefficients are also provided as supplementary material at the website www.gicas.uji.es/SplittingMethods.html. These methods are numerically tested on simple examples in the Appendix. Finally, some relevant applications of splitting methods in different fields are discussed in Section 9.

2. High-order splitting and composition methods

The Lie–Trotter and Strang splitting methods, despite their low order of accuracy, provide a fairly good description of the systems they are approximately solving. In fact, for many problems, including molecular dynamics applications and reaction–diffusion equations, Verlet splitting and Strang splitting are the most popular integrators, perhaps an illustration that, according to MacNamara and Strang (2016), 'it is a meta-theorem of numerical analysis that second-order methods often achieve the right balance between accuracy and complexity'. There are other areas, however, where a higher degree of precision is required, in addition to the preservation of qualitative properties. A classical example is the long-term numerical integration of the Solar System, both forwards (e.g. to analyse the existence of chaos; see Laskar 1989, Sussman and Wisdom 1992) and backwards in time (to study the

insolation quantities of the Earth; see Laskar *et al.* 2004). Thus, in this section, after reviewing a general technique to get high-order integrators by composing low-order ones, we analyse from different perspectives the order conditions that have to be satisfied by a splitting method to achieve a given order. This analysis allows us to provide complementary information about the integrators: e.g. number of order conditions, explicit expressions, and remainders in the asymptotic expansions.

2.1. *Raising the order by composition*

2.1.1. *Composition of Strang maps*

Starting from the Strang splitting $S_h^{[2]} = \varphi_{h/2}^{[1]} \circ \varphi_h^{[2]} \circ \varphi_{h/2}^{[1]}$, the composition

$$\psi_h = S_{\gamma_s h}^{[2]} \circ S_{\gamma_{s-1} h}^{[2]} \circ \cdots \circ S_{\gamma_1 h}^{[2]} \tag{2.1}$$

is at least of order three if

$$\sum_{j=1}^{s} \gamma_j = 1 \quad \text{and} \quad \sum_{j=1}^{s} \gamma_j^3 = 0. \tag{2.2}$$

The smallest value of s for which equations (2.2) admit real solutions is $s = 3$. In that case, by imposing the symmetry $\gamma_1 = \gamma_3$, we indeed get a method of order four, sometimes called the *triple jump*:

$$S_h^{[4]} = S_{\gamma_3 h}^{[2]} \circ S_{\gamma_2 h}^{[2]} \circ S_{\gamma_1 h}^{[2]}, \quad \text{with} \quad \gamma_1 = \gamma_3 = \frac{1}{2 - 2^{1/3}}, \quad \gamma_2 = 1 - 2\gamma_1. \tag{2.3}$$

In general, the recursion

$$S_h^{[2k]} = S_{\gamma_1 h}^{[2k-2]} \circ S_{(1-2\gamma_1)h}^{[2k-2]} \circ S_{\gamma_1 h}^{[2k-2]}, \quad \text{with} \quad \gamma_1 = \frac{1}{2 - 2^{1/(2k-1)}} \tag{2.4}$$

can be used to get methods of arbitrarily high order $2k$ ($k \geq 2$) (Creutz and Gocksch 1989) starting from the Strang map $S_h^{[2]}$ (notice that such methods can be written in the form (2.1)). The price to be paid is the existence of large positive and negative coefficients γ_j and the great number of elementary flows in (2.4) for high orders. The alternative formed by the five maps composition (*quintuple jump*)

$$S_h^{[2k]} = S_{\gamma_1 h}^{[2k-2]} \circ S_{\gamma_1 h}^{[2k-2]} \circ S_{(1-4\gamma_1)h}^{[2k-2]} \circ S_{\gamma_1 h}^{[2k-2]} \circ S_{\gamma_1 h}^{[2k-2]}, \quad \gamma_1 = \frac{1}{4 - 4^{1/(2k-1)}} \tag{2.5}$$

also gives methods $S_h^{[2k]}$ of order $2k$ of the form (2.1) with relatively smaller coefficients γ_j but even larger numbers of elementary flows.

In general, other choices for the coefficients γ_j in (2.1) are more appropriate if we are interested in achieving orders ≥ 6 with a lower number of elementary flows and relatively small coefficients.

Condition (2.2), using the approach based on linear differential operators discussed in Section 1.2, can be derived as follows: the Lie transformation associated

with the Strang map $S_h^{[2]}$, (1.16), can be written as $\mathrm{e}^{h/2F_1}\,\mathrm{e}^{hF_2}\,\mathrm{e}^{h/2F_1} = \mathrm{e}^{Y(h)}$, where

$$\begin{aligned} Y(h) &= \sum_{n=1}^{\infty} h^{2n-1} Y_{2n-1} \\ &:= \log(\mathrm{e}^{h/2F_1}\,\mathrm{e}^{hF_2}\,\mathrm{e}^{h/2F_1}) \\ &= h(F_1+F_2) - \frac{h^3}{24}[F_1,[F_1,F_2]] - \frac{h^3}{12}[F_2,[F_1,F_2]] + \cdots, \end{aligned}$$

that is, $Y_1 = F_1 + F_2$, $Y_3 = -\frac{1}{24}[F_1,[F_1,F_2]] - \frac{1}{12}[F_2,[F_1,F_2]]$, and for each $n > 2$, Y_{2n-1} is a certain linear combination of $(2n-1)$-fold commutators of F_1 and F_2. In consequence, the Lie transformation $\Psi(h)$ of (2.1) formally satisfies

$$\Psi(h) = \mathrm{e}^{Y(\gamma_1 h)} \cdots \mathrm{e}^{Y(\gamma_s h)}, \tag{2.6}$$

so that $g(\psi_h(x)) = (\Psi(h)g)(x)$ for any $x \in \mathbb{R}^D$ and any smooth function $g\colon \mathbb{R}^D \to \mathbb{R}$. It is straightforward to check that

$$\begin{aligned} \Psi(h) &= \mathrm{e}^{\gamma_1 hY_1 + \gamma_1^3 h^3 Y_3 + \cdots} \cdots \mathrm{e}^{\gamma_s hY_1 + \gamma_s^3 h^3 Y_3 + \cdots} \\ &= \mathrm{e}^{h(\sum_{j=1}^s \gamma_j)(F_1+F_2)} + h^3\left(\sum_{j=1}^s \gamma_j^3\right) Y_3 + O(h^4). \end{aligned} \tag{2.7}$$

This shows that the composition (2.1) of Strang maps is of order at least three if condition (2.2) holds.

In fact, (2.7) is also true if in (2.1) the Strang map $S_h^{[2]}$ is replaced by any second-order time-symmetric map. Furthermore, the triple jump recursion (2.4) (resp. the quintuple jump recursion (2.5)) also gives rise to $2k$th-order maps $S_h^{[2k]}$ starting from an arbitrary time-symmetric second-order map $S_h^{[2]}$. Indeed, this is a consequence of the following four statements.

(1) Given an arbitrary near-identity map $\chi_h\colon \mathbb{R}^D \to \mathbb{R}^D$ (i.e. $\chi_h(x) = x + O(h)$ as $h \to 0$), there exists a series

$$Y(h) = \sum_{n\geq 1} h^n Y_n$$

of (first-order) differential operators acting on smooth functions such that, formally, $g(\chi_h(x)) = (\mathrm{e}^{Y(h)}g)(x)$ for each $x \in \mathbb{R}^D$ and $g \in C^\infty(\mathbb{R}^D, \mathbb{R})$. Moreover, χ_h is an rth-order integrator for the ODE system $x' = f_1(x) + f_2(x)$ if and only if

$$Y_1 = F_1 + F_2, \quad Y_n = 0 \quad \text{for } 2 \leq n \leq r. \tag{2.8}$$

This statement can be proved as follows. Given a basic integrator $\chi_h\colon \mathbb{R}^D \to \mathbb{R}^D$, consider the linear differential operators X_n $(n \geq 1)$ acting on smooth

functions $g \in C^\infty(\mathbb{R}^D, \mathbb{R})$ as

$$X_n g(y) = \frac{1}{n!} \frac{\mathrm{d}^n}{\mathrm{d}h^n}\bigg|_{h=0} g(\chi_h(y)), \quad y \in \mathbb{R}^D, \tag{2.9}$$

so that formally $g(\chi_h(x)) = (X(h)g)(x)$, where

$$X(h) = I + \sum_{n\geq 1} h^n X_n, \tag{2.10}$$

and I denotes the identity operator. Each X_n is an nth-order differential operator. Thus, the integrator χ_h is of order r if

$$X_n = \frac{1}{n!}(F_1 + F_2)^n, \quad 1 \leq n \leq r.$$

Now consider the series of differential operators

$$Y(h) = \sum_{n\geq 1} h^n Y_n := \log(X(h)) = \sum_{m\geq 1} \frac{(-1)^{m+1}}{m}(hX_1 + h^2 X_2 + \cdots)^m,$$

that is,

$$Y_n = \sum_{m\geq 1}^{n} \frac{(-1)^{m+1}}{m} \sum_{j_1+\cdots+j_m=n} X_{j_1} \cdots X_{j_m},$$

so that $X(h) = \exp(Y(h))$, and formally, $g(\chi_h(x)) = (\exp(Y(h))g(x)$. It can be shown that each Y_n is a first-order differential operator. Clearly, the basic integrator is of order r if (2.8) holds.

(2) The map χ_h is time-symmetric if and only if $Y(h) = hY_1 + h^3 Y_3 + \cdots$, which implies that time-symmetric methods *are necessarily of even order*. Indeed, for the adjoint integrator $\chi_h^* = \chi_{-h}^{-1}$, we obviously get $g(\chi_h^*(x)) = \mathrm{e}^{-Y(-h)} g(x)$. Hence χ_h is time-symmetric if and only if $-Y(-h) \equiv Y(h)$.

(3) If $S_h^{[2k-2]}$ is a time-symmetric integrator of order at least $2k-2$, then the composition

$$\psi_h = S_{\gamma_s h}^{[2k-2]} \circ S_{\gamma_{s-1} h}^{[2k-2]} \circ \cdots \circ S_{\gamma_1 h}^{[2k-2]} \tag{2.11}$$

is of order at least $2k-1$ if

$$\sum_{j=1}^{s} \gamma_j = 1, \quad \sum_{j=1}^{s} \gamma_j^{2k-1} = 0. \tag{2.12}$$

In fact, if $\mathrm{e}^{h(F_1+F_2)+h^{2k-1}Y_{2k-1}+\cdots}$ is the Lie transform of $S_h^{[2k-2]}$, then $g(\psi_h(x)) = (\Psi(h)g)(x)$ for any $x \in \mathbb{R}^D$ and any smooth function $g\colon \mathbb{R}^D \to \mathbb{R}$, where

$$\begin{aligned}\Psi(h) &= \mathrm{e}^{\gamma_1 h(F_1+F_2)+\gamma_1^{2k-1}h^{2k-1}Y_{2k-1}+O(h^{2k})} \cdots \mathrm{e}^{\gamma_s h(F_1+F_2)+\gamma_s^{2k-1}h^{2k-1}Y_{2k-1}+O(h^{2k})} \\ &= \mathrm{e}^{(\sum_{j=1}^{s}\gamma_j)(F_1+F_2)} + h^{2k-1}\left(\sum_{j=1}^{s}\gamma_j^{2k-1}\right) Y_{2k-1} + O(h^{2k}).\end{aligned}$$

(4) If $S_h^{[2k-2]}$ is a time-symmetric integrator and the sequence $(\gamma_1, \dots, \gamma_s)$ is palindromic, in the sense that for all j

$$\gamma_{s-j+1} = \gamma_j,$$

then clearly the composition (2.11) is time-symmetric, and hence of even order.

2.1.2. Composition of Lie–Trotter maps

We could consider an analogous composition to (2.1), but this time with the Lie–Trotter scheme $\chi_h = \varphi_h^{[1]} \circ \varphi_h^{[2]}$ as the basic method. In that case, the Lie transform of χ_h is of the form $\mathrm{e}^{Y(h)}$ with

$$Y(h) = h(F_1 + F_2) + h^2 Y_2 + h^3 Y_3 + \cdots, \tag{2.13}$$

so that the Lie transform $\Psi(h)$ associated to the composition $\chi_{\gamma_s h} \circ \cdots \circ \chi_{\gamma_2 h} \circ \chi_{\gamma_1 h}$ is

$$\begin{aligned}\Psi(h) &= \mathrm{e}^{Y(\gamma_1 h)} \cdots \mathrm{e}^{Y(\gamma_s h)} \\ &= \mathrm{e}^{(\sum_{j=1}^s \gamma_j)(F_1+F_2)} + h^2 \left(\sum_{j=1}^s \gamma_j^2 \right) Y_2 + O(h^3).\end{aligned}$$

This shows that such a scheme is of order two if

$$\sum_{j=1}^s \gamma_j = 1, \quad \sum_{j=1}^s \gamma_j^2 = 0. \tag{2.14}$$

Obviously, such a system of equations does not admit real solutions.

The situation is different, however, if we compose χ_h with its adjoint $\chi_h^* = \varphi_h^{[2]} \circ \varphi_h^{[1]}$, that is,

$$\psi_h = \chi_{\alpha_{2s} h} \circ \chi^*_{\alpha_{2s-1} h} \circ \cdots \circ \chi_{\alpha_2 h} \circ \chi^*_{\alpha_1 h}. \tag{2.15}$$

If $\mathrm{e}^{Y(h)}$, with $Y(h)$ given by (2.13), is the Lie transform of χ_h, then the Lie transform of its adjoint χ_h^* is $\mathrm{e}^{-Y(-h)}$ and the Lie transform $\Psi(h)$ of (2.15) satisfies

$$\begin{aligned}\Psi(h) &= \mathrm{e}^{-Y(-\alpha_1 h)} \mathrm{e}^{Y(\alpha_2 h)} \cdots \mathrm{e}^{-Y(-\alpha_{2s-1} h)} \mathrm{e}^{Y(\alpha_{2s} h)} \\ &= \mathrm{e}^{(\sum_{j=1}^{2s} \alpha_j)(F_1+F_2)} + h^2 \left(\sum_{j=1}^s \left(\alpha_{2j}^2 - \alpha_{2j-1}^2 \right) \right) Y_2 + O(h^3),\end{aligned} \tag{2.16}$$

whence composition (2.15) is at least of order two if the coefficients satisfy

$$\sum_{j=1}^{2s} \alpha_j = 1, \quad \sum_{j=1}^s \left(\alpha_{2j}^2 - \alpha_{2j-1}^2 \right) = 0.$$

The argument above also holds when χ_h is a first-order integrator other than Lie–Trotter.

As mentioned earlier, the simplest situation corresponds to $s = 1$, in which case we recover the Strang splitting, $S_h^{[2]} = \chi_{h/2} \circ \chi_{h/2}^*$. In fact, the general scheme (2.15) can be rewritten as the splitting method

$$\psi_h = \varphi_{a_{s+1}h}^{[1]} \circ \varphi_{b_s h}^{[2]} \circ \varphi_{a_s h}^{[1]} \circ \cdots \circ \varphi_{a_2 h}^{[1]} \circ \varphi_{b_1 h}^{[2]} \circ \varphi_{a_1 h}^{[1]}, \tag{2.17}$$

where $a_1 = \alpha_1$, and for $j = 1, \ldots, s$,

$$a_{j+1} = \alpha_{2j} + \alpha_{2j+1}, \quad b_j = \alpha_{2j-1} + \alpha_{2j} \tag{2.18}$$

(with $\alpha_{2s+1} = 0$). Conversely, any integrator of the form (2.17) satisfying the condition $\sum_{j=1}^{s+1} a_j = \sum_{j=1}^{s} b_j$ can be expressed in the form (2.15), as shown in McLachlan (1995*b*).

Clearly, any splitting scheme (2.17) with a palindromic sequence of coefficients, that is, satisfying

$$a_{s-j+2} = a_j, \quad b_{s-j+1} = b_j \quad \text{for all } j,$$

is time-symmetric, and thus of even order. Written in the composition format (2.15), it is time-symmetric if

$$\alpha_{2s-j+1} = \alpha_j.$$

2.2. *Order conditions I: splitting schemes with the BCH formula*

In the analysis of the order conditions for splitting methods, and without loss of generality, we will consider the linear case (1.4), so that the treatment is essentially based on matrices, or more generally, on linear operators. Thus, the integrator (2.17) corresponds in this setting to the product of exponentials

$$\Psi(h) = \mathrm{e}^{a_{s+1}hF_1}\, \mathrm{e}^{b_s hF_2}\, \mathrm{e}^{a_s hF_1} \cdots \mathrm{e}^{a_2 hF_1}\, \mathrm{e}^{b_1 hF_2}\, \mathrm{e}^{a_1 hF_1}, \tag{2.19}$$

intended to approximate $\mathrm{e}^{h(F_1+F_2)}$.

In the case of systems of ODEs $x' = f_1(x) + f_2(x)$ with (not necessarily linear) vector fields f_1 and f_2, one may consider the Lie operators F_j of f_j, and compare the Lie transformation $\mathrm{e}^{h(F_1+F_2)}$ with the operator

$$\Psi(h) = \mathrm{e}^{a_1 hF_1}\, \mathrm{e}^{b_1 hF_2}\, \mathrm{e}^{a_2 hF_1} \cdots \mathrm{e}^{a_s hF_1}\, \mathrm{e}^{b_s hF_2}\, \mathrm{e}^{a_{s+1} hF_1} \tag{2.20}$$

associated with the map ψ_h (in the sense that $g(\psi_h(x)) = \Psi(h)g(x)$ for arbitrary smooth $g \in C^\infty(\mathbb{R}^D, \mathbb{R})$). Thus all the formulas derived below for the linear case can be applied to the Lie transformation (2.20) associated with the map (2.17) just by reversing the order of the sequence $(a_1, b_1, a_2, \ldots, a_s, b_s, a_{s+1})$ of coefficients of the splitting scheme. Alternatively, we may reverse the order of the products of the operators in all the expressions involved.

Equivalent results could also be obtained for the ODE case using the concept of *word series* and related techniques introduced in Murua and Sanz-Serna (2017). The Lie transform approach is more general, however, as it can be directly applied

to the case of splitting methods for differential equations on manifolds, as discussed at the end of Section 4.1.

Generally speaking, the order conditions for a method of order r are systems of polynomial equations in the coefficients obtained by requiring that the Taylor expansions in the step size h of both the exact and numerical solution agree up to terms in h^r. A standard approach to obtain the order conditions of scheme (2.19) consists in formally using the Baker–Campbell–Hausdorff (BCH) formula to express $\Psi(h)$ as one exponential of a series of operators in powers of h, and finally to compare this series with $\mathrm{e}^{h(F_1+F_2)}$. In this way, we get

$$\begin{aligned}\log(\Psi(h)) = h(w_1F_1 + w_2F_2) + h^2 w_{12}F_{12} + h^3(w_{122}F_{122} + w_{112}F_{112}) \\ + h^4(w_{1222}F_{1222} + w_{1122}F_{1122} + w_{1112}F_{1112}) + O(h^5),\end{aligned} \tag{2.21}$$

where

$$\begin{aligned} F_{12} &= [F_1, F_2], & F_{122} &= [F_{12}, F_2], & F_{112} &= [F_1, F_{12}], \\ F_{1222} &= [F_{122}, F_2], & F_{1122} &= [F_1, F_{122}], & F_{1112} &= [F_1, F_{112}], \end{aligned}$$

and $w_1, w_2, w_{12}, w_{122}, \ldots$ are polynomials of homogeneous degree in the parameters a_j, b_j. In particular,

$$\begin{gathered} w_1 = \sum_{i=1}^{s+1} a_i, \quad w_2 = \sum_{i=1}^{s} b_i, \quad w_{12} = \frac{1}{2}w_1w_2 - \sum_{1\le i<j\le s+1} b_i a_j, \\ w_{122} = \frac{1}{12}w_1w_2^2 - \frac{1}{2}\sum_{1\le i\le j<k\le s} b_i a_j b_k, \\ w_{112} = \frac{1}{12}w_1^2w_2 - \frac{1}{2}\sum_{1\le i<j\le k\le s+1} a_i b_j a_k, \end{gathered}$$

with $b_{s+1} = 0$. From (2.21), it is clear that a characterization of the order of the splitting scheme (2.17) is obtained by requiring the consistency conditions $w_1 = w_2 = 1$, that is,

$$\sum_{j=1}^{s+1} a_j = \sum_{j=1}^{s} b_j = 1 \tag{2.22}$$

(ensuring that the scheme (2.17) is at least of order one) and $w_{12} = w_{122} = w_{112} = \cdots = 0$ up to the required order. The set of order conditions thus obtained will be independent in the general case if the operators F_1, F_2, F_{12}, F_{122}, $F_{112}, \ldots$ form a basis of the free Lie algebra in the alphabet $\{1, 2\}$. In (2.21) we considered the so-called *Lyndon basis*, associated to the set of Lyndon words in the alphabet $\{1, 2\}$:

$$\{1, 2, 12, 122, 112, 1222, 1122, 1112, \ldots\}. \tag{2.23}$$

Table 2.1. Number of independent order conditions for general splitting methods, c_n, and for RKN-type splitting methods, d_n. The number m_n corresponds to the number of order conditions for compositions of a second-order time-symmetric method.

Order n	1	2	3	4	5	6	7	8	9	10	11
c_n	2	1	2	3	6	9	18	30	56	99	186
d_n	2	1	2	2	4	5	10	14	25	39	69
m_n	1	0	1	1	2	2	4	5	8	11	17

They are defined as follows (Reutenauer 1993): a word $\ell_1 \cdots \ell_m$ is a Lyndon word if $(\ell_1 \cdots \ell_k) \prec (\ell_{k+1} \cdots \ell_m)$ for each $1 \leq k < m$, where $\prec$ is the lexicographical order (i.e. the order used when ordering words in the dictionary) on the set of words in the alphabet $\{1, 2\}$. For instance, 112 is a Lyndon word, while neither 211 nor 121 are, as $2 \not\prec 11$ and as $12 \not\prec 1$, respectively.

The element of the basis associated to a Lyndon word $\ell_1 \cdots \ell_m$ with $m \geq 2$ is given as $F_{\ell_1 \cdots \ell_m} = [F_{\ell_1 \cdots \ell_n}, F_{\ell_{n+1} \cdots \ell_m}]$, where n is the smallest number such that both $\ell_1 \cdots \ell_n$ and $\ell_{n+1} \cdots \ell_m$ are themselves Lyndon words.

Casas and Murua (2009) have presented an efficient algorithm (based on the results in Murua 2006) for the BCH formula and related calculations in the Lyndon basis (and some other basis) that allows us to obtain (2.21) up to terms of arbitrarily high degree.

Of course, if another basis of the free Lie algebra in the alphabet $\{1, 2\}$ is used to expand $\log(\Psi(h))$ in (2.21), a different characterization of the order conditions will be obtained, with a different set of polynomial functions on $(a_1, b_1, \ldots, a_s, b_s, a_{s+1})$. In any case, the number of such independent conditions arising at each order n can be obtained just by determining the dimension of $\mathcal{L}_n(F_1, F_2)$, the linear span of all commutators containing n operators F_1, F_2. This number, denoted c_n, is given in Table 2.1; see Munthe-Kaas and Owren (1999) and McLachlan and Quispel (2002).

2.3. *Order conditions II: splitting schemes with Lyndon words*

Whereas the previous characterization of the order of the splitting scheme (2.19) allows us to easily get the number of order conditions, obtaining explicit expressions for the polynomials $w_{\ell_1 \cdots \ell_n}$ is much more difficult when the considered order increases. The following alternative characterization, based on the direct comparison of the power series expansions of $\mathrm{e}^{h\,(F_1+F_2)}$ and (2.19), tries to ameliorate this difficulty.

2.3.1. Basic expansions and necessary order conditions

Both $\mathrm{e}^{h\,(F_1+F_2)}$ and (2.19) admit an expansion in series indexed by the set

$$\mathcal{W} = \{1, 2, 11, 12, 21, 22, 111, 112, 121, 211, 122, \ldots\}$$

of words in the alphabet $\{1, 2\}$. More precisely, $\mathrm{e}^{h\,(F_1+F_2)}$ can be expanded as

$$I+h\alpha_1 F_1+h\alpha_2 F_2+h^2\alpha_{11} F_1F_1+h^2\alpha_{12} F_1F_2+h^2\alpha_{21} F_2F_1+h^2\alpha_{22} F_2F_2+\cdots, \quad (2.24)$$

with $\alpha_{\ell_1\cdots\ell_n} = 1/n!$. As for (2.19), it can be expanded, for arbitrary s, as (2.24), where for each word $\ell_1\cdots\ell_n \in \mathcal{W}$ with n letters, the corresponding coefficient

$$\alpha_{\ell_1\cdots\ell_n} = u_{\ell_1\cdots\ell_n}(a_1, b_1, \ldots, a_s, b_s, a_{s+1})$$

is a homogeneous polynomial of degree n in the variables $a_1, b_1, \ldots, a_s, b_s, a_{s+1}$.

It is straightforward to check that such polynomials satisfy the following relations, which allows us to compute them recursively.

- If $\ell_1 = \cdots = \ell_j = 1$ and $\ell_{j+1} \neq 1$ with $j \geq 1$,

$$u_{\ell_1\cdots\ell_n}(a_1, b_1, \ldots, a_s, b_s, a_{s+1}) = \sum_{k=0}^{j} u_{\ell_{k+1}\cdots\ell_n}(a_1, b_1, \ldots, a_s, b_s, 0)\,\frac{a_{s+1}^k}{k!},$$

$$u_{\ell_1\cdots\ell_n}(a_1, b_1, \ldots, a_s, b_s, 0) = u_{\ell_1\cdots\ell_n}(a_1, b_1, \ldots, a_s).$$

- If $\ell_1 = \cdots = \ell_j = 2$ and $\ell_{j+1} \neq 2$ with $j \geq 1$,

$$u_{\ell_1\cdots\ell_n}(a_1, b_1, \ldots, a_s, b_s, a_{s+1}) = \sum_{k=0}^{j} u_{\ell_{k+1}\cdots\ell_n}(a_1, b_1, \ldots, a_s)\,\frac{b_s^k}{k!}.$$

- If $\ell_n = \cdots = \ell_1 = 1$,

$$u_{\ell_1\cdots\ell_n}(a_1) = \frac{a_1^n}{n!}.$$

- If $\ell_1 \neq 1$,

$$u_{\ell_1\cdots\ell_n}(a_1) = 0.$$

In this way (2.19) is at least of order r if and only if the conditions

$$u_{\ell_1\cdots\ell_n}(a_1, b_1, \ldots, a_s, b_s, a_{s+1}) = \frac{1}{n!} \quad (2.25)$$

hold for each word $\ell_1\cdots\ell_n$ with $n \leq r$ letters in the alphabet $\{1, 2\}$. For illustration, in Table 2.2 we explicitly give these conditions corresponding to words with up to two letters. Notice that (2.25) for the single-letter words 1 and 2 coincide with the consistency conditions (2.22).

However, such order conditions are not all independent. For instance, from Table 2.2, we can check that

$$u_1^2 = 2u_{11}, \quad u_2^2 = 2u_{22}, \quad u_1u_2 = u_{12} + u_{21}.$$

Table 2.2. Conditions (2.25) corresponding to words with up to two indices.

Word	Condition	Word	Condition
1	$\sum_{j=1}^{s+1} a_j = 1$	2	$\sum_{j=1}^{s} b_j = 1$
11	$\frac{1}{2}\sum_{j=1}^{s+1} \frac{a_j^2}{2} + \sum_{1\le i<j\le s+1} a_i a_j = \frac{1}{2}$	12	$\sum_{1\le i<j\le s+1} b_i a_j = \frac{1}{2}$
21	$\sum_{1\le j\le i\le s} b_i a_j = \frac{1}{2}$	22	$\frac{1}{2}\sum_{j=1}^{s} \frac{b_j^2}{2} + \sum_{1\le j<i\le s} b_j b_i = \frac{1}{2}$

For a consistent method, $u_1 = u_2 = 1$, hence $u_{11} = \frac{1}{2}$, $u_{22} = \frac{1}{2}$, and

$$u_{12} - \frac{1}{2} = \frac{1}{2} - u_{21},$$

which implies that if $u_{12} = \frac{1}{2}$, then automatically $u_{21} = \frac{1}{2}$.

A complete characterization of the relations among the order conditions (2.25) will be obtained in Section 2.3.3 below. As a previous step, we obtain integral representations of both $\mathrm{e}^{h(F_1+F_2)}$ and (2.19) , which in addition give useful expressions for the remainders of their truncated series expansions.

2.3.2. *Integral representation and remainders*

Consider the solution $Y(\tau)$ of the initial value problem

$$\frac{\mathrm{d}}{\mathrm{d}\tau} Y(\tau) = hA(\tau)Y(\tau), \quad Y(0) = I, \tag{2.26}$$

with $A(\tau) = d_1(\tau)F_1 + d_2(\tau)F_2$ and

$$(d_1(\tau), d_2(\tau)) = \begin{cases} (a_i, 0) & \text{if } \tau \in [2i-2, 2i-1],\ i \in \{1, \dots, s+1\}, \\ (0, b_i) & \text{if } \tau \in [2i-1, 2i],\ i \in \{1, \dots, s\}, \\ (0, 0) & \text{if } \tau > 2s+1. \end{cases} \tag{2.27}$$

It is straightforward to check that

$$Y(\tau) = \begin{cases} \mathrm{e}^{(\tau-(2i-2))ha_{i+1}F_1}\, \mathrm{e}^{b_i hF_2}\, \mathrm{e}^{a_i hF_1} \cdots \mathrm{e}^{a_2 hF_1}\, \mathrm{e}^{b_1 hF_2}\, \mathrm{e}^{a_1 hF_1}, & \\ & \tau \in [2i-2, 2i-1], \\ \mathrm{e}^{(\tau-(2i-1))hb_i F_2}\, \mathrm{e}^{a_i hF_1} \cdots \mathrm{e}^{a_2 hF_1}\, \mathrm{e}^{b_1 hF_2}\, \mathrm{e}^{a_1 hF_1}, & \tau \in [2i-1, 2i]. \end{cases}$$

In particular, $\Psi(h) = Y(2s+1)$. The solution $Y(\tau)$ of (2.26) satisfies

$$Y(\tau) = I + h\int_0^\tau A(\tau_1)Y(\tau_1)\,\mathrm{d}\tau_1. \tag{2.28}$$

From that, we obtain

$$Y(\tau) = I + h\int_0^\tau A(\tau_1)\,\mathrm{d}\tau_1 + h^2\int_0^\tau\int_0^{\tau_1} A(\tau_1)A(\tau_2)Y(\tau_2)\,\mathrm{d}\tau_2\,\mathrm{d}\tau_1,$$

and more generally

$$Y(\tau) = I+\sum_{n=1}^m h^n\int_0^\tau\int_0^{\tau_1}\cdots\int_0^{\tau_{n-1}} A(\tau_1)\cdots A(\tau_n)\,\mathrm{d}\tau_n\cdots\mathrm{d}\tau_1+h^{m+1}\mathcal{R}_{m+1}(\tau,h), \tag{2.29}$$

where for each n the remainder $\mathcal{R}_n(\tau,h)$ satisfies

$$\mathcal{R}_n(\tau,h) = \int_0^\tau\int_0^{\tau_1}\cdots\int_0^{\tau_{n-1}} A(\tau_1)\cdots A(\tau_n)Y(\tau_n)\,\mathrm{d}\tau_n\cdots\mathrm{d}\tau_1. \tag{2.30}$$

By substituting $A(\tau_j) = d_1(\tau_j)F_1 + d_2(\tau_j)F_2$, then expanding all the products and taking constant linear operators out from integral signs, we obtain

$$Y(\tau) = I + \sum_{n=1}^m h^n \sum_{\ell_1,\dots,\ell_n\in\{1,2\}} \alpha_{\ell_1\cdots\ell_n}(\tau)\,F_{\ell_1}\cdots F_{\ell_n} + h^{m+1}\mathcal{R}_{m+1}(\tau,h),$$

where

$$\alpha_{\ell_1\cdots\ell_n}(\tau) = \int_0^\tau\int_0^{\tau_1}\cdots\int_0^{\tau_{n-1}} d_{\ell_1}(\tau_1)\cdots d_{\ell_n}(\tau_n)\,\mathrm{d}\tau_n\cdots\mathrm{d}\tau_1. \tag{2.31}$$

In particular, we have

$$\begin{aligned}\Psi(h) = I + \sum_{n=1}^m h^n \sum_{\ell_1,\dots,\ell_n\in\{1,2\}} u_{\ell_1\cdots\ell_n}(a_1,b_1,\dots,a_s,b_s,a_{s+1})\,F_{\ell_1}\cdots F_{\ell_n}\\ + h^{m+1}\mathcal{R}_{m+1}(2s+1,h),\end{aligned}$$

where $u_{\ell_1\cdots\ell_n}(a_1,b_1,\dots,a_s,b_s,a_{s+1}) = \alpha_{\ell_1\cdots\ell_n}(2s+1)$, that is,

$$\begin{aligned}&u_{\ell_1\cdots\ell_n}(a_1,b_1,\dots,a_s,b_s,a_{s+1})\\ &\quad= \int_0^{2s+1}\int_0^{\tau_1}\cdots\int_0^{\tau_{n-1}} d_{\ell_1}(\tau_1)\cdots d_{\ell_n}(\tau_n)\,\mathrm{d}\tau_n\cdots\mathrm{d}\tau_1.\end{aligned}$$

We now consider (2.26) with $d_1(\tau) \equiv d_2(\tau) \equiv 1$. Clearly, in that case, $Y(\tau) = \mathrm{e}^{\tau h(F_1+F_2)}$, and

$$\begin{aligned}
&\mathrm{e}^{h(F_1+F_2)} \\
&\quad = I + \sum_{n=1}^{m} h^n \int_0^1 \int_0^{\tau_1} \cdots \int_0^{\tau_{n-1}} (F_1+F_2)^n \,\mathrm{d}\tau_n \cdots \mathrm{d}\tau_1 + h^{m+1}\overline{\mathcal{R}}_{m+1}(1,h) \\
&\quad = I + \sum_{n=1}^{m} h^n \sum_{\ell_1,\ldots,\ell_n \in \{1,2\}} \frac{1}{n!} F_{\ell_1} \cdots F_{\ell_n} + h^{m+1}\overline{\mathcal{R}}_{m+1}(1,h),
\end{aligned}$$

where for each n, the remainder $\overline{\mathcal{R}}_n(\tau, h)$ is given by

$$\overline{\mathcal{R}}_n(\tau,h) = \int_0^{\tau} \int_0^{\tau_1} \cdots \int_0^{\tau_{n-1}} (F_1+F_2)^n \mathrm{e}^{\tau_n h(F_1+F_2)} \,\mathrm{d}\tau_n \cdots \mathrm{d}\tau_1.$$

We finally arrive at the following expression for the local error of the splitting scheme (2.26):

$$\begin{aligned}
&\Psi(h) - \mathrm{e}^{h(F_1+F_2)} \\
&\quad = \sum_{n=1}^{m} h^n \sum_{\ell_1,\ldots,\ell_n \in \{1,2\}} \left(u_{\ell_1\cdots\ell_n}(a_1,b_1,\ldots,a_s,b_s,a_{s+1}) - \frac{1}{n!} \right) F_{\ell_1} \cdots F_{\ell_n} \\
&\qquad + h^{m+1}(\mathcal{R}_{m+1}(2s+1,h) - \overline{\mathcal{R}}_{m+1}(1,h)).
\end{aligned}$$

Hence, if the scheme is of order r (i.e. (2.25) holds for each word $(\ell_1,\ldots,\ell_n)$ with $n \le r$ letters in the alphabet $\{1,2\}$), then

$$\Psi(h) - \mathrm{e}^{h(F_1+F_2)} = h^{n+1}(\mathcal{R}_{n+1}(2s+1,h) - \overline{\mathcal{R}}_n(1,h)).$$

2.3.3. Iterated integrals and shuffle relations

Iterated integrals of the form (2.31) were first considered and studied in Chen (1957). It is well known that the integration-by-parts formula gives (for an arbitrary integrable path $(d_1(\tau), d_2(\tau))$) the relations

$$\begin{aligned}
\alpha_{\ell_1}(\tau)\alpha_{\ell_2}(\tau) &= \alpha_{\ell_1\ell_2}(\tau) + \alpha_{\ell_2\ell_1}(\tau), \\
\alpha_{\ell_1}(\tau)\alpha_{\ell_2\ell_3}(\tau) &= \alpha_{\ell_1\ell_2\ell_3}(\tau) + \alpha_{\ell_2\ell_1\ell_3}(\tau) + \alpha_{\ell_2\ell_3\ell_1}(\tau),
\end{aligned}$$

and more generally,

$$\alpha_{\ell_1\cdots\ell_n}(\tau)\alpha_{\ell_{n+1}\cdots\ell_{n+m}}(\tau) = \sum_{\sigma \in \mathrm{Sh}(n,m)} \alpha_{\ell_{\sigma(1)}\cdots\ell_{\sigma(n+m)}}(\tau), \tag{2.32}$$

where $\mathrm{Sh}(n,m)$ is the set of the $(n+m)!/(n!m!)$ permutations σ of $(1,\ldots,n+m)$ that are obtained by interleaving $(1,\ldots,n)$ and $(n+1,\ldots,n+m)$ while preserving their respective ordering.

It will be useful to interpret the relations (2.32) in terms of the so-called shuffle product of words: the shuffle product ⧢ of two words $\ell_1\cdots\ell_n$ and $\ell_{n+1}\cdots\ell_{n+m}$ is

defined as the following formal sum of words:

$$\ell_1 \cdots \ell_n \sqcup\!\sqcup \ell_{n+1} \cdots \ell_{n+m} = \sum_{\sigma \in \mathrm{Sh}(n,m)} \ell_{\sigma(1)} \cdots \ell_{\sigma(n+m)}.$$

By extending $\alpha_w(\tau)$ linearly to the case where w is a linear combination of words, the relations (2.32) can be interpreted as

$$\alpha_w(\tau)\alpha_{w'}(\tau) = \alpha_{w \sqcup\!\sqcup w'}$$

for arbitrary words $w = \ell_1 \cdots \ell_n$ and $w' = \ell_{n+1} \cdots \ell_{n+m}$ in the alphabet $\mathcal{A} = \{1, 2\}$.

The shuffle product $\sqcup\!\sqcup$ defines a commutative algebra (the so-called shuffle algebra) over the vector space of formal linear combinations of words in the alphabet $\mathcal{A}$. The shuffle algebra is freely generated by the set of *Lyndon words* (Reutenauer 1993).

The fact that the coefficients of the series expansions of both $\mathrm{e}^{h(F_1+F_2)}$ and $\Psi(h)$ satisfy the shuffle relations, together with the fact that the set of Lyndon words freely generate the shuffle algebra, implies that a set of independent conditions for a consistent splitting scheme to attain order r can be obtained by considering (2.25) for each Lyndon word $(\ell_1, \ldots, \ell_n)$ of length $n \le r$.

2.4. *Order conditions III: splitting methods with Lyndon multi-indices*

Blanes *et al.* (2013*b*) have obtained yet another characterization of the order conditions in terms of explicitly given polynomial equations. We next describe this alternative formulation. To do that, we always assume that the consistency conditions (2.22) hold, so that the method is at least of first order. In this case, the polynomial equations are expressed in terms of the coefficients $b_1, \ldots, b_s$ and the coefficients $c_1, \ldots, c_s$ given by

$$c_i = \sum_{j=1}^{i} a_j, \quad i = 1, 2, \ldots, s. \tag{2.33}$$

We begin by rewriting (2.19) as

$$\begin{aligned}\Psi(h) &= \mathrm{e}^{a_{s+1}hF_1}\,\mathrm{e}^{b_s hF_2}\,\mathrm{e}^{a_s hF_1} \cdots \mathrm{e}^{b_1 hF_2}\,\mathrm{e}^{a_1 hF_1} \\ &= \mathrm{e}^{hF_1}\left(\prod_{j=s}^{1} \mathrm{e}^{-c_j hF_1}\,\mathrm{e}^{b_j hF_2}\,\mathrm{e}^{c_j hF_1}\right) \\ &= \mathrm{e}^{hF_1}\mathrm{e}^{b_s hC(c_s h)} \cdots \mathrm{e}^{b_1 hC(c_1 h)},\end{aligned}$$

where

$$C(h) = \mathrm{e}^{-hF_1}\,F_2\,\mathrm{e}^{hF_1} = \sum_{n=1}^{\infty} h^{n-1}C_n = C_1 + hC_2 + h^2C_3 + h^3C_4 + \cdots, \tag{2.34}$$

with $C_1 = F_2$, and

$$C_k = \frac{1}{(k-1)}[C_{k-1}, F_1] \quad \text{for } k > 1.$$

Now the order of the scheme (2.17) is established by comparing the expansion in powers of h of $\mathrm{e}^{-hF_1}\Psi(h)$ with that of $\mathrm{e}^{-hF_1}\mathrm{e}^{h(F_1+F_2)}$.

Clearly, $Y(\tau) := \mathrm{e}^{-h\tau F_1}\mathrm{e}^{\tau h(F_1+F_2)}$ is the solution of (2.26) with $A(\tau) = C(\tau h)$. Since the solution $Y(\tau)$ of (2.26) admits the representation (2.29) with remainder (2.30), and $\mathrm{e}^{-hF_1}\mathrm{e}^{h(F_1+F_2)} = Y(1)$, we conclude that

$$\begin{aligned}\mathrm{e}^{-hF_1}\mathrm{e}^{h(F_1+F_2)} = I + \sum_{k=1}^{n} h^k \int_0^1 \int_0^{\tau_k} \cdots \int_0^{\tau_2} C(\tau_k h)\cdots C(\tau_1 h)\,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_k \\ + h^{n+1}\mathcal{R}_{n+1}(1,h),\end{aligned}$$

where for each k,

$$\mathcal{R}_k(1,h) = \int_0^1 \int_0^{\tau_k} \cdots \int_0^{\tau_2} C(\tau_k h)\cdots C(\tau_1 h) Y(\tau_1)\,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_k.$$

By substitution of (2.34), we obtain that for each k,

$$\begin{aligned}&h^k \int_0^1 \int_0^{\tau_k} \cdots \int_0^{\tau_2} C(\tau_k h)\cdots C(\tau_1 h)\,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_k \\ &= \sum_{i_1,\ldots,i_k \geq 1} h^{i_1+\cdots+i_k}\left(\int_0^1 \int_0^{\tau_k} \cdots \int_0^{\tau_2} \tau_1^{i_1-1}\cdots\tau_k^{i_k-1}\,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_k\right) C_{i_k}\cdots C_{i_1} \\ &= \sum_{i_1,\ldots,i_k \geq 1} \frac{h^{i_1+\cdots+i_k}}{(i_1+\cdots+i_k)\cdots(i_1+i_2)i_1} C_{i_k}\cdots C_{i_1}.\end{aligned}$$

As for $\mathrm{e}^{-hF_1}\,\Psi(h)$,

$$\begin{aligned}&\mathrm{e}^{-hF_1}\,\Psi(h) \\ &= \prod_{j=s}^{1} \mathrm{e}^{b_j h C(c_j h)} = \prod_{j=s}^{1}\left(I + \sum_{k\geq 1} \frac{h^k b_j^k}{k!} C(c_j h)^k\right) \\ &= I + \sum_{k\geq 1} \frac{h^k}{k!}\sum_{j=1}^{s} b_j^k C(c_j h)^k + \sum_{k_1,k_2\geq 1} \frac{h^{k_1+k_2}}{k_1!k_2!} \sum_{1\leq j_1<j_2\leq s} b_{j_1}^{k_1} b_{j_2}^{k_2} C(c_{j_2}h)^{k_2} C(c_{j_1}h)^{k_1} \\ &\quad + \sum_{k_1,k_2,k_3\geq 1} \frac{h^{k_1+k_2+k_3}}{k_1!k_2!k_3!} \sum_{1\leq j_1<j_2<j_3\leq s} b_{j_1}^{k_1} b_{j_2}^{k_2} b_{j_3}^{k_3} C(c_{j_3}h)^{k_3} C(c_{j_2}h)^{k_2} C(c_{j_1}h)^{k_1} + \cdots \\ &= I + \sum_{k\geq 1} h^k \sum_{1\leq j_1\leq\cdots\leq j_k\leq s} \frac{b_{j_1}\cdots b_{j_k}}{\sigma(j_1,\ldots j_k)} C(c_{j_k}h)\cdots C(c_{j_1}h),\end{aligned} \tag{2.35}$$

where

$$\begin{aligned}\sigma(j_1,\dots,j_k) &= 1 && \text{if } j_1 < \cdots < j_k,\\ \sigma(j_1,\dots,j_k) &= \ell!\,\sigma(j_{\ell+1},\dots,j_k) && \text{if } j_1 = \cdots = j_\ell < j_{\ell+1} \le \cdots \le j_k.\end{aligned}$$

Since

$$h^k C(c_{j_k}h)\cdots C(c_{j_1}h) = \sum_{i_1,\dots,i_k\ge 1} h^{i_1+\cdots+i_k} c_{j_1}^{i_1-1}\cdots c_{j_k}^{i_k-1} C_{i_k}\cdots C_{i_1},$$

we arrive at

$$\mathrm{e}^{-hF_1}\Psi(h) = I + \sum_{k\ge 1}\sum_{i_1,\dots,i_k\ge 1} h^{i_1+\cdots+i_k} v_{i_1,\dots,i_k}(b_1,c_1,\dots,b_s,c_s)\, C_{i_k}\cdots C_{i_1}, \tag{2.36}$$

where

$$v_{i_1,\dots,i_k}(b_1,c_1,\dots,b_s,c_s) = \sum_{1\le j_1\le\cdots\le j_k\le s} \frac{b_{j_1}\cdots b_{j_k}}{\sigma(j_1,\dots,j_k)}\, c_{j_1}^{i_1-1}\cdots c_{j_k}^{i_k-1}. \tag{2.37}$$

In this way, a consistent splitting method is at least of order r if and only if

$$\sum_{1\le j_1\le\cdots\le j_k\le s} \frac{b_{j_1}\cdots b_{j_k}}{\sigma(j_1,\dots,j_k)}\, c_{j_1}^{i_1-1}\cdots c_{j_k}^{i_k-1} = \frac{1}{(i_1+\cdots+i_k)\cdots(i_1+i_2)i_1} \tag{2.38}$$

holds for each multi-index $(i_1,\dots,i_k)$ such that $i_1+\cdots+i_k \le r$. For illustration, in Table 2.3 we give explicit conditions (2.38) corresponding to multi-indices with up to three indices.

However, such order conditions are not independent. The situation is very similar to that of the previous subsection: instead of series indexed by the set of words in the alphabet $\{1,2\}$, now they are indexed by the set of words in the alphabet $\mathbb{N} = \{1,2,3,4,\dots\}$. To distinguish the words of both sets, we will keep referring to the words in the alphabet $\mathbb{N}$ as multi-indices, and will write them as $(i_1,\dots,i_k)$ instead of $i_1\cdots i_k$. Analogously to the previous subsection, the corresponding coefficients of the series expansions of both $\mathrm{e}^{-hF_1}\Psi(h)$ and $\mathrm{e}^{-hF_1}\mathrm{e}^{h(F_1+F_2)}$ satisfy the shuffle relations

$$v_{i_1,\dots,i_n} v_{i_{n+1},\dots,i_{n+m}} = \sum_{\sigma\in\mathrm{Sh}(n,m)} v_{i_{\sigma(1)},\dots,i_{\sigma(n+m)}}. \tag{2.39}$$

This can be seen by showing that such coefficients can be written in both cases as iterated integrals. From the discussion above, we already know that, for each multi-index $(i_1,\dots,i_k)$, the coefficients

$$\frac{1}{(i_1+\cdots+i_k)\cdots(i_1+i_2)i_1} = \int_0^1\int_0^{\tau_k}\cdots\int_0^{\tau_2} \tau_1^{i_1-1}\cdots\tau_k^{i_k-1}\,\mathrm{d}\tau_1\cdots\mathrm{d}\tau_k$$

of the series expansion of $\mathrm{e}^{-hF_1}\mathrm{e}^{h(F_1+F_2)}$ are indeed iterated integrals.

We next show that the coefficients (2.37) of the series expansion (2.36) can also be defined as iterated integrals.

Table 2.3. Conditions (2.38) for multi-indices with up to three indices.

Multi-index	Condition
(i)	$\displaystyle\sum_{j=1}^{s} b_j\, c_j^{i-1} = \frac{1}{i}$
(i_1, i_2)	$\displaystyle\frac{1}{2}\sum_{j=1}^{s} b_j^2 c_j^{i_1+i_2-2} + \sum_{1\le j_1<j_2\le s} b_{j_1} b_{j_2} c_{j_1}^{i_1-1} c_{j_2}^{i_2-1} = \frac{1}{(i_1+i_2)i_1}$
(i_1, i_2, i_3)	$\displaystyle\frac{1}{6}\sum_{j=1}^{s} b_j^3 c_j^{i_1+i_2+i_3-3} + \frac{1}{2}\sum_{1\le j_1<j_3\le s} b_{j_1}^2 b_{j_3} c_{j_1}^{i_1+i_2-2} c_{j_3}^{i_3-1} + \frac{1}{2}\sum_{1\le j_1<j_2\le s} b_{j_1} b_{j_2}^2 c_{j_1}^{i_1-1} c_{j_2}^{i_2+i_3-2} + \sum_{1\le j_1<j_2<j_3\le s} b_{j_1} b_{j_2} b_{j_3} c_{j_1}^{i_1-1} c_{j_2}^{i_2-1} c_{j_2}^{i_3-1} = \frac{1}{(i_1+i_2+i_3)(i_1+i_2)i_1}$

We begin by showing that the infinite series expansion (2.35) of $\mathrm{e}^{-hF_1}\,\Psi(h)$ can be represented as a truncated series plus a remainder. For that purpose, we consider (2.26) with $A(\tau)$ defined as follows:

$$A(\tau) = \begin{cases} b_i C(hc_i) & \text{if } \tau \in [i-1, i],\ i \in \{1, \dots, s\}, \\ 0 & \text{if } \tau > s. \end{cases} \tag{2.40}$$

In that case, if $\tau \in [i, i+1]$ with $i < s$,

$$Y(\tau) = \mathrm{e}^{(\tau-i)hb_{i+1}C(hc_{i+1})}\mathrm{e}^{hb_iC(hc_i)} \cdots \mathrm{e}^{hb_1C(hc_1)},$$

and in particular $\mathrm{e}^{-hF_1}\,\Psi(h) = Y(s)$. Since the solution $Y(\tau)$ of (2.26) admits the representation (2.29) with remainder (2.30), then

$$\mathrm{e}^{-hF_1}\Psi(h) = I + \sum_{k=1}^{n} h^k \int_0^s \int_0^{\tau_k} \cdots \int_0^{\tau_2} A(\tau_k)A(\tau_{k-1})\cdots A(\tau_1)\,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_k + h^{n+1}\mathcal{R}_{n+1}(s,h),$$

where for each n,

$$\mathcal{R}_n(s,h) = \int_0^s \int_0^{\tau_n} \cdots \int_0^{\tau_2} A(\tau_n)A(\tau_{n-1})\cdots A(\tau_1)Y(\tau_1)\,\mathrm{d}\tau_1 \cdots \mathrm{d}\tau_n.$$

By comparison with (2.35), we conclude that

$$\sum_{1\le j_1\le\cdots\le j_k\le s} \frac{b_{j_1}\cdots b_{j_k}}{\sigma(j_1,\dots,j_k)} C(c_{j_k}h)\cdots C(c_{j_1}h)$$
$$= \int_0^s \int_0^{\tau_k} \cdots \int_0^{\tau_2} A(\tau_k)A(\tau_{k-1})\cdots A(\tau_1)\,\mathrm{d}\tau_1\cdots\mathrm{d}\tau_k.$$

Now, $A(\tau)$ can be written as

$$A(\tau) = d_1(\tau)C_1 + hd_2(\tau)C_2 + h^2 d_3(\tau)C_3 + \cdots$$

where $d_j(\tau) = b_i c_i^{j-1}$ if $\tau \in [i, i+1]$ with $i < s$. Proceeding as in the previous section, we obtain

$$v_{i_1,\dots,i_k}(b_1,c_1,\dots,b_s,c_s) = \int_0^s \int_0^{\tau_1} \cdots \int_0^{\tau_{k-1}} d_{i_1}(\tau_1)\cdots d_{i_k}(\tau_k)\,\mathrm{d}\tau_k\cdots\mathrm{d}\tau_1, \tag{2.41}$$

which implies that the shuffle relations (2.39) hold for the polynomials $v_{i_1,\dots,i_k}$.

As in the previous subsection, a set of independent conditions that imply the order conditions (2.38) can be obtained by considering (2.38) for each Lyndon multi-index $(i_1,\dots,i_k)$ such that $1 < i_1+\cdots+i_k \le r$. Here, we exclude the multi-index (1) as in that case (2.38) coincides with the second equality in the consistency condition (2.22). For instance, the subset of Lyndon multi-indices $(i_1,\dots,i_k)$ such that $1 < i_1+\cdots+i_k \le 5$ is

$$\{(2),(3),(4),(5),(1,2),(1,3),(1,4),(2,3),(1,1,2),(1,1,3),(1,2,2),(1,1,1,2)\}.$$

For time-symmetric splitting methods, a set of independent order conditions will be obtained by considering (2.38) restricted to Lyndon multi-indices $(i_1,\dots,i_k)$ with odd weight $i_1+\cdots+i_k$. For instance, the subset of Lyndon multi-indices $(i_1,\dots,i_k)$ such that $1 < i_1+\cdots+i_k \le 5$ with odd weight $i_1+\cdots+i_k$ is

$$\{(3),(5),(1,2),(1,4),(2,3),(1,1,3),(1,2,2),(1,1,1,2)\}.$$

Notice that the treatment carried out in this subsection may also be formally applied when F_1 is an unbounded operator. In that case, however, we have to get rigorous estimates of the remainders to prove stability and convergence of the corresponding schemes, as is done in Thalhammer (2008), for example.

2.5. *Order conditions IV: composition methods with Lyndon multi-indices*

We now turn our attention to compositions (2.1) of a basic second-order time-symmetric scheme $S_h^{[2]}$ with appropriate coefficients $\gamma_1,\dots,\gamma_s$ chosen to achieve higher orders. Of course, a set of conditions that guarantee that the scheme (2.1) attains a given order can be obtained by rewriting that composition in terms of basic maps, such as (2.19), and using the characterization of the order of the splitting method (2.19) described in Section 2.4. The corresponding parameters a_j, b_j can

be obtained in terms of $\gamma_1, \dots, \gamma_s$ as follows: $a_1 = \gamma_1/2$, and

$$a_{j+1} = \frac{\gamma_j + \gamma_{j+1}}{2}, \quad b_j = \gamma_j \quad \text{for } j = 1, \dots, s,$$

with $\gamma_{s+1} = 0$. However, the resulting polynomial equations for any given order r, once written in terms of the coefficients $\gamma_1, \dots, \gamma_s$, are no longer independent. An alternative formulation of the order of (2.1) in terms of explicit independent algebraic equations in the coefficients $\gamma_1, \dots, \gamma_s$ will be presented in Section 2.5.2 below. This characterization is based on the treatment of the more general composition (2.15), which is treated next.

2.5.1. Order conditions of compositions of a basic method and its adjoint

The composition (2.15) is at least of order r if $\Psi(h) - \exp(h(F_1 + F_2)) = O(h^{r+1})$, where $\Psi(h)$ is the associated Lie transformation (2.16).

To get the series expansion of $\Psi(h)$, we first consider the expansion in powers of h of the Lie transformation $\mathcal{X}(h) = I + hX_1 + h^2X_2 + \cdots$ associated to the basic integrator χ_h. If χ_h is the Lie–Trotter scheme, $\chi_h = \varphi_h^{[1]} \circ \varphi_h^{[2]}$, then

$$\mathcal{X}(h) = \mathrm{e}^{hF_2}\mathrm{e}^{hF_1} = I + h(F_1 + F_2) + h^2\left(\frac{1}{2}F_1^2 + F_2F_1 + \frac{1}{2}F_2^2\right) + \cdots .$$

To deal with the most general problem, however, from now on we only assume that χ_h is a smooth consistent integrator, so that $X_1 = F_1 + F_2$, and each X_n can be defined so that for each smooth function g, $X_n g$ is a new smooth function given by (2.9).

Let us consider $\Psi_0(h) = I$, and for each $j \geq 1$,

$$\begin{aligned} \Psi_{2j-1}(h) &= \mathcal{X}(-\alpha_1 h)^{-1}\mathcal{X}(\alpha_2 h)\cdots\mathcal{X}(-\alpha_{2j-1}h)^{-1}, \\ \Psi_{2j}(h) &= \mathcal{X}(-\alpha_1 h)^{-1}\mathcal{X}(\alpha_2 h)\cdots\mathcal{X}(-\alpha_{2j-1}h)^{-1}\mathcal{X}(\alpha_{2j}h), \end{aligned} \tag{2.42}$$

so that in particular, $\Psi(h) = \Psi_{2s}(h)$. Notice that

$$\begin{aligned} \mathcal{X}(-h)^{-1} &= I + \sum_{k\geq 1}(-1)^k(-hX_1 + h^2X_2 - h^3X_3 + \cdots)^k \\ &= I + hX_1 + h^2(X_1^2 - X_2) + h^3(X_1^3 - X_1X_2 - X_2X_1 + X_3) + \cdots , \end{aligned}$$

which implies that for each $k \geq 1$ there exist polynomials $w_{i_1,\dots,i_m}(\alpha_1, \dots, \alpha_k)$ on the coefficients $\alpha_1, \dots, \alpha_k$ such that

$$\Psi_k(h) = I + \sum_{n\geq 1} h^n \sum_{m\geq 1} \sum_{i_1+\cdots+i_m=n} w_{i_1,\dots,i_m}(\alpha_1, \dots, \alpha_k)\, X_{i_1}\cdots X_{i_m}. \tag{2.43}$$

We next determine the polynomial coefficients $w_{i_1,\dots,i_m}(\alpha_1, \dots, \alpha_k)$ recursively from the relations $\Psi_1(h) = \mathcal{X}(-\alpha_{2j-1}h)^{-1}$ and

$$\Psi_{2j-1}(h) = \Psi_{2j-2}(h)\mathcal{X}(-\alpha_{2j-1}h)^{-1}, \quad \Psi_{2j}(h) = \Psi_{2j-1}(h)\mathcal{X}(\alpha_{2j}h),$$

or equivalently, $\Psi_1(h)\mathcal{X}(-\alpha_1 h) = I$ and

$$\Psi_{2j-1}(h)\mathcal{X}(-\alpha_{2j-1}h) = \Psi_{2j-2}(h), \quad \Psi_{2j}(h) = \Psi_{2j-1}(h)\mathcal{X}(\alpha_{2j}h). \tag{2.44}$$

Specifically, for arbitrary coefficients $w_{i_1,\dots,i_m}$ and λ we have

$$\begin{aligned}
&\left(I + \sum_{n\geq 1} h^n \sum_{m\geq 1} \sum_{i_1+\cdots+i_m=n} w_{i_1,\dots,i_m} X_{i_1}\cdots X_{i_m}\right)\mathcal{X}(\lambda h)\\
&\quad = I + \sum_{i\geq 1} h^i (w_i + \lambda^i) X_n\\
&\qquad + \sum_{n\geq 1} h^n \sum_{m\geq 2} \sum_{i_1+\cdots+i_m=n} (w_{i_1,\dots,i_m} + \lambda^{i_m} w_{i_1,\dots,i_{m-1}})\, X_{i_1}\cdots X_{i_m},
\end{aligned}$$

so that, taking this expression into account, (2.44) leads to the following identities:

$$\begin{aligned}
w_i(\alpha_1) &= -(-\alpha_1)^i,\\
w_i(\alpha_1,\dots,\alpha_{2j-1}) &= w_i(\alpha_1,\dots,\alpha_{2j-2}) - (-\alpha_{2j-1})^i,\\
w_i(\alpha_1,\dots,\alpha_{2j}) &= w_i(\alpha_1,\dots,\alpha_{2j-1}) + \alpha_{2j}^i,\\
w_{i_1,\dots,i_m}(\alpha_1) &= -(-\alpha_1)^{i_m} w_{i_1,\dots,i_{m-1}}(\alpha_1),\\
w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_{2j-1}) &= w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_{2j-2})\\
&\quad - (-\alpha_{2j-1})^{i_m} w_{i_1,\dots,i_{m-1}}(\alpha_1,\dots,\alpha_{2j-1}),\\
w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_{2j}) &= w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_{2j-1})\\
&\quad + \alpha_{2j}^{i_m} w_{i_1,\dots,i_{m-1}}(\alpha_1,\dots,\alpha_{2j-1}).
\end{aligned}$$

Clearly, (2.43) holds for the coefficients $w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_k)$ determined by the relations above. Equivalently, the functions $w_{i_1,\dots,i_m}$ can be defined as

$$\begin{aligned}
w_i(\alpha_1,\dots,\alpha_{2\ell}) &= \sum_{j=1}^{s}\left(\alpha_{2j}^i - (-\alpha_{2j-1})^i\right),\\
w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_{2\ell}) &= \sum_{j=1}^{s}\left(\alpha_{2j}^{i_m} - (-\alpha_{2j-1})^{i_m}\right) w_{i_1,\dots,i_{m-1}}(\alpha_1,\dots,\alpha_{2j-1}).
\end{aligned} \tag{2.45}$$

Notice that

$$w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_k,0,\dots,0) = w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_k),$$

as expected from (2.42), (2.43) and $\mathcal{X}(0) = I$.

Comparing the series expansion of $\Psi(h) = \Psi_{2s}(h)$ with

$$\exp(h\,(F_1 + F_2)) = \exp(h\,X_1) = I + \sum_{n\geq 1} \frac{h^n}{n!} X_1^n,$$

we finally conclude that the scheme (2.15) is of order at least r if and only if

$$w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_{2s}) = \begin{cases} \dfrac{1}{m!} & \text{if } (i_1,\dots,i_m) = \overbrace{(1,\dots,1)}^{m}, \\ 0 & \text{otherwise,} \end{cases} \tag{2.46}$$

for each multi-index $(i_1,\dots,i_k)$ such that $1 \le i_1 + \cdots + i_k \le r$.

Furthermore, the order conditions (2.46) are not all independent. For instance, it is straightforward to check from (2.45) that $w_{i_1,i_2} + w_{i_2,i_1} + w_{i_1+i_2} = w_{i_1} w_{i_2}$ for arbitrary indices i_1, i_2. In particular, $2w_{1,1} + w_2 = w_1^2$, which implies that if the order conditions (2.46) for the multi-indices (1) and (2) are satisfied, then the condition for the multi-index $(1,1)$ is automatically fulfilled. Actually, such dependences are similar to the shuffle relations (2.32) that hold for the coefficients $v_{i_1,\dots,i_m}$ considered in Section 2.4. Indeed,

$$w_{i_1,\dots,i_n} w_{i_{n+1},\dots,i_{n+m}} = \sum_{\sigma \in \mathrm{Sh}(n,m)} w_{i_{\sigma(1)},\dots,i_{\sigma(n+m)}} + \cdots, \tag{2.47}$$

where $\cdots$ refers to sums of products of coefficients corresponding to multi-indices with $m-1$ or fewer indices. In fact, as shown in Chartier and Murua (2009), the dependences (2.47) are directly related to the quasi-shuffle product $*$ on the linear span of multi-indices introduced in Hoffman (2000), and due to such dependences, it is enough to consider (2.46) for Lyndon multi-indices $(i_1,\dots,i_m)$ such that $i_1 + \cdots + i_k \le r$. That is, the scheme (2.15) is of order at least r if $\alpha_1 + \cdots + \alpha_{2s} = 1$ and for each Lyndon multi-index $(i_1,\dots,i_k)$ such that $1 < i_1 + \cdots + i_k \le r$,

$$w_{i_1,\dots,i_m}(\alpha_1,\dots,\alpha_{2s}) = 0. \tag{2.48}$$

In particular, a method of order three must satisfy, besides consistency, the conditions $w_2 = w_3 = w_{1,2} = 0$.

This provides an alternative characterization of the order conditions of general splitting schemes (2.17) in terms of the coefficients α_j obtained from the method parameters a_j, b_j from (2.18). Most importantly, this also allows us to characterize the order of the scheme (2.1) obtained by composing Strang maps. This will be presented next in Section 2.5.2.

2.5.2. Explicit characterization of the order conditions of scheme (2.1)
Murua and Sanz-Serna (1999) have obtained such an explicit characterization in terms of a set of polynomials indexed by certain sets of rooted trees decorated by the set of odd positive integers. We now describe a related formulation in terms of polynomials indexed by the set of Lyndon multi-indices with odd indices based on the formalism developed in the previous subsection for the more general composition (2.15).

We begin by considering (2.1) as the particular case of (2.15) with

$$\alpha_{2j-1} = \alpha_{2j} = \gamma_j/2 \quad \text{for } j = 1,\dots,s. \tag{2.49}$$

For each multi-index $(i_1, \ldots, i_k)$, we define the function $u_{i_1,\ldots,i_m}$ on the set of finite sequences $(\gamma_1, \ldots, \gamma_s)$ of real numbers as follows:

$$u_{i_1,\ldots,i_m}(\gamma_1, \ldots, \gamma_s) = 2^{i_1+\cdots+i_m-m} w_{i_1,\ldots,i_m}(\alpha_1, \ldots, \alpha_{2s}), \tag{2.50}$$

with $\alpha_{2j-1} = \alpha_{2j} = \gamma_j/2$. Clearly, scheme (2.1) is of order at least r if and only if $\gamma_1 + \cdots + \gamma_s = 1$ and

$$u_{i_1,\ldots,i_m}(\gamma_1, \ldots, \gamma_s) = 0 \tag{2.51}$$

for each Lyndon multi-index $(i_1, \ldots, i_k)$ such that $1 < i_1 + \cdots + i_k \leq r$. However, by definition, $u_{i_1,\ldots,i_m}(\gamma_1, \ldots, \gamma_s) \equiv 0$ if i_m is even. Moreover, for any multi-index $(i_1, \ldots, i_k)$ with some even index, (2.51) holds provided that it holds for every Lyndon multi-index with fewer indices.

Therefore, scheme (2.1) is of order at least r if and only if $\gamma_1 + \cdots + \gamma_s = 1$ and (2.51) for each Lyndon multi-index $(i_1, \ldots, i_k)$ with odd indices such that $1 < i_1 + \cdots + i_k \leq r$. For instance, the set of Lyndon multi-indices $(i_1, \ldots, i_k)$ of odd indices such that $1 < i_1 + \cdots + i_k \leq 7$ is

$$\{(3), (5), (7), (1,3), (1,5), (1,1,3), (1,1,5), (1,3,3), (1,1,1,3), (1,1,1,1,3)\}.$$

The resulting number of order conditions, denoted as m_n, is gathered in Table 2.1.

For multi-indices $(i_1, \ldots, i_m)$ with odd indices, the functions $u_{i_1,\ldots,i_m}$ can be written more explicitly as follows:

$$u_i(\gamma_1, \ldots, \gamma_s) = \sum_{j=1}^{s} \gamma_j^i, \tag{2.52}$$

$$u_{i_1,i_2}(\gamma_1, \ldots, \gamma_s) = \sum_{j_2=1}^{s} \gamma_{j_2}^{i_2} \sum_{j_1=1}^{j_2*} \gamma_{j_1}^{i_1}, \tag{2.53}$$

$$u_{i_1,i_2,i_3}(\gamma_1, \ldots, \gamma_s) = \sum_{j_3=1}^{s} \gamma_{j_3}^{i_3} \sum_{j_2=1}^{j_3*} \gamma_{j_2}^{i_2} \sum_{j_1=1}^{j_2*} \gamma_{j_1}^{i_1}, \tag{2.54}$$

$$u_{i_1,i_2,i_3,i_4}(\gamma_1, \ldots, \gamma_s) = \sum_{j_4=1}^{s} \gamma_{j_4}^{i_4} \sum_{j_3=1}^{j_4*} \gamma_{j_3}^{i_3} \sum_{j_2=1}^{j_3*} \gamma_{j_2}^{i_2} \sum_{j_1=1}^{j_2*} \gamma_{j_1}^{i_1}, \tag{2.55}$$

and so on. Here, as in Murua and Sanz-Serna (1999), we have used the notation

$$\sum_{j=1}^{k*} A_j = \frac{A_k}{2} + \sum_{j=1}^{k-1} A_j$$

for $A_1, \ldots, A_k \in \mathbb{R}$.

For time-symmetric integration schemes (2.1), a set of independent conditions for even order r will be obtained by considering (2.51) restricted to Lyndon multi-indices $(i_1, \ldots, i_k)$ of odd indices and odd weight $i_1 + \cdots + i_k < r$. For instance, for

order eight, we only need to consider the following subset of Lyndon multi-indices:

$$\{(3), (5), (7), (1, 1, 3), (1, 1, 5), (1, 3, 3), (1, 1, 1, 1, 3)\}.$$

The above characterization of the order of the scheme (2.1) is also true in the more general case where the Strang map $S_h^{[2]}$ is replaced by an arbitrary time-symmetric integrator of order 2ℓ. In that case, only Lyndon multi-indices with indices from the set $\{1, 2\ell + 1, 2\ell + 3, 2\ell + 5, \ldots\}$ have to be taken into account.

2.6. *Negative time steps*

Splitting and composition methods of order $r \geq 3$ necessarily involve some negative coefficients. This can already be observed in the simple triple jump scheme (2.3), and in fact has been established as a general theorem by Goldman and Kaper (1996), Sheng (1989) and Suzuki (1991). A simple proof can be obtained as follows (Blanes and Casas 2005): given the existing relationship between the splitting method (2.17) and the composition (2.15), it is clear that any splitting scheme of order $r \geq 3$ has to verify the condition $w_3 = 0$, where, by virtue of (2.45),

$$w_3(\alpha_1, \ldots, \alpha_{2s}) = \sum_{j=1}^{s} \left(\alpha_{2j}^3 - (-\alpha_{2j-1})^3\right), \tag{2.56}$$

where the coefficients α_j are related to a_j, b_j via (2.18). Since, for all $x, y \in \mathbb{R}$, it is true that $x^3 + y^3 < 0$ implies $x + y < 0$, then there must exist some $j \in \{1, \ldots, s\}$ in the sum of (2.56) such that

$$\alpha_{2j-1}^3 + \alpha_{2j}^3 < 0 \quad \text{and thus} \quad \alpha_{2j-1} + \alpha_{2j} = b_j < 0.$$

Obviously, we can also write (by taking $\alpha_0 = 0$, $\alpha_{2s+1} = 0$)

$$w_3(\alpha_1, \ldots, \alpha_{2s}) = \sum_{i=0}^{s} \left(\alpha_{2j}^3 + \alpha_{2j+1}^3\right) = 0$$

just by grouping terms in a different way, and thus, by repeating the argument, there must exist some $j \in \{0, \ldots, s\}$ such that

$$\alpha_{2j} + \alpha_{2j+1} = a_j < 0.$$

This proof clearly shows the origin of the existence of backward time steps: the equation $w_3 = 0$ can only be satisfied if at least one a_j *and* one b_j are negative.

3. Splitting methods for special problems

Whereas the analysis carried out in Section 2 is completely general, there are important problems arising in applications whose particular structure allows us to simplify the treatment and design schemes without taking into account all the order conditions. Some of them are reviewed in this section, where we also show how to adapt splitting methods to deal with explicitly time-dependent systems.

3.1. RKN splitting methods

Many differential equations of practical interest are of the form

$$y'' = g(y), \tag{3.1}$$

where $y \in \mathbb{R}^d$ and $g : \mathbb{R}^d \longrightarrow \mathbb{R}^d$. An example in point corresponds to Hamiltonian systems of the form $H(q, p) = T(p) + V(q)$, where the kinetic energy $T(p)$ is quadratic in the momenta p, i.e. $T(p) = \frac{1}{2} p^\top M^{-1} p$ for a constant invertible symmetric matrix M, and $V(q)$ is the potential. In that case, the corresponding Hamiltonian system can be written in the form (3.1) with $y = q$, $g(y) = -M^{-1}\nabla V(y)$.

By transforming (3.1) into a first-order ODE system (of dimension $D = 2d$) in the new variables $x = (y, v)^\top$, with $v = y'$, it is clear that the resulting equation

$$x' = \frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} y \\ v \end{pmatrix} = \begin{pmatrix} v \\ g(y) \end{pmatrix}$$

can be expressed as $x' = f_1(x) + f_2(x)$, with

$$f_1(y, v) = (0, g(y))^\top, \quad f_2(y, v) = (v, 0)^\top, \tag{3.2}$$

and splitting methods of the form (2.17) can be applied, with the exact h-flows $\varphi_h^{[1]}$ and $\varphi_h^{[2]}$ given by

$$\varphi_h^{[1]} : \begin{pmatrix} y_0 \\ v_0 \end{pmatrix} \longmapsto \begin{pmatrix} y_0 \\ v_0 + h\, g(y_0) \end{pmatrix}, \quad \varphi_h^{[2]} : \begin{pmatrix} y_0 \\ v_0 \end{pmatrix} \longmapsto \begin{pmatrix} y_0 + h\, v_0 \\ v_0 \end{pmatrix}. \tag{3.3}$$

Just as the class of Runge–Kutta methods can be conveniently adapted to (3.1) to get more efficient schemes (the so-called Runge–Kutta–Nyström or RKN methods; see Hairer, Nørsett and Wanner 1993), special splitting methods can also be designed to improve the accuracy whilst reducing the computational cost with respect to the general composition (2.17). For analogy, they are sometimes called splitting methods of RKN type. The key point here is that the differential operators F_1 and F_2 associated with (3.2) satisfy $[F_1, [F_1, [F_1, F_2]] = 0$ identically. In other words, $F_{1112} = 0$ in (2.21), which introduces linear dependences among higher-order terms in the expansion of $\log(\Psi(h))$ (McLachlan and Quispel 2002) and therefore contributes to a reduction in the number of order conditions. In the notation of Section 2.4, $6C_4 = -[F_1, [F_1, [F_1, F_2]]$, and thus we have that $C_k = 0$ for $k \geq 4$. Hence, the order conditions (2.38) for multi-indices $(i_1, \ldots, i_k)$ with some index $i_j \geq 4$ need not be considered in that case. In particular, a splitting scheme is at least of order five provided that (2.38) holds for the Lyndon multi-indices

$$\{(2), (3), (1, 2), (1, 3), (2, 3), (1, 1, 2), (1, 1, 3), (1, 2, 2), (1, 1, 1, 2)\}.$$

Here, we have excluded the Lyndon multi-indices $(4), (5), (1, 4)$ from the set of Lyndon multi-indices $(i_1, \ldots, i_k)$ with $i_1 + \cdots + i_k \leq 5$. For order six, in addition, we have to consider (2.38) for the following Lyndon multi-indices:

$$(1, 1, 1, 1, 2), (1, 1, 1, 3), (1, 2, 3), (1, 1, 2, 2), (1, 3, 2).$$

They are obtained by excluding from the set of nine Lyndon multi-indices with $i_1 + \cdots + i_k = 6$, the Lyndon multi-indices $(2,4), (1,1,4), (1,5), (6)$. Similarly, for order seven, in addition, we have to consider (2.38) for the following Lyndon multi-indices:

$$(1,1,1,1,1,2), (1,1,1,1,3), (1,1,2,3), (1,1,1,2,2), (1,1,3,2),$$
$$(2,2,3), (1,2,1,3), (1,1,2,3), (1,3,3), (1,2,2,2).$$

For orders higher than seven, more reductions of the order conditions occur, in addition to those obtained by excluding Lyndon multi-indices $(i_1, \ldots, i_k)$ having some index $i_j \geq 4$, due to additional dependences among nested commutators of C_1, C_2, C_3. For instance, it is straightforward to check that $[C_3, [C_2, C_3]]$ vanishes identically, which implies that for order eight we can also exclude the Lyndon multi-index $(2,3,3)$.

The class of problems for which the reduction in order conditions discussed above is in fact more general than (3.2). In particular, it includes the situation where f_1 depends only on y and $f_2(y,v)$ is linear in v. For Hamiltonian systems, this generalization corresponds to $H(q,p) = V(q) + T(q,p)$, where $T(q,p)$ is quadratic in p. In that case, the flow associated with $T(q,p)$ should be easily computed for the RKN splitting methods to be advantageous.

Actually, such a reduction of the order conditions also holds for certain PDEs, and, in particular, for the time-dependent Schrödinger equation considered in Section 1.6 (with $F_1 = \hat{V}$ and $F_2 = \hat{T}$), since the corresponding graded Lie algebra is isomorphic to the classes of problems discussed above (McLachlan and Murua 2019). In fact, McLachlan and Murua (2019) conjectured (and checked up to order 20) that the case (3.2) (where $f_2(y,v) = v$) and the more general case (where $f_2(y,v)$ is linear in v but may depend on y) give rise to the same reduction in the order conditions.

The actual number d_r of order conditions for orders $r \leq 11$ are given in Table 2.1. Since the order conditions up to order three are identical as in the general case, the results for negative time steps still apply.

This reduction in the number of order conditions allows us to design schemes involving a smaller number of elementary flows than in the general case, eventually leading to greater efficiency. We have already illustrated a popular fourth-order method within this class (scheme $\mathrm{RKN}_6 4$ in the examples of Section 1).

3.2. *Methods with commutators*

Another possible way to improve the efficiency of splitting methods consists in incorporating into the scheme not only the flows of F_1 and F_2 but also the flows of some of their commutators $[F_1, F_2]$, $[[F_1, F_2], F_1]$, etc., or convenient approximations of these flows. Of course, the strategy makes sense if the gain in accuracy, stability or any other favourable property compensates the extra computational cost due to the presence of these additional flows. A popular fourth-order method

belonging to this class is

$$\psi_h = \varphi^{[1]}_{a_1 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[112]}_{d_2 h^3} \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_1 h}, \tag{3.4}$$

with

$$a_1 = \frac{1}{6}, \quad b_1 = \frac{1}{2}, \quad a_2 = \frac{1}{3}, \quad d_2 = -\frac{1}{72}, \tag{3.5}$$

first proposed by Koseleff (1993) and Chin (1997). Here $\varphi^{[112]}_h$ denotes the h-flow corresponding to $F_{112} = [F_1, [F_1, F_2]]$ (or $F_{112} = 2C_3$ in the notation of Section 2.4).

We next analyse the family of schemes (3.4). The Lie transformation $\Psi(h)$ is

$$\Psi(h) = \mathrm{e}^{ha_1 F_1}\, \mathrm{e}^{hb_1 F_2}\, \mathrm{e}^{ha_2 F_1}\, \mathrm{e}^{h^3 2 d_2 C_3}\, \mathrm{e}^{ha_2 F_1}\, \mathrm{e}^{hb_1 F_2}\, \mathrm{e}^{ha_1 F_1}.$$

Under the assumption that $a_1 + a_2 = \frac{1}{2}$, we have

$$\mathrm{e}^{-hF_1}\Psi(h) = \mathrm{e}^{hb_1 C(hc_3)}\, \mathrm{e}^{h^3 d_2 D(hc_2)}\, \mathrm{e}^{hb_1 C(hc_1)},$$

where $c_1 = a_1$, $c_2 = a_1 + a_2$, $c_3 = a_1 + 2a_2$, $C(h)$ is given by (2.34), and

$$D(h) = 2\mathrm{e}^{-hF_1}\, C_3\, \mathrm{e}^{hF_1} = \sum_{n=1}^{\infty} h^{n-1}(n+1)nC_{n+2},$$

so that

$$\mathrm{e}^{-hF_1}\Psi(h) = \mathrm{e}^{hb_1(C_1 + hc_3C_2 + h^2c_3^2C_3)}\, \mathrm{e}^{2h^3 d_2 C_3}\, \mathrm{e}^{hb_1(C_1 + hc_1C_2 + h^2c_1^2C_3)} + O(h^4).$$

Expanding the right-hand side in powers of h and comparing the coefficients multiplying hC_1, h^2C_2, $h^3C_1C_2$ and h^3C_3 respectively (corresponding to the Lyndon multi-indices (1), (2), (1, 2), (3)) with those in the expansion

$$\mathrm{e}^{-hF_1}\mathrm{e}^{h(F_1+F_2)} = I + hC_1 + \frac{h^2}{2}C_2 + \frac{h^3}{3}C_3 + \frac{h^3}{6}C_1C_2 + \frac{h^3}{3}C_2C_1 + O(h^4),$$

we conclude that the time-symmetric scheme (3.4) is at least of order four if

$$b_1 = \frac{1}{2}, \quad b_1(c_1 + c_3) = \frac{1}{2}, \quad \frac{1}{2}\, b_1^2(c_1 + c_3) + b_1^2 c_1 = \frac{1}{6},$$
$$b_1\left(c_1^2 + c_3^2\right) + 2d_2 = \frac{1}{3}.$$

That system of polynomial equations has a unique solution, corresponding to the choice (3.5), as expected. The order conditions of more general products of scaled exponentials of hF_1, hF_2 and $h^3[F_1, [F_1, F_2]]$ can be derived similarly.

Recall that in the RKN case $C_4 = [C_3, F_1] = 0$, which implies that $\varphi^{[1]}_h$ and $\varphi^{[112]}_h$ commute. Hence, the three central terms in (3.4) can be merged into one, the h-flow of the differential operator $2a_2\, F_1 + d_2 h^2 F_{112}$, which is of the form

$$\sum_j G_j(q, h)\frac{\partial}{\partial v^j}.$$

In the case of (3.2),

$$\varphi^{[1]}_{a_2 h} \circ \varphi^{[112]}_{d_2 h} \circ \varphi^{[1]}_{a_2 h}(y, v) = \left(y,\ v + 2ha_2\, g(y) + h^3 d_2\, \frac{\partial g}{\partial y}(y) g(y) \right). \tag{3.6}$$

For Hamiltonian systems $H(p, q) = V(q) + \frac{1}{2} p^\top M^{-1} p$, F_{112} is the operator associated to the Hamiltonian function $(\nabla V)^\top M^{-1} \nabla V$, depending only on q. Thus (3.6) is the h-flow of the Hamiltonian

$$2a_2\, V(q) + d_2\, h^2 (\nabla V(q))^\top M^{-1} \nabla V(q), \tag{3.7}$$

which reduces to the potential $V(q)$ when $a_2 = 1/2$ and $d_2 = 0$. This explains the term 'splitting methods with modified potentials' frequently used in the literature (López-Marcos, Sanz-Serna and Skeel 1997, Rowlands 1991, Wisdom, Holman and Touma 1996). One such method has been illustrated in practice in Section 1.6 (Figure 1.5).

In addition to the reduction in the number of force evaluations, including flows associated with commutators has another advantage: since the coefficients a_i, b_i do not have to satisfy all the order conditions at order $r \geq 3$, the results for negative time steps do not apply here, and in fact methods of order greater than two do exist within this class. In addition to (3.4), other 'forward' fourth-order methods (i.e. with all $a_j > 0$ and $b_j > 0$) involving second derivatives of the potential have been published (Omelyan, Mryglod and Folk 2002, 2003) and applied to systems where the presence of negative coefficients leads to severe stability problems (Bader, Blanes and Casas 2013). Although it has been shown that achieving order six in general requires some negative coefficients (Chin 2005), it is indeed possible to construct a sixth-order processed method for cubic potentials with all a_j, b_j being positive (Blanes, Casas, González and Thalhammer 2023).

Additional flows corresponding to commutators involving more operators can in principle be incorporated into the scheme. Thus, for instance, the operator $[F_{12}, F_{112}] = -2[C_2, C_3]$ is also of the form

$$\sum_j g_j^{[5]}(q, h) \frac{\partial}{\partial v^j},$$

which allows us to compute its h-flow explicitly; see Blanes, Casas and Ros (2001*a*) and Blanes, Casas and Murua (2008*b*) for more details. Again, this procedure allows us to introduce additional free parameters into the scheme and construct more efficient integrators as long as the simultaneous evaluation of $g(y)$, $(\partial g/\partial y)(y) g(y)$, $g^{[5]}(y)$, etc., is not substantially more expensive than the evaluation of $g(y)$ itself.

3.3. *Near-integrable systems*

Very often in applications we have to deal with differential equations such as

$$x' = f_1(x) + \varepsilon f_2(x), \tag{3.8}$$

where $|\varepsilon| \ll 1$ and the h-flows $\varphi_h^{[1]}, \varphi_h^{[2]}$ corresponding to f_1 and εf_2, respectively, are readily available. In classical Hamiltonian mechanics, in particular, it is rather common to have a Hamiltonian function H which is a small perturbation of an exactly integrable Hamiltonian H_1, that is, $H = H_1 + \varepsilon H_2$, with $0 < \varepsilon \ll 1$ (Goldstein 1980, Pars 1979). A canonical example corresponds to the gravitational N-body problem in Jacobi coordinates, already considered in Section 1.5.

For this type of problem, splitting methods of the form (2.17),

$$\psi_h = \varphi_{a_{s+1}h}^{[1]} \circ \varphi_{b_s h}^{[2]} \circ \varphi_{a_s h}^{[1]} \circ \cdots \circ \varphi_{a_2 h}^{[1]} \circ \varphi_{b_1 h}^{[2]} \circ \varphi_{a_1 h}^{[1]},$$

are especially well adapted. On the one hand, the error is at most $O(\varepsilon)$ and vanishes with ε, since in that case the scheme reproduces the exact solution. On the other hand, typically, $|\varepsilon| \ll h$ (or at least $|\varepsilon| \approx h$), so that we are mainly interested in eliminating error terms with small powers of ε, and its number grows as a polynomial in the order r, rather than exponentially. Thus there is only one error term of order εh^k (namely, the term $h^k \varepsilon w_{11\cdots 12} F_{11\cdots 12}$ in the expansion (2.21)), $\lfloor \frac{1}{2}(k-1) \rfloor$ terms of order $O(\varepsilon^2 h^k)$ and $\lfloor \frac{1}{6}(k-1)(k-2) \rfloor$ terms of order $O(\varepsilon^3 h^k)$ (McLachlan 1995*a*).

In the treatment of splitting methods for near-integrable systems it is convenient to introduce the notion of *generalized order*, following McLachlan (1995*a*). Thus, we say that ψ_h is of generalized order $(r_1, r_2, \ldots, r_m)$, with $r_1 \geq r_2 \geq \cdots \geq r_m$, if

$$\psi_h(x) - \varphi_h(x) = O\left(\varepsilon h^{r_1+1} + \varepsilon^2 h^{r_2+1} + \cdots + \varepsilon^m h^{r_m+1}\right) \quad \text{as } (h, \varepsilon) \longrightarrow (0, 0).$$

With this notation, a method such that the local error is $O(\varepsilon h^{2n+1} + \varepsilon^2 h^3)$ is said to be of generalized order $(2n, 2)$. In this sense, the Strang scheme is of order $(2, 2)$, whereas the $(10, 6, 4)$ integrator ψ_h used in Sections 1.4 and 1.5 satisfies

$$\psi_h(x) - \varphi_h(x) = O\left(\varepsilon h^{11} + \varepsilon^2 h^7 + \varepsilon^3 h^5\right).$$

The general analysis of the order conditions carried out in Section 2.4 in terms of Lyndon multi-indices readily allows us to characterize the generalized order of a given splitting scheme. We simply need to observe that, by replacing F_2 with εF_2 in the expansions derived in Section 2.4, a series expansion of $\mathrm{e}^{-hF_1}(\Psi(h) - \mathrm{e}^{h(F_1+\varepsilon F_2)})$ is obtained, where the term associated to a multi-index $(i_1, \ldots, i_k)$ with k indices is affected by a kth power of ε. In particular, we get Table 3.1, which contains the Lyndon multi-indices involved in several consistent palindromic splitting methods of the given generalized order.

3.4. *Splitting methods for linear systems*

In the numerical integration of the Schrödinger equation implemented in Section 1.6, we separated the system into kinetic and potential energy and then applied several schemes based on this splitting. It has long been recognized, however, that there exist other possibilities for splitting such a system. Given the N-dimensional

Table 3.1. Lyndon multi-indices corresponding to consistent palindromic splitting methods of a given generalized order.

Generalized order	Lyndon multi-indices
$(2n, 2)$	$(3), (5), \ldots, (2n-1)$
$(8, 4)$	$(3), (5), (7), (1, 2)$
$(10, 4)$	$(3), (5), (7), (9), (1, 2)$
$(8, 6, 4)$	$(3), (5), (7), (1, 2), (1, 4), (2, 3)$
$(10, 6, 4)$	$(3), (5), (7), (9), (1, 2), (1, 4), (2, 3)$

linear ODE

$$\mathrm{i}\frac{\mathrm{d}}{\mathrm{d}t}u(t) = H\,u(t), \quad u(0) = u_0 \in \mathbb{C}^N \tag{3.9}$$

resulting from the space discretization of equation (1.42), with H real and symmetric, Gray and Verosky (1994) and Gray and Manolopoulos (1996) separate u into its real and imaginary parts, $q = \mathrm{Re}\,(u)$, $p = \mathrm{Im}\,(u)$. Then, in terms of q, p, equation (3.9) leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}q = H\,p, \quad \frac{\mathrm{d}}{\mathrm{d}t}p = -H\,q, \tag{3.10}$$

so that they can be seen as the classical evolution equations corresponding to the Hamiltonian function

$$\hat{H}(q, p) = \frac{1}{2}p^\top H p + \frac{1}{2}q^\top H q \tag{3.11}$$

in terms of canonical variables q and p. The exact solution of (3.10) is given by

$$\begin{pmatrix} q(t) \\ p(t) \end{pmatrix} = O(tH)\begin{pmatrix} q_0 \\ p_0 \end{pmatrix}, \quad \text{where} \quad O(tH) = \begin{pmatrix} \cos(tH) & \sin(tH) \\ -\sin(tH) & \cos(tH) \end{pmatrix} \tag{3.12}$$

is an orthogonal and symplectic $2N \times 2N$ matrix. As with the formal solution $u(t) = \mathrm{e}^{-\mathrm{i}tH}u_0$ of (3.9), $O(tH)$ may be very expensive to compute, so that suitable approximations might be necessary, such as those provided by splitting methods applied to (3.11). In this respect, notice that if we introduce the nilpotent matrices A and B,

$$A \equiv \begin{pmatrix} 0 & H \\ 0 & 0 \end{pmatrix}, \quad B \equiv \begin{pmatrix} 0 & 0 \\ -H & 0 \end{pmatrix}, \tag{3.13}$$

then it is clear that the symplectic Euler–VT method of Section 1.4 is simply

$$\begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \mathrm{e}^{hB}\,\mathrm{e}^{hA}\begin{pmatrix} q_n \\ p_n \end{pmatrix}, \tag{3.14}$$

whereas the Störmer–Verlet Algorithm 1.3 corresponds to

$$x_{n+1} = \mathrm{e}^{\frac{h}{2}B}\,\mathrm{e}^{hA}\,\mathrm{e}^{\frac{h}{2}B}\,x_n, \quad x = (q,p)^\top \tag{3.15}$$

for a given time step h. Notice that all schemes based on this splitting are automatically symplectic.

At this point, nothing prevents us from using any of the RKN splitting methods treated in Section 3.1, even with nested commutators as in Section 3.2. It turns out, however, that the particularly simple algebraic structure of the system (3.11) makes it possible to design more efficient schemes. Specifically, there is only one independent condition to increase the order from $r = 2k-1$ to $r = 2k$, and only two to increase the order from $r = 2k$ to $r = 2k+1$ for a given k; see Blanes, Casas and Murua (2008*a*) for more details. As a result, splitting methods of the form

$$x_{n+1} = \mathrm{e}^{ha_{s+1}A}\,\mathrm{e}^{hb_sB}\,\cdots\,\mathrm{e}^{ha_2A}\,\mathrm{e}^{hb_1B}\,\mathrm{e}^{ha_1A}x_n \tag{3.16}$$

of order r for $r = 2, 4, 6, 8, 10$ and 12 can be obtained with $s = r$ exponentials e^{hb_jB} (Gray and Manolopoulos 1996). By contrast, at least 15 and 31 exponentials are needed in general to attain orders eight and ten, respectively.

A couple of comments are worth making. First, this class of symplectic methods do not preserve the orthogonal character of the exact solution given by $O(tH)$ (or alternatively the unitarity of (3.9)). Nevertheless, Blanes *et al.* (2008*a*) have shown that the average relative errors due to the lack of preservation of orthogonality or unitarity do not grow with time, since the schemes are conjugate to orthogonal or unitary methods for sufficiently small values of h. Second, although initially motivated by the time integration of the Schrödinger equation, methods (3.16) can be generalized in several ways. Thus they have been used to construct an algorithm to approximate $\mathrm{e}^{-\mathrm{i}tH}v$ for any real symmetric matrix H and any complex vector v by only carrying out matrix–vector products of the form Hv. As shown in Blanes, Casas and Murua (2015), the algorithm is more efficient than schemes based on Chebyshev polynomials for all tolerances and values of h. These methods can also be adapted for systems of the form

$$x' = My, \quad y' = -Nx,$$

with $x \in \mathbb{R}^{d_1}$, $y \in \mathbb{R}^{d_2}$, $M \in \mathbb{R}^{d_1\times d_2}$ and $N \in \mathbb{R}^{d_2\times d_1}$.

3.5. Splitting methods for non-autonomous systems

So far we have restricted our attention to splitting methods for autonomous differential equations. The question we analyse next is whether the same techniques can be applied when there is an explicit time dependence in the equation to integrate. The ideal situation would be that the methods designed for $x' = f(x)$ could also be used (maybe with only minor modifications) when we have $x' = f(t,x)$. In addition, we would like the schemes previously considered in this section for special problems to remain valid in the non-autonomous case.

Let us first consider the general situation, corresponding to a system of the form

$$x' = f(t,x) = f_1(t,x) + \cdots + f_m(t,x), \quad x(0) = x_0, \tag{3.17}$$

that is, when the explicit time dependence is present in each part. Then we can take t as a new coordinate and transform (3.17) into an equivalent autonomous equation to which standard splitting algorithms can subsequently be applied. More specifically, equation (3.17) is equivalent to the enlarged system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ x_{d+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ 1 \end{pmatrix}}_{\hat{f}_0} + \underbrace{\begin{pmatrix} f_1(x_{d+1},x) \\ 0 \end{pmatrix}}_{\hat{f}_1} + \cdots + \underbrace{\begin{pmatrix} f_m(x_{d+1},x) \\ 0 \end{pmatrix}}_{\hat{f}_m}, \tag{3.18}$$

with $x_{d+1} \in \mathbb{R}$. If the resulting (autonomous) equations

$$y' = \hat{f}_i(y), \quad i = 0, 1, \ldots, m \quad \text{with} \quad y = (x, x_{d+1})$$

can be solved, then we may use any splitting method of the form (2.17), since x_{d+1} advances only with $\hat{f}_0$ and remains constant for the rest of the system.

It turns out that, for problems which are separable into just two parts, that is,

$$x' = f(t,x) = f_1(t,x) + f_2(t,x), \quad x(0) = x_0, \tag{3.19}$$

we can do better: if t is taken as a new coordinate *twice*, and we write

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ x_{d+1} \\ x_{d+2} \end{pmatrix} = \underbrace{\begin{pmatrix} f_1(x_{d+1},x) \\ 0 \\ 1 \end{pmatrix}}_{\hat{f}_1} + \underbrace{\begin{pmatrix} f_2(x_{d+2},x) \\ 1 \\ 0 \end{pmatrix}}_{\hat{f}_2}, \tag{3.20}$$

with $x_{d+1}, x_{d+2} \in \mathbb{R}$, then we can apply the same splitting schemes designed for autonomous systems separable into two pieces to

$$y' = \hat{f}_1(y), \quad y' = \hat{f}_2(y) \quad \text{with} \quad y = (x, x_{d+1}, x_{d+2}).$$

This is so because x_{d+1} is constant when integrating the first equation and x_{d+2} is constant when solving the second one. The procedure can be viewed as a generalization of the one proposed in Sanz-Serna and Portillo (1996) for Hamiltonian systems $H(t,q,p) = T(t,p) + V(t,q)$: by introducing two new coordinates q_{d+1}, q_{d+2} and their associated momenta p_{d+1}, p_{d+2}, we instead deal with the formally autonomous Hamiltonian

$$\tilde{H}(q_{d+1}, q_{d+2}, q, p_{d+1}, p_{d+2}, p) = (T(p_{d+2}, p) + p_{d+1}) + (V(q_{d+1}, q) - q_{d+2}).$$

In the special case of the non-autonomous second-order differential equation

$$y'' = g(t,y), \tag{3.21}$$

it is convenient to split the system in the extended phase space as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} y \\ v \\ y_{d+1} \end{pmatrix} = \begin{pmatrix} v \\ 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 \\ g(y_{d+1}, y) \\ 0 \end{pmatrix},$$

since then we have an autonomous system with the same algebraic structure as those considered in Section 3.1, so that RKN splitting methods (even including commutators) can also be used.

For Hamiltonian systems $H(t, q, p) = T(p) + V(t, q)$, this is equivalent to introducing a new coordinate $q_{d+1} = t$ and its associated momentum $p_{d+1} = -H$, and considering the extended (autonomous) Hamiltonian function

$$\tilde{H}(q_{d+1}, q, p_{d+1}, p) = (T(p) + p_{d+1}) + V(q_{d+1}, q),$$

which is still quadratic in momenta, so that symplectic RKN methods can be used. Notice that $\tilde{H}$ is only linear in p_{d+1}, and modified potentials only involve derivatives of the potential with respect to q but not with respect to q_{d+1}, i.e. they do not require time derivatives. In this case the evolution for p_{d+1} is irrelevant, so there is no need to compute it.

Finally, if we take t as two new coordinates in the non-autonomous near-integrable system

$$x' = f(t, x) = f_1(t, x) + \varepsilon f_2(t, x), \quad x(0) = x_0, \tag{3.22}$$

then the special structure of a near-integrable system is destroyed. A partial remedy consists in separating the system as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} x \\ x_{d+1} \end{pmatrix} = \underbrace{\begin{pmatrix} f_1(x_{d+1}, x) \\ 1 \end{pmatrix}}_{\hat{f}_1} + \underbrace{\varepsilon \begin{pmatrix} f_2(x_{d+1}, x) \\ 0 \end{pmatrix}}_{\varepsilon \hat{f}_2}, \tag{3.23}$$

which requires us to numerically solve the non-autonomous unperturbed system.

We should bear in mind that taking time as an additional coordinate is of interest only if the time dependence in f_i, $i = 1, \ldots, m$ is cheap to compute. Otherwise the resulting algorithm may be computationally costly, since these functions have to be evaluated s times per time step. This drawback can be avoided by approximating the exact solution at each step by a composition of maps that in some sense incorporates average values of the vector fields with different weights on the subinterval $[t_n, t_{n+1}]$ (Blanes and Casas 2006). Specifically, the schemes read

$$\psi_h = \varphi_h^{[\hat{A}_{s+1}]} \circ \varphi_h^{[\hat{B}_s]} \circ \varphi_h^{[\hat{A}_s]} \circ \cdots \circ \varphi_h^{[\hat{A}_2]} \circ \varphi_h^{[\hat{B}_1]} \circ \varphi_h^{[\hat{A}_1]}, \tag{3.24}$$

where the maps $\varphi_h^{[\hat{A}_i]}$, $\varphi_h^{[\hat{B}_i]}$ are the exact 1-flows corresponding to the time-independent differential equations

$$x' = \hat{A}_i(x), \quad x' = \hat{B}_i(x), \quad i = 1, 2, \ldots \tag{3.25}$$

respectively, with

$$\hat{A}_i(x) \equiv h \sum_{j=1}^{k} \rho_{ij} f_1(\tau_j, x), \quad \hat{B}_i(x) \equiv h \sum_{j=1}^{k} \sigma_{ij} f_2(\tau_j, x). \tag{3.26}$$

Here $\tau_j = t_n + c_j h$ and the (real) constants c_j, ρ_{ij}, σ_{ij} are chosen in such a way that ψ_h provides an approximation of order r. These methods have the additional advantage that, when applied to (3.19) with the time frozen, they reproduce the standard splitting (2.17), since $\sum_j \rho_{ij} = a_i$ and $\sum_j \sigma_{ij} = b_i$. The same technique can be applied to the splitting methods analysed in Section 3.4 when the linear system (3.9) is explicitly time-dependent. In that case, the resulting scheme involves linear combinations of H evaluated at some intermediate times (Blanes, Casas and Murua 2017*a*).

4. Qualitative properties of splitting methods

4.1. Changes of variables and differential equations on smooth manifolds

Given a smooth autonomous differential equation in $\mathbb{R}^D$,

$$x' = f(x), \tag{4.1}$$

a smooth change of variables $x = \theta(\hat{x})$ transforms (4.1) into a new autonomous differential equation in $\mathbb{R}^D$,

$$\hat{x}' = \hat{f}(\hat{x}),$$

such that their t-flows are related as follows: for all $x_0 \in \mathbb{R}^D$,

$$\varphi_t^{[f]}(\theta(\hat{x}_0)) = \theta\big(\varphi_t^{[\hat{f}]}(\hat{x}_0)\big), \quad \text{with} \quad \hat{x}_0 = \theta^{-1}(x_0).$$

This implies that a similar property holds for the map $\psi_h \approx \varphi_h^{[f]}$ defined by the splitting method (2.17) applied to (4.1) when $f(x) = \sum_{j=1}^{2} f_j(x)$. That is, if the change of variables $x = \theta(\hat{x})$ transforms each equation $x' = f_j(x)$ into $\hat{x}' = \hat{f}_j(\hat{x})$, then the map $\hat{\psi}_h \approx \varphi_h^{[\hat{f}]}$ obtained by applying the splitting method with the same a_j, b_j coefficients in the new variables $\hat{x}$ is related to ψ_h by

$$\psi_h(\theta(\hat{x}_0)) = \theta(\hat{\psi}_h(\hat{x}_0)).$$

In words, the following two procedures give exactly the same numerical results: (i) applying the splitting method to the ODE corresponding to the new variables $\hat{x}$ with the initial condition $\hat{x}(0) = \hat{x}_0 \in \mathbb{R}^D$ and then transforming the result to the old variables x, and (ii) applying the splitting method to the ODE formulated in the old variables x with the initial condition $x(0) = x_0 := \theta(\hat{x}_0) \in \mathbb{R}^D$. This property does not hold in general for other integration schemes. For instance, in the case of Runge–Kutta methods it is true if the change of variables $\theta\colon \mathbb{R}^D \to \mathbb{R}^D$ is an affine map, but not in general.

The above property can be stated in a more abstract way by saying that splitting methods constitute a particular class of numerical integrators for differential equations defined on smooth manifolds: since the t-flow of a smooth vector field on a smooth manifold $\mathcal{M}$ is coordinate-independent, then splitting methods applied to a differential equation on $\mathcal{M}$ can also be naturally defined in a coordinate-independent way. More precisely, $\psi_h : \mathcal{M} \to \mathcal{M}$ defined by (2.17) gives a natural approximation of the h-flow $\varphi_h^{[f_1+f_2]}$ on the smooth manifold $\mathcal{M}$, described only in terms of objects related to the manifold itself. This is in contrast to other approaches for the numerical integration of differential equations on manifolds, which depend on particular choices of global or local charts, or particular embeddings of the manifold in a higher-dimensional Euclidean space (Hairer *et al.* 2006).

In fact, the treatment of the order conditions for splitting methods carried out in Section 2 is also valid in this more abstract setting. Suppose we have two smooth vector fields f_1 and f_2 on $\mathcal{M}$, and let F_i $(i = 1, 2)$ be the linear operators on $C^\infty(\mathcal{M}, \mathbb{R})$ defined as follows: for each $g \in C^\infty(\mathcal{M}, \mathbb{R})$, $F_i g \in C^\infty(\mathcal{M}, \mathbb{R})$ is the f_i-directional derivative of g. Then the t-flow $\varphi_t^{[f_i]}$ (or $\varphi_t^{[i]}$ for short) satisfies, for all $g \in C^\infty(\mathcal{M}, \mathbb{R})$ and $t \in \mathbb{R}$,

$$\frac{\mathrm{d}}{\mathrm{d}t} g(\varphi_t^{[i]}) = F_i g(\varphi_t^{[i]}).$$

This implies that the series in powers of t of $g(\varphi_t^{[i]})$ can be represented as $\mathrm{e}^{tF_i} g$, that is, we are in the same situation as in the case of differential equations on $\mathbb{R}^D$, discussed in Section 1.2. Consequently, the series in powers of h of $g(\psi_h(x))$ for the map $\psi_h : \mathcal{M} \to \mathcal{M}$ defined by (2.17) can be obtained by expanding $\Psi(h)g$, where $\Psi(h)$ is given by (2.20), so the analysis of the order conditions of Section 2 can be formally applied here as well.

4.2. *Stability*

An important characteristic of numerical integrators is their *stability*. Generally speaking, the numerical solution provided by a stable method does not tend to infinity when the exact solution is bounded. To analyse the (linear) stability of a given integrator, a model problem is typically chosen, so that both the numerical and exact solutions can be explicitly written out. In the case of a splitting method like (2.17), the model problem is the simple harmonic oscillator $y'' + \omega^2 y = 0$, $\omega > 0$ (López-Marcos, Sanz-Serna and Skeel 1996*b*, McLachlan and Gray 1997), with the standard $(x = (q, p) = (\omega y, y'))$ splitting

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \Bigg[\underbrace{\begin{pmatrix} 0 & \omega \\ 0 & 0 \end{pmatrix}}_{A} + \underbrace{\begin{pmatrix} 0 & 0 \\ -\omega & 0 \end{pmatrix}}_{B} \Bigg] \begin{pmatrix} q \\ p \end{pmatrix}, \tag{4.2}$$

and exact solution at time $t = h$

$$\begin{pmatrix} q(h) \\ p(h) \end{pmatrix} = M_z \begin{pmatrix} q(0) \\ p(0) \end{pmatrix}, \quad M_z = \begin{pmatrix} \cos z & \sin z \\ -\sin z & \cos z \end{pmatrix}, \quad z = h\omega. \tag{4.3}$$

There are at least two reasons for this choice of model. First, if a numerical method already provides unbounded numerical solutions for a given h on this system, we cannot expect good behaviour for more general problems. Second, there are physically relevant problems that, once formulated in appropriate coordinates, are expressed as a system of uncoupled harmonic oscillators. Thus, a precise characterization of the stability of a splitting method on (4.2) can be useful to build accurate and stable algorithms for their numerical treatment. The linear system (3.10) considered in Section 3.4 belongs to this category. This is also the case of the more general equation

$$q' = M^{-1}p, \quad p' = -Nq, \tag{4.4}$$

where M and N are $d \times d$ symmetric positive definite matrices. Writing $M = LL^\top$ and introducing new variables $\tilde{q} = L^\top q$, then $\tilde{q}'' = -L^{-1}NL^{-T}\tilde{q}$. Since N is symmetric positive definite, then $L^{-1}NL^{-T}$ is diagonalizable with positive eigenvalues. A new change of variables reduces the system to a set of d uncoupled scalar harmonic oscillators $y_i'' = -\omega_i^2 y_i$, with ω_i^2 the eigenvalues of $L^{-1}NL^{-T}$ (Bou-Rabee and Sanz-Serna 2018).

Application of the splitting method (3.16) to (4.2) results in the map

$$\begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \tilde{M}_z \begin{pmatrix} q_n \\ p_n \end{pmatrix}, \quad \tilde{M}_z = \begin{pmatrix} K_1(z) & K_2(z) \\ K_3(z) & K_4(z) \end{pmatrix}, \tag{4.5}$$

where $K_1(z)$ and $K_4(z)$ (resp. $K_2(z)$, $K_3(z)$) are even (resp. odd) polynomials in z,

$$\det \tilde{M}_z = 1, \quad K_1(0) = K_4(0) = 1,$$

and, if the scheme is time-symmetric, then $K_1(z) = K_4(z)$. An essential role in the analysis is played by the stability polynomial, defined as

$$p(z) = \frac{1}{2}\operatorname{tr} \tilde{M}_z = \frac{1}{2}(K_1(z) + K_4(z)).$$

The eigenvalues of $\tilde{M}_z$ are the zeros of $\lambda^2 - 2p(z)\lambda + 1$ and they determine the stability of the given method: if z is such that $|p(z)| < 1$, then $\tilde{M}_z$ has complex conjugate eigenvalues of modulus 1 and the powers $\tilde{M}_z^n$, $n \geq 0$, remain bounded, whereas if $|p(z)| > 1$, then $\tilde{M}_z^n$ grows exponentially with n. Linear instability occurs when $+1$ or -1 is an eigenvalue with multiplicity 2 and $\tilde{M}_z$ is not diagonalizable.

Notice that the stability polynomial of a consistent splitting method is an even polynomial satisfying

$$p(z) = 1 - \frac{z^2}{2} + O(z^4) \quad \text{as } z \to 0,$$

so that for sufficiently small $z = h\omega > 0$ the scheme will be stable. The stability

interval is defined as the longest interval $(0, z_*)$ such that $\tilde{M}_z^n$ is bounded for all the iterations n.

In the particular case of the Störmer–Verlet method (3.15), we have

$$K_1(z) = K_4(z) = p(z) = 1 - \frac{z^2}{2}, \quad K_2(z) = z, \quad K_3(z) = -z + \frac{z^3}{4},$$

and thus $z_* = 2$. This is also true for Algorithm 1.4, or $x_{n+1} = \mathrm{e}^{\frac{h}{2}A}\,\mathrm{e}^{hB}\,\mathrm{e}^{\frac{h}{2}A}x_n$, since both have the same stability polynomial $p(z)$.

It is important to stress that for systems of the form (4.4) that can be reduced to a collection of d scalar harmonic oscillators with frequencies ω_i, $i = 1, \ldots, d$, the stability interval of Störmer–Verlet is restricted to $h < 2/\omega_{\max}$, where $\omega_{\max}$ is the largest frequency of the system.

Suppose now that, given an integer k, we concatenate k steps of length h/k of the Störmer–Verlet method to our model problem (4.2). The resulting scheme is stable for $0 < z/k < 2$, or alternatively, its stability interval is $(0, 2k)$. It is remarkable that this is in fact the longest stability interval we can achieve by considering any splitting method (2.17) with $s = k$ stages. An elementary proof of this statement is presented in Bou-Rabee and Sanz-Serna (2018). In consequence, the Störmer–Verlet method may be applied with longer scaled time steps z/k than any other splitting method with k stages. This makes it the method of choice in applications such as molecular dynamics, where high accuracy is not required and we are interested in using time steps as large as possible (Leimkuhler, Reich and Skeel 1996).

The problem of designing splitting methods of order, say, $2r$, with extended stability intervals can be addressed by first determining the coefficients c_j in

$$p(z) = \sum_{j=0}^{r} (-1)^j \frac{z^{2j}}{(2j)!} + \sum_{j=r+1}^{s} c_j z^{2j}, \tag{4.6}$$

so that $p(z)$ has the largest possible value of z_*. Thus a fourth-order integrator with maximal stability interval is presented in López-Marcos *et al.* (1996*b*), whereas in McLachlan and Gray (1997) the analysis is generalized to any order and number of stages. On the other hand, Blanes *et al.* (2008*a*) propose a different strategy to determine the coefficients c_j in (4.6), based on interpolatory conditions and minimization of the difference $(p(z) - \cos z)/z^{2r+2}$ in the stability interval. This results in high-order methods with a large number of stages whose stability and accuracy do not deteriorate for larger values of z. It also allows us to construct very efficient second-order methods for linear systems that outperform high-order methods for a wide range of values of the time step.

Problems of the form

$$q' = M^{-1}p, \quad p' = -Nq + f(q) \tag{4.7}$$

derived from the Hamiltonian function

$$H(q,p) = \frac{1}{2}p^\top M^{-1}p + \frac{1}{2}q^\top Nq + U(q), \tag{4.8}$$

with $f(q) = -\nabla_q U(q)$, also appear frequently in applications. The simple pendulum considered in Section 1.4 belongs to this class. Instead of the usual splitting into kinetic and potential energy, it may be advantageous to split H as $H(q,p) = H_1(q,p) + H_2(q)$, with

$$H_1(q,p) = \frac{1}{2}p^\top M^{-1}p + \frac{1}{2}q^\top Nq, \quad H_2(q) = U(q), \tag{4.9}$$

or alternatively to separate (4.7) as

$$\begin{aligned} q' &= M^{-1}p, \\ p' &= -Nq, \end{aligned} \quad \text{and} \quad \begin{aligned} q' &= 0, \\ p' &= f(q), \end{aligned}$$

and consider Strang integrators

$$S_h^{[RKR]} = \varphi_{h/2}^{[R]} \circ \varphi_h^{[K]} \circ \varphi_{h/2}^{[R]}, \quad S_h^{[KRK]} = \varphi_{h/2}^{[K]} \circ \varphi_h^{[R]} \circ \varphi_{h/2}^{[K]}, \tag{4.10}$$

based on the maps

$$\begin{aligned} \varphi_t^{[R]}&: \begin{pmatrix} q_0 \\ p_0 \end{pmatrix} \longmapsto \mathrm{e}^{tA}\begin{pmatrix} q_0 \\ p_0 \end{pmatrix}, \quad \text{with} \quad A = \begin{pmatrix} 0 & M^{-1} \\ -N & 0 \end{pmatrix}, \\ \varphi_t^{[K]}&: \begin{pmatrix} q_0 \\ p_0 \end{pmatrix} \longmapsto \begin{pmatrix} q_0 \\ p_0 + tf(q_0) \end{pmatrix}. \end{aligned} \tag{4.11}$$

Integrators (4.10) and other splitting methods based on sequences of *rotations* $\varphi_t^{[R]}$ and *kicks* $\varphi_t^{[K]}$ are specially suitable when $f(q)$ is a small perturbation of $-Nq$, since they provide the exact solution when the perturbation vanishes. This happens in particular in the Hamiltonian Monte Carlo method, when we deal with target densities that are perturbations of a Gaussian density (see Section 9.6). Section 5 is devoted to the analysis of splitting methods for this type of system.

In view of the applications, it is relevant to analyse the stability of compositions of $\varphi_t^{[R]}$ and $\varphi_t^{[K]}$. As in the previous case, a sequence of linear transformations render (4.7) into a more simplified form which is used as a model problem. Specifically, the one-dimensional oscillator

$$q'' = -q - \varepsilon q, \quad \varepsilon > -1$$

is the appropriate model here (Bou-Rabee 2017). It turns out that the Strang integrators (4.10) are also optimal concerning stability, in the following sense (Casas, Sanz-Serna and Shaw 2023). Let h_k be the smallest positive root of the equation

$$\frac{kh}{2}\sin\left(\frac{h}{k}\right) = \cos\left(\frac{\pi}{k}\right) - \cos\left(\frac{h}{k}\right).$$

Then, for each fixed $h^* < h_k$, $h^* \neq \pi, 2\pi, \ldots, (k-1)\pi$, the intersection of the stability region in the (ε, h) domain with the line $h = h^*$ is strictly larger for a sequence of k integrators (4.10) than for any other splitting method with k stages based on rotations and kicks. One could say, therefore, that for each given step size h, schemes (4.10) remain stable for larger perturbations than any other splitting method based on rotations and kicks. The value of ε where instabilities arise may be very small indeed, as shown in Casas *et al.* (2023) for the particular case of $k = 3$ stages.

4.3. Modified equations

The concept of *backward error analysis* (BEA), arising in several branches of numerical analysis, has also shown its effectiveness for explaining the good behaviour of splitting and composition methods in long-time integrations (Sanz-Serna 1992, Hairer 1994, Reich 1999). Generally speaking, given a problem $\mathcal{P}$ with true solution $\mathcal{S}$, when a suitable numerical solver is applied, we end up with an approximate solution $\tilde{\mathcal{S}}$. Backward error analysis thus consists in showing that $\tilde{\mathcal{S}}$ is indeed the *exact* solution of a problem $\tilde{\mathcal{P}}$ which is in some sense close to $\mathcal{P}$. This is in contrast to *forward error analysis*, where the aim consists in estimating an appropriate distance between $\tilde{\mathcal{S}}$ and $\mathcal{S}$.

In the domain of numerical analysis of differential equations, what lies at the heart of BEA is the idea of a *modified differential equation*: given the initial value problem $x' = f(x)$, $x(0) = x_0$ and a consistent numerical integrator ψ_h producing the sequence of approximations x_n at $t_n = nh$, $n = 0, 1, \ldots$, we look for another differential equation

$$\tilde{x}' = f_h(\tilde{x}) \tag{4.12}$$

whose vector field is defined as a formal series in powers of h,

$$f_h(\tilde{x}) \equiv f(\tilde{x}) + h f^{[2]}(\tilde{x}) + h^2 f^{[3]}(\tilde{x}) + \cdots \tag{4.13}$$

and such that $x_n = \tilde{x}(t_n)$ (Griffiths and Sanz-Serna 1986, Hairer *et al.* 2006). In this way, by analysing the difference of the vector fields $f(x)$ and $f_h(x)$, it is possible to extract useful information about the qualitative behaviour of the numerical solution and the global error $e_n = x_n - x(t_n) = \tilde{x}(nh) - x(t_n)$.

In the case of a splitting method, obtaining f_h is quite straightforward if we use the BCH formula to get the formal operator associated with the whole method, as was done in Section 2.2 for the order conditions. Thus, for the operator $\Psi(h) = \exp(hF(h))$ associated with a scheme (2.17) of order r, we have

$$F(h) = h(F_1 + F_2) + \sum_{i=1}^{\infty} h^{r+i} \sum_{j=1}^{c_{r+i}} w_{i,j} E_{r+i,j}, \tag{4.14}$$

where $E_{r+i,j}$ denotes the element j of the Lyndon basis of the subspace $\mathcal{L}_{r+i}(F_1, F_2)$ and $w_{i,j}$ are fixed real numbers determined by the actual coefficients of the method.

Since $F(h)$ lies in the same Lie algebra as F_1 and F_2, the numerical solution inherits the properties of the exact flow associated with this feature (e.g. Hamiltonian or volume-preserving). Now, from (4.14), we can easily determine the expression of each $f^{[j]}$ in (4.13).

A similar procedure can be applied to composition methods by determining the formal series $h^{-1}\log\Psi(h)$, with $\Psi(h)$ the operator associated to the scheme.

In the case of a Hamiltonian system of the form $H(q,p) = T(p) + V(q)$, the operators F_1 and F_2 are the Lie derivatives associated with the kinetic and potential energy, respectively, so that (4.14) is itself an operator associated with a modified Hamiltonian $\tilde{H}$. This is yet another reflection of the fact that splitting methods applied to a Hamiltonian system produce maps that are symplectic.

For linear problems, the series defining the modified Hamiltonian $\tilde{H}$ associated with the numerical solution is no longer formal, and $\tilde{H}$ can be explicitly determined in closed form. Thus the matrix $\tilde{M}_z$ in the map (4.5), obtained with a time-symmetric splitting method when $|K_1(z)| < 1$, can be expressed as

$$\tilde{M}_z = \begin{pmatrix} \cos\theta_z & \gamma_z \sin\theta_z \\ -\gamma_z^{-1}\sin\theta_z & \cos\theta_z \end{pmatrix},$$

where θ_z and γ_z are real functions such that $p(z) = \cos\theta_z$ and $K_2(z) = -\gamma_z^2 K_3(z)$ (see (4.5) and the subsequent discussion), and $\theta_{-z} = -\theta_z$, $\gamma_{-z} = -\gamma_z$. It is then straightforward to verify that the map $\tilde{M}_z$ is precisely the z-flow of the modified Hamiltonian (Blanes, Casas and Sanz-Serna 2014, Bou-Rabee and Sanz-Serna 2018)

$$\tilde{H}(q,p) = \frac{\theta_z}{2z}\left(\gamma_z p^2 + \frac{1}{\gamma_z}q^2\right).$$

4.4. *Modified equations and long-term behaviour*

Convergence of the series (4.13) defining the modified equation, apart from the linear case, is the exception rather than the general rule. In consequence, an alternative strategy has to be pursued to get rigorous estimates concerning the long-time behaviour of the numerical solutions. Specifically, we first give bounds on the coefficient functions $f^{[j]}(x)$ of the modified equation, then determine an optimal truncation index, and finally estimate the difference between the numerical solution x_n and the exact solution $\tilde{x}(t_n)$ of the truncated modified equation. Here we summarize only the main results, and refer the reader to Hairer *et al.* (2006), Moan (2002) and references therein for a more comprehensive treatment.

Suppose $f(x)$, $f_1(x)$ and $f_2(x)$ are analytic in a complex neighbourhood of x_0 verifying $\|f(x)\| \le K$ for all $x \in B_{2\rho}(x_0)$, where $B_\rho(x_0) = \{x \in \mathbb{C}^d : \|x - x_0\| \le \rho\}$, and the same is true for the functions $f^{[j]}(x)$ of the modified equation on $B_{\rho/2}(x_0)$. If a suitable truncation index for the formal series (4.13) is selected, so that we have

$$\tilde{x}' = f(\tilde{x}) + hf^{[2]}(\tilde{x}) + h^2 f^{[3]}(\tilde{x}) + \cdots + h^{N-1} f^{[N]}(\tilde{x}), \tag{4.15}$$

with $\tilde{x}(0) = x_0$ and exact flow $\tilde{\varphi}_t^{[N]}$, then there exist constants h_0 with $h \leq h_0/4$ and $\gamma > 0$ such that

$$\left\|\psi_h(x_0) - \tilde{\varphi}_h^{[N]}(x_0)\right\| \leq h\gamma K \mathrm{e}^{-h_0/h}. \tag{4.16}$$

In other words, the difference between the numerical solution $\psi_h(x_0)$ and the exact solution $\tilde{\varphi}_h^{[N]}(x_0)$ of the truncated modified equation (4.15) is exponentially small.

Based on this result it is possible to get some insight into the long-time behaviour of the numerical scheme. Thus, for instance, suppose our splitting method of order r is applied to a Hamiltonian system with step size h. Then the modified equation can be derived from a (truncated) Hamiltonian

$$\tilde{H}(x) = H(x) + h^r H_{r+1}(x) + \cdots + h^{N-1} H_N(x),$$

where now $x = (q, p)$. Letting $\tilde{\varphi}_t^{[N]}$ denote the flow of the truncated modified equation as before, it is clear that $\tilde{H}(\tilde{\varphi}_t^{[N]}(x_0)) = \tilde{H}(x_0)$ for all t. Taking into account (4.16) and the bounds on the functions appearing in the modified equation (derivatives of the $\tilde{H}$ in this case), it follows that

$$\tilde{H}(x_{n+1}) - \tilde{H}\big(\tilde{\varphi}_h^{[N]}(x_n)\big) = O(h\mathrm{e}^{-h_0/h})$$

and

$$\tilde{H}(x_n) = \tilde{H}(x_0) + O(\mathrm{e}^{-h_0/2h}) \quad \text{for } nh \leq \mathrm{e}^{h_0/2h}.$$

If we assume in addition that the numerical solution stays in a compact set $\mathcal{K}$, then $H_{r+1}(x) + \cdots + h^{N-r-1} H_N(x)$ is uniformly bounded on $\mathcal{K}$ independently of h and N (Hairer *et al.* 2006, p. 367) and finally

$$H(x_n) = H(x_0) + O(h^r).$$

Equivalently, the error in the energy corresponding to the numerical solution is of order r over exponentially long time intervals when a splitting method is applied with constant step size in a compact region of the phase space (Moan 2004).

With respect to the behaviour of the error in position, as shown in Calvo and Hairer (1995) and Hairer *et al.* (2006), if the Hamiltonian system is integrable and certain conditions on the frequencies at the initial point are satisfied, then

$$\|(q_n, p_n) - (q(t), p(t))\| \leq Cth^r \quad \text{for } t = nh \leq h^{-r},\ C = \text{const.},$$

that is, the global error grows at most linearly in time, whereas first integrals $I(q, p)$ that only depend on the action variables are well preserved on exponentially long-time intervals,

$$\|I(q_n, p_n) - I(q_0, p_0)\| \leq Ch^r \quad \text{for } t = nh \leq h^{-r}.$$

In contrast, for a non-symplectic method (non-conjugate to a symplectic one) of order r we have

$$H(x_n) - H(x_0) = O(nh^{r+1}) = O(th^r),$$

that is, the error of the energy grows linearly, whereas the global error in the solution typically increases quadratically with time. We have seen illustrations of this feature in Section 1.

It is important to remark that the modified differential equation of a numerical scheme depends explicitly on the step size used, so that if h is changed, then we have a different modified equation. This fact helps to explain the poor long-time behaviour observed in practice when a symplectic scheme is implemented directly with a standard variable step size strategy; see e.g. Calvo and Sanz-Serna (1993*a*).

4.5. *Processing and long-term precision*

The concept of *conjugacy* plays a fundamental role in the study of the long term behaviour of both discrete and continuous dynamical systems. In the context of splitting methods (or more generally, numerical integration methods) for systems of ODEs, we replace the h-flow φ_h of the original system with a map ψ_h depending on the small parameter h (the step size) such that $\psi_h \approx \varphi_h$ for h small enough. The precision of the numerical approximations $x_n = \psi_h^n(x_0) \approx \varphi_h^n(x_0) = x(nh)$ can be analysed in one of two ways.

(i) We use standard techniques of numerical integration of ODEs to estimate the local error $\|\psi_h(x) - \varphi_h(x)\|$ and then study how this local error is propagated (Hairer *et al.* 1993),

(ii) As described in the previous subsection, we consider a truncated modified equation (4.15) of the numerical integration map ψ_h, and then study the effect of replacing the original system with the modified one, in addition to the propagation of the modified local error (4.16) (the local error between the map ψ_h and the h-flow $\varphi_t^{[N]}$ of the truncated modified equation).

In both cases, a better understanding of the long-term behaviour of the numerical error (of the application of a given integration scheme with constant step size) can be obtained by combining such techniques with the idea of *processing* a numerical integrator (López-Marcos, Sanz-Serna and Skeel 1996*a*). The main idea consists in analysing how close ψ_h and φ_h are to being conjugate to each other, and using that to estimate the long-term evolution of the errors. This is closely related to the concept of effective order (Butcher 1969, Butcher and Sanz-Serna 1996) and the idea of enhancing numerical integrators with correctors (Wisdom *et al.* 1996). Essentially, the procedure is as follows.

- Given ψ_h, we find a near-identity conjugacy map $\pi_h : \mathbb{R}^D \to \mathbb{R}^D$ (i.e. π_0 is the identity map) such that $\hat{\psi}_h := \pi_h^{-1} \circ \psi_h \circ \pi_h$ is as close as possible to the h-flow φ_h of the original system of ODEs.
- We estimate the propagated error $\|\psi_h^n(x_0) - \varphi_h^n(x_0)\|$ as

$$\|\psi_h^n(x_0) - \varphi_h^n(x_0)\| \le \|\psi_h^n(x_0) - \hat{\psi}_h^n(x_0)\| + \|\hat{\psi}_h^n(x_0) - \varphi_h^n(x_0)\|.$$

- We analyse the propagated error $\|\hat{\psi}_h^n(x_0) - \varphi_h^n(x_0)\|$ of the *processed integrator* $\hat{\psi}_h$, either by standard techniques or with the modified equation of $\hat{\psi}_h$.
- We estimate the difference between the original numerical solution $x_n = \psi_h^n(x_0)$ and the numerical approximation $\hat{x}_n = \hat{\psi}_h^n(x_0)$ that would be obtained if the processed integrator were used instead of the original integration map ψ_h. More precisely, using the notation

  $$x_n = \psi_h^n(x_0), \quad \bar{x}_n = \psi_h^n(\pi_h(x_0)), \quad \hat{x}_n = \hat{\psi}_h^n(x_0) = \pi_h^{-1}(\bar{x}_n), \quad x(nh) = \varphi_h^n(x_0),$$

  for $n \geq 0$, we have

  $$\|\hat{x}_n - x_n\| \leq \|\pi_h^{-1}(\bar{x}_n) - \bar{x}_n\| + \|\bar{x}_n - x_n\|.$$

Summing up, the propagated error of the integrator ψ_h can be estimated as

$$\|x_n - x(nh)\| \leq \|\pi_h^{-1}(\bar{x}_n) - \bar{x}_n\| + \|\bar{x}_n - x_n\| + \|\hat{x}_n - x(nh)\|. \tag{4.17}$$

If the original method is of order r, that is, $\psi_h(x) - \varphi_h(x) = O(h^{r+1})$ as $h \to 0$, then it makes sense to choose a conjugacy map satisfying $\pi_h(x) = x + O(h^r)$. Hence, provided that the sequence $\{\bar{x}_n\}$ stays in a compact set, we can see that $\|\hat{x}_n - \bar{x}_n\| = \|\pi_h^{-1}(\bar{x}_n) - \bar{x}_n\| = O(h^r)$ with a constant independent of n. Therefore, for sufficiently large time intervals, the right-hand side of (4.17) will be dominated either by $\|\bar{x}_n - x_n\| = \|\psi_h^n(\bar{x}_0) - \psi_h^n(x_0)\|$ (the propagation along successive iterations of ψ_h of a perturbation of size $O(h^r)$ in x_0) or by $\|\hat{x}_n - x(nh)\| = \|\hat{\psi}_h^n(x_0) - \varphi_h^n(x_0)\|$ (the sum of the propagated local errors of the processed method $\hat{\psi}_h$). Typically, the latter dominates over the former if the integration interval is sufficiently large. In that case, the precision of the numerical scheme ψ_h for sufficiently long-term integrations will depend on the size of the local errors $\|\pi_h^{-1} \circ \psi_h(x) \circ \pi_h - \varphi_h(x)\|$ of the processed method for an appropriately chosen conjugacy map (or processor map) π_h, rather than on the local errors $\|\psi_h(x) - \varphi_h(x)\|$ of the method itself.

This is illustrated by the evolution of the error in phase space of LT and S_2 displayed in Figures 1.2 and 1.4 for the pendulum problem and the six-body problem, respectively. Recall that the Lie–Trotter method is conjugate to the Strang splitting (see Section 1.3). In this case, $x_n = (q_n, p_n)^\top$ is the numerical solution provided by LT, whereas $\hat{x}_n = (\hat{q}_n, \hat{p}_n)^\top$ corresponds to S_2, with the processor map $\pi_h = \varphi_{h/2}^{[T]}$. In both examples, $\|\pi_h^{-1}(\bar{x}_n) - \bar{x}_n\| = O(h)$, which is bounded for all n provided that $\|\bar{p}_n\|$ remains bounded.

For the pendulum problem, the error $\|\bar{x}_n - x_n\| = \|\psi_h^n(\bar{x}_0) - \psi_h^n(x_0)\|$ due to the propagation of the initial difference $\|(\bar{q}_0 - q_0, \bar{p}_0 - p_0)\| = O(h)$ does not exhibit any significant increment, because (for the considered initial value) the pendulum behaves as a perturbed harmonic oscillator. On the other hand, the global error $\|\hat{x}_n - x(nh)\|$ of S_2 behaves as $O(th^2)$. Therefore the global error $\|x_n - x(nh)\|$ of LT is dominated at the beginning of the integration interval by $\|\hat{x}_n - x_n\| \approx \|\bar{x}_n - x_n\|$ (with no clear growth over time), until it is overcome by the linearly increasing

global error of S_2, resulting in errors of similar size at the end of the interval for LT and S_2.

For the six-body problem, the propagation error $\|\bar{x}_n - x_n\| = \|\psi_h^n(\bar{x}_0) - \psi_h^n(x_0)\|$ increases linearly, because now H_1 corresponds to a collection of Keplerian problems, where perturbations in initial states are propagated linearly. However, the slope of that linear increase is smaller than that of the propagation of the global error $\|\hat{x}_n - x(nh)\|$ of S_2. Therefore, the global error for LT is dominated by $\|\hat{x}_n - x_n\|$ during most of the integration interval in Figure 1.4, and only at the end does it reach the global error of S_2. For longer times (not shown there), the global errors of LT and S_2 will be of similar size.

So far we have focused on studying the long-term performance of a given splitting method ψ_h with the help of a conjugacy map π_h. In practice, we may actually enhance the performance of a given splitting method (Rowlands 1991, Wisdom *et al.* 1996, McLachlan 1996, Laskar and Robutel 2001), effectively integrating the problem with the processed integration map $\hat{\psi}_h = \pi_h^{-1} \circ \psi_h(x) \circ \pi_h$. Indeed, if output is needed only every m steps, the computation of

$$\hat{x}_{n,m} = \pi_h^{-1} \circ \psi_h^m \circ \pi_h(\hat{x}_{(n-1),m}), \quad n = 1, 2, 3, \ldots \tag{4.18}$$

(with $\hat{x}_0 = x_0$), will not require substantially more CPU time than computing

$$x_{n,m} = \psi_h^m(x_{(n-1),m}), \quad n = 1, 2, 3, \ldots,$$

provided that the evaluation of $\pi_h(x)$ is computationally cheap compared to m evaluations of $\psi_h(x)$. Moreover, even if frequent output is required, we might approximate $x(nh)$ by

$$\bar{x}_n = \psi_h^m(\bar{x}_{n-1}), \quad n = 1, 2, 3, \ldots,$$

with $\bar{x}_0 = \pi_h(x_0)$. For sufficiently long integrations, this will cost essentially the same as applying the original integrator ψ_h in a standard way, and will be nearly as accurate as $\hat{x}_n$ (the full application of the processed integrator), since $\|\bar{x}_n - \hat{x}_n\| = \|\pi_h^{-1}(\bar{x}_n) - \bar{x}_n\| = O(h^r)$ will be negligible compared to $\|\bar{x}_n - x(nh)\|$ for n large enough.

This is again illustrated in Figure 1.4 for the six-body problem written in Jacobi coordinates as a perturbation of Keplerian problems. Indeed, if x_n is the numerical solution labelled by $\mathrm{LT}_{\mathrm{pert}}$ and $\pi_h = \varphi_{h/2}^{[H_1]}$, then $\hat{x}_n$ corresponds to scheme $(2, 2)$ while $\mathrm{pLT}_{\mathrm{pert}}$ stands for $\bar{x}_n$. Notice that the position error of $\mathrm{pLT}_{\mathrm{pert}}$ is very similar to that of $(2, 2)$, as expected from the preceding discussion.

The enhancement of splitting integrators by processing is particularly effective for problems of the form $x' = f_1(x) + \varepsilon f_2(x)$ with $|\varepsilon| \ll 1$. Such enhancing was first considered in Wisdom *et al.* (1996) in the context of N-body problems modelling planetary systems. Several processors π_h were constructed for the Strang method, leading to processed (corrected) methods $\hat{\psi}_h$ of generalized order $(2k, 2)$ for several $k > 1$ (see also McLachlan 1996, Laskar and Robutel 2001), so that their local errors are $O(\varepsilon h^{2k+1} + \varepsilon^2 h^3)$ for a prescribed time integration.

In Section 5.7 we construct a different processor map that effectively reduces the local error to $O(\varepsilon^2 h^3)$ in the case where f_1 is derived from a Hamiltonian $H_1(q,p)$ of a harmonic oscillator (or a collection of harmonic oscillators whose frequencies are integer multiples of a basic frequency) and f_2 comes from a Hamiltonian $\varepsilon H_2(q)$ which is a polynomial in q. This means, going back to the results displayed in Figure 1.2 for the pendulum problem, that both methods $\mathrm{LT}_{\mathrm{pert}}$ and $(2,2)$ are conjugate to a more accurate integrator $\hat{\psi}_h$ (with global error of order $O(t\varepsilon^2 h^2)$) obtained from them with an appropriate processor map π_h.

Assume that, in the previous notation, $(2,2)$ (resp. $\mathrm{LT}_{\mathrm{pert}}$) corresponds to the numerical solution x_n, and $\hat{x}_n$ to the processed method. Then the error (4.17) is dominated by $\|\bar{x}_n - x_n\|$, which does not show secular growth in this case. Eventually, for long enough integration intervals, (4.17) will be dominated by the linearly increasing error $\|\hat{x}_n - x(nh)\|$ of the more accurate processed integrator.[4] The situation is very similar for the evolution of errors in position displayed for the six-body problem in Figure 1.4, the only difference being that the propagation $\|\bar{x}_n - x_n\|$ of differences in initial values now grows linearly. The error $\|\hat{x}_n - x(nh)\|$ will eventually dominate in (4.17) because $\|\bar{x}_n - x_n\|$ increases linearly with a smaller slope than $\|\hat{x}_n - x(nh)\|$.

Instead of enhancing a previously existing rth-order integrator by processing, we may also design a processed splitting method from scratch (López-Marcos *et al.* 1997, Blanes, Casas and Ros 1999*b*): determine the h-parametric maps ψ_h and π_h as compositions of basic flows $\varphi^{[1]}_{a_j h}$ and $\varphi^{[2]}_{b_j h}$ with different sequences of coefficients a_j and b_j, such that the processed map $\hat{\psi}_h = \pi_h^{-1} \circ \psi_h \circ \pi_h$ is a good approximation of φ_h for sufficiently small step sizes h. Typically, we require that $\pi_h^{-1} \circ \psi_h \circ \pi_h(x) - \varphi_h(x) = O(h^{r+1})$, so that the processed integrator is of order r. In that case, if we intend to compute the approximations (4.18) of $x(km)$ (for $k = 1, 2, 3 \ldots$), there is no need for the map ψ_h (referred to in this context as the kernel) to be an rth order approximation of φ_h.

In any of the situations considered above (either analysing the performance of a given splitting integrator with the help of a conjugacy map π_h, or enhancing an existing rth-order splitting integrator by processing, or designing a splitting processing integrator from scratch), we need to study the *effective order conditions* of ψ_h. These are the conditions on the parameters a_i, b_i that guarantee that there exists a processor map π_h such that $\pi_h^{-1} \circ \psi_h \circ \pi_h(x) - \varphi_h(x) = O(h^{r+1})$. A general treatment of the effective order conditions of several classes of numerical integrators including splitting methods and composition methods can be found in Blanes, Casas and Murua (2004, 2006*a*). That treatment is based on the series expansion (2.21) of the formal logarithm of the Lie transformation $\Psi(h)$, and it is shown that the conditions for effective order r can be written in terms of the coefficients w_1, w_2, w_{12}, $w_{122}, \ldots$ featuring in (2.21). In addition, the formal

[4] This can also be checked in Figure 5.1, where the same problem is integrated with scheme $(2,2)$ and processed Strang with a larger time step and a longer time interval.

logarithm of the Lie transformation $P(h)$ of the map π_h is determined as

$$\log(P(h)) = h(p_1F_1 + p_2F_2) + h^2 p_{12}F_{12} + h^3(p_{122}F_{122} + p_{112}F_{112}) \\ + h^4(p_{1222}F_{1222} + p_{1122}F_{1122} + p_{1112}F_{1112}) + \cdots + O(h^{r+1}),$$

with the coefficients $p_{\ell_1\cdots\ell_k}$ given as polynomials of the coefficients in (2.21). Alternative ways of obtaining the effective order conditions of splitting methods can be derived following the different approaches considered in Section 2 for analysing the standard order conditions. However, this is beyond the scope of the present work. In any case, the analysis shows that many of the order conditions of the processed method $\hat{\psi}_h$ can be satisfied by π_h, so that ψ_h must fulfil a greatly reduced set of restrictions, also of lower complexity. As a result, it is possible to construct processed schemes as compositions of basic maps with a reduced computational cost.

5. Highly oscillatory problems

In this section we consider Hamiltonian systems of the form (4.8) with the splitting (4.9), that is,

$$H = H_1 + H_2, \quad H_1(q,p) = \frac{1}{2}p^\top M^{-1}p + \frac{1}{2}q^\top N\, q, \quad H_2(q,p) = U(q), \tag{5.1}$$

when M and N are real symmetric matrices, and $U\colon \mathbb{R}^d \to \mathbb{R}$ is a polynomial function. The corresponding equations of motion (4.7) can be rewritten as

$$x' = f_1(x) + f_2(x), \quad \text{with} \quad f_1(x) = Ax,\ f_2(x) = \begin{pmatrix} 0 \\ -\nabla U(q) \end{pmatrix} \tag{5.2}$$

in terms of $x = (q,p)^\top$ and the matrix A given in (4.11). Splitting methods are advantageous for system (5.1) provided $\mathrm{e}^{tA}x$ can be cheaply computed for each $x \in \mathbb{R}^D$, $D = 2d$.

For the time being, we assume that A is fully diagonalizable and the eigenvalues of A are integer multiples of $\omega\,\mathrm{i}$ (with i the imaginary unit). (The more general case where the eigenvalues of A lie on the imaginary axis will be considered in Section 5.8.) This implies that e^{tA} is $2\pi/\omega$-periodic in t. For system (5.1), the present assumption is equivalent to stating that the matrix $M^{-1}N$ is fully diagonalizable with all its eigenvalues of the form $-(\omega k)^2$, with $k \in \mathbb{Z}$. In other words, H_1 in (5.1) is just a collection of harmonic oscillators whose frequencies are integer multiples of a basic frequency ω.

Splitting methods applied to (5.2) can be analysed by considering series expansions in powers of h and using standard tools, in particular the material presented in the previous sections. Thus Section 3.3 is particularly relevant if the basic frequency ω is large compared to the size of the potential $U(q)$ (or more generally, the size of the components of $f_2(x)$). Indeed, rescaling time from t to $\tau = \omega\, t$, system

(5.2) is transformed into

$$\frac{\mathrm{d}}{\mathrm{d}\tau}x = \hat{f}_1(x) + \varepsilon\, f_2(x), \qquad \hat{f}_1(x) = \hat{A}x,$$

where $\varepsilon = 1/\omega$ and $\hat{A} = \varepsilon A$, so that all the eigenvalues of $\hat{A}$ are integer multiples of the imaginary unit i. This approach nevertheless has an important limitation: it is meaningful only when $\omega h = h/\varepsilon$ is sufficiently small.

Different approaches have been adopted in the literature to overcome this limitation and obtain results that remain valid when ωh is large (provided that the step size h is small enough compared to the size of the perturbing potential $U(q)$ and its partial derivatives). Among them, we mention modulated Fourier expansions (see Hairer *et al.* 2006 and references therein) and extended word series (Murua and Sanz-Serna 2017, 2016). Extended word series were introduced in Murua and Sanz-Serna (2017) to analyse splitting methods for a class of problems that is equivalent to (5.2) under the more general assumption that all the eigenvalues of A lie on the imaginary axis. Such expansions were further used in Murua and Sanz-Serna (2016) to analyse normal forms and formal invariants of more general classes of problems. The formalism of extended word series allows us to work with asymptotic expansions valid for step sizes h that are sufficiently small independently of the frequencies of e^{tA}.

In the present section we provide an elementary derivation of second-order versions in h of such expansions, taking the Strang splitting as a case study. In particular, we provide a theoretical justification for the results presented in Section 1.4 for the simple pendulum and analyse the processing technique as a way to further improve those results, which are indeed valid for the general system (5.1). This is done by first constructing an asymptotic expansion of the exact solution of (5.2) and then comparing with the expansion corresponding to the numerical approximation obtained by a general splitting method. We also get the modified equation satisfied by the numerical scheme (exact up to terms in h^3), with explicit formulas for the coefficients, and the corresponding modified Hamiltonian. Furthermore, we explicitly construct a processor for the Strang splitting so that the resulting scheme leads to a better approximation to the solution of (5.2) (in particular, with a better preservation of the energy H for large time intervals). Finally, we indicate how the preceding results can be generalized to the more general case where the eigenvalues of A lie on the imaginary axis (so that, in general, e^{tA} is quasi-periodic in t).

For the analysis it is convenient to reformulate the problem (5.2) into the new variables $y(t)$ given through $x(t) = \mathrm{e}^{tA}y(t)$, so that now $y' = \mathrm{e}^{-tA}f_2(\mathrm{e}^{tA}y)$. Let us write

$$\mathrm{e}^{-t\,A}f_2(\mathrm{e}^{t\,A}x) = \sum_{k\in\mathbb{Z}} \mathrm{e}^{\mathrm{i}k\omega t}\, g_k(x), \tag{5.3}$$

that is, the right-hand side of (5.3) is the Fourier series expansion of $\mathrm{e}^{-t\,A}f_2(\mathrm{e}^{t\,A}x)$.

The map f_2 being real implies that $g_k : \mathbb{R}^D \to \mathbb{C}^D$ is such that each component of $g_{-k}(x)$ is the complex conjugate of the corresponding component of $g_k(x)$.

Since we have assumed that $U(q)$ is a polynomial, each component of $f_2(x)$ is also a polynomial in the variables x, which guarantees that there is a finite number of non-zero terms in (5.3). We will use the notation

$$\mathcal{I} = \{k \in \mathbb{Z} \colon g_k \neq 0\}. \tag{5.4}$$

Remarks.

- The assumption that $U(q)$ and each component of $f_2(x)$ are polynomials in the variables x might seem too restrictive. However, for more general assumptions (e.g. real-analyticity), we can always replace $U(q)$ and $f_2(x)$ with sufficiently accurate polynomial approximations. Furthermore, the material in the present section is formally valid for more general assumptions on $U(q)$ and $f_2(x)$ if we allow the set (5.4) to be infinite. In that case, (5.3) will be an infinite series, several of the formulae derived here will also involve infinite series, and appropriate assumptions should be made so as to guarantee convergence.
- The assumptions that A is a real matrix and that $f_2 \colon \mathbb{R}^D \to \mathbb{R}^D$ are not essential. We could consider a complex matrix A and $f_2 \colon \mathbb{C}^D \to \mathbb{C}^D$, with no changes in the formulae that follow.

By substitution of $t = 0$ into (5.3), we get $f_2(x) = \sum_{k\in\mathcal{I}} g_k(x)$. Thus, the solution of the initial value problem defined by (5.2) and $x(0) = x_0 \in \mathbb{R}^D$ can be expressed as $x(t) = \mathrm{e}^{tA} y(t)$, where $y(t)$ is the solution of

$$\frac{\mathrm{d}}{\mathrm{d}t} y = \sum_{k\in\mathcal{I}} \mathrm{e}^{\mathrm{i}k\omega t} g_k(y), \quad y(0) = x_0. \tag{5.5}$$

From the definition of the Fourier coefficients g_k in (5.3), we can prove that

$$\mathrm{e}^{-tA} g_k(\mathrm{e}^{tA} x) = \mathrm{e}^{\mathrm{i}k\omega t}\, g_k(x), \quad \text{for } k \in \mathcal{I}, \tag{5.6}$$

and, by applying the operator $(\mathrm{d}/\mathrm{d}t)|_{t=0}$ to both sides, this is equivalent to (Murua and Sanz-Serna 2016)

$$(f_1, g_k) = \mathrm{i}k\,\omega\, g_k, \quad \text{for } k \in \mathcal{I}. \tag{5.7}$$

Here, $(\cdot,\cdot)$ represents the usual Lie–Poisson bracket already defined in Section 1: $(f_1, g_k)(x) = g_k'(x) f_1(x) - f_1'(x) g_k(x)$.

5.1. Expansion of the exact solution

We next obtain an approximate representation of the flow $\varphi_h^{[f_1+f_2]}$ of (5.2) with initial condition $x_0 \in \mathbb{R}^D$ that is valid for sufficiently small values of $|h|$ independently of the basic frequency ω. This is done by getting an expansion of the solution $y(t)$ of (5.5) valid for $|t| \leq h$. To begin with, we apply the substitution $y(t) = x_0 + O(t)$ on

the right-hand side of the integral form

$$y(t) = x_0 + \sum_{k\in\mathcal{I}} \int_0^t \mathrm{e}^{\mathrm{i}k\omega s} g_k(y(s))\,\mathrm{d}s$$

of the initial value problem. This gives

$$y(t) = x_0 + \sum_{k\in\mathcal{I}} \int_0^t \mathrm{e}^{\mathrm{i}k\omega s} g_k(x_0)\,\mathrm{d}s + O(t^2). \tag{5.8}$$

Furthermore, $y(h)$ admits the estimate

$$y(h) = x_0 + \sum_{\ell\in\mathcal{I}} \int_0^h \mathrm{e}^{\mathrm{i}\ell\omega t}(g_\ell(x_0) + g'_\ell(x_0)(y(t) - x_0))\,\mathrm{d}t + O(h^3), \tag{5.9}$$

where $g'_\ell(x_0)$ is the value at $x = x_0$ of the Jacobian matrix of $g_\ell(x)$.

Substitution of (5.8) into the right-hand side of (5.9) finally leads to

$$\begin{aligned} y(h) = x_0 &+ \sum_{k\in\mathcal{I}} \left(\int_0^h \mathrm{e}^{\mathrm{i}k\omega t}\,\mathrm{d}t \right) g_k(x_0) \\ &+ \sum_{k,\ell\in\mathcal{I}} \left(\int_0^h \int_0^t \mathrm{e}^{\mathrm{i}(\ell t + ks)\omega}\,\mathrm{d}s\,\mathrm{d}t \right) g'_\ell(x_0) g_k(x_0) + O(h^3). \end{aligned}$$

Equivalently,

$$y(h) = x_0 + h \sum_{k\in\mathcal{I}} \alpha_k(h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \alpha_{k\ell}(h) g'_\ell(x_0) g_k(x_0) + O(h^3), \tag{5.10}$$

where the coefficients are defined as follows.

- For $k \in \mathcal{I}$,

$$\alpha_k(h) = \int_0^1 \mathrm{e}^{\mathrm{i}k\omega h\tau}\,\mathrm{d}\tau = \begin{cases} 1 & \text{if } k = 0, \\ \dfrac{\mathrm{e}^{\mathrm{i}k\omega h} - 1}{\mathrm{i}k\omega h} & \text{otherwise.} \end{cases} \tag{5.11}$$

- For $k, \ell \in \mathcal{I}$,

$$\alpha_{k\ell}(h) = \int_0^1 \int_0^{\tau_2} \mathrm{e}^{\mathrm{i}h\omega(k\tau_1 + \ell\tau_2)}\,\mathrm{d}\tau_1\,\mathrm{d}\tau_2. \tag{5.12}$$

  Specifically, $\alpha_{00}(h) = 1/2$, and the following recursions hold:

$$\alpha_{0k}(h) = \frac{\mathrm{e}^{\mathrm{i}k\omega h} - \alpha_k(h)}{\mathrm{i}k\omega h}, \quad \alpha_{k\ell}(h) = \frac{\alpha_{k+\ell}(h) - \alpha_\ell(h)}{\mathrm{i}k\omega h} \quad \text{for } k, \ell \in \mathcal{I}\backslash\{0\}. \tag{5.13}$$

- The constant in the $O(h^3)$ term depends on upper bounds of the norm of g_k and its partial derivatives, but is independent of the basic frequency ω.

From this we conclude that

$$\varphi_h(x_0) = \mathrm{e}^{hA}\left(x_0 + h\sum_{k\in\mathcal{I}} \alpha_k(h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \alpha_{k\ell}(h) g'_\ell(x_0) g_k(x_0) + O(h^3)\right), \tag{5.14}$$

where we have dropped the upper index in φ_h for clarity. Observe that

$$|\alpha_k(h)| = |\mathrm{sinc}(kh\omega/2)| \le 1, \quad |\alpha_{k\ell}(h)| \le \frac{1}{2}.$$

5.2. *Expansion of the discrete solution given by Strang splitting*

We next proceed to construct an analogous expansion for the (discrete) solution furnished by the Strang splitting given in (4.10), namely

$$S_h^{[RKR]} = \varphi_{h/2}^{[R]} \circ \varphi_h^{[K]} \circ \varphi_{h/2}^{[R]},$$

based on the maps (4.11). In fact, it is straightforward to check that the approximation $\tilde{x}(h) = S_h^{[RKR]}(x_0)$ satisfies $\tilde{x}(h) = \mathrm{e}^{hA}\tilde{y}(h)$, where $\tilde{y}(t)$ is the solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{y} = \mathrm{e}^{-\frac{h}{2}A} g(\mathrm{e}^{\frac{h}{2}A}\tilde{y}(t)) = \sum_{k\in\mathcal{I}} \mathrm{e}^{\mathrm{i}k\omega h/2}\, g_k(\tilde{y}(t)), \quad \tilde{y}(0) = x_0.$$

In consequence, proceeding as in the previous subsection, we get

$$\tilde{y}(h) = x_0 + h\sum_{k\in\mathcal{I}} \tilde{\alpha}_k(h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \tilde{\alpha}_{k\ell}(h) g'_\ell(x_0) g_k(x_0) + O(h^3),$$

where for $k, \ell \in \mathcal{I}$,

$$\tilde{\alpha}_k(h) = \mathrm{e}^{\mathrm{i}k\omega h/2}, \quad \tilde{\alpha}_{k\ell}(h) = \frac{1}{2}\mathrm{e}^{\mathrm{i}(k+\ell)\omega h/2}, \tag{5.15}$$

and finally

$$S_h(x_0) = \mathrm{e}^{hA}\left(x_0 + h\sum_{k\in\mathcal{I}} \tilde{\alpha}_k(h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \tilde{\alpha}_{k\ell}(h) g'_\ell(x_0) g_k(x_0) + O(h^3)\right), \tag{5.16}$$

where we have also dropped the upper index in S_h. Notice that, from (5.15),

$$|\tilde{\alpha}_k(h)| \le 1, \quad |\tilde{\alpha}_{k\ell}(h)| \le \frac{1}{2}$$

and

$$\tilde{\alpha}_{k\ell}(h) + \tilde{\alpha}_{\ell k}(h) = \tilde{\alpha}_k(h)\tilde{\alpha}_\ell(h), \quad k, \ell \in \mathcal{I}. \tag{5.17}$$

5.3. *Composition formulae*

If we are interested in extending the previous analysis to more general splitting methods of the form (2.17), then a composition rule concatenating the expansions

corresponding to different basic methods is clearly needed. This can be obtained as follows.

Suppose the map $\psi_h : \mathbb{R}^D \to \mathbb{R}^D$ can be expanded as

$$\psi_h(x) = x + h \sum_{k\in\mathcal{I}} \kappa_k\, g_k(x) + h^2 \sum_{k,\ell\in\mathcal{I}} \kappa_{k\ell}\, g'_\ell(x) g_k(x) + O(h^3), \tag{5.18}$$

for some coefficients $\kappa_k, \kappa_{k,\ell}$. Then, from (5.6), we have for all $s \in \mathbb{R}$ that

$$\begin{aligned}\psi_h(\mathrm{e}^{sA}x) = \mathrm{e}^{sA}\Bigg(x + h \sum_{k\in\mathcal{I}} \mathrm{e}^{\mathrm{i}k\omega s}\kappa_k\, g_k(x) \\ + h^2 \sum_{k,\ell\in\mathcal{I}} \mathrm{e}^{\mathrm{i}(k+\ell)\omega s}\kappa_{k\ell}\, g'_\ell(x) g_k(x) + O(h^3)\Bigg).\end{aligned} \tag{5.19}$$

If in addition the map $\hat{\psi}_h : \mathbb{R}^D \to \mathbb{R}^D$ can be expanded as

$$\hat{\psi}_h(x) = x + h \sum_{k\in\mathcal{I}} \hat{\kappa}_k\, g_k(x) + h^2 \sum_{k,\ell\in\mathcal{I}} \hat{\kappa}_{k\ell}\, g'_\ell(x) g_k(x) + O(h^3),$$

then we have the following expression for the composition $\psi_h \circ \hat{\psi}_h$:

$$\begin{aligned}\psi_h(\hat{\psi}_h(x)) = x + h \sum_{k\in\mathcal{I}} (\kappa_k + \hat{\kappa}_k)\, g_k(x) \\ + h^2 \sum_{k,\ell\in\mathcal{I}} (\kappa_{k\ell} + \hat{\kappa}_k\, \kappa_\ell + \hat{\kappa}_{k\ell})\, g'_\ell(x) g_k(x) + O(h^3).\end{aligned}$$

This is also valid if the coefficients κ_k, $\hat{\kappa}_k$, $\kappa_{k\ell}$ and $\hat{\kappa}_{k\ell}$ depend on h, although in that case the constant in the $O(h^3)$ term will also depend on the bounds of the coefficients.

By combining the previous results we finally arrive at the following composition rule:

$$\mathrm{e}^{sA}\psi_h\left(\mathrm{e}^{\hat{s}A}\hat{\psi}_h(x)\right) = \mathrm{e}^{(s+\hat{s})A}\left(x + h \sum_{k\in\mathcal{I}} \gamma_k\, g_k(x) + h^2 \sum_{k,\ell\in\mathcal{I}} \gamma_{k\ell}\, g'_\ell(x) g_k(x) + O(h^3)\right), \tag{5.20}$$

where

$$\begin{aligned}\gamma_k &= \mathrm{e}^{\mathrm{i}\hat{s}\omega k}\,\kappa_k + \hat{\kappa}_k, \\ \gamma_{k\ell} &= \mathrm{e}^{\mathrm{i}\hat{s}\omega(k+\ell)}\,\kappa_{k\ell} + \mathrm{e}^{\mathrm{i}\hat{s}\omega\ell}\,\hat{\kappa}_k\,\kappa_\ell + \hat{\kappa}_{k\ell}.\end{aligned} \tag{5.21}$$

5.4. *Expansions for arbitrary splitting methods*

We now have all the required ingredients to extend the expansion (5.16) for Strang to a more general splitting method of the form (2.17) based on kicks and rotations, i.e. on the maps (4.11). Specifically, if ψ_h denotes such a splitting, then $\psi_h(x_0)$ can

be expanded as

$$\mathrm{e}^{h(a_1+\cdots+a_{s+1})A}\left(x_0 + h\sum_{k\in\mathcal{I}} \tilde{\alpha}_k(h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \tilde{\alpha}_{k\ell}(h) g'_\ell(x_0) g_k(x_0) + O(h^3)\right) \tag{5.22}$$

with the following coefficients:

- for $k \in \mathcal{I}$,

$$\tilde{\alpha}_k(h) = \sum_{j=1}^{s} b_j\, \mathrm{e}^{\mathrm{i}kc_j\omega h}, \quad \text{with} \quad c_j = a_1 + \cdots + a_j, \quad \text{for } j = 1, \ldots, s; \tag{5.23}$$

- for $k, \ell \in \mathcal{I}$,

$$\tilde{\alpha}_{k\ell}(h) = \sum_{1\le j<m\le s} b_j b_m\, \mathrm{e}^{\mathrm{i}(c_j k + c_m \ell)\omega h} + \sum_{1\le j\le r} \frac{1}{2} b_j^2\, \mathrm{e}^{\mathrm{i}c_j(k+\ell)\omega h}, \tag{5.24}$$

so that they again satisfy the relation (5.17), since it is preserved under compositions (5.20)–(5.21).

We can now estimate the local error of a consistent splitting method by taking into account (5.22) and (5.14) as

$$\psi_h(x) - \varphi_h(x) = \mathrm{e}^{hA}\left(x_0 + h\sum_{k\in\mathcal{I}} \delta_k(h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \delta_{k\ell}(h) g'_\ell(x_0) g_k(x_0) + O(h^3)\right),$$

where

$$\begin{aligned} \delta_k(h) &= \sum_{j=1}^{s} b_j\, \mathrm{e}^{\mathrm{i}kc_j\omega h} - \int_0^1 \mathrm{e}^{\mathrm{i}k\omega h\tau}\, \mathrm{d}\tau, \\ \delta_{k\ell}(h) &= \sum_{1\le j<m\le s} b_j b_m\, \mathrm{e}^{\mathrm{i}(c_j k + c_m \ell)\omega h} + \sum_{1\le j\le r} \frac{1}{2} b_j^2\, \mathrm{e}^{\mathrm{i}c_j(k+\ell)\omega h} \\ &\quad - \int_0^1 \int_0^{\tau_2} \mathrm{e}^{\mathrm{i}h\omega(k\tau_1 + \ell\tau_2)}\, \mathrm{d}\tau_1\, \mathrm{d}\tau_2. \end{aligned} \tag{5.25}$$

It is worth remarking that if we expand the exponentials in (5.25) in series of powers of h up to a certain degree r, then we recover the order conditions obtained in Section 2.4 for multi-indices with one and two indices. In other words, using the terminology introduced in Section 3.3 to analyse perturbed problems of the form (3.8), we obtain the conditions for a splitting method to be of generalized order $(r_1, r_2, 1)$ (or $(r_1, r_2, 2)$ in the case of time-symmetric splitting methods). Notice, however, that replacing these exponentials with such truncated series expansions will only give useful information about the size of the local error coefficients $\delta_k(h)$ and $\delta_{k\ell}(h)$ provided that the scaled step sizes $\{|k|\omega h\colon k \in \mathcal{I}\}$ are sufficiently small.

5.5. *Modified ODE for the discrete flow furnished by a splitting method*

At this point it is useful to formulate a modified differential equation whose h-flow is closer in a certain sense to the map ψ_h corresponding to a general splitting method admitting an expansion of the form (5.22). More precisely, the idea is to construct a modified ODE whose h-flow is $O(h^3)$ close to ψ_h.

From the composition formula (5.20), the nth iterate of the map ψ_h admits an expansion of the form

$$\psi_h^n(x_0) = \mathrm{e}^{hnA}\left(x_0 + h\sum_{k\in\mathcal{I}} \gamma_k(n,h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \gamma_{k\ell}(n,h) g'_\ell(x_0) g_k(x_0) + O(h^3)\right),$$

where $\gamma_k(0,h) = 0$, $\gamma_{k\ell}(0,h) = 0$, $\gamma_k(1,h) = \tilde{\alpha}_k(h)$, $\gamma_{k\ell}(1,h) = \tilde{\alpha}_{k\ell}(h)$. Thus, to find a suitable modified ODE, it makes sense to assume that its corresponding t-flow $\tilde{\varphi}_t$ can be expanded for all $t \in \mathbb{R}$ as

$$\tilde{\varphi}_t(x_0) = \mathrm{e}^{tA}\left(x_0 + h\sum_{k\in\mathcal{I}} \gamma_k(t/h,h) g_k(x_0) + h^2 \sum_{k,\ell\in\mathcal{I}} \gamma_{k\ell}(t/h,h) g'_\ell(x_0) g_k(x_0) + \cdots\right). \tag{5.26}$$

Then the right-hand side of the corresponding ODE must be of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\varphi}_t(x)\bigg|_{t=0} = A\,x + \sum_{k\in\mathcal{I}} \beta_k(h) g_k(x) + h\sum_{k,\ell\in\mathcal{I}} \beta_{k\ell}(h) g'_\ell(x) g_k(x) + \cdots,$$

where

$$\beta_k(h) = \frac{\mathrm{d}}{\mathrm{d}\tau}\gamma_k(\tau,h)\bigg|_{\tau=0}, \quad \beta_{k\ell}(h) = \frac{\mathrm{d}}{\mathrm{d}\tau}\gamma_{k\ell}(\tau,h)\bigg|_{\tau=0}. \tag{5.27}$$

We are then bound to study the family of perturbed ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t}x = A\,x + \sum_{k\in\mathcal{I}} \beta_k(h)\, g_k(x) + h\sum_{k,\ell\in\mathcal{I}} \beta_{k\ell}(h)\, g'_\ell(x) g_k(x), \tag{5.28}$$

with initial condition $x(0) = x_0$. The coefficients $\beta_k(h)$ and $\beta_{k\ell}(h)$ are, for fixed h, arbitrary complex numbers. Notice that this equation generalizes the original ODE (5.2), which corresponds to the case where $\beta_k(h) = 1$ and $\beta_{k\ell}(h) = 0$.

Assuming that the t-flow $\tilde{\varphi}_t$ of (5.28) can be expanded as in (5.26), the group property of the flow (i.e. $\tilde{\varphi}_{t+s} = \tilde{\varphi}_t \circ \tilde{\varphi}_s$), together with the composition formula (5.20), lead to

$$\begin{aligned}
\gamma_k(\tau+\sigma,h) &= \mathrm{e}^{\mathrm{i}\omega h\tau k}\,\gamma_k(\sigma,h) + \gamma_k(\tau,h),\\
\gamma_{k\ell}(\tau+\sigma,h) &= \mathrm{e}^{\mathrm{i}\omega h\tau(k+\ell)}\,\gamma_{k\ell}(\sigma,h) + \mathrm{e}^{\mathrm{i}\omega h\tau\ell}\,\gamma_k(\tau,h)\,\gamma_\ell(\sigma,h) + \gamma_{k\ell}(\tau,h).
\end{aligned} \tag{5.29}$$

Now, applying the operator $(\mathrm{d}/\mathrm{d}\sigma)|_{\sigma=0}$ to both sides of (5.29) and using (5.27), we obtain the following system of ODEs for the coefficients $\gamma_k(\tau,h)$, $\gamma_{k\ell}(\tau,h)$:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\gamma_k(\tau,h) = \mathrm{e}^{\mathrm{i}\omega h\tau k}\,\beta_k(h),$$
$$\frac{\mathrm{d}}{\mathrm{d}\tau}\gamma_{k\ell}(\tau,h) = \mathrm{e}^{\mathrm{i}\omega h\tau(k+\ell)}\,\beta_{k\ell}(h) + \mathrm{e}^{\mathrm{i}\omega h\tau\ell}\,\beta_\ell(h)\,\gamma_k(\tau,h).$$

This, together with the initial conditions $\gamma_k(0,h) = 0$, $\gamma_{k\ell}(0,h) = 0$, allows us to express $\gamma_k(\tau,h)$, $\gamma_{k\ell}(\tau,h)$ in terms of the coefficients $\beta_k(h)$, $\beta_{k\ell}(h)$ of the perturbed ODE (5.28):

$$\gamma_k(\tau,h) = \beta_k(h)\int_0^\tau \mathrm{e}^{\mathrm{i}\omega h\sigma k}\,\mathrm{d}\sigma = \tau\,\beta_k(h)\,\alpha_k(\tau h),$$
$$\gamma_{k\ell}(\tau,h) = \tau\,\beta_{k\ell}(h)\,\alpha_{k+\ell}(\tau h) + \tau^2\,\beta_\ell(h)\,\alpha_{k\ell}(\tau h).$$

Specifying these equations to the case $\tau = 1$, and replacing $\gamma_k(1,h)$ and $\gamma_{k\ell}(1,h)$ with $\tilde\alpha_k(h)$ and $\tilde\alpha_{k\ell}(h)$ (the coefficients corresponding to ψ_h), respectively, we have

$$\begin{aligned}\tilde\alpha_k(h) &= \beta_k(h)\,\alpha_k(h),\\ \tilde\alpha_{k\ell}(h) &= \beta_{k\ell}(h)\,\alpha_{k+\ell}(h) + \beta_k(h)\,\beta_\ell(h)\,\alpha_{k\ell}(h),\end{aligned} \tag{5.30}$$

where the expression of $\alpha_k(h)$, for all $k\in\mathbb{Z}$, is given by (5.11). Hence $\alpha_0(h) = 1$ and, for any $k\neq 0$, $\alpha_k(h)\neq 0$ if and only if $kh\neq 2\pi j$ for all $j\in\mathbb{Z}$. Therefore, if we assume that $h\in\mathbb{R}$ is such that, for all $k,\ell\in\mathcal{I}\backslash\{0\}$,

$$\frac{k\omega h}{2\pi}\notin\mathbb{Z}\backslash\{0\}\quad\text{and}\quad\frac{(k+\ell)\omega h}{2\pi}\notin\mathbb{Z}\backslash\{0\} \tag{5.31}$$

(so that $\alpha_k(h)\neq 0$ and $\alpha_{k+\ell}(h)\neq 0$), then equations (5.30) can be solved in the $\beta_k(h)$ and $\beta_{k+\ell}(h)$ coefficients. Thus, for each $k,\ell\in\mathcal{I}$, we have

$$\beta_k(h) = \frac{\tilde\alpha_k(h)}{\alpha_k(h)},\quad \beta_{k\ell}(h) = \frac{\tilde\alpha_{k\ell}(h) - \alpha_{k\ell}(h)\,\beta_k(h)\beta_\ell(h)}{\alpha_{k+\ell}(h)}, \tag{5.32}$$

and the h-flow of the corresponding modified equation (5.28) agrees up to terms of order $O(h^3)$ with the expansion (5.22) of the splitting method ψ_h. Moreover, the identity (5.17) satisfied by the coefficients in the expansions of both the splitting method ψ_h and the exact solution, equations (5.11)–(5.12), implies

$$\beta_{k\ell}(h) + \beta_{\ell k}(h) = 0,\quad k,\ell\in\mathcal{I}, \tag{5.33}$$

so that the modified equation (5.28) can also be expressed as

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}x &= A\,x + \sum_{k\in\mathcal{I}}\beta_k(h)\,g_k(x) + \frac{h}{2}\sum_{k,\ell\in\mathcal{I}}\beta_{k\ell}(h)\,(g_\ell'(x)g_k(x) - g_k'(x)g_\ell(x))\\ &= A\,x + \sum_{k\in\mathcal{I}}\beta_k(h)\,g_k(x) + \frac{h}{2}\sum_{k,\ell\in\mathcal{I}}\beta_{k\ell}(h)\,(g_k,g_\ell)(x).\end{aligned} \tag{5.34}$$

In fact, equation (5.34) is itself Hamiltonian, in the sense that there exists a Hamiltonian function $\tilde{H}(x;h)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}x = J\,\nabla\tilde{H}(x;h), \tag{5.35}$$

where J is the canonical symplectic matrix (1.21). This can be seen as follows. First, H_1 in the Hamiltonian (5.1) can be written as

$$H_1(x) = \frac{1}{2}x^\top Qx, \quad \text{with} \quad Q = \begin{pmatrix} N & 0 \\ 0 & M^{-1} \end{pmatrix}, \tag{5.36}$$

whereas $H_2(\mathrm{e}^{tA}x)$ has a Fourier expansion of the form

$$H_2(\mathrm{e}^{tA}x) = \sum_{k\in\mathcal{I}} \mathrm{e}^{\mathrm{i}k\omega t}G_k(x), \quad \text{with} \quad g_k(x) = J\,\nabla G_k(x), \tag{5.37}$$

and furthermore, for all $k \in \mathcal{I}$,

$$G_k(\mathrm{e}^{t\,A}x) = \mathrm{e}^{\mathrm{i}k\omega t}G_k(x), \tag{5.38}$$

so that (5.7) is equivalent to

$$\{H_1, G_k\} = \mathrm{i}k\omega\, G_k, \quad \text{for } k \in \mathcal{I}. \tag{5.39}$$

Here $\{A, B\}$ stands for the Poisson bracket of $A, B \in \mathcal{C}^1(\mathbb{R}^D)$ defined as follows: for each $x \in \mathbb{R}^D$,

$$\{A, B\}(x) = (\nabla A(x))^\top J\,\nabla B(x).$$

It is then clear that the modified ODE (5.34) can be written like (5.35) with

$$\tilde{H}(x;h) = \frac{1}{2}x^\top Qx + \sum_{k\in\mathcal{I}} \beta_k(h)\,G_k(x) + \frac{h}{2}\sum_{k,\ell\in\mathcal{I}} \beta_{k\ell}(h)\,\{G_k, G_\ell\}(x). \tag{5.40}$$

We should stress that both (5.28) and the modified Hamiltonian $\tilde{H}(x;h)$ are well-defined as long as the non-resonance assumptions (5.31) hold.

5.6. *Splitting methods with processing*

In the spirit of Section 4.5, we now consider a *processed* splitting integrator

$$\hat{\psi}_h = \pi_h^{-1} \circ \psi_h \circ \pi_h,$$

where ψ_h is a composition of type (2.17) based on kicks and rotations and $\pi_h\colon \mathbb{R}^D \to \mathbb{R}^D$ is a near-to-identity map with an expansion of the form

$$\pi_h(x) = x + h\sum_{k\in\mathcal{I}} \kappa_k(h)\,g_k(x) + h^2\sum_{k,\ell\in\mathcal{I}} \kappa_{k\ell}(h)\,g'_\ell(x)g_k(x) + O(h^3). \tag{5.41}$$

In contrast to Section 4.5, where the processor map π_h is analysed by its power series expansion in h, the coefficients $\kappa_k(h)$ and $\kappa_{k\ell}(h)$ featuring in (5.41) will depend on h and the frequency ω in a non-polynomial way. This will allow us

to find appropriate processors valid for step sizes h that are not necessarily small compared with ω.

Application of the composition rule (5.20)–(5.21) to both sides of the identity $\pi_h \circ \hat{\psi}_h = \psi_h \circ \pi_h$ implies that $\hat{\psi}_h$ can be expanded as

$$\hat{\psi}_h(x) = \mathrm{e}^{hA}\left(x + h\sum_{k\in\mathcal{I}} \hat{\alpha}_k(h) g_k(x) + h^2 \sum_{k,\ell\in\mathcal{I}} \hat{\alpha}_{k\ell}(h) g'_\ell(x) g_k(x) + O(h^3)\right), \quad (5.42)$$

where

$$\begin{aligned} \hat{\alpha}_k(h) &= (1 - \mathrm{e}^{\mathrm{i}k\omega h})\,\kappa_k(h) + \tilde{\alpha}_k(h), \\ \hat{\alpha}_{k\ell}(h) &= (1 - \mathrm{e}^{\mathrm{i}(k+\ell)\omega h})\,\kappa_{k\ell}(h) + \kappa_k(h)\,\tilde{\alpha}_\ell(h) - \mathrm{e}^{\mathrm{i}\ell\omega h}\kappa_\ell(h)\,\hat{\alpha}_k(h) + \tilde{\alpha}_{k\ell}(h). \end{aligned} \quad (5.43)$$

The expansion of the processed scheme $\hat{\psi}_h$ coincides with that of the exact flow (5.14) if $\hat{\alpha}_k(h) = \alpha_k(h)$ for each $k \in \mathcal{I}\backslash\{0\}$, and this is possible only when the following non-resonance condition holds,

$$\frac{k\omega h}{2\pi} \notin \mathbb{Z}, \quad (5.44)$$

in which case

$$\kappa_k(h) = \frac{\tilde{\alpha}_k(h) - \alpha_k(h)}{\mathrm{e}^{\mathrm{i}k\omega h} - 1}. \quad (5.45)$$

Observe that, for $k = 0$, $\hat{\alpha}_k(h) = \alpha_k(h)$ regardless of the chosen value of $\kappa_0(h)$, since $\tilde{\alpha}_0(h) = \alpha_0(h) = 1$. For simplicity, it makes sense to choose the processor map in such a way that $\kappa_0(h) = 0$.

In consequence, if h satisfies the non-resonance condition (5.44) for all $k \in \mathcal{I}\backslash\{0\}$, and the coefficients $\kappa_k(h)$ for $k \neq 0$ are chosen as (5.45), then the local error of the processed scheme reads

$$\hat{\psi}_h(x) - \varphi_h(x) = \mathrm{e}^{hA}\left(h^2 \sum_{k,\ell\in\mathcal{I}} (\hat{\alpha}_{k\ell}(h) - \alpha_{k\ell}(h)) g'_\ell(x)\, g_k(x) + O(h^3)\right)$$

and the h-flow of $\hat{\psi}_h$ is also Hamiltonian, with the modified Hamiltonian function

$$\hat{H}(x;h) = \frac{1}{2}x^\top Q x + \sum_{k\in\mathcal{I}} \hat{\beta}_k(h)\, G_k(x) + \frac{h}{2}\sum_{k,\ell\in\mathcal{I}} \hat{\beta}_{k\ell}(h)\,\{G_k, G_\ell\}(x),$$

where

$$\hat{\beta}_k(h) = 1, \quad \hat{\beta}_{k\ell}(h) = \frac{\hat{\alpha}_{k\ell}(h) - \alpha_{k\ell}(h)}{\alpha_{k+\ell}(h)}. \quad (5.46)$$

Equivalently,

$$\hat{H}(x;h) = H(x) + \frac{h}{2}\sum_{k,\ell\in\mathcal{I}} \hat{\beta}_{k\ell}(h)\,\{G_k, G_\ell\}(x), \quad (5.47)$$

which is an $O(h)$ perturbation of the original Hamiltonian, and thus we may expect that the value of the $H(x)$ (typically the energy of the original system) will be well approximated for relatively large time intervals. A word of caution is in order here: by construction, the difference between the processed map $\hat{\psi}_h$ and the h-flow of the modified Hamiltonian (5.47) is formally of order $O(h^3)$, but the constant in $O(h^3)$ depends on the size of the modulus of the coefficients

$$\kappa_k(h), \quad \kappa_{k,\ell}(h), \quad \hat{\beta}_{k\ell}(h) \quad \text{for } k, \ell \in \mathcal{I}.$$

For near-resonant step sizes h, that is, for step sizes such that $\mathrm{e}^{\mathrm{i}k\omega h} - 1$ is small for some index k belonging to $\mathcal{I}$ or $\mathcal{I}+\mathcal{I}$, the size of some of these critical coefficients may become large. In such cases, we cannot expect that the processed map $\hat{\psi}_h$ will be close from the h-flow of the modified Hamiltonian (5.47).

5.7. *A processed Strang scheme*

In the particular case in which ψ_h is the Strang splitting $S_h^{[RKR]}$, the coefficients $\kappa_k(h)$ in the corresponding expansion (5.41) verifying (5.45) read

$$\kappa_k(h) = \frac{1}{\mathrm{i}k\omega h}(1 - \mathrm{sinc}(k\omega h/2)^{-1}), \quad k \in \mathcal{I}\backslash\{0\}, \tag{5.48}$$

and $\kappa_{-k}(h) = -\kappa_k(h)$ for all $k \neq 0$.

Assuming that the potential function $U(q)$ in (5.1) is a polynomial of degree m, so that

$$\mathcal{I} \subset \{-m, \dots, -1, 0, 1, \dots m\},$$

we next construct a fully explicit processor π_h for $S_h^{[RKR]}$. We define it as a composition of basic flows as follows:

$$\pi_h = \varphi_\alpha^{[R]} \circ \varphi_{b_{2m}(h)}^{[K]} \circ \varphi_\alpha^{[R]} \circ \varphi_{b_{2m-1}(h)}^{[K]} \circ \cdots \circ \varphi_{b_2(h)}^{[K]} \circ \varphi_\alpha^{[R]} \circ \varphi_{b_1(h)}^{[K]} \circ \varphi_\alpha^{[R]}, \tag{5.49}$$

where $\alpha = (2\pi)/(2m+1)$, and the coefficients $b_j(h)$, $j = 1, \dots, 2m$, depend on h. Since $S_h^{[RKR]}$ is time-symmetric, then it makes sense to construct the processor π_h such that $\hat{\psi}_h = \pi_h^{-1} \circ \psi_h \circ \pi_h$ is also time-symmetric. This can be achieved by requiring that $\pi_{-h} = \pi_h$, or equivalently, by requiring that $b_j(-h) = b_j(h)$ for $j = 1, \dots, 2m$. This condition, together with

$$\kappa_0(h) = 0, \quad \tilde{\alpha}_k(h) = \alpha_k(h) \quad \text{for all } k \in \{-m, \dots, -1, 0, 1, \dots m\},$$

uniquely determines the $b_j(h)$ coefficients as

$$b_j(h) = -b_{2m-j+1}(h) = \frac{2}{2m+1} \sum_{k=1}^{m} \frac{1}{k}(\mathrm{sinc}(k\omega h/2)^{-1} - 1) \sin\left(\frac{2kj\pi}{2m+1}\right), \tag{5.50}$$

for $j = 1, \dots, m$. This can be seen as follows: successive application of the

composition rule (5.21) shows that (5.49) admits the expansion (5.41) with

$$h\,\kappa_k(h) = \sum_{j=1}^{2m} b_j(h)\,\mathrm{e}^{2\mathrm{i}kj\pi/(2m+1)}, \quad k = -m, \ldots, -1, 0, 1, \ldots m, \tag{5.51}$$

and for $k, \ell \in \mathcal{I}$,

$$\begin{aligned} h^2\kappa_{k\ell}(h) = &\sum_{1\le j<n\le 2m} b_j(h)\,b_n(h)\,\mathrm{e}^{2\pi(jk+n\ell)/(2m+1)} \\ &+ \sum_{1\le j\le 2m} \frac{1}{2}\,b_j(h)^2\,\mathrm{e}^{\mathrm{i}2\pi j(k+\ell)/(2m+1)}. \end{aligned} \tag{5.52}$$

We can check that $\kappa_0(h) = 0$ and the symmetry condition $\kappa_{-k}(h) = -\kappa_k(h)$ $(k \neq 0)$ that holds for (5.48) imply that $b_j(h) = -b_{2m-j+1}(h)$ for $j = 1, \ldots, 2m$. Now, (5.51) means that application of the inverse discrete Fourier transform to the vector $(0, b_1, b_2, \ldots, b_m, -b_m, \ldots, -b_1)$ gives the vector

$$\frac{h}{2m+1}(0, \kappa_1, \ldots, \kappa_m, -\kappa_m, \ldots, -\kappa_1).$$

Equivalently, the former is obtained by applying the discrete Fourier transform to the latter. Rearranging terms, we finally arrive at (5.50).

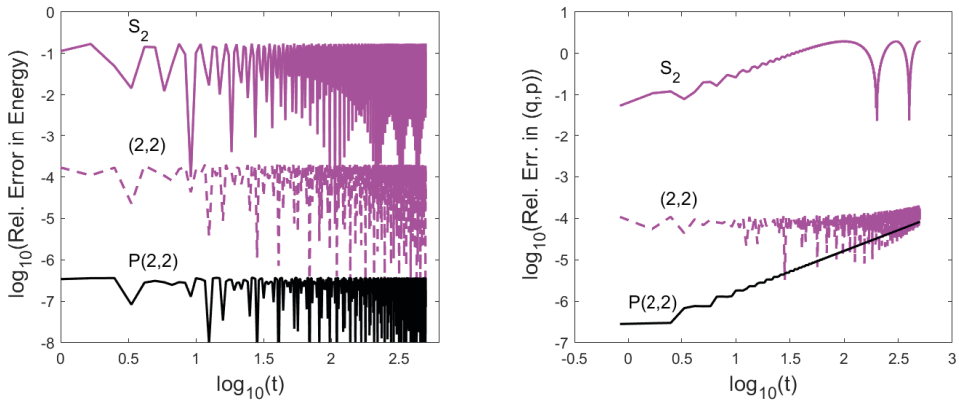
The modified Hamiltonian (5.47) can be obtained in the following way: first the coefficients (5.52) can be used to compute $\hat{\alpha}_{k\ell}(h)$ from (5.43), and then the coefficients $\hat{\beta}_{k\ell}(h)$ are determined from (5.46).

Example: simple pendulum. As an illustrative example we again consider the simple pendulum of Section 1.4, described by the Hamiltonian function (1.19). As we saw there, for initial conditions in a neighbourhood of the stable equilibrium $(0, 0)$ it is advantageous to decompose H as in (1.32), so it constitutes a particular example of system (5.1):

$$H = H_1 + H_2, \quad H_1(q,p) = \frac{1}{2}p^2 + \frac{1}{2}q^2, \quad H_2(q,p) = U(q) = 1 - \frac{1}{2}q^2 - \cos q.$$

Although $U(q)$ is not a polynomial, and therefore the set $\mathcal{I}$ in (5.4) is infinite, we can truncate the Fourier expansion (5.37) and work instead with $\mathcal{I} = \{-m, \ldots, -1, 0, 1, \ldots, m\}$ for a given m to construct a processed Strang scheme based on kicks and rotations as proposed earlier. Specifically, we take $m = 4$, a step size $h = 5/6$, then determine the corresponding coefficients $b_j(h)$ and form the integrator $\hat{\psi}_h = \pi_h^{-1} \circ S_h^{[RKR]} \circ \pi_h$, with π_h given by (5.49) (with $m = 4$), whereas $S_h^{[RKR]}$ corresponds to the map (1.34).

Figure 5.1 shows the relative error in energy (a) and phase space (b) over the time interval $[0, 500]$ corresponding to the solution initiated at $(q_0, p_0) = (1/10, 0)$. As usual, S_2 denotes the Störmer–Verlet method applied to (1.19), (2,2) is the Strang splitting $S_h^{[RKR]}$, and P(2, 2) is the processed $S_h^{[RKR]}$ scheme with step size $h = 5/6$. Since the integrator (2, 2) is conjugate to P(2, 2), the error of the former is eventually



(a) (b)

Figure 5.1. Pendulum. Evolution of relative errors in energy (a) and in phase space (b) obtained with Störmer–Verlet, S_2, scheme $S_h^{[RKR]}$, denoted (2,2), and the processed (2,2), P(2, 2) with initial state $(q_0, p_0) = (1/10, 0)$ and step size $h = 5/6$.

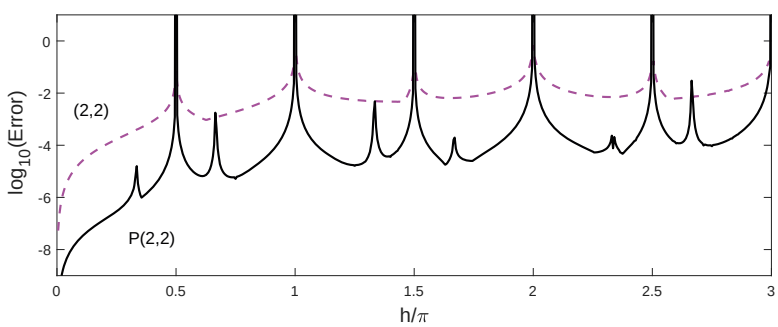


Figure 5.2. Pendulum. Maximum relative energy error in the time interval $[0, 500]$, initial state $(q_0, p_0) = (1/10, 0)$, and step sizes in the range $[0, 3\pi]$ committed by the Strang splitting $S_h^{[RKR]}$ and its processed version P(2, 2).

dominated by the error of the latter, as expected from the discussion in Section 4.5. Observe that the step size used here is larger than those taken in Figure 1.1.

Next, we consider the Strang method $S_h^{[RKR]}$ and its processed version P(2, 2) for the same initial condition and a time interval $[0, 500]$, and show the maximum energy error in this interval for step sizes in the range $[0, 3\pi]$ (Figure 5.2). The spikes in the curve of the error of Strang correspond to the step sizes violating the non-resonance condition (5.44) for $\omega = 1$ and $k = 2, 4$. Indeed, the potential $U(q)$ can be well approximated near the origin by $q^4/24$, which implies that $\mathcal{I} = \{-4, -2, 0, 2, 4\}$. The curve of the error of processed Strang also has spikes for such resonant step sizes, and additionally, for the step sizes h such that $3h/(2\pi) \in \mathbb{Z}$,

which is due to the fact that the processor map π_h has been designed to work well for problems for which $\mathcal{I} = \{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$. Away from such resonant step sizes, the energy error of processed Strang is considerably smaller than that of unprocessed Strang.

5.8. *More general assumptions*

The results given in the present section are also valid with small modifications in the more general setting of systems (5.2) such that the eigenvalues of A lie on the imaginary axis. In the most general case, where such eigenvalues are not integer multiples of $\omega\,\mathrm{i}$, the exponential e^{tA} is quasi-periodic with a finite number of basic frequencies $(\omega_1, \dots, \omega_r) \in \mathbb{R}^r$. Under such conditions, $\mathrm{e}^{-tA} f_2(\mathrm{e}^{tA}x)$ admits a multi-variable Fourier expansion

$$\sum_{k\in\mathbb{Z}^r} \mathrm{e}^{\mathrm{i}\langle k,\omega\rangle t}\, g_k(x), \tag{5.53}$$

where k now denotes an r-tuple of integers $k = (k_1, \dots, k_r) \in \mathbb{Z}^r$, $\omega = (\omega_1, \dots, \omega_r)$ is the vector of basic frequencies, and $\langle k, \omega\rangle = k_1\omega_1 + \cdots + k_r\omega_r$. The set of indices $\mathcal{I}$ is then defined as

$$\mathcal{I} = \{k \in \mathbb{Z}^r : g_k \neq 0\}.$$

Under these more general assumptions and notation, all previous formulae are valid if each occurrence of $k\omega$ with $k \in \mathcal{I}$ is replaced by $\langle k, \omega\rangle$, the only exception being the explicit construction of the processing map π_h carried out in Section 5.7. It is worth remarking that, compared to the periodic case, in the quasi-periodic one there are typically more resonant step sizes, that is, step sizes h such that

$$\frac{\langle k,\omega\rangle h}{2\pi} \in \mathbb{Z}\backslash\{0\} \quad \text{or} \quad \frac{\langle (k+\ell),\omega\rangle h}{2\pi} \in \mathbb{Z}\backslash\{0\}$$

for some $k, \ell \in \mathcal{I}$.

6. Splitting methods for PDEs

6.1. *Splitting, LOD and ADI methods*

Splitting methods can also be applied to partial differential equations (PDEs), in which case equation (1.1) has to be viewed as the abstract system associated with the PDE initial value problem in autonomous form, and f as a spatial partial differential operator. For clarity, in this section we write

$$u_t(x,t) = f(x, u(x,t)), \quad u(x,0) = u_0(x) \tag{6.1}$$

to distinguish between the unknown $u(x,t)$ defined in a certain function space and the spatial variable $x \in \mathbb{R}^d$, but for notational purposes it is convenient to drop the dependence of u and f on x. To introduce the basic concepts and methods it is not necessary at this stage to specify the dimension d, the number of components

of u, the relevant function space and the boundary conditions. These will be considered when analysing particular applications. Moreover, for simplicity in the presentation, only the autonomous case will be treated. If f depends explicitly on time, then we can take t as a new coordinate, as is done in Section 3.5.

Very often, the operator f contains contributions coming from very different physical sources, so we may decompose it into two (or more) parts, and use different schemes to solve each sub-problem approximately. For instance, in a reaction–diffusion system, $f(u) = \nabla \cdot (D\nabla u) + g(u)$, where D and g may also depend on x, it makes sense to split the diffusion from the reaction terms, that is,

$$f(u) = f_1(u) + f_2(u), \quad \text{with} \quad f_1(u) = \nabla \cdot (D\nabla u),\ f_2(u) = g(u).$$

Algorithm 1.1 can of course be used in this setting, but we still have to specify how to solve each initial value sub-problem in practice with appropriate boundary conditions. A simple possibility consists in applying the backward Euler scheme, thus resulting in the so-called *Marchuk–Yanenko* operator-splitting scheme

$$\begin{aligned} u_{n+1/2} &= u_n + hf_1(u_{n+1/2}), \\ u_{n+1} &= u_{n+1/2} + hf_2(u_{n+1}), \quad n = 0, 1, 2, \ldots, \end{aligned} \tag{6.2}$$

where, for consistency with the rest of the paper, we have denoted $h \equiv \Delta t$, the time step size. In spite of its low order of consistency (order one)[5] and the fact that the intermediate stage $u_{n+1/2}$ is not a consistent approximation to the exact solution, its simplicity and robustness make it a useful alternative way to deal with complicated problems and even non-smooth operators (Glowinski *et al.* 2016*a*). It is also appropriate for parabolic problems, since it incorporates the damping properties of the backward Euler method (Hundsdorfer and Verwer 2003).

If, instead of using the backward Euler scheme to integrate each sub-problem in Algorithm 1.1, we apply the second-order implicit trapezoidal rule, it results in Yanenko's Crank–Nicolson method (Hundsdorfer and Verwer 2003, Marchuk 1990):

$$\begin{aligned} u_{n+1/2} &= u_n + \frac{h}{2} f_1(u_n) + \frac{h}{2} f_1(u_{n+1/2}), \\ u_{n+1} &= u_{n+1/2} + \frac{h}{2} f_2(u_{n+1/2}) + \frac{h}{2} f_2(u_{n+1}), \quad n = 0, 1, 2, \ldots \end{aligned} \tag{6.3}$$

In the end, however, it is also of first order of consistency.

Another widely popular class of spitting methods in the domain of PDEs is the Peaceman–Rachford scheme and its variants. Although initially designed for the numerical solution of elliptic and parabolic equations (Peaceman and Rachford 1955, Douglas and Rachford 1956), they also apply to more general situations. The procedure goes as follows. Given an approximation u_n for the solution of (6.1)

[5] Here 'order' should be understood as the order of consistency with respect to the solution of the ODE problem on a fixed spatial grid, not with respect to the underlying PDE solution (Hundsdorfer and Verwer 2003).

at $t = t_n$, the approximation u_{n+1} is computed using the backward (resp. forward) Euler scheme with respect to f_1 (resp. f_2) on the sub-interval $[t_n, t_{n+1/2}]$. Then the roles of f_1 and f_2 are interchanged on the sub-interval $[t_{n+1/2}, t_{n+1}]$. In other words, the *Peaceman–Rachford* scheme corresponds to the sequence

$$\begin{aligned} u_{n+1/2} &= u_n + \frac{h}{2} f_1(u_{n+1/2}) + \frac{h}{2} f_2(u_n), \\ u_{n+1} &= u_{n+1/2} + \frac{h}{2} f_1(u_{n+1/2}) + \frac{h}{2} f_2(u_{n+1}), \quad n = 0, 1, 2, \ldots, \end{aligned} \tag{6.4}$$

which is of second order of consistency. Notice that, in contrast to methods (6.2) and (6.3), both f_1 and f_2 appear in each of the two stages, and thus the intermediate value $u_{n+1/2}$ provides a consistent approximation at $t = t_{n+1/2}$. On the other hand, it does not possess a natural formulation where f is split into more than two operators. In general, f_1 and f_2 can be nonlinear, unbounded and even multi-valued. For a more detailed treatment, the reader is addressed to Glowinski, Pan and Tai (2016*b*) and references therein.

A classical alternative to scheme (6.4), of first order, is the *Douglas–Rachford* method (Douglas and Rachford 1956), which instead reads

$$\begin{aligned} \hat{u}_{n+1} &= u_n + h f_1(\hat{u}_{n+1}) + h f_2(u_n), \\ u_{n+1} &= u_n + h f_1(\hat{u}_{n+1}) + h f_2(u_{n+1}), \quad n = 0, 1, 2, \ldots, \end{aligned} \tag{6.5}$$

and can be generalized to decompositions of f involving more than two operators. Notice that the roles of f_1 and f_2 in (6.5) are not symmetric, in contrast to the Peaceman–Rachford method. On the basis of many numerical experiments, Glowinski *et al.* (2016*b*) conclude that the scheme (6.5) is faster and more robust than (6.4) for problems where one of the operators is non-smooth, in particular when we are interested in approximating steady-state solutions.

Let us now analyse the particular case when f in (6.1) is a linear spatial differential operator. Assuming that an appropriate semidiscretization of (6.1) in the space variable x has been carried out, we end up with the system

$$\frac{\mathrm{d}U}{\mathrm{d}t} = F_1 U + F_2 U, \tag{6.6}$$

where $F_1, F_2 \in \mathbb{C}^{M\times M}$, $F_1 F_2 \neq F_2 F_1$ in general, and $U \in \mathbb{C}^M$ approximates u on the space grid points $x_1, \ldots, x_M$ (see below). Then a step of the Marchuk–Yanenko scheme (6.2) reads

$$U_{n+1} = (I - hF_2)^{-1} (I - hF_1)^{-1} U_n, \tag{6.7}$$

where $U_n \approx (u(x_1, t_n), \ldots, u(x_M, t_n))^\top$. Notice that (6.7) corresponds to applying the $[0/1]$ Padé approximant to the exponentials in the Lie–Trotter scheme $U_{n+1} = \mathrm{e}^{hF_2}\, \mathrm{e}^{hF_1} U_n$, whence the first order of the approximation is obtained at once. On the

other hand, applying (6.3) results in the sequence

$$U_{n+1} = \left(I - \frac{h}{2}F_2\right)^{-1}\left(I + \frac{h}{2}F_2\right)\left(I - \frac{h}{2}F_1\right)^{-1}\left(I + \frac{h}{2}F_1\right)U_n. \tag{6.8}$$

In other words, it corresponds to the application of the $[1/1]$ Padé approximant to the exponentials in the Lie–Trotter scheme. It is then clear that, although a second-order Crank–Nicolson method is carried out for each exponential, their combination (6.8) is of first order, since in the end we are using only a variant of the Lie–Trotter scheme.

By the same token, a straightforward computation shows that the Peaceman–Rachford scheme (6.4) applied to the linear equation (6.6) can be written as

$$U_{n+1} = \left(I - \frac{h}{2}F_2\right)^{-1}\left(I + \frac{h}{2}F_1\right)\left(I - \frac{h}{2}F_1\right)^{-1}\left(I + \frac{h}{2}F_2\right)U_n. \tag{6.9}$$

As a matter of fact, all these algorithms can be formulated by applying properly chosen compositions of the implicit and explicit Euler methods. This observation may eventually lead to the construction of methods of higher order or improved behaviour, but in the same family. Suppose f in (6.1) is of the form $f = f_1 + f_2$ and that the solution of each equation $u_t = f_k(u)$, $k = 1, 2$, is numerically approximated by the maps

$$\begin{aligned} u_{n+1} &= \psi_h^{ke}(u_n) \equiv u_n + hf_k(u_n), && \text{explicit Euler,}\\ u_{n+1} &= \psi_h^{ki}(u_n) \equiv u_n + hf_k(u_{n+1}), && \text{implicit Euler,} \end{aligned}$$

so that by combining all variants, we form the following first-order schemes (and their corresponding adjoints):

$$\begin{aligned} \phi_h^1 &= \psi_h^{1e} \circ \psi_h^{2e}, & \phi_h^{1*} &= \psi_h^{2i} \circ \psi_h^{1i},\\ \phi_h^2 &= \psi_h^{1e} \circ \psi_h^{2i}, & \phi_h^{2*} &= \psi_h^{2e} \circ \psi_h^{1i},\\ \phi_h^3 &= \psi_h^{1i} \circ \psi_h^{2e}, & \phi_h^{3*} &= \psi_h^{2i} \circ \psi_h^{1e},\\ \phi_h^4 &= \psi_h^{1i} \circ \psi_h^{2i}, & \phi_h^{4*} &= \psi_h^{2e} \circ \psi_h^{1e}. \end{aligned} \tag{6.10}$$

Then we can conclude the following.

- The Marchuk–Yanenko operator-splitting scheme (6.2) can be expressed as $u_{n+1} = \phi_h^{1*}(u_n)$. Therefore compositions

  $$u_{n+1} = \phi_{h/2}^{1*} \circ \phi_{h/2}^{1}(u_n) \quad \text{and} \quad u_{n+1} = \phi_{h/2}^{1} \circ \phi_{h/2}^{1*}(u_n)$$

  yield time-symmetric second-order approximations.
- Yanenko's Crank–Nicolson method (6.3) corresponds to the composition

  $$u_{n+1} = \psi_{h/2}^{2i} \circ \psi_{h/2}^{2e} \circ \psi_{h/2}^{1i} \circ \psi_{h/2}^{1e}(u_n), \tag{6.11}$$

  which is not time-symmetric and therefore only of first order. Notice, however,

that (6.11) can also be expressed as

$$u_{n+1} = \psi^{2i}_{h/2} \circ \left(\phi^{2*}_{h/2} \circ \phi^{2}_{h/2}\right) \circ \left(\psi^{2i}_{h/2}\right)^{-1}(u_n),$$

so it is conjugate to a time-symmetric second-order method. This feature could account for its observed good behaviour in practice.

- The Peaceman–Rachford scheme (6.4) is just the symmetric composition

$$u_{n+1} = \phi^{3*}_{h/2} \circ \phi^{3}_{h/2}(u_n),$$

and is therefore of second order. It is also worth mentioning that whereas all symmetric second-order compositions

$$\phi^{\ell*}_{h/2} \circ \phi^{\ell}_{h/2} \quad \text{and} \quad \phi^{\ell}_{h/2} \circ \phi^{\ell*}_{h/2}, \quad \ell = 1, \ldots, 4$$

provide consistent approximations at the midpoint, the steady-state solution is captured only when $\ell = 3$: if $f(w) = 0$, then

$$\phi^{3*}_{h/2} \circ \phi^{3}_{h/2}(w) = \phi^{3}_{h/2} \circ \phi^{3*}_{h/2}(w) = w.$$

The Douglas–Rachford method (6.5) can be alternatively formulated as a splitting method in an extended space as follows. Consider the enlarged system

$$w_t = \begin{pmatrix} \hat{u}_t \\ v_t \end{pmatrix} = \begin{pmatrix} f_1(\hat{u}) + f_2(v) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ f_1(\hat{u}) + f_2(v) \end{pmatrix} = g_1(w) + g_2(w), \tag{6.12}$$

with initial condition $w(0) = (\hat{u}(0), v(0)) = (u_0, u_0)$ and solution $w(t) = (\hat{u}(t), v(t)) = (u(t), u(t))$. Take $w_n = (u_n, u_n)$ and form the composition

$$\hat{w}_{n+1} = \phi^{1*}_{h}(w_n) = \psi^{2i}_{h} \circ \psi^{1i}_{h}(w_n), \quad n \geq 0,$$

providing $\hat{w}_{n+1} = (\hat{u}_{n+1}, v_{n+1})$. Then scheme (6.5) is recovered by taking $u_{n+1} = v_{n+1}$, and considering $w_{n+1} = (u_{n+1}, u_{n+1})$ as the starting point for the next iteration.

Example. The two-dimensional heat equation with source term and Dirichlet boundary conditions on the unit square may serve as an illustration of these methods (Hundsdorfer and Verwer 2003). Specifically, the system reads

$$\begin{aligned} u_t &= u_{xx} + u_{yy} + g(x, y, t) && \text{on } \Omega = (0, 1) \times (0, 1), \\ u(x, y, t) &= u_\Gamma(x, y, t) && \text{on } \Gamma = \partial\Omega, \\ u(x, y, 0) &= u_0(x, y) && \text{on } \Omega. \end{aligned} \tag{6.13}$$

Suppose we take a Cartesian grid in Ω based on $M + 1$ equally spaced intervals in the x and y directions, so that $\Delta x = \Delta y = 1/(M + 1)$ and apply finite differences to approximate the space derivatives. Then we end up with M^2 interior points and the aim is to get approximations of u at these points, i.e. to determine $u_{i,j}(t) \simeq u(x_i, y_j, t)$ for $i, j = 1, 2, \ldots, M$. The problem can be conveniently formulated in terms of the 'supervector' $U = (u_1^\top, \ldots, u_M^\top)^\top$, with $u_i = (u_{i,1}, \ldots, u_{i,M})^\top$, and M^2 components

$$U_\ell(t) = u_{i,j}(t), \quad \ell = j + M(i - 1).$$

Using standard second-order finite differences for ∂_{xx} and ∂_{yy} we get a linear system of the form $U' = F\,U + s(t)$, where $s(t)$ contains sources and boundary data, and $F = F_1 + F_2$, with

$$F_1 = I_M \otimes B_M, \quad F_2 = B_M \otimes I_M, \quad B_M = \frac{1}{(\Delta x)^2}\begin{pmatrix} -2 & 1 & \cdots & 0 \\ 1 & -2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & & 1 & -2 \end{pmatrix}.$$

Here I_M is the identity matrix, B_M the differentiation matrix and $\otimes$ denotes the tensor product. Notice that F_1 acts in the x-direction and F_2 in the y-direction, so both F_1 and F_2 are essentially one-dimensional. Since F_1 is a tridiagonal matrix, the linear system resulting from the application of the previous implicit schemes can be solved in an efficient way, whereas F_2 is equivalent to a tridiagonal matrix, so the same considerations apply (Iserles 1996).

First consider the homogeneous equation with zero Dirichlet boundary conditions. The solution for one time step is

$$U(t_n + h) = \mathrm{e}^{h(F_1+F_2)}\,U(t_n),$$

whereas the application of the Peaceman–Rachford scheme leads to the approximation (6.8):

$$U_{n+1} = (\phi^{3*}_{h/2} \circ \phi^{3}_{h/2})\,U_n = \mathrm{e}^{h(F_1+F_2)}\,U_n + O(h^3).$$

When a dimensional splitting is done, as in this case, applying schemes (6.7) and (6.8) corresponds essentially to carrying out computations in only one dimension. This is the reason why (6.2) and (6.3) are called *locally one-dimensional* (LOD) methods. Analogously, given the alternate use of F_1 and F_2 in this setting, the name *alternating direction implicit* (ADI) is usually attached to methods (6.4) and (6.5).

Regarding these ADI/LOD methods as composition schemes allows us to get approximations in the non-homogeneous case too, with only minor changes, while keeping any favourable properties (if no order reduction occurs due to the Dirichlet boundary conditions). To illustrate this point, we again apply the explicit and implicit Euler methods to the equation $U' = s(t)$ corresponding to the non-homogeneous term

$$U_{n+1} = \psi^{s_e}_h(U_n) = U_n + h\,s(t_n), \quad U_{n+1} = \psi^{s_i}_h(U_n) = U_n + h\,s(t_{n+1}),$$

and consider the first-order scheme $\hat{\phi}^3_h = \psi^{1i}_{h/2} \circ \psi^{2e}_{h/2} \circ \psi^{s_e}_h$. Then

$$\begin{aligned} U_{n+1} &= \left(\hat{\phi}^{3*}_{h/2} \circ \hat{\phi}^{3}_{h/2}\right) U_n = \left(\psi^{s_i}_{h/2} \circ \psi^{2i}_{h/2} \circ \psi^{1e}_{h/2} \circ \psi^{1i}_{h/2} \circ \psi^{2e}_{h/2} \circ \psi^{s_e}_{h/2}\right) U_n \\ &= \frac{h}{2}s(t_{n+1}) + \left(I - \frac{h}{2}F_2\right)^{-1}\left(I + \frac{h}{2}F_1\right)\left(I - \frac{h}{2}F_1\right)^{-1}\left(I + \frac{h}{2}F_2\right)\left(U_n + \frac{h}{2}s(t_n)\right) \end{aligned}$$

produces a symmetric second-order scheme.

It is not only reaction–diffusion problems that can be treated in this way. For instance, systems of hyperbolic conservation laws in three dimensions, such as

$$u_t + \nabla \cdot f(u) = 0, \quad u(x, y, z, 0) = u_0(x, y, z),$$

can be numerically approximated with dimensional splitting by applying a specially tailored numerical scheme to each scalar conservation law $u_t + f(u)_x = 0$, etc. (Holden, Karlsen, Lie and Risebro 2010).

The one-dimensional convection–diffusion problem

$$u_t + f(u)_x = A(u)_{xx}, \quad u(x, 0) = u_0(x)$$

with a scalar non-decreasing function $A(\cdot)$, $A(0) = 0$, possesses a rich set of phenomena depending on the interplay of the different nonlinearities. In this case we can split the problem into a convective and a diffusion part and apply Algorithm 1.1. This formally results in the so-called Godunov split (Holden *et al.* 2010). Specifically, letting $\mathcal{U}_h^{[1]}$ denote the solution operator corresponding to the scalar conservation law $u_t + f(u)_x = 0$, and letting $\mathcal{U}_h^{[2]}$ denote the solution operator corresponding to the (weak) solution of the nonlinear heat equation $u_t = A(u)_{xx}$, the scheme then reads

$$u(x, t_n + h) \approx u_{n+1} = \mathcal{U}_h^{[2]}\big(\mathcal{U}_h^{[1]}(u_n)\big).$$

Of course, we can also use the Strang splitting

$$u(x, t_n + h) \approx u_{n+1} = \mathcal{U}_{h/2}^{[2]}\big(\mathcal{U}_h^{[1]}\big(\mathcal{U}_{h/2}^{[2]}(u_n)\big)\big).$$

To get a numerical approximation, each of the two operators must be approximated. This can be done, for example, by a front-tracking method for $\mathcal{U}_h^{[1]}$ and by a standard implicit finite-difference method for the parabolic operator $\mathcal{U}_h^{[2]}$. A convergence analysis of such schemes has been carried out in Holden *et al.* (2010).

Although the analysis of splitting methods can be done by power series expansions and the formalism of Lie operators, there are fundamental differences with respect to the ODE case. Nonlinear PDEs in general possess solutions that exhibit complex behaviour in small regions of space and time, such as sharp transitions and discontinuities, and thus they lack the usual smoothness required for the analysis. Moreover, even if the exact solution of the original problem is smooth, it may well happen that the composition defining the splitting method provides non-smooth approximations. Therefore, it is necessary to develop an appropriate mathematical framework to analyse the convergence of the numerical solution to the correct solution of the original problem, and this has to be done very often on a case by case basis; see e.g. Holden *et al.* (2010) and references therein. Thus, in particular, the first- and second-order convergence of the Godunov and Strang splitting methods on the Korteweg–de Vries equation $u_t - uu_x + u_{xxx} = 0$ has been proved in Holden, Karlsen, Risebro and Tao (2011) if the initial data are sufficiently regular, whereas the result has been extended in Holden, Lubich and Risebro (2013) to equations

of the form $u_t = Au + uu_x$, when A is a linear differential operator such that the full equation is well-posed. More recently, convergence results and error estimates have also been obtained for initial conditions with low regularity, namely $u_0 \in H^s$ with $0 < s \leq 3/2$, where H^s denotes the Sobolev space (Rousset and Schratz 2022).

Another source of difficulties related to the application of splitting methods to PDEs is the treatment of boundary conditions. In this respect, we should take into account that the boundary conditions are defined for the whole operator f in (6.1), and they do not necessarily hold for the subproblems defined by each part f_1 and f_2. Therefore, we cannot expect the numerical solution obtained by a splitting method to belong to the domain of f. This results in severe order reductions in reaction–diffusion problems when Dirichlet or Neumann boundary conditions are considered (Hansen and Ostermann 2009*b*, Hundsdorfer and Verwer 2003). In particular, the order reduction for the Strang splitting is one in the infinity norm. Similar order reductions for advection–reaction problems have also been reported (Hundsdorfer and Verwer 1995).

Several procedures have been considered in the literature to avoid this order reduction in the case of reaction–diffusion problems. One possibility, proposed in Einkemmer and Ostermann (2015, 2016), consists in introducing a smooth correction function in such a way that the new reaction flow is compatible with the prescribed boundary conditions. For time-invariant Dirichlet boundary conditions, this correction can be computed only once at the beginning of the simulation, but for time-dependent Dirichlet, Neumann or Robin boundary conditions, the correction is time-dependent and has to be computed at each time step. Various techniques to deal with this problem are explored in Einkemmer, Moccaldi and Ostermann (2018). An alternative approach requiring additional calculations with grid values on the boundaries, and not on grid values on the total domain, is proposed in Alonso-Mallo, Cano and Reguera (2018, 2019).

6.2. *IMEX methods*

IMEX schemes are suitable combinations of implicit and explicit schemes and constitute a popular technique for approximating the solution of PDEs that involve terms of a different nature (Ascher, Ruuth and Wetton 1995, Ascher, Ruuth and Spiteri 1997, Hundsdorfer and Verwer 2003). Thus, for convection–diffusion or reaction–diffusion problems where the convection or reaction terms are moderately stiff, it might be appropriate to use an explicit scheme for these parts and an implicit scheme for the diffusion term. We next analyse the connections of some popular IMEX methods with splitting and composition methods.

Suppose we have the semidiscrete system $u_t = f_1(u) + f_2(u)$, where f_1 is (still) a diffusion term and f_2 a nonlinear term suitable for explicit integration. Then the simple composition (see (6.10))

$$u_{n+1} = \phi_h^3(u_n) = \psi_h^{1i} \circ \psi_h^{2e}(u_n) = u_n + h\left(f_1(u_{n+1}) + f_2(u_n)\right)$$

corresponds to the linear one-step IMEX scheme of Ascher *et al.* (1997).

Letting $\phi_h^{[2,k]}$, $k \geq 2$, denote an explicit kth-order (Runge–Kutta or multistep) method for the equation $u_t = f_2(u)$, it turns out that many IMEX methods from the literature have the structure

$$u_{n+1} = \psi_{h/2}^{1i} \circ \phi_h^{[2,k]} \circ \psi_{h/2}^{1e}(u_n), \tag{6.14}$$

thus yielding second-order approximations.

Let us consider, for instance, the popular Crank–Nicolson–leapfrog (IMEX-CNLF) method (Hundsdorfer and Verwer 2003)

$$u_{n+1} = u_{n-1} + 2hf_2(u_n) + h(f_1(u_{n+1}) + f_1(u_{n-1})), \tag{6.15}$$

to be initiated with, for example, $u_1 = u_0 + h(f_1(u_0) + f_2(u_0))$. It is equivalent to the one-step method

$$u_{n+1} = u_n + h^* f_2(u_{n+1/2}) + \frac{h^*}{2}(f_1(u_{n+1}) + f_1(u_n)), \tag{6.16}$$

with $h^* = 2h$, which requires us to compute the approximation at the midpoint and so advances every half time step. If we take the explicit second-order midpoint rule as $\phi_h^{[2,2]}$ in (6.14), then we get the following sequence of maps:

$$\begin{aligned}
\psi_{h/2}^{1e}: &\quad U = u_n + \frac{h}{2} f_1(u_n),\\
\phi_h^{[2,2]}: &\quad \left\{\begin{aligned} V &= U + \frac{h}{2} f_2(U)\\ \hat{V} &= U + hf_2(V)\end{aligned}\right\},\\
\psi_{h/2}^{1i}: &\quad u_{n+1} = \hat{V} + \frac{h}{2} f_1(u_{n+1}),
\end{aligned}$$

or equivalently

$$\begin{aligned}
U &= u_n + \frac{h}{2} f_1(u_n),\\
V &= U + \frac{h}{2} f_2(U),\\
u_{n+1} &= u_n + hf_2(V) + \frac{h}{2}(f_1(u_n) + f_1(u_{n+1})).
\end{aligned} \tag{6.17}$$

This scheme is in fact quite similar to (6.16) for $h = h^*$ since V is a first-order approximation to $u_{n+1/2}$, the solution at the midpoint. Notice that by replacing $f_2(u_{n+1/2})$ in (6.16) with $f_2(V)$, where V depends explicitly on u_n, it allows us to advance from u_{n+1} to u_{n+2} without evaluating the solutions at $u_{n+3/2}$ and therefore halving the computational cost to solve the implicit equations involved.

Although the method is not symmetric (due to the lack of symmetry of the explicit scheme $\phi_h^{[2,2]}$), by instead using an explicit method of order $k > 2$, the overall scheme will be time-symmetric up to this order k and therefore we can apply extrapolation to get a method of order k in an efficient way.

Another IMEX Runge-Kutta method that combines the implicit and explicit trapezoidal methods and shows a fairly good performance on examples is (Hundsdorfer and Verwer 2003)

$$\begin{aligned} U &= u_n + h(f_1(u_n) + f_2(u_n)), \\ u_{n+1} &= u_n + \frac{h}{2}(f_1(u_n) + f_2(u_n)) + \frac{h}{2}(f_1(u_{n+1}) + f_2(U)). \end{aligned} \tag{6.18}$$

If we take the system

$$w' = \begin{pmatrix} u \\ U \end{pmatrix}' = \begin{pmatrix} f_1(u) + f_2(U) \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ f_1(u) + f_2(u) \end{pmatrix} = g_1(w) + g_2(w), \tag{6.19}$$

and form, as before, the composition

$$w_{n+1} = \psi^{1i}_{h/2} \circ \phi^{[2]}_h \circ \psi^{1e}_{h/2}(w_n), \tag{6.20}$$

where $\phi^{[2]}_h$ denotes the exact solution of $w' = g_2(w)$, and $w_n = (U_n, u_n) = (u_n, u_n)$, then we get

$$\begin{aligned} u_{1/2} &= u_n + \frac{h}{2}(f_1(u_n) + f_2(u_n)), \\ U_{n+1} &= u_n + h(f_1(u_{1/2}) + f_2(u_{1/2})), \\ u_{n+1} &= u_n + \frac{h}{2}(f_1(u_n) + f_2(u_n)) + \frac{h}{2}(f_1(u_{n+1}) + f_2(U_{n+1})). \end{aligned} \tag{6.21}$$

Scheme (6.18) is recovered by replacing $u_{1/2}$ with u_n in the expression of U_{n+1}. However, if in (6.21) we do not restart the value of U_n at each step, the method will be symmetric and therefore we can apply extrapolation to increase its order.

Higher-order IMEX methods have been built involving implicit multistep or Runge–Kutta methods. Although splitting methods are no longer appropriate to advance the diffusion term due to the presence of negative coefficients for orders higher than two, we can incorporate higher derivatives or complex coefficients and form new higher-order splitting IMEX methods (or consider extrapolation from a basic symmetric second-order method).

6.3. *Schrödinger equations*

T–V splitting. Until now, in the context of PDEs, we have dealt with splitting methods of orders one and two. There are, however, relevant problems where high-order splitting methods can be and have been safely used, and where a rigorous convergence analysis can be established. This is the case, in particular, for the time-dependent Schrödinger equation, already considered in Section 1.6. The numerical experiments presented there clearly indicate that the Strang splitting based on kinetic and potential energy in combination with a pseudo-spectral space discretization, i.e. method (1.45), provides approximations of order two in the time

step, in accordance with

$$\mathrm{e}^{\frac{\tau}{2}V}\,\mathrm{e}^{\tau T}\,\mathrm{e}^{\frac{\tau}{2}V} = \mathrm{e}^{\tau(T+V)} + O(\tau^3(\|T\| + \|V\|)^3), \tag{6.22}$$

where $\tau = -\mathrm{i}h$. Be aware, however, that this error estimate only makes sense for bounded T and V. In fact, when the norm of T or V is very large (as is usually the case when the number of space discretization points M is large), then (6.22) is of no practical use, and thus other estimates are necessary to explain the observed good behaviour.

Error bounds for the Strang splitting are actually derived in Jahnke and Lubich (2000). They clearly show that the method is indeed of order two when applied to pseudo-spectral discretizations of the time-dependent Schrödinger equation under some regularity conditions and periodic boundary conditions. Specifically, assume that the potential $V(x)$ is $\mathcal{C}^5$-smooth, periodic and bounded, $\|V\psi\| \le \beta\|\psi\|$, $\beta > 0$. Then, if $u(x,t)$ denotes the trigonometric interpolation polynomial of the solution of the pseudo-spectral method and $u_n(x)$ is the corresponding trigonometric interpolation polynomial built from the numerical approximations obtained from the Strang splitting (1.45) at time $t = t_n = nh$, for the local and global errors we get the following bounds:

$$\begin{aligned} \|u_1 - u(\cdot,\tau)\|_{L^2} &\le C_1|\tau|^3\|u_0\|_{H^2}, \\ \|u_n - u(\cdot,t_n)\|_{L^2} &\le C_2|\tau|^2\|u_0\|_{H^2}, \end{aligned} \tag{6.23}$$

respectively. Here $\|\cdot\|_{H^2}$ denotes the usual Sobolev norm, and the constants C_1, C_2 are independent of the initial data u_0 and the discretization parameters M, n and τ, with $0 \le t_n \le t_f$ for some finite t_f. The case when V is time-dependent and bounded for any t has been recently treated in del Valle and Kropielnicka (2023), where new schemes are proposed and analysed.

The previous results for the Strang splitting have been extended in Thalhammer (2008) to splitting methods of the general form

$$u_{n+1} = \prod_{j=1}^{s} \mathrm{e}^{b_j\tau V}\,\mathrm{e}^{a_j\tau T}\,u_n = \mathrm{e}^{b_s\tau V}\,\mathrm{e}^{a_s\tau T}\,\cdots\,\mathrm{e}^{b_1\tau V}\,\mathrm{e}^{a_1\tau T}\,u_n, \tag{6.24}$$

whose coefficients a_j, b_j satisfy the order conditions up to order r. In that case

$$\|u_n - u(\cdot,t_n)\|_{L^2} \le C\|u(\cdot,0) - u_0\|_{L^2} + C|\tau|^r\|u(\cdot,0)\|_{H^r}, \quad 0 \le t_n \le t_f \tag{6.25}$$

is valid with some constant C depending on t_f, but not on n and h. This error bound implies, in particular, that the splitting methods of Section 8 retain their order of convergence when applied to the Schrödinger equation with periodic boundary conditions, provided that the data are sufficiently differentiable; see also Hansen and Ostermann (2009*a*). Otherwise, an order reduction may occur.

We can also take advantage of the property (1.41) and include the commutator $[V,[T,V]]$ in the composition (6.24), as in RKN splitting methods, so that we end

up with

$$u_{n+1} = \prod_{j=1}^{s} \mathrm{e}^{b_j \tau V + c_j \tau^3 [V,[T,V]]}\, \mathrm{e}^{a_j \tau T}\, u_n. \tag{6.26}$$

Even in this case the resulting schemes retain their order of convergence if the solution is sufficiently regular, as shown in Kieri (2015). It is then also possible to apply the RKN schemes presented in Section 8.

Symplectic splitting. In Section 3.4 we reviewed the symplectic structure involved in the (semidiscretized) Schrödinger equation and illustrated how the Strang splitting (and, for that matter, all methods presented in Section 8) can be applied if H is a real and symmetric matrix. They are formulated as products of exponentials of the nilpotent matrices A and B given in (3.13),

$$\begin{pmatrix} q_{n+1} \\ p_{n+1} \end{pmatrix} = \prod_{j=1}^{s} \mathrm{e}^{b_j hB}\, \mathrm{e}^{a_j hA} \begin{pmatrix} q_n \\ p_n \end{pmatrix} = \exp\left[h \begin{pmatrix} 0 & H \\ -H & 0 \end{pmatrix} \right] \begin{pmatrix} q_n \\ p_n \end{pmatrix} + O(h^{r+1}), \tag{6.27}$$

and orders $r = 2, 4, 6, 8, 10$ and 12 have been achieved with only $s = r$ exponentials $\mathrm{e}^{a_j hA}$ and $\mathrm{e}^{b_j hB}$ (Gray and Manolopoulos 1996, Zhu, Zhao and Tang 1996, Liu, Ding, Hong and Wang 2005).

The processing technique has also been used to construct splitting schemes with two different goals in mind: to attain maximal stability and maximal accuracy. They have the general structure $P^{-1}(hH)K(hH)P(hH)$, where K (the kernel) is built as a composition (6.27) with a large number of stages s, and P (the processor) is taken as a polynomial. Although these methods are neither unitary nor unconditionally stable, they are symplectic and conjugate to unitary schemes. In consequence, neither the average error in energy nor the norm of the solution increases with time. Specifically, Blanes, Casas and Murua (2006*b*, 2008*a*) have proposed kernels with up to 19, 32 and 38 stages, either to construct methods of orders $r = 10$, 16 and 20, or to bring highly accurate *second*-order methods with an enlarged stability domain.

This approach to approximating $\mathrm{e}^{\tau H} u$ is closely related to other polynomial approximations of the form

$$\mathrm{e}^{\tau H} u \approx P_m(hH)u, \tag{6.28}$$

where $P_m(y)$ is a polynomial in y approximating the exponential $\mathrm{e}^{-\mathrm{i}y}$. Different choices for such $P_m(y)$ are available: truncated Taylor or Chebyshev series expansions of $\mathrm{e}^{-\mathrm{i}y}$ for an appropriate real interval of y, or a Lanczos approximation, where the polynomial is determined by a Galerkin approximation on the Krylov space spanned by $u, Hu, \ldots, H^{m-1}v$ (Lubich 2008).

Given a prescribed error tolerance, some appropriate estimates of the upper and lower bounds of the eigenvalues of the matrix H, $E_{\min}$ and $E_{\max}$, and a time integration interval, $[t_0, t_f]$, in the Chebyshev approach we choose, according to

some known error bounds, the lowest-degree polynomial that provides the solution with such accuracy. The coefficients of the polynomial are determined for each case and the action of the polynomial on a vector is computed recursively using the Clenshaw algorithm (Lubich 2008). On the other hand, in the Taylor approach we have to adjust the maximum degree allowed with the time step h to reach the desired accuracy with the minimum number of matrix–vector products. As a result, in general, Chebyshev turns out to be between two and three times faster than Taylor, depending on the final time at which the output is desired. Chebyshev and Lanczos approximations have quite similar error bounds (see Lubich 2008) and their relative performance depends on the particular problem considered.

It is worth remarking that, whereas in the approach (6.28) the approximation of $\mathrm{e}^{\tau H}u$ is constructed by computing products of the form Hu, where $u \in \mathbb{C}^M$, with symplectic splitting methods of the form (6.27) we proceed by successively computing *real* matrix–vector products Hq and Hp with different weights. With splitting methods, the real and imaginary parts of $\mathrm{e}^{\tau H}u \equiv \mathrm{e}^{-\mathrm{i}hH}(q + \mathrm{i}p)$ are approximated in a different way, with a considerably reduced computational cost.

Blanes *et al.* (2015) construct several optimized symplectic splitting methods and present an algorithm that automatically selects the most efficient one for a prescribed error tolerance under the same conditions as when using the Chebyshev method. The resulting algorithm is between 1.4 and 2 times faster than the Chebyshev method for the same accuracy, with reduced energy and unitarity errors for large values of h. The computation of the coefficients of the schemes is largely based on the stability and error analysis of splitting methods carried out in Blanes, Casas and Murua (2008*a*, 2011).

In contrast to T–V splitting methods, which preserve unitarity by construction and are thus unconditionally stable, the previous polynomial approximations suffer from a step size restriction. Given h, Δx and m, the degree of the polynomial, these methods must satisfy the restriction

$$\frac{h}{m\Delta x^2} \leq C,$$

so the time interval that one can advance per matrix–vector product is proportional to Δx^2 or, equivalently, the number of matrix–vector products to reach the final time is inversely proportional to Δx^2.

Time-dependent potentials. We have assumed so far that the potential in the Schrödinger equation (1.40) does not depend explicitly on time or, if it does, it only varies slowly with time, so that in each sub-interval $[t_n, t_{n+1}]$ the corresponding matrix V is obtained from the average of $V(x,t)$ on this interval. In general, however, we have to deal with situations in which this approximation is no longer valid. In that case, and in contrast to other approaches based on Chebyshev or Lanczos approximations, splitting methods can still be applied with some appropriate modifications.

In a similar way to classical Hamiltonian problems, one may take t in the potential as an additional coordinate, $x_{d+1} = t$, introduce its canonical momentum, $p_{d+1} = -i\,\partial_{x_{d+1}}$, and deal with the system in the extended phase space, $\tilde{H} = (T + p_{d+1}) + V(x, x_{d+1})$. Since the action of the operator p_{d+1} just corresponds to a shift in the variable x_{d+1}, i.e. $\mathrm{e}^{ah\partial_t} V(t,x) = V(t+ah,x)$, and moreover

$$\mathrm{e}^{ah\partial_t}\,\mathrm{e}^{-\mathrm{i}bhV(t,x)}\,\mathrm{e}^{-ah\partial_t} = \mathrm{e}^{-\mathrm{i}bh\mathrm{e}^{ah\partial_t}V(t,x)} = \mathrm{e}^{-\mathrm{i}bhV(t+ah,x)},$$

composition (6.24) applied to the corresponding non-autonomous problem now simply reads (notice that T and p_{d+1} commute)

$$u_{n+1} = \prod_{j=1}^{s} \mathrm{e}^{b_j \tau V(t_n + c_j h)}\,\mathrm{e}^{a_j \tau T}\,u_n, \quad \text{with} \quad c_j = \sum_{k=1}^{j} a_k,\ j \in \{1,\dots,s\}, \tag{6.29}$$

and analogously if the scheme includes modified potentials; see also e.g. Chin and Chen (2002) and references therein.

For a generic time-dependent Hamiltonian $H(t)$, the Schrödinger equation $\mathrm{i}\partial_t\psi = H(t)\psi$ can be recast as a non-autonomous evolution equation of the form

$$u'(t) = A(t)u(t), \quad u(t_0) = u_0, \quad t \in [t_0, t_f], \tag{6.30}$$

defined by a family of time-dependent linear operators $(A(t))_{t\in[t_0,t_f]}$, which, assuming a spatial discretization has been carried out, are generally complex matrices of large dimension and large norm.

It turns out that standard rth-order splitting methods defined by coefficients $(a_j, b_j)_{j=1}^{s}$ can be applied in this setting simply by adding the trivial relation $(\mathrm{d}/\mathrm{d}t)t = 1$ to equation (6.30). This results in the scheme

$$u_{n+1} = \prod_{j=1}^{s} \mathrm{e}^{hb_j A(t_n + c_j h)} u_n, \quad \text{with} \quad c_j = \sum_{k=1}^{j} a_k,\ j \in \{1,\dots,s\} \tag{6.31}$$

of the same formal order r as the method originally designed for autonomous problems.

A different approach is based on the use of the Magnus expansion (Magnus 1954) to get a formal solution representation of (6.30) as the exponential of an infinite series:

$$u(t_n + h) = \mathrm{e}^{\Omega(h)} u_n, \quad \Omega(h) = \sum_{m=1}^{\infty} \Omega_m(h), \tag{6.32}$$

where each term Ω_m involves multiple integrals of nested matrix-commutators (Blanes, Casas, Oteo and Ros 2009). By appropriately truncating this series and

approximating the integrals by quadratures, efficient integrator schemes can be constructed (Iserles, Munthe-Kaas, Nørsett and Zanna 2000). Thus, for instance, taking the two-stage Gauss–Legendre quadrature rule with nodes

$$c_{1,2} = \frac{1}{2} \mp \frac{\sqrt{3}}{6}$$

results in the scheme

$$\begin{aligned} \Omega^{[4]}(h) &= \frac{1}{2}h(A_1 + A_2) + \frac{\sqrt{3}}{12}h^2[A_2, A_1], \\ u_{n+1} &= \mathrm{e}^{\Omega^{[4]}(h)} u_n, \end{aligned} \tag{6.33}$$

with $A_j = A(t_n + c_j h)$, $j = 1, 2$. For the Schrödinger equation with $A(t) = -\mathrm{i}(T + V(t))$ and a smooth time-dependent potential $V(t)$, Hochbruck and Lubich (2003) showed that Magnus integrators retain their full order of convergence (without bounds on T in the error bound) for sufficiently regular solutions, uniformly with respect to the space discretization.

There are, however, several issues related to Magnus integrators due to the presence of iterated commutators. Thus, computing the action of iterated commutators on vectors can be very costly due to the number of matrix–vector products required. This is particularly relevant when considering problems in two and three space dimensions (Bader, Iserles, Kropielnicka and Singh 2016). In addition, the evolution equations defining high-order Magnus integrators in general involve differential operators of different nature from the original problem (Blanes, Casas, González and Thalhammer 2021*b*).

A different class of exponential integrators that circumvent these difficulties whilst still retaining the favourable properties of Magnus integrators is formed by the so-called commutator-free quasi-Magnus (CFQM) methods: the basic idea is to replace the single exponential in (6.32) with a composition of several exponentials involving linear combinations of the values of the operator A at certain nodes, c_k, of a quadrature rule:

$$\begin{gathered} u_{n+1} = \mathrm{e}^{h\,B_{nJ}} \cdots \mathrm{e}^{h\,B_{n1}}\, u_n \;\approx\; u(t_{n+1}) = \mathrm{e}^{\Omega(h)}\, u(t_n), \\ c_k \in [0,1], \quad A_{nk} = A(t_n + c_k h), \quad k \in \{1, \dots, K\}, \\ B_{nj} = a_{j1}\, A_{n1} + \cdots + a_{jK}\, A_{nK}, \quad j \in \{1, \dots, J\}. \end{gathered}$$

Particular examples of CFQM exponential integrators are the exponential midpoint rule (order two)

$$\begin{gathered} J = K = 1, \quad c_1 = \tfrac{1}{2}, \quad a_{11} = b_1 = 1, \\ u_{n+1} = \mathrm{e}^{h\,A(t_n + \frac{1}{2}h)}\, u_n \end{gathered} \tag{6.34}$$

and the fourth-order scheme

$$\begin{gathered} J = K = 2, \quad \alpha = \tfrac{\sqrt{3}}{6}, \quad c_1 = \tfrac{1}{2} - \alpha, \quad c_2 = \tfrac{1}{2} + \alpha, \\ a_{11} = a_{22} = \tfrac{1}{4} + \alpha, \quad a_{12} = a_{21} = \tfrac{1}{4} - \alpha, \\ B_j(h) = a_{j1}\, A(t_n + c_1 h) + a_{j2}\, A(t_n + c_2 h), \quad j \in \{1, 2\}, \\ u_{n+1} = \mathrm{e}^{h\, B_2(h)}\, \mathrm{e}^{h\, B_1(h)}\, u_n. \end{gathered} \tag{6.35}$$

A detailed treatment and specific schemes up to order six can be found in Blanes, Casas and Thalhammer (2017*b*, 2018) and up to order eight in Alvermann and Fehske (2011), whereas Blanes *et al.* (2021*b*) have proved that CFQM methods applied to the Schrödinger equation with Hamiltonian $H(t) = -\frac{1}{2}\Delta + V(t)$ are unconditionally stable in the underlying Hilbert space and retain full order of convergence under low regularity requirements on the initial state.

Semiclassical regime. The so-called semiclassical Schrödinger equation

$$\mathrm{i}\varepsilon \partial_t \psi(x, t) = \left(-\frac{\varepsilon^2}{2} \Delta + V(x) \right) \psi(x, t) \tag{6.36}$$

(in atomic units), with a small parameter $\varepsilon \ll 1$, arises in particular when applying the time-dependent Born–Oppenheimer approximation for the motion of nuclei as driven by the potential energy surface of the electrons (Lubich 2008). In that case ε^2 represents the mass ratio of nuclei and electrons. Recall that (6.36) has highly oscillatory solutions with wavelength $\sim \varepsilon$, so grid-based numerical schemes require a resolution of this order in both space and time, which is computationally very expensive. One of the challenges, therefore, is to construct numerical methods that are robust in the limit $\varepsilon \to 0$.

Several options have been proposed and analysed in detail; see, for instance, the recent review by Lasser and Lubich (2020). Among others, we can recount the following.

- Split the equation into the usual kinetic and potential energy parts and apply the Strang splitting in time in combination with trigonometric spectral methods (Bao, Jin and Markowich 2002). Although the resulting scheme is unconditionally stable, time-reversible and preserves the position density, it requires very fine resolution, in both space and time, for small ε (Jin, Markowich and Sparber 2011).
- Use a Gaussian wave packet as an approximation for the wave function $\psi(x, t)$ depending on certain parameters and apply a variational splitting to get approximate solutions for the differential equations they satisfy (Faou and Lubich 2006). The resulting algorithm is symplectic, time-reversible and preserves the unit L^2-norm of the wave packets.
- Another variant of this approach instead consists in taking Hagedorn wave packets. They provide a spectral approximation in space with a time-dependent set of basis functions giving the exact solution of the Schrödinger

equation with the potential locally approximated by a quadratic function. The potential $V(x)$ is split into the quadratic term $U(q(t), x)$ in the Taylor expansion of V, around the time-dependent classical position $q(t)$ and the remainder (Faou, Gradinaru and Lubich 2009). The overall algorithm has a number of conservation and limit properties, as listed in Faou and Lubich (2006), for example. In addition, highly efficient splitting methods for perturbed problems can also be applied (Blanes and Gradinaru 2020).

- A different class of exponential splittings is proposed in Bader, Iserles, Kropielnicka and Singh (2014) for the one-dimensional case. Essentially, the formal solution of the space-discretized equation $u' = \mathrm{i}(\varepsilon T - \varepsilon^{-1}V)u$ is approximated as

$$\mathrm{e}^{\mathrm{i}h(\varepsilon T - \varepsilon^{-1}V)} \approx \mathrm{e}^{R_0}\mathrm{e}^{R_1}\cdots\mathrm{e}^{R_s}\mathrm{e}^{T_{s+1}}\mathrm{e}^{R_s}\cdots\mathrm{e}^{R_1}\mathrm{e}^{R_0}$$

with error $O(\varepsilon^{2s+2})$. Here $R_0 = O(\varepsilon^0)$, $R_k = O(\varepsilon^{2k-2})$, $k \geq 1$, and $T_{s+1} = O(\varepsilon^{2s})$. In this approach the number of exponentials grows linearly with s and the exponentials can be computed efficiently, although the terms R_k and T_{s+1} contain nested commutators.

Nonlinear Schrödinger equations. Introducing nonlinear effects in the Schrödinger equation allows us to model some relevant physical phenomena taking place in nonlinear optics, quantum superfluids, plasmas, water waves, etc.; see e.g. Sulem and Sulem (1999) and references therein. Consider in particular a Bose–Einstein condensate (BEC), the ground state of a system of interacting bosons very close to zero temperature. It was first predicted by Einstein in 1925 and experimentally realized by Anderson *et al.* (1995). Mathematically, a BEC of an atomic species trapped in an external potential $V(x)$ is modelled by the (normalized) Gross–Pitaevskii equation (GPE)

$$\mathrm{i}\partial_t\psi(x,t) = \left(-\frac{1}{2}\Delta + V(x) + \sigma|\psi(x,t)|^2\right)\psi(x,t), \tag{6.37}$$

with asymptotic boundary conditions $\psi(x,t) \to 0$ as $|x| \to \infty$. Here the parameter σ originates from the mean-field interaction between the particles: repulsive forces lead to $\sigma > 0$, whereas $\sigma < 0$ represents attractive forces. Equation (6.37) has been the subject of many different studies, including the existence of solutions and its numerical treatment. Concerning the first aspect, we refer to Cazenave (2003), Carles (2008) and references therein.

With respect to the numerical integration of the GPE equation, a combination of spectral discretization in space with splitting methods in time constitutes a natural option, and in fact has been explored in detail in the literature; see e.g. Bao, Jin and Markowich (2003*b*), Bao, Jaksch and Markowich (2003*a*, 2004) and Thalhammer, Caliari and Neuhauser (2009). If (6.37) is expressed as

$$\mathrm{i}\partial_t\psi(x,t) = (A + B(x,\psi))\psi(x,t), \quad \psi(x,0) = \psi_0(x), \tag{6.38}$$

with

$$A = -\frac{1}{2}\Delta, \quad B(x,\psi) = V(x) + \sigma|\psi|^2, \tag{6.39}$$

it is clear that the solution of the initial value problem $\mathrm{i}\psi_t = A\psi$, $\psi(x,0) = \psi_0(x)$ is

$$\psi(x,t) = \mathrm{e}^{-\mathrm{i}tA}\,\psi_0(x)$$

and (an approximation to) $\psi(x,t)$ is obtained by representing the initial value with respect to the (truncated) Fourier basis functions.

On the other hand, given a real function G, the solution of

$$\mathrm{i}\partial_t\psi(x,t) = G(x,|\psi(x,t)|)\,\psi(x,t) \tag{6.40}$$

leaves the norm invariant, i.e. $|\psi(x,t)| = |\psi(x,0)|$, and therefore

$$\psi(x,t) = \mathrm{e}^{-\mathrm{i}tG(x,|\psi(x,0)|)}\,\psi(x,0). \tag{6.41}$$

In consequence, the initial value problem $\mathrm{i}\partial_t\psi = B\psi$, $\psi(x,0) = \psi_0(x)$ is also solvable. Splitting methods can also be applied in the more general situation when the potential V is explicitly time-dependent, as explained earlier for the linear Schrödinger equation.

6.4. Parabolic evolution equations

Let us now consider the evolution equation

$$u'(t) = Lu(t) = Au(t) + Bu(t), \quad t \geq 0, \quad u(0) = u_0, \tag{6.42}$$

where the linear, possibly unbounded, operators A, B and L generate C_0 semigroups over an infinite-dimensional Banach space X. We recall that if L is the infinitesimal generator of the C_0 semigroup $T(t)$ on X and $u_0 \in \mathcal{D}(L)$, the domain of L (which is dense in X), then $u(t) = T(t)u_0$ is a classical solution of (6.42) (Partington 2004). Since in the special case of a bounded linear operator L the solution is given by the familiar expression $u(t) = \mathrm{e}^{tL}u_0$ (Engel and Nagel 2006), the semigroup $T(t)$ is also denoted by the symbol e^{tL}; see Engel and Nagel (2006), Pazy (1983) and Yosida (1971) for an introduction to the theory of C_0 semigroups.

A prototypical example is the linear heat equation with potential

$$u_t(t,x) = \frac{1}{2}\Delta u(t,x) - V(x)u(t,x). \tag{6.43}$$

Here $V(x) \geq 0$, $t \geq 0$ and $x \in \mathbb{R}^d$ (or $x \in \mathbb{T}^d$). In that case $(Au)(x) = \frac{1}{2}\Delta u(x)$, $(Bu)(x) = -V(x)u(x)$ and A generates only a C_0 semigroup. This can be seen by considering the equation $u_t = \frac{1}{2}\Delta u$ on $0 < x < 1$ with Dirichlet boundary conditions: the kth Fourier mode of the solution is $c_k\mathrm{e}^{-\frac{1}{2}(k\pi)^2 t}$, which is generally not well-defined for $t < 0$.

Hansen and Ostermann (2009*a*) have established the following result. Assuming that $\|\mathrm{e}^{tA}\| \leq \mathrm{e}^{\omega t}$, $\|\mathrm{e}^{tB}\| \leq \mathrm{e}^{\omega t}$ for the same value of $\omega \geq 0$ and all $t \geq 0$, and

that for any operator E_{r+1} obtained as the product of exactly $r+1$ factors chosen amongst A and B, there is a constant $\tilde{C} > 0$ such that

$$\|E_{r+1} \mathrm{e}^{t(A+B)} u_0\| \leq \tilde{C}.$$

Then a splitting method $\Psi(h) = \prod_{j=1}^{s} \mathrm{e}^{b_j hB}\, \mathrm{e}^{a_j hA}$, with all $a_j \geq 0$ and $b_j \geq 0$ and of classical order r, retains its order when applied to (6.42):

$$\|(\Psi(h)^n - \mathrm{e}^{nhL})u_0\| \leq Ch^r, \quad \text{for } n\,h \leq t_f\,, \tag{6.44}$$

where the constant C is independent of n and h on the bounded time interval $[0, t_f]$.

In practice, however, the positivity requirement on the coefficients restricts the splitting method to be of at most order two. If, in addition, $[B, [A, B]]$ is a bounded operator and $\|\mathrm{e}^{t\,[B,[A,B]]}\| \leq \mathrm{e}^{\omega t}$, then the same result (6.44) also holds for splitting methods involving double commutators,

$$\Psi(h) = \prod_{j=1}^{s} \mathrm{e}^{b_j hB + c_j h^3 [B,[A,B]]}\, \mathrm{e}^{a_j hA}$$

and positive a_j, b_j (Kieri 2015). Efficient schemes within this class up to order four specially tailored to the problem at hand have recently been proposed in Blanes *et al.* (2023).

Example: imaginary time propagation. The imaginary-time evolution method is a well-known approach to computing the ground state (and its corresponding eigenvalue) of a quantum system with Hamiltonian $H = -\frac{1}{2}\Delta + V$ (Auer, Krotscheck and Chin 2001, Lehtovaara, Toivanen and Eloranta 2007, Bader *et al.* 2013). Essentially, under the time transformation $t = -\mathrm{i}s$, the time-dependent Schrödinger equation (1.40) is transformed into ($\hbar = 1$)

$$\partial_s \psi(x,s) = \frac{1}{2}\Delta\psi(x,s) - V(x)\psi(x,s), \quad \psi(x,0) = \psi_0(x), \tag{6.45}$$

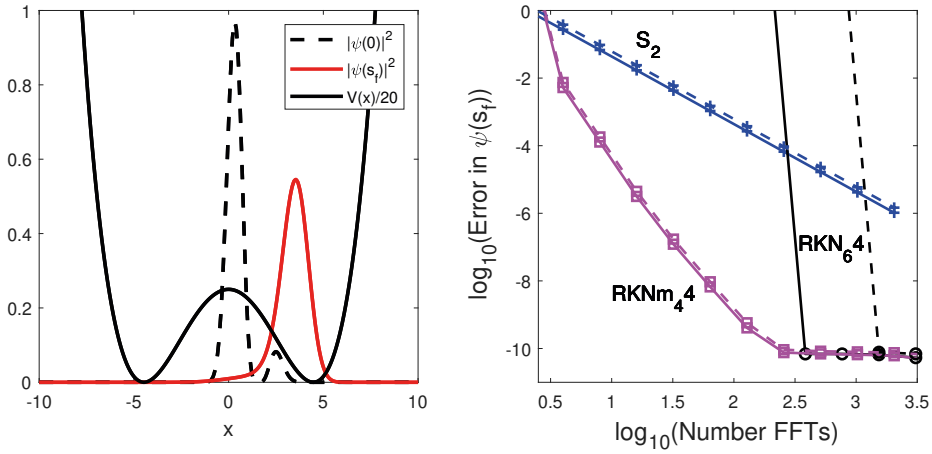
that is, a linear heat equation of the form (6.43). If we denote the (real) eigenvalues of H as E_j, with $E_0 < E_1 < \cdots$, and the corresponding eigenfunctions as ϕ_j, $j = 0, 1, 2, \ldots$, the initial wave function $\psi_0(x)$ can be expanded in the orthonormal basis $\{\phi_j\}$,

$$\psi_0(x) = \sum_{j\geq 0} c_j\, \phi_j(x), \quad c_j = \langle \phi_j \,|\, \psi(\cdot, 0)\rangle,$$

where $\langle \cdot \,|\, \cdot \rangle$ is the usual L^2-scalar product. Then the solution of (6.45) can be written as

$$\psi(x,s) = \mathrm{e}^{-sH}\psi(x,0) = \sum_{j\geq 0} \mathrm{e}^{-sE_j}\, c_j\, \phi_j(x). \tag{6.46}$$

Notice that for sufficiently large s we get $\psi(x,s) \to \mathrm{e}^{-sE_0} c_0 \phi_0$, since the other exponentials decay more rapidly. In other words, any given wave function at $s = 0$ for which $c_0 \neq 0$, converges towards the ground state solution when $s \to \infty$.



(a) (b)

Figure 6.1. Imaginary time propagation of the Schrödinger equation. (a) Double-well potential, initial and (normalized) final wave function obtained with a mesh with $N = 256$ and repeated with $N = 510$ (the two results overlap visually). (b) Discrete L_2-norm error in the normalized solution at the final time, $s_f = 1$, vs. number of FFTs for different values of the time step obtained with the same methods as in Figure 1.5: solid lines for $M = 256$ and dashed lines for $M = 512$ discretization points of the space interval.

Once an accurate approximation to ϕ_0 is obtained, the associated eigenvalue E_0 is easily obtained by computing $E_0 = \langle \phi_0 \,|\, H\phi_0 \rangle$. Other functions ϕ_j can also be approximated, for example by propagating different wave functions simultaneously in time (Aichinger and Krotscheck 2005).

To illustrate the technique, we next consider the same example as in Section 1.6, but now integrating in imaginary time up to $s_f = 1$. We take the same initial conditions (which is slightly closer to one of the minima of the potential) and normalize the solution at the final time. Figure 6.1(a) shows the potential, initial conditions and the normalized wave function at the final time when the spatial interval, $x \in [-13, 13]$, is divided into $M = 256$ and $M = 512$ parts, leading to the same visual results. In Figure 6.1(b) we show the L_2-norm error in the normalized wave function at the final time versus the number of FFTs required by the same methods as previously, both for $M = 256$ (solid lines) and $M = 512$ (dashed lines). We notice that one scheme suffers from step size restriction because it involves negative coefficients, and this is inversely proportional to Δx^2. However, the schemes with positive coefficients are insensitive to the spatial mesh.

One possible way to circumvent this order barrier for splitting methods when applied to problem (6.42) consists in considering schemes with *complex* coefficients a_i, b_i having *positive* real part. As shown in Castella, Chartier, Descombes and

Vilmart (2009) and Hansen and Ostermann (2009*b*), any splitting method

$$\Psi(h) = \prod_{j=1}^{s} \mathrm{e}^{b_j hB}\, \mathrm{e}^{a_j hA}$$

within this class still retains its classical order if the previous assumptions are conveniently modified. Specifically, we have to extend the notion of a C_0 semigroup $T(t)$ to the sector Σ_θ in the complex plane, for some angle $0 < \theta < \pi/2$,

$$\Sigma_\theta = \{t \in \mathbb{C} \colon |\arg(t)| < \theta\},$$

so that $T(t)$ is analytic in t for all $t \in \Sigma_\theta$ (Engel and Nagel 2006, Pazy 1983). Thus, if

- L, A and B generate analytic semigroups on X (now a complex Banach space) in the sector Σ_θ, $0 < \theta < \pi/2$, with
$$\|\mathrm{e}^{tA}\| \le \mathrm{e}^{\omega|t|}, \quad \|\mathrm{e}^{tB}\| \le \mathrm{e}^{\omega|t|}$$
for $\omega \ge 0$ and all $t \in \Sigma_\theta$,
- $\|E_{r+1}\mathrm{e}^{t(A+B)}u_0\| \le C$, where E_{r+1} is a composition of the operators A and B that consist of exactly $r+1$ factors,

then the splitting method of classical order r with coefficients $a_i, b_i \in \Sigma_\theta \subset \mathbb{C}$ satisfies

$$\|(\Psi(h)^n - \mathrm{e}^{nhL})u_0\| \le Ch^r, \quad 0 \le nh < t_f\,. \tag{6.47}$$

Since $\Psi(h)u_0$ is complex-valued, this approach cannot be applied in principle when the operators A and B are real, i.e. for problems defined in a real Banach space X. The most straightforward remedy consists in projecting the numerical solution after each time step on the real axis, i.e. computing the approximations $u_n \approx u(t_n)$, as $u_n = \mathrm{Re}(\Psi(h)u_{n-1})$. In that case we still have the bound (6.47) for the resulting integration scheme (Hansen and Ostermann 2009*b*).

Although this is only valid in the linear case, similar results are observed in practice in the nonlinear heat equation $\partial_t u = \Delta u + F(u)$ with periodic boundary conditions, at least when F is analytic (Castella *et al.* 2009, Blanes, Casas, Chartier and Murua 2013*a*), so they constitute a strong motivation to study splitting methods with complex coefficients in general. This is precisely the subject of the following section.

7. Splitting methods with complex coefficients

It has been known for a long time that, besides real solutions, the order conditions arising from splitting and composition methods (see Section 2) also admit complex solutions (Bandrauk and Shen 1991, Suzuki 1990, 1991, 1995). In fact, some of the resulting methods were explored in the context of Hamiltonian (Chambers 2003)

and quantum mechanics (Bandrauk, Dehghanian and Lu 2006, Prosen and Pizorn 2006) before being used more recently to overcome the order barrier in parabolic partial differential equations (Castella *et al.* 2009, Hansen and Ostermann 2009*b*).

The use of the complex plane to solve problems formulated in the real line has proved to be very fruitful in many branches of mathematics, as illustrated by Painlevé's famous dictum.[6] By applying the same logic in the particular case of splitting methods, it is reasonable to ask what benefits (if any) might result from carrying out the integration along paths in the complex plane.

We have already seen at the end of Section 6 that a strategy to develop effective methods of order higher than two for systems evolving in a semigroup, such as the heat equation,[7] consists precisely in applying splitting schemes with complex coefficients having positive real part. Moreover, the large number of complex solutions for the order conditions might offer more flexibility in the final choice of coefficients, and perhaps lead to schemes with smaller truncation errors and new symmetries.

On the other hand, when dealing with problems formulated in the real line, the use of complex arithmetic introduces an additional computational cost with respect to methods with purely real coefficients. From a theoretical point of view, the vector field appearing in the differential equation has to be analytic at least in a domain containing the path where the actual integration is carried out in the complex plane. Otherwise, order reductions are to be expected unless the implementation is not conveniently adapted.

In this section we summarize some of the issues involved in the construction and analysis of splitting methods with complex coefficients, and also review some of their most salient properties with regard to preservation of qualitative properties, in both classical Hamiltonian and quantum systems.

For brevity, henceforth we denote a complex number a with positive real part by writing $a \in \mathbb{C}_+$, so we are only interested in methods whose coefficients $a_j, b_j \in \mathbb{C}_+$.

7.1. Compositions

Most of the existing splitting methods with complex coefficients have been constructed by applying the composition technique of Section 2.1. Here, for completeness, we review the process, starting with the Lie–Trotter scheme $\chi_h = \varphi_h^{[2]} \circ \varphi_h^{[1]}$ as the basic method and composing it with different weights. The simplest situation corresponds of course to

$$\phi_h^{[2]} \equiv \chi_{\gamma_{1,2}h} \circ \chi_{\gamma_{1,1}h}. \tag{7.1}$$

[6] 'Il apparut que, entre deux vérités du domaine réel, le chemin le plus facile et le plus court passe bien souvent par le domaine complexe' (Painlevé 1900).

[7] This in fact constitutes Problem 10 in the list of open problems posed in McLachlan and Quispel (2002).

In accordance with Section 2.1.2, $\phi_h^{[2]}$ is of order two if the coefficients satisfy equation (2.14), that is,

$$\gamma_{1,1} = \frac{1}{2} \pm \frac{1}{2}\mathrm{i}, \quad \gamma_{1,2} = \overline{\gamma}_{1,1} = \frac{1}{2} \mp \frac{1}{2}\mathrm{i}.$$

If we try to construct a method of order three, a composition of at least four maps χ_h is needed, since in that case there are four order conditions. Although there are eight solutions (+cc), only four have positive real part. In fact, two of these solutions result from composing $\phi_h^{[2]}$, namely

$$\phi_h^{[3]} = \phi_{\gamma_{2,2}h}^{[2]} \circ \phi_{\gamma_{2,1}h}^{[2]}, \tag{7.2}$$

by requiring that $\gamma_{2,1} + \gamma_{2,2} = 1$, $\gamma_{2,1}^3 + \gamma_{2,2}^3 = 0$. Notice that in (7.2) $\gamma_{2,2} = \overline{\gamma}_{2,1}$. This procedure can in principle be repeated by considering the recurrence

$$\phi_h^{[r+1]} = \phi_{\gamma_{r,m_r}h}^{[r]} \circ \cdots \circ \phi_{\gamma_{r,1}h}^{[r]}, \quad r = 1, 2, \ldots \tag{7.3}$$

to construct higher-order methods with *any* first-order integrator $\phi_h^{[1]}$ (not necessarily the Lie–Trotter scheme). It turns out, however, that $\phi_h^{[4]}$ and higher-order schemes obtained from (7.3) have at least one coefficient with negative real part (Blanes *et al.* 2013*a*). Thus, if we are interested in schemes having only coefficients in $\mathbb{C}_+$, $r = 3$ indeed constitutes an order barrier for this type of composition (Hansen and Ostermann 2009*b*).

Let us now consider the Strang splitting as the basic method. Then composition (2.1) with $s = 2$ already provides a method of order three:

$$S_h^{[3]} = S_{\gamma_2 h}^{[2]} \circ S_{\gamma_1 h}^{[2]}, \quad \text{with} \quad \gamma_1 = \frac{1}{2} \pm \mathrm{i}\frac{\sqrt{3}}{6}, \ \gamma_2 = \overline{\gamma}_1. \tag{7.4}$$

Again, higher-order methods can be obtained by applying the recursive procedure

$$S_h^{[r+1]} = S_{\gamma_{r,2}h}^{[r]} \circ S_{\gamma_{r,1}h}^{[r]}, \quad r = 2, 3, \ldots, \tag{7.5}$$

with $\gamma_{r,1} + \gamma_{r,2} = 1$, $\gamma_{r,1}^{r+1} + \gamma_{r,2}^{r+1} = 0$ (see (2.52)). The solution with the smallest phase is given by

$$\gamma_{k,1} = \frac{1}{2} \pm \frac{\mathrm{i}}{2} \tan\left(\frac{\pi}{2(r+1)}\right), \quad \gamma_{r,2} = \overline{\gamma}_{r,1}, \tag{7.6}$$

but only schemes up to order six with coefficients in $\mathbb{C}_+$ are possible with this approach (Hansen and Ostermann 2009*b*).

As we have already seen, the triple jump (2.4) allows us to raise the order by two. Thus, starting from $S_h^{[2]}$, the composition

$$S_h^{[4]} = S_{\gamma_3 h}^{[2]} \circ S_{\gamma_2 h}^{[2]} \circ S_{\gamma_1 h}^{[2]} \tag{7.7}$$

has, apart from the real solution (2.3),

$$\gamma_1 = \gamma_3 = \frac{1}{2 - 2^{1/3}\mathrm{e}^{2\mathrm{i}\ell\pi/3}}, \quad \gamma_2 = 1 - 2\alpha_1, \quad \ell = 1, 2, \tag{7.8}$$

corresponding to a time-symmetric composition, and

$$\gamma_1 = \overline{\gamma}_3 = \frac{1}{4} \pm \mathrm{i}\,\frac{1}{4}\sqrt{\frac{5}{3}}, \quad \gamma_2 = \frac{1}{2}, \tag{7.9}$$

possessing the same symmetry as (7.2) and (7.4). Palindromic methods up to order eight with coefficients having positive real part are possible by applying the triple jump composition, whereas if we instead consider the *quadruple jump*,

$$S^{[2k]}_{\gamma_{k,1}h} \circ S^{[2k]}_{\gamma_{k,2}h} \circ S^{[2k]}_{\gamma_{k,2}h} \circ S^{[2k]}_{\gamma_{k,1}h},$$

we can achieve order 14 with all coefficients in $\mathbb{C}_+$. These order barriers have been rigorously proved in Blanes *et al.* (2013*a*).

Compositions (7.4) and (7.7) can be reformulated as splitting methods when $S^{[2]}_h$ is the Strang splitting. Thus

$$S^{[3]}_h = \varphi^{[1]}_{\overline{a}_1 h} \circ \varphi^{[2]}_{\overline{b}_1 h} \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_1 h}, \tag{7.10}$$

with

$$a_1 = \frac{1}{4} \pm \mathrm{i}\frac{\sqrt{3}}{12}, \quad b_1 = \frac{1}{2} \pm \mathrm{i}\frac{\sqrt{3}}{2}, \quad a_2 = \frac{1}{2},$$

whereas the palindromic version of (7.7) reads

$$S^{[4]}_{h,P} = \varphi^{[1]}_{a_1 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_2 h} \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_1 h}, \tag{7.11}$$

with

$$a_1 = \frac{\gamma_1}{2}, \quad b_1 = \gamma_1, \quad a_2 = \frac{1}{2} - a_1, \quad b_2 = 1 - 2b_1,$$

respectively. Finally, solution (7.9) leads to

$$S^{[4]}_{h,C} = \varphi^{[1]}_{\overline{a}_1 h} \circ \varphi^{[2]}_{\overline{b}_1 h} \circ \varphi^{[1]}_{\overline{a}_2 h} \circ \varphi^{[2]}_{b_2 h} \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_1 h}, \tag{7.12}$$

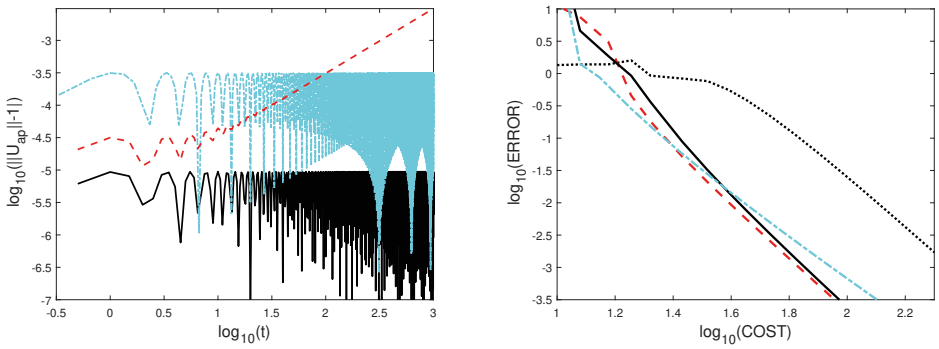
with

$$a_1 = \frac{b_1}{2}, \quad b_1 = \frac{1}{4} \pm \mathrm{i}\,\frac{1}{4}\sqrt{\frac{5}{3}}, \quad a_2 = \frac{1}{4}(2b_1 + 1), \quad b_2 = \frac{1}{2}.$$

Before proceeding further, it is useful to illustrate how the different symmetries of these third- and fourth-order schemes manifest in practice on a simple example.

Example: two-level system. We consider the time evolution of a two-level quantum system, described by (Messiah 1999)

$$\mathrm{i}\frac{\mathrm{d}U}{\mathrm{d}t} = HU = (\sigma_1 + \sigma_2)\,U, \quad \text{with} \quad \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 0 & -\mathrm{i} \\ \mathrm{i} & 0 \end{pmatrix} \tag{7.13}$$



(a) (b)

Figure 7.1. (a) Evolution of error in unitarity for system (7.13) integrated with third- and fourth-order splitting methods with complex coefficients. (b) Efficiency diagram computed for a final time $t_f = 10$. Different patterns in the coefficients lead to different qualitative behaviour.

and $U(0) = I$. Clearly, $U(t)$ is a 2×2 unitary matrix and $\det U(t) = 1$. We test both the preservation of unitarity in the numerical approximations and the computational efficiency of the previous third- and fourth-order schemes. To this end, we take $\varphi_h^{[j]} = \mathrm{e}^{\tau\sigma_j}$, $j = 1, 2$, with $\tau = -\mathrm{i}h$, and compare with the exact solution $\mathrm{e}^{\tau(\sigma_1+\sigma_2)}$. Notice that for problems formulated in complex variables, the use of complex coefficients does not generally increase the overall cost of the algorithm.

We fix $t_f = 1000$ as the final time, and adjust h (and therefore the number of steps n) so that all the methods require the same computational effort. Specifically, $n = 6000$ ($h = 1/6$) for scheme (7.10) and $n = 4000$ ($h = 1/4$) for the fourth-order methods (7.11) and (7.12). Finally we compute $|\|U_{\mathrm{app}}(jh)\| - 1|$, $j = 1, 2, \ldots, n$, where $U_{\mathrm{app}}(jh)$ denotes the approximate solution after j steps, which is shown in Figure 7.1(a). The dashed (red) line corresponds to scheme (7.11), the solid (black) line is produced by (7.12), and finally the dot-dashed line is the result obtained with the third-order scheme (7.10). Notice the different behaviour of different classes of integrators with complex coefficients: whereas error in the preservation of unitarity does not grow with time for methods (7.12) and (7.10), it certainly does for the palindromic scheme (7.11).

Figure 7.1(b) shows an efficiency diagram of these integrators, together with the classical triple jump with real coefficients (dotted line). It is produced by integrating (7.13) with different values of h and computing the error of the approximation (in the 2-norm) at the final time $t_f = 10$ as a function of the computational cost (estimated as the number of exponentials involved in the whole integration). We observe the surprisingly good performance of scheme (7.10) and the superior efficiency of the fourth-order methods with complex coefficients in comparison to their real counterpart. This is related to the difference in size of the main error term in

the expansion of the modified operator associated with each fourth-order method, which is typically associated with the coefficient multiplying $\mathcal{E}_5 = |\sum_{j=1}^{3} \gamma_j^5|$. Thus, $\mathcal{E}_5$ is about 200 times smaller for the schemes (7.8) and (7.9) than for the triple jump with real coefficients (Blanes, Casas and Escorihuela-Tomàs 2022*a*).

7.2. *Symmetric-conjugate methods and unitary problems*

The previous example shows that the time-symmetric character of a method with complex coefficients does not necessarily guarantee good preservation properties, whereas the situation is different for schemes (7.12) and (7.10). They belong to the general class of methods

$$\psi_h = \varphi^{[1]}_{a_{s+1}h} \circ \varphi^{[2]}_{b_s h} \circ \varphi^{[1]}_{a_s h} \circ \cdots \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_1 h}, \tag{7.14}$$

whose coefficients verify

$$a_{s+2-j} = \overline{a}_j, \quad b_{s+1-j} = \overline{b}_j, \quad j = 1, 2, \ldots \tag{7.15}$$

and can be properly called *symmetric-conjugate*. They possess the following distinctive feature: assume our differential equation $x' = f(x)$ is reversible with respect to complex conjugation, in the sense that $\overline{f(x)} = -f(\overline{x})$ for all $x \in \mathbb{C}$, and the same holds for each piece f_j in $f = f_1 + f_2$. Then scheme (7.14)–(7.15) verifies

$$\overline{\psi}_h = \psi_h^{-1}, \tag{7.16}$$

so that the map ψ_h is also reversible with respect to complex conjugation (Hairer *et al.* 2006, sect. V.1). Notice that (7.16) is not satisfied by palindromic compositions involving complex coefficients.

Property (7.16) has some major consequences. Suppose we are dealing with the linear problem

$$\mathrm{i}\frac{\mathrm{d}u}{\mathrm{d}t} = Hu, \quad u(0) = u_0, \quad u \in \mathbb{C}^N, \tag{7.17}$$

where H is an $N \times N$ real matrix of the form $H = H_1 + H_2$, with H_1, H_2 also real, so that the symmetric-conjugate method (7.14)–(7.15) reads

$$\psi_\tau = \mathrm{e}^{\overline{a}_1 \tau H_1}\, \mathrm{e}^{\overline{b}_1 \tau H_2}\, \mathrm{e}^{\overline{a}_2 \tau H_1} \cdots \mathrm{e}^{a_2 \tau H_1}\, \mathrm{e}^{b_1 \tau H_2}\, \mathrm{e}^{a_1 \tau H_1} \tag{7.18}$$

with $\tau = -\mathrm{i}h$. Bernier, Blanes, Casas and Escorihuela-Tomàs (2023) have proved that if all the eigenvalues of H are real and simple, then, for sufficiently small h, there exist real matrices D_τ (diagonal) and P_τ (invertible) such that $\psi_\tau^n = P_\tau\, \mathrm{e}^{n\tau D_\tau} P_\tau^{-1}$, all the eigenvalues of the linear map ψ_τ have modulus 1 and the norm $|u|^2$ and the energy $\overline{u}^\top H u$ are almost preserved for long times. In other words, any symmetric-conjugate splitting method ψ_τ applied to (7.17) is similar to a unitary method for sufficiently small values of the step size h.

We should recall that the time-dependent Schödinger equation, once discretized in space, leads to an equation of the form (7.17), with both H_1 and H_2 real and symmetric matrices: $H_1 = T$ is associated with the second-order periodic spectral

differentiation matrix and $H_2 = V$ is the (diagonal) matrix corresponding to the potential evaluated at the grid points. It is then admissible to use symmetric-conjugate methods in this setting, since they also guarantee preservation of invariants for sufficiently small values of h. We should notice, however, that both positive and negative imaginary parts are present in the scheme if $a_j \in \mathbb{C}$ (since the method is consistent), and this may lead to severe instabilities due to the unboundedness of the Laplace operator (Castella *et al.* 2009, Hansen and Ostermann 2009*b*). It then seems appropriate to consider in this setting only symmetric-conjugate splitting methods with $0 < a_j < 1$ and $b_j \in \mathbb{C}_+$. Different schemes of orders three to six within this family have been proposed and tested in Bernier *et al.* (2023), also showing promising results in terms of efficiency.

If, on the other hand, the equation to integrate is of the form

$$\frac{\mathrm{d}u}{\mathrm{d}t} = Hu = (H_1 + H_2)u, \quad u(0) = u_0, \quad u \in \mathbb{R}^N \tag{7.19}$$

with H_1, H_2 real symmetric matrices, then a symmetric-conjugate splitting method ψ_h of order r satisfies

$$(\overline{\psi}_h)^\top = \psi_h,$$

for all values of h. In consequence, there exists a family of unitary matrices U_h and a family of real diagonal matrices D_h depending smoothly on h such that

$$(\psi_h)^n = U_h \, \mathrm{e}^{nhD_h} \, (\overline{U}_h)^\top,$$

where $\mathrm{Im}(U_h) = O(h^r)$ and D_h is a perturbation of order r of the matrix D_0 diagonalizing H. If $u_n = (\psi_h)^n u_0$, then the quantity $\|\mathrm{Im}(u_n)\|/\|u_n\|$ remains bounded along the numerical trajectory, since the error in the imaginary part is only due to the transformation U_h. Again, this is not true for palindromic schemes with complex coefficients.

We can then conclude that symmetric-conjugate splitting methods can be safely used to integrate equation (7.19) when H_1 and H_2 are real symmetric, as for palindromic (or time-symmetric) schemes with real coefficients, with one important difference: whereas in splitting methods with real coefficients at least one a_j and at least one b_j are negative when the order $r \geq 3$, symmetric-conjugate methods of order $r \geq 3$ do exist with coefficients having positive real part.

Several parabolic PDEs lead to equation (7.19) after space discretization. This is the case, in particular, for the time-dependent Schrödinger equation in imaginary time (6.45).

7.3. Projecting on the real axis

In the general case of an equation $x' = f(x) = f_1(x) + f_2(x)$ with f_1, f_2 real, the usual practice consists in projecting the numerical evolution after each time step to its real part, since we are interested in getting real approximations to the exact

solution. Thus, if ψ_h is a splitting method of order r, by following this approach, we are actually applying the scheme

$$R_h = \frac{1}{2}(\psi_h + \overline{\psi}_h), \tag{7.20}$$

which is also at least of order r, in agreement with the comments at the end of Section 6 (Hansen and Ostermann 2009*b*). In fact, the order of R_h is actually $r+1$ if ψ_h is symmetric-conjugate and r is odd. This was already noticed by Chambers (2003) for the third-order method (7.10) and proved in general by Blanes, Casas, Chartier and Escorihuela-Tomàs (2022*c*).

This feature of symmetric-conjugate methods has some effects, especially when constructing high-order schemes to be used by projecting on the real axis. Although apparently a symmetric-conjugate composition requires us to solve the same number of order conditions to achieve order r as a general composition (and more than a palindromic one for orders higher than four), additional reductions take place for the projected method (7.20) (Blanes *et al.* 2022*c*), and in fact the resulting R_h requires fewer stages. Thus, in particular, it is possible to construct a composition of the form

$$\psi_h = S^{[2]}_{\overline{\alpha}_1 h} \circ S^{[2]}_{\overline{\alpha}_2 h} \circ \cdots \circ S^{[2]}_{\alpha_2 h} \circ S^{[2]}_{\alpha_1 h}, \tag{7.21}$$

with $S^{[2]}_h$ a time-symmetric second-order method, with only five stages so that its projected part R_h is of order six, whereas schemes based on palindromic compositions require at least seven stages. The reduction is more notable for higher orders: thus, a composition (7.21) of order five involving nine appropriately chosen stages leads to a projected method R_h of order eight. By contrast, 15 stages are required by palindromic compositions.

One of the salient features of splitting and composition methods is that they preserve by construction qualitative features of the exact solution, as seen in Section 4. These favourable properties are of course lost when the special linear combination (7.20) is considered, but nevertheless the resulting scheme R_h still preserves them up to an order much higher than the order of the method itself. More specifically, suppose $S^{[2]}_h$ is a time-symmetric second-order and symplectic method applied to a Hamiltonian system. Then we have the following result (Blanes *et al.* 2022*c*).

- If the symmetric-conjugate composition ψ_h given in (7.21) is of odd order $r = 2k-1$, $k \geq 2$, then the method R_h (of order $2k$) preserves time-symmetry up to order $4k-1$, that is,

  $$R_h \circ R_{-h} = \mathrm{id} + O(h^{4k}),$$

  and symplecticity up to the same order,

  $$(R'_h)^\top J R'_h = J + O(h^{4k}),$$

  where J denotes the canonical symplectic matrix and R'_h is the Jacobian.

- If ψ_h is symmetric-conjugate and of even order $r = 2k$, then R_h (which is also of order $2k$) preserves time-symmetry and symplecticity up to order $4k + 3$.
- If ψ_h is palindromic of order $r = 2k$, then the resulting R_h (also of order $2k$) preserves time-symmetry and symplecticity up to order $4k + 1$.

8. A collection of splitting methods

Once the order conditions for a splitting or composition method of a given order r have been explicitly obtained, the next step in the construction of particular schemes consists of course in solving these polynomial equations to determine the coefficients. To begin, we consider compositions with as many parameters as equations and try to use a computer algebra system to determine all real solutions. Nevertheless, solving the order conditions in this way is only possible for moderate values of r, and thus we have to turn to numerical techniques. Since the number of real solutions usually increases with the order, the problem is how to select the particular solution expected to give the best performance when the integrator is applied to practical problems. This is typically done by minimizing some objective function, depending on the particular class of schemes considered. Thus, in the case of splitting methods of the form (2.17), we try to minimize the leading error term in the asymptotic expansion of $\log(\Psi(h))$ in (2.21), namely $\sum_{j=1}^{c_{r+1}} \beta_j F_{r+1,j}$, where $F_{r+1,j}$ denote the elements of the Lyndon basis associated with Lyndon words with $r + 1$ letters. The objective function is then

$$E_{r+1} = \left(\sum_{j=1}^{c_{r+1}} |\beta_j|^2 \right)^{1/2}.$$

However, we have to take into account that E_{r+1} will change if another basis of Lie brackets is considered. We are also assuming that all these brackets contribute in a similar way, something that is not guaranteed to take place in general. It makes sense, then, to introduce other quantities as possible estimators of the error committed. In particular, it has been noticed that large coefficients a_j, b_j in the splitting method usually lead to large truncation errors, since higher-order terms in $\Psi(h)$ depend on increasingly higher powers of these coefficients. For this reason, it is also convenient to keep track of the quantities

$$\Delta \equiv \sum_j (|a_j| + |b_j|) \quad \text{and} \quad \delta \equiv \max_j (|a_j|, |b_j|) \tag{8.1}$$

and eventually discard solutions with large values of Δ or δ. In the case of compositions (2.15) of a basic first-order method and its adjoint, and compositions of Strang maps, (2.1), a frequently used criterion is to choose the solution that minimizes $\sum_{j=1}^{2s} |\alpha_j|$ and $\sum_{j=1}^{s} |\gamma_j|$, respectively. Other possibilities include minimizing only those terms in the truncation error that are not removable by a processor, in accordance with the analysis carried out in Sections 4 and 5.

Including additional maps (stages) in the composition provides additional parameters that may lead to smaller values of the chosen objective function. It is important at this point to remark that the efficiency of a method is measured by taking into account the computational cost required to achieve a given accuracy. Thus, if we have several methods of order r with different computational cost (usually measured as the number of stages or evaluations of the functions involved), the most efficient method does not necessarily correspond to the cheapest method: the extra cost of some methods can be compensated by an improvement in the accuracy obtained. In fact, this is what usually happens in practice, although solving the polynomial equations with additional stages (and free parameters) is by no means a trivial task. Continuation techniques have been shown to be very useful in this context; see e.g. Blanes *et al.* (2013*b*) and Alberdi, Antoñana, Makazaga and Murua (2019).

The use of the processing technique also allows us to construct methods of a given order typically requiring a reduced computational effort, so that the overall efficiency of the resulting schemes is enlarged if the output is not frequently computed.

With all these considerations in mind, our purpose in this section is to present a comprehensive overview of (most of) the existing methods with real coefficients by classifying them into different families and giving the appropriate references. The corresponding coefficients can be found at www.gicas.uji.es/SplittingMethods.html.

8.1. Symmetric compositions of time-symmetric second-order schemes

Perhaps the first method in this family corresponds to the fourth-order method obtained from the triple jump composition (2.4) when $k = 2$. It was known and internally used in the accelerator physics community during the 1980s (Forest and Ruth 1990) but was independently discovered afterwards in several settings: in Candy and Rozmus (1991), in Campostrini and Rossi (1990) as an algorithm for hybrid Monte Carlo simulations, and also in Creutz and Gocksch (1989) as the more general composition

$$S_h^{[2k+2]} = \left(S_{\gamma_1 h}^{[2k]}\right)^p \circ S_{(1-2p\gamma_1)h}^{[2k]} \circ \left(S_{\gamma_1 h}^{[2k]}\right)^p, \quad \text{with} \quad \gamma_1 = \frac{1}{2p - (2p)^{1/(2k+1)}}. \tag{8.2}$$

This recursion was also obtained in Suzuki (1990, 1991), whereas the case $p = 1$ became immensely popular following Yoshida (1990). It was in a certain sense generalized by McLachlan (2002), who provided a rule of thumb to select the optimal value of p for different orders, also valid for processed methods.

As stated in Section 2.1, other choices for the coefficients γ_j in the more general composition

$$\psi_h = S_{\gamma_s h}^{[2]} \circ S_{\gamma_{s-1} h}^{[2]} \circ \cdots \circ S_{\gamma_1 h}^{[2]} \tag{8.3}$$

lead to more efficient schemes when $r \geq 6$. It turns out that virtually all published methods of this form correspond to time-symmetric compositions, i.e. $\gamma_{s+1-j} = \gamma_j$,

Table 8.1. Symmetric compositions of second-order symmetric methods of order $r = 4, 6, 8, 10$ published in the literature. We indicate the number of stages **s** and the pertinent reference. Processed methods are preceded by **P**. The recommended methods, in accordance with the experiments in the Appendix, are $\boxed{\text{framed}}$.

$\psi_h = S^{[2]}_{\gamma_s h} \circ S^{[2]}_{\gamma_{s-1} h} \circ \cdots \circ S^{[2]}_{\gamma_1 h}$			
$r = 4$	$r = 6$	$r = 8$	$r = 10$
3-CR90 $\boxed{\mathbf{5}}$-SUZ90,MCL95b	**7**-YOS90 **9**-MCL95b,KR97 **11**-$\boxed{\mathbf{13}}$-SS05	**15**-YOS90,SUZ93, MCL95b,KR97 **17**-MCL95b,KR97 **19**-$\boxed{\mathbf{21}}$-SS05 **24**-CSS93	**31**-SUZ93,KR97,SS05 **33**-KR97,TSI99, HLW02,SS05 $\boxed{\mathbf{35}}$-SS05
P:3-17-MCL02	**P:**$\boxed{\mathbf{11}}$**-13**-BCM06	**P:**$\boxed{\mathbf{13}}$**-19**-BCM06	**P:19-**$\boxed{\mathbf{23}}$-BCM06

and are therefore of even order. There are at least two reasons for that: (i) the task of constructing new schemes is simplified, since the number of order conditions to be solved is much reduced, and (ii) the resulting methods also require a smaller number of stages s, since the number of order conditions at even orders (automatically satisfied by a time-symmetric composition) is greater than the total number of order conditions divided by 2, for sufficiently large r. For instance, constructing a non-symmetric eighth-order scheme requires at least $s = 16$ (and therefore to solve a system of 16 polynomial equations), whereas $s = 15$ is the minimum number for a symmetric composition, so that only eight polynomial equations have to be solved.

In Table 8.1 we present the most relevant schemes of this type found in the literature. At each order, r, we label each method by the number of stages **s** and an acronym indicating the author(s) and year when it was first published. We also include processed methods, referred to as **P**:**s**, where **s** is the number of stages of the kernel. The following list provides additional information about the collected schemes.

CR90. Collective name given to the triple jump composition, (8.2) with $k = p = 1$.

SUZ90. Recursion (8.2) with $k = 1$, $p = 2$, proposed in Suzuki (1990).

YOS90. Yoshida (1990) gives three sixth-order compositions with $s = 7$ and five eighth-order methods with $s = 15$.

CSS93. Calvo and Sanz-Serna (1993*b*) obtain a seventh-order non-symmetric scheme which can be written as composition (8.3) with $s = 12$. After symmetrization, it provides a 24-stage eighth-order method.

SUZ93. Suzuki and Umeno (1993) provide one sixth-order method with $s = 14$, six eighth-order methods with $s = 15$ and four tenth-order schemes with $s = 31$.

MCL95b. McLachlan (1995*b*) gives several optimized schemes: one fourth-order method with $s = 5$, one sixth-order method with $s = 9$, and two eighth-order methods with $s = 15$ and $s = 17$.

KR97. Kahan and Li (1997) also present several optimized schemes: two sixth-order methods with $s = 9$, three eighth-order methods (one with $s = 15$ and two with $s = 17$), and five tenth-order methods (two with $s = 31$ and three with $s = 33$).

TSI99. Tsitouras (1999) constructs one optimized tenth-order method with $s = 33$.

MCL02. Construction of kernel with recurrence (8.2) for several values of p in McLachlan (2002).

HLW02. Hairer, Lubich and Wanner (2006, first edition published in 2002) provide one optimized tenth-order method with $s = 33$, the most efficient at the time.

SS05. Sofroniou and Spaletta (2005) carry out an exhaustive search of sixth-order schemes with $s = 11, 13$, eighth-order methods with $s = 19, 21$, and tenth-order methods with $s = 31, 33, 35$.

BCM06. Blanes, Casas and Murua (2006*a*) give several processed methods: two sixth-order kernels with $s = 11$ and $s = 13$, two eighth-order kernels with $s = 13$ and $s = 19$, and two tenth-order kernels with $s = 19$ and $s = 23$.

8.2. *Splitting into two parts / composition of a basic first-order method and its adjoint*

Although the order conditions for the two types of scheme,

$$\psi_h = \varphi^{[1]}_{a_{s+1}h} \circ \varphi^{[2]}_{b_s h} \circ \varphi^{[1]}_{a_s h} \circ \cdots \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_1 h} \tag{8.4}$$

and

$$\psi_h = \chi_{\alpha_{2s}h} \circ \chi^*_{\alpha_{2s-1}h} \circ \cdots \circ \chi_{\alpha_2 h} \circ \chi^*_{\alpha_1 h}, \tag{8.5}$$

are equivalent by virtue of the relationship (2.18), the optimization procedures to get the most efficient schemes may differ. In consequence, a particular method optimized for systems that are separable into two parts is not necessarily the best scheme when written as (8.5), although in practice their performances are closely related.

Since schemes of order $r \geq 6$ require more stages than taking the composition (8.3) with $S^{[2]}_h = \chi_{h/2} \circ \chi^*_{h/2}$ or $S^{[2]}_h = \varphi^{[1]}_{h/2} \circ \varphi^{[2]}_h \circ \varphi^{[1]}_{h/2}$, only methods with $r \leq 4$ seem promising. Note, however, that a sixth-order symmetric composition (8.3) with $s = 7$ has only three real solutions for the γ_j, whereas symmetric versions of (8.4) and (8.5) require at least $s = 9$ stages to solve the nine order conditions. This

Table 8.2. Symmetric composition schemes of the form (8.4) (appropriate when the ODE is split in two parts) and (8.5) (composition of a first-order method and its adjoint) of order $r = 2, 3, 4, 6$ with real coefficients. The notation is the same as in Table 8.1.

$$\psi_h = \chi_{\alpha_{2s}h} \circ \chi^*_{\alpha_{2s-1}h} \circ \cdots \circ \chi_{\alpha_2 h} \circ \chi^*_{\alpha_1 h},$$
$$\psi_h = \varphi^{[1]}_{a_{s+1}h} \circ \varphi^{[2]}_{b_s h} \circ \varphi^{[1]}_{a_s h} \circ \cdots \circ \varphi^{[1]}_{a_2 h} \circ \varphi^{[2]}_{b_1 h} \circ \varphi^{[1]}_{a_1 h}$$

$r = 2$	$r = 3$	$r = 4$	$r = 6$
$\boxed{\mathbf{2}}$-MCL95b,OMF03,BCS14 **3**-BCS14	**3**-RUT83,SUZ92	**4-5**-MCL95b $\boxed{\mathbf{6}}$-BM02	**9**-FOR92 $\boxed{\mathbf{10}}$-BM02
		P:3, $\boxed{\mathbf{4}}$-BCR99 **P:** $\boxed{\mathbf{6}}$-BCM06	**P:5**-BCR99 **P:** $\boxed{\mathbf{9}}$-BCM06

enlarged number of equations might then provide some solution leading to smaller error terms.

The schemes presented in Table 8.2 have been specifically designed to deal with these problems, and cannot be obtained as particular cases of the composition (8.3). In more detail, they correspond to the following.

RUT83. Ruth (1983) gives the first three-stage third-order non-symmetric method for systems separable into two parts.

SUZ92. Suzuki (1992) presents a family of three-stage third-order methods for the composition (8.5) with $s = 3$ and $\alpha_6 = 0$.

FOR92. Forest (1992) gives a sixth-order method for separable systems with $s = 9$.

MCL95b. McLachlan (1995*b*) gives an optimized two-stage second-order method and two optimized fourth-order schemes of the form (8.4) with $s = 4$ and $s = 5$.

BCR99. Blanes, Casas and Ros (1999*b*) present several processed methods for separable systems: fourth-order methods with $s = 3$ and $s = 4$, one fifth-order method with a non-symmetric kernel and $s = 4$, and finally one sixth-order scheme with $s = 5$.

BM02. Blanes and Moan (2002) obtain one fourth-order method with $s = 6$ and one sixth-order method with $s = 10$, both optimized for separable systems.

OMF03. Omelyan, Mryglod and Folk (2003) rediscover the optimized two-stage second-order scheme of McLachlan (1995*b*) as an efficient method with positive coefficients.

BCM06. Blanes, Casas and Murua (2006*a*) provide one fourth-order processed method with $s = 6$ and two sixth-order processed schemes with $s = 9$ and $s = 10$.

BCS14. Blanes, Casas and Sanz-Serna (2014) propose efficient symmetric second-order methods with $s = 2$ and $s = 3$, designed for hybrid Monte Carlo simulations.

8.3. Runge–Kutta–Nyström methods

Given the relevance of the problem $y'' = g(y)$, and the simplifications arising when splitting methods are applied to this system (see Section 3.1), many RKN splitting schemes have been obtained in the literature. Since F_1 and F_2 play different roles here, it is convenient to classify them according to the sequence of coefficients as

$$\begin{aligned} \text{AB}: &\quad \varphi^{[A]}_{a_s h} \circ \varphi^{[B]}_{b_s h} \circ \cdots \circ \varphi^{[A]}_{a_1 h} \circ \varphi^{[B]}_{b_1 h}, \\ \text{BA}: &\quad \varphi^{[B]}_{b_s h} \circ \varphi^{[A]}_{a_s h} \circ \cdots \circ \varphi^{[B]}_{b_1 h} \circ \varphi^{[A]}_{a_1 h}, \end{aligned} \tag{8.6}$$

although in fact the two types are conjugate to each other, thus providing similar efficiency. A word of caution is necessary: *in all the RKN splitting methods published in the literature the operator F_1 as defined in Section 3.1 is in fact associated with the map $\varphi_t^{[B]}$, and F_2 is associated with $\varphi_t^{[A]}$*, so that $[F_1, [F_1, [F_1, F_2]]] = 0$. Thus, in particular, in classical and quantum mechanics, A corresponds to the kinetic energy and B to the potential energy. This should be taken into account when implementing the schemes collected here.

It is also convenient to take profit of the FSAL (first same as last) property, and thus we may consider the non-equivalent compositions

$$\begin{aligned} \text{ABA}: &\quad \varphi^{[A]}_{a_{s+1} h} \circ \varphi^{[B]}_{b_s h} \circ \varphi^{[A]}_{a_s h} \circ \cdots \circ \varphi^{[B]}_{b_1 h} \circ \varphi^{[A]}_{a_1 h}, \\ \text{BAB}: &\quad \varphi^{[B]}_{b_{s+1} h} \circ \varphi^{[A]}_{a_s h} \circ \varphi^{[B]}_{b_s h} \circ \cdots \circ \varphi^{[A]}_{a_1 h} \circ \varphi^{[B]}_{b_1 h}, \end{aligned} \tag{8.7}$$

as well as their time-symmetric versions ($a_{s+2-i} = a_i, b_{s+1-i} = b_i$ for ABA and $b_{s+2-i} = b_i, a_{s+1-i} = a_i$ for BAB, respectively).

In Table 8.3 we present the most representative methods within this class. We add **S** or **N** to distinguish symmetric from non-symmetric schemes and the subscript AB, ABA and BAB to denote the different alternatives (8.6), (8.7). Processed methods have also been included. Explicitly, they correspond to the following.

FOR92. Forest (1992) gives a sixth-order method with $s = 7$.

MA92. McLachlan and Atela (1992) present non-symmetric fourth- and fifth-order methods with $s = 4$ and $s = 6$, respectively.

CSS93. Calvo and Sanz-Serna (1993*a*) obtain one fourth-order non-symmetric BAB method with $s = 4$.

OS94. Okunbor and Skeel (1994) obtain four fifth-order methods with $s = 5$ and sixteen sixth-order symmetric methods with $s = 7$.

Table 8.3. RKN splitting integrators of order $r = 4, 5, 6, 8$. Since the role of the flows $\varphi_h^{[A]}$ and $\varphi_h^{[B]}$ is not interchangeable here, we distinguish symmetric **S** and non-symmetric **N** compositions with a subscript AB, ABA, BAB. As usual, processed methods are preceded by **P** and the recommended schemes are $\boxed{\text{framed}}$.

RKN splitting methods			
$r = 4$	$r = 5$	$r = 6$	$r = 8$
4N$_{\text{AB}}$-MA92 **4N**$_{\text{BAB}}$-CSS93 **4,5S**$_{\text{ABA}}$-MCL95b $\boxed{\textbf{6}}$**S**$_{\text{ABA,BAB}}$-BM02 **4,5S**$_{\text{ABA,BAB}}$-OMF03	**5N**$_{\text{ABA}}$-OS94 $\boxed{\textbf{6}}$**N**$_{\text{AB}}$-MA92 **6N**$_{\text{AB}}$-CHO00	**7S**$_{\text{ABA}}$-FOR92,OS94 **7S**$_{\text{BAB}}$-FOR92 **11S**$_{\text{BAB}}$-BM02 $\boxed{\textbf{14}}$**S**$_{\text{ABA}}$-BM02	**17S**$_{\text{ABA}}$-OL94 **17,18,**$\boxed{\textbf{19}}$**S**$_{\text{ABA,BAB}}$-BCE22
P:$\boxed{\textbf{2}}$**N**$_{\text{AB}}$-BCR99		**P:4-6S**$_{\text{ABA,BAB}}$-BCR01a **P:**$\boxed{\textbf{7}}$**S**$_{\text{BAB}}$-BCR01b	**P:9S**$_{\text{ABA}}$-BCR01a **P:**$\boxed{\textbf{11}}$**S**$_{\text{BAB}}$-BCR01b

OL94. Okunbor and Lu (1994) obtain one eighth-order ABA method with $s = 17$.

MCL95b. McLachlan (1995*b*) gives optimized RKN methods of order four with $s = 4$ and $s = 5$, and of order six with $s = 7$.

BCR99. Blanes, Casas and Ros (1999*b*) present a processed fourth-order method with a non-symmetric kernel and $s = 2$.

CHO00. Chou and Sharp (2000) get a non-symmetric scheme of order five with $s = 6$.

BCR01a. Blanes, Casas and Ros (2001*a*) present one processed sixth-order method with $s = 6$

BCR01b. Blanes, Casas and Ros (2001*b*) give one sixth- and one eighth-order processed method with $s = 7$ and $s = 11$, respectively.

BM02. Blanes and Moan (2002) obtain one BAB fourth-order method with $s = 6$ and two sixth-order methods: one BAB with $s = 11$ and one ABA with $s = 14$.

OMF03. Omelyan, Mryglod and Folk (2003) propose fourth-order methods of type ABA and BAB with $s = 4$ and $s = 5$.

BCE22. Blanes, Casas and Escorihuela-Tomàs (2022*b*) obtain optimized eighth-order methods with $s = 17, 18$ and 19 of type ABA and BAB.

As we have seen in Section 3.2, the particular structure of problem $y'' = g(y)$ allows us to include the flows corresponding to nested commutators in the previous compositions (8.6)–(8.7). This may on one hand lead to more efficient schemes, and on the other hand allow us to achieve order four with only positive coefficients.

Table 8.4. RKN splitting methods of order $r = 4, 6, 8$ with modified potentials. Schemes are coded as in Table 8.3.

RKN splitting methods with commutators		
$r = 4$	$r = 6$	$r = 8$
2S$_{\mathrm{ABA,BAB}}$-KOS93,CHI97 **4S**$_{\mathrm{ABA,BAB}}$-SUZ95 **3,$\boxed{4}$,5S**$_{\mathrm{ABA,BAB}}$-CHI06,OMF02	**4,$\boxed{5}$S**$_{\mathrm{ABA,BAB}}$-OMF02	**11S**$_{\mathrm{ABA,BAB}}$-OMF02
P:$\boxed{1}$S$_{\mathrm{BAB}}$-TI84,ROW91, WHT96,BCR99 **P:2S**$_{\mathrm{BAB}}$-LSS97	**P:$\boxed{3}$S**$_{\mathrm{ABA,BAB}}$-BCR99	**P:4S**$_{\mathrm{ABA}}$-BCR01a **P:5S**$_{\mathrm{BAB}}$-BCR01a,BCR01b

In Table 8.4 we present some relevant methods within this class, both processed and non-processed. They correspond to the following.

TI84. Takahashi and Imada (1984) obtain the kernel of a fourth-order processed method with $s = 1$.

ROW91. Rowlands (1991) rediscovers the kernel of the fourth-order processed method with $s = 1$.

KOS93. Koseleff (1993) obtains the first fourth-order BAB method with positive coefficients with $s = 2$ (but with a misprint in one of the coefficients).

SUZ95. Suzuki (1995) gives fourth-order ABA and BAB methods with $s = 3$.

WHT96. Wisdom, Holman and Touma (1996) obtain one fourth-order processed method with $s = 1$.

LSS97. López-Marcos, Sanz-Serna and Skeel (1997) obtain fourth-order processed methods with $s = 2$.

CHI97. Chin (1997) rediscovers the fourth-order method of Koseleff (1993) (but now with the correct coefficients) and shows its efficiency in practice.

BCR99. Blanes, Casas and Ros (1999*b*) obtain the following processed methods: one fourth-order method with $s = 1$, one non-symmetric fifth-order method with $s = 2$ and one sixth-order method with $s = 3$.

BCR01a. Blanes, Casas and Ros (2001*a*) present one sixth-order method with $s = 2$ and higher-order commutators and two sixth-order methods with $s = 3$.

BCR01b. Blanes, Casas and Ros (2001*b*) propose one optimized sixth-order method with $s = 3$ and one eighth-order method with $s = 5$.

OMF03. Omelyan, Mryglod and Folk (2003) present a detailed treatment of fourth-order methods with $2 \leq s \leq 5$ and sixth-order methods with $s = 4, 5$ both ABA and BAB and different number of modified potentials.

CHI06. Chin (2006) obtains families of fourth-order schemes with coefficients written analytically in terms of free parameters for $s > 2$.

8.4. Methods for near-integrable systems

As we have seen in Section 3.3, splitting methods are outstanding for near-integrable systems of the form $x' = f_1(x) + \varepsilon f_2(x)$: the presence of two parameters, h (the step size) and ε (the size of the perturbation), allows us to reduce the number of order conditions and still construct highly efficient schemes. There are problems where, in addition, we can use the same techniques as for RKN methods and also incorporate the flows of nested commutators into the algorithm. This is the case, in particular, for Hamiltonian systems of the form $H = H_1 + \varepsilon H_2$, when H_1 is the sum of two-body Kepler problems and H_2 depends only on coordinates. If we let F_1 denote the operator associated with H_1, and let F_2 denote the operator associated to the perturbation H_2, then $[F_2, [F_2, [F_2, F_1]]] = 0$, and the analysis done in Section 3.1 can be applied here (with the obvious interchange $F_1 \leftrightarrow F_2$).

In Table 8.5 we separate, as usual, non-processed from processed schemes (preceded by **P**). We also include methods when $[F_2, [F_2, [F_2, F_1]]] = 0$ (second row) and schemes with nested commutators (denoted by a label *) in the last two rows of processed methods.

WH91. Wisdom and Holman (1991) present the first $(2, 2)$ method.

MCL95a. McLachlan (1995*a*) gives families of symmetric $(2s, 2)$ schemes of type ABA and BAB with $s \leq 5$ and positive coefficients, one BAB (6,4) method with $s = 4$ and two (8,4) ABA and BAB schemes with $s = 5$.

WHT96. Wisdom, Holman and Touma (1996) present one processed $(k, 2)$ method with $k = 16$ and $s = 1$.

BCR00. Blanes, Casas and Ros (2000) give several processed methods: (i) one non-symmetric (6,4,3) method with $s = 2$, one (7,6,4) method with $s = 3$, one (7,6,5,4) method with $s = 4$, and (ii) several RKN methods, one (6,4) with $s = 2$ and one (7,6,5) with $s = 3$ both non-symmetric. In addition, with modified potentials, they give one (6,4) with $s = 1$ and a non-symmetric (7,6,5) with $s = 2$.

LR01. Laskar and Robutel (2001) carry out a systematic study of $(2s, 2)$ methods, and obtain new schemes up to $s = 10$ with positive coefficients. The order of these methods is increased to $(2s, 4)$ by including an appropriate modified potential.

BCF13. Blanes *et al.* (2013*b*) present several schemes of type ABA: one (10,4) with $s = 7$, one (8,6,4) with $s = 7$ and one (10,6,4) with $s = 8$.

Table 8.5. Splitting methods for near-integrable systems of the form $x' = f_1 + \varepsilon f_2$. For processed methods we also include methods applicable when $[F_2, [F_2, [F_2, F_1]]] = 0$ (the last two rows of processed methods) and schemes with modified flows (labelled *).

Splitting methods of generalized order		
$(n, 2)$	$(n, 4)$	$(n, 6, 4)$
$\boxed{\mathbf{1}}$(2,2)**S**-WH91 $\boxed{\mathbf{n}}$(2n,2)$\mathbf{S}_{\mathrm{ABA,BAB}}$-MCL95a,LR01	**4**(6,4)$\mathbf{S}_{\mathrm{BAB}}$-MCL95a $\boxed{\mathbf{5}}$(8,4)$\mathbf{S}_{\mathrm{ABA,BAB}}$-MCL95a **7**(10,4)$\mathbf{S}_{\mathrm{ABA}}$-BCF13	$\boxed{\mathbf{7}}$(8,6,4)$\mathbf{S}_{\mathrm{ABA}}$-BCF13 $\boxed{\mathbf{8}}$(10,6,4)$\mathbf{S}_{\mathrm{ABA}}$-BCF13
P:1(17,2)-WHT96		**P:**$\boxed{\mathbf{3}}$(7,6,4)$\mathbf{S}_{\mathrm{ABA}}$-BCR00
	P:$\boxed{\mathbf{2}}$(6,4)$\mathbf{S}_{\mathrm{AB}}$-BCR00	**P:**$\boxed{\mathbf{3}}$(7,6,5)$\mathbf{S}_{\mathrm{AB}}$-BCR00
	P:1(6,4)$\mathbf{S}^*_{\mathrm{ABA}}$-BCR00 **P:n**(n,4)$\mathbf{S}^*_{\mathrm{ABA}}$-LR01	**P:2**(7,6,5)$\mathbf{S}^*_{\mathrm{AB}}$-BCR00

9. Splitting everywhere: some relevant applications

9.1. Splitting and extrapolation

Extrapolation methods constitute a class of efficient high-order schemes for solving initial value problems when the local error of the basic integrator has an asymptotic expansion containing only even powers of the step size h. In this subsection we review how some of the previous techniques used in the construction and analysis of splitting and composition methods can also be applied to simplify the analysis of extrapolation.

9.1.1. The Gragg/GBS method

Given the initial value problem $x' = f(x)$, $x(t_0) = x_0$, Gragg proposed the following algorithm to produce the quantity $T_{h^*}(t)$. Denoting $t = t_0 + 2nh^*$, $t_i = t_0 + \mathrm{i}h^*$, it reads

$$\begin{aligned} x_1 &= x_0 + h^* f(x_0), \\ x_{i+1} &= x_{i-1} + 2h^* f(x_i), \quad i = 1, 2, \ldots, 2n, \\ T_{h^*}(t) &= \frac{1}{4}(x_{2n-1} + 2x_{2n} + x_{2n+1}). \end{aligned} \tag{9.1}$$

Subsequently, he proved that $T_{h^*}(t)$ possesses an asymptotic expansion in even powers of h^* (Gragg 1965), so that it can then be used for Richardson extrapolation. The original proof was simplified by Stetter (1970) when he realized that (9.1) can

be interpreted as a one-step algorithm. Later on, Hairer *et al.* (1993) showed that this scheme is consistent with the differential equation in the extended phase space

$$\begin{aligned} u' &= f(v), \qquad u(t_0) = u_0 = y_0, \\ v' &= f(u), \qquad v(t_0) = v_0 = y_0, \end{aligned} \tag{9.2}$$

whose exact solution is $u(t) = v(t) = x(t)$. In fact, to prove that $T_{h^*}(t)$ only contains even powers of h^* is trivial by noticing that system (9.2) can be expressed as

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} 0 \\ f(u) \end{pmatrix} + \begin{pmatrix} f(v) \\ 0 \end{pmatrix} = f_1(u) + f_2(v), \qquad \begin{pmatrix} u(t_0) \\ v(t_0) \end{pmatrix} = \begin{pmatrix} x_0 \\ x_0 \end{pmatrix}, \tag{9.3}$$

and that the Strang splitting with step size $h = 2h^*$ gets

$$\begin{aligned} v_{1/2} &= v_{i-1} + \frac{h}{2} f(u_{i-1}), \\ u_i &= u_{i-1} + h f(v_{1/2}), \qquad i = 1, 2, \ldots, n, \\ v_i &= v_{1/2} + \frac{h}{2} f(u_i). \end{aligned}$$

Then, clearly, $T_{h^*} = \frac{1}{2}(u_n + v_n)$, and the result is obtained by noticing that the Strang splitting is time-symmetric.

On the other hand, application of the Gragg method (9.1) to the second-order equation

$$y'' = g(y), \qquad y(t_0) = y_0, \qquad y'(t_0) = y'_0, \tag{9.4}$$

produces the equivalent formula (Hairer *et al.* 1993)

$$y_{i+1} - 2y_i + y_{i-1} = h^2 g(y_i), \qquad i = 1, 2, 3, \ldots, \tag{9.5}$$

again with $h = 2h^*$. As is well known, by considering the equivalent first-order system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} y \\ y' \end{pmatrix} = \begin{pmatrix} y' \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ g(y) \end{pmatrix} = f_1(y') + f_2(y), \tag{9.6}$$

and then applying the Strang splitting $\varphi^{[2]}_{h/2} \circ \varphi^{[1]}_{h} \circ \varphi^{[2]}_{h/2}$,

$$\begin{aligned} y_{n+1} &= y_n + h y'_n + \frac{h^2}{2} g(y_n), \\ y'_{n+1} &= y'_n + \frac{h}{2}(g(y_n) + g(y_{n+1})), \end{aligned}$$

reproduces Störmer's rule (9.5) after simplification. In fact, one can also apply the other variant of the Strang splitting, i.e. $\varphi^{[1]}_{h/2} \circ \varphi^{[2]}_{h} \circ \varphi^{[1]}_{h/2}$, to (9.6), thus yielding

$$\begin{aligned} y'_{n+1} &= y'_n + h g\left(y_n + \frac{h}{2} y'_n\right), \\ y_{n+1} &= y_n + \frac{h}{2}(y'_n + y'_{n+1}). \end{aligned} \tag{9.7}$$

This (slightly cheaper) scheme could also be used as the basic method for extrapolation, as for the Störmer rule, but since $f_1(y')$ and $f_2(y)$ may have different qualitative properties, the resulting schemes would also be different.

9.1.2. Multi-product expansions

Linear combinations of splitting methods

$$\sum_k c_k \left(\varphi^{[2]}_{b_{k,i_k}h} \circ \varphi^{[1]}_{a_{k,i_k}h} \circ \cdots \circ \varphi^{[2]}_{b_{k,1}h} \circ \varphi^{[1]}_{a_{k,1}h}\right)$$

offer much freedom in the choice of parameters $\{c_k, a_{k,i_j}, b_{k,i_j}\}$ to approximate the flow of $x' = f_1(x) + f_2(x)$, although they no longer provide approximations with the same preservation properties as the original system. As shown in Sheng (1989), if all $\{a_{k,i_j}, b_{k,i_j}\}$ were to be positive, then any individual composition has to be at most of second order, and some coefficients c_k must be negative. The choice of the Strang splitting $S^{[2]}_h$ (or more generally a time-symmetric second-order integrator) as the basic scheme in the composition is particularly relevant due to the structure of the truncation error. Suppose $\{k_1, k_2, \ldots, k_m\}$ denotes a given set of m integer numbers, and form the linear combination

$$\phi_h \equiv \sum_{i=1}^{m} c_i \left(S^{[2]}_{h/k_i}\right)^{k_i}. \tag{9.8}$$

Then ϕ_h furnishes an approximation of order $2m$ if the coefficients satisfy the linear equations

$$G_0 = \sum_{i=1}^{m} c_i = 1, \quad G_{2+2j} = \sum_{i=1}^{m} \frac{c_i}{k_i^{2+2j}} = 0, \quad j = 0, 1, \ldots, m-2,$$

with solutions

$$c_i = \prod_{j=1(\neq i)}^{m} \frac{k_i^2}{k_i^2 - k_j^2}.$$

Schemes of the form (9.8) are referred to as multi-product expansions (Chin and Geiser 2011) and allow us to construct in an easy way efficient high-order integrators. In general, if we start with a basic time-symmetric geometric method of order $2n$, the corresponding linear combination (9.8) provides a scheme of order $2n + 2\ell$, $\ell = 1, \ldots, n$, which, quite remarkably, preserves the geometric properties of the system (e.g. symplecticity for Hamiltonian problems) up to order $4n + 1$ or higher (Blanes, Casas and Ros 1999*a*, Chan and Murua 2000). This can be shown by appropriately rewriting the differential operator associated with (9.8) as a product of exponentials of operators (Blanes and Casas 2016).

9.2. *Crouch–Grossman methods*

The nonlinear differential equation

$$Y' = A(Y)\,Y, \quad Y(0) = Y_0 \in \mathcal{G}, \tag{9.9}$$

where $A \in \mathbb{R}^{d\times d}$ and $\mathcal{G}$ is a matrix Lie group, appears in relevant physical fields such as rigid body mechanics, in the calculation of Lyapunov exponents ($\mathcal{G} \equiv \mathrm{SO}(d)$) and other problems arising in Hamiltonian dynamics ($\mathcal{G} \equiv \mathrm{Sp}(d)$). In fact, it can be shown that every differential equation evolving on a matrix Lie group $\mathcal{G}$ can be written in the form (9.9) (Iserles *et al.* 2000).

A class of methods providing by construction approximations in $\mathcal{G}$ was proposed in Crouch and Grossman (1993) by appropriately modifying Runge–Kutta schemes as follows; see also Hairer *et al.* (2006, sect. IV.8).

Let $b_i, a_{i,j}$, $(i, j = 1, \dots, s)$ be real numbers. Then an explicit s-stage Crouch–Grossman method is given by

$$\begin{aligned} Y^{(1)} &= Y_n, & K_1 &= A(Y^{(1)}), \\ &\vdots \\ Y^{(j)} &= \mathrm{e}^{ha_{j,j-1}K_{j-1}} \cdots \mathrm{e}^{ha_{j,1}K_1}\, Y_n, & K_j &= A(Y^{(j)}), \quad 2 \le j \le s, \\ Y_{n+1} &= \mathrm{e}^{hb_sK_s} \cdots \mathrm{e}^{hb_1K_1} Y_n. \end{aligned} \tag{9.10}$$

As usual, the order conditions to be satisfied by the coefficients $b_i, a_{i,j}$, $(i, j = 1, \dots, s)$ can be found by comparing the Taylor series expansions of the exact and numerical solutions. It has been shown that the order conditions for classical Runge–Kutta methods form a subset of those for the Crouch–Grossman methods (Owren and Marthinsen 1999). Notice that the s-stage scheme (9.10) involves a total of $s(s+1)/2$ matrix exponentials.

In Crouch and Grossman (1993) there are several three-stage methods of order three, whereas Owren and Marthinsen (1999) present a fourth-order scheme with $s = 5$ stages (or equivalently 15 exponentials). Achieving higher order within this approach is by no means simple, and the resulting methods require a large number of exponentials. It turns out, however, that splitting can also be used here to construct schemes of any order with a reduced number of matrix exponentials. To illustrate the technique, we consider the separable system in the extended phase space

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} A(V)\,U \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ A(U)\,V \end{pmatrix}, \tag{9.11}$$

to which we apply the Strang splitting. This leads to

$$\begin{aligned} U_{1/2} &= \mathrm{e}^{\frac{h}{2}A(V_n)}\, U_n, \\ V_{n+1} &= \mathrm{e}^{hA(U_{1/2})}\, V_n, \\ U_{n+1} &= \mathrm{e}^{\frac{h}{2}A(V_{n+1})}\, U_{1/2}. \end{aligned} \tag{9.12}$$

Notice that (9.12) is simply a particular example of (9.10) with $s = 3$, that is,

$$\begin{aligned} Y^{(1)} &= Y_n, & K_1 &= A(Y^{(1)}),\\ Y^{(2)} &= \mathrm{e}^{ha_{2,1}K_1}\, Y_n, & K_2 &= A(Y^{(2)}),\\ Y^{(3)} &= \mathrm{e}^{ha_{3,2}K_2}\, \mathrm{e}^{ha_{3,1}K_1}\, Y_n, & K_3 &= A(Y^{(3)}),\\ Y_{n+1} &= \mathrm{e}^{hb_3K_3}\, \mathrm{e}^{hb_2K_2}\, \mathrm{e}^{hb_1K_1}\, Y_n, \end{aligned} \tag{9.13}$$

with

$$a_{2,1} = \frac{1}{2}, \quad a_{3,2} = 1, \quad a_{3,1} = 0, \quad b_1 = \frac{1}{2}, \quad b_2 = 0, \quad b_3 = \frac{1}{2}.$$

Since $a_{3,1} = b_2 = 0$ and $a_{2,1} = b_1$, then the whole scheme can be computed with only three exponentials, instead of six by applying the recursion (9.10). Higher-order methods can be obtained in the same way by applying appropriate splitting schemes to the enlarged system (9.11). Thus, in particular, it is possible to get a Crouch–Grossman scheme of order four with seven exponentials.

9.2.1. Positivity-preserving splitting methods

Differential equations $x' = f(x)$ modelling population dynamics must satisfy positivity and frequently mass preservation, that is,

(i) $x_i(0) \geq 0 \quad \Rightarrow \quad x_i(t) \geq 0 \;\; \forall\, t,\; i = 1, \ldots, D$ (positivity preservation)

(ii) $\mathbf{1}^\top x(t) = \mathbf{1}^\top x(0), \quad \text{with} \quad \mathbf{1} = (1, \ldots, 1)^\top$ (mass conservation)

Both of them are automatically fulfilled if the differential equation can be written as $x' = A(x)x$, where A is a matrix such that

- if $x_i \geq 0$, $i = 1, \ldots, D$, then $A(x)_{k,\ell} \geq 0$ when $k \neq \ell$, and $A_{k,k} \leq 0$, for $k, \ell = 1, \ldots, D$ (to guarantee positivity);
- $\sum_{k=1}^{D} A_{k,\ell} = 0$ for $\ell = 1, \ldots, D$ (to ensure mass conservation).

In such situations, it is of course expedient to use a numerical integrator that also satisfies these features at the discrete level when carrying out simulations. Unfortunately, there is also an order barrier here for two of the most used families of schemes, namely Runge–Kutta and multistep methods: if they unconditionally preserve positivity, then they cannot be better than first-order (Bolley and Crouzeix 1978).

It turns out, however, that splitting methods are able to overcome this order barrier. To see this point, again consider the problem in the extended phase space

$$\begin{aligned} y' &= A(z)y, \quad y(0) = y_0 = x_0,\\ z' &= A(y)z, \quad z(0) = z_0 = x_0, \end{aligned} \tag{9.14}$$

where $x(t) = y(t) = z(t)$. Now, as in the previous subsection, we apply the Strang

splitting to the system (9.14), that is,

$$\begin{aligned} y_{1/2} &= \mathrm{e}^{\frac{1}{2}hA(z_0)}y_0, \\ z_1 &= \mathrm{e}^{hA(y_{1/2})}z_0, \\ y_1 &= \mathrm{e}^{\frac{1}{2}hA(z_1)}y_{1/2}. \end{aligned} \tag{9.15}$$

Then both y_1 and z_1 are second-order approximations to the exact solution $x(h)$ preserving mass and positivity. This can be easily checked by realizing that $\mathbf{1}^\top A = 0$ and so $\mathbf{1}^\top \mathrm{e}^A x = \mathbf{1}^\top x$. On the other hand, for $h > 0$, if $x_i \geq 0$ then $(\mathrm{e}^{hA(x)})_{jk} \geq 0$ for all i, j, k.

If $h < 0$, positivity is no longer preserved with this approach; see Blanes, Iserles and Macnamara (2022*d*) for a more detailed treatment of this topic.

9.3. *Exact splitting*

There are equations for which a clever decomposition of the vector field in fact provides the exact solution. The resulting factorization is called *exact splitting* and, in the case of partial differential equations, allows us to get very accurate approximations once each subsystem is solved by pseudo-spectral methods and pointwise multiplication. Early applications of this idea to the time integration of the Gross–Pitaevskii equation are contained in Chin and Krotscheck (2005) and Bader and Blanes (2011), and a convenient generalization to a large class of PDEs, called inhomogeneous quadratic differential equations, has been recently proposed in Bernier (2021). These PDEs are of the form

$$u_t(t,x) = -p^w u(t,x), \quad u(0,x) = u_0(x), \quad t \geq 0,\ x \in \mathbb{R}^d, \tag{9.16}$$

where $d \geq 1$, $u_0 \in L^2(\mathbb{R}^d)$ and p^w is the quadratic differential operator

$$p^w = (x,\ -\mathrm{i}\nabla_x)\, Q \begin{pmatrix} x \\ -\mathrm{i}\nabla_x \end{pmatrix} + Y^\top \begin{pmatrix} x \\ -\mathrm{i}\nabla_x \end{pmatrix} + c.$$

Here Q is a constant $2d \times 2d$ symmetric matrix with complex coefficients, $Y \in \mathbb{C}^{2d}$ and $c \in \mathbb{C}$ is constant. If $(-p^w)$ is such that the real part of the polynomial $p(X) = X^\top Q X + Y^\top X + c$ is bounded from below on $\mathbb{R}^{2d}$, then it generates a strongly continuous semigroup, denoted $\mathrm{e}^{-tp^w}u_0$. It was shown in Bernier (2021) that e^{-tp^w} can be factorized as a product of operators that can be exactly evaluated. Notice that some relevant equations are of the form (9.16), so this technique allows us to construct numerical methods that are spectral in space and exact in time using only a small number of fast Fourier transforms. This has been illustrated in Bernier, Crouseilles and Li (2021) in the case of kinetic and nonlinear Schrödinger equations.

As an illustrative example of this technique, let us consider the Schrödinger equation with a quadratic potential

$$\mathrm{i}u_t = \frac{1}{2}(-\partial_{xx} + x^2)\, u \equiv \frac{1}{2}(P^2 + X^2)\, u.$$

Then, for $|h| < \pi$, we have

$$u(h,x) = \mathrm{e}^{-\mathrm{i}h(P^2/2+X^2/2)}\, u(0,x) = \mathrm{e}^{-\mathrm{i}f(h)(X^2/2)}\, \mathrm{e}^{-\mathrm{i}g(h)(P^2/2)}\, \mathrm{e}^{-\mathrm{i}f(h)(X^2/2)}\, u_0(x),$$

with

$$f(h) = \frac{1-\cos(h)}{\sin(h)}, \quad g(h) = \sin(h),$$

so the action of each exponential can be easily evaluated.

9.4. *Gravitational N-body problem*

The long-time numerical integrations of the whole Solar System carried out in Sussman and Wisdom (1992) and Wisdom and Holman (1991), suggesting its chaotic nature, represented a major boost for the development of new and efficient symplectic algorithms. The integrator used in these simulations, referred to in the dynamical astronomy literature as the Wisdom–Holman map, is simply the Strang splitting method applied to the Hamiltonian system

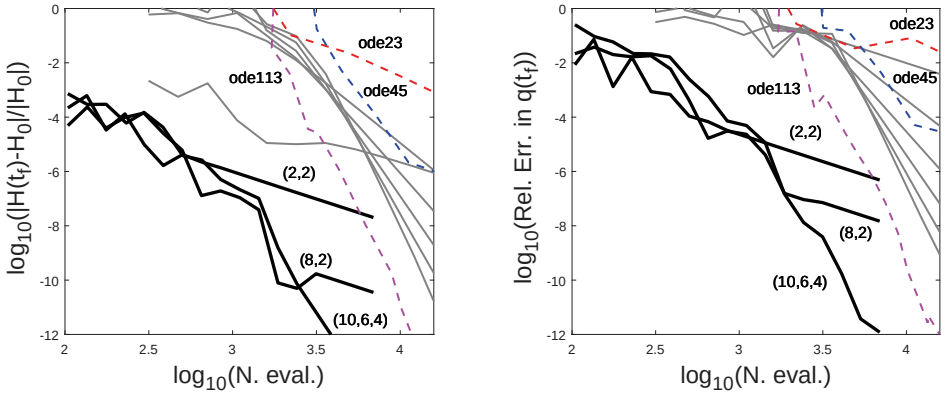
$$\hat{H}(\hat{q},\hat{p}) = H_1(\hat{q},\hat{p}) + H_2(\hat{q}), \tag{9.17}$$

where H_1 and H_2 are given by (1.39) in terms of Jacobi coordinates, as described in Section 1.5. Here H_1 represents the $N-1$ independent Keplerian motions of the planets (which can be efficiently solved) and H_2 accounts for the gravitational interactions among the planets. Moreover, as shown in Section 1.5, $H_2(\hat{q}) = O(\varepsilon)$ with $\varepsilon \equiv (1/m_0)\max_{1\le i\le N-1} m_i$. For the Solar System, $\varepsilon \approx 10^{-3}$. In consequence, splitting methods for near-integrable systems, such as those presented in Sections 3.3 and 8.4, are particularly appropriate. In fact, the Wisdom–Holman map, in the terminology of Section 3.3, corresponds to a method of generalized order $(2,2)$.

Very often, splitting methods are used when very long-time integrations are involved and getting high accuracy is not of particular concern, but rather when having good behaviour of the error over the whole integration period is the crucial point. For instance, in the situation analysed in Sussman and Wisdom (1992), the important issue was to decide whether the motion is regular or chaotic, and not to compute accurately the state of the Solar System after a long time.

The experiment we describe next illustrates that even when short-time integrations are involved and very high accuracy is required, splitting methods can outperform standard integrators once the particular structure of the problem is incorporated into their very formulation.

Specifically, we again deal with the simplified model of the outer Solar System addressed in Section 1.5, with the same initial conditions and an integration interval of 2×10^5 days, and compare splitting methods specially tailored for near-integrable systems with some popular standard integrators. Thus, Figure 9.1 shows efficiency diagrams for the error in energy (a) and in positions (b) at the final time vs. the number of force evaluations, obtained with three different splitting methods



(a) (b)

Figure 9.1. Outer Solar System. Efficiency diagrams showing the relative error in energy (a) and in positions q (b) at the final time $t_f = 2 \times 10^5$ days vs. the number of force evaluations, obtained with different schemes: extrapolation of order $r = 2$, 4, 6, 8, 10, 12 (grey curves), Matlab routines `ode23`, `ode45` and `ode113`, and splitting methods of generalized order (2,2), (8,2) and (10,6,4).

of generalized orders (2,2), (8,2) and (10,6,4), in comparison with extrapolation methods of orders 2-12 (grey lines) and the Matlab routines `ode23`, `ode45` and `ode113`.

As stated before, extrapolation methods are particularly appropriate for second-order differential equations when high accuracy is required (Hairer *et al.* 1993). They take the symmetric second-order scheme (9.5) as the basic method in the linear combination (9.8) with the harmonic sequence $k_i = i$, $i = 1, 2, \ldots, m$, to get a method of order $r = 2m$ at the cost of $m(m+1)/2 + 1$ evaluations of the force,[8] On the other hand, `ode23` and `ode45` solve the ODE with variable time step using embedded Runge–Kutta methods of orders 2–3 and 4–5, respectively, whereas `ode113` does it with variable time step and variable order using multistep methods up to order 13. Due to the smoothness of the problem and the low cost of evaluating the highest order, the latter procedure provides better results than extrapolation methods of high order, as shown in the figure. Nevertheless, none of them can compete with the specially designed splitting methods for this particular problem at all accuracies. This is so even when a very short integration interval is considered. In fact, this relative superiority will only increase for longer integration intervals due to their different error propagation mechanism.

[8] One evaluation can be saved if the basic scheme (9.7) is taken instead. This is the actual implementation we use in our experiments here.

9.5. Molecular simulations

In the simulation of the dynamics of large molecules, two classes of algorithms are commonly used: (i) molecular dynamics, and (ii) those based on stochastic differential equations and the Monte Carlo method.

In classical molecular dynamics, the motion of the atoms in the molecule is determined by integrating Newton's second law,

$$Mq'' = -\nabla_q U(q),$$

where q is a vector containing all positions (in Cartesian coordinates), M is a diagonal matrix whose elements are the atomic masses and $U(q)$ is the (empirical) potential function modelling the inter-particle interactions. Popular models include the Lennard-Jones and Morse potentials, although many other choices are being used for different molecules; see e.g. Leimkuhler and Matthews (2015) for more details. In any case, the number of atoms may be exceedingly large (up to 100 000) and the resulting system is highly nonlinear, exhibiting sensitive dependence on perturbations. In addition, the initial velocities are typically assigned randomly, so there is no point in trying to obtain accurate trajectories. Thus, although in principle Runge–Kutta–Nyström splitting methods can be used in this setting, the method of choice in practice is the Störmer–Verlet scheme, given its excellent stability properties and low computational cost (Leimkuhler *et al.* 1996). Other variants include the use of multiple time-stepping (and in particular the mollified impulse method: García-Archilla, Sanz-Serna and Skeel 1999), and symplectic schemes specially designed for Hamiltonian systems with constraints, such as SHAKE (Ryckaert, Ciccotti and Berendsen 1977) and RATTLE (Andersen 1983).

To take into account the influence of a medium (say, a solvent or air) on a molecular system, typically as external random impacts, a common practice consists in extending the molecular dynamics approach by incorporating a stochastic component in the equations of motion. The resulting system is referred to as Langevin dynamics and is described by the stochastic differential equation

$$\begin{aligned} \mathrm{d}q &= M^{-1} p \,\mathrm{d}t, \\ \mathrm{d}p &= -\nabla_q U(q)\,\mathrm{d}t - \gamma p \,\mathrm{d}t + \sqrt{2\gamma k_B T} M^{1/2}\,\mathrm{d}W \end{aligned} \tag{9.18}$$

in terms of coordinates $q \in \mathbb{R}^d$ and momenta $p \in \mathbb{R}^d$. Here $\gamma > 0$ is the friction coefficient or collision frequency, k_B is the Boltzmann constant, T is the temperature and W is a Wiener process. We may regard (9.18) as modelling a system of particles immersed in a fluid bath consisting of many particles, giving rise to much weaker interactions than those modelled by the potential U (Leimkuhler and Matthews 2015).

As with deterministic equations, splitting methods can be constructed to deal with Langevin dynamics. One possible approach consists in decomposing (9.18)

into three parts as follows:

$$\begin{pmatrix} \mathrm{d}q \\ \mathrm{d}p \end{pmatrix} = \underbrace{\begin{pmatrix} M^{-1}p \\ 0 \end{pmatrix} \mathrm{d}t}_{A} + \underbrace{\begin{pmatrix} 0 \\ -\nabla_q U(q) \end{pmatrix} \mathrm{d}t}_{B} + \underbrace{\begin{pmatrix} 0 \\ -\gamma p\,\mathrm{d}t + \sigma M^{1/2}\,\mathrm{d}W \end{pmatrix}}_{O}, \tag{9.19}$$

with $\sigma \equiv \sqrt{2\gamma k_B T}$. Each of the parts can be exactly solved: the solution of the first is a *drift* in position,

$$\varphi_h^{[A]}(q,p) = (q + hM^{-1}p, p),$$

the second corresponds to a *kick* in momentum,

$$\varphi_h^{[B]}(q,p) = (q, p - h\nabla_q U(q)),$$

and the third piece defines an Ornstein–Uhlenbeck process in p which can be exactly sampled,

$$\varphi_h^{[O]}(q,p) = (q, \mathrm{e}^{-\gamma h}p + \sqrt{k_B T(1-\mathrm{e}^{-2\gamma h})}M^{1/2}R),$$

where R is a vector of independent and identically distributed normal random numbers. In Leimkuhler and Matthews (2013*a*,*b*), several splitting methods constructed along these lines are designed and tested on numerical experiments. They are denoted by the acronym resulting from concatenating the previous symbols A, B and O. Thus, BAOAB corresponds to

$$\varphi_{h/2}^{[B]} \circ \varphi_{h/2}^{[A]} \circ \varphi_h^{[O]} \circ \varphi_{h/2}^{[A]} \circ \varphi_{h/2}^{[B]},$$

and so on. The analysis carried out in Leimkuhler and Matthews (2013*a*,*b*) and Leimkuhler, Matthews and Stolz (2016) for the large friction limit, and in Alamo and Sanz-Serna (2016) for the general case, shows that BAOAB offers an improved behaviour with respect to other members of this family. Other Langevin integrators are given in Leimkuhler and Matthews (2015).

Another approach to molecular simulations, rather than approximating Hamiltonian trajectories, consists in studying the paths originating from the collection of all initial conditions within a given set. This perspective allows us to apply statistical mechanics for calculating averages.

Specifically, if $H(q,p)$ denotes the Hamiltonian function of the system in thermal equilibrium at temperature T, then the probability measure μ in $\mathbb{R}^{2d}$ with density

$$Z^{-1}\mathrm{e}^{-\beta H(q,p)}, \quad \text{where} \quad Z = \int_{\mathbb{R}^{2d}} \mathrm{e}^{-\beta H(q,p)}\,\mathrm{d}q\,\mathrm{d}p < \infty, \tag{9.20}$$

is preserved by the flow of $H(q,p)$. Here $\beta = 1/(k_B T)$. The measure μ is called the Boltzmann–Gibbs distribution and Z is the partition function. Intuitively, μ provides the distribution of (q,p) over an ensemble of many copies of the given system when it is embedded in a much larger system acting as a 'heat bath' at constant temperature. In other words, $Z^{-1}\mathrm{e}^{-\beta H(q,p)}\,\mathrm{d}q\,\mathrm{d}p$ represents the fraction

of copies with momenta between p and $p + \mathrm{d}p$, and configuration between q and $q + \mathrm{d}q$ (Bou-Rabee and Sanz-Serna 2018). Then the average energy is given by

$$\tilde{E} = Z^{-1} \int_{\mathbb{R}^{2d}} H(q,p)\, \mathrm{e}^{-\beta H(q,p)}\, \mathrm{d}q\, \mathrm{d}p.$$

In general, $\tilde{E}$ cannot be obtained analytically, and even the use of numerical quadratures is unfeasible if the dimension d is large. A common practice to approximate this class of integrals then consists in applying Monte Carlo methods, and more specifically, Markov chain Monte Carlo methods (Brooks, Gelman, Jones and Meng 2011).

9.6. *Hamiltonian Monte Carlo*

For Hamiltonian systems of the form

$$H(q,p) = \frac{1}{2} p^\top M^{-1} p + U(q), \tag{9.21}$$

we can factorize the (non-normalized) density (9.20) as

$$\mathrm{e}^{-\beta H(q,p)} = \mathrm{e}^{-\frac{1}{2}\beta p^\top M^{-1} p}\, \mathrm{e}^{-\beta U(q)},$$

so that q and p are stochastically independent (Bou-Rabee and Sanz-Serna 2018): the marginal distribution of the configuration variables q has probability density $\propto \mathrm{e}^{-\beta U(q)}$, whereas the momenta p obey a Gaussian distribution with zero mean and covariance matrix M. Therefore, samples from the p-marginal are easily obtained, and thus one may concentrate on generating samples with probability density

$$Z_q^{-1} \mathrm{e}^{-\beta U(q)}, \quad \text{where} \quad Z_q = \int_{\mathbb{R}^d} \mathrm{e}^{-\beta U(q)}\, \mathrm{d}q. \tag{9.22}$$

This can be carried out with the so-called Hamiltonian (or hybrid) Monte Carlo method (HMC). In fact, HMC was proposed in the landmark paper by Duane, Kennedy, Pendleton and Roweth (1987), not in the context of molecular simulation but in lattice quantum chromodynamics. Later on, it was used in data science and statistics (Liu 2008, Neal 2011). More generally, HMC can be used to sample from any continuous probability distribution on $\mathbb{R}^d$ for which the density function can be evaluated (perhaps up to an unknown normalizing constant): given a target distribution $\Pi(q)$, if $U(q)$ denotes the negative logarithm of the (not necessarily normalized) probability function of the target, then it is clear that $\Pi(q)$ is given by (9.22) (with $\beta = 1$). HMC then generates samples $(q_i, p_i) \in \mathbb{R}^{2d}$ from the Boltzmann–Gibbs distribution corresponding to H, that is,

$$P(q,p) = (2\pi)^{-d/2}\, |\det M|^{-1/2} \mathrm{e}^{-\frac{1}{2} p^\top M^{-1} p}\, Z_q^{-1}\, \mathrm{e}^{-U(q)}, \tag{9.23}$$

by means of a Markov chain so that $P(q,p)$ is an invariant of this chain. The corresponding marginal $q_i \in \mathbb{R}^d$ chain then leaves invariant the target distribution

Table 9.1. HMC algorithm. The function $H = (1/2)p^\top M^{-1}p + U(q)$ is the Hamiltonian. The algorithm generates a Markov chain $q^{(0)} \mapsto q^{(1)} \mapsto \cdots \mapsto q^{(m_{\max})}$ reversible with respect to the target probability distribution $\propto \exp(-U(q))$.

Given $q^{(0)} \in \mathbb{R}^d$, $m_{\max} \geq 1$, set $m = 0$.

1 (Momentum refreshment.) Draw $p^{(m)} \sim \mathcal{N}(0, M)$.

2 (Integration leg.) Compute (q^*, p^*) (q^* is the proposal) by integrating, by means of a reversible, volume-preserving integrator with step size h, the equations of motion derived from the Hamiltonian (9.21) over an interval $0 \leq t \leq Nh$. The initial condition is $(q^{(m)}, p^{(m)})$.

3 (Accept/reject.) Calculate

$$a^{(m)} = \min(1, \exp(H(q^{(m)}, p^{(m)}) - H(q^*, p^*)))$$

and draw $u^{(m)} \sim \mathcal{U}(0, 1)$. If $a^{(m)} > u^{(m)}$, set $q^{(m+1)} = q^*$ (acceptance); otherwise set $q^{(m+1)} = q^{(m)}$ (rejection).

4 Set $m = m + 1$. If $m = m_{\max}$ stop; otherwise go to step 1.

$\Pi(q)$, and this allows us to estimate the multidimensional integral of a certain function F with respect to Π by averaging F at the points (q_i) of the Markov chain.

Since the HMC algorithm has been thoroughly reviewed in the excellent survey by Bou-Rabee and Sanz-Serna (2018), we only summarize here those aspects most closely related to splitting methods. The basic HMC procedure is described in Table 9.1, although more elaborate possibilities exist. Notice that, when generating the proposal, the integrator to be used has to be both reversible and volume-preserving. Given the structure of H, it is clear that palindromic splitting methods constitute the natural choice for this task.

One of the most salient features of the algorithm is that it is able to generate proposal moves that, while being far away from the current state of the Markov chain, has a high probability of acceptance, thus reducing the correlation between samples.

As the main contribution to the computational cost of HMC resides in the numerical integration of the equations of motion, it is of paramount interest to use methods requiring as few evaluations of the force $-\nabla_q U(q)$ as possible, with small energy errors (to avoid rejections) and able to use large step sizes h. Taking these considerations into account, it is hardly surprising that the Störmer–Verlet algorithm is the method of choice, especially in low dimensions. If d increases, however, there are more favourable alternatives. In particular, the three-stage method

$$\varphi^{[T]}_{a_1 h} \circ \varphi^{[V]}_{b_1 h} \circ \varphi^{[T]}_{a_2 h} \circ \varphi^{[V]}_{b_2 h} \circ \varphi^{[T]}_{a_2 h} \circ \varphi^{[V]}_{b_1 h} \circ \varphi^{[T]}_{a_1 h},$$

with

$$a_1 = 0.11888010966548, \quad b_1 = 0.29619504261126, \quad a_2 = \frac{1}{2} - a_1, \quad b_2 = 1 - 2b_1,$$

provides better results at the same computational cost as the standard Störmer–Verlet method (Blanes *et al.* 2014). The rationale behind this and other multi-stage splitting methods presented in Blanes *et al.* (2014) is that the coefficients a_j, b_j are chosen to minimize the energy error in the proposal for relatively large values of h (and not, as in the usual numerical integration domain, to increase the accuracy in the limit $h \to 0$). An extension of this idea is the so-called adaptive integration approach (AIA) (Fernández-Pendás, Akhmatskaya and Sanz-Serna 2016, Akhmatskaya, Fernández-Pendás, Radivojević and Sanz-Serna 2017): here the user chooses the value of h to be used, and then the AIA algorithm itself finds the coefficients of the method within a given family of s-stage splitting integrators providing the best acceptance rate. The efficiency can be further enhanced by using a conveniently modified version of processing so that time-reversible kernels provide time-reversible integrations (Blanes, Calvo, Casas and Sanz-Serna 2021*a*).

In many situations of interest in statistics, the target density is a perturbation of a Gaussian density, so the corresponding Hamiltonian is given by (5.1), that is,

$$H(q, p) = \frac{1}{2} p^\top M^{-1} p + \frac{1}{2} q^\top N q + U(q),$$

where N is a constant symmetric, positive definitive $d \times d$ matrix whose spectral radius typically grows with the dimension d. In that case it is advantageous to precondition the dynamics by choosing the mass matrix (which is free in this setting) as $M = N$, since now all the d frequencies of the preconditioned system are 1, and the only restriction on the step size used by Störmer–Verlet is $h < 2$, independently of h (Bou-Rabee and Sanz-Serna 2018). Furthermore, the considerations on stability exposed in Section 4.2 indicate that the Strang integrators (4.10) based on the exact solution of the quadratic part of H and the perturbation $U(q)$ constitute the best option. In fact, the numerical experiments collected in Casas, Sanz-Serna and Shaw (2022) show that these methods, together with preconditioning, dramatically reduce the computational cost in all test problems and all observables considered, in comparison with the standard Störmer–Verlet scheme. It is worth remarking that all the theory developed in Section 5 is valid here, and in particular, the explicit processor built in Section 5.7 for the Strang splitting is expected to reduce the error in H, and thus it might contribute to increasing the acceptance probability.

9.7. Quantum statistical mechanics

The description of a quantum system in thermal equilibrium at temperature T with Hamiltonian H is based on the thermal density matrix

$$\rho = \mathrm{e}^{-\beta H},$$

where, again, $\beta = 1/(k_B T)$ (Feynman 1972). In fact, most of the properties of the system can be obtained from ρ, but now as an average on the different quantum states. Thus the equilibrium value of an operator $\hat{O}$ corresponding to a physical observable O for a system of N quantum particles in a volume V is given by

$$\langle \hat{O} \rangle = Z^{-1}\,\mathrm{Tr}(\mathrm{e}^{-\beta H}\hat{O}) = Z^{-1}\sum_n \langle n|\mathrm{e}^{-\beta H}\hat{O}|n\rangle, \tag{9.24}$$

where the partition function Z now reads

$$Z = \mathrm{Tr}\,(\mathrm{e}^{-\beta H}) = \sum_n \langle n|\mathrm{e}^{-\beta H}|n\rangle, \tag{9.25}$$

and the states $|n\rangle$ form a complete, orthonormal basis set (Ceperley 1995, Landau and Binder 2005). Since the eigenvalues of the Hamiltonian H are not generally known, we try to evaluate the traces in (9.24) and (9.25) without diagonalizing the Hamiltonian. This can be done with the Feynman path-integral approach. To proceed, we consider the position representation where the particle is labelled. Then the density matrix is given by $\rho(R, R'; \beta) \equiv \langle R|\mathrm{e}^{-\beta H}|R'\rangle$, where $R \equiv \{r_1, \ldots, r_N\}$, r_i is the position of the ith particle and the elements of $\rho(R, R'; \beta)$ are positive and can be interpreted as probabilities. The partition function is then

$$Z = \int \mathrm{d}R\,\langle R|\mathrm{e}^{-\beta H}|R\rangle = \int \mathrm{d}R\,\rho(R, R; \beta). \tag{9.26}$$

Since $\mathrm{e}^{-\beta H} = (\mathrm{e}^{-\varepsilon H})^M$, with $\varepsilon = \beta/M$ for any positive integer M, the density matrix can be expressed as

$$\begin{aligned}\rho(R_0, R_M; \beta) = \int \cdots \int\, & \mathrm{d}R_1\,\mathrm{d}R_2 \cdots \mathrm{d}R_{M-1}\,\rho(R_0, R_1; \varepsilon) \\ & \times\ \rho(R_1, R_2; \varepsilon) \cdots \rho(R_{M-1}, R_M; \varepsilon).\end{aligned} \tag{9.27}$$

The action for a given link k is defined as

$$S_k \equiv S(R_{k-1}, R_k; \varepsilon) = -\ln(\rho(R_{k-1}, R_k; \varepsilon)),$$

so (9.27) becomes

$$\rho(R_0, R_M; \beta) = \int \cdots \int\, \mathrm{d}R_1\,\mathrm{d}R_2 \cdots \mathrm{d}R_{M-1}\,\exp\left(-\sum_{k=1}^{M} S_k\right), \tag{9.28}$$

and the goal is then to construct a sufficiently accurate approximation to ρ while minimizing the number of integrals involved in (9.28) (the number of *beads* M).

On the other hand, the partition function (9.26) can be written as

$$\begin{aligned}Z = \int \cdots \int\, & \mathrm{d}R_0\,\mathrm{d}R_1\,\mathrm{d}R_2 \cdots \mathrm{d}R_{M-1}\,\rho(R_0, R_1; \varepsilon) \\ & \times\ \rho(R_1, R_2; \varepsilon) \cdots \rho(R_{M-1}, R_0; \varepsilon),\end{aligned} \tag{9.29}$$

where the first $|R_0\rangle$ and the last $|R_M\rangle$ elements are identified as required by the trace operations.

In practical applications, one must generally use approximations to ρ. To this end, notice that, typically

$$H = T + V = -\frac{1}{2}\sum_{i=1}^{N} \Delta_i + V. \tag{9.30}$$

It then makes sense to approximate $\mathrm{e}^{-\varepsilon H}$ by symmetric products of $\mathrm{e}^{-\varepsilon T}$ and $\mathrm{e}^{-\varepsilon V}$. The simplest approximation is given, of course, by the Lie–Trotter scheme, known in this setting as the *primitive action*, in which case

$$\rho(R_0, R_2; \varepsilon) \approx \int \mathrm{d}R_1 \langle R_0|\mathrm{e}^{-\varepsilon T}|R_1\rangle \langle R_1|\mathrm{e}^{-\varepsilon V}|R_2\rangle.$$

The operator V is diagonal in the position representation, whereas the kinetic matrix can be evaluated by using the eigenfunction expansion of T (Ceperley 1995). It is then possible to arrive at the discrete path-integral expression for the density matrix

$$\begin{aligned}\rho(R_0, R_M; \beta) = \int \cdots \int \mathrm{d}R_1\, \mathrm{d}R_2 \cdots \mathrm{d}R_{M-1} \left(\frac{1}{2\pi\varepsilon\hbar^2}\right)^{3NM/2} \\ \times \exp\left(-\sum_{j=1}^{M}\left(\frac{1}{2\varepsilon\hbar^2}\|R_{j-1} - R_j\|^2 + \varepsilon V(R_j)\right)\right),\end{aligned} \tag{9.31}$$

providing an approximation of order $O(\varepsilon^2)$, since the Lie–Trotter method is of effective order two. A significant improvement with respect to the primitive approximation is achieved by considering the so-called Takahashi–Imada action (Takahashi and Imada 1984), i.e. the approximation

$$\mathrm{e}^{-\varepsilon(T+V)} \approx \mathrm{e}^{-\frac{\varepsilon}{2}T}\, \mathrm{e}^{-\varepsilon V - \frac{\varepsilon^3}{24}[V,[T,V]]}\, \mathrm{e}^{-\frac{\varepsilon}{2}T}$$

of effective order four in ε.

A typical method to compute the multidimensional integrals appearing in (9.26) and (9.29) is to apply a Monte Carlo sampling according to the probability density $\pi \propto \exp(-\sum_{k=1}^{M} S_k)$, where Z normalizes π. In this respect, notice that for the Hamiltonian (9.30) we have $\langle R_{k-1}|\mathrm{e}^{-a_i\varepsilon T}|R_k\rangle \propto \exp(-\|R_{k-1} - R_k\|^2/(2a_i\varepsilon))$, so the coefficient a_j cannot be negative for a probabilistic based simulation. Here again the order barrier for splitting methods having positive coefficients is important: no splitting method of order higher than two can be used for doing quantum statistical calculations, unless nested commutators enter into their formulation. In this context, the fourth-order scheme (3.4) is widely used, as well as the more general two-parameter family of fourth-order methods

$$\mathrm{e}^{a_1\varepsilon T}\, \mathrm{e}^{\varepsilon W_{b_1,c_1}}\, \mathrm{e}^{a_2\varepsilon T}\, \mathrm{e}^{\varepsilon W_{b_2,c_2}}\, \mathrm{e}^{a_2\varepsilon T}\, \mathrm{e}^{\varepsilon W_{b_1,c_1}}\, \mathrm{e}^{a_1\varepsilon T}, \tag{9.32}$$

with modified potential

$$W_{b_i,c_i} = b_i V + c_i \varepsilon^2 [V, [T, V]]. \tag{9.33}$$

In this case, the presence of two free parameters makes it possible to minimize some of the sixth-order error terms and thus yield more efficient schemes. Numerical simulations carried out in Sakkos, Casulleras and Boronat (2009) show that the required number M in (9.28) to reproduce the exact energy of the system at low temperatures is much smaller with scheme (9.32) than with the Lie–Trotter and Takahashi–Imada methods; see Chin (2023) for a recent review.

9.8. Vlasov–Poisson equations

When a gas is brought to a very high temperature, electrons leave the atoms, thus leading to an overall mixture of electrically charged particles (ions and electrons) usually called *plasma*. There is a hierarchy of models to describe plasmas, ranging from those based on N particles evolving with the laws of classical relativistic mechanics and forces due to external and self-consistent electromagnetic fields, to kinetic equations and fluid models, more appropriate when the plasma is in thermodynamic equilibrium.

In kinetic models each species s in the plasma is characterized by a distribution function $f_s(x, v, t)$, so that $f_s \, \mathrm{d}x \, \mathrm{d}v$ is the average number of particles of species s with position and velocity in a box of volume $\mathrm{d}x \, \mathrm{d}v$ centred at (x, v). In the limit where the collective effects are dominant over collisions between particles, the kinetic equation describing the system (in the non-relativistic regime) is the Vlasov equation (Vlasov 1961)

$$\frac{\partial f_s}{\partial t} + v \cdot \nabla_x f_s + \frac{q_s}{m_s}(E + v \times B) \cdot \nabla_v f_s = 0, \tag{9.34}$$

where q_s and m_s denote the charge and mass of the particles of species s, and E and B stand for the electric and magnetic field, respectively. Equation (9.34) just expresses the fact that the distribution function f_s is conserved along the trajectories of the particles, and is typically coupled with the Maxwell equations to take into account the self-consistent electromagnetic field generated by the particles (Sinitsyn, Dulov and Vedenyapin 2011).

The so-called Vlasov–Poisson equation describes a plasma with only one atomic species (alternatively, electrons and positive ions) in the mean electric field derived from the potential ϕ created by the particles. Since $m_e \ll m_i$, then the effect of the ions can be treated as a uniform neutralizing background. Taking all constants equal to 1 and denoting the distribution by f, the relevant system of equations describing the electron dynamics reads

$$\begin{aligned} &\frac{\partial f}{\partial t} + v \cdot \nabla_x f - \nabla_x \phi(f) \cdot \nabla_v f = 0 \\ &\Delta_x \phi(f)(x) = -\left(\int_{\mathbb{R}^d} f(x, v) \, \mathrm{d}v - \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d \times \mathbb{R}^d} f(x, v) \, \mathrm{d}x \, \mathrm{d}v \right) \end{aligned} \tag{9.35}$$

in the domain $(x, v) \in \mathbb{T}^d \times \mathbb{R}^d$, $d = 1, 2, 3$. The energy associated with equations (9.35) is

$$\begin{aligned}\mathcal{H}(f) &= \int_{\mathbb{T}^d\times\mathbb{R}^d} \frac{|v|^2}{2} f(x, v)\,\mathrm{d}x\,\mathrm{d}v + \int_{\mathbb{T}^d} \frac{1}{2}|E(f)(x)|^2\,\mathrm{d}x \\ &= \mathcal{T}(f) + \mathcal{U}(f)\end{aligned} \tag{9.36}$$

and is preserved along the solution. In fact the system has infinitely many invariants, but not all of them can be preserved by numerical schemes. Taking into account that the solution of the equations associated with $\mathcal{T}$ and $\mathcal{U}$ can be solved exactly (up to a phase space discretization), it is then natural to consider splitting methods for the time integration. Specifically, letting $\varphi_t^{\mathcal{T}}(f)$ denote the solution of $\partial_t f + v\cdot\nabla_x f = 0$, that is,

$$f(t, x, v) = f(0, x - tv, v),$$

and letting $\varphi_t^{\mathcal{U}}(f)$ be the solution of $\partial_t f - \nabla_x\phi(f)\cdot\nabla_v f = 0$, which reads

$$f(t, x, v) = f(0, x, v - tE(f(0))),$$

where $E(f(0))$ is the value of the electric field at time $t = 0$, the numerical integrators are of the form

$$\varphi^{\mathcal{U}}_{b_{s+1}h}(f) \circ \varphi^{\mathcal{T}}_{a_s h}(f) \circ \varphi^{\mathcal{U}}_{b_s h}(f) \circ \cdots \circ \varphi^{\mathcal{U}}_{b_2 h}(f) \circ \varphi^{\mathcal{T}}_{a_1 h}(f) \circ \varphi^{\mathcal{U}}_{b_1 h}(f).$$

The convergence of these schemes can be established by requiring the appropriate smoothness of f and using the Hamiltonian structure of the system. Moreover, the functionals $\mathcal{T}$ and $\mathcal{U}$ in the decomposition (9.36) satisfy the formal relation

$$[[[\mathcal{T},\mathcal{U}],\mathcal{U}],\mathcal{U}](f) = 0 \quad \text{for all } f,$$

where $[\cdot,\cdot]$ is the Poisson bracket associated with the infinite-dimensional Poisson structure of the system (see Casas, Crouseilles, Faou and Mehrenberger 2017 for a detailed treatment), so that Runge–Kutta–Nyström splitting methods can be safely used in this setting. In addition, when $d = 1$,

$$[[\mathcal{T},\mathcal{U}],\mathcal{U}](f) = 2m(f)\mathcal{U}(f), \quad \text{where} \quad m(f) = \frac{1}{2\pi}\int_{\mathbb{T}\times\mathbb{R}} f(x, v)\,\mathrm{d}x\,\mathrm{d}v.$$

Since $m(f)$ is a constant of the motion, this introduces additional simplifications. In particular, methods involving nested commutators only require the evaluation of $\mathcal{U}$ with appropriate coefficients, and thus methods up to order six can be designed involving a reduced number of maps that, when combined with semi-Lagrangian techniques in phase space, provides high efficiency (Casas *et al.* 2017)

9.9. Quantum simulation of quantum systems

As noted by Feynman (1982), simulating the full time evolution of arbitrary quantum systems on a classical computer requires exponential amounts of computational resources: the states of the system are wave functions that belong to a

vector space whose dimension grows exponentially with the size of the system, so that merely to record the state of the system is already cumbersome. For this reason, he conjectured the possibility of using a specific quantum system (a quantum computer) to simulate the behaviour of arbitrary quantum systems whose dynamics are determined by local interactions. This conjecture was later shown to be correct (Lloyd 1996). Today, quantum computers can simulate a variety of systems arising in quantum chemistry, quantum field theory and many-body physics, etc.; see e.g. Berry *et al.* (2015), Childs and Su (2019) and the recent review by Miessen, Ollitrault, Tacchino and Tavernelli (2023).

Simulating the time evolution of a quantum system with Hamiltonian H requires approximating $\mathrm{e}^{-\mathrm{i}tH}$ ($\hbar = 1$). If the Hamiltonian is the sum over many local interactions, then one can also use splitting methods for this purpose. Specifically, suppose that for a system composed of n variables, H can be decomposed as

$$H = \sum_{j=1}^{\ell} H_j, \tag{9.37}$$

where each H_j acts on a space of dimension m_j encompassing at most k of the variables, and ℓ is a polynomial in n. For instance, the well-known Hubbard, Ising and Heisenberg models belong to this class (Nielsen and Chuang 2010). The important point is that, whereas $\mathrm{e}^{-\mathrm{i}tH}$ is difficult to compute, each $\mathrm{e}^{-\mathrm{i}tH_j}$ acts on a much smaller subsystem and can be straightforward to evaluate by quantum circuits. In fact, the explicit quantum simulation algorithm proposed in Lloyd (1996) is based on the Lie–Trotter method

$$\chi_t = \mathrm{e}^{-\mathrm{i}tH_\ell} \cdots \mathrm{e}^{-\mathrm{i}tH_2}\, \mathrm{e}^{-\mathrm{i}tH_1},$$

whereas subsequent proposals include the Strang splitting

$$S_t^{[2]} = \mathrm{e}^{-\mathrm{i}\frac{t}{2}H_1}\, \mathrm{e}^{-\mathrm{i}\frac{t}{2}H_2} \cdots \mathrm{e}^{-\mathrm{i}tH_\ell} \cdots \mathrm{e}^{-\mathrm{i}\frac{t}{2}H_2}\, \mathrm{e}^{-\mathrm{i}\frac{t}{2}H_1},$$

and especially the quintuple jump recursion (2.5)

$$S_t^{[2k+2]} = \left(S_{\gamma_{2k}t}^{[2k]}\right)^2 \circ S_{(1-4\gamma_{2k})t}^{[2k]} \circ \left(S_{\gamma_{2k}t}^{[2k]}\right)^2 \tag{9.38}$$

with $\gamma_{2k} = 1/(4 - 4^{1/(2k+1)})$. These schemes are known in this setting as *product formulas* (Childs and Su 2019, Chen *et al.* 2022) and the procedure is called *Trotterization*.

Product formulas provide approximations U_{app} to the exact evolution $\mathrm{e}^{-\mathrm{i}tH}$ and the goal is, given a time t and a maximum simulation error ε, to find an algorithm (a quantum circuit) U_{app} such that $\|U_{\mathrm{app}} - \mathrm{e}^{-\mathrm{i}tH}\| < \varepsilon$. In consequence, it is of paramount importance to analyse and eventually provide tight bounds for the error ε committed by product formulas / splitting methods when applied to Hamiltonian

systems of the form (9.37). Thus, Childs *et al.* (2021) have shown that

$$\left\| S_t^{[2k]} - \mathrm{e}^{-\mathrm{i}tH} \right\| = O\big((B_H)^{2k+1}\big), \quad \text{with} \quad B_H \equiv \sum_{j=1}^{\ell} \|H_j\|\, t$$

if H_j are Hermitian. As usual, if t is large, then the whole interval is divided into N steps and (9.38) is applied within each step. In that case,

$$\left\| \big(S_{t/N}^{[2k]}\big)^N - \mathrm{e}^{-\mathrm{i}tH} \right\| = O(\varepsilon)$$

provided that

$$N = O\left(\frac{(B_H)^{1+1/(2k)}}{\varepsilon^{1/(2k)}} \right).$$

As we have already pointed out, whereas the recursion (9.38) (as well as the triple jump) constitutes a systematic way to achieve high-order approximations, it is not necessarily the most efficient, both in terms of errors and the number of exponentials involved. It makes sense, then, to consider some other methods in this setting, such as those collected in Section 8. This requires, in particular, a detailed analysis to achieve more stringent bounds for the corresponding errors than those obtained in Childs *et al.* (2021).

Hamiltonian systems of the form $H = -\frac{1}{2}\Delta + V(x)$ can also be simulated on quantum computers by using an appropriate representation of the states and the previous product formulas with the quantum Fourier transform (Nielsen and Chuang 2010).

9.10. Other topics

Space and time constraints prevent us from including in this review additional relevant applications where splitting methods have shown their merits, as well as other closely related important issues. Let us briefly mention some of them.

- Except for a brief incursion into Langevin dynamics, we have restricted our treatment to deterministic problems, although splitting methods have been widely applied to both ordinary and partial stochastic differential equations. Recent references in this area include Bréhier, Cohen and Jahnke (2023), Bréhier, Cohen and Giordano (2024) and Foster, dos Reis and Strange (2024).
- The evolution of a particle of mass m and charge q in a given electromagnetic field is modelled by the Lorentz equation $mx'' = q(E + x' \times B)$. By introducing the velocity $v = x'$ as a new variable, the resulting system of first-order ODEs can be split into three explicitly solvable parts, so that composition methods can be applied preserving volume in phase space (He, Sun, Liu and Qin 2015, 2016). The treatment can also be generalized to relativistic charged particles (Zhang *et al.* 2015).

- In addition to the Schrödinger equation, both linear and nonlinear, splitting methods have been applied to other relevant partial differential equations arising in quantum physics. This is the case, in particular, for the Dirac equation (Bao, Cai and Yin 2020) and the Klein–Gordon equation (Bao, Cai and Feng 2022).
- One important theoretical aspect not treated here refers to the convergence analysis of splitting methods applied to the nonlinear Schrödinger equation and other semi-linear Hamiltonian PDEs, as well as the use of the Birkhoff normal forms. This fascinating topic is the subject of much attention in the recent literature (see e.g. Faou 2012, Faou and Grébert 2011, Bernier and Grébert 2022) and probably deserves a review in its own right. It might be of interest to explore whether the formalism of extended word series, successfully applied in Murua and Sanz-Serna (2016) to construct formal invariants and normal forms of general classes of finite-dimensional Hamiltonian systems, could be adapted and applied for semi-linear Hamiltonian PDEs.
- Here we have only considered the application of splitting and composition methods with constant step size h. This is essential if we are interested in preserving qualitative properties of the system. As we have seen in Section 4, the modified equation corresponding to the numerical scheme depends explicitly on h, so that if h is changed then so is the modified equation, and the preserving properties that geometric integrators possess when h is constant are no longer guaranteed. In other problems, such as those defined by PDEs with no particular structure, this is not a matter of concern. There are systems which do possess a geometric structure that is advantageous to preserve by the numerical scheme and where the use of an adaptive step size is of the utmost importance to get efficient approximations. A relevant example is the gravitational N-body problem when there are close encounters between some of the bodies. In that case, one may apply splitting methods with variable step size by using some specifically designed transformations involving the time variable, in such a way that in the new variables the resulting time step is constant; see e.g. Mikkola (1997), Calvo, López-Marcos and Sanz-Serna (1998), Blanes and Budd (2005), Blanes and Iserles (2012) and Hairer *et al.* (2006, sect. VIII.2).

Forse altro canterà con miglior plectro.

Acknowledgements

The authors wish to thank J. Bernier, J. Boronat, A. Iserles, A. Pérez, J. M. Sanz-Serna, L. Shaw and M. Thalhammer for reading parts of this manuscript and for their comments and suggestions. They would also like to express their gratitude to their long-suffering families for their patience, understanding and wholehearted support. This work has been funded by Ministerio de Ciencia e Innovación (Spain) through projects PID2022-136585NB-C21 and PID2022-136585NB-C22,

MCIN/AEI/10.13039/501100011033/FEDER, UE, and also by Generalitat Valenciana (Spain) through project CIAICO/2021/180. AM has also received funding from the Department of Education of the Basque Government through the Consolidated Research Group MATHMODE (ITI456-22).

Appendix: Testing splitting methods on matrices

In Section 8 we collected more than 80 different splitting and composition methods that are available in the literature. Although they have been classified into different categories according to the structure of the system to be integrated (RKN splitting methods, near-integrable systems, etc.) and the type of scheme (composition of a basic second-order time-symmetric method, splitting for a system separated into two parts, etc.), it is not obvious in advance which particular method is the most suitable for carrying out a given integration of a specific problem. There are many factors involved in the final choice: type of problem, the qualitative properties we are interested in preserving, computational cost, desired accuracy, time integration interval and even initial condition.

For these reasons, it might be illustrative to fix a list of examples and apply to them the most representative numerical integrators gathered in Section 8, just to get some clues about their relative performances. To this end, we take the linear system

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F\,X, \quad X(0) = I, \tag{A.1}$$

where F is a constant real $d \times d$ matrix and I is the identity, so that the exact solution $X(t) = \mathrm{e}^{tF}$ can be easily computed with any desired accuracy. Obviously, none of the methods we test here should be used, in general, to approximate the exponential of a matrix or its action on a vector (unless additional information on the structure of the matrices is known which could make splitting an efficient technique). Our purpose is rather to compare the relative performance of the different schemes.

Let Φ_h^N, with $t_f = Nh$, denote the matrix obtained with a given method applied N times with step size h that approximates e^{tF} at $t = t_f$. We then compute the following relative errors:

$$\mathcal{E}_1 = \frac{\|\mathrm{e}^{tF} - \Phi_h^N\|}{\|\mathrm{e}^{tF}\|}, \quad \mathcal{E}_2 = \frac{|\mathrm{tr}(\mathrm{e}^{tF}) - \mathrm{tr}(\Phi_h^N)|}{|\mathrm{tr}(\mathrm{e}^{tF})|}.$$

The first, $\mathcal{E}_1$, measures the accuracy of the method, whereas $\mathcal{E}_2$ provides an estimate of the accuracy of the scheme in the case where it is used as a kernel with an ideal processor. In all cases, $\|\cdot\|$ refers to the 2-norm of the matrices considered.

The numerical experiments reported here have been carried out with MATLAB, and the function `randn( )` has been used to construct the matrix elements randomly from a normal distribution, initiated with the seed `rng(1)` for reproducibility. Additional material related to these simulations can be found at our website accompanying this paper.

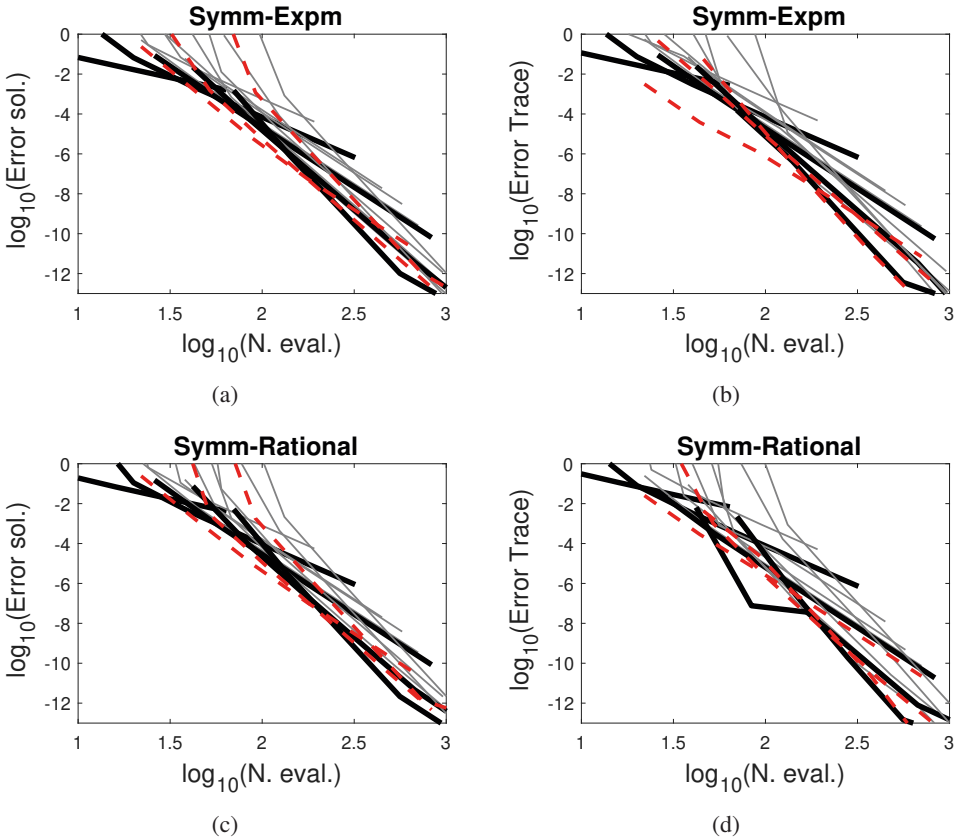


Figure A.1. Errors $\mathcal{E}_1$ (a,c) and $\mathcal{E}_2$ (b,d) vs. the number of evaluations of the basic scheme for $S_h^{[2,1]}$ (a,b) and $S_h^{[2,2]}$ (c,d). The non-processed schemes selected in Table 8.1 are shown in thick solid lines and the selected processed methods are drawn as thick red dashed lines.

Symmetric compositions of time-symmetric second-order schemes. We take $F = A + B + C$, with A, B, C matrices of dimension 50×50 constructed as previously indicated, that is,

```
d=50; rng(1);  A=randn(d);   A=A/norm(A); B=randn(d); ...
```

To illustrate the role that the basic scheme may play in the overall performance of the composition, we take the following two second-order time-symmetric methods:

$$S_h^{[2,1]} = \mathrm{e}^{hA/2}\mathrm{e}^{hB/2}\mathrm{e}^{hC}\mathrm{e}^{hB/2}\mathrm{e}^{hA/2} \tag{A.2}$$

and

$$S_h^{[2,2]} = \left(I - \frac{h}{2}A\right)^{-1}\left(I - \frac{h}{2}B\right)^{-1}\left(I - \frac{h}{2}C\right)^{-1}\left(I + \frac{h}{2}C\right)\left(I + \frac{h}{2}B\right)\left(I + \frac{h}{2}A\right). \tag{A.3}$$

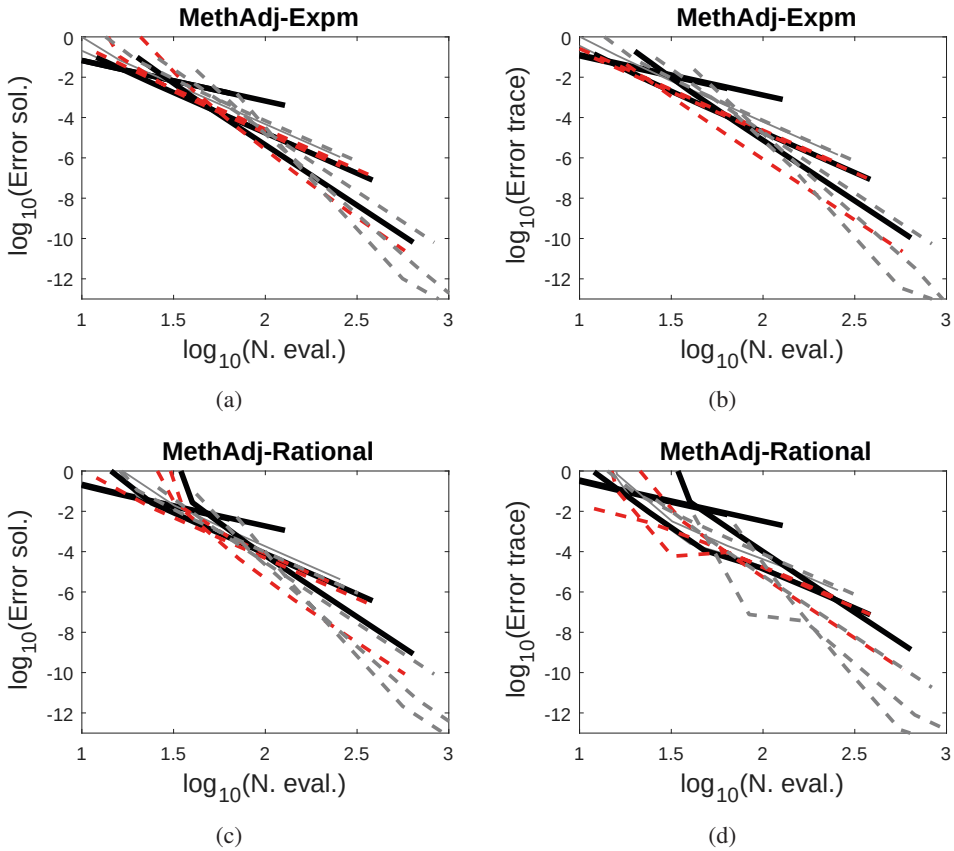


Figure A.2. Errors $\mathcal{E}_1$ (a,c) and $\mathcal{E}_2$ (b,d) vs. the number of evaluations of the basic scheme for $\chi_h^{[1,1]}$ (a,b) and $\chi_h^{[1,2]}$ (c,d). Now the schemes are compositions of $\chi_h^{[1,j]}$ and $\left(\chi_h^{[1,j]}\right)^*$.

The two differ in their computational cost and error terms. For the most relevant methods presented in Table 8.1, we plot the errors $\mathcal{E}_1$ and $\mathcal{E}_2$ as a function of the total number of evaluations of the basic scheme. Figure A.1 shows the results obtained for $t_f = 10$ when the basic scheme is $S_h^{[2,1]}$ (a,b) and $S_h^{[2,2]}$ (c,d). Compositions with $s = 5$, 13, 19 and 35 stages of order $r = 4$, 6, 8 and 10, respectively, which on average show good performance in most tested examples, correspond to the thick black lines, whereas the remaining non-processed methods are shown as thin lines. On the other hand, the recommended processed methods in Table 8.1 with kernels having $s = 13$, 19 and 23 stages of order 6, 8 and 10, respectively, are represented by thick red dashed lines.

Looking at these figures, it is hardly possible to recommend a particular scheme as the one leading to the best results for the example considered. We may only conclude that for some intervals of accuracy, there are certain methods exhibiting

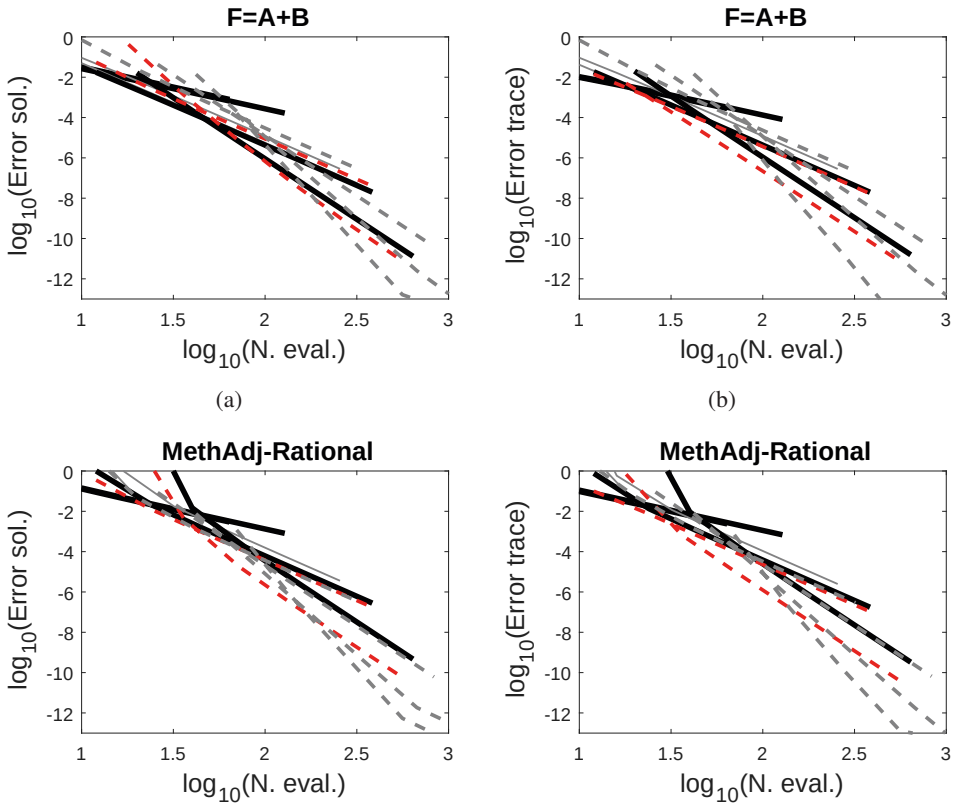


Figure A.3. The same as in Figure A.2 but now for the splitting $F = A+B$.

the best performance. Moreover, this performance also depends on the particular basic scheme chosen as $S_h^{[2]}$ in the compositions and the nature of the composition itself: whether it is intended to be used as the kernel of a processed method (with an optimal processor). In many situations, the relative performance displayed in Figure A.1(a,c) should be close to that of Figure A.1(b,d), when long-time integrations are considered and the contribution from the processable error terms can be neglected.

Splitting into two parts / composition of a basic method and its adjoint. We again consider $F = A+B+C$, with A, B, C the same matrices as in the previous example, but instead we take as basic methods the first-order compositions

$$\chi_h^{[1,1]} = \mathrm{e}^{hA}\mathrm{e}^{hB}\mathrm{e}^{hC}, \quad \left(\chi_h^{[1,1]}\right)^* = \mathrm{e}^{hC}\mathrm{e}^{hB}\mathrm{e}^{hA},$$

and

$$\chi_h^{[1,2]} = (I-hA)^{-1}(I-hB)^{-1}(I-hC)^{-1}, \quad \left(\chi_h^{[1,2]}\right)^* = (I+hC)(I+hB)(I+hA).$$

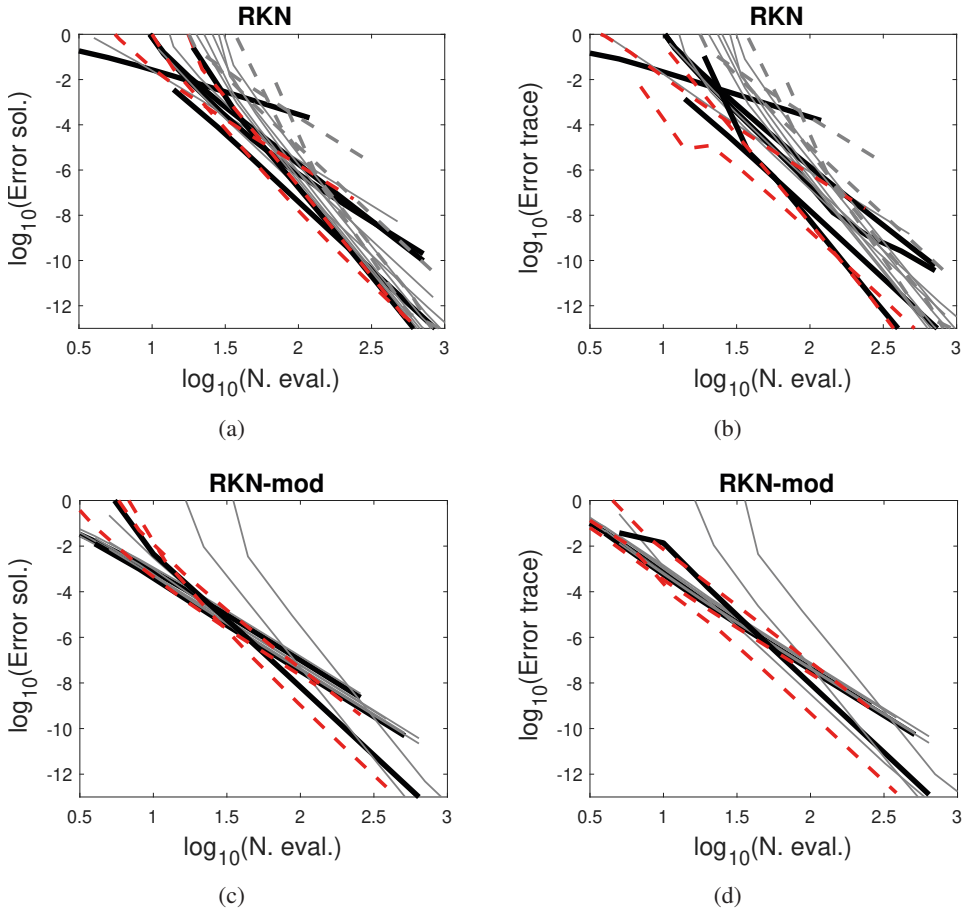


Figure A.4. RKN methods without (a,b) and with (c,d) modified potentials.

We plot the errors $\mathcal{E}_1$ and $\mathcal{E}_2$ versus the total number of evaluations of the basic scheme for most of the methods presented in Table 8.2. Figure A.2 shows the results obtained for $t_f = 10$. As before, the thick black solid lines correspond to the selected non-processed schemes and the thick red dashed lines correspond to the selected processed ones. The dashed grey lines correspond to the most efficient non-processed symmetric–symmetric schemes of orders 6, 8 and 10.

We repeat the same numerical experiments, but now taking $C = 0$, while keeping A, B the same matrices as previously. Hence, the problem has to be seen as separable into two parts. Notice that some methods have been optimized for this particular case. Figure A.3 shows the results obtained, where the superiority of the methods for this class of problems is clear for low-to-medium accuracy.

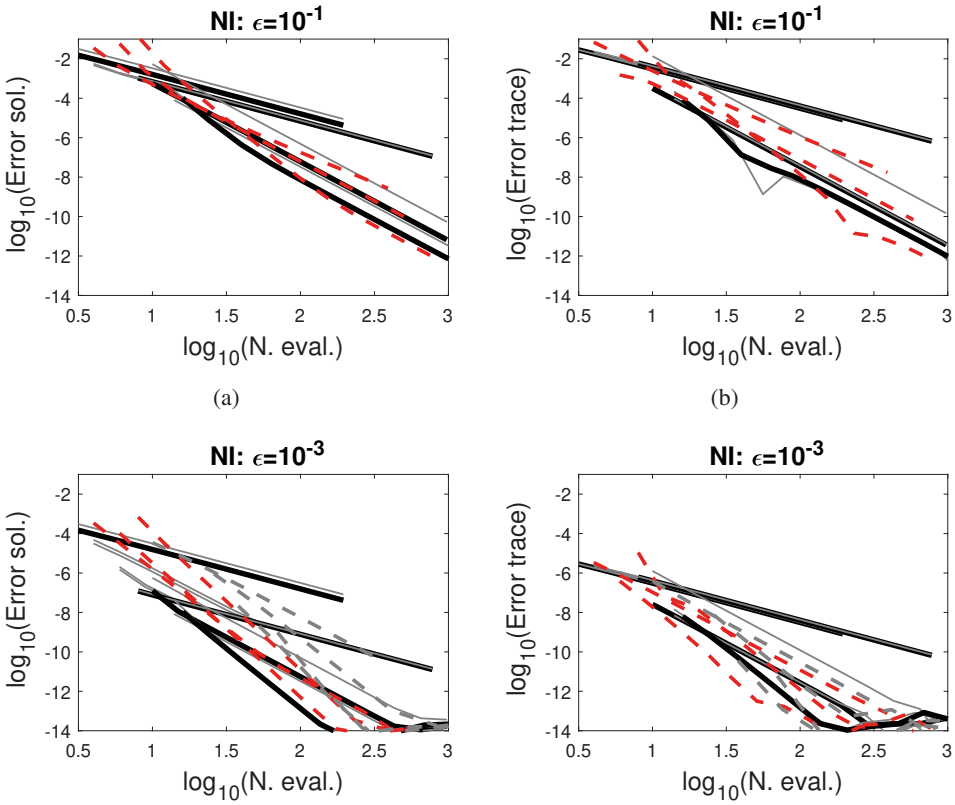


Figure A.5. Methods for near-integrable systems.

Runge–Kutta–Nyström methods. To test the RKN splitting methods gathered in Tables 8.3 and 8.4, we take $F = A + B$, with

$$A = \begin{pmatrix} O_d & O_d \\ A_1 & O_d \end{pmatrix}, \quad B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix},$$

where A_1, B_i, $i = 1, 2, 3, 4$ are matrices of dimension $d = 50$ with elements chosen as in the previous cases, with the same seed, and O_d is the null matrix. With this choice we have

$$[A, [A, B]] = \begin{pmatrix} O_d & O_d \\ 2A_1B_2A_1 & O_d \end{pmatrix}$$

and $[A, [A, [A, B]]] = 0$. For this particular choice, the computation of e^{hB} dominates the total cost of the method and the cost of evaluating e^{hA} and $\mathrm{e}^{hA+h^3[A,[A,B]]}$ can be neglected. Obviously, we can add an artificial cost to these exponentials as a test for different problems where this term can be more expensive to evaluate.

Figure A.4(a,b) shows the results for most of the methods from Table 8.3 for $t_f = 10$ where, as previously, the thick black lines correspond to the selected non-processed schemes, the thin lines correspond to the remaining non-processed schemes, the thick red dashed lines are the processed ones and the dashed grey lines correspond to the selected high-order symmetric–symmetric methods. Figure A.4(c,d) shows, for the same matrices, the results for the methods from Table 8.4 with modified potentials.

Methods for near-integrable systems. Let $F = A+\varepsilon B$ with A, B the same matrices as for the separable problem in two parts and ε a small parameter which corresponds to the relative norm of the matrices. We analyse the performance of methods tailored for perturbed problems for two choices of the small parameter: $\varepsilon = 10^{-1}$ and $\varepsilon = 10^{-3}$. Figure A.5 shows the results for the methods from Table 8.5 for $t_f = 10$. For small values of ε none of the previous splitting methods are competitive against the most efficient ones from this family.

References

M. Aichinger and E. Krotscheck (2005), A fast configuration space method for solving local Kohn–Sham equations, *Comput. Mater. Sci.* **34**, 188–212.

E. Akhmatskaya, M. Fernández-Pendás, T. Radivojević and J. M. Sanz-Serna (2017), Adaptive splitting integrators for enhancing sampling efficiency of modified Hamiltonian Monte Carlo methods in molecular simulations, *Langmuir* **33**, 11530–11542.

A. Alamo and J. M. Sanz-Serna (2016), A technique for studying strong and weak local errors of splitting stochastic integrators, *SIAM J. Numer. Anal.* **54**, 3239–3257.

E. Alberdi, M. Antoñana, J. Makazaga and A. Murua (2019), An algorithm based on continuation techniques for minimization problems with highly non-linear equality constraints. Available at arXiv:1909.07263.

I. Alonso-Mallo, B. Cano and N. Reguera (2018), Avoiding order reduction when integrating linear initial boundary value problems with exponential splitting methods, *IMA J. Numer. Anal.* **38**, 1294–1323.

I. Alonso-Mallo, B. Cano and N. Reguera (2019), Avoiding order reduction when integrating reaction–diffusion boundary value problems with exponential splitting methods, *J. Comput. Appl. Math.* **357**, 228–250.

A. Alvermann and H. Fehske (2011), High-order commutator-free exponential time-propagation of driven quantum systems, *J. Comput. Phys.* **230**, 5930–5956.

H. C. Andersen (1983), Rattle: A 'velocity' version of the Shake algorithm for molecular dynamics calculations, *J. Comput. Phys.* **52**, 24–34.

M. H. Anderson, J. R. Ensher, M. R. Mathews, C. E. Wieman and E. A. Cornell (1995), Observations of Bose–Einstein condensation in a dilute atomic vapor, *Science* **269**, 198–201.

V. I. Arnold (1989), *Mathematical Methods of Classical Mechanics*, second edition, Springer.

U. Ascher, S. Ruuth and R. Spiteri (1997), Implicit–explicit Runge–Kutta methods for time-dependent partial differential equations, *Appl. Numer. Math.* **25**, 151–167.

U. Ascher, S. Ruuth and B. Wetton (1995), Implicit–explicit methods for time-dependent partial differential equations, *SIAM J. Numer. Anal.* **32**, 797–823.

J. Auer, E. Krotscheck and S. A. Chin (2001), A fourth-order real-space algorithm for solving local Schrödinger equations, *J. Chem. Phys.* **115**, 6841–6846.

P. Bader and S. Blanes (2011), Fourier methods for the perturbed harmonic oscillator in linear and nonlinear Schrödinger equations, *Phys. Rev. E* **83**, art. 046711.

P. Bader, S. Blanes and F. Casas (2013), Solving the Schrödinger eigenvalue problem by the imaginary time propagation technique using splitting methods with complex coefficients, *J. Chem. Phys.* **139**, art. 124117.

P. Bader, A. Iserles, K. Kropielnicka and P. Singh (2014), Effective approximation for the semiclassical Schrödinger equation, *Found. Comput. Math.* **14**, 689–720.

P. Bader, A. Iserles, K. Kropielnicka and P. Singh (2016), Efficient methods for linear Schrödinger equation in the semiclassical regime with time-dependent potential, *Proc. R. Soc. A* **472**, art. 20150733.

A. D. Bandrauk and H. Shen (1991), Improved exponential split operator method for solving the time-dependent Schrödinger equation, *Chem. Phys. Lett.* **176**, 428–432.

A. D. Bandrauk, E. Dehghanian and H. Lu (2006), Complex integration steps in decomposition of quantum exponential evolution operators, *Chem. Phys. Lett.* **419**, 346–350.

W. Bao, Y. Cai and Y. Feng (2022), Improved uniform error bounds on time-splitting methods for long-time dynamics of the nonlinear Klein–Gordon equation with weak nonlinearity, *SIAM J. Numer. Anal.* **60**, 1962–1984.

W. Bao, Y. Cai and J. Yin (2020), Super-resolution of time-splitting methods for the Dirac equation in the nonrelativistic regime, *Math. Comp.* **89**, 2141–2173.

W. Bao, D. Jaksch and P. A. Markowich (2003*a*), Numerical solution of the Gross–Pitaevskii equation for Bose–Einstein condensation, *J. Comput. Phys.* **187**, 318–342.

W. Bao, D. Jaksch and P. A. Markowich (2004), Three dimensional simulation of jet formation in collapsing condensates, *J. Phys. B* **37**, 329–343.

W. Bao, S. Jin and P. A. Markowich (2002), On time-splitting spectral approximations for the Schrödinger equation in the semiclassical regime, *J. Comput. Phys.* **175**, 487–524.

W. Bao, S. Jin and P. A. Markowich (2003*b*), Numerical study of time-splitting spectral discretizations of nonlinear Schrödinger equations in the semiclassical regimes, *SIAM J. Sci. Comput.* **25**, 27–64.

R. Bellman (1970), *Introduction to Matrix Analysis*, second edition, McGraw-Hill.

J. Bernier (2021), Exact splitting methods for semigroups generated by inhomogeneous quadratic differential operators, *Found. Comput. Math.* **21**, 1401–1439.

J. Bernier and B. Grébert (2022), Birkhoff normal forms for Hamiltonian PDEs in their energy space, *J. Éc. Polytech. Math.* **9**, 681–745.

J. Bernier, S. Blanes, F. Casas and A. Escorihuela-Tomàs (2023), Symmetric-conjugate splitting methods for linear unitary problems, *BIT* **63**, art. 58.

J. Bernier, N. Crouseilles and Y. Li (2021), Exact splitting methods for kinetic and Schrödinger equations, *J. Sci. Comput.* **86**, art. 10.

D. W. Berry, A. M. Childs, R. Cleve, R. Kothari and R. D. Somma (2015), Simulating Hamiltonian dynamics with a truncated Taylor series, *Phys. Rev. Lett.* **114**, art. 090502.

S. Blanes and C. J. Budd (2005), Adaptive geometric integrators for Hamiltonian problems with approximate scale invariance, *SIAM J. Sci. Comput.* **26**, 1089–1113.

S. Blanes and F. Casas (2005), On the necessity of negative coefficients for operator splitting schemes of order higher than two, *Appl. Numer. Math.* **54**, 23–37.

S. Blanes and F. Casas (2006), Splitting methods for non-autonomous separable dynamical systems, *J. Phys. A* **39**, 5405–5423.

S. Blanes and F. Casas (2016), *A Concise Introduction to Geometric Numerical Integration*, CRC Press.

S. Blanes and V. Gradinaru (2020), High order efficient splittings for the semiclassical time-dependent Schrödinger equation, *J. Comput. Phys.* **405**, art. 109157.

S. Blanes and A. Iserles (2012), Explicit adaptive symplectic integrators for solving Hamiltonian systems, *Celestial Mech. Dynam. Astronom.* **114**, 297–317.

S. Blanes and P. C. Moan (2002), Practical symplectic partitioned Runge–Kutta and Runge–Kutta–Nyström methods, *J. Comput. Appl. Math.* **142**, 313–330.

S. Blanes, M. P. Calvo, F. Casas and J. M. Sanz-Serna (2021*a*), Symmetrically processed splitting integrators for enhanced Hamiltonian Monte Carlo sampling, *SIAM J. Sci. Comput.* **43**, A3357–A3371.

S. Blanes, F. Casas and A. Escorihuela-Tomàs (2022*a*), Applying splitting methods with complex coefficients to the numerical integration of unitary problems, *J. Comput. Dyn.* **9**, 85–101.

S. Blanes, F. Casas and A. Escorihuela-Tomàs (2022*b*), Runge–Kutta–Nyström symplectic splitting methods of order 8, *Appl. Numer. Math.* **182**, 14–27.

S. Blanes, F. Casas and A. Murua (2004), On the numerical integration of ordinary differential equations by processed methods, *SIAM J. Numer. Anal.* **42**, 531–552.

S. Blanes, F. Casas and A. Murua (2006*a*), Composition methods for differential equations with processing, *SIAM J. Sci. Comput.* **27**, 1817–1843.

S. Blanes, F. Casas and A. Murua (2006*b*), Symplectic splitting operator methods tailored for the time-dependent Schrödinger equation, *J. Chem. Phys.* **124**, art. 234105.

S. Blanes, F. Casas and A. Murua (2008*a*), On the linear stability of splitting methods, *Found. Comput. Math.* **8**, 357–393.

S. Blanes, F. Casas and A. Murua (2008*b*), Splitting and composition methods in the numerical integration of differential equations, *Bol. Soc. Esp. Mat. Apl.* **45**, 89–145.

S. Blanes, F. Casas and A. Murua (2011), Error analysis of splitting methods for the time dependent Schrödinger equation, *SIAM J. Sci. Comput.* **33**, 1525–1548.

S. Blanes, F. Casas and A. Murua (2015), An efficient algorithm based on splitting for the time integration of the Schrödinger equation, *J. Comput. Phys.* **303**, 396–412.

S. Blanes, F. Casas and A. Murua (2017*a*), Symplectic time-average propagators for the Schrödinger equation with a time-dependent Hamiltonian, *J. Chem. Phys.* **146**, art. 114109.

S. Blanes, F. Casas and J. Ros (1999*a*), Extrapolation of symplectic integrators, *Celestial Mech. Dynam. Astronom.* **75**, 149–161.

S. Blanes, F. Casas and J. Ros (1999*b*), Symplectic integrators with processing: A general study, *SIAM J. Sci. Comput.* **21**, 711–727.

S. Blanes, F. Casas and J. Ros (2000), Processing symplectic methods for near-integrable Hamiltonian systems, *Celestial Mech. Dynam. Astronom.* **77**, 17–35.

S. Blanes, F. Casas and J. Ros (2001*a*), High-order Runge–Kutta–Nyström geometric methods with processing, *Appl. Numer. Math.* **39**, 245–259.

S. Blanes, F. Casas and J. Ros (2001*b*), New families of symplectic Runge–Kutta–Nyström integration methods, in *Numerical Analysis and its Applications*, Vol. 1988 of Lecture Notes in Computer Science, Springer, pp. 102–109.

S. Blanes, F. Casas and J. M. Sanz-Serna (2014), Numerical integrators for the Hybrid Monte Carlo method, *SIAM J. Sci. Comput.* **36**, A1556–A1580.

S. Blanes, F. Casas and M. Thalhammer (2017*b*), High-order commutator-free quasi-Magnus exponential integrators for non-autonomous linear evolution equations, *Comput. Phys. Commun.* **220**, 243–262.

S. Blanes, F. Casas and M. Thalhammer (2018), Convergence analysis of high-order commutator-free quasi-Magnus integrators for nonautonomous linear evolution equations, *IMA J. Numer. Anal.* **38**, 743–778.

S. Blanes, F. Casas, P. Chartier and A. Escorihuela-Tomàs (2022*c*), On symmetric-conjugate composition methods in the numerical integration of differential equations, *Math. Comp.* **91**, 1739–1761.

S. Blanes, F. Casas, P. Chartier and A. Murua (2013*a*), Optimized high-order splitting methods for some classes of parabolic equations, *Math. Comp.* **82**, 1559–1576.

S. Blanes, F. Casas, A. Farrés, J. Laskar, J. Makazaga and A. Murua (2013*b*), New families of symplectic splitting methods for numerical integration in dynamical astronomy, *Appl. Numer. Math.* **68**, 58–72.

S. Blanes, F. Casas, C. González and M. Thalhammer (2021*b*), Convergence analysis of high-order commutator-free quasi-Magnus integrators for nonautonomous linear Schrödinger equations, *IMA J. Numer. Anal.* **41**, 594–617.

S. Blanes, F. Casas, C. González and M. Thalhammer (2023), Efficient splitting methods based on modified potentials: Numerical integration of linear parabolic problems and imaginary time propagation of the Schrödinger equation, *Commun. Comput. Phys.* **33**, 937–961.

S. Blanes, F. Casas, J. A. Oteo and J. Ros (2009), The Magnus expansion and some of its applications, *Phys. Rep.* **470**, 151–238.

S. Blanes, A. Iserles and S. Macnamara (2022*d*), Positivity-preserving methods for ordinary differential equations, *ESAIM Math. Model. Numer. Anal.* **56**, 1843–1870.

C. Bolley and M. Crouzeix (1978), Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques, *RAIRO Anal. Numér.* **12**, 237–245.

N. Bou-Rabee (2017), Cayley splitting for second-order Langevin stochastic partial differential equations. Available at arXiv:1707.05603.

N. Bou-Rabee and J. M. Sanz-Serna (2018), Geometric integrators and the Hamiltonian Monte Carlo method, *Acta Numer.* **27**, 113–206.

C.-E. Bréhier, D. Cohen and G. Giordano (2024), Splitting schemes for Fitzhugh–Nagumo stochastic partial differential equations, *Discrete Contin. Dyn. Syst. Ser. B* **29**, 214–244.

C.-E. Bréhier, D. Cohen and T. Jahnke (2023), Splitting integrators for stochastic Lie–Poisson systems, *Math. Comp.* **92**, 2167–2216.

S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds (2011), *Handbook of Markov Chain Monte Carlo*, CRC Press.

J. C. Butcher (1969), The effective order of Runge–Kutta methods, in *Proceedings of the Conference on the Numerical Solution of Differential Equations* (J. L. Morris, ed.), Vol. 109 of Lecture Notes in Mathematics, Springer, pp. 133–139.

J. C. Butcher and J. M. Sanz-Serna (1996), The number of conditions for a Runge–Kutta method to have effective order p, *Appl. Numer. Math.* **22**, 103–111.

S. T. Butler and M. H. Friedman (1955), Partition function for a system of interacting Bose–Einstein particles, *Phys. Rev.* **98**, 287–293.

M. P. Calvo and E. Hairer (1995), Accurate long-term integration of dynamical systems, *Appl. Numer. Math.* **18**, 95–105.

M. P. Calvo and J. M. Sanz-Serna (1993*a*), The development of variable-step symplectic integrators, with application to the two-body problem, *SIAM J. Sci. Comput.* **14**, 936–952.

M. P. Calvo and J. M. Sanz-Serna (1993*b*), High-order symplectic Runge–Kutta–Nyström methods, *SIAM J. Sci. Comput.* **14**, 1237–1252.

M. P. Calvo, M. A. López-Marcos and J. M. Sanz-Serna (1998), Variable step implementation of geometric integrators, *Appl. Numer. Math.* **28**, 1–16.

M. Campostrini and P. Rossi (1990), A comparison of numerical algorithms for dynamical fermions, *Nuclear Phys. B* **329**, 753–764.

J. Candy and W. Rozmus (1991), A symplectic integration algorithm for separable Hamiltonian functions, *J. Comput. Phys.* **92**, 230–256.

R. Carles (2008), *Semi-Classical Analysis for Nonlinear Schrödinger Equations*, World Scientific.

F. Casas and A. Murua (2009), An efficient algorithm for computing the Baker–Campbell–Hausdorff series and some of its applications, *J. Math. Phys.* **50**, art. 033513.

F. Casas, N. Crouseilles, E. Faou and M. Mehrenberger (2017), High-order Hamiltonian splitting for the Vlasov–Poisson equations, *Numer. Math.* **135**, 769–801.

F. Casas, J. M. Sanz-Serna and L. Shaw (2022), Split Hamiltonian Monte Carlo revisited, *Statist. Comput.* **32**, art. 86.

F. Casas, J. M. Sanz-Serna and L. Shaw (2023), A new optimality property of Strang's splitting, *SIAM J. Numer. Anal.* **61**, 1369–1385.

F. Castella, P. Chartier, S. Descombes and G. Vilmart (2009), Splitting methods with complex times for parabolic equations, *BIT* **49**, 487–508.

T. Cazenave (2003), *Semilinear Schrödinger Equations*, American Mathematical Society.

D. M. Ceperley (1995), Path integrals in the theory of condensed helium, *Rev. Modern Phys.* **67**, 279–355.

J. E. Chambers (2003), Symplectic integrators with complex time steps, *Astron. J.* **126**, 1119–1126.

R. Chan and A. Murua (2000), Extrapolation of symplectic methods for Hamiltonian problems, *Appl. Numer. Math.* **34**, 189–205.

P. J. Channell and J. C. Scovel (1990), Symplectic integration of Hamiltonian systems, *Nonlinearity* **3**, 231–259.

P. Chartier and A. Murua (2009), An algebraic theory of order, *ESAIM Math. Model. Numer. Anal.* **43**, 607–630.

K.-T. Chen (1957), Integration of paths, geometric invariants and a generalized Baker–Hausdorff formula, *Ann. of Math.* **65**, 163–178.

Y.-A. Chen, A. M. Childs, M. Hafezi, Z. Jiang, H. Kim and Y. Xu (2022), Efficient product formulas for commutators and applications to quantum simulations, *Phys. Rev. Res.* **4**, art. 013191.

A. M. Childs and Y. Su (2019), Nearly optimal lattice simulation by product formulas, *Phys. Rev. Lett.* **123**, art. 050503.

A. M. Childs, Y. Su, M. C. Tran, N. Wiebe and S. Zhu (2021), Theory of Trotter error with commutator scaling, *Phys. Rev. X* **11**, art. 011020.

S. A. Chin (1997), Symplectic integrators from composite operator factorizations, *Phys. Lett. A* **226**, 344–348.

S. A. Chin (2005), Structure of positive decomposition of exponential operators, *Phys. Rev. E* **71**, art. 016703.

S. A. Chin (2006), Complete characterization of fourth-order symplectic integrators with extended-linear coefficients, *Phys. Rev. E* **73**, art. 026705.

S. A. Chin (2023), Anatomy of path integral Monte Carlo: Algebraic derivation of the harmonic oscillator's universal discrete imaginary-time propagator and its sequential optimization, *J. Chem. Phys.* **159**, art. 134109.

S. A. Chin and C. R. Chen (2002), Gradient symplectic algorithms for solving the Schrödinger equation with time-dependent potentials, *J. Chem. Phys.* **117**, 1409–1415.

S. A. Chin and J. Geiser (2011), Multi-product operator splitting as a general method of solving autonomous and nonautonomous equations, *IMA J. Numer. Anal.* **31**, 1552–1577.

S. A. Chin and E. Krotscheck (2005), Fourth-order algorithms for solving the imaginary-time Gross–Pitaevskii equation in a rotating anisotropic trap, *Phys. Rev. E* **72**, art. 036705.

A. Chorin, T. J. R. Huges, J. E. Marsden and M. McCracken (1978), Product formulas and numerical algorithms, *Commun. Pure Appl. Math.* **31**, 205–256.

L. Y. Chou and P. W. Sharp (2000), Order 5 symplectic explicit Runge–Kutta–Nyström methods, *J. Appl. Math. Decision Sci.* **4**, 143–150.

P. G. Ciarlet and J. L. Lions, eds (1990), *Handbook of Numerical Analysis*, Vol. I, North-Holland.

J. E. Cohen, S. Friedland, T. Kato and F. P. Kelly (1982), Eigenvalue inequalities for products of matrix exponentials, *Linear Algebra Appl.* **45**, 55–95.

M. Creutz and A. Gocksch (1989), Higher-order hybrid Monte Carlo algorithms, *Phys. Rev. Lett.* **63**, 9–12.

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

J. M. A. Danby (1988), *Fundamentals of Celestial Mechanics*, Willmann-Bell.

H. de Raedt and B. de Raedt (1983), Applications of the generalized Trotter formula, *Phys. Rev. A* **28**, 3575–3580.

R. de Vogelaere (1956), Methods of integration which preserve the contact transformation property of the Hamiltonian equations. Report no. 4, Department of Mathematics, University of Notre Dame, IN.

J. C. del Valle and K. Kropielnicka (2023), Family of Strang-type exponential splitting in the presence of unbounded and time dependent operators. Available at arXiv:2310.01556.

J. Douglas and H. H. Rachford (1956), On the numerical solution of heat conduction problems in two and three space variables, *Trans. Amer. Math. Soc.* **82**, 421–439.

S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth (1987), Hybrid Monte-Carlo, *Phys. Lett. B* **195**, 216–222.

L. Einkemmer and A. Ostermann (2015), Overcoming order reduction in diffusion–reaction splitting, Part I: Dirichlet boundary conditions, *SIAM J. Sci. Comput.* **37**, A1577–A1592.

L. Einkemmer and A. Ostermann (2016), Overcoming order reduction in diffusion–reaction splitting, Part II: Oblique boundary conditions, *SIAM J. Sci. Comput.* **38**, A3741–A3757.

L. Einkemmer, M. Moccaldi and A. Ostermann (2018), Efficient boundary corrected Strang splitting, *Appl. Math. Comput.* **332**, 76–89.

K.-J. Engel and R. Nagel (2006), *A Short Course on Operator Semigroups*, Springer.

E. Faou (2012), *Geometric Numerical Integration and Schrödinger Equations*, European Mathematical Society.

E. Faou and B. Grébert (2011), Hamiltonian interpolation of splitting approximation for Hamiltonian PDEs, *Found. Comput. Math.* **11**, 381–415.

E. Faou and C. Lubich (2006), A Poisson integrator for Gaussian wavepacket dynamics, *Comput. Vis. Sci.* **9**, 45–55.

E. Faou, V. Gradinaru and C. Lubich (2009), Computing semi-classical quantum dynamics with Hagedorn wavepackets, *SIAM J. Sci. Comput.* **31**, 3027–3041.

M. D. Feit, J. A. Fleck Jr and A. Steiger (1982), Solution of the Schrödinger equation by a spectral method, *J. Comput. Phys.* **47**, 412–433.

K. Feng (1992), Formal power series and numerical algorithms for dynamical systems, in *Proc. Conf. Scientific Computation Hangzhou, 1991* (T. Chan and Z.-C. Shi, eds), World Scientific, pp. 28–35.

K. Feng and M. Z. Qin (1987), The symplectic methods for the computation of Hamiltonian equations, in *Numerical Methods for Partial Differential Equations* (Y. L. Zhu and B. Y. Gao, eds), Vol. 1297 of Lecture Notes in Mathematics, Springer, pp. 1–37.

M. Fernández-Pendás, E. Akhmatskaya and J. M. Sanz-Serna (2016), Adaptive multi-stage integrators for optimal energy conservation in molecular simulations, *J. Comput. Phys.* **327**, 434–449.

R. P. Feynman (1972), *Statistical Mechanics: A Set of Lectures*, Addison-Wesley.

R. P. Feynman (1982), Simulating physics with computers, *Internat. J. Theoret. Phys.* **21**, 467–488.

W. Folkner, J. Williams, D. Boggs, R. Park and P. Kuchynka (2014), The planetary and lunar ephemerides DE430 and DE431. IPN Progress Report 42-196.

E. Forest (1992), Sixth-order Lie group integrators, *J. Comput. Phys.* **99**, 209–213.

E. Forest and R. Ruth (1990), Fourth-order symplectic integration, *Phys. D* **43**, 105–117.

J. Foster, G. dos Reis and C. Strange (2024), High order splitting methods for SDEs satisfying a commutativity condition, *SIAM J. Numer. Anal.* **62**, 500–532.

B. García-Archilla, J. M. Sanz-Serna and R. D. Skeel (1999), Long-time-step methods for oscillatory differential equations, *SIAM J. Sci. Comput.* **20**, 930–963.

R. Glowinski, S. J. Osher and W. Yin, eds (2016*a*), *Splitting Methods in Communication, Imaging, Science, and Engineering*, Springer.

R. Glowinski, T.-W. Pan and X.-C. Tai (2016*b*), Some facts about operator-splitting and alternating direction methods, in *Splitting Methods in Communication, Imaging, Science, and Engineering* (R. Glowinski, S. J. Osher and W. Yin, eds), Springer, pp. 19–94.

S. Golden (1957), Statistical theory of many-electron systems: Discrete bases of representation, *Phys. Rev.* **107**, 1283–1290.

D. Goldman and T. J. Kaper (1996), *N*th-order operator splitting schemes and nonreversible systems, *SIAM J. Numer. Anal.* **33**, 349–367.

H. Goldstein (1980), *Classical Mechanics*, second edition, Addison-Wesley.

W. B. Gragg (1965), On extrapolation algorithms for ordinary differential equations, *SIAM J. Numer. Anal.* **2**, 384–403.

S. Gray and D. E. Manolopoulos (1996), Symplectic integrators tailored to the time-dependent Schrödinger equation, *J. Chem. Phys.* **104**, 7099–7112.

S. Gray and J. M. Verosky (1994), Classical Hamiltonian structures in wave packet dynamics, *J. Chem. Phys.* **100**, 5011–5022.

D. F. Griffiths and J. M. Sanz-Serna (1986), On the scope of the method of modified equations, *SIAM J. Sci. Statist. Comput.* **7**, 994–1008.

E. Hairer (1994), Backward analysis of numerical integrators and symplectic methods, *Ann. Numer. Math.* **1**, 107–132.

E. Hairer, C. Lubich and G. Wanner (2003), Geometric numerical integration illustrated by the Störmer–Verlet method, *Acta Numer.* **12**, 399–450.

E. Hairer, C. Lubich and G. Wanner (2006), *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, second edition, Springer.

E. Hairer, S. P. Nørsett and G. Wanner (1993), *Solving Ordinary Differential Equations I: Nonstiff Problems*, second revised edition, Springer.

E. Hansen and A. Ostermann (2009*a*), Exponential splitting for unbounded operators, *Math. Comp.* **78**, 1485–1496.

E. Hansen and A. Ostermann (2009*b*), High order splitting methods for analytic semigroups exist, *BIT* **49**, 527–542.

Y. He, Y. Sun, J. Liu and H. Qin (2015), Volume-preserving algorithms for charged particle dynamics, *J. Comput. Phys.* **281**, 135–147.

Y. He, Y. Sun, J. Liu and H. Qin (2016), Higher order volume-preserving schemes for charged particle dynamics, *J. Comput. Phys.* **305**, 172–184.

M. Hochbruck and C. Lubich (2003), On Magnus integrators for time-dependent Schrödinger equations, *SIAM J. Numer. Anal.* **41**, 945–963.

M. Hoffman (2000), Quasi-shuffle products, *J. Algebraic Combin.* **11**, 49–68.

H. Holden, K. H. Karlsen, K.-A. Lie and N. H. Risebro (2010), *Splitting Methods for Partial Differential Equations with Rough Solutions*, European Mathematical Society.

H. Holden, K. H. Karlsen, N. E. Risebro and T. Tao (2011), Operator splitting for the KdV equation, *Math. Comp.* **80**, 821–846.

H. Holden, C. Lubich and N. E. Risebro (2013), Operator splitting for partial differential equations with Burgers nonlinearity, *Math. Comp.* **82**, 173–185.

W. Hundsdorfer and J. G. Verwer (1995), A note on splitting errors for advection–reaction equations, *Appl. Numer. Math.* **18**, 191–199.

W. Hundsdorfer and J. G. Verwer (2003), *Numerical Solution of Time-Dependent Advection–Diffusion–Reaction Equations*, Springer.

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

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

T. Jahnke and C. Lubich (2000), Error bounds for exponential operator splittings, *BIT* **40**, 735–744.

S. Jin, P. Markowich and C. Sparber (2011), Mathematical and computational methods for semiclassical Schrödinger equations, *Acta Numer.* **20**, 121–209.

W. Kahan and R. C. Li (1997), Composition constants for raising the order of unconventional schemes for ordinary differential equations, *Math. Comp.* **66**, 1089–1099.

E. Kieri (2015), Stiff convergence of force-gradient operator splitting methods, *Appl. Numer. Math.* **94**, 33–45.

P.-V. Koseleff (1993), Calcul formel pour les méthodes de Lie en mécanique hamiltonienne. PhD thesis, École Polytechnique.

D. P. Landau and K. Binder, eds (2005), *A Guide to Monte Carlo Simulations in Statistical Physics*, Cambridge University Press.

F. M. Lasagni (1988), Canonical Runge–Kutta methods, *Z. Angew. Math. Phys.* **39**, 952–953.

J. Laskar (1989), A numerical experiment on the chaotic behaviour of the solar system, *Nature* **338**, 237–238.

J. Laskar and P. Robutel (2001), High order symplectic integrators for perturbed Hamiltonian systems, *Celestial Mech. Dynam. Astronom.* **80**, 39–62.

J. Laskar, P. Robutel, F. Joutel, M. Gastineau, A. C. M. Correia and B. Levrard (2004), A long-term numerical solution for the insolation quantities of the Earth, *Astron. Astrophys.* **428**, 261–285.

L. J. Laslett (1986), Nonlinear dynamics: A personal perspective, in *Nonlinear Dynamics Aspects of Particle Accelerators* (J. M. Jowett, M. Month and S. Turner, eds), Vol. 247 of Lecture Notes in Physics, Springer, pp. 519–580.

C. Lasser and C. Lubich (2020), Computing quantum dynamics in the semiclassical regime, *Acta Numer.* **29**, 229–401.

L. Lehtovaara, J. Toivanen and J. Eloranta (2007), Solution of time-independent Schrödinger equation by the imaginary time propagation method, *J. Comput. Phys.* **221**, 148–157.

B. Leimkuhler and C. Matthews (2013*a*), Rational construction of stochastic numerical methods for molecular sampling, *Appl. Math. Res. Express* **2013**, 34–56.

B. Leimkuhler and C. Matthews (2013*b*), Robust and efficient configurational molecular sampling via Langevin dynamics, *J. Chem. Phys.* **138**, art. 174102.

B. Leimkuhler and C. Matthews (2015), *Molecular Dynamics*, Springer.

B. Leimkuhler and S. Reich (2004), *Simulating Hamiltonian Dynamics*, Cambridge University Press.

B. Leimkuhler, C. Matthews and G. Stolz (2016), The computation of averages from equilibrium and nonequilibrium Langevin molecular dynamics, *IMA J. Numer. Anal.* **36**, 13–79.

B. Leimkuhler, S. Reich and R. D. Skeel (1996), Integration methods for molecular dynamics, in *Mathematical Approaches to Biomolecular Structure and Dynamics* (J. P. Mesirov, K. Schulten and D. W. Sumners, eds), Springer, pp. 161–185.

S. Lie (1888), *Theorie der Transformationsgruppen*, Druck un Verlag.

J. S. Liu (2008), *Monte Carlo Strategies for Scientific Computing*, second edition, Springer.

X. Liu, P. Ding, J. Hong and L. Wang (2005), Optimization of symplectic schemes for time-dependent Schrödinger equations, *Comput. Math. Appl.* **50**, 637–644.

S. Lloyd (1996), Universal quantum simulators, *Science* **273**, 1073–1078.

M. A. López-Marcos, J. M. Sanz-Serna and R. D. Skeel (1996*a*), Cheap enhancement of symplectic integrators, in *Proceedings of the Dundee Conference on Numerical Analysis* (D. F. Griffiths and G. A. Watson, eds), Longman, pp. 107–122.

M. A. López-Marcos, J. M. Sanz-Serna and R. D. Skeel (1996*b*), An explicit symplectic integrator with maximal stability interval, in *Numerical Analysis: A.R. Mitchell 75th Birthday Volume* (D. F. Griffiths and G. A. Watson, eds), World Scientific, pp. 163–176.

M. A. López-Marcos, J. M. Sanz-Serna and R. D. Skeel (1997), Explicit symplectic integrators using Hessian-vector products, *SIAM J. Sci. Comput.* **18**, 223–238.

C. Lubich (2008), *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*, European Mathematical Society.

S. MacNamara and G. Strang (2016), Operator splitting, in *Splitting Methods in Communication, Imaging, Science, and Engineering* (R. Glowinski, S. J. Osher and W. Yin, eds), Springer, pp. 95–114.

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

G. I. Marchuk (1990), Splitting and alternating direction methods, in *Handbook of Numerical Analysis* (P. G. Ciarlet and J. L. Lions, eds), Vol. 1, North-Holland, pp. 197–462.

R. I. McLachlan (1995*a*), Composition methods in the presence of small parameters, *BIT* **35**, 258–268.

R. I. McLachlan (1995*b*), On the numerical integration of ODE's by symmetric composition methods, *SIAM J. Sci. Comput.* **16**, 151–168.

R. I. McLachlan (1996), More on symplectic correctors, in *Integration Algorithms and Classical Mechanics* (J. E. Marsden, G. W. Patrick and W. F. Shadwick, eds), Vol. 10 of Fields Institute Communications, American Mathematical Society, pp. 141–149.

R. I. McLachlan (2002), Families of high-order composition methods, *Numer. Algorithms* **31**, 233–246.

R. I. McLachlan and P. Atela (1992), The accuracy of symplectic integrators, *Nonlinearity* **5**, 541–562.

R. I. McLachlan and S. K. Gray (1997), Optimal stability polynomials for splitting methods, with applications to the time-dependent Schrödinger equation, *Appl. Numer. Math.* **25**, 275–286.

R. I. McLachlan and A. Murua (2019), The Lie algebra of classical mechanics, *J. Comput. Dyn.* **6**, 198–213.

R. I. McLachlan and R. Quispel (2002), Splitting methods, *Acta Numer.* **11**, 341–434.

R. I. McLachlan and R. Quispel (2006), Geometric integrators for ODEs, *J. Phys. A* **39**, 5251–5285.

A. Messiah (1999), *Quantum Mechanics*, Dover.

A. Miessen, P. J. Ollitrault, F. Tacchino and I. Tavernelli (2023), Quantum algorithms for quantum dynamics, *Nat. Comput. Sci.* **3**, 25–37.

S. Mikkola (1997), Practical symplectic methods with time transformation for the few-body problem, *Celestial Mech. Dynam. Astronom.* **67**, 145–165.

P. C. Moan (2002), On backward error analysis and Nekhoroshev stability in the numerical analysis of conservative systems of ODEs. PhD thesis, University of Cambridge.

P. C. Moan (2004), On the KAM and Nekhorosev theorems for symplectic integrators and implications for error growth, *Nonlinearity* **17**, 67–83.

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

A. Murua (2006), The Hopf algebra of rooted trees, free Lie algebras, and Lie series, *Found. Comput. Math.* **6**, 387–426.

A. Murua and J. M. Sanz-Serna (1999), Order conditions for numerical integrators obtained by composing simpler integrators, *Philos. Trans. Roy. Soc. A* **357**, 1079–1100.

A. Murua and J. M. Sanz-Serna (2016), Computing normal forms and formal invariants of dynamical systems by means of word series, *Nonlinear Anal.* **138**, 326–345.

A. Murua and J. M. Sanz-Serna (2017), Word series for dynamical systems and their numerical integrators, *Found. Comput. Math.* **17**, 675–712.

R. M. Neal (2011), MCMC using Hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo* (S. Brooks *et al.*, eds), CRC Press, pp. 113–174.

F. Neri (1988), Lie algebras and canonical integration. Technical report, University of Maryland.

M. A. Nielsen and I. L. Chuang (2010), *Quantum Computation and Quantum Information*, 10th anniversary edition, Cambridge University Press.

D. I. Okunbor and E. J. Lu (1994), Eighth-order explicit symplectic Runge–Kutta–Nyström integrators. Technical report CSC 94-21, University of Missouri-Rolla.

D. I. Okunbor and R. D. Skeel (1994), Canonical Runge–Kutta–Nyström methods of orders five and six, *J. Comput. Appl. Math.* **51**, 375–382.

I. P. Omelyan, I. M. Mryglod and R. Folk (2002), On the construction of high order force gradient algorithms for integration of motion in classical and quantum systems, *Phys. Rev. E* **66**, art. 026701.

I. P. Omelyan, I. M. Mryglod and R. Folk (2003), Symplectic analytically integrable decomposition algorithms: Classification, derivation, and application to molecular dynamics, quantum and celestial mechanics simulations, *Comput. Phys. Commun.* **151**, 272–314.

B. Owren and A. Marthinsen (1999), Runge–Kutta methods adapted to manifolds and based on rigid frames, *BIT* **39**, 116–142.

P. Painlevé (1900), *Analyse des Travaux Scientifiques*, Gauthier-Villars.

L. A. Pars (1979), *A Treatise on Analytical Dynamics*, Ox Bow Press.

J. Partington (2004), *Linear Operators and Linear Systems: An Analytical Approach to Control Theory*, Cambridge University Press.

A. Pazy (1983), *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer.

D. W. Peaceman and H. H. Rachford (1955), The numerical solution of parabolic and elliptic differential equations, *J. Soc. Indust. Appl. Math.* **3**, 28–41.

T. Prosen and I. Pizorn (2006), High order non-unitary split-step decomposition of unitary operators, *J. Phys. A* **39**, 5957–5964.

M. Reed and B. Simon (1980), *Methods of Modern Mathematical Physics I: Functional Analysis*, Academic Press.

S. Reich (1999), Backward error analysis for numerical integrators, *SIAM J. Numer. Anal.* **36**, 1549–1570.

C. Reutenauer (1993), *Free Lie Algebras*, Oxford University Press.

F. Rousset and K. Schratz (2022), Convergence error estimates at low regularity for time discretizations of KdV, *Pure Appl. Anal.* **4**, 127–152.

G. Rowlands (1991), A numerical algorithm for Hamiltonian systems, *J. Comput. Phys.* **97**, 235–239.

R. Ruth (1983), A canonical integration technique, *IEEE Trans. Nucl. Sci.* **30**, 26–69.

J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen (1977), Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes, *J. Comput. Phys.* **23**, 327–341.

K. Sakkos, J. Casulleras and J. Boronat (2009), High order Chin actions in path integral Monte Carlo, *J. Chem. Phys.* **130**, art. 204109.

J. M. Sanz-Serna (1988), Runge–Kutta schemes for Hamiltonian systems, *BIT* **28**, 877–883.

J. M. Sanz-Serna (1992), Symplectic integrators for Hamiltonian problems: An overview, *Acta Numer.* **1**, 243–286.

J. M. Sanz-Serna (1997), Geometric integration, in *The State of the Art in Numerical Analysis* (I. S. Duff and G. A. Watson, eds), Vol. 63 of Institute of Mathematics and its Applications Conference Series: New Ser., Oxford University Press, pp. 121–143.

J. M. Sanz-Serna and M. P. Calvo (1994), *Numerical Hamiltonian Problems*, Chapman & Hall.

J. M. Sanz-Serna and A. Portillo (1996), Classical numerical integrators for wave-packet dynamics, *J. Chem. Phys.* **104**, 2349–2355.

T. Schlick (2010), *Molecular Modelling and Simulation: An Interdisciplinary Guide*, second edition, Springer.

Q. Sheng (1989), Solving linear partial differential equations by exponential splitting, *IMA J. Numer. Anal.* **9**, 199–212.

A. Sinitsyn, E. Dulov and V. Vedenyapin (2011), *Kinetic Boltzmann, Vlasov and Related Equations*, Elsevier.

R. D. Skeel and J. L. Cieśliński (2020), On the famous unpublished preprint 'methods of integration which preserve the contact transformation property of the Hamilton equations'. Available at arXiv:2003.12268.

M. Sofroniou and G. Spaletta (2005), Derivation of symmetric composition constants for symmetric integrators, *Optim. Methods Softw.* **20**, 597–613.

R. Speth, W. Green, S. MacNamara and G. Strang (2013), Balanced splitting and rebalanced splitting, *SIAM J. Numer. Anal.* **51**, 3084–3105.

H. J. Stetter (1970), Symmetric two-step algorithms for ordinary differential equations, *Computing* **5**, 267–280.

C. Störmer (1907), Sur les trajectoires des corpuscules électrisés, *Arch. Sci. Phys. Nat. Genève* **24**, 5–18, 113–158, 221–247.

G. Strang (1968), On the construction and comparison of difference schemes, *SIAM J. Numer. Anal.* **5**, 506–517.

C. Sulem and P.-L. Sulem (1999), *The Nonlinear Schrödinger Equation: Self-Focusing and Wave Collapse*, Springer.

Y. B. Suris (1988), Preservation of symplectic structure in the numerical solution of Hamiltonian systems, in *Numerical Solution of Differential Equations* (S. S. Filippov, ed.), pp. 148–160. In Russian.

G. Sussman and J. Wisdom (1992), Chaotic evolution of the Solar System, *Science* **257**, 56–62.

M. Suzuki (1990), Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations, *Phys. Lett. A* **146**, 319–323.

M. Suzuki (1991), General theory of fractal path integrals with applications to many-body theories and statistical physics, *J. Math. Phys.* **32**, 400–407.

M. Suzuki (1992), General nonsymmetric higher-order decomposition of exponential operators and symplectic integrators, *J. Phys. Soc. Japan* **61**, 3015–3019.

M. Suzuki (1995), Hybrid exponential product formulas for unbounded operators with possible applications to Monte Carlo simulations, *Phys. Lett. A* **201**, 425–428.

M. Suzuki and K. Umeno (1993), High-order decomposition theory of exponential operators and its applications to QMC and nonlinear dynamics, in *Computer Simulation Studies in Condensed Matter Physics VI* (D. P. Landau, K. K. Mon and H. B. Schüttler, eds), Springer, pp. 74–86.

M. Takahashi and M. Imada (1984), Monte carlo calculation of quantum system II. Higher order correction, *J. Phys. Soc. Japan* **53**, 3765–3769.

M. Thalhammer (2008), High-order exponential operator splitting methods for time-dependent Schrödinger equations, *SIAM J. Numer. Anal.* **46**, 2022–2038.

M. Thalhammer, M. Caliari and C. Neuhauser (2009), High-order time-splitting Hermite and Fourier spectral methods, *J. Comput. Phys.* **228**, 822–832.

L. N. Trefethen (2000), *Spectral Methods in MATLAB*, SIAM.

H. F. Trotter (1959), On the product of semi-groups of operators, *Proc. Amer. Math. Soc.* **10**, 545–551.

C. Tsitouras (1999), A tenth order symplectic Runge–Kutta–Nyström method, *Celestial Mech. Dynam. Astronom.* **74**, 223–230.

V. S. Varadarajan (1984), *Lie Groups, Lie Algebras, and Their Representations*, Springer.

L. Verlet (1967), Computer 'experiments' on classical fluids I: Thermodynamical properties of Lennard-Jones molecules, *Phys. Rev.* **159**, 98–103.

A. A. Vlasov (1961), *Many-Particle Theory and its Application to Plasma*, Gordon & Breach.

J. Wisdom and M. Holman (1991), Symplectic maps for the n-body problem, *Astron. J.* **102**, 1528–1538.

J. Wisdom, M. Holman and J. Touma (1996), Symplectic correctors, in *Integration Algorithms and Classical Mechanics* (J. E. Marsden, G. W. Patrick and W. F. Shadwick, eds), Vol. 10 of Fields Institute Communications, American Mathematical Society, pp. 217–244.

N. N. Yanenko (1971), *The Method of Fractional Steps*, Springer.

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

K. Yosida (1971), *Functional Analysis*, third edition, Springer.

R. Zhang, J. Liu, H. Qin, Y. Wang, Y. He and Y. Sun (2015), Volume-preserving algorithm for secular relativistic dynamics of charged particles, *Phys. Plasmas* **22**, art. 044501.

W. Zhu, X. Zhao and Y. Tang (1996), Numerical methods with a high order of accuracy applied in the quantum system, *J. Chem. Phys.* **104**, 2275–2286.