Chapter II.
Numerical Integrators

After having seen in Chap. I some simple numerical methods and a variety of numerical phenomena that they exhibited, we now present more elaborate classes of numerical methods. We start with Runge–Kutta and collocation methods, and we introduce discontinuous collocation methods, which cover essentially all high-order implicit Runge–Kutta methods of interest. We then treat partitioned Runge–Kutta methods and Nyström methods, which can be applied to partitioned problems such as Hamiltonian systems. Finally we present composition and splitting methods.

II.1 Runge–Kutta and Collocation Methods

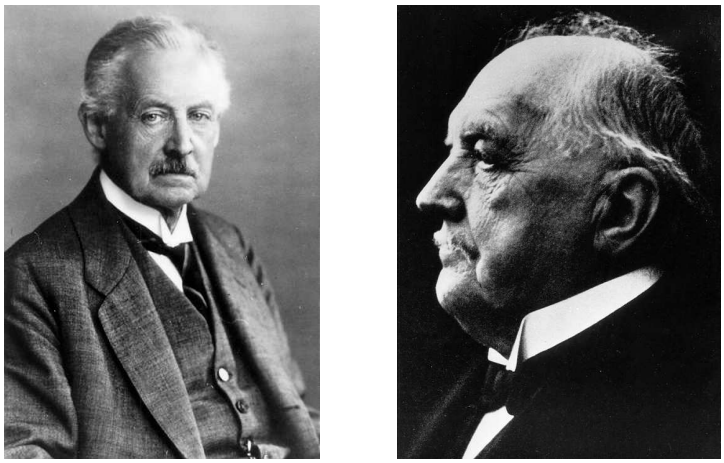


Fig. 1.1. Carl David Tolmé Runge (left picture), born: 30 August 1856 in Bremen (Germany), died: 3 January 1927 in Göttingen (Germany).
Wilhelm Martin Kutta (right picture), born: 3 November 1867 in Pitschen, Upper Silesia (now Byczyna, Poland), died: 25 December 1944 in Fürstenfeldbruck (Germany).

Runge–Kutta methods form an important class of methods for the integration of differential equations. A special subclass, the collocation methods, allows for a particularly elegant access to order, symplecticity and continuous output.

II.1.1 Runge–Kutta Methods

In this section, we treat non-autonomous systems of first-order ordinary differential equations

$$\dot{y} = f(t, y), \qquad y(t_0) = y_0. \tag{1.1}$$

The integration of this equation gives $y(t_1) = y_0 + \int_{t_0}^{t_1} f(t, y(t))\, dt$, and replacing the integral by the trapezoidal rule, we obtain

$$y_1 = y_0 + \frac{h}{2}\big(f(t_0, y_0) + f(t_1, y_1)\big). \tag{1.2}$$

This is the *implicit trapezoidal rule*, which, in addition to its historical importance for computations in partial differential equations (Crank–Nicolson) and in A-stability theory (Dahlquist), played a crucial role even earlier in the discovery of Runge–Kutta methods. It was the starting point of Runge (1895), who "predicted" the unknown y_1-value to the right by an Euler step, and obtained the first of the following formulas (the second being the analogous formula for the midpoint rule)

$$\begin{aligned} k_1 &= f(t_0, y_0) & k_1 &= f(t_0, y_0) \\ k_2 &= f(t_0 + h, y_0 + hk_1) & k_2 &= f(t_0 + \tfrac{h}{2}, y_0 + \tfrac{h}{2}k_1) \\ y_1 &= y_0 + \tfrac{h}{2}\big(k_1 + k_2\big) & y_1 &= y_0 + hk_2. \end{aligned} \tag{1.3}$$

These methods have a nice geometric interpretation (which is illustrated in the first two pictures of Fig. 1.2 for a famous problem, the Riccati equation): they consist of polygonal lines, which assume the slopes prescribed by the differential equation evaluated at previous points.

Idea of Heun (1900) and Kutta (1901): compute *several* polygonal lines, each starting at y_0 and assuming the various slopes k_j on portions of the integration interval, which are proportional to some given constants a_{ij}; at the final point of each polygon evaluate a new slope k_i. The last of these polygons, with constants b_i, determines the numerical solution y_1 (see the third picture of Fig. 1.2). This idea leads to the class of *explicit* Runge–Kutta methods, i.e., formula (1.4) below with $a_{ij} = 0$ for $i \le j$.

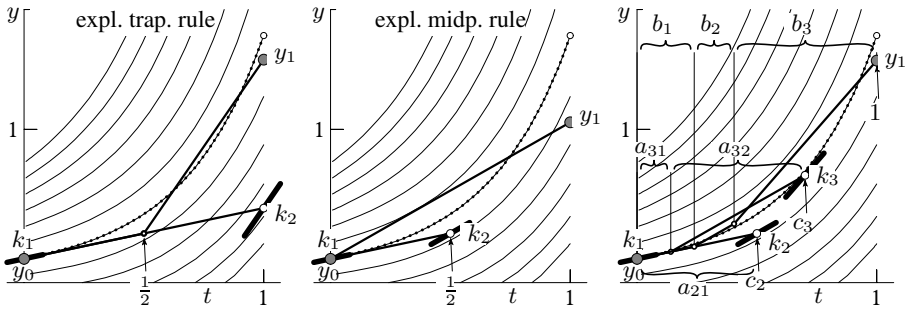


Fig. 1.2. Runge–Kutta methods for $\dot{y} = t^2 + y^2$, $y_0 = 0.46$, $h = 1$; dotted: exact solution.

Much more important for our purpose are *implicit* Runge–Kutta methods, introduced mainly in the work of Butcher (1963).

Definition 1.1. Let b_i, a_{ij} $(i, j = 1, \dots, s)$ be real numbers and let $c_i = \sum_{j=1}^{s} a_{ij}$. An *$s$-stage Runge–Kutta method* is given by

$$\begin{aligned} k_i &= f\Big(t_0 + c_i h, y_0 + h \sum_{j=1}^{s} a_{ij} k_j\Big), \quad i = 1, \dots, s \\ y_1 &= y_0 + h \sum_{i=1}^{s} b_i k_i. \end{aligned} \tag{1.4}$$

Here we allow a full matrix (a_{ij}) of non-zero coefficients. In this case, the slopes k_i can no longer be computed explicitly, and even do not necessarily exist. For example, for the problem set-up of Fig. 1.2 the implicit trapezoidal rule has no solution. However, the implicit function theorem assures that, for sufficiently small h, the nonlinear system (1.4) for the values $k_1, \dots, k_s$ has a locally unique solution close to $k_i \approx f(t_0, y_0)$.

Since Butcher's work, the coefficients are usually displayed as follows:

$$\begin{array}{c|ccc} c_1 & a_{11} & \dots & a_{1s} \\ \vdots & \vdots & & \vdots \\ c_s & a_{s1} & \dots & a_{ss} \\ \hline & b_1 & \dots & b_s \end{array} . \tag{1.5}$$

Definition 1.2. A Runge–Kutta method (or a general one-step method) has *order p*, if for all sufficiently regular problems (1.1) the *local error* $y_1 - y(t_0 + h)$ satisfies

$$y_1 - y(t_0 + h) = \mathcal{O}(h^{p+1}) \quad \text{as} \quad h \to 0.$$

To check the order of a Runge Kutta method, one has to compute the Taylor series expansions of $y(t_0 + h)$ and y_1 around to $h = 0$. This leads to the following algebraic conditions for the coefficients for orders 1, 2, and 3:

$$\begin{aligned} & \textstyle\sum_i b_i = 1 && \text{for order 1;} \\ \text{in addition} \quad & \textstyle\sum_i b_i c_i = 1/2 && \text{for order 2;} \\ \text{in addition} \quad & \textstyle\sum_i b_i c_i^2 = 1/3 && \\ \text{and} \quad & \textstyle\sum_{i,j} b_i a_{ij} c_j = 1/6 && \text{for order 3.} \end{aligned} \tag{1.6}$$

For higher orders, however, this problem represented a great challenge in the first half of the 20th century. We shall present an elegant theory in Sect. III.1 which allows order conditions to be derived.

Among the methods seen up to now, the explicit and implicit Euler methods

$$\begin{array}{c|c} 0 & \\ \hline & 1 \end{array} \qquad\qquad \begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array} \tag{1.7}$$

are of order 1, the implicit trapezoidal and midpoint rules as well as both methods of Runge

$$\begin{array}{c|cc} 0 & & \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array} \qquad \begin{array}{c|c} 1/2 & 1/2 \\ \hline & 1 \end{array} \qquad \begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & 1/2 & 1/2 \end{array} \qquad \begin{array}{c|cc} 0 & & \\ 1/2 & 1/2 & \\ \hline & 0 & 1 \end{array}$$

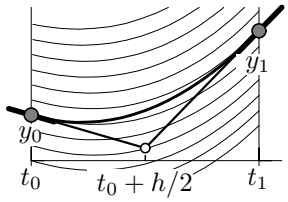
are of order 2. The most successful methods during more than half a century were the 4th order methods of Kutta:

$$\begin{array}{c|cccc} 0 & & & & \\ 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ 1 & 0 & 0 & 1 & \\ \hline & 1/6 & 2/6 & 2/6 & 1/6 \end{array} \qquad \begin{array}{c|cccc} 0 & & & & \\ 1/3 & 1/3 & & & \\ 2/3 & -1/3 & 1 & & \\ 1 & 1 & -1 & 1 & \\ \hline & 1/8 & 3/8 & 3/8 & 1/8 \end{array} \tag{1.8}$$

II.1.2 Collocation Methods

> The high speed computing machines make it possible to enjoy the advantages of intricate methods. (P.C. Hammer & J.W. Hollingsworth 1955)

Collocation methods for ordinary differential equations have their origin, once again, in the implicit trapezoidal rule (1.2): Hammer & Hollingsworth (1955) discovered that this method can be interpreted as being generated by a *quadratic function* "which agrees in direction with that indicated by the differential equation at two points" t_0 and t_1 (see the picture to the right). This idea allows one to "see much-used methods in a new light" and allows various generalizations (Guillou & Soulé (1969), Wright (1970)). An interesting feature of collocation methods is that we not only get a discrete set of approximations, but also a *continuous approximation* to the solution.



Definition 1.3. Let $c_1, \dots, c_s$ be distinct real numbers (usually $0 \le c_i \le 1$). The *collocation polynomial* $u(t)$ is a polynomial of degree s satisfying

$$\begin{aligned} u(t_0) &= y_0 \\ \dot{u}(t_0 + c_i h) &= f\big(t_0 + c_i h, u(t_0 + c_i h)\big), \quad i = 1, \dots, s, \end{aligned} \tag{1.9}$$

and the numerical solution of the *collocation method* is defined by $y_1 = u(t_0 + h)$.

For $s = 1$, the polynomial has to be of the form $u(t) = y_0 + (t - t_0)k$ with

$$k = f(t_0 + c_1 h, y_0 + h c_1 k).$$

We see that the explicit and implicit Euler methods and the midpoint rule are collocation methods with $c_1 = 0$, $c_1 = 1$ and $c_1 = 1/2$, respectively.

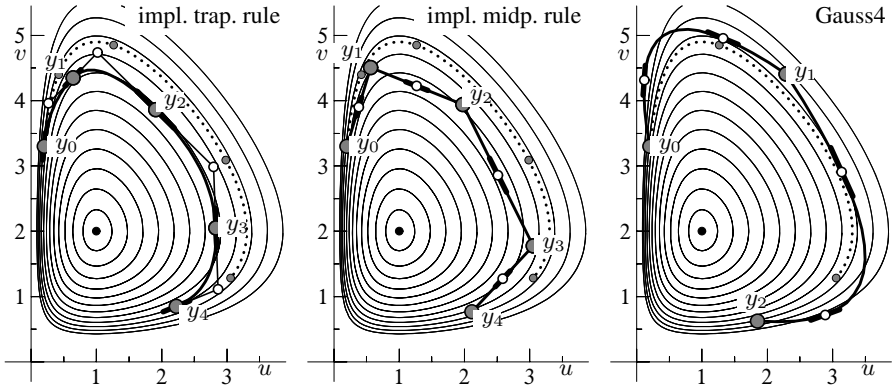
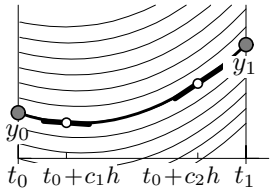


Fig. 1.3. Collocation solutions for the Lotka–Volterra problem (I.1.1); $u_0 = 0.2$, $v_0 = 3.3$; methods of order 2: four steps with $h = 0.4$; method of order 4: two steps with $h = 0.8$; dotted: exact solution.

For $s = 2$ and $c_1 = 0, c_2 = 1$ we find, of course, the implicit trapezoidal rule. The choice of Hammer & Hollingsworth for the collocation points is $c_{1,2} = 1/2 \pm \sqrt{3}/6$, the *Gaussian quadrature nodes* (see the picture to the right). We will see that the corresponding method is of order 4.



In Fig. 1.3 we illustrate the collocation idea with these methods for the Lotka–Volterra problem (I.1.1). One can observe that, in spite of the extremely large step sizes, the methods are quite satisfactory.

Theorem 1.4 (Guillou & Soulé 1969, Wright 1970). *The collocation method of Definition 1.3 is equivalent to the s-stage Runge–Kutta method (1.4) with coefficients*

$$a_{ij} = \int_0^{c_i} \ell_j(\tau)\, d\tau, \qquad b_i = \int_0^1 \ell_i(\tau)\, d\tau, \tag{1.10}$$

where $\ell_i(\tau)$ is the Lagrange polynomial $\ell_i(\tau) = \prod_{l \neq i} (\tau - c_l)/(c_i - c_l)$.

Proof. Let $u(t)$ be the collocation polynomial and define

$$k_i := \dot{u}(t_0 + c_i h).$$

By the Lagrange interpolation formula we have $\dot{u}(t_0 + \tau h) = \sum_{j=1}^{s} k_j \cdot \ell_j(\tau)$, and by integration we get

$$u(t_0 + c_i h) = y_0 + h \sum_{j=1}^{s} k_j \int_0^{c_i} \ell_j(\tau)\, d\tau.$$

Inserted into (1.9) this gives the first formula of the Runge–Kutta equation (1.4). Integration from 0 to 1 yields the second one. □

The above proof can also be read in reverse order. This shows that a Runge–Kutta method with coefficients given by (1.10) can be interpreted as a collocation method. Since $\tau^{k-1} = \sum_{j=1}^{s} c_j^{k-1} \ell_j(\tau)$ for $k = 1, \dots, s$, the relations (1.10) are equivalent to the linear systems

$$\begin{aligned} C(q): &\qquad \sum_{j=1}^{s} a_{ij} c_j^{k-1} = \frac{c_i^k}{k}, \quad k = 1, \dots, q, \ \text{ all } i \\ B(p): &\qquad \sum_{i=1}^{s} b_i c_i^{k-1} = \frac{1}{k}, \quad k = 1, \dots, p, \end{aligned} \tag{1.11}$$

with $q = s$ and $p = s$. What is the order of a Runge–Kutta method whose coefficients b_i, a_{ij} are determined in this way?

Compared to the enormous difficulties that the first explorers had in constructing Runge–Kutta methods of orders 5 and 6, and also compared to the difficult algebraic proofs of the first papers of Butcher, the following general theorem and its proof, discovered in this form by Guillou & Soulé (1969), are surprisingly simple.

Theorem 1.5 (Superconvergence). *If the condition $B(p)$ holds for some $p \geq s$, then the collocation method (Definition 1.3) has order p. This means that the collocation method has the same order as the underlying quadrature formula.*

Proof. We consider the collocation polynomial $u(t)$ as the solution of a perturbed differential equation

$$\dot{u} = f(t, u) + \delta(t) \tag{1.12}$$

with defect $\delta(t) := \dot{u}(t) - f\big(t, u(t)\big)$. Subtracting (1.1) from (1.12) we get after linearization that

$$\dot{u}(t) - \dot{y}(t) = \frac{\partial f}{\partial y}\Big(t, y(t)\Big)\Big(u(t) - y(t)\Big) + \delta(t) + r(t), \tag{1.13}$$

where, for $t_0 \leq t \leq t_0 + h$, the remainder $r(t)$ is of size $\mathcal{O}\big(\|u(t) - y(t)\|^2\big) = \mathcal{O}(h^{2s+2})$ by Lemma 1.6 below. The variation of constants formula (see e.g., Hairer, Nørsett & Wanner (1993), p. 66) then yields

$$y_1 - y(t_0+h) = u(t_0+h) - y(t_0+h) = \int_{t_0}^{t_0+h} R(t_0+h, s)\Big(\delta(s) + r(s)\Big)\, ds, \tag{1.14}$$

where $R(t, s)$ is the resolvent of the homogeneous part of the differential equation (1.13), i.e., the solution of the matrix differential equation $\partial R(t,s)/\partial t = A(t)R(t,s)$, $R(s,s) = I$, with $A(t) = \partial f/\partial y(t, y(t))$. The integral over $R(t_0 + h, s)r(s)$ gives a $\mathcal{O}(h^{2s+3})$ contribution. The main idea now is to apply the quadrature formula $(b_i, c_i)_{i=1}^{s}$ to the integral over $g(s) = R(t_0 + h, s)\delta(s)$; because the defect $\delta(s)$ vanishes at the collocation points $t_0 + c_i h$ for $i = 1, \dots, s$, this gives zero as the numerical result. Thus, the integral is equal to the quadrature error, which is bounded by h^{p+1} times a bound of the pth derivative of the function $g(s)$. This derivative is bounded independently of h, because by Lemma 1.6 all derivatives of the collocation polynomial are bounded uniformly as $h \to 0$. Since, anyway, $p \leq 2s$, we get $y_1 - y(t_0 + h) = \mathcal{O}(h^{p+1})$ from (1.14). □

Lemma 1.6. *The collocation polynomial $u(t)$ is an approximation of order s to the exact solution of (1.1) on the whole interval, i.e.,*

$$\|u(t) - y(t)\| \le C \cdot h^{s+1} \qquad \textit{for} \quad t \in [t_0, t_0 + h] \tag{1.15}$$

and for sufficiently small h.

Moreover, the derivatives of $u(t)$ satisfy for $t \in [t_0, t_0 + h]$

$$\|u^{(k)}(t) - y^{(k)}(t)\| \le C \cdot h^{s+1-k} \qquad \textit{for} \quad k = 0, \ldots, s.$$

Proof. The collocation polynomial satisfies

$$\dot{u}(t_0 + \tau h) = \sum_{i=1}^{s} f\Big(t_0 + c_i h, u(t_0 + c_i h)\Big)\, \ell_i(\tau),$$

while the exact solution of (1.1) satisfies

$$\dot{y}(t_0 + \tau h) = \sum_{i=1}^{s} f\Big(t_0 + c_i h, y(t_0 + c_i h)\Big)\, \ell_i(\tau) + h^s E(\tau, h),$$

where the interpolation error $E(\tau, h)$ is bounded by $\max_{t\in[t_0,t_0+h]} \|y^{(s+1)}(t)\|/s!$ and its derivatives satisfy

$$\|E^{(k-1)}(\tau, h)\| \;\le\; \max_{t\in[t_0,t_0+h]} \frac{\|y^{(s+1)}(t)\|}{(s-k+1)!}\,.$$

This follows from the fact that, by Rolle's theorem, the differentiated polynomial $\sum_{i=1}^{s} f\big(t_0 + c_i h, y(t_0 + c_i h)\big)\, \ell_i^{(k-1)}(\tau)$ can be interpreted as the interpolation polynomial of $h^{k-1} y^{(k)}(t_0 + \tau h)$ at $s - k + 1$ points lying in $[t_0, t_0 + h]$. Integrating the difference of the above two equations gives

$$y(t_0 + \tau h) - u(t_0 + \tau h) = h \sum_{i=1}^{s} \Delta f_i \int_0^\tau \ell_i(\sigma)\, d\sigma + h^{s+1} \int_0^\tau E(\sigma, h)\, d\sigma \tag{1.16}$$

with $\Delta f_i = f\big(t_0 + c_i h, y(t_0 + c_i h)\big) - f\big(t_0 + c_i h, u(t_0 + c_i h)\big)$. Using a Lipschitz condition for $f(t, y)$, this relation yields

$$\max_{t\in[t_0,t_0+h]} \|y(t) - u(t)\| \le h\, C\, L \max_{t\in[t_0,t_0+h]} \|y(t) - u(t)\| + \mathit{Const} \cdot h^{s+1},$$

implying the statement (1.15) for sufficiently small $h > 0$.

The proof of the second statement follows from

$$h^k \Big(y^{(k)}(t_0 + \tau h) - u^{(k)}(t_0 + \tau h)\Big) = h \sum_{i=1}^{s} \Delta f_i\, \ell_i^{(k-1)}(\tau) + h^{s+1}\, E^{(k-1)}(\tau, h)$$

by using a Lipschitz condition for $f(t, y)$ and the estimate (1.15). □

II.1.3 Gauss and Lobatto Collocation

Gauss Methods. If we take $c_1, \dots, c_s$ as the zeros of the sth shifted Legendre polynomial

$$\frac{d^s}{dx^s}\Big(x^s(x-1)^s\Big),$$

the interpolatory quadrature formula has order $p = 2s$, and by Theorem 1.5, the Runge–Kutta (or collocation) method based on these nodes has the same order $2s$. For $s = 1$ we obtain the implicit midpoint rule. The Runge–Kutta coefficients for $s = 2$ (the method of Hammer & Hollingsworth 1955) and $s = 3$ are given in Table 1.1. The proof of the order properties for general s was a sensational result of Butcher (1964a). At that time these methods were considered, at least by the editors of *Math. of Comput.*, to be purely academic without any practical value; 5 years later their A-stability was discovered, 12 years later their B-stability, and 25 years later their symplecticity. Thus, of all the papers in issue No. 85 of *Math. of Comput.*, the one most important to us is the one for which publication was the most difficult.

Table 1.1. Gauss methods of order 4 and 6

$$\begin{array}{c|cc}
\frac{1}{2}-\frac{\sqrt{3}}{6} & \frac{1}{4} & \frac{1}{4}-\frac{\sqrt{3}}{6} \\
\frac{1}{2}+\frac{\sqrt{3}}{6} & \frac{1}{4}+\frac{\sqrt{3}}{6} & \frac{1}{4} \\
\hline
 & \frac{1}{2} & \frac{1}{2}
\end{array}$$

$$\begin{array}{c|ccc}
\frac{1}{2}-\frac{\sqrt{15}}{10} & \frac{5}{36} & \frac{2}{9}-\frac{\sqrt{15}}{15} & \frac{5}{36}-\frac{\sqrt{15}}{30} \\
\frac{1}{2} & \frac{5}{36}+\frac{\sqrt{15}}{24} & \frac{2}{9} & \frac{5}{36}-\frac{\sqrt{15}}{24} \\
\frac{1}{2}+\frac{\sqrt{15}}{10} & \frac{5}{36}+\frac{\sqrt{15}}{30} & \frac{2}{9}+\frac{\sqrt{15}}{15} & \frac{5}{36} \\
\hline
 & \frac{5}{18} & \frac{4}{9} & \frac{5}{18}
\end{array}$$

Radau Methods. Radau quadrature formulas have the highest possible order, $2s-1$, among quadrature formulas with either $c_1 = 0$ or $c_s = 1$. The corresponding collocation methods for $c_s = 1$ are called Radau IIA methods. They play an important role in the integration of stiff differential equations (see Hairer & Wanner (1996), Sect. IV.8). However, they lack both *symmetry* and *symplecticity*, properties that will be the subjects of later chapters in this book.

Lobatto IIIA Methods. Lobatto quadrature formulas have the highest possible order with $c_1 = 0$ and $c_s = 1$. Under these conditions, the nodes must be the zeros of

$$\frac{d^{s-2}}{dx^{s-2}}\Big(x^{s-1}(x-1)^{s-1}\Big) \tag{1.17}$$

and the quadrature order is $p = 2s-2$. The corresponding collocation methods are called, for historical reasons, Lobatto IIIA methods. For $s=2$ we have the implicit trapezoidal rule. The coefficients for $s=3$ and $s=4$ are given in Table 1.2.

Table 1.2. Lobatto IIIA methods of order 4 and 6

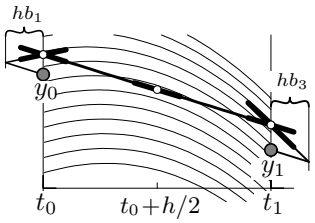
0	0	0	0
$\frac{1}{2}$	$\frac{5}{24}$	$\frac{1}{3}$	$-\frac{1}{24}$
1	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$
	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{1}{6}$

0	0	0	0	0
$\frac{5-\sqrt{5}}{10}$	$\frac{11+\sqrt{5}}{120}$	$\frac{25-\sqrt{5}}{120}$	$\frac{25-13\sqrt{5}}{120}$	$\frac{-1+\sqrt{5}}{120}$
$\frac{5+\sqrt{5}}{10}$	$\frac{11-\sqrt{5}}{120}$	$\frac{25+13\sqrt{5}}{120}$	$\frac{25+\sqrt{5}}{120}$	$\frac{-1-\sqrt{5}}{120}$
1	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$
	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{5}{12}$	$\frac{1}{12}$

II.1.4 Discontinuous Collocation Methods

Collocation methods allow, as we have seen above, a very elegant proof of their order properties. By similar ideas, they also admit strikingly simple proofs for their A- and B-stability as well as for symplecticity, our subject in Chap. VI. However, not all method classes are of collocation type. It is therefore interesting to define a modification of the collocation idea, which allows us to extend all the above proofs to much wider classes of methods. This definition will also lead, later, to important classes of *partitioned* methods.

Definition 1.7. Let $c_2, \dots, c_{s-1}$ be distinct real numbers (usually $0 \le c_i \le 1$), and let b_1, b_s be two arbitrary real numbers. The corresponding *discontinuous collocation method* is then defined via a polynomial of degree $s-2$ satisfying



$$\begin{aligned} u(t_0) &= y_0 - hb_1\big(\dot u(t_0) - f(t_0,u(t_0))\big) \\ \dot u(t_0+c_ih) &= f\big(t_0+c_ih, u(t_0+c_ih)\big), \quad i=2,\dots,s-1, \\ y_1 &= u(t_1) - hb_s\big(\dot u(t_1) - f(t_1,u(t_1))\big). \end{aligned} \tag{1.18}$$

The figure gives a geometric interpretation of the correction term in the first and third formulas of (1.18). The motivation for this definition will become clear in the proof of Theorem 1.9 below. Our first result shows that discontinuous collocation methods are equivalent to implicit Runge–Kutta methods.

Theorem 1.8. *The discontinuous collocation method of Definition 1.7 is equivalent to an s-stage Runge–Kutta method (1.4) with coefficients determined by $c_1 = 0$, $c_s = 1$, and*

$$\begin{aligned} &a_{i1} = b_1, \qquad a_{is} = 0 \qquad \textit{for } i = 1, \dots, s, \\ &C(s-2) \qquad \textit{and} \qquad B(s-2), \end{aligned} \tag{1.19}$$

with the conditions $C(q)$ and $B(p)$ of (1.11).

Proof. As in the proof of Theorem 1.4 we put $k_i := \dot{u}(t_0 + c_i h)$ (this time for $i = 2, \dots, s-1$), so that $\dot{u}(t_0+\tau h) = \sum_{j=2}^{s-1} k_j \cdot \ell_j(\tau)$ by the Lagrange interpolation formula. Here, $\ell_j(\tau)$ corresponds to $c_2, \dots, c_{s-1}$ and is a polynomial of degree $s-3$. By integration and using the definition of $u(t_0)$ we get

$$\begin{aligned} u(t_0 + c_i h) &= u(t_0) + h \sum_{j=2}^{s-1} k_j \int_0^{c_i} \ell_j(\tau)\, d\tau \\ &= y_0 + h b_1 k_1 + h \sum_{j=2}^{s-1} k_j \Big(\int_0^{c_i} \ell_j(\tau)\, d\tau - b_1 \ell_j(0) \Big) \end{aligned}$$

with $k_1 = f(y_0)$. Inserted into (1.18) this gives the first formula of the Runge–Kutta equation (1.4) with $a_{ij} = \int_0^{c_i} \ell_j(\tau)\, d\tau - b_1 \ell_j(0)$. As for collocation methods, one checks that the a_{ij} are uniquely determined by the condition $C(s-2)$. The formula for y_1 is obtained similarly. □

Table 1.3. Survey of discontinuous collocation methods

type	characteristics	prominent examples
$b_1 = 0,\ b_s = 0$	$(s-2)$-stage collocation	Gauss, Radau IIA, Lobatto IIIA
$b_1 = 0,\ b_s \neq 0$	$(s-1)$-stage with $a_{is} = 0$	methods of Butcher (1964b)
$b_1 \neq 0,\ b_s = 0$	$(s-1)$-stage with $a_{i1} = b_1$	Radau IA, Lobatto IIIC
$b_1 \neq 0,\ b_s \neq 0$	s-stage with $a_{i1} = b_1,\ a_{is} = 0$	Lobatto IIIB

If $b_1 = 0$ in Definition 1.7, the entire first column in the Runge–Kutta tableau vanishes, so that the first stage can be removed, which leads to an equivalent method with $s-1$ stages. Similarly, if $b_s = 0$, we can remove the last stage. Therefore, we have all classes of methods, which are "continuous" either to the left, or to the right, or on both sides, as special cases in our definition.

In the case where $b_1 = b_s = 0$, the discontinuous collocation method (1.18) is equivalent to the $(s-2)$-stage collocation method based on $c_2, \dots, c_{s-1}$ (see Table 1.3). The methods with $b_s = 0$ but $b_1 \neq 0$, which include the Radau IA and

Table 1.4. Lobatto IIIB methods of order 4 and 6

$$\begin{array}{c|ccc} 0 & \frac{1}{6} & -\frac{1}{6} & 0 \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{3} & 0 \\ 1 & \frac{1}{6} & \frac{5}{6} & 0 \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array} \qquad \begin{array}{c|cccc} 0 & \frac{1}{12} & \frac{-1-\sqrt{5}}{24} & \frac{-1+\sqrt{5}}{24} & 0 \\ \frac{5-\sqrt{5}}{10} & \frac{1}{12} & \frac{25+\sqrt{5}}{120} & \frac{25-13\sqrt{5}}{120} & 0 \\ \frac{5+\sqrt{5}}{10} & \frac{1}{12} & \frac{25+13\sqrt{5}}{120} & \frac{25-\sqrt{5}}{120} & 0 \\ 1 & \frac{1}{12} & \frac{11-\sqrt{5}}{24} & \frac{11+\sqrt{5}}{24} & 0 \\ \hline & \frac{1}{12} & \frac{5}{12} & \frac{5}{12} & \frac{1}{12} \end{array}$$

Lobatto IIIC methods, are of interest for the solution of stiff differential equations (Hairer & Wanner 1996). The methods with $b_1 = 0$ but $b_s \neq 0$, introduced by Butcher (1964a, 1964b), are of historical interest. They were thought to be computationally attractive, because their last stage is explicit. In the context of geometric integration, much more important are methods for which both $b_1 \neq 0$ and $b_s \neq 0$.

Lobatto IIIB Methods (Table 1.4). We consider the quadrature formulas whose nodes are the zeros of (1.17). We have $c_1 = 0$ and $c_s = 1$. Based on $c_2, \dots, c_{s-1}$ and b_1, b_s we consider the discontinuous collocation method. This class of methods is called Lobatto IIIB (Ehle 1969), and it plays an important role in geometric integration in conjunction with the Lobatto IIIA methods of Sect. II.1.3 (see Theorem IV.2.3 and Theorem VI.4.5). These methods are of order $2s-2$, as the following result shows.

Theorem 1.9 (Superconvergence). *The discontinuous collocation method of Definition 1.7 has the same order as the underlying quadrature formula.*

Proof. We follow the lines of the proof of Theorem 1.5. With the polynomial $u(t)$ of Definition 1.7, and with the defect

$$\delta(t) := \dot{u}(t) - f\big(t, u(t)\big)$$

we get (1.13) after linearization. The variation of constants formula then yields

$$\begin{aligned} u(t_0+h) - y(t_0+h) &= R(t_0+h, t_0)\big(u(t_0) - y_0\big) \\ &\quad + \int_{t_0}^{t_0+h} R(t_0+h, s)\Big(\delta(s) + r(s)\Big)\, ds, \end{aligned}$$

which corresponds to (1.14) if $u(t_0) = y_0$. As a consequence of Lemma 1.10 below (with $k = 0$), the integral over $R(t_0+h, s)r(s)$ gives a $\mathcal{O}(h^{2s-1})$ contribution. Since the defect $\delta(t_0 + c_i h)$ vanishes only for $i = 2, \dots, s-1$, an application of the quadrature formula to $R(t_0+h, s)\delta(s)$ yields $hb_1 R(t_0+h, t_0)\delta(t_0) + hb_s\delta(t_0+h)$ in addition to the quadrature error, which is $\mathcal{O}(h^{p+1})$. Collecting terms suitably, we obtain

$$u(t_1) - hb_s\delta(t_1) - y(t_1) \;=\; R(t_1,t_0)\big(u(t_0) + hb_1\delta(t_0) - y_0\big) \\ +\mathcal{O}(h^{p+1}) + \mathcal{O}(h^{2s-1}),$$

which, after using the definitions of $u(t_0)$ and $u(t_1)$, proves $y_1 - y(t_1) = \mathcal{O}(h^{p+1}) + \mathcal{O}(h^{2s-1})$. □

Lemma 1.10. *The polynomial $u(t)$ of the discontinuous collocation method (1.18) satisfies for $t \in [t_0, t_0+h]$ and for sufficiently small h*

$$\|u^{(k)}(t) - y^{(k)}(t)\| \le C \cdot h^{s-1-k} \qquad \textit{for} \quad k = 0, \dots, s-2.$$

Proof. The proof is essentially the same as that for Lemma 1.6. In the formulas for $\dot u(t_0+\tau h)$ and $\dot y(t_0+\tau h)$, the sum has to be taken from $i=2$ to $i=s-1$. Moreover, all h^s become h^{s-2}. In (1.16) one has an additional term

$$y_0 - u(t_0) = hb_1\big(\dot u(t_0) - f(t_0, u(t_0))\big),$$

which, however, is just an interpolation error of size $\mathcal{O}(h^{s-1})$ and can be included in $Const \cdot h^{s-1}$. □

II.2 Partitioned Runge–Kutta Methods

Some interesting numerical methods introduced in Chap. I (symplectic Euler and the Störmer–Verlet method) do not belong to the class of Runge–Kutta methods. They are important examples of so-called partitioned Runge–Kutta methods. In this section we consider differential equations in the partitioned form

$$\dot y = f(y,z), \qquad \dot z = g(y,z), \tag{2.1}$$

where y and z may be vectors of different dimensions.

II.2.1 Definition and First Examples

The idea is to take two different Runge–Kutta methods, and to treat the y-variables with the first method (a_{ij}, b_i), and the z-variables with the second method $(\widehat a_{ij}, \widehat b_i)$.

Definition 2.1. Let b_i, a_{ij} and $\widehat b_i, \widehat a_{ij}$ be the coefficients of two Runge–Kutta methods. A *partitioned Runge–Kutta method* for the solution of (2.1) is given by

$$\begin{aligned} k_i &= f\Big(y_0 + h\sum_{j=1}^{s} a_{ij}k_j,\; z_0 + h\sum_{j=1}^{s}\widehat a_{ij}\ell_j\Big), \\ \ell_i &= g\Big(y_0 + h\sum_{j=1}^{s} a_{ij}k_j,\; z_0 + h\sum_{j=1}^{s}\widehat a_{ij}\ell_j\Big), \\ y_1 &= y_0 + h\sum_{i=1}^{s} b_i k_i, \qquad z_1 \;=\; z_0 + h\sum_{i=1}^{s}\widehat b_i \ell_i. \end{aligned} \tag{2.2}$$

Methods of this type were originally proposed by Hofer in 1976 and by Griepentrog in 1978 for problems with stiff and nonstiff parts (see Hairer, Nørsett & Wanner (1993), Sect. II.15). Their importance for Hamiltonian systems (see the examples of Chap. I) has been discovered only in the last decade.

An interesting example is the symplectic Euler method (I.1.9), where the implicit Euler method $b_1 = 1, a_{11} = 1$ is combined with the explicit Euler method $\widehat{b}_1 = 1, \widehat{a}_{11} = 0$. The Störmer–Verlet method (I.1.17) is of the form (2.2) with coefficients given in Table 2.1.

Table 2.1. Störmer–Verlet as a partitioned Runge–Kutta method

0	0	0
1	1/2	1/2
	1/2	1/2

1/2	1/2	0
1/2	1/2	0
	1/2	1/2

The theory of Runge–Kutta methods can be extended in a straightforward manner to partitioned methods. Since (2.2) is a one-step method $(y_1, z_1) = \Phi_h(y_0, z_0)$, the Definition 1.2 of the order applies directly. Considering problems $\dot{y} = f(y)$, $\dot{z} = g(z)$ without any coupling terms, we see that the order of (2.2) cannot exceed $\min(p, \widehat{p})$, where p and $\widehat{p}$ are the orders of the two methods.

Conditions for Order Two. Expanding the exact solution of (2.1) and the numerical solution (2.2) into Taylor series, we see that the method is of order 2 if the coupling conditions

$$\sum_{ij} b_i \widehat{a}_{ij} = 1/2, \qquad \sum_{ij} \widehat{b}_i a_{ij} = 1/2 \tag{2.3}$$

are satisfied in addition to the usual Runge–Kutta order conditions for order 2. The method of Table 2.1 satisfies these conditions, and it is therefore of order 2. We also remark that (2.3) is automatically satisfied by partitioned methods that are based on the same quadrature nodes, i.e.,

$$c_i = \widehat{c}_i \qquad \text{for all } i \tag{2.4}$$

where, as usual, $c_i = \sum_j a_{ij}$ and $\widehat{c}_i = \sum_j \widehat{a}_{ij}$.

Conditions for Order Three. The conditions for order three already become quite complicated, unless (2.4) is satisfied. In this case, we obtain the additional conditions

$$\sum_{ij} b_i \widehat{a}_{ij} c_j = 1/6, \qquad \sum_{ij} \widehat{b}_i a_{ij} c_j = 1/6. \tag{2.5}$$

The order conditions for higher order will be discussed in Sect. III.2.2. It turns out that the number of coupling conditions increases very fast with order, and the proofs for high order are often very cumbersome. There is, however, a very elegant proof of the order for the partitioned method which is the most important one in connection with 'geometric integration', as we shall see now.

II.2.2 Lobatto IIIA - IIIB Pairs

These methods generalize the Störmer–Verlet method to arbitrary order. Indeed, the left method of Table 2.1 is the trapezoidal rule, which is the Lobatto IIIA method with $s = 2$, and the method to the right is equivalent to the midpoint rule and, apart from the values of the c_i, is the Lobatto IIIB method with $s = 2$. Sun (1993b) and Jay (1996) discovered that for general s the combination of the Lobatto IIIA and IIIB methods are suitable for Hamiltonian systems. The coefficients of the methods for $s = 3$ are given in Table 2.2. Using the idea of discontinuous collocation, we give a direct proof of the order for this pair of methods.

Table 2.2. Coefficients of the 3-stage Lobatto IIIA - IIIB pair

0	0	0	0
1/2	5/24	1/3	−1/24
1	1/6	2/3	1/6
	1/6	2/3	1/6

0	1/6	−1/6	0
1/2	1/6	1/3	0
1	1/6	5/6	0
	1/6	2/3	1/6

Theorem 2.2. *The partitioned Runge–Kutta method composed of the s-stage Lobatto IIIA and the s-stage Lobatto IIIB method, is of order $2s - 2$.*

Proof. Let $c_1 = 0, c_2, \dots, c_{s-1}, c_s = 1$ and $b_1, \dots, b_s$ be the nodes and weights of the Lobatto quadrature. The partitioned Runge–Kutta method based on the Lobatto IIIA - IIIB pair can be interpreted as the discontinuous collocation method

$$\begin{aligned}
&u(t_0) = y_0 \\
&v(t_0) = z_0 - hb_1\big(\dot v(t_0) - g(u(t_0), v(t_0))\big) \\
&\dot u(t_0 + c_i h) = f\big(u(t_0 + c_i h), v(t_0 + c_i h)\big), \qquad i = 1, \dots, s \\
&\dot v(t_0 + c_i h) = g\big(u(t_0 + c_i h), v(t_0 + c_i h)\big), \qquad i = 2, \dots, s-1 \\
&y_1 = u(t_1) \\
&z_1 = v(t_1) - hb_s\big(\dot v(t_1) - g(u(t_1), v(t_1))\big),
\end{aligned} \tag{2.6}$$

where $u(t)$ and $v(t)$ are polynomials of degree s and $s-2$, respectively. This is seen as in the proofs of Theorem 1.4 and Theorem 1.8. The superconvergence (order $2s - 2$) is obtained with exactly the same proof as for Theorem 1.9, where the functions $u(t)$ and $y(t)$ have to be replaced with $(u(t), v(t))^T$ and $(y(t), z(t))^T$, etc. Instead of Lemma 1.10 we use the estimates (for $t \in [t_0, t_0 + h]$)

$$\begin{aligned}
\|u^{(k)}(t) - y^{(k)}(t)\| &\le c \cdot h^{s-k} \quad \text{for} \quad k = 0, \dots, s, \\
\|v^{(k)}(t) - z^{(k)}(t)\| &\le c \cdot h^{s-1-k} \quad \text{for} \quad k = 0, \dots, s-2,
\end{aligned}$$

which can be proved by following the lines of the proofs of Lemma 1.6 and Lemma 1.10. □

II.2.3 Nyström Methods

> Da bis jetzt die *direkte* Anwendung der Rungeschen Methode auf den wichtigen Fall von Differentialgleichungen zweiter Ordnung nicht behandelt war . . . (E.J. Nyström 1925)

Second-order differential equations

$$\ddot{y} = g(t, y, \dot{y}) \tag{2.7}$$

form an important class of problems. Most of the differential equations in Chap. I are of this form (e.g., the Kepler problem, the outer solar system, problems in molecular dynamics). This is mainly due to Newton's law that forces are proportional to second derivatives (acceleration). Introducing a new variable $z = \dot{y}$ for the first derivative, the problem (2.7) becomes equivalent to the partitioned system

$$\dot{y} = z, \qquad \dot{z} = g(t, y, z). \tag{2.8}$$

A partitioned Runge–Kutta method (2.2) applied to this system yields

$$\begin{aligned} k_i &= z_0 + h\sum_{j=1}^{s} \widehat{a}_{ij}\ell_j, \\ \ell_i &= g\Big(t_0 + c_i h,\, y_0 + h\sum_{j=1}^{s} a_{ij}k_j,\, z_0 + h\sum_{j=1}^{s} \widehat{a}_{ij}\ell_j\Big), \\ y_1 &= y_0 + h\sum_{i=1}^{s} b_i k_i, \qquad z_1 = z_0 + h\sum_{i=1}^{s} \widehat{b}_i\ell_i. \end{aligned} \tag{2.9}$$

If we insert the formula for k_i into the others, we obtain Definition 2.3 with

$$\overline{a}_{ij} = \sum_{k=1}^{s} a_{ik}\widehat{a}_{kj}, \qquad \overline{b}_i = \sum_{k=1}^{s} b_k\widehat{a}_{ki}. \tag{2.10}$$

Definition 2.3. Let $c_i, \overline{b}_i, \overline{a}_{ij}$ and $\widehat{b}_i, \widehat{a}_{ij}$ be real coefficients. A *Nyström method* for the solution of (2.7) is given by

$$\begin{aligned} \ell_i &= g\Big(t_0 + c_i h,\, y_0 + c_i h\dot{y}_0 + h^2\sum_{j=1}^{s} \overline{a}_{ij}\ell_j,\, \dot{y}_0 + h\sum_{j=1}^{s} \widehat{a}_{ij}\ell_j\Big), \\ y_1 &= y_0 + h\dot{y}_0 + h^2\sum_{i=1}^{s} \overline{b}_i\ell_i, \qquad \dot{y}_1 = \dot{y}_0 + h\sum_{i=1}^{s} \widehat{b}_i\ell_i. \end{aligned} \tag{2.11}$$

For the important special case $\ddot{y} = g(t, y)$, where the vector field does not depend on the velocity, the coefficients $\widehat{a}_{ij}$ need not be specified. A Nyström method is of order p if $y_1 - y(t_0 + h) = \mathcal{O}(h^{p+1})$ and $\dot{y}_1 - \dot{y}(t_0 + h) = \mathcal{O}(h^{p+1})$. It is not sufficient to consider y_1 alone. The order conditions will be discussed in Sect. III.2.3.

Notice that the Störmer–Verlet scheme (I.1.17) is a Nyström method for problems of the form $\ddot{y} = g(t, y)$. We have $s = 2$, and the coefficients are $c_1 = 0, c_2 = 1$, $\overline{a}_{11} = \overline{a}_{12} = \overline{a}_{22} = 0$, $\overline{a}_{21} = 1/2$, $\overline{b}_1 = 1/2$, $\overline{b}_2 = 0$, and $\widehat{b}_1 = \widehat{b}_2 = 1/2$. With $q_{n+1/2} = q_n + \frac{h}{2}v_{n+1/2}$ the step $(q_{n-1/2}, v_{n-1/2}) \mapsto (q_{n+1/2}, v_{n+1/2})$ of (I.1.17) becomes a one-stage Nyström method with $c_1 = 1/2$, $\overline{a}_{11} = 0$, $\overline{b}_1 = \widehat{b}_1 = 1$.

II.3 The Adjoint of a Method

We shall see in Chap. V that *symmetric* numerical methods have many important properties. The key for understanding symmetry is the concept of the *adjoint* method.

The flow φ_t of an autonomous differential equation

$$\dot{y} = f(y), \qquad y(t_0) = y_0 \tag{3.1}$$

satisfies $\varphi_{-t}^{-1} = \varphi_t$. This property is *not*, in general, shared by the one-step map Φ_h of a numerical method. An illustration is presented in the upper picture of Fig. 3.1 (a), where we see that the one-step map Φ_h for the explicit Euler method is different from the inverse of Φ_{-h}, which is the implicit Euler method.

Definition 3.1. The *adjoint method* Φ_h^* of a method Φ_h is the inverse map of the original method with reversed time step $-h$, i.e.,

$$\Phi_h^* := \Phi_{-h}^{-1} \tag{3.2}$$

(see Fig. 3.1 (b)). In other words, $y_1 = \Phi_h^*(y_0)$ is implicitly defined by $\Phi_{-h}(y_1) = y_0$. A method for which $\Phi_h^* = \Phi_h$ is called *symmetric*.

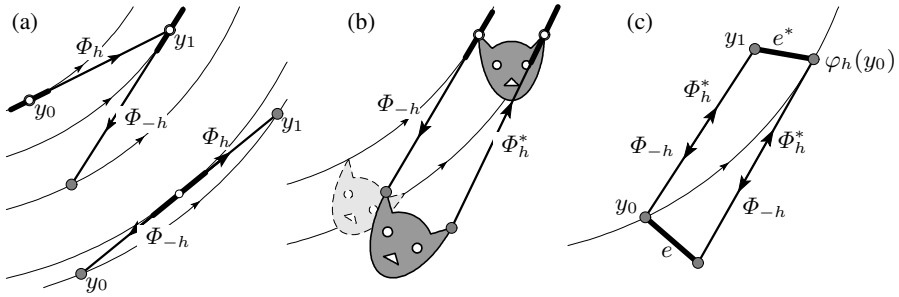


Fig. 3.1. Definition and properties of the adjoint method

The consideration of adjoint methods evolved independently from the study of symmetric integrators (Stetter (1973), p. 125, Wanner (1973)) and from the aim of constructing and analyzing stiff integrators from explicit ones (Cash (1975) calls them "the backward version" which were the first example of mono-implicit methods and Scherer (1977) calls them "reflected methods").

The adjoint method satisfies the usual properties such as $(\Phi_h^*)^* = \Phi_h$ and $(\Phi_h \circ \Psi_h)^* = \Psi_h^* \circ \Phi_h^*$ for any two one-step methods Φ_h and Ψ_h. The implicit Euler method is the adjoint of the explicit Euler method. The implicit midpoint rule is symmetric (see the lower picture of Fig. 3.1 (a)), and the trapezoidal rule and the Störmer–Verlet method are also symmetric.

The following theorem shows that the adjoint method has the same order as the original method, and, with a possible sign change, also the same leading error term.

Theorem 3.2. *Let φ_t be the exact flow of (3.1) and let Φ_h be a one-step method of order p satisfying*

$$\Phi_h(y_0) = \varphi_h(y_0) + C(y_0)h^{p+1} + \mathcal{O}(h^{p+2}). \tag{3.3}$$

The adjoint method Φ_h^* then has the same order p and we have

$$\Phi_h^*(y_0) = \varphi_h(y_0) + (-1)^p C(y_0)h^{p+1} + \mathcal{O}(h^{p+2}). \tag{3.4}$$

If the method is symmetric, its (maximal) order is even.

Proof. The idea of the proof is exhibited in drawing (c) of Fig. 3.1. From a given initial value y_0 we compute $\varphi_h(y_0)$ and $y_1 = \Phi_h^*(y_0)$, whose difference e^* is the local error of Φ_h^*. This error is then 'projected back' by Φ_{-h} to become e. We see that $-e$ is the local error of Φ_{-h}, i.e., by hypothesis (3.3),

$$e = (-1)^p C(\varphi_h(y_0))h^{p+1} + \mathcal{O}(h^{p+2}). \tag{3.5}$$

Since $\varphi_h(y_0) = y_0 + \mathcal{O}(h)$ and $e = (I + \mathcal{O}(h))e^*$, it follows that

$$e^* = (-1)^p C(y_0)h^{p+1} + \mathcal{O}(h^{p+2})$$

which proves (3.4). The statement for symmetric methods is an immediate consequence of this result, because $\Phi_h = \Phi_h^*$ implies $C(y_0) = (-1)^p C(y_0)$, and therefore $C(y_0)$ can be different from zero only for even p. □

II.4 Composition Methods

The idea of composing methods has some tradition in several variants: composition of different Runge–Kutta methods with the same step size leading to the Butcher group, which is treated in Sect. III.1.3; cyclic composition of multistep methods for breaking the 'Dahlquist barrier' (see Stetter (1973), p. 216); composition of low order Runge–Kutta methods for increasing stability for stiff problems (Gentzsch & Schlüter (1978), Iserles (1984)). In the following, we consider the composition of a given basic one-step method (and, eventually, its adjoint method) with *different* step sizes. The aim is to increase the order while preserving some desirable properties of the basic method. This idea has mainly been developed in the papers of Suzuki (1990), Yoshida (1990), and McLachlan (1995).

Let Φ_h be a basic method and $\gamma_1, \ldots, \gamma_s$ real numbers. Then we call its composition with step sizes $\gamma_1 h, \gamma_2 h, \ldots, \gamma_s h$, i.e.,

$$\Psi_h = \Phi_{\gamma_s h} \circ \ldots \circ \Phi_{\gamma_1 h}, \tag{4.1}$$

the corresponding *composition method* (see Fig. 4.1 (a)).

Theorem 4.1. *Let Φ_h be a one-step method of order p. If*

$$\begin{aligned} \gamma_1 + \ldots + \gamma_s &= 1 \\ \gamma_1^{p+1} + \ldots + \gamma_s^{p+1} &= 0, \end{aligned} \tag{4.2}$$

then the composition method (4.1) is at least of order $p + 1$.

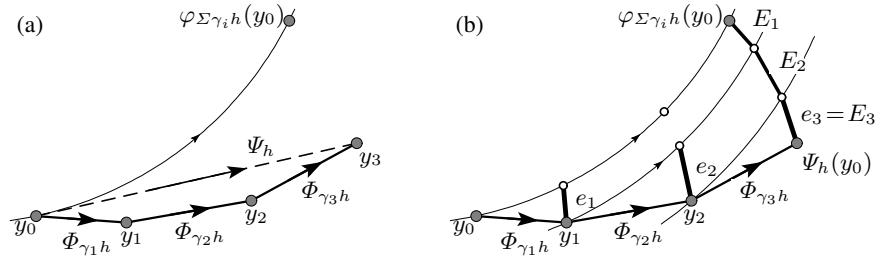


Fig. 4.1. Composition of method Φ_h with three step sizes

Proof. The proof is presented in Fig. 4.1 (b) for $s = 3$. It is very similar to the proof of Theorem 3.2. By hypothesis

$$\begin{aligned} e_1 &= C(y_0) \cdot \gamma_1^{p+1} h^{p+1} + \mathcal{O}(h^{p+2}) \\ e_2 &= C(y_1) \cdot \gamma_2^{p+1} h^{p+1} + \mathcal{O}(h^{p+2}) \\ e_3 &= C(y_2) \cdot \gamma_3^{p+1} h^{p+1} + \mathcal{O}(h^{p+2}). \end{aligned} \tag{4.3}$$

We have, as before, $y_i = y_0 + \mathcal{O}(h)$ and $E_i = (I + \mathcal{O}(h))e_i$ for all i and obtain, for $\sum \gamma_i = 1$,

$$\varphi_h(y_0) - \Psi_h(y_0) = E_1 + E_2 + E_3 = C(y_0)(\gamma_1^{p+1} + \gamma_2^{p+1} + \gamma_3^{p+1})h^{p+1} + \mathcal{O}(h^{p+2})$$

which shows that under conditions (4.2) the $\mathcal{O}(h^{p+1})$-term vanishes. □

Example 4.2 (The Triple Jump). Equations (4.2) have no real solution for odd p. Therefore, the order increase is only possible for even p. In this case, the smallest s which allows a solution is $s = 3$. We then have some freedom for solving the two equations. If we impose symmetry $\gamma_1 = \gamma_3$, then we obtain (Creutz & Gocksch 1989, Forest 1989, Suzuki 1990, Yoshida 1990)

$$\gamma_1 = \gamma_3 = \frac{1}{2 - 2^{1/(p+1)}}, \qquad \gamma_2 = -\frac{2^{1/(p+1)}}{2 - 2^{1/(p+1)}} \,. \tag{4.4}$$

This procedure can be repeated: we start with a symmetric method of order 2, apply (4.4) with $p = 2$ to obtain order 3; due to the symmetry of the γ's this new method is in fact of order 4 (see Theorem 3.2). With this new method we repeat (4.4) with $p = 4$ and obtain a symmetric 9-stage composition method of order 6, then with $p = 6$ a 27-stage symmetric composition method of order 8, and so on. One obtains in this way *any* order, however, at the price of a terrible zig-zag of the step points (see Fig. 4.2).

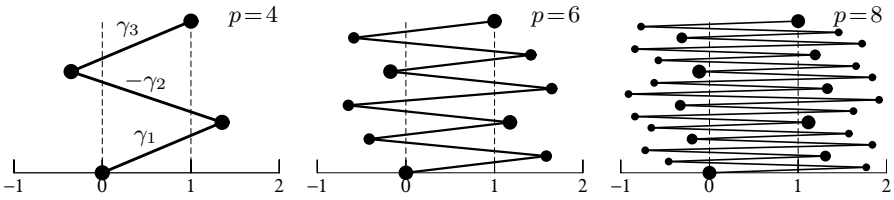


Fig. 4.2. The Triple Jump of order 4 and its iterates of orders 6 and 8

Example 4.3 (Suzuki's Fractals). If one desires methods with smaller values of γ_i, one has to increase s even more. For example, for $s = 5$ the best solution of (4.2) has the sign structure $++-++$ with $\gamma_1 = \gamma_2$ (see Exercise 7). This leads to (Suzuki 1990)

$$\gamma_1 = \gamma_2 = \gamma_4 = \gamma_5 = \frac{1}{4 - 4^{1/(p+1)}}, \qquad \gamma_3 = -\frac{4^{1/(p+1)}}{4 - 4^{1/(p+1)}} \; . \tag{4.5}$$

The repetition of this algorithm for $p = 2, 4, 6, \ldots$ leads to a fractal structure of the step points (see Fig. 4.3).

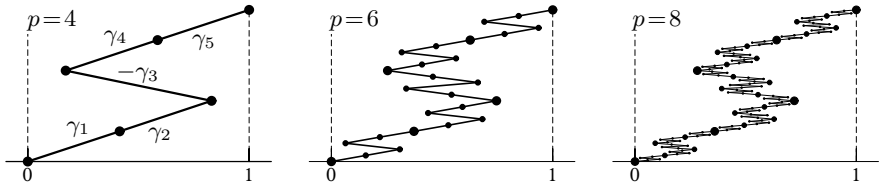


Fig. 4.3. Suzuki's 'fractal' composition methods

Composition with the Adjoint Method. If we replace the composition (4.1) by the more general formula

$$\Psi_h = \Phi_{\alpha_s h} \circ \Phi^*_{\beta_s h} \circ \ldots \circ \Phi^*_{\beta_2 h} \circ \Phi_{\alpha_1 h} \circ \Phi^*_{\beta_1 h}, \tag{4.6}$$

the condition for order $p+1$ becomes, by using the result (3.4) and a similar proof as above,

$$\begin{aligned} \beta_1 + \alpha_1 + \beta_2 + \ldots + \beta_s + \alpha_s &= 1 \\ (-1)^p \beta_1^{p+1} + \alpha_1^{p+1} + (-1)^p \beta_2^{p+1} + \ldots + (-1)^p \beta_s^{p+1} + \alpha_s^{p+1} &= 0. \end{aligned} \tag{4.7}$$

This allows an order increase for odd p as well. In particular, we see at once the solution $\alpha_1 = \beta_1 = 1/2$ for $p = s = 1$, which turns every consistent one-step method of order 1 into a second-order symmetric method

$$\Psi_h = \Phi_{h/2} \circ \Phi^*_{h/2}. \tag{4.8}$$

Example 4.4. If Φ_h is the explicit (resp. implicit) Euler method, then Ψ_h in (4.8) becomes the implicit midpoint (resp. trapezoidal) rule.

Example 4.5. In a second-order problem $\dot q = p$, $\dot p = g(q)$, if Φ_h is the symplectic Euler method, which discretizes q by the implicit Euler and p by the explicit Euler method, then the composed method Ψ_h in (4.8) is the Störmer–Verlet method (I.1.17).

A Numerical Example. To demonstrate the numerical performance of the above methods, we choose the Kepler problem (I.2.2) with $e = 0.6$ and the initial values from (I.2.11). As integration interval we choose $[0, 7.5]$, a bit more than one revolution. The exact solution is obtained by carefully evaluating the integral (I.2.10), which gives

$$\varphi = 8.67002632314281495159108828552, \tag{4.9}$$

with the help of which we compute r, $\dot\varphi$, $\dot r$ from (I.2.8) and (I.2.6). This gives

$$\begin{aligned} q_1 &= -0.82816440269077081820475758537 0 \\ q_2 &= 0.77889809565863544708165448079 6 \\ p_1 &= -0.85638471534339535152448621503 0 \\ p_2 &= -0.16055215079983843525441910410 2\ . \end{aligned} \tag{4.10}$$

As the basic method we use the Verlet scheme and compare in Fig. 4.4 the performances of the composition sequences of the Triple Jump (4.4) and those of Suzuki (4.5) for a large number of different equidistant basic step sizes and for orders $p = 4, 6, 8, 10, 12$. Each basic step is then divided into $3, 9, 27, 81, 243$ respectively $5, 25, 125, 625, 3125$ composition steps and the maximal final error is compared with the total number of function evaluations in double logarithmic scales. For each method and order, all the points lie asymptotically on a straight line with slope $-p$. Therefore, theoretically, a higher order method will become superior when the precision requirements become sufficiently high. But we see that for orders 10 and 12 these 'break even points' are far beyond any precision of practical interest, after some 40 or 50 digits. We also observe that the wild zig-zag of the Triple Jump (4.4) is a more serious handicap than the enormous number of small steps of the Suzuki sequence (4.5).

For later reference we have also included, in black symbols, the results obtained by the two methods (V.3.11) and (V.3.13) of orders 6 and 8, respectively, which will be the outcome of a more elaborate order theory of Chap. III.

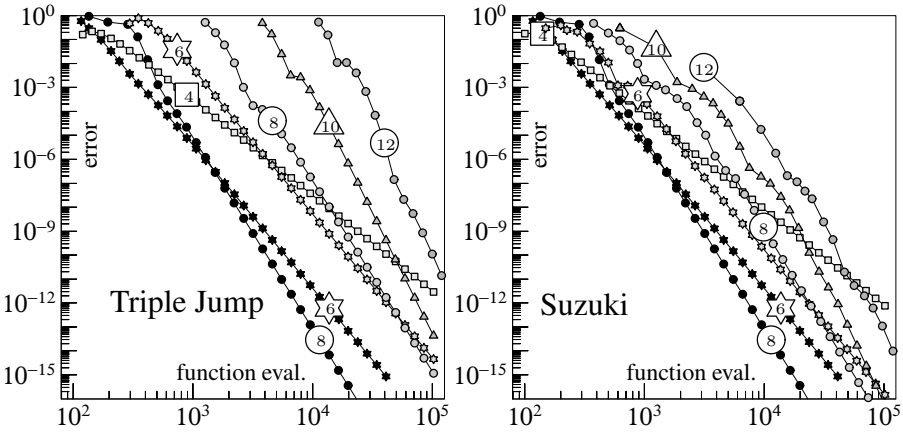


Fig. 4.4. Numerical results of the Triple Jump and Suzuki step sequences (grey symbols) compared to optimal methods (black symbols)

II.5 Splitting Methods

The splitting idea yields an approach that is completely different from Runge–Kutta methods. One decomposes the vector field into integrable pieces and treats them separately.

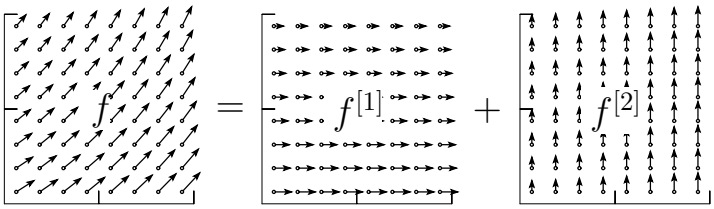


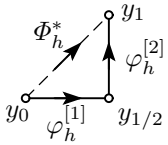
Fig. 5.1. A splitting of a vector field.

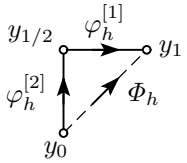
We consider an arbitrary system $\dot{y} = f(y)$ in $\mathbb{R}^n$, and suppose that the vector field is "split" as (see Fig. 5.1)

$$\dot{y} = f^{[1]}(y) + f^{[2]}(y). \tag{5.1}$$

If then, by chance, the exact flows $\varphi_t^{[1]}$ and $\varphi_t^{[2]}$ of the systems $\dot{y} = f^{[1]}(y)$ and $\dot{y} = f^{[2]}(y)$ can be calculated explicitly, we can, from a given initial value y_0, first solve the first system to obtain a value $y_{1/2}$, and from this value integrate the second system to obtain y_1. In this way we have introduced the numerical methods

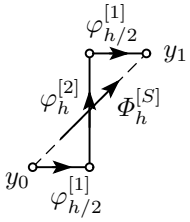
$$\begin{aligned} \Phi_h^* &= \varphi_h^{[2]} \circ \varphi_h^{[1]} \\ \Phi_h &= \varphi_h^{[1]} \circ \varphi_h^{[2]} \end{aligned} \qquad \text{(5.2)}$$





where one is the adjoint of the other. These formulas are often called the *Lie–Trotter splitting* (Trotter 1959). By Taylor expansion we find that $(\varphi_h^{[1]} \circ \varphi_h^{[2]})(y_0) = \varphi_h(y_0) + \mathcal{O}(h^2)$, so that both methods give approximations of order 1 to the solution of (5.1). Another idea is to use a symmetric version and put

$$\Phi_h^{[S]} = \varphi_{h/2}^{[1]} \circ \varphi_h^{[2]} \circ \varphi_{h/2}^{[1]}\,, \qquad \text{(5.3)}$$



which is known as the *Strang splitting*[1] (Strang 1968), and sometimes as the *Marchuk splitting* (Marchuk 1968). By breaking up in (5.3) $\varphi_h^{[2]} = \varphi_{h/2}^{[2]} \circ \varphi_{h/2}^{[2]}$,

[1] The article Strang (1968) deals with spatial discretizations of partial differential equations such as $u_t = Au_x + Bu_y$. There, the functions $f^{[i]}$ typically contain differences in only one spatial direction.

we see that the Strang splitting $\Phi_h^{[S]} = \Phi_{h/2} \circ \Phi_{h/2}^*$ is the composition of the Lie-Trotter method and its adjoint with halved step sizes. The Strang splitting formula is therefore symmetric and of order 2 (see (4.8)).

Example 5.1 (The Symplectic Euler and the Störmer–Verlet Schemes). Suppose we have a Hamiltonian system with separable Hamiltonian $H(p,q) = T(p) + U(q)$. We consider this as the sum of *two* Hamiltonians, the first one depending only on p, the second one only on q. The corresponding Hamiltonian systems

$$\begin{aligned} \dot p &= 0 \\ \dot q &= T_p(p) \end{aligned} \qquad \text{and} \qquad \begin{aligned} \dot p &= -U_q(q) \\ \dot q &= 0 \end{aligned} \tag{5.4}$$

can be solved without problem to yield

$$\begin{aligned} p(t) &= p_0 \\ q(t) &= q_0 + t\, T_p(p_0) \end{aligned} \qquad \text{and} \qquad \begin{aligned} p(t) &= p_0 - t\, U_q(q_0) \\ q(t) &= q_0. \end{aligned} \tag{5.5}$$

Denoting the flows of these two systems by φ_t^T and φ_t^U, we see that the symplectic Euler method (I.1.9) is just the composition $\varphi_h^T \circ \varphi_h^U$. Furthermore, the adjoint of the symplectic Euler method is $\varphi_h^U \circ \varphi_h^T$, and by Example 4.5 the Verlet scheme is $\varphi_{h/2}^U \circ \varphi_h^T \circ \varphi_{h/2}^U$, the Strang splitting (5.3). Anticipating the results of Chap. VI, the flows φ_h^T and φ_h^U are both symplectic transformations, and, since the composition of symplectic maps is again symplectic, this gives an elegant proof of the symplecticity of the 'symplectic' Euler method and the Verlet scheme.

General Splitting Procedure. In a similar way to the general idea of composition methods (4.6), we can form with arbitrary coefficients $a_1, b_1, a_2, \ldots, a_m, b_m$ (where, eventually, a_1 or b_m, or both, are zero)

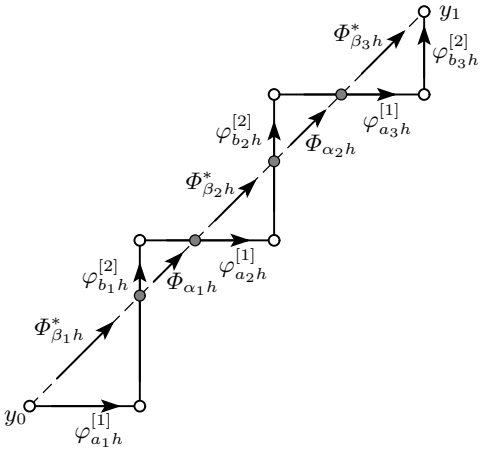
$$\Psi_h = \varphi_{b_m h}^{[2]} \circ \varphi_{a_m h}^{[1]} \circ \varphi_{b_{m-1} h}^{[2]} \circ \ldots \circ \varphi_{a_2 h}^{[1]} \circ \varphi_{b_1 h}^{[2]} \circ \varphi_{a_1 h}^{[1]} \tag{5.6}$$

and try to increase the order of the scheme by suitably determining the free coefficients. An early contribution to this subject is the article of Ruth (1983), where, for the special case (5.4), a method (5.6) of order 3 with $m = 3$ is constructed. Forest & Ruth (1990) and Candy & Rozmus (1991) extend Ruth's technique and construct methods of order 4. One of their methods is just (4.1) with $\gamma_1, \gamma_2, \gamma_3$ given by (4.4) ($p = 2$) and Φ_h from (5.3). A systematic study of such methods started with the articles of Suzuki (1990, 1992) and Yoshida (1990).

A close connection between the theories of splitting methods (5.6) and of composition methods (4.6) was discovered by McLachlan (1995). Indeed, if we put $\beta_1 = a_1$ and break up $\varphi_{b_1 h}^{[2]} = \varphi_{\alpha_1 h}^{[2]} \circ \varphi_{\beta_1 h}^{[2]}$ (group property of the exact flow) where α_1 is given in (5.8), further $\varphi_{a_2 h}^{[1]} = \varphi_{\beta_2 h}^{[1]} \circ \varphi_{\alpha_1 h}^{[1]}$ and so on (cf. Fig. 5.2), we see, using (5.2), that Ψ_h of (5.6) is identical with Ψ_h of (4.6), where

$$\Phi_h = \varphi_h^{[1]} \circ \varphi_h^{[2]} \qquad \text{so that} \qquad \Phi_h^* = \varphi_h^{[2]} \circ \varphi_h^{[1]}. \tag{5.7}$$

A necessary and sufficient condition for the existence of α_i and β_i satisfying (5.8) is that $\sum a_i = \sum b_i$, which is the consistency condition anyway for method (5.6).



$$\begin{aligned} a_1 &= \beta_1 \\ b_1 &= \beta_1 + \alpha_1 \\ a_2 &= \alpha_1 + \beta_2 \\ b_2 &= \beta_2 + \alpha_2 \\ a_3 &= \alpha_2 + \beta_3 \\ b_3 &= \beta_3 \end{aligned} \tag{5.8}$$

Fig. 5.2. Equivalence of splitting and composition methods.

Combining Exact and Numerical Flows. It may happen that the differential equation $\dot y = f(y)$ can be split according to (5.1), such that only the flow of, say, $\dot y = f^{[1]}(y)$ can be computed exactly. If $f^{[1]}(y)$ constitutes the dominant part of the vector field, it is natural to search for integrators that exploit this information. The above interpretation of splitting methods as composition methods allows us to construct such integrators. We just consider

$$\Phi_h = \varphi^{[1]}_h \circ \Phi^{[2]}_h, \qquad \Phi^*_h = \Phi^{[2]*}_h \circ \varphi^{[1]}_h \tag{5.9}$$

as the basis of the composition method (4.6). Here $\varphi^{[1]}_t$ is the exact flow of $\dot y = f^{[1]}(y)$, and $\Phi^{[2]}_h$ is some first-order integrator applied to $\dot y = f^{[2]}(y)$. Since Φ_h of (5.9) is consistent with (5.1), the resulting method (4.6) has the desired high order. It is given by

$$\Psi_h = \varphi^{[1]}_{\alpha_s h} \circ \Phi^{[2]}_{\alpha_s h} \circ \Phi^{[2]*}_{\beta_s h} \circ \varphi^{[1]}_{(\beta_s+\alpha_{s-1})h} \circ \Phi^{[2]}_{\alpha_{s-1} h} \circ \ldots \circ \Phi^{[2]*}_{\beta_1 h} \circ \varphi^{[1]}_{\beta_1 h}. \tag{5.10}$$

Notice that replacing $\varphi^{[2]}_t$ with a low-order approximation $\Phi^{[2]}_t$ in (5.6) would not retain the high order of the composition, because $\Phi^{[2]}_t$ does not satisfy the group property.

Splitting into More than Two Vector Fields. Consider a differential equation

$$\dot y = f^{[1]}(y) + f^{[2]}(y) + \ldots + f^{[N]}(y), \tag{5.11}$$

where we assume that the flows $\varphi^{[j]}_t$ of the individual problems $\dot y = f^{[j]}(y)$ can be computed exactly. In this case there are many possibilities for extending (5.6) and for writing the method as a composition of $\varphi^{[1]}_{a_j h}, \varphi^{[2]}_{b_j h}, \varphi^{[3]}_{c_j h}, \ldots$. This makes it difficult to find optimal compositions of high order. A simple and efficient way is to consider the first-order method

$$\Phi_h = \varphi_h^{[1]} \circ \varphi_h^{[2]} \circ \ldots \circ \varphi_h^{[N]}$$

together with its adjoint as the basis of the composition (4.6). Without any additional effort this yields splitting methods for (5.11) of arbitrary high order.

II.6 Exercises

1. Compute all collocation methods with $s = 2$ as a function of c_1 and c_2. Which of them are of order 3, which of order 4?
2. Prove that the collocation solution plotted in the right picture of Fig. 1.3 is composed of arcs of parabolas.
3. Let $b_1 = b_4 = 1/8$, $c_2 = 1/3$, $c_3 = 2/3$, and consider the corresponding discontinuous collocation method. Determine its order and find the coefficients of the equivalent Runge–Kutta method.
4. Show that each of the symplectic Euler methods in (I.1.9) is the adjoint of the other.
5. (Additive Runge–Kutta methods). Let b_i, a_{ij} and $b_i, \widehat{a}_{ij}$ be the coefficients of two Runge–Kutta methods. An additive Runge–Kutta method for the solution of $\dot y = f^{[1]}(y) + f^{[2]}(y)$ is given by

$$\begin{aligned} k_i &= f^{[1]}\Big(y_0 + h\sum_{j=1}^{s} a_{ij}k_j\Big) + f^{[2]}\Big(y_0 + h\sum_{j=1}^{s} \widehat{a}_{ij}k_j\Big) \\ y_1 &= y_0 + h\sum_{i=1}^{s} b_i k_i. \end{aligned}$$

   Show that this can be interpreted as a partitioned Runge–Kutta method (2.2) applied to

$$\dot y = f^{[1]}(y) + f^{[2]}(z), \qquad \dot z = f^{[1]}(y) + f^{[2]}(z)$$

   with $y(0) = z(0) = y_0$. Notice that $y(t) = z(t)$.
6. Let Φ_h denote the Störmer–Verlet scheme, and consider the composition

$$\Phi_{\gamma_{2k+1}h} \circ \Phi_{\gamma_{2k}h} \circ \ldots \circ \Phi_{\gamma_2 h} \circ \Phi_{\gamma_1 h}$$

   with $\gamma_1 = \ldots = \gamma_k = \gamma_{k+2} = \ldots = \gamma_{2k+1}$. Compute γ_1 and γ_{k+1} such that the composition gives a method of order 4. For several differential equations (pendulum, Kepler problem) study the global error of a constant step size implementation as a function of k.
7. Consider the composition method (4.1) with $s = 5$, $\gamma_5 = \gamma_1$, and $\gamma_4 = \gamma_2$. Among the solutions of

$$2\gamma_1 + 2\gamma_2 + \gamma_3 = 1, \qquad 2\gamma_1^3 + 2\gamma_2^3 + \gamma_3^3 = 0$$

   find the one that minimizes $|2\gamma_1^5 + 2\gamma_2^5 + \gamma_3^5|$.
   Remark. This property motivates the choice of the γ_i in (4.5).