Numerical solution of ordinary differential equations

L. P.

November 2012

1 Euler method

Let us consider an ordinary differential equation of the form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x,t), \tag{1}$$

where $f(x,t)$ is a function defined in a suitable region D of the plane (x,t). Suppose that we wish to evaluate the solution $x(t)$ of this equation, which satisfies the initial condition

$$x(t_0) = x_0, \tag{2}$$

where (x_0, t_0) belongs to the interior of D.

We wish to set up a numerical method for the solution of this problem. Of course, in general, we do not expect to obtain an analytical expression as the result. What we can expect to achieve is to obtain an array containing some values t_n of the independent variable and the corresponding values x_n of $x(t_n)$:

$$\begin{array}{cc} t_0 & x_0 \\ t_1 & x_1 \\ t_2 & x_2 \\ \vdots & \vdots \end{array}$$

To evaluate the quantities x_n, we shall apply different methods which allow us to evaluate $x(t+h)$, where h is a small increment of t which we shall assume to be positive, when $x(t)$ (and in case earlier values of x) is known.

Suppose we know $x(t_0) = x_0$. Let us consider the Taylor expansion of the solution $x(t)$ of equation (1) around t_0:

$$x(t_0 + h) = x_0 + hx_1 + \mathrm{O}\left(h^2\right). \tag{3}$$

We have of course

$$x_1 = \left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_{t=t_0}. \tag{4}$$

On the other hand, since $x(t)$ is by hypothesis the solution of equation (1) around $t = t_0$, we have

$$\left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_{t=t_0} = f(x_0, t_0). \tag{5}$$

We thus obtain an estimate of $x(t_0 + h)$:

$$\tilde{x}(t_0 + h) = x_0 + hf(x_0, t_0). \tag{6}$$

We see that the difference between this estimation and the exact solution $x(t)$ is proportional to h^2.

This reasoning suggests a first method to solve the differential equation (1) on the interval $[t_0, t_0 + T]$ with the initial condition $x(t_0)$. We divide the interval into N subintervals $[t_{n-1}, t_n]$, $n = 1, \ldots, N$, with $t_N = t_0 + T$, $t_n - t_{n-1} = T/N = h$. We let $x(t_0) = x_0$ and evaluate in turn $x(t_n)$, $n = 1, \ldots, N$ via the expression

$$x(t_n) = x(t_{n-1}) + hf(x(t_{n-1}), t_{n-1}), \qquad n = 1, \ldots, N. \tag{7}$$

We thus obtain an array (t_n, x_n) that can be further thickened by interpolation. This method of numerical solution is known as the **Euler method**.

To evaluate the error implied by this method, let us assume to know the exact solution $x(t)$ in the interval $[t_0, t_0 + h]$. We then have

$$x(t_0 + h) = x(t_0) + \int_{t_0}^{t_0+h} \mathrm{d}t \; f(x(t), t). \tag{8}$$

By the theorem of the mean, there is then a value $\bar{t}$ of t between t_0 and $t_0 + h$ such that

$$x(t_0 + h) = x(t_0) + h\, f(x(\bar{t}), \bar{t}). \tag{9}$$

Now $x(t)$ has a derivative and if $f(x, t)$ also possesses derivatives with respect to its arguments, and since $|\bar{t} - t_0| < h$, we have

$$|f(x(\bar{t}), \bar{t}) - f(x_0, t_0)| < Kh, \tag{10}$$

for some positive constant K. We thus obtain

$$|\delta x| < Kh^2. \tag{11}$$

On the other hand, the number of intervals of length h in which we have to divide an interval of fixed length T is given by T/h. Thus the error on the estimate of $x(t_0 + T)$ is proportional to h^1. One expresses this result by saying that the Euler method is a **first-order** method.

Let us consider, e.g., the equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = x, \tag{12}$$

with the initial condition

$$x(0) = 1. \tag{13}$$

It is well known that the solution of this problem is given by

$$x(t) = \mathrm{e}^t. \tag{14}$$

Evaluating the solution by the Euler method, it is easy to see that one obtains

$$x(t + nh) = (1 + h)^n. \tag{15}$$

Thus, if we fix T and N we obtain the approximation

$$x_N(T) = \left(1 + \frac{T}{N}\right)^N. \tag{16}$$

We have of course

$$\lim_{N\to\infty} x_N(T) = x(T). \tag{17}$$

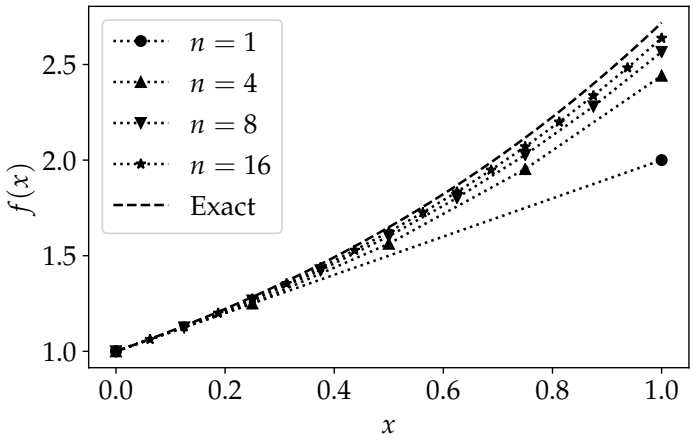

Figure 1: Successive approximations to $x(t) = \exp(x)$, by the Euler method, with $n = 1, 4, 8, 16$ points. The exact solution is also shown.

Although this expression in fact converges to the exact solution, figure 1 shows that it does it quite slowly. Let us in fact imagine that we wish to evaluate $x(T)$ at a fixed value of T. If we divide the interval $[0, T]$ into N intervals, the "step" h will be equal to $1/(TN)$, and the error is proportional to h^2 at each step. Thus the error on $x(T)$, *since the errors add up*, will be proportional to $Nh^2 \sim 1/N$. To obtain one more figure in $x(T)$ we'll have to introduce 10 times more points. To make things explicit, with $N = 32$ we have $x_N(1) = 2.6770$ instead of the exact result $x(1) = 2.7183$, while with 320 points we have $x_N(1) = 2.7140$.

There are even more serious problems. Let us consider the following system of ordinary differential equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = y; \qquad \frac{\mathrm{d}y}{\mathrm{d}t} = -x. \tag{18}$$

The exact solution of this equation, satisfying the initial condition ($x(0) = 0$, $y(0) = 1$) is given by the pair $(\sin t, \cos t)$, so that we obviously have

$$x^2(t) + y^2(t) = x_0^2 + y_0^2; \qquad \forall t. \tag{19}$$

By the Euler method we obtain

$$x(t_0 + h) = x(t_0) + hy(t_0); \tag{20}$$

$$y(t_0 + h) = y(t_0) - hx(t_0). \tag{21}$$

It is useful to write down this relation in matrix form:

$$\mathsf{X}(t_0 + h) = (1 + h\mathsf{F})\,\mathsf{X}(t_0), \tag{22}$$

where

$$\mathsf{X}(t_0) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \tag{23}$$

1 is the unit matrix, and

$$\mathsf{F} = \begin{pmatrix} 0, & 1 \\ -1, & 0 \end{pmatrix}. \tag{24}$$

We thus obtain the following estimate of $\mathsf{X}(T)$:

$$\mathsf{X}_N(T) = \begin{pmatrix} 1, & T/N \\ -T/N, & 1 \end{pmatrix}^N \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{25}$$

Now, if we evaluate $x_N^2(T) + y_N^2(T)$, we obtain

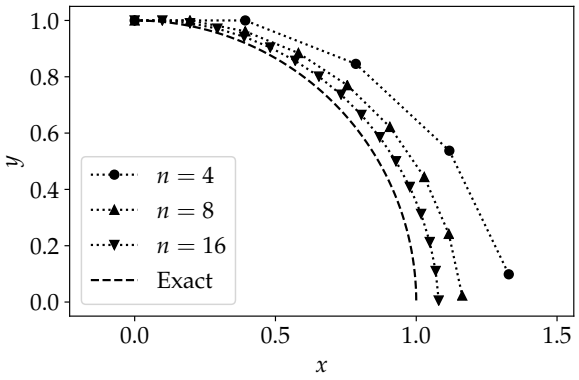

Figure 2: Solution of equation (18) by the Euler method for $0 \le t \le \pi/2$, with initial condition $x(0) = 0$, $y(0) = 1$, and for $N = 4, 8, 16$. The exact solution $x(t) = \sin t$, $y(t) = \cos t$ is also shown.

$$x_N^2(T) + y_N^2(T) = \left(1 + \left(\frac{T}{N}\right)^2\right)^N \left(x^2(0) + y^2(0)\right) \simeq \mathrm{e}^{T^2/N} \left(x^2(0) + y^2(0)\right). \tag{26}$$

We see from figure 2 that the numerical solution wanders away from the exact one as T grows, forming a logarithmic spiral, as shown, e.g., in figure 3.

2 Heun method

To have a better approximation one could think of using more terms in the Taylor expansion (3) in the right-hand side of equation (1). However this approach is not very convenient, since it requires to evaluate the higher derivatives of the $f(x, t)$,

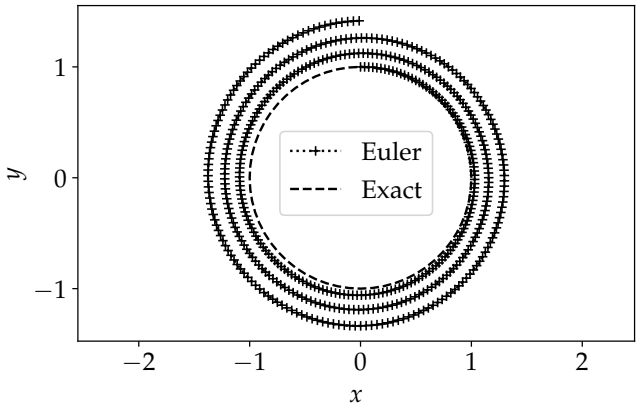

Figure 3: Solution of the differential equation (18) by the Euler method, for $0 \leq t \leq 12\pi$, with the initial condition $x(0) = 1$, $y(0) = 0$, and for $N = 1024$. The exact solution $x(t) = \cos t$, $y(t) = \sin t$ is also shown.

first analytically and then numerically. It would be safer to evaluate just the $f(x, t)$, in case more than one time.

We can find a method of this kind by the following reasoning. Let us consider a differential equation $x'(t) = f_x(x, y)$, whose exact solution $x(t)$ is represented in figure 4. Suppose that for a given value t_0 of t the solution lies at the point $\mathrm{P} = (t_0, x_0)$. Then $f(x_0, t_0)$ is the slope of the tangent to the curve drawn in P. For $t = t_0 + h$ the exact solution will lie, for instance, in R, while the Euler-method solution will lie along the tangent, for instance in Q. In our case, in which the exact solution is convex, Q lies below the curve: otherwise said, Euler's method underestimates the solution.

Suppose that we knew the exact solution for $t = t_0 + h$, and therefore the exact value of $f_1 = f(x(t_0 + h), t_0 + h)$, which corresponds to the slope of the line RS. If, instead of evaluating $x(t_0 + h)$ by the Euler method via $f(x_0, t_0)$ we were to use f_1, we would obtain for $x(t_0 + h)$ the estimate Q_1 (the vectors PQ_1 and RS are equal): then we would have overestimated $x(t_0 + h)$. (Let us remark that if the solution $x(t)$ had been concave rather than convex, the Euler method would have provided an overestimation, while the vector PQ_1 would have been an underestimate.)

It appears that we could obtain a better estimate of $x(t_0 + h)$ by taking the average of these estimates:

$$x(t_0 + h) \simeq x(t_0) + \frac{1}{2} \left(f(x_0, t_0) + f(x(t_0 + h), t_0 + h) \right) h. \tag{27}$$

The problem is that we do not know $x(t_0 + h)$! We can however take advantage of the Euler method to obtain a first estimate of $x(t_0 + h)$, and then use this estimate in (27). We obtain in this way the **Heun method**:

$$x_1 = x_0 + f(x_0, t_0)h; \tag{28}$$

$$x(t_0 + h) \simeq x(t_0) + \frac{h}{2} \left[f(x_0, t_0) + f(x_1, t_0 + h) \right]. \tag{29}$$

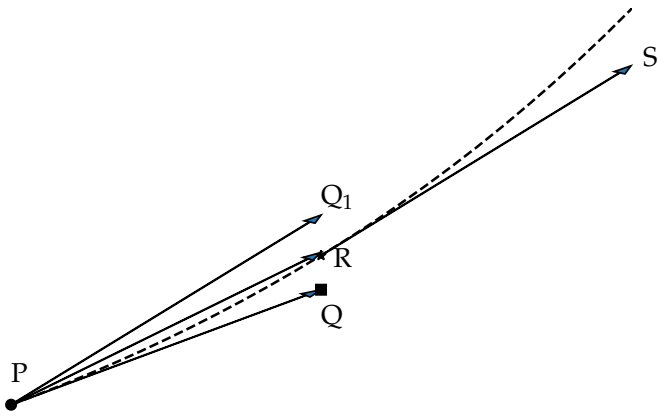

Figure 4: Towards the definition of the Heun method. Given the known solution on point P, the point Q represents the estimate of the solution according to the Euler method. The red curve represents the exact solution. The point Q_1 represents the solution obtained starting from P, but with the tangent evaluated on the point R corresponding to the exact solution. The point R lies near the midpoint of the line QQ_1.

Let us evaluate the solution of $x'(t) = x(t)$ by this method. We obtain

$$x_1 = (1+h)\, x(t_0); \tag{30}$$

$$x(t_0+h) \simeq x(t_0)\left[1+\frac{1}{2}\,(1+1+h)\,h\right] = x(t_0)\left(1+h+\frac{h^2}{2}\right). \tag{31}$$

Thus the first three terms in the Taylor expansion of $x(t_0+h)$ are correctly retrieved. In figure 5 I compare the results of the Euler and the Heun methods for the equation (12). Let us remark that the computational cost lies most often with the evaluation of the function $f(x,t)$. Now the Euler method requires one evaluation for each step, while the Heun method requires two evaluations. It is therefore fairer to compare the two methods with the same number of evaluations of $f(x,t)$, and thus when the number of points of the Euler method is twice that of the Heun one. We now see that on this scale the Heun method with 8 points yields results that cannot be set apart from the exact ones, while the Euler method with 16 points look rather different.

Let us now look at the equation (18). We obtain

$$\mathsf{X}(t+h) = \mathsf{G}\mathsf{X}(t), \tag{32}$$

dove

$$\mathsf{G} = \begin{pmatrix} 1-h^2/2, & -h \\ h, & 1-h^2/2 \end{pmatrix}. \tag{33}$$

We have therefore

$$x^2(t+h) + y^2(t+h) = \left(1+h^4/4\right)\left(x^2(t)+y^2(t)\right). \tag{34}$$

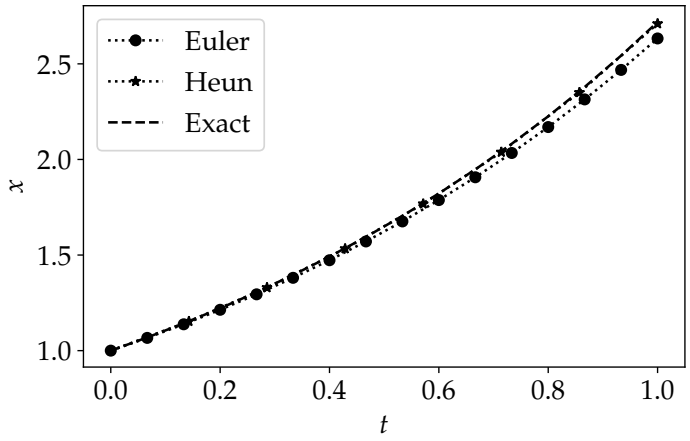

Figure 5: Comparison between the Euler and the Heun method in the solution of equation (12), with the initial condition $x(0) = 1$. The Euler-method solution with 16 points, and the Heun-method one with 8 points (which requires 16 evaluations of $f(x,t)$), and the exact solution are shown.

Thus $x^2 + y^2$ is much better conserved, but it remains true that the errors add up and that the estimated solution is a logarithmic spiral rather than a closed curve.

3 Implicit midpoint method

Suppose again that we knew the exact solution $x(t)$ of the differential equation (1). By the theorem of the mean we have, as we have already seen,

$$x(t_0 + h) = x(t_0) + hf(x(\bar{t}), \bar{t}), \tag{35}$$

where $t_0 < \bar{t} < t_0 + h$. This relation suggests another possible estimate of $x(t_0 + h)$: the midpoint of the interval $[t_0, t_0 + h]$ is most likely closer to $\bar{t}$ than either of the endpoints. We can thus set

$$\tilde{x}(t_0 + h) = x(t_0) + hf(\bar{x}, \bar{t}), \tag{36}$$

where $\bar{x} = x(t_0 + h/2)$. The problem is, of course, that we do not know $x(t_0 + h/2)$. We solved this problem in the Heun method by evaluating $x(t_0 + h)$ via the Euler method, and evaluating the mean increment of $x(t)$ in the interval $[t_0, t_0 + h]$ by averaging the increments estimated at the beginning and at the end of the interval. Another possibility is to set

$$\bar{x} = \frac{x(t_0 + h) + x(t_0)}{2}, \tag{37}$$

i.e., by averaging the values of $x(t)$ at the endpoints. We thus obtain

$$\tilde{x}(t_0 + h) = x(t_0) + hf((\tilde{x}(t_0 + h) + x(t_0))/2, t_0 + h/2). \tag{38}$$

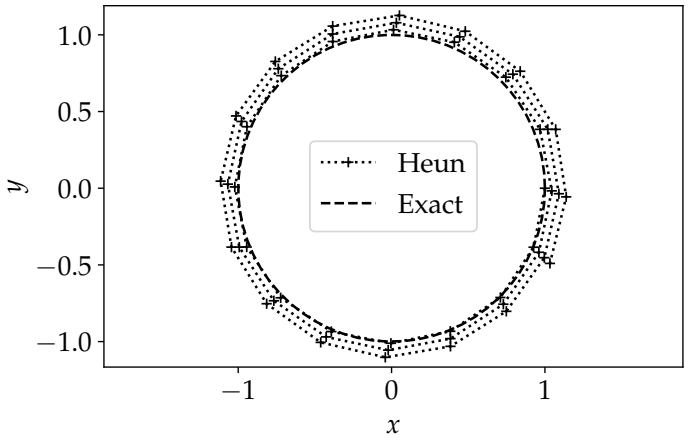

Figure 6: Solution of the differential equation (18) by the Heun method, for $0 \leq t \leq 6\pi$ with the initial condition $x(0) = 1$, $y(0) = 0$, and for $N = 50$. The exact solution is also shown.

This expression should be considered as an equation for $\tilde{x}(t_0 + h)$. Thus, instead of an explicit equation for the estimate of $x(t_0 + h)$, this method provides an equation (which will be most often non linear) which must be solved to obtain the estimate. It is therefore an **implicit method**.

What is its advantage? Let us rewrite (38) in the form

$$\tilde{x}(t_0 + h) - x(t_0) = hf((\tilde{x}(t_0 + h) + x(t_0))/2, t_0 + h/2). \tag{39}$$

We see that $x(t_0)$ and $\tilde{x}(t_0 + h)$ play a perfectly symmetric role in this equation: we can consider it as an equation in $\tilde{x}(t_0 + h)$, where $x(t_0)$ is known, or just as well as an equation in $x(t_0)$, if $\tilde{x}(t_0 + h)$ is known. This symmetry is especially advantageous if the differential equation is invariant under the transformation $t \longrightarrow -t$, as is the case of the equations for the dynamics of particles.

Of course the method requires a fast and cheap algorithm to solve the equation (38). This can be found easily, exploiting the fact that h is small. Let us consider, e.g., the sequence defined by

$$x_0 = x(t_0); \qquad x_{n+1} = x_0 + hf((x_n + x_0)/2, t_0 + h/2); \quad n = 1, 2, \ldots \tag{40}$$

This sequence (for h small enough!) approaches x^*, which satisfies the equation

$$x^* = x_0 + hf((x^* + x_0)/2, t_0 + h/2). \tag{41}$$

Indeed, let us set, e.g., $x_n = x^* + \delta x$. We then have

$$x_{n+1} = x_0 + hf((x_n + x_0)/2, t_0 + h/2) \simeq x^* + h\, f'(x^*)\, \delta x. \tag{42}$$

Thus, if $|h\, f'(x^*)| < 1$, we have $|x_{n+1} - x^*| < |\delta x| = |x_n - x^*|$, and the sequence (x_n) tends to x^*. This method is easy to implement, but is not necessarily the most effective.

Let us now look at the behavior of this method for our differential equations. For (12) we have the equation

$$\tilde{x}(t_0 + h) = \tilde{x} = x_0 + h\left(\frac{x_0 + \tilde{x}}{2}\right). \tag{43}$$

This is a linear equation whose solution reads

$$\tilde{x} = x_0 \frac{1 + h/2}{1 - h/2}. \tag{44}$$

The error of this solution is comparable to that of the Heun method.

It is more interesting to look at the equation (18). Letting

$$\mathsf{X} = \begin{pmatrix} x \\ y \end{pmatrix}; \qquad \|\mathsf{X}\|^2 = x^2 + y^2, \tag{45}$$

and introducing the notations $\mathsf{X}_0 = \mathsf{X}(t_0)$ and $\tilde{\mathsf{X}} = \mathsf{X}(t_0 + h)$, we have

$$\mathsf{A}\tilde{\mathsf{X}} = \mathsf{B}\mathsf{X}_0, \tag{46}$$

where

$$\mathsf{A} = \begin{pmatrix} 1, & -h/2 \\ h/2, & 1 \end{pmatrix}, \tag{47}$$

$$\mathsf{B} = \begin{pmatrix} 1, & h/2 \\ -h/2, & 1 \end{pmatrix}. \tag{48}$$

We thus obtain

$$\tilde{\mathsf{X}} = \mathsf{A}^{-1}\mathsf{B}\,\mathsf{X}_0 = \mathsf{T}\,\mathsf{X}_0, \tag{49}$$

where

$$\mathsf{T} = \frac{1}{1 + h^2/4}\begin{pmatrix} 1 - h^2/4, & h \\ -h, & 1 - h^2/4 \end{pmatrix}. \tag{50}$$

It is then easy to verify that

$$\|\tilde{\mathsf{X}}\|^2 = \|\mathsf{X}_0\|^2. \tag{51}$$

Thus the trajectory of the estimated solution will not wander away from the exact one. This does not imply of course that the solution is exact, as one can see, e.g., in figure 7. However the error lies mostly in the delay with respect to the exact solution, rather than in the trajectory.

The implicit midpoint method as a symplectic integrator

This property of the implicit midpoint method is the consequence of a deeper property. Let us consider a canonical differential equation for the pair of variables (x, p):

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\partial H}{\partial p}; \tag{52}$$

$$\frac{\mathrm{d}p}{\mathrm{d}t} = -\frac{\partial H}{\partial x}. \tag{53}$$

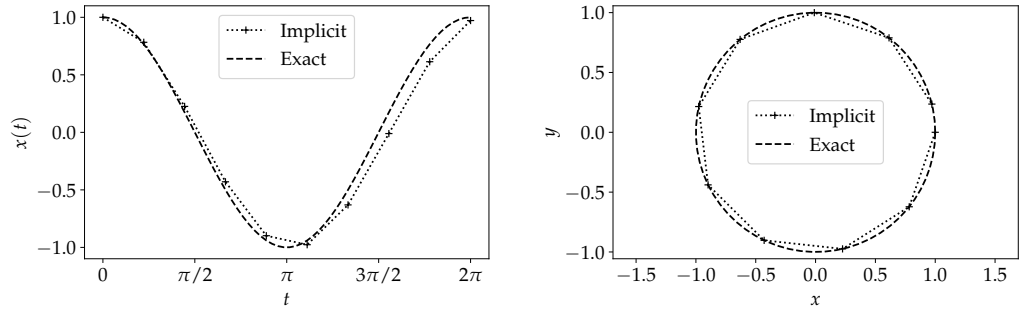

Figure 7: Implicit midpoint method in the solution of equation (18). Left panel: $x(t)$ as a function of t for $0 \leq t \leq 2\pi$, with $N = 10$ points, together with the exact solution $\cos t$. Right panel: the trajectory in the (x, y)-plane.

In this equation, $H(x, p)$ is the hamiltonian. Let us remark that equation (18) takes this shape, if we set $p = y$ and $H = (p^2 + y^2)/2$. Then $H(x, p)$ is conserved in the sense that if $(x(t), p(t))$ is a solution of the (52,53) which satisfies $(x(t_0), p(t_0)) = (x_0, p_0)$, we have

$$H(x(t), p(t)) = H(x_0, p_0) = \text{const.} \tag{54}$$

We have indeed

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial H}{\partial p} = \frac{\partial H}{\partial x}\frac{\partial H}{\partial p} - \frac{\partial H}{\partial p}\frac{\partial H}{\partial x} = 0. \tag{55}$$

Let us now evaluate the change of H for the solution obtained via the implicit midpoint method. Let us denote by $\mathsf{X} = (x, p)$ the generic point of the (x, p)-plane, by $\mathsf{X}_0 = (x_0, p_0)$ the solution at the time t_0, and by $\tilde{\mathsf{X}} = \mathsf{X}_0 + \delta\tilde{\mathsf{X}}$ the estimate of $\mathsf{X}(t_0 + h)$ obtained by the implicit midpoint method:

$$\tilde{\mathsf{X}} = (x_0 + \delta\tilde{x}, p_0 + \delta\tilde{p}), \tag{56}$$

where $(\delta\tilde{x}, \delta\tilde{p})$ are solutions of the equation

$$\delta\tilde{x} = hH_x(x_0 + \delta\tilde{x}/2, p_0 + \delta\tilde{p}/2); \tag{57}$$

$$\delta\tilde{p} = -hH_p(x_0 + \delta\tilde{x}/2, p_0 + \delta\tilde{p}/2) \tag{58}$$

We have introduced the shorthand $H_x = \partial H/\partial x$, etc. Let us denote by $\bar{\mathsf{X}} = (\bar{x}, \bar{p})$ the midpoint. We then have

$$H(\tilde{\mathsf{X}}) = H(\bar{\mathsf{X}}) + H_p\delta\tilde{x}/2 + H_x\delta\tilde{p}/2 + \mathrm{O}\left(\delta\tilde{\mathsf{X}}^2\right); \tag{59}$$

$$H(\mathsf{X}_0) = H(\bar{\mathsf{X}}) - H_p\delta\tilde{x}/2 - H_x\delta\tilde{p}/2 + \mathrm{O}\left(\delta\tilde{\mathsf{X}}^2\right), \tag{60}$$

where the derivatives are evaluated in in $\bar{\mathsf{X}}$. Subtracting these equations, and taking into account the fact that the second-order terms in $\delta\tilde{\mathsf{X}}$ annihilate, we obtain

$$\begin{aligned}\Delta H &= H(\tilde{\mathsf{X}}) - H(\mathsf{X}_0) = H_p\delta\tilde{x}/2 + H_x\delta\tilde{p}/2 + \mathrm{O}\left(\delta\tilde{\mathsf{X}}^3\right) \\ &= h\left(H_xH_p - H_pH_x\right) + \mathrm{O}\left(\delta\tilde{\mathsf{X}}^3\right) = \mathrm{O}\left(\delta\tilde{\mathsf{X}}^3\right).\end{aligned} \tag{61}$$

The conservation of H is thus verified up to order $\delta\tilde{\mathsf{X}}^3$, and thus h^3.

In the case of differential equations of canonical form we have a subtler conservation law. Suppose that the system finds itself in $\mathsf{X}_0 = (x_0, p_0)$ at the initial time$t = t_0$. Consider small perturbations, $\delta\mathsf{X}_1 = (\delta x_0, 0)$ and $\delta\mathsf{X}_1 = (0, \delta p_0)$ of the initial state. In the (x, p)-plane these perturbations identify a parallelogram of area $A = \delta\mathsf{X}_1 \times \delta\mathsf{X}_2 = \delta_0\, \delta p_0$, where "$\times$" denotes the cross product. Let us now follow the solution of the differential equation to a generic instant t. Let us denote by $\mathsf{X}(t)$ the solution satisfying the initial condition $\mathsf{X}(t_0) = \mathsf{X}_0$, and by $\mathsf{X}_1(t)$ and $\mathsf{X}_2(t)$ respectively those which satisfy the initial conditions $\mathsf{X}(t_0) = \mathsf{X}_0 + \delta\mathsf{X}_1$ and $\mathsf{X}(t_0) = \mathsf{X}_0 + \delta\mathsf{X}_2$. Then the vectors $\delta\mathsf{X}_i(t) = \mathsf{X}_i(t) - \mathsf{X}(t)$, for $i = 1, 2$, identify a parallelogram of area $A(t) = \delta\mathsf{X}_1(t) \times \delta\mathsf{X}_2(t)$. Now we have

$$A(t) = A, \qquad \forall t. \tag{62}$$

Indeed, letting

$$\delta\mathsf{X}_i(t) = (\delta x_i(t), \delta p_i(t)), \tag{63}$$

we have, for $i = 1, 2$,

$$\frac{\mathrm{d}\delta x_i}{\mathrm{d}t} = H_{px}\delta x_i + H_{pp}\delta p_i; \tag{64}$$

$$\frac{\mathrm{d}\delta p_i}{\mathrm{d}t} = -H_{xx}\delta x_i - H_{xp}\delta p_i. \tag{65}$$

In this expression, the derivatives of H are evaluated in $\mathsf{X}(t)$. On the other hand we have

$$\begin{aligned}\frac{\mathrm{d}A}{\mathrm{d}t} &= \frac{\mathrm{d}}{\mathrm{d}t}\det\begin{pmatrix} \delta x_1(t), & \delta p_1(t) \\ \delta x_2(t), & \delta p_2(t) \end{pmatrix} = \delta\dot{x}_1\delta p_2 + \delta x_1\delta\dot{p}_2 - \left(\delta\dot{x}_2\delta p_1 + \delta x_2\dot{\delta}p_1\right) \\ &= H_{px}\delta x_1\delta p_2 + H_{pp}\delta p_1\delta p_2 - H_{xx}\delta x_1\delta x_2 - H_{xp}\delta x_1\delta x_2 \\ &\quad - \left(H_{px}\delta x_2\delta p_1 + H_{pp}\delta p_2\delta p_1 - H_{xx}\delta x_2\delta x_1 - H_{xp}\delta x_2\delta p_1\right) = 0.\end{aligned} \tag{66}$$

It is straightforward to generalize this result to a system of $2N$ differential equations, with N pairs of conjugate variables (x_i, p_i) $(i = 1, \ldots, N)$, where the equations have the canonical form (52,53). For each pair the area $A_i(t)$ of the corresponding parallelogram is conserved. As a corollary, sif we consider $2N$ small perturbations $\delta\mathsf{X}_i$ $(i = 1, \ldots, N)$, which encompass a small region of the space $(x_1, p_1, \ldots, x_N, p_N)$, and follow the evolution of this region in time, the corresponding volume remains constant. In mechanics, this result is known as the **Liouville theorem**.

Let us now show that the implicit midpoint method conserves A, of course provided that the $\delta\mathsf{X}_i$ are small enough. Given $\mathsf{X}_0 = \mathsf{X}(t_0)$ and the increment h of t, the estimate $\tilde{\mathsf{X}} = (\tilde{x}, \tilde{p})$ of $\mathsf{X}(t_0 + h)$ is the solution of

$$\tilde{x} = x_0 + hH_p\left(\frac{x_0 + \tilde{x}}{2}, \frac{p_0 + \tilde{p}}{2}\right); \tag{67}$$

$$\tilde{p} = p_0 - hH_x\left(\frac{x_0 + \tilde{x}}{2}, \frac{p_0 + \tilde{p}}{2}\right). \tag{68}$$

Given an increment $\delta \mathsf{X}_0 = (\delta x_0, \delta p_0)$ of the condition at $t = 0$, the corresponding increment $\delta \tilde{\mathsf{X}}$ of $\tilde{\mathsf{X}}$ is the solution of

$$\delta \tilde{x} = \delta x_0 + \frac{h}{2} \left[H_{px}(\delta x_0 + \delta \tilde{x}) + H_{pp}(\delta p_0 + \delta \tilde{p}) \right]; \tag{69}$$

$$\delta \tilde{p} = \delta p_0 - \frac{h}{2} \left[H_{xx}(\delta x_0 + \delta \tilde{x}) + H_{xp}(\delta p_0 + \delta \tilde{p}) \right]. \tag{70}$$

These equations can be written in the following form:

$$\mathsf{A} \delta \tilde{\mathsf{X}} = \mathsf{B} \delta \mathsf{X}_0, \tag{71}$$

where the matrices A and B are defined by

$$\mathsf{A} = \begin{pmatrix} \frac{h}{2} H_{px}, & -\frac{h}{2} H_{pp} \\ \frac{h}{2} H_{xx}, & 1 + \frac{h}{2} H_{xp} \end{pmatrix}; \qquad \mathsf{B} = \begin{pmatrix} 1 + \frac{h}{2} H_{px}, & \frac{h}{2} H_{pp} \\ -\frac{h}{2} H_{xx}, & 1 - \frac{h}{2} H_{xp} \end{pmatrix}. \tag{72}$$

They allow therefore for the solution

$$\delta \tilde{\mathsf{X}} = \mathsf{A}^{-1} \mathsf{B} \delta \mathsf{X}_0 = \mathsf{T} \delta \mathsf{X}_0, \tag{73}$$

where, letting

$$\mathcal{H} = \det \begin{pmatrix} H_{xx}, & H_{xp} \\ H_{px}, & H_{pp} \end{pmatrix}, \tag{74}$$

we have (taking into account the equality of the mixed derivatives)

$$\mathsf{T} = \frac{1}{|\mathsf{A}|} \begin{pmatrix} 1 + hH_{xp} - (h/2)^2 \mathcal{H}, & -hH_{pp} \\ hH_{xx}, & 1 - hH_{px} - (h/2)^2 \mathcal{H} \end{pmatrix}; \tag{75}$$

$$|\mathsf{A}| = 1 - (h/2)^2 \mathcal{H}. \tag{76}$$

Let us consider the initial vectors $\delta \mathsf{X}_1 = (\delta x_0, 0)$ and $\delta \mathsf{X}_2 = (0, \delta p_0)$. The corresponding increments $\delta \tilde{\mathsf{X}}_1$, $\delta \tilde{\mathsf{X}}_2$, are given by

$$\delta \tilde{\mathsf{X}}_1 = \frac{1}{|\mathsf{A}|} \begin{pmatrix} \left(1 + hH_{xp} - (h/2)^2 \mathcal{H}\right) \delta x_0 \\ hH_{xx}, \delta x_0 \end{pmatrix}; \tag{77}$$

$$\delta \tilde{\mathsf{X}}_2 = \frac{1}{|\mathsf{A}|} \begin{pmatrix} -hH_{pp} \, \delta p_0 \\ \left(1 - hH_{px} - (h/2)^2 \mathcal{H}\right) \delta p_0 \end{pmatrix}. \tag{78}$$

With these initial vectors we have $A(t_0) = \delta x_0 \, \delta p_0$. The area encompassed by the vectors $\delta \tilde{\mathsf{X}}_1$ and $\delta \tilde{\mathsf{X}}_2$ is given by the cross product of the vectors, i.e., by the determinant of the matrix one obtains my writing them side by side. An easy calculation shows that one has

$$\delta \tilde{\mathsf{X}}_1 \times \delta \tilde{\mathsf{X}}_2 = |\mathsf{T}| \, \delta x_0 \delta p_0, \tag{79}$$

where $|\mathsf{T}| = \det \mathsf{T}$. It is also easy to verify that

$$|\mathsf{T}| = 1. \tag{80}$$

Thus the implicit midpoint method conserves the areas of the kind of $\delta \tilde{\mathsf{X}}_1 \times \delta \tilde{\mathsf{X}}_2$. These expressions are called **symplectic forms**, and the conservation law is therefore known as the **symplectic property** of the implicit midpoint method. The method is therefore especially adapted for the integration of equations in the canonical form. In the special case of particle dynamics there are however also simpler (explicit) methods that possess this property.

4 Runge-Kutta methods

One of the problems with the implicit midpoint method is that the number of calls of the function $f(x,t)$ needed to solve the equation $\bar{t} = x(t_0 + h/2)$ cannot be forecast. One can ask what happens if one makes just one call to estimate $\bar{x}$, i.e., one lets

$$\bar{x} = x_0 + \frac{h}{2} f(x_0, t_0), \tag{81}$$

and then evaluates the estimate $\tilde{x}$ via equation (36). One thus obtains an algorithm that can be written as follows:

$$k_1 = f(x_0, t_0); \tag{82}$$
$$k_2 = f(x_0 + (h/2)k_1, t_0 + h/2); \tag{83}$$
$$\tilde{x} = x_0 + hk_2. \tag{84}$$

This algorithm is known as the **explicit midpoint method**, oar also as the **second-order Runge-Kutta method**, or RK2. One can show that the errors of this methods are larger than those of the Heun method, while the number of calls of the function $f(x,t)$ is the same. However it allows us to introduce a discussion of several integration methods known as **Runge-Kutta** methods, some of which are the most popular in the applications.

Let us consider equation (8). The integral on the right-hand side can be estimated by means of any method used to evaluate integrals numerically. Let us denote$f(x(t),t)$ by $F(t)$. We then have, e.g.:

The rectangle rule: $I = \int_{t_0}^{t_0+h} \mathrm{d}t\ F(t) \simeq h\, F(t_0 + h/2)$;

The trapezoid rule: $I \simeq \frac{h}{2}\left(F(t_0) + F(t_0 + h)\right)$;

The Simpson rule: $I \simeq \frac{h}{6}\left(F(t_0) + 4F(t_0 + h/2) + F(t_0 + h)\right)$.

These prescriptions also require a method to evaluate $F(t^*) = f(x(t^*), t^*)$, and thus $x^* = x(t^*)$, where t^* is a value of t lying between t_0 and $t_0 + h$. This method can be implicit (i.e., requires the solution of an equation) or explicit (yielding an estimate of x^* as a function of already known quantities). Thus the implicit midpoint method corresponds to the rectangle rule with implicit evaluation, the RK2 to the rectangle rule with explicit evaluation, while the Heun method corresponds to the trapezoidal rule with explicit evaluation. (The advantage of the Heun method with respect to the RK2 one corresponds to the advantage of the trapezoidal rule with respect to the rectangle one.)

In general, these methods require the successive evaluation of quantities of the form

$$k_i = f(x_0 + h\sum_j a_{ij}k_j, t_0 + c_i h), \qquad i = 1, \ldots, s, \tag{85}$$

to produce an estimate of $x(t_0 + h)$ via an expression of the form

$$\tilde{x}(t_0 + h) = x_0 + h\sum_i b_i k_i. \tag{86}$$

These expressions can be conveniently summarized by a **Butcher table**:

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array} \tag{87}$$

In this scheme we have, e.g.,

- For the Euler method:

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array} \tag{88}$$

- For the Heun method:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array} \tag{89}$$

- For the RK2 method:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \hline & 0 & 1 \end{array} \tag{90}$$

- For the implicit midpoint method:

$$\begin{array}{c|c} \frac{1}{2} & \frac{1}{2} \\ \hline & 1 \end{array} \tag{91}$$

Note that in the explicit methods, the elements a_{ij} with $j \geq i$ vanish. In these cases the Butcher table can be simplified in the form

$$\begin{array}{c|ccccc} 0 & & & & & \\ c_2 & a_{21} & & & & \\ c_3 & a_{31} & a_{32} & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ c_s & a_{s1} & a_{s2} & \cdots & a_{s,s-1} & \\ \hline & b_1 & b_2 & \cdots & b_{s-1} & b_s \end{array} \tag{92}$$

Since the k_i can be interpreted as the slope of the solution $x(t)$ evaluated at the time $t_0 + c_i h$, one requires for consistency that the following relations are satisfied:

$$c_i = \sum_j a_{ij}, \quad \text{for } i = 1, \ldots, s; \tag{93}$$

$$\sum_i b_i = 1. \tag{94}$$

The most popular Runge-Kutta method is known as RK4. It is an explicit method described by the following simplified Butcher table:

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array} \tag{95}$$

This corresponds in practice to the following sequence of evaluations:

$$k_1 = f(x_0, t_0); \tag{96}$$
$$k_2 = f(x_0 + (h/2)k_1, t_0 + h/2); \tag{97}$$
$$k_3 = f(x_0 + (h/2)k_2, t_0 + h/2); \tag{98}$$
$$k_4 = f(x_0 + hk_3, t_0 + h); \tag{99}$$
$$\tilde{x}(t_0 + h) = x_0 + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right). \tag{100}$$

The method corresponds to the Simpson rule, with two evaluations of $F(t_0 + h/2)$, where the second one is used to estimate $F(t_0 + h)$. The error on $x(t + h)$ at each step is of order h^5, and thus the total error in a finite interval is of order h^4. Thus it is a fourth-order method, what justifies denoting it by RK4.

Truncation error in Runge-Kutta methods

To evaluate the truncation error in these methods, one can use the following approach. Let us consider the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, t). \tag{101}$$

Assume that we wish to evaluate the solution $x(t)$ in the point $t = t_0 + h$. Let us write the generic Taylor expansion $x(t_0 + h)$ in powers of h:

$$x(t_0 + h) = x_0 + \frac{h}{2}x_1 + \frac{h^2}{2}x_2 + \frac{h^3}{6}x_3 + \mathrm{O}\left(h^4\right). \tag{102}$$

Expanding the right-hand side of equation (101) in powers of h, taking into account the expansion of $x(t_0 + h)$, we obtain

$$f(x(t_0 + h), t_0 + h) = f_0 + h\left(f_x x_1 + f_t\right) + \frac{h^2}{2}\left(f_{xx}(x_1^2 + x_2) + 2f_{xt}x_1 + f_{tt}\right) + \mathrm{O}\left(h^3\right), \tag{103}$$

where $f_0 = f(x_0, t_0)$, $f_x = \partial f/\partial x|_{x_0,t_0}$, etc. Integrating this expression between t_0 and h, and adding $x(t_0) = x_0$ we obtain the expansion of the (101):

$$\begin{aligned} x(t_0 + h) &= x_0 + \int_0^h \mathrm{d}h'\, f(x(t_0 + h'), h') \\ &= x_0 + hf_0 + \frac{h^2}{2}\left(x_1 f_x + f_t\right) + \frac{h^3}{6}\left((x_2 + x_1^2)f_{xx} + 2x_1 f_{xt} + f_{tt}\right) + \mathrm{O}\left(h^4\right). \end{aligned} \tag{104}$$

Comparing equations (104) and (102) we obtain the expressions of the Taylor coefficients x_i:

$$x_1 = f_0; \tag{105}$$
$$x_2 = f_0 f_x + f_t; \tag{106}$$
$$x_3 = f_{xx}f_0^2 + 2f_{xt}f_0 + f_{tt} + f_x\left(f_t + f_0 f_x\right). \tag{107}$$

Let us evaluate an estimate of the RK2 form:

$$k_1 = f(x_0, t_0); \tag{108}$$

$$k_2 = f(x_0 + ak_1, t_0 + ah); \tag{109}$$

$$\tilde{x}(t_0 + h) = x_0 + (1 - b)k_1 + bk_2. \tag{110}$$

We can now evaluate the Taylor expansion in powers of h of the estimate $\tilde{x}(t_0 + h)$:

$$\tilde{x}(t_0 + h) = x_0 + hf_0 + h^2 ab\,(f_x f_0 + f_t) + \frac{h^3}{2} a^2 b\,\left(f_{xx} f_0^2 + 2f_{xt} f_0 + f_{tt}\right) + \mathrm{O}\left(h^4\right). \tag{111}$$

Thus in order for the two expansions to coincide to order h^2 we must have $ab = \frac{1}{2}$. Thus we are only free to fix a. We can then evaluate the difference between (104) and (111):

$$x(t_0 + h) - \tilde{x}(t_0 + h) = \frac{h^3}{12}\left[2f_t f_x + 2f_0 f_x^2 + (2 - 3a)(f_{tt} + 2f_{xt} f_0 + f_{xx} f_0^2)\right] + \mathrm{O}\left(h^4\right). \tag{112}$$

We cannot evaluate the error of order h^3. However, choosing $a = 2/3$, we can simplify its expression:

$$x(t_0 + h) - \tilde{x}(t_0 + h) = \frac{h^3}{6} f_x(f_t + f_0 f_x) + \mathrm{O}\left(h^4\right). \tag{113}$$

Conclusion: the RK2 scheme, with $b = 1/(2a)$, has a truncation error of order h^3 aat each step, and thus a total error of order h^2. In a sense the optimal scheme corresponds to $a = 2/3$.

The corresponding evaluation for RK4 is straightforward but extremely tedious. One can find it in the following *Mathematica* notebook: rk4Eng.nb. The calculation confirms that the error is of order h^5 at each step, and that the total error is of order h^4. Let us stress that the method does not possess the symplectic property, and therefore it is not adapted for the integration of equations in the canonical form. On the other hand it is well adapted for the integration of dissipative systems.

5 Verlet methods

The differential equations governing the evolution of a system of particles interacting via forces which depend only on their reciprocal positions can be solved by a class of very powerful methods known as **Verlet methods**. They are third-order symplectic methods whose implementation is especially simple. Suppose that we wish to solve the differential equation

$$\frac{\mathrm{d}^2 x}{\mathrm{d}t^2} = f(x, t). \tag{114}$$

Let us consider the Taylor expansion of $x(t)$ around t_0, where $x(t_0) = x_0$:

$$x(t_0 + h) = x_0 + hx_1 + \frac{h^2}{2} x_2 + \frac{h^3}{3!} x_3 + \mathrm{O}\left(h^4\right). \tag{115}$$

Evaluating this expansion in $-h$, we obtain

$$x(t_0 - h) = x_0 - hx_1 + \frac{h^2}{2}x_2 - \frac{h^3}{3!}x_3 + \mathrm{O}\left(h^4\right). \tag{116}$$

Summing these two equations we obtain

$$x(t_0 + h) + x(t_0 - h) = 2x_0 + h^2 x_2 + \mathrm{O}\left(h^4\right). \tag{117}$$

On the other hand, since $x(t)$ is a solution of equation (114), we have

$$x_2 = \left.\frac{\mathrm{d}^2 x}{\mathrm{d}t^2}\right|_{t=t_0} = f(x_0, t_0). \tag{118}$$

We obtain therefore

$$x(t_0 + h) = 2x(t_0) - x(t_0 - h) + h^2 f(x_0, t_0) + \mathrm{O}\left(h^4\right) = \tilde{x}(t_0 + h) + \mathrm{O}\left(h^4\right). \tag{119}$$

This equation defines the "standard" Verlet method.

Notice that the method requires at each step the knowledge of the value $x(t_0)$ of the solution at time t_0 and of its value $x(t_0 - h)$ at the *previous step*. On the other hand the initial conditions of equation (114) are usually given in the form $(x(t_0), v(t_0))$, where $v_0 = \mathrm{d}x/\mathrm{d}t|_{t=t_0}$. Thus in order to start evaluating $x(t)$ by the Verlet method, one needs to obtain an estimate of $x(t_0 - h)$ by some other method. In this sense, the standard Verlet method is not **self sufficient**. However, since this evaluation is needed only at the first step, one can use also a rather expensive method to obtain this estimate, such as the RK4 one. Once $x(t_0 - h)$ has been estimated, the standard Verlet method allows for the estimation of $x(t)$ up to order h^3 (global) with *just one* evaluation of $f(x, t)$ aat each step.

Let us also remark that the relation (117) is perfectly symmetric between $x(t_0 + h)$ and $x(t_0 - h)$. Thus the Verlet method, just like the implicit midpoint method, is invariant with respect to the transformation $t \longrightarrow -t$. Therefore, if we evaluate the solution $x(t)$ for $t_0 \leq t \leq t_0 + T$, we can trace it back by solving the differential equation by the Verlet method with negative time increments h, starting from the initial condition $x(t_0 + T + h)$, $x(t_0 + T)$: this reasoning does not take into account the effects of rounding errors, related to the finite precision of the representation of real numbers in the computer.

It is well known that the evolution of a particle system conserves the energy if all forces are conservative. In practice, if $x = (x_i)$, $(i = 1, \ldots, N)$, and equation (114) has the form

$$\frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2} = -\frac{1}{m_i}\frac{\partial U}{\partial x_i}, \tag{120}$$

where $U(x) = U(x_1, \ldots, x_N)$ is some function, we have

$$E = \sum_{i=1}^{N} \frac{1}{2} m_i \left(\frac{\mathrm{d}x_i}{\mathrm{d}t}\right)^2 + U(x) = \text{const.} \tag{121}$$

To evaluate this quantity, we need an estimate of the velocity $v = (\mathrm{d}x_i/\mathrm{d}t)$. Subtracting equations (118) and (119) from each other, we obtain

$$v = \left.\frac{\mathrm{d}x}{\mathrm{d}t}\right|_{t=t_0} = \frac{x(t_0 + h) - x(t_0 - h)}{2h} + \mathrm{O}\left(h^2\right). \tag{122}$$

This expression contains the difference of quantities that are very close to each other, and will be marred by sizable rounding errors. One can still verify that E remains constant up to order h^2 at each step, which suggests that the global error is of order h. In fact the error is smaller, since it is due mostly to the error in the estimate of v, while the estimated solution approaches the exact one with a global error of order h^3. Thus conservation is verified up to order h^3 (globally), even if the estimate of E, step by step, has a h^2 error. However, the fact that the velocity is badly estimated can be corrected by introducing a slightly more complex scheme that also yields a good evaluation of the velocity. This method is known as the **velocity Verlet** method.

The velocity Verlet method

The **velocity Verlet** method formally involves two evaluations of $f(x,t)$ per step, but one can see that each evaluation can be used in two successive steps, so that the computational effort is essentially the same as in the standard Verlet method. Its advantage is that it also yields a good estimate of the velocity at each step. Suppose we know $x(t_0) = x_0$ and $v(t_0) = v_0$. Let us estimate $v(t_0 + h/2)$:

$$v(t_0 + h/2) = v_0 + \frac{h}{2} f(x_0, t_0) + \mathrm{O}\left(h^2\right). \tag{123}$$

To evaluate $x(t_0 + h)$, we use the approximation

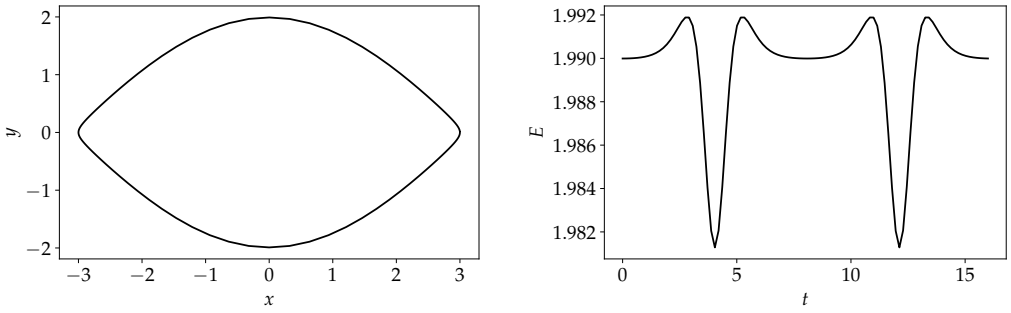

Figure 8: Solution of the pendulum differential equation $\mathrm{d}^2x/\mathrm{d}t^2 = -\sin x$ with initial condition $x_0 = 3$, $v_0 = 0$ by the velocity Verlet algorithm. Left panel: trajectory in the (x, v) plane. Right panel: value of the energy $E = v^2/2 + (1 - \cos x)$. One sees that the value of the energy is not constant, but repeats periodically. Indeed the conserved quantity is a *shadow hamiltonian* whose difference from the "true" hamiltonian can be estimated from the graph.

$$x(t_0 + h) = x(t_0) + hv(t_0 + h/2). \tag{124}$$

To evaluate $v(t_0 + h)$, we use again (123):

$$v(t_0 + h) = v(t_0 + h/2) + (h/2) f(x(t_0 + h), t_0 + h). \tag{125}$$

Eliminating $v(t_0 + h/2)$ by equation (123), we obtain the following algorithm:

$$x(t_0 + h) \simeq x(t_0) + hv(t_0) + (h^2/2)f(x(t_0), t_0); \tag{126}$$

$$v(t_0 + h) \simeq v_0 + \frac{h}{2}\left(f(x_0, t_0) + f(x(t_0 + h), t_0 + h)\right). \tag{127}$$

This method yields an update of the *pair* (x, v) at each step. Note that the second evaluation of the force which appears in the velocity equation can be used for the next update of $x(t)$.

One can show (via the usual tedious method) that the error on $v(t_0 + h)$ and $x(t_0 + h)$ is of order h^3. We can also show explicitly that the method has the symplectic property. Indeed, considering the perturbation $\delta\mathsf{X} = (\delta x_0, \delta v_0)$ we obtain the following expression of the change $\delta\tilde{\mathsf{X}}$:

$$\delta\tilde{\mathsf{X}} = \mathsf{T}\delta\mathsf{X}, \tag{128}$$

where the matrix T is given by

$$\mathsf{T} = \begin{pmatrix} 1 + \frac{h^2}{2}f_x, & h \\ \frac{h}{2}\left[f_x + \left(1 + \frac{h^2}{2}\right)f_x(\tilde{x}(t_0 + h), t_0 + h)\right], & 1 + \frac{h^2}{2}f_x(\tilde{x}(t_0 + h), t_0 + h) \end{pmatrix}. \tag{129}$$

One easily verifies that, to the order evaluated above, one has

$$|\mathsf{T}| = 1. \tag{130}$$

The velocity Verlet algorithm is closely related to the so-called **leapfrog method**, in which the velocity is evaluated at times $t_{k+1/2} = t_0 + (k + 1/2)h$, while positions are evaluated at times $t_k = t + kh$. The leapfrog algorithm reads as follows:

$$v(t_{k+1/2}) \simeq v(t_{k-1/2}) + hf(x(t_k), t_k); \tag{131}$$

$$x(t_{k+1}) \simeq x(t_k) + hv(t_{k+1/2}). \tag{132}$$

Let us remark that the method is explicitly time-reversal invariant. In fact, let us assume that we have obtained the estimates of $x(t_{k+1})$ and $v(t_{k+1/2})$ by the above expressions, and let us evaluate $x(t_k)$ and $v(t_{k-1/2})$. We obtain

$$x(t_k) = x(t_{k+1}) - hv(t_{k+1/2}) = x(t_k) + hv(t_{k+1/2}) - hv(t_{k+1/2}) = x(t_k); \tag{133}$$

$$\begin{aligned} v(t_{k-1/2}) &= v(t_{k+1/2}) - hf(x(t_k), t_k) = v(t_{k-1/2}) + hf(x(t_k), t_k) - hf(x(t_k), t_k) \\ &= v(t_{k-1/2}). \end{aligned} \tag{134}$$

One can also check that the velocity Verlet algorithm defined by equations (126,127) possesses the same invariance. We have indeed, writing x_0 for $x(t_0)$ etc., and x_1 for $x(t_0 + h)$ etc.,

$$\begin{aligned} x_0 &= x_1 - hv_1 + (h^2/2)f(x_1, t_1) \\ &= x_0 + hv_0 + (h^2/2)f(x_0, t_0) - h\left[v_0 + (h/2)\left(f(x_0, t_0) + f(x_1, t_1)\right)\right] + (h^2/2)f(x_1, t_1) \\ &= x_0; \end{aligned} \tag{135}$$

$$\begin{aligned} v_0 &= v_1 - (h/2)\left[f(x_0, t_0) + f(x_1, t_1)\right] \\ &= v_0 + (h/2)\left[f(x_0, t_0) + f(x_1, t_1)\right] - (h/2)\left[f(x_0, t_0) + f(x_1, t_1)\right] \\ &= v_0. \end{aligned} \tag{136}$$

Indeed one can obtain the velocity Verlet algorithm from the leapfrog method by introducing the following estimate of the velocity at times t_k:

$$v(t_k) = \frac{v(t_{h+1/2}) - v(t_{k-1/2})}{h}, \tag{137}$$

and by eliminating $v(t_{k+1/2})$ at each step. Thus, in a sense, the velocity Verlet method is just a different way to write down the leapfrog algorithm.

6 Richardson extrapolation

It is possible to improve our approximations to the exact solution $x(t)$ by taking advantage of the dependence of the estimator $\tilde{x}(t_0 + h)$ on the step size h. Suppose that we have a method which allows us to estimate $x(t_0 + h)$, given $x(t_0)$, with an error of order h^k, where k is some positive integer. We then have

$$x(t_0 + h) = \tilde{x}_h + ah^k + \cdots, \tag{138}$$

where we have neglected term od order higher than h^k. Let us now evaluate $x(t_0 + h)$ using a step size $h/2$ and applying the algorithm twice. We obtain an estimate that we denote by $\tilde{x}(h/2)$, and we have

$$x(t_0 + h) = \tilde{x}_{h/2} + \frac{a}{2^k}h^k + \cdots \tag{139}$$

Let us multiply this equation by 2^k and subtract it form equation (138). We obtain

$$\left(2^k - 1\right) x(t_0 + h) = 2^k \tilde{x}_{h/2} - \tilde{x}_h + \cdots \tag{140}$$

Thus we have

$$x(t_0 + h) = \frac{2^k \tilde{x}_{h/2} - \tilde{x}_h}{2^k - 1} + \cdots, \tag{141}$$

where the neglected terms are of order higher than h^k. This technique can be easily generalized to factors different from 2 and is called **Richardson extrapolation**.

Let us investigate its working for the differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = f(x, t) \tag{142}$$

via the Euler method. We have

$$\tilde{x}_h(t_0 + h) = x_0 + hf_0, \tag{143}$$

where $f_0 = f(x_0, t_0)$, and

$$\tilde{x}_{h/2}(t_0 + h/2) = x_0 + \frac{h}{2}f_0; \tag{144}$$

$$\tilde{x}_{h/2}(t_0 + h) = x_0 + \frac{h}{2}f_0 + \frac{h}{2}f\left(x_0 + \frac{h}{2}f_0, t_0 + \frac{h}{2}\right). \tag{145}$$

Since the *global* error of the Euler method is of order h, let us set $k = 1$. We thus have the extrapolation

$$\begin{aligned} \tilde{x}(t_0 + h) &= 2\tilde{x}_{h/2}(t_0 + h) - \tilde{x}_h(t_0 + h) \\ &= x_0 + hf\left(x_0 + \frac{h}{2}f_0, t_0 + \frac{h}{2}\right). \end{aligned} \tag{146}$$

We have thus recovered the explicit midpoint method.