Patryk Kwoczak

On the Numeric Techniques for Problems in Physics

There are several ways to approach a given problem and translate it into the appropriate information for a program to correctly, up to a given margin, answer with the solution. The topic of this paper is to offer several techniques that programs can use which will produce results, and four sample problems that form the basis of our analysis. This way, the computed results can be considered along with the 'true,' analytic answer to examine the accuracy and demands of the various methods. These problems are the shape of a hanging chain, the general path of fastest descent for a particle affected only by gravity between two points which are not directly vertical, a cannonball shot from a height, and the orbit of a planet around a heavier center of mass. Among these problems, the code produced will account for initial states of the system, and compute solution curves; these solutions aren't equations or polynomial approximations, but discrete arrays which claim to be close to the real solution, where every index is the numeric solution at some particular moment in time.

The methods discussed in the paper are the traditional Euler method, the fourth-order Runge-Kutta method, which will be called RK4 for short, and the symplectic integrators Symplectic Euler, otherwise known as Modified Euler, Euler-Cromer, or semi-implicit Euler, and Verlet integration, or Verlet-Störmer, both of which rely on the geometry which the physical system produces. The objectives of the project are multiple, including to illustrate the varying ranges of 'how close' in how close the different methods produce their solution curves relative to the real solution. Moreover, it is necessary to establish how the laws of physics play into the numeric methods, as well as translating why the methods work.

We begin with a few preliminary notions surrounding every method. The step-size, $\Delta t,$ is a necessary part of the process; because the solution is given in the form of an array, then there must be a uniform treatment over how each element is related to the index, as well as how each element of the array relates to its adjacent neighbors. In fact, before the solution array is constructed, there must first be an array which records how time is progressing. Step size therefore is the interval between each element; if the system begins at time 0 and has a step size of 1, then the elements will be ordered as 0,1,2,3…. If the step size is a quarter, then they increase by the same amount. Therefore, it is clearly seen that the smaller the step size, the more information can be collected at every interval along the array. The systems dealt with in this paper are governed by laws that can be translated into mathematics by differential equations, which provide exact information over how a system evolves with time. However, it is not always easy to find an exact solution to these equations, and so approximations are often developed to give answers which are close enough to the true solutions. Euler's method is the simplest answer to this. Starting from some initial condition, meaning at a certain time, what is the necessary information we need to get things started, then we make predictions over how it will develop between now and the next step: we go there, and then ask once more, and move again.

To illustrate this example, consider the shape of a hanging rope, with uniform weight, held at both ends at a height of approximately 7.715 units, 10 units apart. These, along with the length of the rope, would be the initial conditions. Due to the symmetry, it is sufficient to only find the shape of half the rope, from the bottom to the top, as then it can easily be reflected to produce an entire rope. In this example, the height from the x-axis to the bottom of the rope is 5 units. By the laws of physics governing the problem, the differential equation is as follows: $\frac{d^2y}{dx^2} = \frac{1}{5}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}$, and after a simple substitution of $\frac{dy}{dx} = u$, then it becomes the first-order equation $\frac{du}{dx} = \frac{1}{5}\sqrt{1 + u^2} = f(u)$ (*Wolfram, Catenary*). As can be seen, the exact value of how u changes at any instantaneous time is given by this function, which only needs the current value of u. What Euler's method reasons is to approximate the way u changes from now, at t_0 to the next interval $t_1 = t_0 + \Delta t$, by assuming it is close enough of a change as to the way it changes at t_0. In this example, note that the tangent to the bottom of the rope is parallel to the floor, and therefore $\frac{dy}{dx}$ at t_0 is 0, so $u_0 = 0$ and $\frac{du}{dx} = f(0) = \frac{1}{5}\sqrt{1 + 0^2} = \frac{1}{5}$. Then, the method proceeds to take the current value of u at t_0, and add this change, resulting in $u_1 = u_0 + f(u_0) * \Delta t = \frac{1}{5}$. This continues through iteration until the final value is calculated. However, this will find an array for values of u as time increments. Therefore, once more Euler's method will be used to calculate how y changes given $\frac{dy}{dx} = u$ at each point in time u_i. Notice that in this case, there is no longer a function to insert the current values of y into; instead, at an arbitrary moment in time, $y_{i+1} = y_i + u_i * \Delta t$.

RK4 applies the same philosophy of investigating the differential equation locally around the point to produce an informed guess of the next value. The difference lies in the details.

Instead of just finding the rate of change at the current point, it adapts a weighted strategy of considering the derivative at the current interval, two certain points between the current value and next, and at the next interval. Specifically, the next point is calculated as follows:

$$\text{given} \frac{dy}{dt} = f(t, y),$$

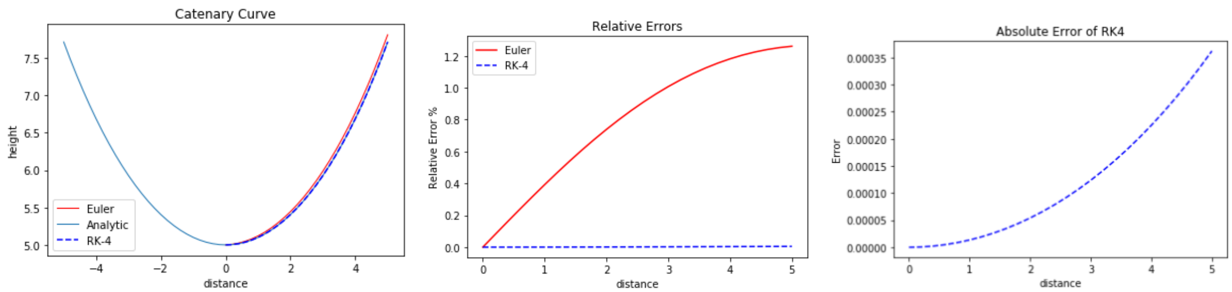
$$k_1 = f(t_n, y_n);\ k_2 = f\left(t_n + \frac{\Delta t}{2}, y_n + \Delta t * \frac{k_1}{2}\right);$$

$$k_3 = f\left(t_n + \frac{\Delta t}{2}, y_n + \Delta t * \frac{k_2}{2}\right);\ k_4 = f(t_{n+1}, y_n + \Delta t * k_3)$$

$$y_{n+1} = y_n + \frac{\Delta t}{6} * (k_1 + 2k_2 + 2k_3 + k_4)$$

(Chai). In the example of the hanging chain, t is not present in the differential equation and so the calculations are reduced to only the dependent variable y. This problem offers insight into the approach of tackling circumstances not usually mentioned; because the problem is a second-order differential equation, then after the solution has been found for u, it must still be found for y. However, there is no longer a convenient function one can evaluate, only discrete values of the derivative at each interval. If the second set of values were to be approached by Euler's method, then they would lose their superior accuracy and be reduced to the lower approximation order. However, it is impossible to apply the same technique of inserting unique values of y, between the two intervals, to obtain weighted calculations of the way the rate of change influences the location of the next point. Therefore, the compromise chosen for the project is to take the average of the derivatives at the current point and next and use that as the basis for the values. Alternatively, it is also possible to tune the step size for the values of u, and use those results to calculate purer version of discretized RK4, or perhaps instead of choosing the average, take weighted values at one-third of the interval and two-thirds.

In the following three images are the approximated solutions for the given problem using Euler's Method and RK4, as well as the errors of the methods when compared to the real analytic solution, known mathematically as the catenary curve, shown in light blue. The step size is 0.2.

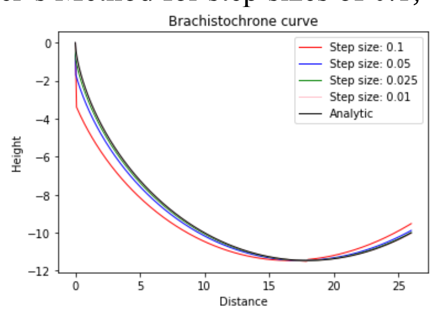


The next problem involves finding the path of shortest time for some object to descend down, influenced only by gravity, from some point A to a lower point B which is not directly below it. For the purposes of the initial conditions, let point A rest at the origin and B at (26,-10). The laws which model such a system produce a differential equation assuming the y-axis is increasing downwards as the following:

$$\left[\frac{dy}{dx}\right]^2 = \frac{k^2}{y} - 1$$

(*Wolfram, Brachistochrone*). Where k is some constant deduced through the initial conditions and the value of the gravity constant. Therefore $\frac{dy}{dx} = f(y)$ must be piecewise based on the sign of $\frac{dy}{dx}$, i.e. when it is decreasing then take the negative value of the square-root of the right hand sign, and positive when increasing. This differential equation is a good example of how one must be very careful when taking numeric integration: note that it must be known when the path is decreasing and increasing, which also means the solution of $f(y) = 0$ must be either calculated or inserted into the program, as well as the singularity at $y = 0$ must be somehow handled, and the issue that the y-axis is in reverse, meaning the values must be flipped before graphing. Moreover, careful designation of the initial conditions is needed to prevent any negative value from entering the square-roots.

The problem illustrates a good reason for lower step sizes: the derivative is critically sensitive at points very close to $y = 0$. What this means is that $\frac{dy}{dx}$ explodes to a large number for small y, as the reciprocal $\frac{1}{y}$ inside of $f(y)$ is very large. The following image represents the solution curve produced by Euler's Method for step sizes of 0.1, 0.05, 0.025, and 0.01:



The red curve representing the largest step size shoots off very quickly far away, while the following smaller sizes grow in increasing care.

The next method which will be discussed arises from systems in physics which follow certain conservation laws. For systems which are not explicitly time-dependent, i.e. the system does not have time as a variable, the Hamiltonian of the system, H, is expressed as the total energy. Then, the system conserves that quantity as time elapses, which gives rise to the canonical equations of motion which describe the time evolution for the coordinates, $\bar{q}_i$, and momenta, $\bar{p}_i$, of the system as $\frac{d\boldsymbol{p}}{dt} = -\frac{\partial H}{\partial \boldsymbol{q}}, \frac{d\boldsymbol{q}}{dt} = +\frac{\partial H}{\partial \boldsymbol{p}}$. Geometrically, a conserved quantity can be considered as a measure-preserving flow. For 2-dimensional problems, this can be expressed as two pairs of two first-order differential equations, for $\boldsymbol{q} = [x, y]$ and $\boldsymbol{p} = [p_x, p_y]$. Additionally, note that velocity v_i is obtained from momentum p_i by the definition $v = \frac{p}{mass}$.

Symplectic integrators are designed to numerically solve Hamiltonian Equations of this nature. An important note is that the symplectic integrators are *not* ordinary numeric integration methods applied to Hamilton's equations of motion (or Lagrangian, for that matter). As such, they can handle a wide range of physical phenomena, but they cannot be as accessible as the general Euler's Method and RK4.

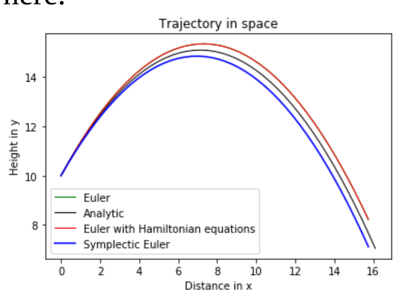
For an example, consider the standard trajectory of a cannonball shot from a bridge, and neglect all air friction, taking into account only the force of gravity on the ball. If one were to apply the standard Euler's method to the above equation, the results would be identically to first calculating the v array for y from the given acceleration of gravity, and then solving for the y array given an initial height. This can be seen as equivalent to solving the following canonical pair given by the previous formulation, where mass $= m$ and gravity is g:

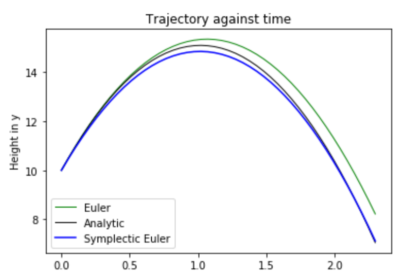
$$\left\{\frac{dy}{dt} = \frac{p_y}{m}, \frac{dp_y}{dt} = mg\right\}, \left\{\frac{dx}{dt} = \frac{p_x}{m}, \frac{dp_x}{dt} = 0\right\}$$

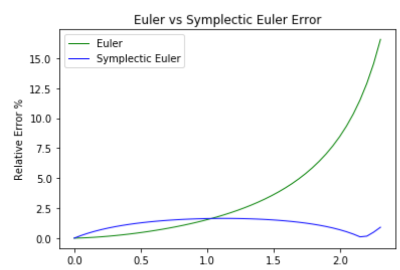
It is identical due to the first set $\frac{dq_i}{dt}$ saying nothing new, it is the identity, and the second set just expressing $F = ma$.

The Symplectic Euler method works as follows: for any arbitrary point y_i, and $\frac{dy}{dt} = f(y)$, then

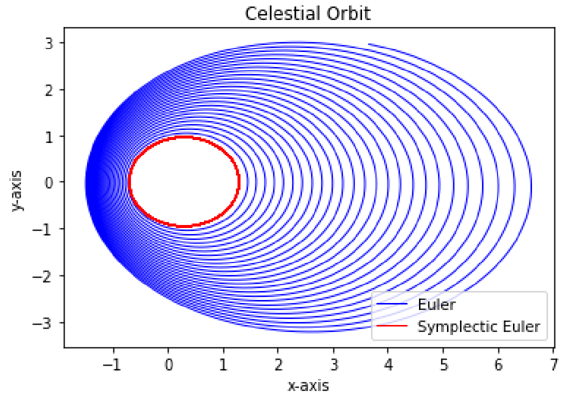
$y_{i+1} = y_i + \Delta t * v_i; v_{i+1} = v_i + \Delta t * f(y_{i+1})$(Rosenthal). In words, the next value for the velocity should be evaluated using y_{i+1} as opposed to y_i. Then the Symplectic Euler should be closer to the true solution than the standard Euler; however, in this problem the application is only partially representing the benefit of the Symplectic algorithm to its naïve counterpart. The solution curve the code produces in only more accurate when considering y as a function of time, when both y and x are plotted, then the two methods are approximately equidistant from the analytic solution, one from above and the second from below. The following graphs are seen here.

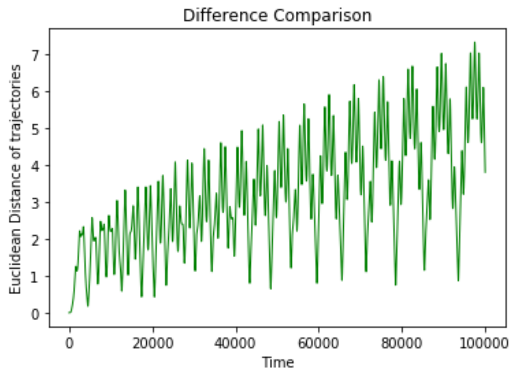






The study of celestial mechanics motivated the advances of Hamilton and his predecessors as methods were investigated to simplify the equations of Newtonian physics. There are many complicated and intricate examples of chaotic behavior among bodies interacting with one another purely through the force of gravity, where if the number of bodies is greater than 3, then periodic orbits are very unstable. Moreover, objective proof of the potency of Symplectic Integration is apparent in the solution trajectory of a celestial body orbiting a center of mass at the origin. The standard Euler fails to conserve the energy of the system, and the body's orbit increasingly grows inaccurate. Given enough time, in the example a total range of 10^5, the Euclidean distance between the two trajectories grows apart by a considerable distance.





On the other hand, the Symplectic integrator is able to preserve the orbit by constructing a system which has a bound on the error related to energy, meaning that it is constructed to solve an approximate Hamiltonian exactly, opposed to solving the exact system approximately. They are constructed to remain invarient by solving the equation which emphasizes staying on the approximate trajectory. In other words, the Symplectic Euler of this example will prioritize preserving the orbit which will go back to its original position.

Another method which is adapted to the symplectic structure of Hamiltonian systems is the Verlet Method. It is constructed as two Taylor Series expansion of third order:

$$y(t+\Delta t) = y(t) + \frac{dy}{dt}\Delta t + \frac{1}{2}\frac{d^2y}{dt^2}\Delta t^2 + \frac{1}{6}\frac{d^3y}{dt^3}\Delta t^3$$

$$y(t-\Delta t) = y(t) - \frac{dy}{dt}\Delta t + \frac{1}{2}\frac{d^2y}{dt^2}\Delta t^2 - \frac{1}{6}\frac{d^3y}{dt^3}\Delta t^3$$

Which combine to produce:

$$y(t+\Delta t) = 2 * y(t) - y(t-\Delta t) + \frac{1}{2}\frac{d^2y}{dt^2}\Delta t^2$$

(Chai).This is the standard Verlet algorithm, which does not rely on the velocity, and directly from accelaration one can produce calculations of position. This is very handy, however one issue can be the loss of precision apparent when calculating the velocity explicitly, for example if it is needed for other properties. Because the algorithm results in velocty found as

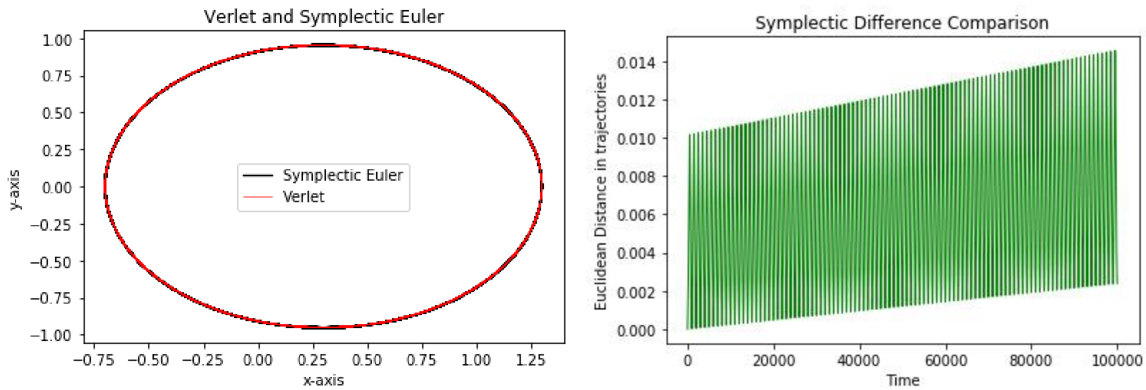
$$v(t) = \frac{y(t+\Delta t) - y(t-\Delta t)}{2\Delta t}.$$

(Chai). A solution is to use the Verlet-Velocity scheme, which incorporates velocity directly into the algorithm. This is the algorithm used in the code.

$$y(t+\Delta t) = y(t) + v(t)\Delta t + \frac{1}{2}a(t)\Delta t^2$$

$$v(t+\Delta t) = v(t) + \frac{a(t) + a(t+\Delta t)}{2}\Delta t$$

(Chai). It is a very simple algorithm to use. Both methods produce very similar trajectories and their relative distance varies by only less than 0.015.



The conclusion of this paper is that every system needs attention and an understanding of the equations being applied to produce a numeric solution which will be accurate without expending uneccesary resources. Several physical systems, such as the Catenary problem, Brachistochrone problem, trajectory of an object in the air, and the two-body problem, have been studied under various integration methods which demonstrated the unique challenges and benefits of prefering certain methods over others, such as RK4 to Euler, and when applicable, Symplectic over the traditional integrators. Credit as well to Rosenthal for explaining the framework in Python for the celestial trajectory for the two Euler methods. When dealing with problems from physics, it is necessary to analyze the system first to consider the most beneficial method to apply, as well as where to be careful.

1: “Brachistochrone Problem.” *From Wolfram MathWorld*, mathworld.wolfram.com/BrachistochroneProblem.html.

2: “Catenary.” *From Wolfram MathWorld*, mathworld.wolfram.com/Catenary.html.

4: Chai, Patrick, and Evan Anzalone. “Numerical Integration Techniques in Orbital Mechanics Applications.” *Research Gate*, doi:10.13140/RG.2.1.1474.2243.

4:Rosenthal, Mike, “Numeric Integration of Orbits – A Crash Course” presentation, https://people.ucsc.edu/~mmrosent/talks_notes/num_int_pres.pdf