PHY256/PHY411 Lecture notes on Integrators

A. C. Quillen

March 10, 2021

Contents

1 Integrating Ordinary Differential Equations

1.1 Euler's method

We start with a dynamical system described by the equation

$$\dot{x} = f(x,t)$$

Starting with initial condition x_0 at t_0 we would like to *advance* and compute an approximate solution for x at a later time $t = t_0 + h$. The ideal solution would be a function $x(t)$ but we only start known the slope of this function at x_0, t_0. We could expand this ideal solution using a Taylor series

$$x(t_0+h) = x(t_0) + x'(t_0)h + x''(t_0)\frac{h^2}{2} + x'''(t_0)\frac{h^3}{3!}.... \tag{1}$$

Here the function $x(t)$ with $x(t_0) = x_0$ and the derivatives are with respect to time. We note that the first order term has slope $x'(t_0)$ that is equal to $f(x_0, t_0)$. Ignoring second order terms

$$x(t_0+h) \sim x(t_0) + hf(x_0,t_0). \tag{2}$$

This is known as the Euler method and it is first order which means that the error is second order or proportional to h^2.

How large is the error? Looking at the Taylor series and comparing equation 1 to equation 2, the error has size

$$\texttt{error} \sim x''(t_0)h^2/2. \tag{3}$$

It makes sense that the error depends on the second derivative because if the actual solution was linear the Euler method would have given the correct answer.

The size of the error in a singe step depends on the actual solution as a function of time.

$$\begin{aligned} x''(t_0) = \frac{d}{dt}x'(t_0) &= \left.\frac{d}{dt}f(x,t)\right|_{t_0} \\ &= \left.\frac{\partial f}{\partial x}\frac{dx}{dt}\right|_{t_0} + \left.\frac{\partial f}{\partial t}\right|_{x_0,t_0} \\ &= \left.\frac{\partial f}{\partial x}\right|_{x_0,t_0} f(x_0,t_0) + \left.\frac{\partial f}{\partial t}\right|_{x_0,t_0} \end{aligned}$$

The error for the first order method depends on the slope of f and the square of the time-step h^2. As $f(x)$ gives the time derivative of x, the second derivative of $x(t)$ depends on the first derivative of $f()$. You can increase the accuracy of the integration by reducing the time-step. If $f'(x)$ is large, then your integration will not be accurate.

With $f(x)$, only indirectly dependent on time, the error in a single step (following equation 3)

$$\texttt{error} \sim f(x_0)f'(x_0)\frac{h^2}{2}.$$

Note this is the error in a single step. The total error after a series of steps depends upon how the individual errors in each step propagate. If they tend to have the same sign they will add, whereas if the errors behave ergodically then they might add in quadrature. Round-off error should add in quadrature and should, on-average, cancel out. However errors in integration often consecutively add giving worse and worse performance the longer the integration. Increasing the accuracy of the integrator only delays and slows the long term drift away from the real solution. It is possible to mitigate the effect of drifts by using *time reversible* or *symplectic* integrators.

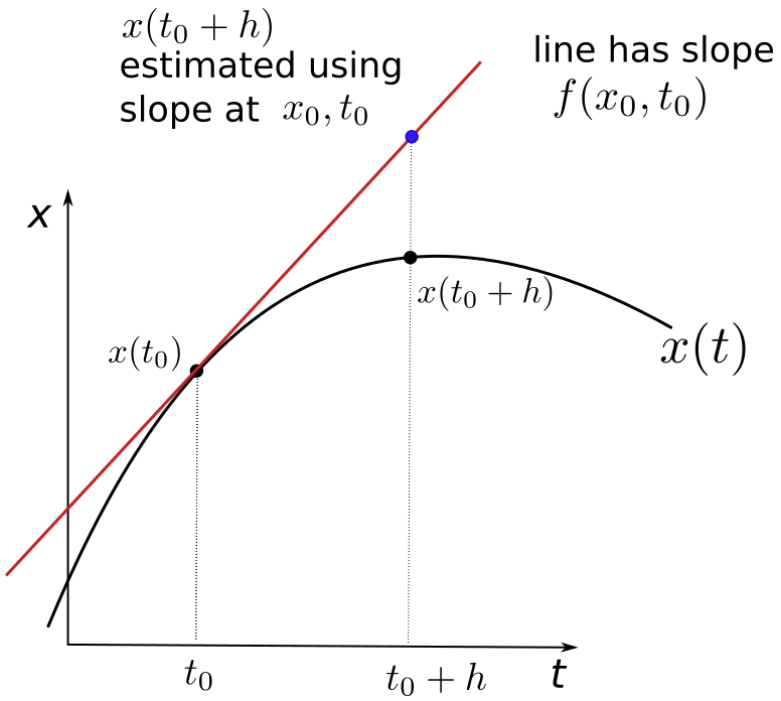

Figure 1: Illustrating using the Euler method to integrate an ordinary differential equation. The size of the error is $x''(t_0)h^2/2$.

1.2 Using the midpoint and the Runge-Kutta method

Let us consider expanding the solution for $x(t)$ around time $t = t_0 + h/2$, the midpoint in time!

$$x(t_0 + h) = x(t_0 + h/2) + \frac{h}{2}\frac{dx}{dt}\bigg|_{t_0+h/2} + \frac{h^2}{8}\frac{d^2x}{dt^2}\bigg|_{t_0+h/2} + O(h^3)$$
$$x(t_0) = x(t_0 + h/2) - \frac{h}{2}\frac{dx}{dt}\bigg|_{t_0+h/2} + \frac{h^2}{8}\frac{d^2x}{dt^2}\bigg|_{t_0+h/2} + O(h^3).$$

Subtracting these two equations gives

$$\begin{aligned} x(t_0 + h) &= x(t_0) + h\,\frac{dx}{dt}\bigg|_{t_0+h/2} + O(h^3) \\ &= x(t_0) + hf\big(x\,(t_0 + h/2)\,, t_0 + h/2\big) + O(h^3). \end{aligned} \tag{4}$$

Notice that the second order terms cancel. To use this we need an estimate for $x(t_0+h/2)$, the x value at the midpoint in time. We approximate x at the midpoint using our first order Euler method

$$x(t_0 + h/2) \sim x(t_0) + \frac{h}{2}f(x_0, t_0) + O(h^2)$$

We want to insert this into equation 4. Because

$$x(t_0 + h/2) \qquad \sim \qquad x(t_0) + \frac{h}{2}f(x_0, t_0)$$

are equal to first order in h

$$hf\big(x(t_0 + h/2), t_0 + h/2\big) \qquad \sim \qquad hf\big(x(t_0) + h/2f(x_0, t_0), t_0 + h/2\big)$$

are equal to second order in h. This means that when we can sub in $x(t_0)+h/2f(x_0, t_0)$ for $x(t_0 + h/2)$ in equation 4 the second order terms are equivalent and so the approximation is still good to second order. The error is $O(h^3)$ which means it is $\propto h^3$.

This approximation using the midpoint is also called a second order Runge-Kutta method. We can also write the method as follows

$$\begin{aligned} k_1 &= hf(x_0, t_0) \\ k_2 &= hf(x_0 + k_1/2, t_0 + h/2) \\ x(t_0 + h) &= x(t_0) + k_2. \end{aligned}$$

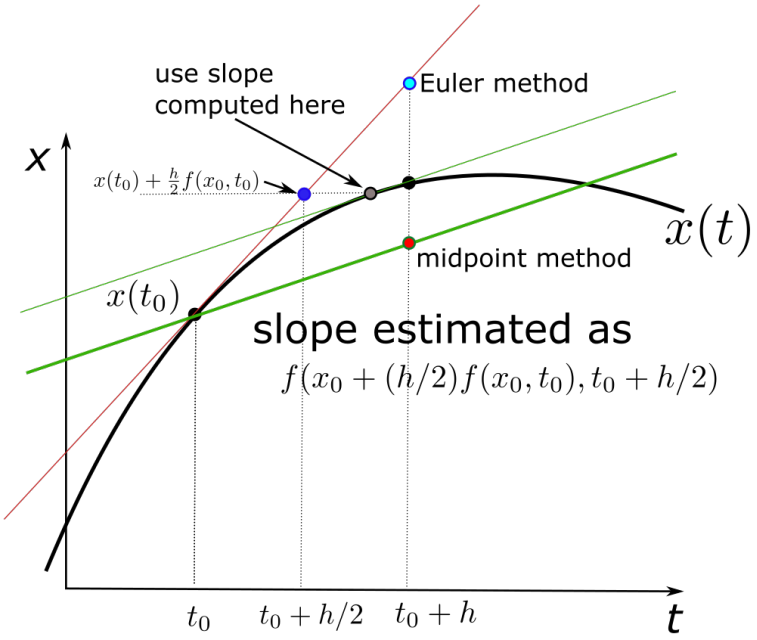

Figure 2: Using the midpoint of the time interval, the estimate for the slope can be improved, giving the second order Runge-Kutta method. Recall that $\dot{x} = f(x)$ and $\dot{x}$ is the slope of $x(t)$ on this plot. This explains why the slope is computed where $x(t)$ is on the same horizontal line as $x(t_0) + \frac{h}{2} f(x_0, t_0)$.

1.2.1 The 4th order Runge-Kutta method

The idea of using midpoints to improve the order the computation it is possible to make a higher order approximation, still using only the ability to evaluate the function for different x, t values. The Fourth order Runge-Kutta looks like this

$$\begin{aligned} k_1 &= hf(x_0, t_0) \\ k_2 &= hf(x_0 + k_1/2, t_0 + h/2) \\ k_3 &= hf(x_0 + k_2/2, t_0 + h/2) \\ k_4 &= hf(x_0 + k_3, t_0 + h) \\ x(t_0 + h) &= x(t_0) + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \end{aligned} \tag{5}$$

Using Taylor expansions it is possible to show that this is accurate to 4th order so the error $\propto h^5$. For a few more computation steps, the fourth order Runge-Kutta vastly improves the accurate of the integration. Due to its simplicity and accuracy the 4th order Runge-Kutta integrator is a *workhorse* of solvers.

- k_1 is the slope at the beginning of the interval, and is computed with Euler's method
- k_2 is the slope computed at the midpoint of the interval evaluated the position given by k_1
- k_3 is the slope computed at the midpoint of the interval evaluated the position given by k_2
- k_4 is the slope at the end of the interval computed using the position given by k_3

The final step is computed using a weighted sum of these slopes at the end point of the interval.

The calculation is fairly tedious, so we will simplify it by taking $f(x)$ and dropping the t dependence of f.

$$\begin{aligned} k_2 &= hf(x_0 + k_1/2) \\ &= h\left[f(x_0) + f'(x_0)\frac{k_1}{2} + f''(x_0)\frac{k_1^2}{4}\frac{1}{2} + f'''(x_0)\frac{k_1^3}{8}\frac{1}{6} + ...\right] \\ &= hf + \frac{h^2}{2}f'f + \frac{h^3}{8}f''f^2 + \frac{h^4}{48}f'''f^3 + O(h^5) \end{aligned}$$

$$
\begin{aligned}
k_3 &= hf(x_0 + k_2/2) \\
&= hf + hf'\frac{k_2}{2} + hf''\frac{k_2^2}{4}\frac{1}{2} + hf'''\frac{k_2^3}{8}\frac{1}{6}... \\
&= hf + \frac{h}{2}f'\left(hf + \frac{h^2}{2}f'f + \frac{h^3}{8}f''f^2 + ..\right) + \frac{h}{8}f''\left(hf + \frac{h^2}{2}f'f + ..\right)^2 + \frac{hf'''}{48}(hf)^3 + ... \\
&= hf + h^2\frac{f'f}{2} + h^3\left(\frac{f'f'f}{4} + \frac{f''f^2}{8}\right) + h^4\left(\frac{f'f''f}{16} + \frac{f''f'f^2}{8} + \frac{f'''f^3}{48}\right) + O(h^5)
\end{aligned}
$$

$$
\begin{aligned}
k_4 &= hf(x_0 + k_3) \\
&= hf + hf'k_3 + hf''k_3^2\frac{1}{2} + hf'''k_3^3\frac{1}{6}... \\
&= hf + hf'\left(hf + h^2\frac{f'f}{2} + h^3\left(\frac{f'f'f}{4} + \frac{f''f^2}{8}\right)\right) + \frac{hf''}{2}\left(hf + \frac{h^2}{2}f'f\right)^2 + \frac{hf'''}{6}(hf)^3 + ... \\
&= hf + h^2f'f + h^3\left(\frac{f'f'f}{2} + \frac{f''f^2}{2}\right) + h^4\left(\frac{f'f'f'f}{4} + \frac{f'f''f^2}{8} + \frac{f''f'f^2}{2} + \frac{f'''f^3}{6}\right) + O(h^5)
\end{aligned}
$$

All these expressions are evaluated at x_0. We now have what we need to evaluate the right hand side of equation 5.

We need to expand the left hand side of equation 5.

$$
x(t_0 + h) = x(t_0) + x'h + x''\frac{h^2}{2} + x'''\frac{h^3}{3!} + x''''\frac{h^4}{4!} + O(h^5) \tag{6}
$$

$$
\begin{aligned}
x'(t_0) &= \frac{dx}{dt}\Big|_{t_0} = f(x_0) \\
x''(t_0) &= \frac{d}{dt}f(x(t))\Big|_{t_0} = f'\dot{x} = f'(x_0)f(x_0) \\
x'''(t_0) &= \frac{d}{dt}f'(x(t))f(x(t))\Big|_{t_0} \\
&= f''\dot{x}f + f'f'\dot{x} = f''(x_0)f^2(x_0) + f'(x_0)f'(x_0)f(x_0) \\
x''''(t_0) &= \frac{d}{dt}\left(f''(x(t))f^2(x(t)) + f'(x(t))f'(x(t))f(x(t))\right) \\
&= f'''\dot{x}f^2 + f''2ff'\dot{x} + 2f'f''\dot{x}f + f'f'f'\dot{x} \\
&= f'''f^3 + 2f''f'f^2 + 2f'f''f^2 + f'f'f'f \\
&= f'''f^3 + 4f''f'f^2 + f'f'f'f
\end{aligned}
$$

Inserting these into the expansion for $x(t_0 + h)$ (equation 6)

$$
x(t_0 + h) = x_0 + hf + h^2\frac{f'f}{2} + \frac{h^3}{6}\left(f''f^2 + f'f'f\right) + \frac{h^4}{4!}\left(f'''f^3 + 4f''f'f^2 + f'f'f'f\right) + O(h^5)
$$

and the functions are evaluated at x_0. If I have not made any mistakes we should find that the coefficients in equation 5 are consistent with a 4th order accurate integrator. I have checked and find that there are no mistakes up to and including terms proportional to h^4!

The 4-th order Runge-Kutta method is in the class of **explicit** Runge-Kutta methods. These involve sequential use of slopes computed at different points in the interval. To specify the method, you provide the number of stages, the positions in the interval and how the slopes depend on each other in the computation. This and the desired order of the integrator place constraints on the coefficients used in the method.

1.3 Numbers of integration steps: comparing second order to fourth order integrators

Suppose we desire the most accurate possible integration. The order of the integrator n relates the size of the error to the step size, $e_h \propto h^{n+1}$. We set the error in each step to be the minimum possible given the double floating point precision level of $\sim 10^{-16}$ and this lets choose our step size h, with

$$h \sim e_h^{\frac{1}{n+1}}.$$

The number of steps we need to integrate an interval T is

$$N = T/h \propto e_h^{-\frac{1}{n+1}}$$

For a second order integrator $N \propto e_h^{-1/3}$ and for a fourth order integrator $N \propto e_h^{-1/5}$. The minus third power of 10^{-15} is 10^5 but the minus fifth power of 10^{-15} is 1000. Even if we need 10 times the number of computations for each step in the fourth order integrator than for the second order one, it will likely take less time to do the total integration with the fourth order method than if we use a second order method.

A large number of small steps are needed to integrate a function that is rapidly changing. If the function is not continuous (or *stiff*) then integrator routines can give an error. An example of a stiff setting might be a mass on a frictional surface where the force on the body is not a smooth function of velocity. In such a setting you might approximate your force with a continuous function so that your integrator is better behaved.

1.4 Adaptive Step-Sizes

Looking at the error in the first order Newtonian method for integration of an ordinary differential equation

$$\texttt{Err} \sim x''(t_0)h^2/2 \sim f'(x_0)f(x_0)h^2/2$$

If $x(t)$ varies rapidly, then the error is large. We might want to adjust our step size so that we achieve the same level of accuracy in different regions of the trajectory.

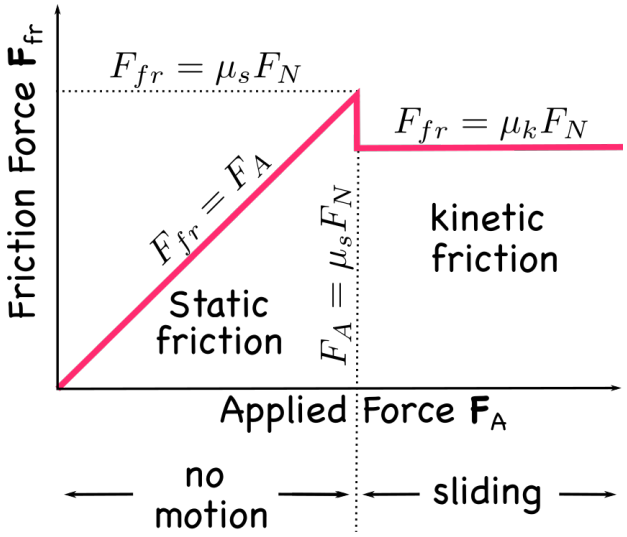

Figure 3: The friction force F_{fr} as a function of applied horizontal force, F_A. We show static and kinetic friction regimes. This is an example of a non-continuous force that might give errors in numerical integration.

An example setting might be a highly eccentric Keplerian orbit where the body is moving fast nearing pericenter and slowly at apocenter. To integrate the orbit carefully you would need to take small stepsizes (in time) at pericenter, but at apocenter, where the body is moving slowly, one could take extremely large stepsizes without compromising the accuracy.

To implement this we take two steps of h and compare the result to what we get if we take a single step of $2h$. Suppose we get x_1 after taking the single large step and x_2 after taking the two smaller steps. The difference between the two computations tells us the size of the error. If the error is very small we can increase the step size before we take the next step. And if the error is too big we would reduce the step size and try again, only taking the step size if we are happy with the size of the error.

A single step of $2h$ gives an error $\propto (2h)^{n+1}$ where n is the order of the integrator. Whereas a single step of h gives an error of size $\propto h^{n+1}$. It would be nice to add two of these in quadrature but errors could add during consecutive timesteps. We compare

$$(2h)^{n+1} \qquad \texttt{vs} \qquad 2h^{n+1}$$

The error is much larger for the single large step than for the two smaller steps. This means that $|x_2 - x_1|$ essentially measures the size of the error for the single large timestep of size $2h$.

There is an unknown constant c relating the size of an error to the step size. For a

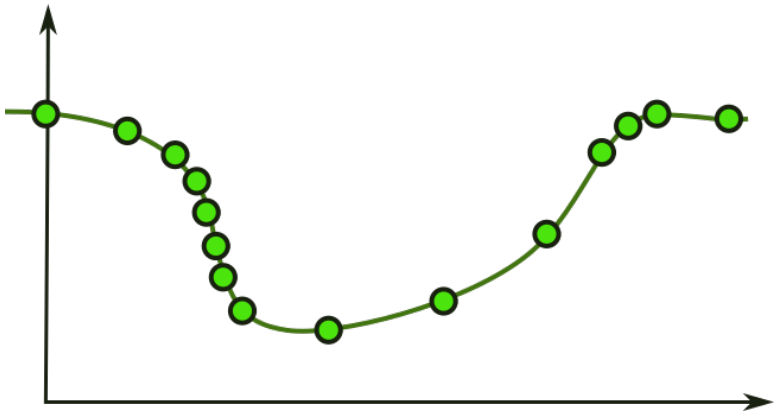

Figure 4: Illustrating that smaller step sizes are taken when the function changes rapidly.

single step of size h

$$\texttt{err} = ch^{n+1}$$

We can estimate the coefficient c using the difference $|x_2 - x_1|$

$$|x_2 - x_1| \sim c(2h)^{n+1} \tag{7}$$

Solving for c

$$c \approx \frac{|x_2 - x_1|}{(2h)^{(n+1)}}$$

If our desired level of accuracy per unit time is δ then we want to chose our new stepsize h' so that

$$\texttt{err} = ch'^{n+1} \sim \delta h'$$

or

$$h' \sim \left(\frac{\delta}{c}\right)^{\frac{1}{n}}$$

Taking our estimated value for c and put it in this equation we find

$$h' = \left(\frac{\delta}{|x_2 - x_1|}\right)^{\frac{1}{n}} (2h)^{1+\frac{1}{n}}$$

And this gives us a way to adjust the step size to the desired level of accuracy.

An adaptive step size integrator is more computationally expensive than a fixed step size integrator. However, you would speed up the calculation overall by taking larger steps

when the system is slowly varying. Furthermore if you by mistake take steps that are too large when the solution is rapidly varying, you might get an unstable or unbelievable solution. For single particle dynamics where there is a large dynamic range, one would preferentially choose a variable step size integrator.

1.5 More than one degree of freedom and dynamics with a potential force

Much of the above discussion can be generalized for multidimensional systems where $\mathbf{x}$ is vector.

For example Newtonian dynamics in one dimension, the acceleration $\ddot{y}$ is equal to a force divided by a mass

$$m\ddot{y} = f(y)$$

and we can make this look like a dynamical system with

$$\mathbf{x} = \begin{pmatrix} y \\ \dot{y} \end{pmatrix}$$

and

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$$

with

$$\mathbf{F}(\mathbf{x}) = \begin{pmatrix} \dot{y} \\ f(y)/m \end{pmatrix}$$

If we have a potential force then the force

$$f(y) = -\frac{\partial U(y)}{\partial y}$$

where $U(y)$ is the potential energy.

It is convenient to work with potential energy per unit mass and force per unit mass, so henceforth we set $m = 1$. Let $\dot{y} = v$. We compute

$$\begin{aligned} \dot{v} &= -\frac{dU(y)}{dy} \\ v\dot{v} &= -\frac{dU}{dy}v = -\frac{dU}{dy}\dot{y} = -\frac{d}{dt}U(y) \\ \frac{d}{dt}\frac{v^2}{2} &= -\frac{d}{dt}U(y) \\ \frac{v^2}{2} + U(y) &= \texttt{constant} \end{aligned}$$

The term on the left is the kinetic energy per unit mass and the term on the right is the potential energy per unit mass. We recognize energy as a conserved quantity. On a plot of v vs y, trajectories ($v(t)$ and $y(t)$) would be curves of constant energy. In other words, you can plot constant energy contours as a function of v and y and these contours would also be orbits.

However when we call our integrator, it is not obvious whether energy is conserved to high order. In fact, it is unlikely. The integrators we have discussed are designed to match the equations of motion to a certain order and so to low order in step-size they will conserve energy. Using a first order Newton method or a Runge-Kutta method, we have not required energy to be conserved at each step, so on long timescales there will be a drift in energy.

1.6 The meaning of incompressible flow

Let us expand on the connection between being divergence free and being incompressible. Consider a distribution of particles in **phase space**. In three Cartesian dimensions, phase space is x, y, z, v_x, v_y, v_z and is six dimensional as both coordinate and velocity are three-dimensional spaces. The number of particles per unit volume would be $\rho(\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{v_x}, \mathbf{v_y}, \mathbf{v_z})$ or $\rho(\mathbf{x})$ where $\mathbf{x} = (x, y, z, v_x, v_y, v_z)$. This the number of particles at position $\mathbf{x}$ in a region with volume $dV = dx^{2N} = dp^N dq^N$. We can consider the flux $\rho\mathbf{v}$ of particles out of each surface of a box that has volume dV. Integrating the flux over all surfaces of a box gives us the number of particles leaving the box per unit time.

$$\frac{dN}{dt} = \int_S \rho \mathbf{v} d\mathbf{A}$$

over the surface, S, of the box. Here the velocity $\mathbf{v} = \dot{\mathbf{x}}$. Using Gaus' law we can rewrite this as

$$\frac{dN}{dt} = \int_V \boldsymbol{\nabla} \cdot (\rho \mathbf{v}) dV$$

If particles are leaving the box, then the number density in the box must decrease.

$$\frac{dN}{dt} = -\int \frac{\partial \rho}{\partial t} dV$$

We find a conservation law

$$\frac{\partial \rho}{\partial t} + \boldsymbol{\nabla} \cdot (\rho \mathbf{v}) = 0$$

which can be recognized as conservation of mass for a fluid flow. Consider a bowling ball moving through water. As the bowling ball moves past an observer, it would appear that the density at some location is changing. However the bowling ball is not being compressed. The problem is that the above equation is written respect to a fixed coordinate system and we need to think about the density in a volume element that is moving with the fluid.

We can rewrite this equation as

$$\begin{aligned} \frac{\partial \rho}{\partial t} + (\mathbf{v}\cdot\boldsymbol{\nabla})\rho &= -\rho \boldsymbol{\nabla}\cdot\mathbf{v} \\ \frac{D\rho}{Dt} &= -\rho\boldsymbol{\nabla}\cdot\mathbf{v} \end{aligned} \tag{8}$$

where I have used

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} + \mathbf{v}\cdot\boldsymbol{\nabla}$$

is the advective derivative or that moving with the fluid. Consider $\rho(\mathbf{x}, t)$ as it varies

$$\begin{aligned} \frac{D\rho}{Dt} &= \frac{\partial\rho}{\partial x}\dot{x} + \frac{\partial\rho}{\partial y}\dot{y} + \frac{\partial\rho}{\partial z}\dot{z} + \frac{\partial\rho}{\partial t} \\ &= (\mathbf{v}\cdot\boldsymbol{\nabla})\rho + \frac{\partial\rho}{\partial t} \end{aligned}$$

The equation 8 implies that when $\boldsymbol{\nabla}\cdot\mathbf{v} = 0$ then $\frac{D\rho}{dt} = 0$ and the density of a distribution of particles remains fixed even as the distribution of particles is deformed.

Let us think about this in another way. What does it mean to have $\nabla\cdot\mathbf{v} = 0$ in a uniform density medium? In one dimension this means sense as particles never bunch together and the gradient of v is zero.

1.7 Volume conservation in Phase Space

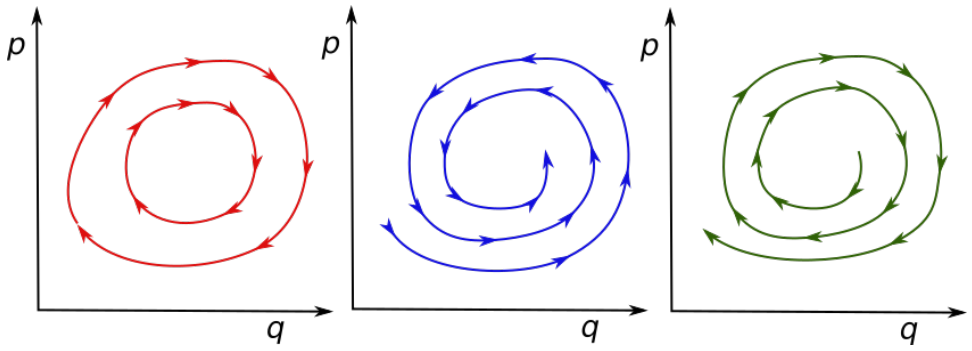

Figure 5: On left we show trajectories in a system where area is preserved. In the middle we show a system where volume contracts, as would occur when energy dissipation takes place. On the right the system gains energy.

In a two-dimensional system on the plane with density $\rho(x, y, t)$ conservation of mass is

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho \mathbf{v})$$

with $\mathbf{v} = (\dot{x}, \dot{y})$. We can show this by thinking of mass flux. We can also use Gaus's law to show this.

If the system is incompressible and density stays the same at all times then

$$\nabla \cdot \mathbf{v} = 0$$

or

$$\partial_x v_x + \partial_y v_y = \partial_x \dot{x} + \partial_y \dot{y} = 0$$

Now going back to our 2d integrated system with $\mathbf{x} = (y, v)$. The condition of incompressibility is

$$\partial_y \dot{y} + \partial_v \dot{v} = 0$$

Recall that $\dot{y} = v$ and so $\partial_y \dot{y} = 0$. Also $\dot{v} = -\partial_y U(y)$ is independent of v and so $\partial_v \dot{v} = 0$.

We can think of our two dimensional system as conserving density in phase space or being incompressible in phase space. This is related to energy conservation and that we have used a conservative force. If the force depends on velocity, then $\partial_v \dot{v} \neq 0$ and volume in phase space is not conserved. In a dissipating system, volume contracts and in a system that gains energy volume increases. See Figure 4.

1.8 Comparison between symplectic and non-symplectic first order integration for the Harmonic oscillator

A simple example is that of a harmonic oscillator. Setting the spring constant $k = 1$ and momentum per unit mass $p = \dot{q}$, the energy per unit mass is

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

The force per unit mass is $f(q) = -q$ and potential energy $U(q) = q^2/2$. The equations of motion are

$$\dot{q} = p$$
$$\dot{p} = -q$$

and together $\ddot{q} = -q$ giving us oscillation with angular frequency $\omega = 1$. These equations of motion are consistent with Newtons's equation with the potential U, or $\ddot{q} = -\partial_q U(q)$.

A general solution is

$$q(t) = A\cos(t - t_0)$$
$$p(t) = -A\sin(t - t_0) \tag{9}$$

We can expand this using trigonometric identities for the difference of sines and cosines

$$\begin{aligned} q(t) &= A\cos t\cos t_0 + A\sin t\sin t_0 \\ p(t) &= -A\sin t\cos t_0 + A\cos t\sin t_0 \end{aligned} \tag{10}$$

The initial condition consistent with equation 9

$$\begin{aligned} q_0 &= A\cos t_0 \\ p_0 &= A\sin t_0 \end{aligned}$$

lets us write equations 10 as

$$\begin{aligned} q(t) &= q_0\cos t + p_0\sin t \\ p(t) &= -q_0\sin t + p_0\cos t. \end{aligned}$$

The exact evolution of the system in matrix form

$$\begin{pmatrix} q(\tau) \\ p(\tau) \end{pmatrix} = \begin{pmatrix} \cos\tau & \sin\tau \\ -\sin\tau & \cos\tau \end{pmatrix} \begin{pmatrix} q(0) \\ p(0) \end{pmatrix}$$

where I replaced t with τ because τ will become our timestep!

We note that

$$\left|\det \begin{pmatrix} \cos\tau & \sin\tau \\ -\sin\tau & \cos\tau \end{pmatrix}\right| = \cos^2\tau + \sin^2\tau = 1$$

The matrix is the Jacobian of the transformation. Why is this? Think of the transformation as $\mathbf{x}' = A\mathbf{x}$. The Jacobian is a matrix composed of elements $\frac{\partial x_i'}{\partial x_j}$ and as the relation between $\mathbf{x}'$ and $\mathbf{x}$ is linear $\frac{\partial x_i'}{\partial x_j} = A_{ij}$. So the matrix is the Jacobian matrix of the transformation. And consequently if $|\det A| = 1$ the transformation is volume preserving. Another way to think about volume preservation is

$$\int_V dx^n = \int_V |\det J| dx'^n$$

If we expand the cosine and sine to first order in τ we can find a first order approximation to the evolution

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} 1 & \tau \\ -\tau & 1 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} \tag{11}$$

This is a first order integration scheme! However,

$$\det \begin{pmatrix} 1 & \tau \\ -\tau & 1 \end{pmatrix} = 1 + \tau^2 \neq 1.$$

The Jacobian is not equal to 1 so area is not preserved by the transformation.

After one timestep

$$q' = q + \tau p$$
$$p' = -\tau q + p$$

and the energy is

$$H' = \frac{1}{2}(p'^2 + q'^2) = \frac{1}{2}\left[(q+\tau p)^2 + (p - \tau q)^2\right]$$
$$= \frac{1}{2}(1+\tau^2)(p^2+q^2).$$

Since $1 + \tau^2$ must be greater than 1, the energy will increase every timestep. Volume in phase space has not been conserved. After many timesteps the trajectory will spiral outwards. Note that the determinant of the matrix in equation (11) is equal to $1 + \tau^2$ and is greater than 1 implying that the area increases.

A *symplectic* or area preserving in phase space scheme can be constructed with

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} 1 & \tau \\ -\tau & 1-\tau^2 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}. \tag{12}$$

The determinant of the transformation matrix is 1. The transformation is area preserving and so symplectic. This scheme and that of equation 11 only differ by τ^2. Since the scheme of equation 11 is accurate to first order this implies that the one in equation 12 is also accurate to first order.

Since the transformation of equation 12 is symplectic it does preserve *an* energy. However the quantity that is conserved is not the original energy H. The transformation gives

$$\begin{aligned} q' &= q + \tau p \\ p' &= -\tau q + (1-\tau^2)p \end{aligned}$$

giving energy after a single timestep of

$$2H' = q^2 + p^2 + \tau^2(q^2 - p^2) + \tau^4 p^2 + 2\tau^3 qp$$

Energy is not conserved. However by inserting q', p' into this

$$H_{integrated} = \frac{1}{2}(p^2+q^2) + \frac{\tau}{2}pq \tag{13}$$

we can show that this quantity is conserved and does not change. The integrated Hamiltonian $H_{integrated}$ (which is preserved) differs from the true Hamiltonian. The above Hamiltonian $H_{integrated}$ is called the "modified Hamiltonian." Just as finite differencing

techniques can better approximate a modified differential equation, a symplectic method preserves a Hamiltonian that differs from that intended.

$$H_{integrated} = H_{true} + H_{error}$$

where H_{true} is that for the harmonic oscillator, $H_{integrated}$ is that in equation (13) and the difference depends on the timestep

$$H_{error} = \frac{\tau}{2} pq$$

Since the $H_{integrated}$ is conserved, it is likely that the difference $H_{integrated} - H_{true}$ never gets very big. In this sense we expect the error to be *bounded* and not grow forever.

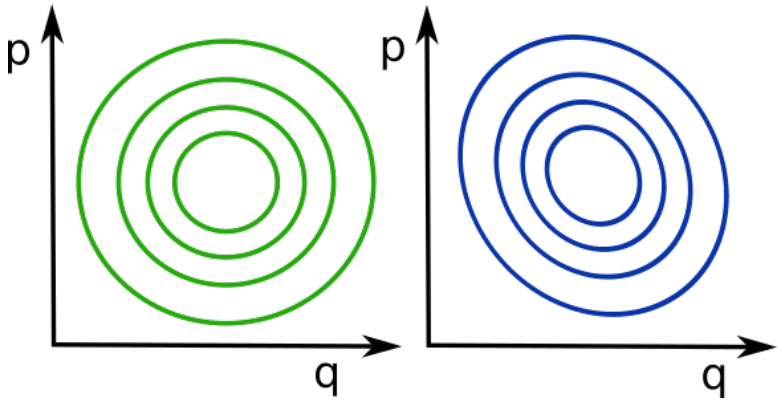

Figure 6: On the left, we show level curves of the Hamiltonian $H = (p^2 + q^2)/2$ for the harmonic oscillator. On the right we show level curves for $H_{integrated} = (p^2 + q^2)/2 + \frac{\tau}{2} pq$. Trajectories on the integrated Hamiltonian do not continuously diverge from those of the real system because they are confined to level curves.

The integrated Hamiltonian (equation 13) depends on the timestep used. If the timestep is changed then the integrated Hamiltonian *changes*. That means adaptive step-size integrators do not preserve a Hamiltonian (approximate to the true one) and so are no longer symplectic. This presents a limitation for symplectic integrators.

1.9 Comparison between 2-nd order Runge Kutta and leap-frog midpoint methods

The second order Runge-Kutta estimates the slope at the step midpoint and then advances to the next time. It always uses the position at beginning of each timestep to compute

the mid-point slope. In contrast the leap-frog method uses the previous midpoint position to estimate the position of the next midpoint. The slopes are taken from the midpoint to compute the full step and from the full step to advance the midpoints.

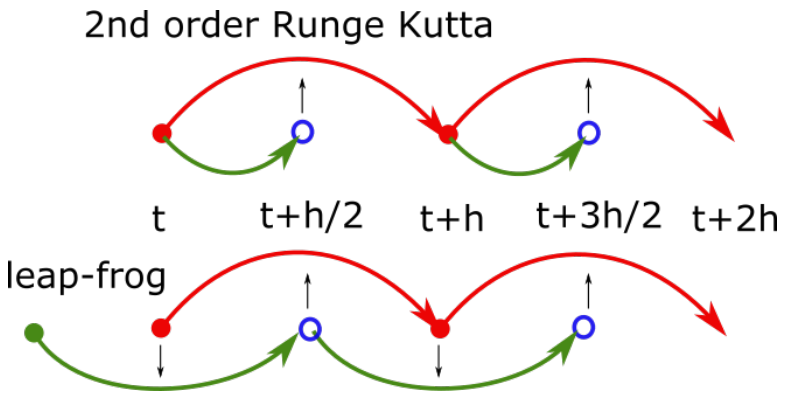

Figure 7: Illustrating the difference between midpoint methods. Both methods are second order, however the leapfrog is time reversible and the second order Runge-Kutta mehod is not.

Here is our previous midpoint (2nd order Runge-Kutta method)

$$x(t+h/2) = x(t) + \frac{1}{2}hf(x,t)$$
$$x(t+h) = x(t) + hf(x(t+h/2, t+h/2)).$$

This starts with $x(t)$.

The leap-frog method does this

$$x(t+h) = x(t) + hf(x(t+h/2, t+h/2))$$
$$x(t+\frac{3}{2}h) = x(t+h/2) + hf(x(t+h, t+h)).$$

It starts with values for $x(t)$ and $x(t+h/2)$.

Both methods are second order. However the leap-frog method has an advantage, it is time symmetric, which means that you integrate forwards and then backwards and get back to where you started. Since it is time symmetric, energy is unlikely to increase (or decrease) as that would imply a direction for time. Systems that conserve energy or volume in phase space are time symmetric. Whereas dissipative systems are not.

1.10 Integrating many particles - Accuracy and chaotic behavior

Consider the setting where there are a number of bodies interacting, for example massive bodies interacting via gravity (this is called N-body). To carry out an integration we need to compute accelerations on every particle. That means we need to know where every particle is when computing accelerations. One way to do this is to integrate all particles with the same step-size. It can be challenging to use a variable step size integrator, so more commonly a low order integrator is used on all particles during each timestep. The step size can be chosen so that it is appropriate for the particle with the largest acceleration. A single particle getting close to another particle can slow down an entire integration. If the step size is too large, then when a particle gets close to another particle it will see all of a sudden a huge force and it can be ejected from the system. This is particularly a problem if the force law is *steep* (such as for the Lennard-Jones potential).

When doing an N-body integration how would we know if we have an accurate simulation? We can check energy conservation summing all particles. We can check that total angular momentum is roughly conserved. When running the integration there will be drifts in both quantities and we will have to decide if they are preventing us from understanding the modeled system. Inaccuracy in the integration will give noise that should not be there.

N-body integrations are extremely sensitive to initial conditions (have short Lyapunov exponents). When the number of particles is more than a few, N-body simulations are swiftly diverging systems, so they rapidly loose memory of their initial conditions. An error is made during each timestep but we can ask: Is the system giving a trajectory that exists if we could integrate it exactly? If so then we don't necessarily need to worry about the inaccuracy of the integration. Arguments like these were hotly debated during the 80s when people started doing N-body integrations and wanted to interpret their results as being relevant for systems like galaxies.

2 Symplectic Integrators

2.1 Setting

Symplectic integrators are designed to preserve the geometry of phase space. A Hamiltonian system is integrated that is an approximation to the desired Hamiltonian. Because a Hamiltonian system is integrated and a quantity conserved (though not the real energy), errors are bounded. The integration exhibits stability on long (exponentially long) timescales. The primary astrophysical application of symplectic integrators has been long timescale integration of the few body problem, that is a few massive bodies interacting solely with Newtonian gravity.

2.2 Review of Hamiltonian Dynamics

A Hamiltonian system can be described in terms of coordinates $\mathbf{q}$ and momenta $\mathbf{p}$, both vectors. The Hamiltonian itself is a function of both and possibly time $H(\mathbf{p}, \mathbf{q}, t)$. Momenta and coordinates evolve according to

$$\frac{\partial H}{\partial p_i} = \frac{dq_i}{dt} \qquad \frac{\partial H}{\partial q_i} = -\frac{dp_i}{dt}$$

Hamiltonian systems have a special geometric property that volume in phase space is conserved during evolution. At each time coordinates and momenta are mapped to new coordinates and momenta in an area preserving manner.

Coordinates can be transformed into other coordinates via canonical transformations which preserve the symplectic property. Evolution itself is a canonical transformation. The Hamiltonian itself generates symplectic evolution of coordinates and momenta and a series of canonical transformations.

We can describe a Poisson bracket

$$\{x, y\} = \frac{\partial x}{\partial q_i}\frac{\partial y}{\partial p_i} - \frac{\partial x}{\partial p_i}\frac{\partial y}{\partial q_i}$$

Time derivatives of a quantity can be described in terms of a Poisson bracket with H

$$\frac{dz}{dt} = \{z, H\}$$

This follows because

$$\frac{dz}{dt} = \frac{\partial z}{\partial q}\dot{q} + \frac{\partial z}{\partial p}\dot{p} = \frac{\partial z}{\partial q}\frac{\partial H}{\partial p} - \frac{\partial z}{\partial p}\frac{\partial H}{\partial q}$$

Let us for short hand call D_H an evolution operator

$$\frac{dz}{dt} = \{z, H\} = D_H z$$

On longer timescales we can write z as a function of τ in terms of an exponential operator

$$z(\tau) = \exp(\tau D_H) z(0)$$

Suppose we can split our Hamiltonian

$$H = H_0 + H_1$$

Evolution can be described

$$z(\tau) = \exp[\tau(D_{H_0} + D_{H_1})] z(0)$$

Symplectic integration methods often involve splitting the Hamiltonian into more manageable pieces. by manageable we usually mean pieces that can be directly integrated and are easily solvable. One approach to developing a symplectic numerical approximation is to separate the Hamiltonian and construct a symplectic algorithm that approximates the original Hamiltonian based on these solvable pieces.

2.3 Examples of Evolution operators

2.3.1 Kinetic energy – Drifting

Consider

$$H = \frac{\mathbf{p}^2}{2}$$

consisting only of a kinetic energy term. Hamilton's equations give

$$\dot{\mathbf{q}} = \mathbf{p}$$

with solution at a later time τ

$$\begin{aligned} \mathbf{q}(\tau) &= \mathbf{q}_0 + \mathbf{p}_0\tau \\ \mathbf{p}(\tau) &= \mathbf{p}_0 \end{aligned}$$

where p_0, q_0 are the initial momenta and coordinates. The transformation happens to be linear. We can also write this as

$$\mathbf{z}(\tau) = \begin{pmatrix} 1 & \tau \\ 0 & 1 \end{pmatrix} \mathbf{z}_0 = \left(\mathbf{I} + \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \tau \right) \mathbf{z}_0$$

where $\mathbf{z} = (\mathbf{q}, \mathbf{p})$. The above form makes it clearer that the infinitesimal operator for drifting is the matrix

$$L_{drift} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

It is probably better to use an operator to describe the infinitesimal operator as in general we don't expect to get a linear operator. Taking derivatives

$$\begin{aligned} \frac{\partial \mathbf{q}(\tau)}{\partial \tau} &= \mathbf{p}_0 \\ \frac{\partial \mathbf{p}(\tau)}{\partial \tau} &= 0 \end{aligned}$$

so our infinitesimal drift operator is

$$L_{drift} = \mathbf{p} \frac{\partial}{\partial \mathbf{q}} \tag{14}$$

(a dot product).

2.3.2 Potential energy – Interactions or Kicks

Potential energy terms arise from coordinate dependent interactions.

$$H = V(\mathbf{q})$$

For example gravitational interactions between N particles with masses m_i and coordinates $\mathbf{q}_i$

$$V = \sum_{i>j}^{N} \frac{G m_i m_j}{|\mathbf{q}_i - \mathbf{q}_j|}$$

where G is the gravitational constant. Hamilton's equations gives

$$\dot{\mathbf{p}}_i = -\frac{\partial V}{\partial \mathbf{q}_i}$$

After time τ

$$\begin{aligned} \mathbf{q}_i(\tau) &= \mathbf{q}_{0,i} \\ \mathbf{p}_i(\tau) &= \mathbf{p}_{0,i} - \frac{\partial V}{\partial \mathbf{q}_i}\tau \end{aligned}$$

We note here that only momenta are changed. We sometimes call this evolution a "kick." This contrasts with the evolution operator due to the kinetic energy which only changes coordinates that we called a "drift". Hamiltonians that divide into two pieces, one only dependent on momenta and the other only dependent on coordinates are called "separable." Symplectic algorithms can be simpler for separable systems. In some cases implicit algorithms can be more easily expressed as explicit.

Taking derivatives

$$\begin{aligned} \frac{\partial \mathbf{q}(\tau)}{\partial \tau} &= 0 \\ \frac{\partial \mathbf{p}(\tau)}{\partial \tau} &= -\frac{\partial V}{\partial \mathbf{q}} \end{aligned}$$

giving an operator

$$L_{kick} = -\frac{\partial V}{\partial \mathbf{q}} \frac{\partial}{\partial \mathbf{p}} \tag{15}$$

2.3.3 Keplerian Evolution

We consider the N-body problem with gravity. For planetary systems one mass dominates the others. With a suitable choice of coordinates the Hamiltonian can be written as a sum of terms

$$H = H_{Kep} + H'$$

where the dominant term

$$H_{Kep} = \sum_i \frac{\mathbf{p}_i^2}{2m_i} - \frac{Gm_0 m_i}{|\mathbf{q}_i - \mathbf{q}_0|}$$

Here m_i are the masses of planets with positions $\mathbf{q}_i$. The central mass (the Sun) has coordinate $\mathbf{q}_0$ and mass m_0. The deviation H' is smaller than the dominant Keplerian evolution term. Keplerian evolution is done with fewer computations using f, g functions. A differential form of Kepler's equation must still be solved but not all orbital elements determined.

Starting with position $\mathbf{x}_0$ and velocity $\mathbf{v}_0$ we calculate the new position $\mathbf{x}_1$ and $\mathbf{v}_1$ after time τ.

$$\begin{aligned} \mathbf{x}_1 &= f\mathbf{x}_0 + g\mathbf{v}_0 \\ \mathbf{v}_1 &= \dot{f}\mathbf{x}_0 + \dot{g}\mathbf{v}_0 \end{aligned}$$

where the functions f, g need to be calculated and depend on the timestep τ. Here I summarize some of the equations used to find f, g in the form of a recipe taken from the book by Prussing and Conway.

Given the initial conditions and the timestep we first must solve a differential Kepler's equation for the quantity x. If you don't know if all particles in your problem are bound then it is a good idea to use a version that does not assume that the particle is in an eccentric orbit. This means your solver should work for parabolic or hyperbolic orbits as well as eccentric orbits. An iterative procedure must be used to solve the universal differential form of Kepler's equation. However the procedure for doing this can converge rapidly. I have found that 6 to 7 iterations of the Laguerre method as described by Prussing and Conway converges within double precision for all orbital positions that I have tried.

We first define some quantities that will be used more than once The inverse of the original semi-major axis

$$\frac{1}{a} = \alpha = \frac{2}{r_0} - \frac{v_0^2}{m_0}$$

where r_0 is the original radius.

$$\sigma_0 = \frac{\mathbf{x}_0 \cdot \mathbf{v}_0}{\sqrt{m_0}}$$

Solve iteratively the following version of the differential Kepler's equation to find x such that

$$F = \sigma_0 x^2 C_p(\alpha x^2) + (1 - r_0\alpha)x^3 S_p(\alpha x^2) + r_0 x\sqrt{m_0}\tau = 0$$

The f, g functions can be written in terms of x.

$$\begin{aligned} r &= \sigma_0 x(1-\alpha x^2 S_p(\alpha x^2)) + (1-r_0\alpha)x^2 C_p(\alpha x^2) + r_0 \\ f &= 1 - C_p(\alpha x^2)\frac{x^2}{r_0} \\ g &= \tau - S_p(\alpha x^2)\frac{x^3}{\sqrt{m_0}} \\ \dot{f} &= \frac{x\sqrt{m_0}}{r r_0}\left(\alpha x^2 S_p(\alpha x^2) - 1\right) \\ \dot{g} &= 1 - C_p(\alpha x^2)\frac{x^2}{r} \end{aligned}$$

though angular momentum conservation can be used to solve for one of these from the other three. The angular momentum after the timestep

$$\mathbf{L}_1 = (f\dot{g} - g\dot{f})(\mathbf{x}_0 \times \mathbf{v}_0) = (f\dot{g} - g\dot{f})\mathbf{L}_0$$

Conservation of angular momentum gives the condition

$$f\dot{g} - g\dot{f} = 1$$

which can be used to solve for one of the functions in terms of the others.

The associated functions needed to compute the f, g functions using $u \equiv \sqrt{|y|}$

$$C_p(y) = \begin{cases} \frac{\cosh u - 1}{y} & \\ \frac{1}{2}(1 - \frac{y}{12}(1 - \frac{y}{30}(1 - \frac{y}{56}(1 - \frac{y}{90})))) & \text{for} \\ \frac{1-\cos u}{y} & \end{cases} \begin{array}{l} y < 0 \\ y \sim 0 \\ y > 0 \end{array}$$

$$S_p(y) = \begin{cases} \frac{\sinh u - u}{u^3} & \\ \frac{1}{6}(1 - \frac{y}{20}(1 - \frac{y}{42}(1 - \frac{y}{72}(1 - \frac{y}{110})))) & \text{for} \\ \frac{u - \sin u}{u^3} & \end{cases} \begin{array}{l} y < 0 \\ y \sim 0 \\ y > 0 \end{array}$$

This recipe should be checked for accuracy.

2.4 Comparison between symplectic and non-symplectic first order integration for the Harmonic oscillator

A simple example is that of a harmonic oscillator

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

The exact evolution of the system is

$$\begin{pmatrix} q(\tau) \\ p(\tau) \end{pmatrix} = \begin{pmatrix} \cos\tau & \sin\tau \\ -\sin\tau & \cos\tau \end{pmatrix} \begin{pmatrix} q(0) \\ p(0) \end{pmatrix}$$

If we expand the cosine and sine to first order in τ we can find a first order approximation to the evolution

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} 1 & \tau \\ -\tau & 1 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix}$$

However after one timestep the energy is

$$H' = \frac{1}{2}(p'^2 + q'^2) = \frac{1}{2}(1+\tau^2)(p^2+q^2)$$

and we note that the energy has increased. Since $1+\tau^2$ must be greater than 1, the energy will increase every timestep. Volume in phase space has not been conserved. After many timesteps the trajectory will spiral outwards. Note that the determinant of the matrix in equation (2.4) is equal to $1+\tau^2$ and is greater than 1 implying that the area increases.

A symplectic scheme can be constructed with

$$\begin{pmatrix} q' \\ p' \end{pmatrix} = \begin{pmatrix} 1 & \tau \\ -\tau & 1-\tau^2 \end{pmatrix} \begin{pmatrix} q \\ p \end{pmatrix} \tag{16}$$

Because we modified our previous scheme (equation 11) with τ^2, the scheme is still accurate to first order. The determinant of the transformation matrix is 1. The transformation is area preserving and so is symplectic.

Since the transformation is symplectic it does preserve *an* energy. However the quantity that is conserved is not the original H. The transformation gives

$$\begin{aligned} q' &= q + \tau p \\ p' &= -\tau q + (1-\tau^2)p \end{aligned}$$

giving energy after a single timestep of

$$2H' = q^2 + p^2 + \tau^2(q^2 - p^2) + \tau^4 p^2 + 2\tau^3 qp.$$

Energy is not conserved. However by inserting q', p' into this

$$H_{integrated} = \frac{1}{2}(p^2+q^2) + \frac{\tau}{2}pq, \tag{17}$$

we can show that this quantity $H_{integrated}$ is conserved and does not change. The integrated Hamiltonian $H_{integrated}$ and that is preserved differs from the true Hamiltonian. The above Hamiltonian is called the "modified Hamiltonian." Just as finite differencing techniques can better approximate a modified differential equation, a symplectic method preserves a Hamiltonian that differs from that intended.

$$H_{integrated} = H_{true} + H_{error}$$

where H_{true} is that for the harmonic oscillator, $H_{integrated}$ is that in equation (17) and the difference depends on the timestep

$$H_{error} = \frac{\tau}{2} pq$$

The integrated Hamiltonian depends on the timestep used. If the timestep is changed then the integrated Hamiltonian *changes*. That means adaptive step-size integrators do not preserve a Hamiltonian (approximate to the true one) and so are no longer symplectic. This presents a limitation for symplectic integrators.

2.5 Low Order Symplectic integrators – Störmer-Verlet (aka leapfrog)

For a simple evolution equation

$$\frac{dz}{dt} = f(z)$$

An Euler method is

$$z^{n+1} = z^n + \tau f'(z^n)$$

It is explicit rather than implicit. For a Hamiltonian system we need to update both momenta and positions. The order of the update is important. For first order integrations if the coordinates are updated first the method is called the asymmetric Euler A method. If the momenta are updated first the method is called the Euler B method.

For a separable Hamiltonian

$$H(\mathbf{p}, \mathbf{q}) = \mathbf{p}^T \mathbf{M} \mathbf{p} + V(\mathbf{q})$$

with symmetric inverse mass matrix $\mathbf{M}$. Hamilton's equations give

$$\dot{\mathbf{q}} = \mathbf{M}\mathbf{p} \qquad \dot{\mathbf{p}} = -\nabla_q V(\mathbf{q})$$

The Störmer-Verlet method

$$\begin{aligned} \mathbf{p}^{n+\frac{1}{2}} &= \mathbf{p}^n - \frac{\tau}{2} \nabla_q V(\mathbf{q}^n) \\ \mathbf{q}^{n+1} &= \mathbf{q}^n + \tau \mathbf{M} \mathbf{p}^{n+1/2} \\ \mathbf{p}^{n+1} &= \mathbf{p}^{n+\frac{1}{2}} - \frac{\tau}{2} \nabla_q V(\mathbf{q}^{n+1}) \end{aligned}$$

Momenta are evaluated at half steps. When the half step at either end is combined, the method can be called the leapfrog method and only requires evaluating the forces once per step. It turns out this method is 2nd order and symplectic.

You will notice that the first piece is a kick step, the second is a drift step and the third is a kick step. The above is a series of transformations, each which is symplectic. The first and last transformations evolve the potential energy Hamiltonian by $\tau/2$. The

middle transformation evolve the kinetic energy term by τ. Each piece is a symplectic transformation so the series is a symplectic transformation. We will show below that if the Hamiltonian is partitioned into any two pieces that the above procedure is a second order approximation with error dependent on the commutator of the two pieces.

The order of the half steps can be reversed giving

$$\begin{aligned} \mathbf{q}^{n+\frac{1}{2}} &= \mathbf{q}^n + \frac{\tau}{2}\mathbf{M}\mathbf{p}^n \\ \mathbf{p}^{n+1} &= \mathbf{p}^n - \tau\nabla_q V(\mathbf{q}^{n+1/2}) \\ \mathbf{q}^{n+1} &= \mathbf{q}^{n+\frac{1}{2}} + \frac{\tau}{2}\mathbf{M}\mathbf{p}^{n+1}. \end{aligned}$$

In this case the forces must be evaluated in the middle of the computation.

2.6 More on the exponential operator

A linear dynamical system

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$$

with constant matrix $\mathbf{A}$ and $\mathbf{x} \in \mathbf{R}^n$ has solution

$$\mathbf{x}(t) = e^{t\mathbf{A}}\mathbf{x}_0$$

where the matrix

$$e^{t\mathbf{A}} = 1 + t\mathbf{A} + \frac{1}{2!}t^2\mathbf{A}^2 + \frac{1}{3!}t^3\mathbf{A}^3 ...$$

Now for a non-linear dynamical system

$$\dot{\mathbf{x}} = \mathbf{A}(\mathbf{x})$$

we can consider the Taylor series of a function $f(\mathbf{x}(t))$. To first order in t

$$\begin{aligned} f(\mathbf{x}(t)) &= f(x_0) + \boldsymbol{\nabla} f(x_0) \cdot \dot{\mathbf{x}} t + ... \\ &= f(x_0) + \boldsymbol{\nabla} f(x_0) \cdot \mathbf{A}(x_0) t + ... \end{aligned}$$

With $\mathbf{A} = \sum_i A_i \hat{\mathbf{e}}_i$ notice that we can write

$$\boldsymbol{\nabla} f(x_0) \cdot \mathbf{A}(x_0) t = \sum_i A_i \frac{\partial}{\partial x_i} f$$

Let us think of our vector function as an operator

$$\mathbf{L}_\mathbf{A} = \sum_i A_i \frac{\partial}{\partial x_i}$$

It is possible to expand further and write

$$f(\mathbf{x}(t)) = f(x_0) + t\mathbf{L_A} f(x_0) + \frac{1}{2!}t^2 \mathbf{L_A}^2 f(x_0) + \frac{1}{3!}t^3 \mathbf{L_A}^3 f(x_0) \quad ...$$

where the operator $\mathbf{L_A}$

$$\mathbf{L_A} f = A_i \frac{\partial f}{\partial x_i} \qquad \mathbf{L_A}^2 f = A_i \frac{\partial}{\partial x_i}\left(A_j \frac{\partial f}{\partial x_j}\right) ...$$

We can consider the exponential $e^{t\mathbf{L_A}}$ so that

$$f(\mathbf{x}(t)) = e^{t\mathbf{L_A}} f(x_0)$$

Time evolution operators don't in general commute.

Often notation is used where the vector itself is considered an operator $\mathbf{A}$.

2.7 An example of a second order symplectic integrator

Consider evolution over a timestep τ. The actual evolution over the timestep involves the evolution of

$$\exp[\tau(A+B)]$$

where A, B are evolution operators that arise from splitting a Hamiltonian. Any product of symplectic evolution operators will give a symplectic transformation. However not all products will give a good approximation to the desired Hamiltonian. It is possible to construct an order $o(\tau^n)$ approximation with individual evolution operators applied sequentially

$$\exp[\tau(A+B)] = \prod_{i=1}^{n} \exp(c_i \tau A) \exp(d_i \tau B) + o(\tau^{n+1})$$

The coefficients c_i, d_i can be found for any order approximation (e.g., Yoshida). High order (greater than 2nd) symplectic methods were first developed by E. Forest, Campostrini & Rossi, Candy & Rozmous, Suzuki, Yoshida, and Creutz and Gocksch at around 1990.

For example if a second order approximation is desired, the following can be used as an approximation

$$\exp\frac{\tau}{2}A \ \exp\tau B \ \exp\frac{\tau}{2}A = \tag{18}$$

$$\begin{aligned} &= (1 + \frac{\tau}{2}A + \frac{\tau^2}{8}A^2)(1 + \tau B + \frac{\tau^2}{2}B^2)(1 + \frac{\tau}{2}A + \frac{\tau^2}{8}A^2) + o(\tau^3) \\ &= 1 + \tau(A+B) + \frac{\tau^2}{2}(A^2 + B^2) + \frac{\tau^2}{2}(AB + BA) + o(\tau^3) \end{aligned}$$

To order τ^3 this is equivalent to

$$\exp[\tau(A+B)] + o(\tau^3) \tag{19}$$

Thus the difference between the above approximation (equation 18) and the actual evolution of the Hamiltonian is $o(\tau^3)$.

For the separable Hamiltonian, the Stormer-Verlet method can be described in terms of evolution by $\tau/2$ of the momenta followed by evolution by τ of the coordinates and then again by $\tau/2$ of the momenta.

Let us look at the remaining terms. The difference between the real Hamiltonian and the modified one will consist of commutators involving 3 operators. For the Keplerian systems, one operator, the Keplerian one is not small. The others, the drift term (if any) and the interaction terms depend on the rate of the planet masses to the stellar mass. Thus the difference between the real Hamiltonian and the modified one is order $\mu\tau^3$ unless the interaction terms get large (close encounters).

How large is the error? We need to consider the third order terms in equation 18.

$$o(\tau^3) = \tau^3 \left[\frac{ABA}{4} + \frac{1}{6}(A^3 + B^3) + \frac{1}{4}(AB^2 + B^2A) + \frac{1}{8}(A^2B + BA^2) \right]$$

Compare this to

$$\frac{\tau^3}{3!}(A+B)^3 = \frac{\tau^3}{6}\left[A^3 + B^3 + BAB + ABA + AB^2 + B^2A + A^2B + BA^2\right]$$

Subtract the two

$$\begin{aligned} o(\tau^3) - \frac{\tau^3}{3!}(A+B)^3 &= \tau^3 \left(\frac{1}{12}ABA - \frac{1}{6}BAB + \frac{1}{12}(AB^2 + BA^2) - \frac{1}{24}(A^2B + BA^2) \right) \\ &= \tau^3 \left(\frac{1}{24}[A,[B,A] - \frac{1}{12}[B,[A,B]] \right) \end{aligned}$$

with coefficients that we can find using

$$[A,[B,A]] = 2ABA - A^2B - BA^2$$
$$[B,[A,B]] = 2BAB - B^2A - AB^2$$

Thus the error in the approximated evolution operator is

$$\tau^3 \left(\frac{1}{24}[A,[B,A] - \frac{1}{12}[B,[A,B]] \right) \tag{20}$$

and this can be computed using Poisson brackets.

Remark The difference will always depend on commutators as there is no difference if A, B commute. The error of the integrator depends on the commutators.

Remark The Störmer-Verlet integrator is in the form of Equation 18 and so is a second order symplectic integrator. The error can be estimated using the commutators in equation 20 and the operators for the kick and drift steps (Equations 14 and 15).

2.8 Regularization

In their first incarnation in celestial mechanics symplectic integrators were used with large timesteps (orbit long ones, e.g., Wisdom and Holman) and interactions were small. As long as interactions remain small, their effect is primarily important over many orbits. However if interactions vary rapidly then a long timestep integrator will fail to emulate the dynamics of the real system. One example is stars near a black hole in a galactic center. When stars approach the black hole the dynamical timescale can be much shorter than needed for integration outside the black hole's sphere of influence. Another example is an integration of a planetary system with an object on a highly eccentric orbit or when there are close encounters.

One procedure for constructing an integrator with a timestep that depends on approach distance is called "regularization." Papers that discuss this are Preti & Tremaine (1999), Mikkola & Tanikawa (1999), but also see Dehnen (2009) on improving the epicyclic approximation using regularization. Also relevant is the KS transformation where KS is Kustaanheimo-Stiefel regularization which employs a transformation known as a Levi-Civita transformation.

2.8.1 Extended Phase space

Cannonical transformations involve transformation of coordinates and momenta. However it is harder to transform time. For example let's look at a Hamiltonian and multiply it by a constant, a. Consider the new Hamiltonian

$$K(p,q) = aH(p,q)$$

Hamilton's equations

$$\begin{aligned} \frac{dq}{dt} &= \frac{\partial H}{\partial p} = \frac{1}{a}\frac{\partial K}{\partial p} \\ \frac{dp}{dt} &= -\frac{\partial H}{\partial p} = -\frac{1}{a}\frac{\partial K}{\partial q} \end{aligned}$$

Let

$$\frac{dt}{d\tau} = a \tag{21}$$

Multiplying both sides of Hamilton's equation by $\frac{dt}{d\tau}$ we find

$$\begin{aligned} \frac{dq}{d\tau} &= \frac{\partial K}{\partial p} \\ \frac{dp}{d\tau} &= -\frac{\partial K}{\partial q} \end{aligned}$$

We find that K satisfies Hamilton's equation but with time rescaled by the factor a. So K is equivalent to H but with a rescaling of time. The procedure of regularization is similar but we will multiply the original Hamiltonian by a function $g(p,q)$.

$$\frac{dt}{d\tau} = g(p,q) \tag{22}$$

Now consider the function

$$K(p,q) = g(p,q)(H - H_0) \tag{23}$$

where H_0 is a constant that we will discuss below. Taking derivatives

$$\begin{aligned} \frac{\partial K}{\partial q} &= \frac{\partial g}{\partial q}(H-H_0) + g\frac{\partial H}{\partial q} = \frac{\partial g}{\partial q}(H-H_0) - g\frac{dp}{dt} = \frac{\partial g}{\partial q}(H-H_0) - \frac{dp}{d\tau} \\ \frac{\partial K}{\partial p} &= \frac{\partial g}{\partial p}(H-H_0) + g\frac{\partial H}{\partial p} = \frac{\partial g}{\partial p}(H-H_0) + g\frac{dq}{dt} = \frac{\partial g}{\partial p}(H-H_0) + \frac{dq}{d\tau} \end{aligned} \tag{24}$$

If we choose H_0 to be equal to H at $t = 0$ then we find that K satisfies Hamilton's equation but with time rescaled by the function g. In doing an integration we can log the initial energy and so chose to develop a symplectic integrator for K rather than for H. This will allow us to take equal τ timesteps rather than equal t timesteps. This allows us to make an integrator that has timestep depend on radius or semi-major axis, for example.

More formally if we desire coordinate transformations involving time we can make time act like a new coordinate. This means we need a new momentum. This description is called **extended phase space**. Let t be a new coordinate with associated momentum T. Choose our extended Hamiltonian

$$K(p,q,t,T) = H(p,q,t) + T \tag{25}$$

The extended Hamiltonian is only a function of coordinates and momentum.

$$\frac{\partial K}{\partial T} = 1 = \frac{dt}{dt} \tag{26}$$

$$\frac{\partial K}{\partial t} = \frac{\partial H}{\partial t} = -\frac{dT}{dt} \tag{27}$$

However Hamilton's equations on H implies that $\frac{dH}{dt} = \frac{\partial H}{\partial t}$.

$$T(t) = T(0) + \int_0^t \frac{dT}{d\tau} d\tau = T(0) - \int_0^t \frac{dH}{d\tau} d\tau \tag{28}$$

If we choose that $T(0) = -H(0)$ then the above implies that

$$T(t) = -H(t) \tag{29}$$

As long as we set $T(0)$ equal to the negative initial energy (even if system is time dependent) we will find that $K = 0$ at all times.

Remark A brilliant application of regularization is to find good low order approximations for eccentric galactic orbits, by Dehnen 99.

2.9 Force-Gradient Algorithms

If you separate the Hamiltonian into kinetic and potential energy parts

$$H(\mathbf{p},\mathbf{q}) = \frac{1}{2}p^2 + V(\mathbf{q})$$

with $T = p^2/2$ and operators

$$A = \{, T\} = p_i \frac{\partial}{\partial q_i}$$

$$B = \{, V\} = f_i \frac{\partial}{\partial p_i}$$

with force $f_i = -\frac{\partial V}{\partial q_i}$. The commutator

$$\{, \{\{V,T\}, V\}\} = [B,[A,B]] = [V,[T,V]] = \nabla_i |\mathbf{F}|^2 \frac{\partial}{\partial p_i}$$

where the force $\mathbf{F} = -\boldsymbol{\nabla} V(\mathbf{q})$. The commutator has a gradient of the force in it and so is used to make what is called force gradient algorithms that are accurate to fourth order.

We know that

$$T_2(\tau) = e^{\frac{T\tau}{2}} e^{V\tau} e^{\frac{T\tau}{2}} = e^{\tau(T+V)} + O(\tau^3)$$

is a second order approximation. Note that T_2 is time reversible as

$$T(\tau)T(-\tau) = 1$$

This also implies that terms in the expansion for τ must be odd. Time reversal symmetry can be used to cancel a higher order term by considering steps forward and backward. However it is desirable to make an integrator that does involve forwards and backwards steps. The third order terms involve commutators such as a given above. Taking into account the commutator it is possible to devise 4th order algorithms that are remarkably simple to compute (Forest and Ruth and work by Chin).

For example, an integrator that exploits the force gradient is

$$e^{\tau(T+V)} = e^{\frac{T\tau}{6}} e^{\frac{3V\tau}{8}} e^{\frac{T\tau}{3}} e^{\frac{\tilde{V}\tau}{4}} e^{\frac{T\tau}{3}} e^{\frac{3V\tau}{8}} e^{\frac{T\tau}{6}} + O(\tau^5)$$

with

$$\tilde{V} = V + \tau^2 [V,[T,V]]$$

The algorithm can be read off directly and is

$$\begin{aligned} q_1 &= q_0 + \frac{\tau}{6} p_0 \\ p_1 &= p_0 + \frac{\tau 3}{8} F(q_1) \\ q_2 &= q_1 + \frac{\tau}{3} p_1 \\ p_2 &= p_1 + \frac{\tau}{4}\left[F(q_2) + \frac{\tau^2}{48} \nabla |F(q)|^2 \right] \\ q_3 &= q_2 + \frac{\tau}{3} p_2 \\ p_3 &= p_2 + \frac{\tau 3}{8} F(q_3) \\ q_4 &= q_3 + \frac{\tau}{6} p_3 \end{aligned}$$

The algorithm (by Chin) is remarkably efficient in terms of calculations for a high order algorithm.

See work by Forrest, Ruth, Chin, Yoshida, Susuki ...

2.10 Symplectic Runge Kutta Methods

The general formulation for an s-stage Runge Kutta method is

$$\mathbf{Q}_i = \mathbf{q}_i^n + \tau \sum_{j=1}^{s} a_{ij} \mathbf{F}_j \qquad \mathbf{F}_i = \nabla_p H(\mathbf{Q}_i, \mathbf{P}_i), \qquad i \in 1, ...s \tag{30}$$

$$\mathbf{P}_i = \mathbf{p}_i^n + \tau \sum_{j=1}^{s} a_{ij} \mathbf{G}_j \qquad \mathbf{G}_i = -\nabla_q H(\mathbf{Q}_i, \mathbf{P}_i), \qquad i \in 1, ...s \tag{31}$$

$$\mathbf{q}^{n+1} = \mathbf{q}^n + \tau \sum_{i=1}^{s} b_i \mathbf{F}_i \tag{32}$$

$$\mathbf{p}^{n+1} = \mathbf{p}^n + \tau \sum_{i=1}^{s} b_i \mathbf{G}_i \tag{33}$$

Here $\mathbf{P}, \mathbf{Q}$ are computed at intermediate stages for a large step updating $\mathbf{p}, \mathbf{q}$. The system is explicit if $a_{ij} = 0$ for $j \geq i$. The condition for such a method to be symplectic is

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0 \tag{34}$$

as shown by Sanz-Serna, Lasagni and Suris. This condition implies that Runge Kutta methods are necessarily implicit.

If different coefficients are used for p, q separately then the method is called partitioned Runge-Kutta method. For these explicit techniques can be found when the Hamiltonian is separable.

2.11 Democratic heliocentric coordinates

Heliocentric coordinates $\mathbf{Q}_i$ are

$$\mathbf{Q}_0 = \frac{1}{m_{tot}} \sum_{j=0}^{N} m_j \mathbf{q}_j \tag{35}$$

$$\mathbf{Q}_i = \mathbf{q}_i - \mathbf{q}_0 \tag{36}$$

The reverse transform

$$\mathbf{q}_0 = \mathbf{Q}_0 - \frac{1}{m_{tot}} \sum_{j=1}^{N} m_j \mathbf{Q}_j \tag{37}$$

$$\mathbf{q}_i = \mathbf{Q}_i + \mathbf{Q}_0 - \frac{1}{m_{tot}} \sum_{j=1}^{N} m_j \mathbf{Q}_j \tag{38}$$

So that the coordinates are cannonical the associated momenta

$$\mathbf{P}_0 = \sum_{j=0}^{N} \mathbf{p}_j \tag{39}$$

$$\mathbf{P}_i = \mathbf{p}_i - \frac{m_i}{m_{tot}} \sum_{j=0}^{N} \mathbf{p}_j \tag{40}$$

The inverse transform is

$$\mathbf{p}_0 = \frac{m_0}{m_{tot}} \mathbf{P}_0 - \sum_{j=1}^{N} \mathbf{P}_j \tag{41}$$

$$\mathbf{p}_i = \mathbf{P}_i + \frac{m_i}{m_{tot}} \mathbf{P}_0. \tag{42}$$

How are the above transformations found? The total Hamiltonian for N interacting bodies

$$H = \sum_i \frac{\mathbf{p}_i^2}{2m_i} - \sum_{i \neq j} \frac{G m_i m_j}{2|\mathbf{q}_i - \mathbf{q}_j|} \tag{43}$$

Take a generating function

$$F_2(\mathbf{r}_i, \mathbf{P}_i) = \sum_{i>0} (\mathbf{q}_i - \mathbf{q}_0) \cdot \mathbf{P}_i + \mathbf{r}_0 \cdot \mathbf{P}_0 \tag{44}$$

New coordinates and momenta can be found in terms of old coordinates and momenta using the generating function

$$\begin{aligned}
\frac{\partial F_2}{\partial \mathbf{P}_i} &= (\mathbf{q}_i - \mathbf{q}_0) = \mathbf{Q}_i \qquad i \neq 0 && (45)\\
\frac{\partial F_2}{\partial \mathbf{q_i}} &= \mathbf{P}_i = \mathbf{p}_i \qquad i \neq 0 && (46)\\
\frac{\partial F_2}{\partial \mathbf{P}_0} &= \mathbf{q}_0 && (47)\\
\frac{\partial F_2}{\partial \mathbf{q}_0} &= \mathbf{P}_0 - \sum_{i>0} \mathbf{P}_i = \mathbf{p}_0 && (48)
\end{aligned}$$

$$H = \sum_{i>0} \left(\frac{P_i^2}{2m_i} - \frac{Gm_i m_0}{|\mathbf{Q}_i|} \right) - \sum_{i \neq j \;\; i,j>0} \frac{Gm_i m_j}{2|\mathbf{q}_i - \mathbf{q}_j|} + \frac{1}{2m_0} (P_0 - \sum_{i>0} P_i)^2 \tag{49}$$

Setting the center of momentum to be zero and we find

$$H = \sum_{i>0} \left(\frac{P_i^2}{2m_i} - \frac{Gm_i m_0}{|Q_i|} \right) - \sum_{i \neq j \;\; i,j>0} \frac{Gm_i m_j}{2|r_i - r_j|} + \frac{1}{2m_0} (\sum_{i>0} P_i)^2 \tag{50}$$

We see that the Hamiltonian is a sum of three terms, a Keplerian part, an interaction part and a drifting part due to the non-inertial coordinate system. This subdivision is used in the so called democractic heliocentric and barycentric momenta symplectic integrators by Duncan et al. and Chambers et al. The Holman and Wisdom integrator instead used *Jacobi coordinates* and resulting in only two terms, a Keplerian part and an interaction part.

2.12 Using Delta functions

Consider the simple system

$$H(I, \phi) = I\omega + A(I) \cos \phi \tag{51}$$

Think about adding a series of fast higher order terms that all average to zero

$$H \rightarrow I\omega + A(I) \sum_m \cos(\phi + 2\phi m t) = I\omega + A(I) 2\pi \cos\phi \; \delta(t = 2\pi m) \tag{52}$$

When we are not at a time where the delta function contributes the system is integrable

$$\dot{\phi} = \omega \tag{53}$$

When we are at the delta function $t = 0, 2\pi, 4\pi..$ we integrate over the delta function

$$\dot{I} = -A(I_0) 2\pi \sin\phi \; \delta(t = 2\pi m) \qquad \rightarrow \qquad \Delta I = A(I_0) 2\pi \sin\phi \tag{54}$$

Procedure: integrate integrable Hamiltonian between momenta kicks. The difference between the real Hamiltonian and the modified one is the sum of the extra terms added. This procedure is used to create the standard map.

Remark Conjunction integrators are those where planetary perturbations are only taken into account during conjunctions. These give remarkably similar phenomena to more accurately integrated systems.

2.13 Discussion

Symplectic integrators have been very successful at long timescale integrations of the solar system. There are a variety of reasons for this. The system is approximately Keplerian. The Keplerian evolution operator can be computed within double precession for any step size. Of order 10^{15} steps need to be taken before round off error (if drifting in 1 direction) gives a significant error. Planetary perturbations are not combined with the perturbation from the star or Sun. The force of the Sun does not swamp the planetary perturbations.

The difference between the modified Hamiltonian and the true Hamiltonian is small. We can describe the system as $H_{mod} = H_{true} + \epsilon H_{err}$. For a small ϵ set by the step size the divergence between the two systems takes an exponentially long timescale. This I think can be proven in ways similar to proofs of Nekhoroshev's theorem. This has been done by Neistadt and Lochak and others. This is also shown numerically and discussed in terms of a bounded error. The difference between the energy of the integrated system and the initial energy tends not to increase significantly after a time period.

In a splitting approach the error depends on the timestep but also on commutators of the individual pieces of the Hamiltonian. If there is a strong interaction then the energy difference between the modified Hamiltonian and real one can be large. This has led to the development of hybrid integrators where terms are shifted between pieces of the Hamiltonian during a close encounter.

The geometry of phase space is maintained by a symplectic integrator. This is often discussed in terms of stability. If the true Hamiltonian contains a periodic orbit, the modified one should also, but it may be shifted. If there is an unstable point (like a separatrix) the modified Hamiltonian should contain a similar unstable orbit. Resonances in one system should be present in the other. However the strengths of resonances might not be the same in both systems and so chaotic trajectories might be present in the modified system but not the original or vice versa.

Symplectic integrators for planetary systems are parallizeable. In particular it is possible to combine an N-body approach toward calculating interactions with Keplerian operators in a parallel fashion. Keplerian evolution can be done without interactions.

Higher order symplectic methods exist. Runge-Kutta methods tend to be implicit and so require additional calculation. Operator splitting methods can be generalized to any order. Of recent interest is the forced gradient method which reduces the number of computation steps and achieves similar accuracy to the operator splitting methods.

3 References

Dehnen, W. 1999 Astronomical Journal, 118, 1190, *Approximating Stellar Orbits: Improving on Epicycle Theory*

Chambers, J. E. 1999, MNRAS, 304, 793, A hybrid symplectic integrator that permites close encounters between massive bodies

Duncan, M., Levison, H. & Lee, M.H. 1998, Astronomical Journal, 116, 2067, *A Multiple Timestep Symplectic Algorithm for Integrating Close Encounters*

Preti, M., & Tremaine, S. 1999, Astronomical Journal, 118, 2532, *Class of Symplectic Integrators with Adaptive Time Step for Separable Hamiltonian Systems*

Mikkola, S., & Tanikawa, K. 1999, Celest. Mech. Dyn. Astron., 74, 287, *Explicit Symplectic Algorithms For Time-Transformed Hamiltonians*

Wisdom, J., & Holman, M. 1991, Astronomical Journal, 102, 1528, *Symplectic Maps for the N-body problem*

Yoshida, H. 1993, Celestial Mechanics and Dynamical Astronomy, 56, 27, *Recent Progress in the Theory and Application of Symplectic Integrators*

Yoshida, H. 1990, Phys. Lett. A, 150, 262

S. A. Chin, Phys. Lett. A226 (1997) 344

E. Forest and R. D. Ruth, Physica D 43 (1990) 105.

M. Suzuki, J. Math. Phys. 32 (1991)400.

Symplectic Integrators on the Shearing Sheet, Rein, H. and Tremaine, S. 2011 and?

4 Acknowledgements

Following Yoshida's review on the harmonic oscillator. Some notes from Rajeev's book on discrete time. Some notes from one of Wiggin's book on reversible systems. Some notes from one of Chin's recent papers on force gradient algorithms.

Force gradient integration discussion should be improved. I would like to better explain the exponential operator in the context of Lie derivatives. I would like to create some problems, perhaps on shearing sheet and on evaluating exponentials and commutators. For time reversal there is a nice paper by Dave Merritt on how to create integrators that obey time reversal symmetry.