BMB/Bi/Ch 173 – Winter 2018
Homework Set 9.1 – Assigned 3-6-18, due 3-14-18 by 10:30 a.m
TA – Taylor Stevens tstevens@caltech.edu
Office hours – Friday Mar 9 1:00pm - 3:00pm (in SFL 229) and Monday Mar 12 3:30pm - 5:00pm (in SFL 229), or by appointment

1. Methods of Integration (50 points)
Several methods exist to numerically integrate Newton's equations of motion and solve molecular dynamics problems. Two of the most commonly used are Euler integration and Velocity Verlet integration.

Please refer to the following resources to understand and compare these methods:
https://en.wikipedia.org/wiki/Verlet_integration
https://en.wikipedia.org/wiki/Euler_method
http://gafferongames.com/game-physics/integration-basics/

Note: There are three common flavors of Verlet integration schemes: the Basic Verlet algorithm, the Verlet Leapfrog algorithm, and the Velocity Verlet algorithm. This problem set asks about the Velocity Verlet algorithm.

a. (15 points) The Euler and Velocity Verlet integration methods follow different algorithms to numerically integrate. The Velocity Verlet algorithm has 4 steps, listed below. Referring to the above sources, fill in the missing components (denoted by light gray boxes) for the Velocity Verlet integration algorithm:

1. Calculate: $\vec{v}\left(t+\frac{1}{2}\Delta t\right) = \vec{v}(t) +$ [ ]
2. Calculate: [ ] $= \vec{x}(t) + \vec{v}\left(t+\frac{1}{2}\Delta t\right)\Delta t$
3. Derive $\vec{a}(t+\Delta t)$ from the interaction potential using $\vec{x}(t+\Delta t)$ where $\vec{a} = \frac{\vec{F}(x)}{m}$
4. Calculate: $\vec{v}(t+\Delta t) = \vec{v}(t) +$ [ ]

1. Calculate: $\vec{v}\left(t+\frac{1}{2}\Delta t\right) = \vec{v}(t) + \frac{1}{2}\vec{a}(t)\Delta t$
2. Calculate: $\vec{x}(t+\Delta t) = \vec{x}(t) + \vec{v}\left(t+\frac{1}{2}\Delta t\right)\Delta t$
3. Derive $\vec{a}(t+\Delta t)$ from the interaction potential using $\vec{x}(t+\Delta t)$ where $\vec{a} = \frac{\vec{F}(x)}{m}$
4. Calculate: $\vec{v}(t+\Delta t) = \vec{v}(t) + \frac{1}{2}\left(\vec{a}(t) + \vec{a}(t+\Delta t)\right)\Delta t$

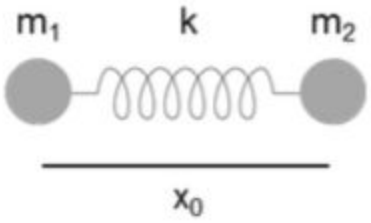
b. (5 points) Which of these two integration methods is simpler?
Euler method is the simplest of numerical methods – it is simpler than Velocity Verlet.

c. (10 points) Under what condition will Euler integration yield a correct result?
Euler integration will only yield a correct result under the condition that the rate of change is constant over the time step. In most cases, however, when simulating classical Newtonian mechanics, this condition is not met. For example, velocity can be accurately integrated using the Euler method in the case that acceleration is constant, but if you then wanted to integrate the velocity to get the position, the Euler integration will fail as the velocity is not constant (changing over timestep due to constant acceleration).

d. (10 points) How can one adjust Euler Integration to get a more accurate result? What is the tradeoff in doing this?
Reducing the time step for the Euler integration will get you a more accurate result, but the error will still persist and increase over time. Because of this, the Euler integration method often is used as a framework upon which more complex methods can be built for greater accuracy. The tradeoff is that adding more time points will increase the computational time required to simulate the system, which is not ideal.

e. (10 points) As you can see, MD simulation is constrained by the choice of the numerical integration algorithm used. What are three other limits of MD simulation?
Many possible answers: Simulation times (10's of μs) System sizes (10^9 atoms) Each trajectory is only one of all possible trajectories – requires many simulations Representations of atoms as points and force fields are just approximations (not relativistic/quantum mechanical)

2. Harmonic Potential (50 points)
Next, we will use molecular dynamics to solve a classic two-body problem: simulation of the positions of two masses connected by a spring. Consider m_1 and m_2, shown below, on a spring with a force constant k describing the stiffness of the spring. The initial distance between the two masses is x_0.



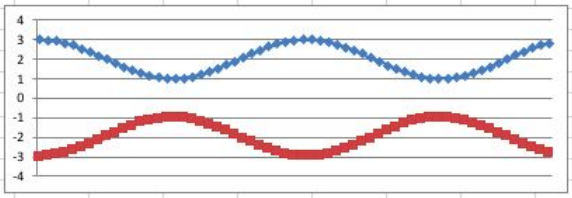
The potential energy can be modeled as the following harmonic potential:

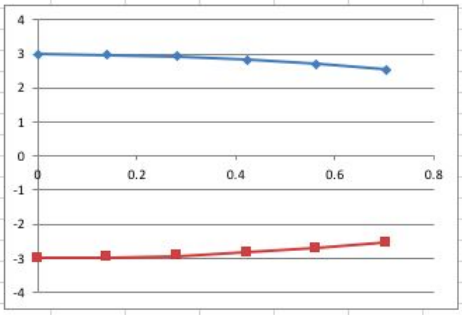
$$U(x) = \left(\frac{k}{2}\right)(x - x_0)^2$$

In the spreadsheet associated with this homework, the Velocity Verlet algorithm can be used to simulate the positions of the two masses. An example of the filled in spreadsheet is below. The

orange box denotes the input parameters.

	A	B	C	D	E	F	G
1		m1	1				
2		m2	1		resonance	1.414214	
3		x0	4				
4		k	1		MAX t	8.46	
5		Dt	0.141				
6							
7	t	x_1	v_1	x_2	v_2	F_{12}	F_{21}
8	0	3	0	-3	0	-2	2
9	0.0705		-0.141		0.141		
10	0.141	2.980119	-0.2792	-2.98012	0.27919678	-1.96024	1.960238
11	0.2115		-0.41739		0.41739356		
12	0.282	2.921267	-0.54729	-2.92127	0.54729214	-1.84253	1.842533
13	0.3525		-0.67719		0.67719071		
14	0.423	2.825783	-0.79363	-2.82578	0.79362606	-1.65157	1.651565
15	0.4935		-0.91006		0.91006141		
16	0.564	2.697464	-1.0084	-2.69746	1.00840383	-1.39493	1.394928
17	0.6345		-1.10675		1.10674625		
18	0.705	2.541413	-1.18309	-2.54141	1.18308544	-1.08283	1.082825
19	0.7755		-1.25942		1.25942464		
20	0.846	2.363834	-1.31073	-2.36383	1.31072521	-0.72767	0.727668
21	0.9165		-1.36203		1.36202579		
22	0.987	2.171788	-1.38625	-2.17179	1.38624793	-0.34358	0.343576
23	1.0575		-1.41047		1.41047007		
24	1.128	1.972912	-1.40665	-1.97291	1.40665065	0.054176	-0.05418
25	1.1985		-1.40283		1.40283124		
26	1.269	1.775113	-1.37112	-1.77511	1.37112214	0.449775	-0.44977
27	1.3395		-1.33941		1.33941303		
28	1.41	1.586256	-1.28108	-1.58626	1.28107506	0.827489	-0.82749
29	1.4805		-1.22274		1.22273708		
30	1.551	1.41385	-1.14009	-1.41385	1.14008987	1.172301	-1.1723
31	1.6215		-1.05744		1.05744267		
32	1.692	1.26475	-0.95377	-1.26475	0.95377244	1.4705	-1.4705





a. (5 points) Note that in the table, all of the values in rows 8 and 10 are filled in, while values are missing in row 9. This is on purpose. Explain the reason for this choice in terms of Velocity Verlet integration algorithm. What is the computational benefit of this choice?
The only values filled in for row 9 are the half-step velocities between rows 8 and 10 (positions and forces are excluded). This is used in Velocity Verlet because it offers greater accuracy and requires less memory because it does not require storing velocity. Instead, you can get positions and forces from the velocity half time step calculation without having to calculate all the values at the half time step – minimizing computing cost while maintaining accuracy.

b. (25 points) Now, let's use Velocity Verlet integration to simulate this system.
i. (10 points) First, derive the force (F(x)) for this harmonic oscillator.

$$F(x) = -\frac{dU(x)}{dx} = -2\left(\frac{k}{2}\right)(x - x_0) = -k(x - x_0)$$

ii. (15 points) Next, write equations to update the velocity and position as well. Use these equations to fill in the missing rows of the spreadsheet. Take the time out to cell A128. You will know you have the correct solution when your values match those in the screenshot above.

To calculate the first time step (1/2Δt) enter the following:
Cell A9 = A8+0.5*$C$5

Next calculate the velocities at this time step:
Cell C9 = C8+0.5*F8/$C$1*$C$5
Cell E9 = E8+0.5*G8/$C$2*$C$5

To calculate the next time step (Δt) enter the following:
Cell A10 = A9+0.5*$C$5

Next we calculate the positions at this time step:
Cell B10 = B8+C9*$C$5
Cell D10 = D8+E9*$C$5

Next we calculate the forces at this time step:
Cell F10 = -$C$4*(B10-D10-$C$3)
Cell G10 = -F10

Lastly we calculate the velocities at this time step:
Cell C10 = C8+0.5*((F8/$C$1)+(F10/$C$1))*$C$5
Cell E10 = E8+0.5*((G8/$C$1)+(G10/$C$1))*$C$5

These two rows of cells can now be copied continuously until the desired time point is reached.

c. (10 points) The equation for the resonance frequency, ω, is found in cell F2.

By the way, the resonance frequency is found by solving the equations of motion analytically:

$$F(x) = ma = m\frac{d^2x}{dt^2}$$

Substituting in what we know about our system, we can arrive at the following second order linear differential equation:

$$\frac{d^2x}{dt^2} = \omega^2 x$$

$$\text{where } \omega = \sqrt{\frac{k(m_1+m_2)}{m_1 m_2}} = \sqrt{\frac{k}{\mu}}\ .$$

The solution to this differential equation is of the form:

$$x(t) = A\cos\omega t + B\sin\omega t$$

Why is the resonance frequency important for this system? What constraint does it place on an important MD parameter?
The resonance frequency can be understood as the characteristic frequency for the mass/spring system. The resonance frequency is important for this system because it is the inverse characteristic time for this oscillating system based on its approximation as a harmonic potential. It is also important for this system because it defines the limitation on the time step to maintain numerical stability. If you set the time step to be greater than the resonance frequency, the simulation becomes numerically unstable.

d. (10 points) Now try tweaking the time step in the spreadsheet to see what happens as the time step approaches the resonant frequency, and then what happens when the time step is greater than the resonant frequency. What happens? Based on this what rules would you follow to pick an appropriate time step for a molecular dynamics simulation?

As you approach the resonant frequency the graph takes on a weird appearance the amplitude of each peak changes in an oscillating pattern. After you pass the resonant frequency the amplitude of oscillation increases continuously, eventually reaching massive values. This suggests that we need to pick a time step that is below the resonant frequency of any harmonic oscillators in an MD system.

3. Coarse-grained molecular dynamics (25 points)
To gain a better understanding of how coarse-grained simulation can be applied, let's read through the following publication (https://www.sciencedirect.com/science/article/pii/S0006349506721701) from Klaus Schulten's group. In this paper, the authors use a coarse-grained simulation to analyze the motions of a bacterial flagellum.

a. (6 points) What did the authors want to analyze with their simulations and why couldn't they do this with an all-atom MD simulation? What limitations could be by-passed in a coarse grained simulation?

They hypothesized that the rotation of the flagella in a particular direction would lead to the formation of a specific superhelical conformation due to solvent friction. A segment of the filament that would be sufficient to test this hypothesis would contain way to many atoms to allow simulation on a relevant time scale with all-atom methods. With a coarse grained simulation they were able to recreate the behavior of a single monomer with 15 beads, greatly reducing the computational requirement of each time step. The coarse-graining also allowed them to use a much larger time step.

b. (6 points) How many beads are used to model a flagellin monomer? What is the total number of beads in the largest simulation? What do the authors need to do in order to determine the parameters for describing the physical interaction between each of the beads?

- 15 beads are in a single monomer
- 1100 monomers are in the largest simulation; 15*1100 = 16500 beads
- The authors first run an all-atom MD simulation of a single monomer and from this they are able to derive parameters for the coarse grained simulation that will most closely match the behavior of the all atom simulation

c. (7 points) All-atom MD simulations typically use time steps of 1-2 fs. For the simulations described in this publication, the authors use a much larger time step. Explain why it might be acceptable to use a larger time step for a coarse grained simulation than for an all atom simulation. (Hint: look at the expression for resonance frequency of a harmonic oscillator in the preceding problem).

The resonance frequency of a harmonic oscillator is inversely proportional to the square root of the reduced mass. Since each bead represents several atoms its mass is much higher than a single atom, therefore the resonance frequency of the bond vibrations will be much lower, allowing longer time steps.

d. (6 points) List 2 drawbacks to the approach used in these simulations.

- They could not observe precise conformational change within the protein such as the switching interactions implicated in the L-R superhelical form transitions
- They still had to speed up the rotation of the filament 500x the normal biological rate to fit a full rotation into their simulations, and after 20μs they actually saw the the base tear away from the rest of the filament probably due to this unnatural rotation
- The filament compressed to a length that is shorter than the actual filament should be during their simulation