Hamiltonian Notes

There are three main approaches to mechanics, Newtonian, Lagrangian, and Hamiltonian. Newtonian we are all familiar with as that is what is taught in highschool. The other two reformat mechanical systems using objects called the Lagrangian and Hamiltonian to construct systems of differential equations with which physical systems can be solved.

Both the Lagrangian and Hamiltonian are equations related to the total energy of a system and how it is distributed between kinetic and potential. The thing is that they are only defined by the fact that they produce the correct equations of motion when used in physics. As such there are multiple possible equations. However, tradtionally,

$$L = T - V \qquad H = T + V$$

Where T is the sum of kinetic forces on a system and V is some method of measuring potential energy.

Intuitively it kind of makes sense to draw inspiration for descent algorithims from how the Hamiltonian operates as it is governed by two differential equations, and gradient descent algorithms involve following the gradient downwards. Which is kind of like a flow through a slope field.

Proximal Point Methodology

Roughly speaking, the Proximal Point Method is a generalization of numerical optimization methods for minimizing cost functions. The idea is that each step in the discrete optimization algorithm is characterized as follows:

$$x_{t+1} \leftarrow \underset{x \in \mathbb{R}^d}{argmin}\{f(x) + \frac{1}{2\eta_{t+1}}||x - x_t||^2\}$$

In layman's terms, we take a point x and then consider all other possible points in the domain of f. We look for the point where the function plus some expression related to the distance from the initial point is minized and then step to that point. Since evaluating f at every point x might be computationally taxing, often f is approximated by a first or second order approximation. The first order approximation directly yields gradient descent.

Situation

We are going to evaluate descent methods and discretizations of hamiltonian mechanics in the context of a 1d oscillator. As such the following are taken to be true:

$$\frac{dx}{dt} = \nabla f(x) = v(t)$$
$$\frac{dv}{dt} = \frac{-kx}{m}$$

Descent Methods

- $x_t =$ position vector after t iterations
- $p_t =$ momentum vector after t iterations
- $h =$ stepsize
- $\mu =$ weight for momentum
- $f =$ objective function we are trying to optimize, in this case we take $f = \frac{1}{2}kx^2$
- Hence, $\nabla f = kx$

Classical Momentum

$$x_{t+1} = x_t + \mu p_t - h_t \nabla f(x_t) = x_t + p_{t+1}$$
$$p_{t+1} = \mu p_t - h_t \nabla f(x_t) = \mu p_t - h_t k x_t$$

Nesterov

$$x_{t+1} = x_t + \mu p_t - h_t \nabla f(x_t + \mu p_t) = x_t + p_{t+1}$$
$$p_{t+1} = \mu p_t - h_t \nabla f(x_t + \mu p_t) = \mu p_t - h_t k(x_t + \mu p_t)$$

Dissaptive Discretizations of Hamiltonian Systems

Using a symplectic euler method with an additional dissaptive element, we can achieve similar results and paths to both classical and nestrov descent methods. (Springer Page 189). Take our example to start in one dimension, governed by the following equations:

- Because we are already using p as the momentum of our discretization, we will label traditional physical momentum with
- Then, the Hamiltonian H will be defined as follows: $H(x,p) = \frac{p^2}{2m} + \frac{1}{2}kx^2$
- Thus, $\frac{\partial H}{\partial x} = -\frac{dp}{dt} = -kx$ and $\frac{\partial H}{\partial p} = \frac{dx}{dt} = \frac{p}{m}$

Discretization via Euler Method

$$x_{t+1} = x_t + h\frac{p_{t+1}}{m}$$
$$p_{t+1} = p_t - hkx_t$$

Similarities

Clearly these two completely seperate numerical techinques share a lot of similarities when applied to the same problem. The gradient descent algorithims are solving for a minimum of the objective function $\frac{1}{2}kx^2$, or the potential energy of a spring. On the other hand, the Symplectic Euler Method is applied to the same system to try and solve for a flow. As such, when we rum the symplectic euler method, it will return an approximation of the oscillator, never settling to the minimum. This is because the Hamiltonian is being preserved, as such when $x_k = 0$, there is some value in p_k and so it will never truly descend until we dampen it in some way. Since we are trying to find the bottom of the potential curve, it makes sense to dampen the kinetic term of the Hamiltonian. For example, the Hamiltonian descent method suggests that we should reset the momentum term to 0 on every iteration. This yields the equation below. As you can see, this is pretty much a recovery of regular gradient descent. If we want the concept of momentum to persist, we can include the μ term from above applied to the previous momentum instead of just eliminating it. This nearly recovers the formula for classical momentum.

Hamiltonian Descent via Euler Method

$$x_{t+1} = x_t + h\frac{p_{t+1}}{m} = x_t - \frac{h^2}{m}\nabla f$$
$$p_{t+1} = -h\nabla f = -hkx_t$$

$$x_{t+1} = x_t + h\frac{p_{t+1}}{m}$$
$$p_{t+1} = \mu p_t - h\nabla f$$

Computation

Starting Values:

- $x_0 = 1$
- $p_0 = 0$
- $h = .1$
- $\mu = 1$

- $m = 1$
- $k = 1$

Outline:

1. Introduce Hamiltonian Mechanics
2. Lemma: Time derivative of H is 0 on Hamiltonian Flow: Hamiltonian Descent Paper
3. Introduce concept of removing momentum and following flow
4. Examine duration of flow(Should not be near infinity) and prove appendix a from:Hamiltonian descent
5. Lemma: Hamiltonian descent satisfies:

   $$f(x_{k+1}) + \frac{1}{2}||v_{k+1}^2|| = x_k$$

   Hamiltonian descent paper
6. Consider difficulty of calculating flow
7. Introduce numerical integrators and symplecity
8. Prove that a flow ϕ_t over $H(p,q)$ is symplectic (Theorem 2.4 of Symplectic Transforms Hairer)
9. Introduce Euler Symplectic method
10. Modify it to produce gradient and CM descent
11. Applications of CM to physical systems(Tracing water runoff?)