Hamiltonian Learner ReadMe

March 25, 2018

1 Mathematical Background

1.1 Hamiltonian Mechanics

Let $\mathbb{R}^{2n}$ be our phase spaces with coordinates and conjugate momenta x_i and p_i

$$
\begin{aligned}
\frac{dx_i}{dt} &= \{H, x_i\} = \frac{dH}{dp_i} \\
\frac{dp_i}{dt} &= \{H, p_i\} = -\frac{dH}{dx_i}
\end{aligned}
$$

This is not strictly mechanics despite the name. As an example:

1.1 Example (Chemical Rate Equations) *A Volterra equation is of the form:*

$$
\begin{aligned}
\dot{x}_j &= \epsilon_j x_j + \frac{1}{\beta_j} \sum_k a_{jk} x_j x_k \\
\dot{x}_j &= \epsilon_j x_j + \sum_k a_{jk} x_j x_k \\
a_{jk} &= -a_{kj}
\end{aligned}
$$

where the second line is by rescaling variables. This can be turned into a Hamiltonian system by:

$$Q_j \equiv \int_0^t x_j(\tau)d\tau$$

$$\ddot{Q}_j = \epsilon_j Q_j + \sum_{k=1}^n a_{jk}\dot{Q}_j\dot{Q}_k$$

$$H \equiv \sum \epsilon_j Q_j - \dot{Q}_j$$

$$\dot{H} = 0$$

$$P_j \equiv \log\dot{Q}_j - \frac{1}{2}\sum a_{jk}Q_k$$

$$H = \sum_j \epsilon_j Q_j - \sum_j e^{P_j + \frac{1}{2}\sum_k a_{jk}Q_k}$$

$$\dot{P}_j = \frac{\partial H}{\partial Q_j} = \epsilon_j - \sum_l e^{P_l + \frac{1}{2}\sum_k a_{lk}Q_k}\frac{1}{2}a_{lj}$$

$$\dot{Q}_j = -\frac{\partial H}{\partial P_j} = \sum_l e^{P_l + \frac{1}{2}\sum a_{lk}Q_k}$$

$$I_j \equiv P_j - \frac{1}{2}\sum a_{jk}Q_k - \epsilon_j t$$

$$\dot{I}_j = 0$$

$$I_j = I_j(0) = \log x_j(0)$$

$$\{I_j, I_k\} = a_{jk}$$

This means that we may extract the underlying system by using an exponential model Hamiltonian.

1.2 Algebraic Varieties

Let I be an ideal in $k[x_1\cdots x_d]$. Demanding that all the functions in that ideal vanish determines a variety in k^d. k will be $\mathbb{R}$ but I don't want to get into the subtleties of base change to algebraic closure so some of the statements aren't strctly true.

1.2 Example (Conic Sections) *We can get examples such as a circle with $I = (x^2 + y^2 - 1) \subset \mathbb{R}[x, y]$. Ellipses, hyperbolas and parabolas are also visible with other degree 2 equations.*

An important factor here is that we are not necessarily using them as functions. When finding the equation $y - x^2 = 0$ in the ideal is not the same as doing a regression problem. There is no notion of dependent or independent variables. No causality is implied or even desired.

1.3 Lagrangian Correspondences

1.3 Definition (Lagrangian Submanifold) *Isotropic submanifolds have the property that $\omega \mid_I = 0$. A Lagrangian is maximal dimensional among these.*

Consider a pair of symplectic manifolds $X_{in,out}$. This stands for the inputs and outputs of a system. For example, it could be a black boxed electronic circuit. For each of the input wires there is an $\mathbb{R}^2$ for the current and voltage. These form a canonical conjugate pair for each of the wires. So for a circuit with n input wires and m output wires we get $\mathbb{R}^{2n}$ and $\mathbb{R}^{2m}$. To specify what the circuit is at this black boxed level we specify a morphism $X_{in} \to X_{out}$ in the following sense.

1.4 Definition (LagCor) *A morphism between symplectic manifolds in the "category" LagCor is a Lagrangian correspondence. That is a Lagrangian submanifold of $X_{in} \times \bar{X}_{out}$. Composition of L_{12} and L_{23} is given by pullback if possible. That is a transversality condition with the diagonal in $X_2 \times \bar{X}_2$.*
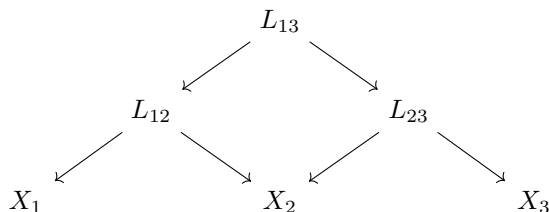
Composition is simply piping some outputs into other inputs. By moving some wires around you are allowed to perform partial gluings so that you can feed some outputs of system 1 into system 2 and leave some outputs as free to go all the way to the end. Similarly for inputs to system 2 that don't come from system 1 but instead straight from the begining.

1.5 Definition (Graphs of symplectomorphisms) *For $X_1 = X_2$ a large class of morphisms $X_1 \to X_2$ is given by the graphs of symplectomorphisms $\phi \, X_1 \to X_2$. A symplectomorphism mean it is a diffeomorphism that preserve ω. But these are special in the sense that if you specify the input, you uniquely know the output. Sets and relations is better than sets and functions.*

1.6 Example *Suppose we observe the state of a system at times t_1 and t_2. There is some symplectomorphism that implements this discretized dynamics. In fact it might be the time $t_2 - t_1$ flow of a (possibly time dependent) Hamiltonian. If we observe enough such state pairs, we have enough information to approximately reconstruct a Lagrangian correspondence (At least the closest approximation in the class of low degree polynomial varieties).*

Nicely provided in `https://ncatlab.org/nlab/show/prequantized+Lagrangian+correspondence` with more context/generality. Mostly unneeded for now, though it provides the mindset.

1.3.1 Singular Lagrangians

We might have systems where the vanishing locus is no longer a manifold but is stratified. An example that shows up often is hysterisis loops.

1.7 Definition (Magnetic Hysterisis) *Let th symplectic manifold be $\mathbb{R}^2$ with magnetization and applied field. Then a material defines a singular "1-dimensional" subset. A picture describes it best.*
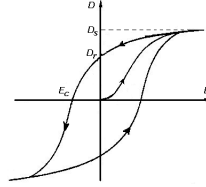

Figure 1: Hysterisis loop

This is thought of as a system with 0 inputs and $\mathbb{R}^2$ output. The dynamics is then constrained to this Lagrangian because once you specify the applied field you are stuck with either 1 or 2 values for the magnetization because of the equations of state. It is not a Lagrangian submanifold only because of those triple points.

1.4 Doubling Trick

Recall from Baez-Fong-... how certain linear circuit elements provide affine Lagrangian subspaces of current and voltage for inputs and outputs. This straightforward construction won't work for a memristor. However, we can double the number of variables by appending $\int_{-\infty}^{t} V$ and $\int_{-\infty}^{t} I$ as well. Because they have a fundamentally different nature compared to the original V and I give them homological degree 1. That is our new phase space is $T^*[1]\mathbb{R}^2$. This is the same trick used in the Poisson Sigma Model. You take a symplectic manifold $\mathbb{R}^2$, regard it merely as Poisson and then form a 1-shifted symplectic space by taking the odd cotangent bundle. Now instead of looking for Lagrangian Correspondences between $\mathbb{R}^2_{in}$ and $\mathbb{R}^2_{out}$, you look for Lagrangian correspondences between $T^*[1]\mathbb{R}^2_{in}$ and $T^*[1]\mathbb{R}^2_{out}$. This is the general situation that subsumes Baez et al, but also allows more general circuit elements.

This is especially satisfying because that is the first step in Kontsevich quantization. It is the construction you needed to do anyway when quantizing, so it is satisfying to see it being useful even in the classical world.

2 Symmetric Hamiltonians

Consider the example of n particles in $\mathbb{R}^3$. Then the phase space is $(\mathbb{R}^{6n})$ and there is a natural $Euc(3)$ action that we can expect to be a symmetry of the system. One can immediately write down that polynomials in x_i and p_i that correspond to the Noether charges. The Hamiltonian will then Poisson commute with these. This drastically reduces the space of potential choices for the gradient descent to operate in.

3   Down from Infinite Dimensional

Consider the problem for a PDE instead of an ODE. Let the space be M

$$\omega \;=\; \int_M \delta A \wedge \delta B$$

See BF in BV-BFV for more explanation.

Give a basis of $L^2(\Omega^\bullet(M))$ so that the integration pairing is adapted. Will explain by example on $\Omega^0(S^2)$ and $\Omega^2(S^2)$

$$
\begin{aligned}
e_{lm,0} &= Y_l^m(\theta,\phi)^* \\
e_{lm,2} &= Y_l^m(\theta,\phi)\,dvol_{S^2}
\end{aligned}
$$

A complexified 0 form A and a complexified 2 form B can be decomposed in this basis and as can their variations.

$$\omega \;=\; \sum_{lm} \delta_{lm0} A \wedge \delta_{lm2} B + \cdots$$

So this is taking a function on $S^2 \times \mathbb{R}$, decomposing each time slice into spherical harmonics and then asking if we can analyze the associated infinite dimensional Hamiltonian system. Of course realistically, it should be reasonable to approximate with some sufficient L cutoff. There will be some relatively small $n(L)$ modes kept. This can then be fed into the machine with $\mathbb{R}^{2(n(L))}$.

3.1 Example (Weather) *Too bad typically don't have the data of the canonically conjugate field.*

3.2 Definition (Kaluza-Klein Reduction) *This is in particular a reduction from a $M \times \mathbb{R}$ spacetime to just mechanics with only $\mathbb{R}$ for time and no space. The cost is the infinite tower that is introduced like the fact that lm run over a countably infinite range.*

4   Implementation

4.1   Time Trajectory

Input some traces of trajectories in phase space. That is $x_i(t_j) = x_{ij}$ and $p_i(t_j) = p_{ij}$ for t_j a discetization of a time interval. The loss function is given by comparing to Hamilton's equations.

$$\Delta x_{ij} = \frac{x_{i,j+1} - x_{ij}}{t_{j+1} - t_j} \approx \frac{dx_i}{dt}(t_j)$$

$$\Delta p_{ij} = \frac{p_{i,j+1} - p_{ij}}{t_{j+1} - t_j} \approx \frac{dp_i}{dt}(t_j)$$

$$L = \sum_j \sum_i (\Delta x_i(t_j) - \frac{dH}{dp_i}(x_{ij}, p_{ij}))^2 + (\Delta p_i(t_j) + \frac{dH}{dx_i}(x_{ij}, p_{ij}))^2$$

The possible Hamiltonians are specified as giving the terms that are allowed to show up. For exampe:

$$H_{pol,3} = A + Bx + Cp + Dx^2 + Exp + Fp^2 + Gx^3 + Hx^2p + Ixp^2 + Jp^3$$

$$H_{trig,1} = A\cos Bx + C\sin Dx + E\cos Fx + G\sin Hx$$

$$H_{exp,1} = Ae^{Bx} + Ce^{Dp}$$

These capital letter unknowns are what we optimize over.

4.2 Algebraic Variety Learning

First ignore the dynamical system aspect and consider the problem of giving a series of points x_{ij} where i indexes the space $\mathbb{R}^n$ while j indexes which data point this is. We are told that they come from an algebraic variety of low degree and the task is to determine the defining functions that cut out the variety. For example, if we know it is cut out by a real plane quadric.

$$f = A + Bx_1 + Cx_2 + Dx_1^2 + Ex_1x_2 + Fx_2^2$$

$$\begin{pmatrix} 1 & x_{1,j} & x_{2,j} & x_{1,j}^2 & x_{1,j}x_{2,j} & x_{2,j}^2 \\ \dots \end{pmatrix} \begin{pmatrix} A \\ B \\ C \\ D \\ E \\ F \end{pmatrix} = \begin{pmatrix} 0 \\ \dots \end{pmatrix}$$

Giving one function means cutting a codimension 1 variety typically. So to give a codimension m variety, we should give m such functions which are independent of each other. Then our variety is the vanishing set $V(I)$ for the ideal $I = (f_1 \cdots f_m)$.

However because of noise we cannot expect exactly 0. Instead we say that we seek to minimize the squares of each of the rows of the RHS. That turns into a quadratic optimization.

4.1 Corollary *We can parallelize over the rows. That is we have some solutions that work for a subset of rows, and another bunch of solutions that work for another subset of the rows. If there is something in common there, we have a solution for the composite.*
This may or may not provide any advantage.

4.3 Lagrangian Subvariety Learning

If we are told that we are learning a coisotropic variety of codimension m that amounts to giving m such functios as above but now there is the constraint that all of the f_i Poisson commute. These are quadratic constraints in the capital letter unknown variables.

4.2 Example (Codimension 2 in 4d)

$$
\begin{aligned}
f_1 &= A_1 + B_1 x_1 + C_1 x_2 + D_1 p_1 + E_1 p_2 + F_1 x_1^2 + G_1 x_2^2 + H_1 x_1 x_2 \\
f_2 &= A_2 + B_2 x_1 + C_2 x_2 + D_2 p_1 + E_2 p_2 + F_2 x_1^2 + G_2 x_2^2 + H_2 x_1 x_2 \\
\{f_1, f_2\} &= B_1 D_2 + C_1 E_2 - D_1 B_2 - E_1 C_2 + \cdots
\end{aligned}
$$

At the very extreme for a Lagrangian submanifold in $\mathbb{R}^{2n}$ we give a codimension n coisotropic variety. This is the case of an integrable system where the f_i are the conserved quantities that define the action coordinates in action-angle variable system formalism.

4.4 Bootstrapping

Now the easier problem of finding many conserved quantities is accomplished we can proceed to finding which is the actual Hamiltonian. Suppose the result from the previous step gave $f_1 \cdots f_n$ each with total degree $\mid f \mid_i$. For example some of them might be L_z for a rotationally symmetric system with a conserved angular momentum. So we know many potential Hamiltonians. Parameterize them with unknown coefficents.

$$
H = \sum_i a_i f_i + \sum_{ij} a_{ij} f_i f_j + \cdots
$$

Use the total degree to make this list finite and manageable. The Hamiltonian is unlikely to need a term like L_z^{100} even though it is conserved so that is not even considered. With this reduced space of potential Hamiltonians, we proceed with the loss function of the previous section using the trajectory data. Before $H_{pol,3}$ had 10 unknown parameters to do gradient descent with. But there is only one conserved quantity found by the simpler method so this is reduced to looking at $A + B f_1 + C f_1^2 + D f_1^3$ which is significantly easier.

5   Applications

5.1   Circuits

Already described

5.2   Robotic Control and other Optimal Control problems

For a robot moving in 2 or 3 dimensions the configuration space is $SE(2)$ or $SE(3)$ or some products thereof for describing position and framing orientation. This means the phase space will be $T^*SE(3)$ etc. The dynamics will take place on that. To implement this we must restrict to a symplectic coordinate chart. If we stay within that one chart there is no problem with reducing to this already established problem.

Often the problem is presented within the realm of optimal control. That would require adjoining the control variables to the Hamiltonian as well.

5.1 Theorem (Hamilton Jacobi) *Let P α be an exact symplectic manifold not necessarily a cotangent bundle with Liouville form. Get a function on parameterized paths γ by:*

$$S(\gamma) \;\; = \;\; \int_{[\gamma]} \alpha - \int_{t_0}^{t_1} H(\gamma(t))$$

where $[\gamma]$ is the unparameterized path. When not exact, α is now a connection on a line bundle so S is now valued as a section instead.

$$\mathcal{L}$$
$$\downarrow$$
$$Paths(P) \Longrightarrow (P, \omega)$$

5.2 Theorem (Hamilton-Jacobi-Bellman)

$$
\begin{aligned}
V(x(0), 0) &= min_u L \\
L &= \int_0^T C(x(t), u(t))dt + D(x(T)) \\
\dot{x}(t) &= F(x(t), u(t)) \\
\dot{V}(x, t) &+ min_u \nabla V(x, t) \cdot F(x, u) + C(x, u) = 0 \\
V(x, T) &= D(x)
\end{aligned}
$$

We then solve this PDE with that terminal condition for $V(x,t)$

Proof

$$
\begin{aligned}
V(x(t),t) &= min_u\left(V(x(t+dt),t+dt) + \int_t^{t+dt} C(x(s),u(s))ds\right) \\
&\approx min_u\left(V(x(t),t) + \dot{V}(x(t),t)dt + \nabla V(x(t),t)\cdot\dot{x}(t)dt + \int_t^{t+dt} C(x(s),u(s))ds\right) \\
0 &\approx min_u\left(\dot{V}(x(t),t)dt + \nabla V(x(t),t)\cdot\dot{x}(t)dt + C(x(t),u(t))dt\right)
\end{aligned}
$$

where the error is $o(dt)$. Taking this limit then provides the required PDE.

5.3 Lemma (Principle of Optimality) *An optimal control policy $u(t)$ has the property that if we drop the first step, the remaining policy is optimal for the new state. In other words, an optimal policy is like a shark it doesn't need to look back. It doesn't have a neck.* $http://www.imdb.com/title/tt0584440/quotes$.

5.4 Definition (LQR)

$$
\begin{aligned}
\dot{x}(t) &= A(t)x(t) + B(t)u(t) \\
J = \frac{1}{2}x^T(t_1)F(t_1)x(t_1) &+ \frac{1}{2}\int_{t_0}^{t_1}(x^T(t)Q(t)x(t) + u^T(t)R(t)u(t))dt \\
x(t_0) &= x_0 \\
u(t) &= -K(t)x(t) \\
K(t) &= R^{-1}(t)B^T(t)P(t) \\
A^T(t)P(t) + P(t)A(t) &- (P(t)B(t))R^{-1}(t)(B^T(t)P(t)) + Q(t) = -\dot{P}(t) \\
P(t_1) &= P_f
\end{aligned}
$$

5.5 Definition (Infinite Horizon Problem)

$$
\begin{aligned}
\dot{x} &= Ax + Bu \\
J &= \frac{1}{2}\int_0^\infty (x^T Q x + u^T R u)\,dt \\
x(0) &= x_0 \\
u &= -Kx \\
K &= R^{-1}(B^T P) \\
A^T P + PA - (PB)R^{-1}(B^T P) + Q &= 0
\end{aligned}
$$

The Hamilton-Jacobi-Bellman equations become

5.6 Definition (Differential Riccati Equation)

$$
A^T P(t) + P(t)A - (P(t)B)R^{-1}(B^T P(t)) + Q = -\dot{P}(t)
$$

5.7 Definition (Algebraic Riccati Equation) *If we assume the dynamics and cost functions have no explicit time dependence, this simplifies*

$$
A^T P + PA - (PB)R^{-1}(B^T P) + Q = 0
$$

5.8 Lemma *We would like to solve for P such that the eigenvalues of $A - BK = A - BR^{-1}B^T P$ are all in the negative left half plane. That is the stability for $x = e^{\lambda t} x_0$ solutions.*

$$
Z \equiv \begin{pmatrix} A & -BR^{-1}B^T \\ -Q & -A^T \end{pmatrix}
$$

Assume no eigenvalues with real part 0. Then have n eigenvectors of Z with negative real part. Write them as columns $\begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$ as a $2n$ by n matrix. Then take $P = U_2 U_1^{-1}$.

So in our setup, we do not know A or B, those are determined from observing a trajectory first. Then once we have found them, we solve for the policy P.

Recht - `https://arxiv.org/pdf/1802.08334.pdf`

Jha,Bhasin - `https://www.sciencedirect.com/science/article/pii/S2405896317312582`

5.3   Climate Data

First must reduce from the infinite dimensional situation of functions on a sphere to a finite dimensional phase space which only keeps track of the low lying modes in some frame (generalized basis) of $L^2(S^2, d\Omega)$. You plot the dynamics of these coefficents which is a lossy compression and hope to learn the dynamics there. This can be done iteratrively by starting with only keeping the 0th order, learning the dynamics then using that information for the initial guess in the next step and so on until the compression loss is sufficiently low.

5.4   Financial/Biodiversity

Is there a candidate for the conjugate variables for price data? Could just use the time derivative. Typically these are modeled as ARIMA processes, but the interactions between various entities should be close to generalized Lotka-Volterra type just like other predator-prey-cooperation systems. Already mentioned Hamiltonian realization of these sorts of networks.