

# Machine Learning of Dynamical Systems

April 24, 2018

## 1 Introduction

Dynamical system identification has a long history [HGH15, WWJ16]. Here we take the insights gained from the symplectic “category” of Weinstein called *LagCor* below and some inspiration from holomorphic symplectic geometry [Jin15] in order to design a system that identifies algebraic dynamical systems from their trajectories.

## 2 Mathematical Background

### 2.1 Hamiltonian Mechanics

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

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

Despite the name, these systems do not have to strictly be mechanical. As an example:

**2.1 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:

$$\begin{aligned}
Q_j &\equiv \int_0^t x_j(\tau) d\tau \\
\ddot{Q}_j &= \epsilon_j Q_j + \sum_{k=1}^n a_{jk} \dot{Q}_j \dot{Q}_k \\
H &\equiv \sum_j \epsilon_j Q_j - \dot{Q}_j \\
\dot{H} &= 0 \\
P_j &\equiv \log \dot{Q}_j - \frac{1}{2} \sum_k a_{jk} Q_k \\
H &= \sum_j \epsilon_j Q_j - \sum_j e^{P_j + \frac{1}{2} \sum_k a_{jk} Q_k} \\
\dot{P}_j &= \frac{\partial H}{\partial Q_j} = \epsilon_j - \sum_l e^{P_l + \frac{1}{2} \sum_k a_{lk} Q_k} \frac{1}{2} a_{lj} \\
\dot{Q}_j &= -\frac{\partial H}{\partial P_j} = \sum_l e^{P_l + \frac{1}{2} \sum_k a_{lk} Q_k} \\
I_j &\equiv P_j - \frac{1}{2} \sum_k a_{jk} Q_k - \epsilon_j t \\
\dot{I}_j &= 0 \\
I_j = I_j(0) &= \log x_j(0) \\
\{I_j, I_k\} &= a_{jk}
\end{aligned}$$

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

## 2.2 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 strictly true.

**2.2 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.

## 2.3 Lagrangian Correspondences

**2.3 Definition (Lagrangian Submanifold)** *Isotropic submanifolds have the property that  $\omega|_I = 0$ . 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  $X_{in} \rightarrow X_{out}$  in the following sense.

**2.4 Definition (Graphs of symplectomorphisms)** *For  $X_1 = X_2$  a large class of morphisms  $X_1 \rightarrow X_2$  is given by the graphs of symplectomorphisms  $\phi: X_1 \rightarrow X_2$ . A symplectomorphism means it is a diffeomorphism that preserves  $\omega$ . 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.*

**2.5 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 [nLa18] with more context/generalizability. Mostly unneeded for now, though it provides the mindset.

### 2.3.1 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 hysteresis loops.

**2.6 Definition (Magnetic Hysteresis)** *Let the 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.*

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.

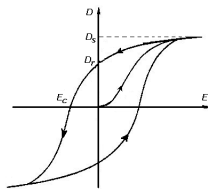
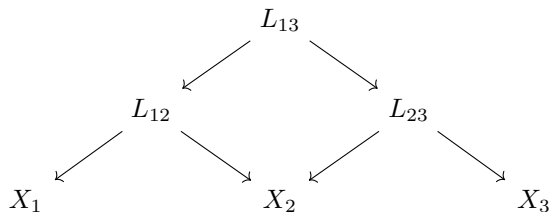


Figure 1: Hysteresis loop

## 2.4 The “category” LagCor

**2.7 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 beginning.

For us the Lagrangians are varieties of degree  $\leq d$  so we need to restrict our attention and use facts from intersection of algebraic cycles. These are all inspired from the theory of the derived category of motives. For practical implementability, we wish to find ways to guarantee transversality. This would allow us to restrict to a category level of 2 rather than  $\infty$  [WW09].

## 2.5 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. We can restrict to the space of polynomials in those  $6n$  variables of degrees  $\leq d$  that Poisson commute with all the Noether charges. A good start for these is simply polynomials in the Noether charges themselves.

## 3 Implementation

### 3.1 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 discretization of a time interval. The loss function is given by comparing to Hamilton's equations.

$$\begin{aligned}\Delta x_{ij} &= \frac{x_{i,j+1} - x_{ij}}{t_{j+1} - t_j} \approx \frac{dx_i}{dt}(t_j) \\ \Delta p_{ij} &= \frac{p_{i,j+1} - p_{ij}}{t_{j+1} - t_j} \approx \frac{dp_i}{dt}(t_j) \\ L &= \sum_j \sum_i \left( \Delta x_i(t_j) - \frac{dH}{dp_i}(x_{ij}, p_{ij}) \right)^2 + \left( \Delta p_i(t_j) + \frac{dH}{dx_i}(x_{ij}, p_{ij}) \right)^2\end{aligned}$$

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

$$\begin{aligned}H_{pol,3} &= A + Bx + Cp + Dx^2 + Exp + Fp^2 + Gx^3 + Hx^2p + Ixp^2 + Jp^3 \\ H_{trig,1} &= A \cos Bx + C \sin Dx + E \cos Fx + G \sin Hx \\ H_{exp,1} &= Ae^{Bx} + Ce^{Dp}\end{aligned}$$

These capital letter unknowns are what we optimize over.

### 3.2 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.

**3.1 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 speedup depending on the data size.*

Similar problems were studied in [BVSU18]

### 3.3 Lagrangian Subvariety Learning

If we are told that we are learning a coisotropic variety of codimension  $m$  that amounts to giving  $m$  such functions 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.

#### 3.2 Example (Codimension 2 in 4d)

$$\begin{aligned} f_1 &= A_1 + B_1x_1 + C_1x_2 + D_1p_1 + E_1p_2 + F_1x_1^2 + G_1x_2^2 + H_1x_1x_2 \\ f_2 &= A_2 + B_2x_1 + C_2x_2 + D_2p_1 + E_2p_2 + F_2x_1^2 + G_2x_2^2 + H_2x_1x_2 \\ \{f_1, f_2\} &= B_1D_2 + C_1E_2 - D_1B_2 - E_1C_2 + \dots \end{aligned}$$

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.

### 3.4 Bootstrapping

Now the easier problem of finding many conserved quantities is accomplished we can proceed to finding which is the actual Hamiltonian. This is thanks to the fact that we know the Hamiltonian is a conserved quantity itself, so it must be in the ideal found in the previous step. Suppose the result from the previous step gave  $f_1 \cdots f_n$  each with total degree  $|f|_i$ . 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 coefficients.

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

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  $A + Bf_1 + Cf_1^2 + Df_1^3$  which is significantly easier.

## References

- [BPK16] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.
- [BVSW18] Paul Breiding, Sara Kalisnik Verovsek, Bernd Sturmfels, and Madeleine Weinstein. Learning algebraic varieties from samples. *arXiv preprint arXiv:1802.09436*, 2018.
- [DEHH18] Emilie Dufresne, Parker B Edwards, Heather A Harrington, and Jonathan D Hauenstein. Sampling real algebraic varieties for topological data analysis. *arXiv preprint arXiv:1802.07716*, 2018.
- [FS18] S. Fattahi and S. Sojoudi. Data-Driven Sparse System Identification. *ArXiv e-prints*, March 2018.
- [FTM17] Mohamad Kazem Shirani Faradonbeh, Ambuj Tewari, and George Michailidis. Finite time identification in unstable linear systems. *arXiv preprint arXiv:1710.01852*, 2017.
- [HGH15] Daniel JA Hills, Adrian M Grütter, and Jonathan J Hudson. An algorithm for discovering lagrangians automatically from data. *PeerJ Computer Science*, 1:e31, 2015.
- [HMR16] Moritz Hardt, Tengyu Ma, and Benjamin Recht. Gradient descent learns linear dynamical systems. *arXiv preprint arXiv:1609.05191*, 2016.
- [Jin15] Xin Jin. Holomorphic lagrangian branes correspond to perverse sheaves. *Geometry & Topology*, 19(3):1685–1735, 2015.
- [JRB17] Sumit Kumar Jha, Sayan Basu Roy, and Shubhendu Bhasin. Data-driven adaptive lqr for completely unknown lti systems. *IFAC-PapersOnLine*, 50(1):4156–4161, 2017.

- [KM18] V. Kuznetsov and M. Mohri. Theory and Algorithms for Forecasting Time Series. *ArXiv e-prints*, March 2018.
- [nLa18] nLab authors. Lagrangian correspondence. <http://ncatlab.org/nlab/show/Lagrangian%20correspondence>, April 2018. Revision 11.
- [SKK<sup>+</sup>17] Inkyu Sa, Mina Kamel, Raghav Khanna, Marija Popovic, Juan Nieto, and Roland Siegwart. Dynamic system identification, and control for a cost effective open-source vtol mav. *arXiv preprint arXiv:1701.08623*, 2017.
- [SMT<sup>+</sup>18] M. Simchowitz, H. Mania, S. Tu, M. I. Jordan, and B. Recht. Learning Without Mixing: Towards A Sharp Analysis of Linear System Identification. *ArXiv e-prints*, February 2018.
- [VTK04] Henning U Voss, Jens Timmer, and Jürgen Kurths. Nonlinear dynamical system identification from uncertain and indirect measurements. *International Journal of Bifurcation and Chaos*, 14(06):1905–1933, 2004.
- [WGB<sup>+</sup>18] P. Wenk, A. Gotovos, S. Bauer, N. Gorbach, A. Krause, and J. M. Buhmann. Fast Gaussian Process Based Gradient Matching for Parameter Identification in Systems of Nonlinear ODEs. *ArXiv e-prints*, April 2018.
- [WW09] K. Wehrheim and C. T. Woodward. Floer Cohomology and Geometric Composition of Lagrangian Correspondences. *ArXiv e-prints*, May 2009.
- [WWJ16] Andre Wibisono, Ashia C Wilson, and Michael I Jordan. A variational perspective on accelerated methods in optimization. *Proceedings of the National Academy of Sciences*, 113(47):E7351–E7358, 2016.