

DPhil: Thesis

Shaan Desai

June 30, 2020

Contents

1	Introduction	1
2	Literature Review	2
2.1	Artificial Intelligence	2

1 Introduction

The Euler-Lagrange formulation equates to:

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{q}} \right) = \frac{\partial L}{\partial q}$$

where:

$$L = T - U$$

In the example of a pendulum:

$$T = 1/2mv^2 = 1/2m(l\dot{\theta})^2 = 1/2ml^2\dot{\theta}^2$$

$$U = mgh = mgl(1 - \cos \theta)$$

$$L = 1/2ml^2\dot{\theta}^2 - mgl(1 - \cos \theta)$$

Placing the above into the euler-lagrange form gives us:

LHS:

$$\frac{\partial L}{\partial \dot{q}} = ml^2\dot{\theta}$$

$$\frac{\partial}{\partial t} \left(\frac{\partial L}{\partial \dot{q}} \right) = ml^2\ddot{\theta}$$

RHS:

$$\frac{\partial L}{\partial q} = -mgl \sin \theta$$

Therefore:

$$\ddot{\theta} = \frac{g}{l} \sin \theta$$

Now, to transition to the Hamiltonian version:

We can write the hamiltonian as the sum of the energies:

$$H = T + U$$

$$H = p^2/2m + U$$

$$H = -L + 2K$$

$$H = -L + mv^2$$

$$H = -L + p^2/m = -L + p * (m\dot{q})/m = -L + p\dot{q}$$

$$p = \frac{\partial L}{\partial \dot{q}}$$

2 Literature Review

2.1 Artificial Intelligence

- Humans have long tried to understand how the mind works - from a biological perspective it is clear that connectomics paves a pathway to understanding how we think - With the advent of computers and 'large' processing power, many researchers began to ask whether the human mind can be replicated using a computer. - Early work by walter and warren was the first to point out a means to build neural network logic. - Followed by Hebbian - Then an AI winter - Geof Hinton backprop - slew of advents in neural processing, convnets, rnns, graphs, tensor nets, deep RL, - all seeking to solve more tasks/increase generalizability. - to date, there has been no general purpose solution (AGI) for a vast array of tasks e.g. a human can learn to create art, sit math exams, play sports, yet a robot so far can only do one of these tasks well. - Numerous limitations in terms of processing power and memory access still exist

- Despite these limitations, neural networks have brought about significant advancements to society in robotics, self driving cars, cancer detection.

- In what follows, we outline some of the major advances, those which have been fundamental pillars to research conducted during this phd. We follow no particular order.

2.2 Inductive Biases

An inductive bias in machine learning is prior information used to guide model building. In simple linear regression, for example, we assume the distribution of the noise in our data follows a Gaussian. This is a natural inductive bias because prior to seeing any data, we have assumed our noise model follows this form. In neural networks, architectures with variable depth, form and activations for example, serve as inductive biases. In essence, we encode initial assumptions about the complexity of data via inductive biases.

A feedforward deep neural network naturally encodes 'nonlinearity' as a prior and arguably makes the least assumptions about the underlying data. Convolutional Neural Networks assume spatial representations can be captured. Recurrent Neural Networks assume temporal data as inputs.

In what follows, we highlight some of the recent advances in inductive biases, particularly as they relate to solving problems in physics.

2.3 Graph Neural Networks

Graph Neural Networks are neural networks which use graphs as inputs. As such, they can be used to understand and learn the relationships which exist between the nodes of a graph. More formally, a graph is a tuple (Vertices, Edges, Globals). These are attributes of the graph where the nodes contain information, edges contain relationship weights and globals consist of macro variables general to all graphs.

interaction-physics based problems naturally benefit from graphs. We see that HOGNs and OGNs achieve great results in rolling out trajectories of N body systems.

Although they account for a new set of relational inductive biases, graphs still involve complex message passing schemes that involve significant fine-tuning. In addition, they assume data streams are already structured. However, graphs have shown significant promise in physics e.g. in explaining the phase transition of glassy materials. Their use in physics opens up a tried-and-tested pathway to solve more complex problems particularly in accounting for material interactions.

They're even shown to model complex fluid like systems from visual data.

Most systems to date have looked at graphs for classical physics but literature from 2004 suggests graphs, inherent in their relational structure, can also capture Ising-like hamiltonian structure.

A brief look at residual nets illustrates that these iterative methods look like euler discretization steps. Adding more layers and taking smaller steps can result in a parametrization of a continuous dynamics:

$$\frac{dh}{dt} = f(h(t), t, \theta)$$

Essentially the neural network hidden layers contain the components of the differential equation. The input layer $h(0)$ gives us an IVP which we can evolve

to time T . But this is a 'continuous-depth' network and we need a way to backprop. The method to backprop is *adjoint sensitivity method* which computes gradients by solving a second, augmented ODE back in time.

Neural networks have shown remarkable success at solving an array of problems in vision, robotics and healthcare. However, these networks still struggle to learn the basic laws of physics. [?] show that we can take a physical system (i.e. position and momentum values or images of a moving pendulum) and map it to a latent space which can encode the dynamics of the system. One way to learn these dynamics is to train the network to learn a latent space which is equivalent to the rate of change of momentum/position. However, [?] show that using this simple approach is not enough in being able to accurately predict energy conserving phase space trajectories. Instead, they show that if we treat the latent space as a vector field (symplectic gradient) or as a parametric function for the hamiltonian then we can differentiate it with respect to the input (p and q) and obtain energy conservation laws. In this way, the neural network is parametrized to encode the hamiltonian of the system.

An alternative approach to this method is that of variational integrator networks proposed by [?]. Inspired by neural-ODE's, they propose a neural net whose architecture matches the discrete equation of motion governing the dynamical system. Essentially, the embedding learned by the network is a dynamical system itself. They use a variational integration discretization - this method only introduces 3rd order errors w.r.t energy and closely resembles the continuous time dynamics. The approach follows the neural ODE approach with two changes, the ODE we model is the Euler-Lagrange from free-form dynamical system. In addition the discretization is a VI.