Where Is Noether's Principle in Machine Learning?

Christopher Gadzinski (me@cgad.ski)

UNIVERSITÉ DU LUXEMBOURG

Overview

Physics likes optimization! Subject to its boundary conditions, the time-evolution of a physical system is a stationary point for a quantity called an **action**. Furthermore, continuous invariances of the action turn out to imply **conservation laws** of the system, like conservation of energy and momentum. This is called Noether's principle.

In machine learning, we often deal with discrete "processes" whose control parameters are chosen to minimize some quantity. For example, we can see a deep residual network as a process where the role of "time" is played by depth. We may ask:

1. Does Noether's principle apply?
2. Can we find meaningful conserved quantities?

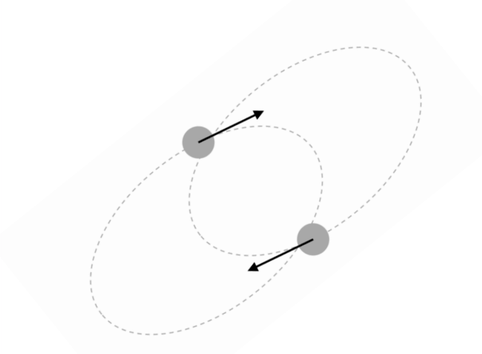
Our answers: "**yes**," and "**not sure!**"

interactive version

Noether's Principle in Physics

The two body problem has an action invariant under rotation and translation…

$$S = \int \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2) - \frac{G}{\|q_1 - q_2\|}\, dt$$

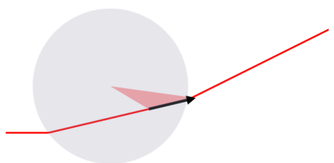


…and therefore has conservation of **momentum:**

$$\dot{q}_1 + \dot{q}_2 = \text{constant},$$
$$\dot{q}_1 \wedge q_1 + \dot{q}_2 \wedge q_2 = \text{constant}.$$

When a beam of light passes through a rotationally symmetric lens…

$$S = \int \frac{1}{v(q)} \|\dot{q}\|\, dt,$$

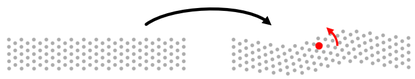


…the **angular component of the gradient** of position with respect to time is conserved:

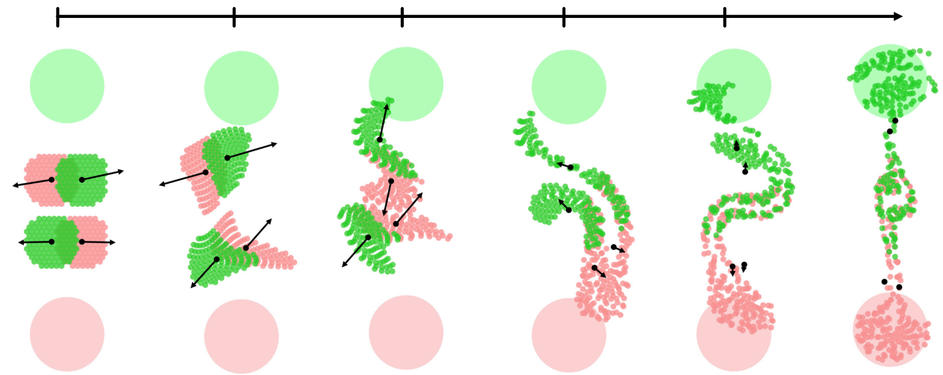
$$\frac{\dot{q} \wedge q}{v(q)\|\dot{q}\|} = \text{constant}.$$

Noether's Principle in ML: A Toy Example

Let's compose many small "distortions" of the plane so that some red and green points are mapped close to their destinations. How do our clusters "evolve" as more distortions are composed?



We'll also keep track of the gradients of the loss at each intermediate point. Black arrows show the **average gradient** over each cluster.



Average gradient is **conserved!** Why? We can explain using a version of Noether's principle:

Gradient is conserved because our family of maps is **invariant under translation.**

General Statement

Define the **action** $S = \langle g_x, x\rangle - \langle g_y, \varphi_\theta(x) - y\rangle$ and suppose that some transformation group leaves the second term **invariant**.

For example, suppose that

"Noether equivariance"

$$\varphi_{g\cdot\theta}(g \cdot x) = g \cdot \varphi_\theta(x).$$

Then we have a "conservation law"

$$\left\langle g_x, \frac{\partial x}{\partial \epsilon} \right\rangle = \left\langle g_y, \frac{\partial y}{\partial \epsilon} \right\rangle$$

for each dimension of the transformation group.

Significance

In the physical world, interactions are often transfers of conserved quantities.

If the layers of a deep model have meaningful "Noether equivariances," then we can build conserved quantities and quantify "interaction" between inputs through an optimized model. (See the example above.)

I don't know if this idea works! Can you think of non-trivial equivariances that we might find in practice? (Keep in mind they might only be valid near the actual parameters of the model.)