Proceedings Track

Dynamical symmetries in the fluctuation-driven regime: an application of Noether's theorem to noisy dynamical systems

Abstract

Noether's theorem provides a powerful link between continuous symmetries and conserved quantities for systems governed by some variational principle. Perhaps unfortunately, most dynamical systems of interest in neuroscience and artificial intelligence cannot be described by any such principle. On the other hand, nonequilibrium physics provides a variational principle that describes how fairly generic noisy dynamical systems are most likely to transition between two states; in this work, we exploit this principle to apply Noether's theorem, and hence learn about how the continuous symmetries of dynamical systems constrain their most likely trajectories. We identify analogues of the conservation of energy, momentum, and angular momentum, and briefly discuss examples of each in the context of models of decision-making, recurrent neural networks, and diffusion generative models.

Keywords: symmetry, invariance, Noether's theorem, stochastic processes, diffusion

1. Introduction

In physics, Noether's theorem provides a fundamental link between the symmetries of physical systems on the one hand, and conserved quantities like energy and momentum on the other hand (Noether, 1918; Kosmann-Schwarzbach et al., 2011; Neuenschwander, 2017). In its modern form, it uniquely associates (equivalence classes of) independent continuous symmetries, which can be formalized in terms of Lie groups and algebras, with (equivalence classes of) independent conserved quantities (Martinez Alonso, 1979; Olver, 1986, 1993; Brown, 2020). Importantly, it allows us to make nontrivial statements about the dynamics of systems which may be otherwise extremely difficult to analyze.

The utility of Noether's theorem has inspired a variety of extensions and applications, including in the context of classical mechanics (Sarlet and Cantrijn, 1981), non-variational partial differential equations (Anco, 2017), statistical mechanics (Hermann and Schmidt, 2021), and Markov models (Baez and Fong, 2013). But it is worth emphasizing that the physical systems considered in typical applications of Noether's theorem—whose dynamics can be characterized as extremizing a certain functional—are extremely special. Most dynamical systems, including those most relevant to neuroscience and artificial intelligence (AI), *cannot* be described in terms of some nontrivial variational principle. Moreover, many systems of interest exhibit features that suggest they are not even 'almost' describable by some variational principle. For example, most dynamical systems lack a notion of 'inertia'; that is, their dynamics are completely determined by their current state, rather their current state and some notion of state velocity. Additionally, most dynamical systems are not gradient-like, i.e., the function governing how states evolve in time is not the gradient of some scalar potential. Either feature can yield major qualitative differences in dynamics: for example, 'normal' versus 'non-normal' dynamics (Hennequin et al., 2012; Kerg et al., 2019) maps onto the gradient-like versus non-gradient-like distinction.

Nonetheless, one still hopes that symmetries can be exploited to say *something* about the dynamics of many systems of interest, including neural networks. Of special note is the work of Tanaka and Kunin (2021), which says that neural network learning dynamics can be associated with a certain Lagrangian when gradient descent is performed in discrete steps, and hence that a version of Noether's theorem holds. But what can be said about more general dynamics not necessarily tied to learning? Consider, for example, the firing rate dynamics of some trained recurrent neural network (RNN); given symmetries, what can we say about the network's trajectory through state space?

Perhaps surprisingly, there is a sense in which Noether's theorem applies to extremely general noisy dynamical systems, including noisy RNNs and diffusion models; in this work, our goal is to explore this application and interpret the associated conserved quantities. The idea is that the transient and fluctuation-driven behavior (e.g., jumps between neighboring attractor basins) of such systems is well-described by a certain variational principle, which allows us to apply Noether's theorem to learn about how state transitions are most likely to occur. Noether's theorem tells us that even though noise plays a crucial role in many of these transitions, their form is not arbitrary, but in fact constrained by a system's symmetries.

The paper is organized as follows. First, we review Noether's theorem and some of its well-known consequences. Next, we (re-)derive a variational principle that allows us to apply Noether's theorem to noisy dynamical systems. Finally, we discuss various consequences of this application of Noether's theorem, including analogues of energy and momentum conservation. We consider a few example systems relevant to neuroscience and AI, including drift-diffusion models of decision-making, noisy RNNs, and diffusion generative models.

2. Noether's theorem, symmetries, and conserved quantities

What is colloquially referred to as "Noether's theorem" was first presented in a paper by Emmy Noether in 1918 (Noether, 1918). Two subtleties are worth mentioning: first, Noether's paper actually contains two theorems, and "Noether's theorem" without additional qualifications usually refers to the first; second, although the theorem is usually invoked to show that continuous symmetries imply conserved quantities, the converse direction (conserved quantity $\implies$ symmetry) actually also holds when one is careful about how one defines "symmetry" and "conserved quantity". See Brown (2020) and Olver (2018) for helpful discussions of the history and precise formulation of the converse direction.

For our purposes, the simple version of Noether's (first) theorem presented in physics textbooks on mechanics and field theory will suffice. The standard story is usually told in terms of actions and Lagrangians; while an alternative formulation which exploits a Hamiltonian rather than Lagrangian viewpoint is also available, and arguably simplifies the picture in a number of ways (Baez, 2020), we will not consider it here. Since our applications involve trajectories parameterized only by one variable, time, we present the 'particle' rather than the 'field' version of Noether's theorem.

Noether's theorem applies to systems described by a variational principle. Let $\boldsymbol{x}(t) := (x_1(t), ..., x_N(t))^T \in \mathbb{R}^N$ denote the state of one such system at time $t \in [t_0, t_f]$. We assume that the system's trajectory $\boldsymbol{x}(t)$ extremizes the 'action' functional

$$\mathcal{S}[\boldsymbol{x}] := \int_{t_0}^{t_f} L(\boldsymbol{x}(t), \dot{\boldsymbol{x}}(t), t) \; dt \tag{1}$$

where the function L is the corresponding Lagrangian. By a standard calculus of variations result, this assumption implies the system's trajectory satisfies the Euler-Lagrange equations

$$\frac{\partial L}{\partial \boldsymbol{x}} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\boldsymbol{x}}}\right) \qquad \text{or equivalently} \qquad \frac{\partial L}{\partial x_i} = \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_i}\right) \quad \text{for } i = 1, ..., N \ . \qquad (2)$$

A continuous symmetry of such a system is a continuous transformation that leaves the action $\mathcal{S}$ invariant (regardless of whether the equations of motion are satisfied). The more general notion of a continuous *quasi*-symmetry refers to a continuous transformation that leaves the action *almost* invariant—i.e., up to a surface term. Noether's theorem is about the consequences of quasi-symmetry given that the equations of motion are satisfied.

Theorem 1 (Noether's theorem) *Let $\delta t \geq 0$ and $\delta \boldsymbol{x} \in \mathbb{R}^N$, and consider a transformation that takes $t \to t' := t + \epsilon\,\delta t$ and $\boldsymbol{x} \to \boldsymbol{x}' := \boldsymbol{x} + \epsilon\,\delta\boldsymbol{x}$. If the Lagrangian L satisfies*

$$L(\boldsymbol{x}', \dot{\boldsymbol{x}}', t') = L(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) + \frac{d}{dt}K(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) + \mathcal{O}(\epsilon^2) \qquad (3)$$

for all $\epsilon > 0$ sufficiently small, where K is some function, then the quantity

$$J := \sum_i \frac{\partial L}{\partial \dot{x}_i}\delta x_i - K \qquad (4)$$

is conserved in the sense that $dJ/dt = 0$ when the equations of motion are satisfied.

The proof of this version of Noether's theorem is standard, so we do not include it here. Well-known consequences of Noether's theorem include the conservation of energy, which follows from (quasi-)symmetry with respect to time translations; the conservation of momentum, which follows from symmetry with respect to translations along one or more directions of state space; and the conservation of angular momentum, which follows from symmetry with respect to state space rotations within one or more planes.

Importantly, although the aforementioned quantities have specific forms for typical physical systems—e.g., energy is kinetic plus potential energy, and momentum is mass times velocity—one can derive *analogues* of the more familiar energy, momentum, and so on as long as the associated symmetries still hold for a non-physical system that can be characterized by a Lagrangian. We exploit this fact in what follows.

3. Applying Noether's theorem to stochastic dynamical systems

Let us now consider a somewhat different type of system defined by a set of (Ito-interpreted) stochastic differential equations (SDEs)

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, t) + \boldsymbol{G}(\boldsymbol{x}, t)\,\boldsymbol{\eta}_t \qquad (5)$$

where $\boldsymbol{x}(t) := (x_1(t), ..., x_N(t))^T \in \mathbb{R}^N$ denotes the system's state at time $t \in [t_0, t_f]$, $\boldsymbol{f} : \mathbb{R}^N \times [t_0, t_f] \to \mathbb{R}^N$ determines the drift term, $\boldsymbol{G} : \mathbb{R}^N \times [t_0, t_f] \to \mathbb{R}^{N \times M}$ determines the noise amplitude, and $\boldsymbol{\eta}_t$ is an M-dimensional vector of independent Gaussian white noise terms. For technical reasons, we assume that the $N \times N$ diffusion tensor $\boldsymbol{D}(\boldsymbol{x}, t) := \frac{1}{2}\boldsymbol{G}(\boldsymbol{x}, t)\boldsymbol{G}(\boldsymbol{x}, t)^T$ is positive definite for all $\boldsymbol{x}$ and t, and hence always invertible.

To apply Noether's theorem to systems described by Eq. 5, we must identify a relevant variational principle. But from the form of Eq. 5 alone, it is not clear why we should expect one to exist, or what form it should take if it does. A path forward becomes clearer if one exploits the rich analogy between quantum mechanics and stochastic dynamics.

In one view, quantum mechanical systems are characterized by the time evolution of their wave function $\psi(\boldsymbol{x}, t)$, which encodes the probability of each possible state occupancy observation. The probability *amplitude* associated with a transition from state $\boldsymbol{x}_0$ at time t_0 to a state $\boldsymbol{x}$ at time t_f can be written as (Feynman et al., 2010; Kleinert, 2006)

$$K(\boldsymbol{x}, t_f; \boldsymbol{x}_0, t_0) = \int \mathcal{D}[\boldsymbol{x}] \exp\left\{\frac{i}{\hbar} \mathcal{S}[\boldsymbol{x}]\right\} \tag{6}$$

where the integral is over all possible paths from $\boldsymbol{x}_0$ to $\boldsymbol{x}$, and where $\mathcal{S}$ is the system's associated action functional. What makes this particular description of a quantum system's behavior interesting is that it makes the connection between quantum and classical mechanics more explicit: each possible path from $\boldsymbol{x}_0$ to $\boldsymbol{x}$ contributes to the above path integral, but the path that extremizes $\mathcal{S}$—i.e., the classical path—contributes the most, and dominates in the $\hbar \to 0$ limit, or equivalently when quantum effects are taken to be negligible. Said differently, the above path integral description provides us with a variational principle that describes what path the system is most likely to take from $\boldsymbol{x}_0$ to $\boldsymbol{x}$.

A path integral description of stochastic dynamics also exists, and says the probability (rather than probability amplitude) associated with an $\boldsymbol{x}_0$ to $\boldsymbol{x}$ transition can be written

$$p(\boldsymbol{x}, t_f; \boldsymbol{x}_0, t_0) = \int \mathcal{D}[\boldsymbol{x}] \exp\left\{-\mathcal{S}[\boldsymbol{x}]\right\} \tag{7}$$

where the Lagrangian associated with *this* action is

$$L(\boldsymbol{x}, \dot{\boldsymbol{x}}, t) := \frac{1}{4} \left[\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, t)\right]^T \boldsymbol{D}^{-1}(\boldsymbol{x}, t)[\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, t)] . \tag{8}$$

This construction is usually called the Onsager-Machlup path integral (Onsager and Machlup, 1953; Cugliandolo and Lecomte, 2017), and is essentially equivalent to the Martin-Siggia-Rose (Martin et al., 1973) path integral. The Lagrangian L here is a quadratic form which can be interpreted as follows: the system *usually* follows its drift term, but can fluctuate in other directions, with the strength and direction of those fluctuations being controlled by the diffusion tensor. In directions with higher noise, fluctuations incur a smaller action cost, and hence they are more likely. Unlike in the case of quantum and classical mechanics, the most likely path is the one which *minimizes* $\mathcal{S}$, rather than makes it stationary. This path dominates the path integral in the small noise limit—yielding a stochastic dynamics analogue to classical mechanics—as one can formally show by exploiting Laplace's method.

Now that we have a variational principle associated with Eq. 5, we can apply Noether's theorem; in the next section, we introduce analogues of familiar conserved quantities.

4. Energy, momentum, and angular momentum conservation

Applying Noether's theorem to our Lagrangian (Eq. 8) tells us how a system's continuous symmetries and quasi-symmetries constrain the most likely path it takes between any two

states, including paths that require noise, like transitions between attractor basins. Since this Lagrangian can exhibit some of the same (quasi-)symmetries as physical systems, one can derive analogues of energy, momentum, and angular momentum conservation. Given that we started with a fairly arbitrary stochastic dynamical system (e.g., $\boldsymbol{f}$ is not necessarily the gradient of any potential function), this may be somewhat surprising.

A caveat, however: one should not get too greedy. While one might hope that applying Noether's theorem here produces conserved quantities that remain interesting as one takes the noise in the system to zero, since many dynamical systems of interest are deterministic, we will see that this does not happen. All of the conservation laws we obtain reduce to trivialities in the zero noise limit, which might be expected in light of the fact that we are not making any assumptions about $\boldsymbol{f}$. One cannot get something from nothing, after all.

4.1. Energy conservation

Assume our Lagrangian does not depend explicitly on time, which is equivalent to demanding that neither $\boldsymbol{f}$ nor $\boldsymbol{D}$ depends explicitly on time. Then an infinitesimal time translation $t' := t + \epsilon$ produces a state perturbation $\boldsymbol{x}(t') = \boldsymbol{x}(t) + \dot{\boldsymbol{x}}(t)\,\epsilon$ and changes L as

$$L \to L + \epsilon \left[\frac{\partial L}{\partial \boldsymbol{x}} \cdot \dot{\boldsymbol{x}} + \frac{\partial L}{\partial \dot{\boldsymbol{x}}} \cdot \ddot{\boldsymbol{x}} \right] = L + \epsilon \left[\frac{dL}{dt} - \frac{\partial L}{\partial t} \right] = L + \epsilon \frac{dL}{dt} \tag{9}$$

where we used the fact that L does not depend explicitly on time in the last step. By Noether's theorem, the quantity

$$E := \frac{\partial L}{\partial \dot{\boldsymbol{x}}} \cdot \dot{\boldsymbol{x}} - L = \frac{1}{4}\dot{\boldsymbol{x}}^T \boldsymbol{D}^{-1} \dot{\boldsymbol{x}} - \frac{1}{4}\boldsymbol{f}^T \boldsymbol{D}^{-1} \boldsymbol{f} \tag{10}$$

is conserved, i.e., $dE/dt = 0$ when the equations of motion are satisfied. Considering special cases helps us make sense of this quantity. Suppose $\boldsymbol{D}$ is a diagonal matrix and independent of $\boldsymbol{x}$. If $\boldsymbol{f} = -\boldsymbol{D}\frac{\partial V}{\partial \boldsymbol{x}}$ for some scalar potential $V(\boldsymbol{x})$, our original L is equivalent to

$$L' = \frac{1}{4}\,\dot{\boldsymbol{x}}\boldsymbol{D}^{-1}\dot{\boldsymbol{x}} + \frac{1}{4}\,\boldsymbol{f}\boldsymbol{D}^{-1}\boldsymbol{f}\,, \tag{11}$$

since the cross term $-\frac{1}{2}\dot{\boldsymbol{x}}\frac{\partial V}{\partial \boldsymbol{x}}$ is equal to a total time derivative, and hence does not contribute to the equations of motion. This Lagrangian 'looks' like a kinetic energy term (the mass associated with x_i is $\frac{1}{2D_{ii}}$) and a potential term $W(\boldsymbol{x}) := \frac{1}{4}\boldsymbol{f}(\boldsymbol{x})^T \boldsymbol{D}^{-1} \boldsymbol{f}(\boldsymbol{x})$, except that the potential term has the 'wrong' sign. Similarly, our expression for E becomes kinetic *minus* potential energy, which is opposite how E looks for physical systems. Another informative special case assumes $\boldsymbol{D}$ is state-independent and isotropic, i.e., $\boldsymbol{D} = D\boldsymbol{I}_N$, so

$$E = \frac{\|\dot{\boldsymbol{x}}\|_2^2 - \|\boldsymbol{f}(\boldsymbol{x})\|_2^2}{4D} = \sum_i \frac{\dot{x}_i^2 - f_i(\boldsymbol{x})^2}{4D}\,. \tag{12}$$

The above expression makes clear that the conservation of energy is a statement about how the system's fluctuation-driven trajectory differs from the deterministic trajectory determined only by its drift term $\boldsymbol{f}$. The squared norm of the 'state velocity' $\dot{\boldsymbol{x}}$ deviates from that of the drift term in general, but *by the same amount* throughout its most likely transition between two states. When the system follows its deterministic trajectory, $E = 0$.

Since E is greater when the difference between the norms of $\dot{\boldsymbol{x}}$ and $\boldsymbol{f}(\boldsymbol{x})$ is large, we might expect the size of E to be related to the time the system takes to transition between two states. (As a related point, note that E has units of inverse time.) This turns out to be true, and for 1D systems one has a one-to-one relationship through the easy-to-derive

$$t_* = \int_{x_0}^{x_f} \frac{dx}{\sqrt{f(x)^2 + 4D(x)E}} \ . \tag{13}$$

In other words, the transition time t_* is strictly decreasing as a function of energy.

4.2. Momentum conservation

Assume our Lagrangian does not depend explicitly on x_i, i.e., that neither $\boldsymbol{f}$ nor $\boldsymbol{D}$ does. Then an infinitesimal translation $x_i' \to x_i + \epsilon$ along direction i does not change L, and

$$p_i := \frac{\partial L}{\partial \dot{x}_i} = \frac{1}{2} \sum_j D_{ij}^{-1} \left[\dot{x}_j - f_j(\boldsymbol{x}, t) \right] \tag{14}$$

is conserved. More generally, momentum is defined as the N-dimensional vector

$$\boldsymbol{p} := \frac{\partial L}{\partial \dot{\boldsymbol{x}}} = \frac{1}{2} \boldsymbol{D}^{-1}(\boldsymbol{x}, t) \left[\dot{\boldsymbol{x}} - \boldsymbol{f}(\boldsymbol{x}, t) \right] \tag{15}$$

and can be conserved in up to N independent directions. While energy appears to measure the overall extent to which the deterministic dynamics are violated, different components of $\boldsymbol{p}$ quantify how badly this happens along each direction of state space. As in the case of energy, when the system follows its deterministic trajectory, $\dot{\boldsymbol{x}} = \boldsymbol{f}$, so $\boldsymbol{p} = \boldsymbol{0}$.

4.3. Angular momentum conservation

Assume our Lagrangian is invariant to rotations within the x_i-x_j plane, and in particular to infinitesimal rotations of the form $x_i' = x_i - \epsilon x_j$, $x_j' = x_j + \epsilon x_i$. By Noether's theorem,

$$L_{ij} := \frac{\partial L}{\partial \dot{x}_i} \delta x_i + \frac{\partial L}{\partial \dot{x}_j} \delta x_j = x_i p_j - x_j p_i \tag{16}$$

is conserved. More generally, we can define an $N \times N$ angular momentum tensor

$$\boldsymbol{L} := \boldsymbol{x} \boldsymbol{p}^T - \boldsymbol{p} \boldsymbol{x}^T = \frac{1}{2} \boldsymbol{x} \boldsymbol{D}^{-1} (\dot{\boldsymbol{x}} - \boldsymbol{f})^T - \frac{1}{2} (\dot{\boldsymbol{x}} - \boldsymbol{f}) \boldsymbol{D}^{-1} \boldsymbol{x}^T \tag{17}$$

which is antisymmetric and has up to $N(N-1)/2$ distinct entries (not coincidentally, the number of generators of the rotation group SO(N)), each of which may be conserved. The conservation of L_{ij} implies that the extent to which dynamics deviate from the deterministic trajectory along the x_i direction constrains the extent to which they can deviate along the x_j direction. When the system follows its drift term, $p_i = p_j = 0$ and hence $L_{ij} = 0$.

5. Case studies

It is finally time to apply the conservation laws we identified to example settings relevant to neuroscience and AI: (i) drift-diffusion models of decision-making, (ii) RNNs with point attractors, and (iii) diffusion generative models. Where numerical solutions for transition paths are required, we use the action minimization approach of Strang et al. (2023).
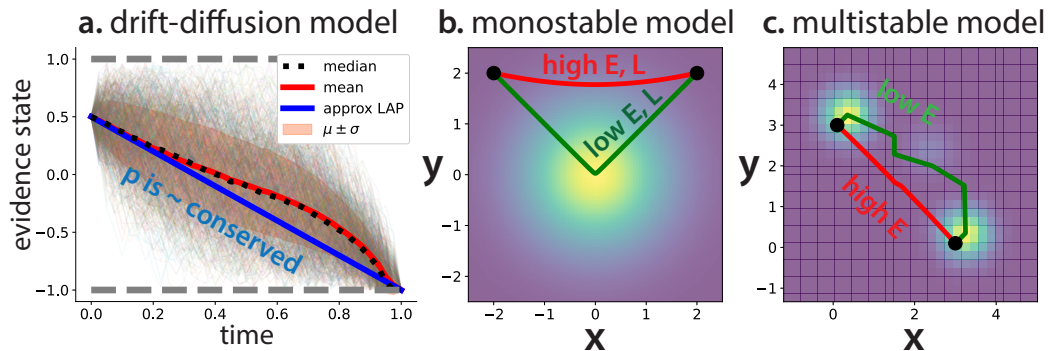
Figure 1: Conservation laws relevant to simple decision-making and decision memory models. **a.** For a drift-diffusion model with fixed decision bounds at ± 1, the most likely path from any evidence state to the opposite boundary is (approximately) a straight line (blue, 'LAP'). Raw paths ($n = 981$) with same start state, end state, and transition time also shown. **b.** Transitions between two states of a single-attractor model (heatmap: steady state distribution) are direct when energy and angular momentum are high, and involve a diversion to the attractor state when they are low. **c.** Transitions between two attractor basins in a model with three attractors (heatmap: steady state distribution) are more direct when energy is high, and involve visiting an intermediate attractor when energy is low.

5.1. Drift-diffusion models of decision-making

Drift-diffusion models (Ratcliff et al., 2016) purport to describe the dynamics of decision-making in various multiple-alternative forced choice tasks (e.g., are dots on a screen moving mostly to the left or right?). They involve two components: Brownian-like particle dynamics

$$\dot{x} = v + \sigma \, \eta_t \qquad\qquad t \in [0, \infty) \qquad\qquad (18)$$

where $v = $ const. is the drift term (or 'bias') and σ is the noise strength, and boundary conditions that specify how the particle's state determines a decision. In the simplest case, which we will assume here, a decision happens when x reaches either an upper or lower threshold, say, at ± 1. More complex boundary shapes are often considered, e.g., time-dependent bounds that model decision urgency. But even simple drift-diffusion models explain available behavioral and neural data surprisingly well (Gold and Shadlen, 2007).

The decision usually matches the sign of the drift term, but fluctuations, which can be viewed as representing random variations in the decision-maker's incoming evidence, can cause the particle to move in the opposite direction. One can ask: given that the system makes the 'wrong' decision at some time $T \geq 0$, what path through state space did the system most likely take? Or equivalently: did the decision-maker most likely receive a sudden burst of contrary evidence, a steady stream of it, or something else?

For $f = v$ and $D = \sigma^2/2$, if we neglect decision-boundary-related effects, we have a Lagrangian that does not depend explicitly on x, so momentum is conserved. Then

$$p = \frac{\dot{x} - v}{\sigma^2} = \text{const.} \implies x(t) = x_0 + (v + \sigma^2 p)t = x_0 + \frac{(x_f - x_0)}{T}t \; , \qquad (19)$$

i.e., that the system is most likely to travel from x_0 to x_f in an amount of time T with a constant velocity. In other words, the most likely way the decision-maker makes the wrong choice is by unluckily receiving a steady stream of contrary evidence. In practice, the presence of the decision boundaries modifies this conclusion, but only slightly (Fig. 1a).

5.2. Changes of mind in an attractor network model of decision memory

Animal decision-making is thought to involve at least three processes: (i) evidence accumulation, (ii) weighing evidence to come to a decision, and (iii) storing that decision for later use. Attractor network models, where each possible decision corresponds to a different attractor basin, provide a popular possible means by which neural circuits could achieve some combination of the above three aims, but especially the last (Machens et al., 2005; Kopec et al., 2015; Piet et al., 2017).

Our conservation laws provide a way to think about movement both within and between basins, assuming no explicit time-dependence. Consider a 2D Ornstein-Uhlenbeck process as a model of within-attractor dynamics, and note both energy E and angular momentum L are conserved. State transitions tend to be more direct when E and L are high (i.e., when the transition is fast), and to visit the attractor's fixed point when they are low (Fig. 1b).

Interestingly, multistable attractor network models predict that 'changes of mind' are possible depending on the distance between and relative shallowness of different attractor basins (Albantakis and Deco, 2011). For example, if the available evidence is somewhat ambiguous during a two-alternative forced choice task, the basins may be relatively shallow and close together, and one might expect a relatively high rate of 'hops' between them.

Consider a modified version of the model from Piet et al. (2017) with

$$\tau\dot{x} = \mu_0 + \frac{A}{2}\left[\tanh(\frac{x-c}{n}) + 1\right] - \frac{I}{2}\left[\tanh(\frac{y-c}{n}) + 1\right] - x + \sqrt{2\tau}\sigma\ \eta_x$$
$$\tau\dot{x} = \mu_0 + \frac{A}{2}\left[\tanh(\frac{y-c}{n}) + 1\right] - \frac{I}{2}\left[\tanh(\frac{x-c}{n}) + 1\right] - y + \sqrt{2\tau}\sigma\ \eta_y \tag{20}$$

which is a multistable attractor network. The parameters μ_0, A, I, and σ control the size and number of attractor basins. We consider a parameter set with three stable states, and interpret one as an 'undecided' state. Given a transition between the two choice-related basins, will this system stop at the 'undecided' state, or not? The answer depends on the system's energy E. When E is high, the system takes a direct path; when E is low, the system must visit the intermediate state en route to the other choice basin (Fig. 1c).

5.3. Data manifold symmetries and reverse diffusion dynamics

Diffusion generative models provide a performant method for sampling from realistic and high-dimensional data distributions (Sohl-Dickstein et al., 2015; Song et al., 2021; Yang et al., 2023). The core idea is to corrupt a data distribution via a forward stochastic process, and then to learn to reverse the mapping, and hence generate samples from noise:

$$\dot{\boldsymbol{x}} = \sqrt{2t}\ \boldsymbol{\eta}_t \qquad t = 0 \to t = T \qquad \text{forward process, } p_{data} \text{ to } p_{noise} \tag{21}$$
$$\dot{\boldsymbol{x}} = -t\ \boldsymbol{s}(\boldsymbol{x}, t) \qquad t = T \to t = 0 \qquad \text{reverse process A, } p_{noise} \text{ to } p_{data}$$
$$\dot{\boldsymbol{x}} = -2t\ \boldsymbol{s}(\boldsymbol{x}, t) + \sqrt{2t}\ \boldsymbol{\eta}_t \quad t = T \to t = 0 \qquad \text{reverse process B, } p_{noise} \text{ to } p_{data}$$

8

Proceedings Track

Here, $p(\boldsymbol{x}|t)$ is the noise-corrupted data distribution and $\boldsymbol{s}(\boldsymbol{x},t) := \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}|t)$ is the score function. Two reverse processes are shown: the popular probability-flow ODE (PF-ODE), and the reverse SDE, which involves noise (Song et al., 2021). In practice, either can outperform the other sample-quality-wise depending on the data set (Karras et al., 2022). How do the reverse diffusion trajectories of the two approaches differ?

We can exploit (rotation) symmetry to provide an answer in an interesting special case. Consider a Gaussian defined on a 2D ring of radius R, whose noise-corrupted version is

$$p_{data}(\boldsymbol{x}|t) := \frac{1}{2\pi(\sigma_0^2 + t^2)} \int_0^{2\pi} e^{-\frac{(x - R\cos\theta)^2}{2(\sigma_0^2 + t^2)} - \frac{(y - R\sin\theta)^2}{2(\sigma_0^2 + t^2)}} \frac{d\theta}{2\pi} = \frac{e^{-\frac{(\|\boldsymbol{x}\|_2^2 + R^2)}{2(\sigma_0^2 + t^2)}}}{2\pi t^2} I_0\left(\frac{R}{\sigma_0^2 + t^2}\|\boldsymbol{x}\|_2\right)$$

where I_0 denotes the modified Bessel function of the first kind. One can use this expression to analytically compute the score function. One finds that PF-ODE trajectories are not curved, but move towards the ring along the line perpendicular to its nearest tangent. Meanwhile, the stochastic sampler can produce trajectories that end at various other locations. The differences between these trajectories can be conveniently characterized in terms of angular momentum L, which is conserved by the rotation symmetry of the data distribution, and hence of the score; reaching points near the deterministic trajectory's end is 'easier', and hence requires less angular momentum, while reaching points on the other side of the ring requires greater angular momentum (Fig. 2).

6. Conclusion

The preceding investigation is a proof of principle that Noether's theorem can be applied to much more general settings, and that the conserved quantities we obtain by applying it (e.g., energy, momentum, angular momentum) recognizably modulate the most likely transitions of noisy dynamical systems. As in physics, a system's symmetries constrain its most likely behavior.

Several points suggest interesting directions for future work. First, it appears that Noether's theorem can be used to derive conserved quantities associated with neural networks that are not invariant, but equivariant (Cohen and Welling, 2016), to at least some transformations; our Lagrangian is rotation-invariant as long as $\boldsymbol{f}$ is rotation-equivariant, for example. Second, since the appropriately-formulated converse direction of Noether's theorem is true, it may be pos-
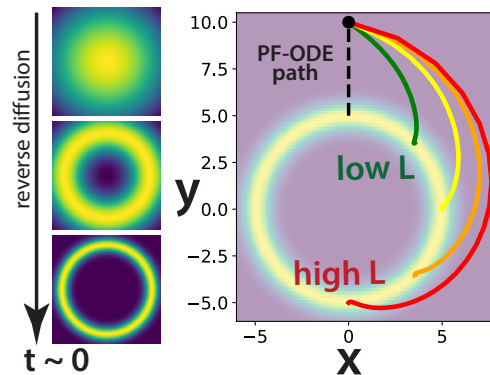

Figure 2: Angular momentum conservation in reverse diffusion. Left: depiction of reverse diffusion for a rotationally-symmetric data distribution. Right: most likely transition paths given a fixed starting point (black dot) and different angular momentum values (green is lowest, red is highest). Dashed black line is the PF-ODE trajectory.

sible to learn symmetries from empirically-identified conserved quantities. Finally, given the variety of geometric structures and symmetries relevant to modern AI (Sanborn et al., 2024), there are many more conserved quantities to explore beyond energy and momentum.

References

Larissa Albantakis and Gustavo Deco. Changes of mind in an attractor network of decision-making. *PLOS Computational Biology*, 7(6):1–13, 06 2011. doi: 10.1371/journal.pcbi. 1002086. URL https://doi.org/10.1371/journal.pcbi.1002086.

Stephen C. Anco. *Generalization of Noether's Theorem in Modern Form to Non-variational Partial Differential Equations*, pages 119–182. Springer New York, New York, NY, 2017. ISBN 978-1-4939-6969-2. doi: 10.1007/978-1-4939-6969-2_5. URL https://doi.org/10.1007/978-1-4939-6969-2_5.

John C. Baez. Getting to the Bottom of Noether's Theorem. *arXiv e-prints*, art. arXiv:2006.14741, June 2020. doi: 10.48550/arXiv.2006.14741.

John C. Baez and Brendan Fong. A Noether theorem for Markov processes. *Journal of Mathematical Physics*, 54(1):013301, 01 2013. ISSN 0022-2488. doi: 10.1063/1.4773921. URL https://doi.org/10.1063/1.4773921.

Harvey R. Brown. Do symmetries "explain" conservation laws? The modern converse Noether theorem vs pragmatism. *arXiv e-prints*, art. arXiv:2010.10909, October 2020. doi: 10.48550/arXiv.2010.10909.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In Maria Florina Balcan and Kilian Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 2990–2999, New York, New York, USA, 20–22 Jun 2016. PMLR. URL https://proceedings.mlr.press/v48/cohenc16.html.

Leticia F Cugliandolo and Vivien Lecomte. Rules of calculus in the path integral representation of white noise langevin equations: the onsager–machlup approach. *Journal of Physics A: Mathematical and Theoretical*, 50(34):345001, jul 2017. doi: 10.1088/1751-8121/aa7dd6. URL https://dx.doi.org/10.1088/1751-8121/aa7dd6.

Richard P Feynman, Albert R Hibbs, and Daniel F Styer. *Quantum mechanics and path integrals*. Courier Corporation, 2010.

Joshua I. Gold and Michael N. Shadlen. The neural basis of decision making. *Annual Review of Neuroscience*, 30(Volume 30, 2007):535–574, 2007. ISSN 1545-4126. doi: https://doi.org/10.1146/annurev.neuro.29.051605.113038. URL https://www.annualreviews.org/content/journals/10.1146/annurev.neuro.29.051605.113038.

Guillaume Hennequin, Tim P. Vogels, and Wulfram Gerstner. Non-normal amplification in random balanced neuronal networks. *Phys. Rev. E*, 86:011909, Jul 2012. doi: 10.1103/PhysRevE.86.011909. URL https://link.aps.org/doi/10.1103/PhysRevE.86.011909.

Sophie Hermann and Matthias Schmidt. Noether's theorem in statistical mechanics. *Communications Physics*, 4(1):176, Aug 2021. ISSN 2399-3650. doi: 10.1038/s42005-021-00669-2. URL https://doi.org/10.1038/s42005-021-00669-2.

Proceedings Track

Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=k7FuTOWMOc7.

Giancarlo Kerg, Kyle Goyette, Maximilian Puelma Touzel, Gauthier Gidel, Eugene Vorontsov, Yoshua Bengio, and Guillaume Lajoie. Non-normal recurrent neural network (nnrnn): learning long time dependencies while improving expressivity with transient dynamics. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/file/9d7099d87947faa8d07a272dd6954b80-Paper.pdf.

Hagen Kleinert. *Path integrals in quantum mechanics, statistics, polymer physics, and financial markets*. World Scientific Publishing Company, 2006.

Charles D. Kopec, Jeffrey C. Erlich, Bingni W. Brunton, Karl Deisseroth, and Carlos D. Brody. Cortical and subcortical contributions to short-term memory for orienting movements. *Neuron*, 88(2):367–377, Oct 2015. ISSN 0896-6273. doi: 10.1016/j.neuron.2015.08.033. URL https://doi.org/10.1016/j.neuron.2015.08.033.

Yvette Kosmann-Schwarzbach, Bertram E Schwarzbach, and Yvette Kosmann-Schwarzbach. *The Noether theorems*. Springer, 2011.

Christian K. Machens, Ranulfo Romo, and Carlos D. Brody. Flexible control of mutual inhibition: A neural model of two-interval discrimination. *Science*, 307(5712):1121–1124, 2005. doi: 10.1126/science.1104171. URL https://www.science.org/doi/abs/10.1126/science.1104171.

P. C. Martin, E. D. Siggia, and H. A. Rose. Statistical dynamics of classical systems. *Phys. Rev. A*, 8:423–437, Jul 1973. doi: 10.1103/PhysRevA.8.423. URL https://link.aps.org/doi/10.1103/PhysRevA.8.423.

L. Martinez Alonso. On the noether map. *Letters in Mathematical Physics*, 3(5):419–424, Sep 1979. ISSN 1573-0530. doi: 10.1007/BF00397216. URL https://doi.org/10.1007/BF00397216.

Dwight E Neuenschwander. *Emmy Noether's wonderful theorem*. JHU Press, 2017.

E. Noether. Invariante variationsprobleme. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, 1918:235–257, 1918. URL http://eudml.org/doc/59024.

Peter J Olver. *Applications of Lie groups to differential equations*, volume 107. Springer Science & Business Media, 1993.

Peter J Olver. Emmy noether's enduring legacy in symmetry. *Symmetry: Culture and Science*, 29(4):475–485, 2018.

P.J. Olver. Noether's theorems and systems of Cauchy-Kovalevskaya type. In *Nonlinear Systems of Partial Differential Equations in Applied Mathematics*, volume 23 of *Lectures in Applied Mathematics*, pages 81–104. American Mathematical Society, Providence, R.I., 1986.

L. Onsager and S. Machlup. Fluctuations and irreversible processes. *Phys. Rev.*, 91:1505–1512, Sep 1953. doi: 10.1103/PhysRev.91.1505. URL https://link.aps.org/doi/10.1103/PhysRev.91.1505.

Alex T. Piet, Jeffrey C. Erlich, Charles D. Kopec, and Carlos D. Brody. Rat Prefrontal Cortex Inactivations during Decision Making Are Explained by Bistable Attractor Dynamics. *Neural Computation*, 29(11):2861–2886, 11 2017. ISSN 0899-7667. doi: 10.1162/neco_a_01005. URL https://doi.org/10.1162/neco_a_01005.

Roger Ratcliff, Philip L. Smith, Scott D. Brown, and Gail McKoon. Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, 20(4):260–281, Apr 2016. ISSN 1364-6613. doi: 10.1016/j.tics.2016.01.007. URL https://doi.org/10.1016/j.tics.2016.01.007.

Sophia Sanborn, Johan Mathe, Mathilde Papillon, Domas Buracas, Hansen J Lillemark, Christian Shewmake, Abby Bertics, Xavier Pennec, and Nina Miolane. Beyond Euclid: An Illustrated Guide to Modern Machine Learning with Geometric, Topological, and Algebraic Structures. *arXiv e-prints*, art. arXiv:2407.09468, July 2024. doi: 10.48550/arXiv.2407.09468.

Willy Sarlet and Frans Cantrijn. Generalizations of noether's theorem in classical mechanics. *SIAM Review*, 23(4):467–494, 1981. doi: 10.1137/1023098. URL https://doi.org/10.1137/1023098.

Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2256–2265, Lille, France, 07–09 Jul 2015. PMLR. URL https://proceedings.mlr.press/v37/sohl-dickstein15.html.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Tim Strang, Isabella Caruso, and Sam Greydanus. Nature's Cost Function: Simulating Physics by Minimizing the Action. *arXiv e-prints*, art. arXiv:2303.02115, March 2023. doi: 10.48550/arXiv.2303.02115.

Hidenori Tanaka and Daniel Kunin. Noether's learning dynamics: Role of symmetry breaking in neural networks. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 25646–25660. Curran Associates, Inc.,

2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/d76d8deea9c19cc9aaf2237d2bf2f785-Paper.pdf.

Ling Yang, Zhilong Zhang, Yang Song, Shenda Hong, Runsheng Xu, Yue Zhao, Wentao Zhang, Bin Cui, and Ming-Hsuan Yang. Diffusion models: A comprehensive survey of methods and applications. *ACM Comput. Surv.*, 56(4), nov 2023. ISSN 0360-0300. doi: 10.1145/3626235. URL https://doi.org/10.1145/3626235.