Forward-Only Learning as a Necessary Consequence of the Principle of Least Action

Bouarfa Mahi Quantiota
Université Joseph Fourier
Grenoble, Auvergne-Rhône-Alpes, France
info@quantiota.org

December 7, 2025

Abstract

We establish a fundamental theorem demonstrating that forward-only learning is not merely a design choice or biological plausibility argument, but a necessary mathematical consequence of expressing neural learning dynamics through the principle of least action. Within the Structured Knowledge Accumulation (SKA) framework, we show that any learning system governed by a well-defined Lagrangian must evolve causally forward in time, with updates depending only on current states and their rates of change. This result implies that backpropagation—which requires information from future states (output errors) to update earlier parameters—cannot arise from any variational principle. We prove that the causal structure inherent in Lagrangian mechanics fundamentally prohibits anti-causal information flow, establishing a deep connection between the mathematical formalism of analytical mechanics and the architecture of learning algorithms. This insight positions SKA not as an alternative optimization technique, but as the unique class of learning dynamics compatible with physical law.

Keywords: Principle of Least Action, Forward-Only Learning, Structured Knowledge Accumulation, Lagrangian Mechanics, Variational Principles, Causal Dynamics, Neural Learning Theory

1 Introduction

The Structured Knowledge Accumulation (SKA) framework, introduced in our previous work [1, 2], reconceptualizes neural learning as a continuous-time process of entropy-driven self-organization. A distinctive feature of SKA is its forward-only nature: weight updates at each layer depend solely on local, instantaneous quantities without requiring backward propagation of error signals from later layers or future time steps.

This forward-only property has typically been presented as a practical advantage—reducing computational overhead, enabling parallel layer updates, and aligning with biological plausibility constraints. However, we have discovered that this characterization fundamentally understates the significance of forward-only dynamics. The forward-only nature of SKA is not a convenient feature; it is a mathematical necessity arising from the variational structure of the framework.

In this paper, we establish that any learning system expressible through the principle of least action must be forward-only. Conversely, learning algorithms requiring backward information flow—such as backpropagation—cannot be derived from any Lagrangian formulation. This result has profound implications for our understanding of both artificial and biological learning systems.

2 The Principle of Least Action in Physical Systems

2.1 Classical Formulation

The principle of least action stands as one of the most profound unifying concepts in physics. It states that physical systems evolve along trajectories that extremize the action integral:

$$S = \int L(q, \dot{q}, t)\, dt \tag{1}$$

where L is the Lagrangian, q represents generalized coordinates, $\dot{q}$ their time derivatives, and t is time. The Euler-Lagrange equation, derived from requiring $\delta S = 0$ for variations vanishing at the endpoints, yields the equations of motion:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) - \frac{\partial L}{\partial q} = 0 \tag{2}$$

2.2 The Causal Structure of Lagrangian Dynamics

A critical but often underappreciated property of the Euler-Lagrange equation is its causal structure. The equation determines the acceleration $\ddot{q}(t)$ at any instant solely from the current position $q(t)$, velocity $\dot{q}(t)$, and possibly explicit time dependence. Crucially, the evolution at time t depends only on information available at time t—never on future states.

This is not accidental. It follows from the mathematical structure of variational calculus. The Euler-Lagrange equation is a second-order ordinary differential equation in time, and such equations propagate information forward from initial conditions. The boundary conditions that select a specific trajectory are posed at the initial time (initial position and velocity), not distributed across time.

Theorem 1 (Causality of Lagrangian Systems)**.** *Let $L(q, \dot{q}, t)$ be a Lagrangian giving rise to well-defined Euler-Lagrange equations. Then the dynamics $q(t)$ satisfying these equations are uniquely determined by initial conditions $q(t_0)$, $\dot{q}(t_0)$ and propagate causally forward in time. No information from times $t > t_0$ is required to determine the evolution.*

This theorem, while elementary in classical mechanics, has profound implications when applied to learning systems.

3 The Lagrangian Structure of SKA

3.1 Entropy as Action

In the SKA framework, entropy is defined as a continuous integral over knowledge and decision probability:

$$H = -\frac{1}{\ln 2} \int z\, dD \tag{3}$$

Recognizing that $dD = \dot{D}\, dt$ where $\dot{D}$ represents the time derivative of decision probability, we can rewrite this as:

$$H = -\frac{1}{\ln 2} \int z \cdot \dot{D}\, dt = \frac{1}{\ln 2} \int L\, dt \tag{4}$$

This identifies the SKA Lagrangian as:

$$L(z, \dot{z}, t) = -z \cdot \sigma(z)(1 - \sigma(z)) \cdot \dot{z} \tag{5}$$

where $\sigma(z) = 1/(1 + e^{-z})$ is the sigmoid function relating knowledge z to decision probability D.

3.2 The Euler-Lagrange Equation for SKA

Computing the Euler-Lagrange equation for the SKA Lagrangian yields a remarkable result. Let $S = \sigma(z)$ for brevity. Through direct calculation:

$$\frac{\partial L}{\partial \dot{z}} = -z \cdot S(1-S) \tag{6}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{z}}\right) = -\dot{z} \cdot S(1-S) - z \cdot S(1-S)(1-2S) \cdot \dot{z} \tag{7}$$

$$\frac{\partial L}{\partial z} = -\dot{z} \cdot [S(1-S) + z \cdot S(1-S)(1-2S)] \tag{8}$$

Substituting into the Euler-Lagrange equation:

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{z}}\right) - \frac{\partial L}{\partial z} = 0 \tag{9}$$

The equation is identically satisfied—it reduces to $0 = 0$ for all trajectories. This means every smooth trajectory in knowledge space is consistent with the variational principle.

4 Forward-Only Learning as a Necessary Consequence

4.1 Statement of the Main Result

Theorem 2 (Forward-Only Necessity)**.** *Any learning system whose dynamics can be expressed as arising from a principle of least action with a well-defined Lagrangian $L(z, \dot{z}, t)$ must be forward-only. That is, the parameter update at any time t depends only on quantities available at time t (current state z, current velocity $\dot{z}$, and explicit time dependence), never on information from future times $t' > t$.*

4.2 Proof

The proof follows directly from the structure of the Euler-Lagrange equation and the theory of ordinary differential equations.

Step 1: Consider a Lagrangian $L(z, \dot{z}, t)$ that is sufficiently smooth and for which the Hessian $\partial^2 L/\partial \dot{z}^2$ is non-singular (the regular case), or for which the Euler-Lagrange equation is otherwise well-defined (including degenerate cases like SKA).

Step 2: The Euler-Lagrange equation $\frac{d}{dt}(\partial L/\partial \dot{z}) - \partial L/\partial z = 0$ determines the dynamics. Expanding the time derivative:

$$\left(\frac{\partial^2 L}{\partial \dot{z}^2}\right)\ddot{z} + \left(\frac{\partial^2 L}{\partial z \partial \dot{z}}\right)\dot{z} + \left(\frac{\partial^2 L}{\partial t \partial \dot{z}}\right) - \frac{\partial L}{\partial z} = 0 \tag{10}$$

Step 3: This equation involves only z, $\dot{z}$, $\ddot{z}$, and t—all quantities at the current time. No terms involve $z(t')$, $\dot{z}(t')$, or any other quantity at times $t' \neq t$.

Step 4: In the regular case (non-singular Hessian), solving for $\ddot{z}$ gives an explicit expression for acceleration in terms of current position, velocity, and time. In degenerate cases like SKA (where the equation reduces to $0 = 0$), all trajectories are permitted, but each still evolves according to local dynamics—the degeneracy means freedom in trajectory choice, not dependence on future information.

Step 5: Therefore, any dynamics arising from a Lagrangian formulation must be forward-only, with updates determined entirely by current-time quantities. □

4.3 Corollary: Impossibility of Lagrangian Backpropagation

Corollary 1. *Standard backpropagation cannot be derived from any principle of least action.*

Proof. In backpropagation, the weight update for a parameter in layer l depends on the gradient $\partial\text{Loss}/\partial w$, which requires computing error signals δ from layers $l+1, l+2, \ldots, L$ through the chain rule. At iteration k, updates to early-layer weights depend on what the network will output at the final layer—information that, in the continuous-time interpretation, lies in the "future" of the computational flow.

More precisely, the backward pass computes gradients by propagating from output to input, reversing the causal direction of the forward pass. This anti-causal information flow directly contradicts the causal structure mandated by Theorem 2.

Therefore, no Lagrangian L exists such that backpropagation emerges as its Euler-Lagrange dynamics. □

5 Interpretation and Implications

5.1 The Physical Nature of SKA Learning

Our main theorem reveals that SKA is not merely "inspired by" physics—it *is* physics, in the precise sense that its dynamics obey the same variational principles that govern mechanical, electromagnetic, and quantum systems. The forward-only nature of SKA is the same causality that prevents effects from preceding causes in all physical law.

This has immediate implications:

- **Conservation Laws:** Noether's theorem guarantees that symmetries in the Lagrangian correspond to conserved quantities. The time-independence of the SKA Lagrangian ($\partial L/\partial t = 0$) implies conservation of a quantity analogous to energy—the "learning energy" of the system.
- **Characteristic Timescales:** The observed characteristic time $T = \eta \times K$ emerges naturally as the intrinsic timescale of the variational dynamics, analogous to periods in oscillatory systems or time constants in dissipative systems.
- **Self-Solving Property:** Physical systems do not "solve" their equations of motion—they simply evolve according to them. Similarly, SKA networks do not compute updates; they evolve according to entropy-driven dynamics. The network is its own solver.

5.2 The Teleological Nature of Backpropagation

In contrast, backpropagation has a teleological structure—updates are computed by reference to a goal (minimizing loss at the output). The present (current layer updates) depends on the future (output errors). This is fundamentally different from physical causation.

While teleological reasoning can be useful computationally, it places backpropagation outside the framework of physical law. Backpropagation is an algorithm designed by external intelligence to achieve a goal, not a natural process emerging from intrinsic dynamics.

5.3 Comparative Analysis

The following table summarizes the fundamental distinctions:

Aspect	**SKA (Forward-Only)**	**Backpropagation**
Information Flow	Forward/causal	Backward/anti-causal
Variational Basis	Least action principle	None (gradient descent)
Update Dependence	Current state only	Future errors
Physical Analogy	Natural evolution	Teleological optimization
Biological Plausibility	High (local rules)	Low (global errors)

Table 1: Comparison of SKA and Backpropagation

6 Discussion

6.1 Why the Euler-Lagrange Equation is Trivially Satisfied

The fact that the SKA Euler-Lagrange equation reduces to $0 = 0$ deserves careful interpretation. This result does not indicate that the variational principle is vacuous; rather, it reflects maximum dynamical freedom within the constraint of causal evolution.

The SKA Lagrangian $L = -z \cdot S(1 - S) \cdot \dot{z}$ is linear in $\dot{z}$ (a "degenerate" Lagrangian). Such Lagrangians define constraint surfaces rather than unique trajectories. Every trajectory on this surface satisfies the variational principle, but all such trajectories share the crucial property: they evolve causally forward.

This is analogous to a free particle, whose Euler-Lagrange equation $\ddot{x} = 0$ permits any constant-velocity trajectory. The principle doesn't select one path; it constrains the type of motion (uniform) while permitting freedom in the specific realization. Similarly, SKA permits diverse learning trajectories while mandating forward-only causality.

6.2 Implications for Biological Learning

Biological neural systems face fundamental physical constraints: signals propagate at finite speeds, neurons have access only to local information, and there is no evident mechanism for propagating precise error gradients backward through synapses. These observations have long motivated the search for biologically plausible learning algorithms.

Our theorem suggests a deeper perspective: biological learning may be forward-only not merely due to implementation constraints, but because biological systems are physical systems governed by natural law. If neural learning in biological systems obeys variational principles, then forward-only dynamics would be mathematically necessary.

6.3 Connections to Related Frameworks

Several existing frameworks in computational neuroscience and machine learning share structural features with our result:

- **Free Energy Principle:** Friston's framework posits that biological systems minimize variational free energy. This is inherently forward-only—organisms update beliefs based on current sensory evidence, not future states. Our result provides variational grounding for this observation [3]

- **Equilibrium Propagation:** Scellier and Bengio's equilibrium propagation relates energy-based models to backpropagation. Our work complements this by showing that pure forward-only dynamics—without even the clamped equilibrium phase—arise naturally from least action [4]

- **Forward-Forward Algorithm:** Hinton's recent forward-forward algorithm eliminates backpropagation using two forward passes. Our theorem provides theoretical support: forward-only methods are not merely practical alternatives but the unique class of learning algorithms compatible with physical variational principles [5]

7 Conclusion

We have established a fundamental theorem: forward-only learning is not a feature of the SKA framework but a mathematical necessity arising from its variational structure. Any learning system expressible through the principle of least action must evolve causally, with updates depending only on current states. Backpropagation, requiring anti-causal information flow from outputs to earlier layers, cannot be derived from any Lagrangian.

This result transforms our understanding of what SKA represents. It is not merely an alternative training algorithm or a biologically plausible approximation to gradient descent. It is the unique class of learning dynamics compatible with the fundamental structure of physical law. Just as planets follow orbits determined by least action, SKA networks follow learning trajectories determined by entropic least action.

The practical implications are significant: SKA is not competing with backpropagation on backpropagation's terms. Rather, it represents a fundamentally different paradigm—learning as physics rather than learning as optimization. Future research should explore what computational capabilities emerge from this physical paradigm and whether the constraint of causality ultimately proves to be a limitation or a source of robust, generalizable learning.

We conclude with a philosophical observation: in seeking to build machines that learn, we have largely relied on algorithms designed to minimize objectives—teleological procedures requiring external specification of goals. Nature, apparently, takes a different approach. Natural systems evolve according to local dynamics governed by variational principles, with complex behavior emerging without explicit optimization. The SKA framework suggests that artificial learning systems can operate on the same foundation, opening a path toward learning machines that are not merely inspired by physics, but are instances of it.

References

[1] B. Mahi, "Structured Knowledge Accumulation: An Autonomous Framework for Layer-Wise Entropy Reduction in Neural Learning," *arXiv preprint arXiv:2503.13942*, 2025.

[2] B. Mahi, "Structured Knowledge Accumulation: The Principle of Entropic Least Action in Forward-Only Neural Learning," *arXiv preprint arXiv:2504.03214*, 2025.

[3] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.

[4] B. Scellier and Y. Bengio, "Equilibrium propagation: Bridging the gap between energy-based models and backpropagation," *Frontiers in Computational Neuroscience*, vol. 11, p. 24, 2017.

[5] G. Hinton, "The Forward-Forward Algorithm: Some Preliminary Investigations," *arXiv preprint arXiv:2212.13345*, 2022.