The Action Operator Formalism
A Deep Dive into Generalized Reinforcement Learning (GRL)

GRL Framework Analysis

November 23, 2025

Abstract

This document formalizes the concept of the **Action Operator** proposed in the GRL.04822 framework. We transition from the classical Reinforcement Learning (RL) view of an action as a discrete selection index to a GRL view where an action is a parameterized mathematical operator $\hat{O}$ acting on the state space Hilbert space $\mathcal{H}$. We present formal definitions, the connection to the Principle of Least Action, and three distinct working examples: Affine Operators, Flow Operators, and Hamiltonian Operators.

1 The Core Shift: From Index to Operator

In classical RL, an action $a \in \mathcal{A}$ is a symbol. The environment E interprets this symbol to mutate the state s:

$$s_{t+1} \leftarrow E(s_t, a) \tag{1}$$

In Generalized RL (GRL), the agent constructs the transition dynamics itself. The action is an operator $\hat{O}$ drawn from an operator space $\mathfrak{O}$. The state update is the direct application of this operator:

$$s_{t+1} \leftarrow \hat{O}(s_t) + \xi \tag{2}$$

where ξ represents irreducible environmental noise. The agent effectively learns "local physics" to navigate the state space.

2 Formal Definitions

Definition 1 (The Operator Space $\mathfrak{O}$). *Let $\mathcal{S}$ be the state space, assumed to be a manifold or vector space. The Operator Space $\mathfrak{O}$ is a family of mappings:*

$$\mathfrak{O} = \{\hat{O} : \mathcal{S} \to \mathcal{S}\}$$

where each $\hat{O}$ is a smooth, potentially non-linear transformation.

Definition 2 (Parametric Action Manifold Θ). *Since $\mathfrak{O}$ is infinite-dimensional, we restrict the agent to a parametric submanifold. Let $\Theta \subseteq \mathbb{R}^d$ be the parameter space. We define a generator function Φ:*

$$\Phi : \Theta \times \mathcal{S} \to \mathcal{S}$$

An action is the selection of a parameter vector $\theta \in \Theta$. The resulting operator $\hat{O}_\theta$ is defined as:

$$\hat{O}_\theta(s) \triangleq \Phi(\theta, s)$$

3 The Principle of Least Action in GRL

The inspiration from physics suggests that state transitions should not be arbitrary. We introduce an **Action Functional** $\mathcal{J}$ that penalizes the "energy" of the operator.

Principle 1 (Least Action GRL). *The optimal policy π^* maximizes the expected cumulative reward while minimizing the complexity (or energy) of the action operators applied.*

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \Big(\underbrace{R(s_t)}_{Goal} - \lambda \underbrace{||\hat{O}_{\theta_t} - \mathbf{I}||}_{Least\ Action}\Big)\right] \tag{3}$$

where $\mathbf{I}$ is the identity operator. The term $||\hat{O} - \mathbf{I}||$ measures how much the operator distorts the state space.

4 Working Examples

4.1 Example 1: The Affine Operator (Linear Dynamics)

Scenario: Robotic control or navigation in continuous space.

- **State Space:** $s \in \mathbb{R}^n$.

- **Action Parameters:** $\theta = \{\mathbf{W}, \mathbf{b}\}$, where $\mathbf{W} \in \mathbb{R}^{n \times n}$ and $\mathbf{b} \in \mathbb{R}^n$.

- **The Operator:**
$$\hat{O}_{\theta}(s) = \mathbf{W}s + \mathbf{b} \tag{4}$$

- **Least Action Cost:** We want the transformation to be minimal (close to Identity matrix and zero bias).
$$\mathcal{L}_{energy}(\theta) = ||\mathbf{W} - \mathbf{I}_n||_F^2 + ||\mathbf{b}||_2^2 \tag{5}$$

Interpretation: The agent outputs a matrix $\mathbf{W}$ that stretches/rotates the state space and a vector $\mathbf{b}$ that translates it.

4.2 Example 2: The Flow Operator (Vector Fields)

Scenario: Controlling a fluid system or a high-dimensional trajectory where instantaneous velocity matters more than position steps.

- **State Space:** $s \in \mathbb{R}^n$.

- **Action Parameters:** θ are the weights of a hyper-network or coefficients of a polynomial vector field.

- **The Operator:** The operator evolves the state by integrating a differential equation defined by θ for a time step Δt.

$$\dot{s}(t) = f(s(t); \theta) \tag{6}$$

$$\hat{O}_{\theta}(s_t) = s_t + \int_t^{t+\Delta t} f(s(\tau); \theta) d\tau \tag{7}$$

- **Least Action Cost:** Minimize the kinetic energy of the flow.

$$\mathcal{L}_{energy}(\theta) = \frac{1}{2}||f(s_t; \theta)||^2 \tag{8}$$

4.3 Example 3: The Hamiltonian Operator (Physics Embedding)

Scenario: Controlling physical systems where conservation of energy is desirable.

- **State Space:** $s = (\mathbf{q}, \mathbf{p})$ (Position and Momentum).

- **Action Parameters:** θ parameterizes a scalar Hamiltonian function $H_\theta(\mathbf{q}, \mathbf{p})$.

- **The Operator:** The operator evolves the state according to Hamilton's equations:

$$\frac{d\mathbf{q}}{dt} = \frac{\partial H_\theta}{\partial \mathbf{p}}, \quad \frac{d\mathbf{p}}{dt} = -\frac{\partial H_\theta}{\partial \mathbf{q}} \tag{9}$$

- **Least Action Cost:**

$$\mathcal{L}_{energy} = |H_\theta(\mathbf{q}, \mathbf{p}) - H_{ref}| \tag{10}$$

  Here, the agent doesn't choose "force"; it chooses the *energy landscape H_θ* that naturally guides the system to the target.

5 The Modified Bellman Equation

For a GRL agent, the Bellman equation must integrate over the parameter manifold Θ rather than a discrete set of actions.

Let $Q(s, \theta)$ be the value of applying operator parameterized by θ to state s.

$$Q(s, \theta) = \mathcal{R}(s, \hat{O}_\theta(s)) - \lambda \mathcal{E}(\theta) + \gamma \int_\mathcal{S} P(s'|\hat{O}_\theta(s))V(s')ds' \tag{11}$$

Where:

- $\mathcal{R}$ is the extrinsic reward.

- $\mathcal{E}(\theta)$ is the intrinsic "operator cost" (Least Action penalty).

- P is the environmental noise kernel.

6 Conclusion

By formalizing actions as operators, GRL allows agents to learn the *laws of motion* required to solve a task. This is a higher-order learning process compared to standard RL. The inclusion of the Least Action Principle ($\lambda \mathcal{E}$) serves as a crucial regularizer, ensuring that among the infinite operators that could reach a goal, the agent selects the most efficient and physically plausible one.