Center for Research in Computer Vision
UCF | UNIVERSITY OF CENTRAL FLORIDA

Training Dynamics and Decision Stability in Multi-Objective Reinforcement Learning via Spectral Sensitivity Analysis

David Tang, Mentor: Dr. Yue Wang
University of Central Florida, Orlando, FL

Introduction

Problem and Background

- Standard Multi-Objective Reinforcement Learning assumes smooth, convex Pareto frontiers. In practice, discovered empirical frontiers are noisy, non-convex, and discrete.
- Existing selection rules force a choice between either efficiency (Nash Social Welfare) or guaranteeing the worst-case floor (Chebyshev).
- Standard evaluation techniques apply arbitrary additive noise or random point removal, which break fundamental Pareto properties such as weak monotonicity or boundary conditions.

Contributions

- **Hybrid Decision Rule:** Interpolates continuously between Nash efficiency and Chebyshev robustness.
- **Sensitivity Pipeline:** Uses a Type-II Discrete Cosine Transform to systematically measure decision stability.

Preliminaries

Standard deep reinforcement learning objectives require a scalar optimization signal to perform gradient updates. Therefore, vector-valued returns must be mapped to a scalar target using a scalarization function f.

For vector $\mathbf{u} \in \mathbb{R}^K$:

Linear: $f_{\text{linear}}(\boldsymbol{u}) = \sum_{k=1}^{K} w_k u_k$

Nash Social Welfare: $f_{\text{Nash}}(\boldsymbol{u}) = \prod_{k=1}^{K} u_k$

Chebyshev: $f_{\text{Chebyshev}}(\boldsymbol{u}) = -\max_{k\in\{1,\dots,K\}} [w_k (z_k^* - u_k)]$

where $\sum_{k=1}^{K} w_k = 1$

As multiple objectives frequently conflict, optimizing every objective simultaneously is often impossible. Instead, evaluation is done on the Pareto frontier. We consider a two objective setting.

N choices sorted $[x_n, y_n]$ sorted by ascending x

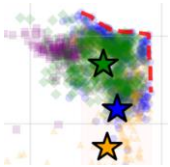
$$\mathbf{y} = [y_0, y_1, \dots, y_{N-1}]^T \in \mathbb{R}^N$$

subject to $y_n \geq y_{n+1}$

Method

$$f_{\text{hybrid}}(\mathbf{u};\alpha) = \alpha f_{\text{Nash}}(\mathbf{u}) + (1-\alpha) f_{\text{Chebyshev}}(\mathbf{u}), \quad \alpha \in (0,1).$$

Raw Pareto Frontier



Discovered Pareto Frontier of mo-lunar-lander-v3-continuous Shaping (y) vs fuel (x)

1. Uniform Grid Interpolation and [0, 1] Normalization

$$c_k = \beta_k \sum_{n=0}^{N-1} y_n \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right] \quad k = 0,\dots,N-1,$$

2. Forward DCT-II: Spectral Truncation at Energy cutoff E(k)

$$E(k) = \frac{\sum_{i=0}^{k-1} c_i^2}{\sum_{i=0}^{N-1} c_i^2}.$$

$$\mathbf{c}_{\text{filt}} = [c_0, c_1, \dots, c_{k-1}, 0, \dots, 0]^T$$

3. Inverse DCT-II (DCT-III): Reconstruct Simplified Frontier

$$\hat{y}_{\text{recon},n} = \sum_{i=0}^{k-1} \beta_i c_i \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)i\right], \quad n = 0,\dots,N-1.$$

4. Ensure Frontier Validity: Enforce Monotonicity and Boundary Conditions

$$\tilde{y}_{N-1} = y_{N-1},$$
$$\tilde{y}_n = \max\left(\hat{y}_{\text{recon},n}, \tilde{y}_{n+1}\right) \quad \forall n \in \{N-2,\dots,0\},$$
$$\hat{y}_{E,n} = y_{N-1} + (y_0 - y_{N-1}) \cdot \left(\frac{\tilde{y}_n - \tilde{y}_{N-1}}{\tilde{y}_0 - \tilde{y}_{N-1}}\right).$$

5. Evaluate Stability: Compute Sensitivty S(E) and Integrated Score A

$$S(E) = \|\mathbf{v}^* - \mathbf{v}_E^*\|_2$$

$$A = \int_0^1 S(E)\, dE$$

Experiment

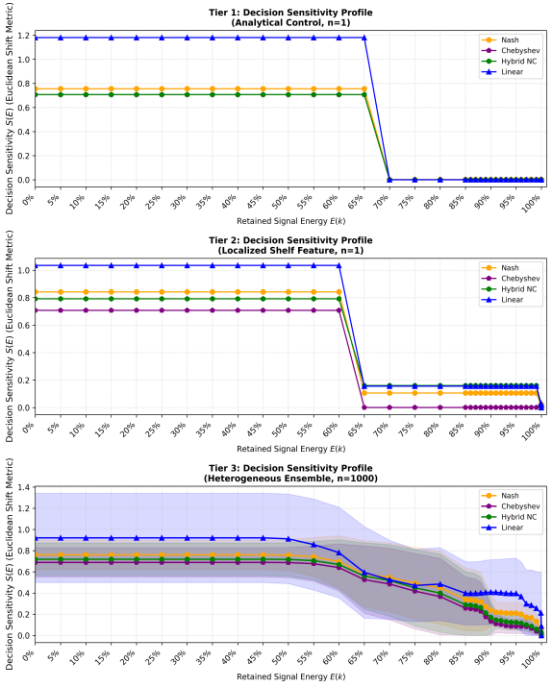


Table 1: Integrated Sensitivity Score Summary

Decision Metric	Tier 1 ($n = 1$) Mean (± Std)	Tier 2 ($n = 1$) Mean (± Std)	Tier 3 Ensemble ($n = 1000$) Mean (± Std)
MaxMin	0.4774 (±0.0000)	0.4427 (±0.0000)	0.5430 (±0.1549)
Chebyshev	0.4774 (±0.0000)	0.4427 (±0.0000)	0.5430 (±0.1549)
Hybrid NMM	0.4774 (±0.0000)	0.4675 (±0.0000)	0.5598 (±0.1716)
Hybrid NC	0.4774 (±0.0000)	0.5543 (±0.0000)	0.5736 (±0.1696)
Nash	0.5107 (±0.0000)	0.5657 (±0.0000)	0.6175 (±0.1535)
Linear	0.7972 (±0.0000)	0.7047 (±0.0000)	0.7336 (±0.1867)

Intuition: Less Area means more stable

Conclusion

We propose a hybrid decision rule that combines the strengths of Nash and Chebyshev to improve stability while preserving desirable decision properties. We also introduce a spectral sensitivity pipeline that provides a principled framework for evaluating the stability of Pareto decision rules.