SKKU

Knothe-Rosenblatt Quantile Regression for Risk-sensitive Multi-objective Reinforcement Learning

Gwangpyo Yoo, Wookyung Kim, and Honguk Woo

Introduction: MORL and Risk-sensitive MORL

SKKU

- Real-world domains (e.g., robotics, finance) require agents to balance conflicting objectives while mitigating risks.
- Evaluating risk in MORL necessitates multivariate quantile formulations.
- Existing approaches' limitations: marginal methods ignore objective correlations, and MMD-based methods fail to guarantee a proper transport map.
- We introduce the theoretical and empirical results in risk-sensitive MORL framework that aligns with vector-risk measures and guarantees critic convergence.

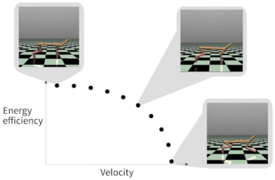

Figure 1: Locomotion Example: We may prioritize the robot's energy efficiency.

Image source: MO-Gymnasium (Felten et al., 2023)

Introduction: MORL and Risk-sensitive MORL

SKKU

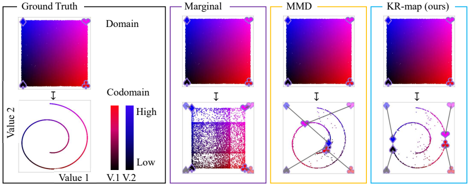

Figure 2: Comparison of marginal, MMD, and KR map approaches.

- Computing vector-risk requires a valid **multivariate quantile formulation**.
- **Marginal-based:** Fail to capture joint distributions and correlations among objectives.
- **MMD-based:** Do not guarantee an order-preserving property (not transport map).
- **Our Solution:** The Knothe-Rosenblatt (KR) map provides a valid quantile formulation as an autoregressive map.

Problem Formulation

SKKU

Definition (KR map). For a d-dim random vector $\mathbf{Z}$, the Knothe-Rosenblatt map $\mathbf{Q_Z} : [0,1]^d \to \mathbb{R}^d$ is defined as:

$$[\mathbf{Q_Z}(\mathbf{u})]_1 = F^{-1}_{\mathbf{Z}_1}(\mathbf{u}_1),$$
$$[\mathbf{Q_Z}(\mathbf{u})]_k = F^{-1}_{\mathbf{Z}_k}\left(\mathbf{u}_k; [\mathbf{Q_Z}(\mathbf{u})]_{<k}\right) \text{ for } k = 2, \ldots, d. \tag{1}$$

Theorem $\mathbf{Q_Z}(\mathbf{u}) \sim \mathbf{Z}$ for $\mathbf{u} \sim U[0,1]^d$.

RL-objective Given univariate risk measures $\{\rho(\cdot;\phi_i)\}_{i=1}^d$, we define $\Phi^{-1} : [0,1]^d \to [0,1]^d$ as:

$$\Phi^{-1}(\mathbf{u}) = [\Phi_1^{-1}(\mathbf{u}_1), \ldots, \Phi_d^{-1}(\mathbf{u}_d)], \tag{2}$$

where Φ_i^{-1} is the quantile of ϕ_i. Letting $\mathbf{Q}_{\mathbf{Z}^\pi}$ be the KR map for $\mathbf{Z}^\pi$, our objective for preference ω is:

$$\pi^*_\omega = \arg\min_\pi -\omega^T \mathbb{E}_{\mathbf{u}\sim U[0,1]^d}\left[\mathbf{Q}_{\mathbf{Z}^\pi} \circ \Phi^{-1}(\mathbf{u})\right] \tag{3}$$

Theoretical & Empirical Approach

Dynamic Vector-Risk Measure

The risk formulation using the KR map perfectly satisfies the rigorous axioms (Zero, Monotonicity, Translation Invariance) of a dynamic vector-risk measure:

$$\rho_\Phi(\mathbf{X}) := -\mathbb{E}_{\mathbf{u}\sim U[0,1]^d}[\mathbf{Q}_\mathbf{X} \circ \Phi^{-1}(\mathbf{u})] = \rho_d$$

Convergence of KR-Quantile Regression

Under the distributional MO-Bellman operator $\mathscr{T}^\pi$, we guarantee contraction in the adapted Wasserstein metric ensuring the critic converges:

$$\bar{d}_\infty(\Pi_{\text{KR}}\mathscr{T}^\pi\mathbf{Z},\, \Pi_{\text{KR}}\mathscr{T}^\pi\mathbf{Z}') \;\le\; \gamma\bar{d}_\infty(\mathbf{Z},\mathbf{Z}')$$

KR-IQN Architecture

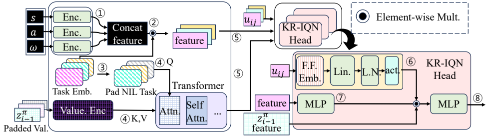

Figure 3: Overall Implementation

- Task Embeddings guide the objective inference (padded with NIL task initially).
- Transformer without PE prevents overfitting to the KR map's artificial ordering.
- KR-IQN Head computes the final risk using context, base features, and quantile u_{ij}.

Experiments

SKKU

Table 1: Main results.

Environment	Hopper($d=3, i=2$)				Cheetah ($i=2$)			
Algorithm	EU	EU_{Risk}	HV($\times 10^8$)	HV_{Risk}($\times 10^8$)	EU	EU_{Risk}	HV($\times 10^6$)	HV_{Risk}($\times 10^6$)
DPMORL	988.18$\pm$56.91	680.74$\pm$102.91	6.59$\pm$14.24	3.49$\pm$1.67	558.89$\pm$32.38	67.44$\pm$123.78	4.05$\pm$0.32	4.31$\pm$0.27
EWP	429.75$\pm$0.56	376.39$\pm$48.82	1.18$\pm$0.01	1.18$\pm$0.01	1578.34$\pm$147.46	1575.96$\pm$126.87	9.59$\pm$0.80	9.49$\pm$0.70
MarginalIQN (N)	909.52$\pm$140.61	384.95$\pm$170.54	7.73$\pm$2.47	6.67$\pm$3.18	2097.21$\pm$96.63	2113.80$\pm$179.73	11.48$\pm$0.36	11.44$\pm$0.75
MarginalIQN (RS)	702.57$\pm$253.62	181.46$\pm$84.92	4.59$\pm$3.39	3.10$\pm$2.74	1797.59$\pm$394.39	1933.87$\pm$164.57	10.10$\pm$1.96	10.61$\pm$1.05
KR-IQN(N)	1247.39$\pm$9.55	934.89$\pm$232.81	15.75$\pm$0.29	15.33$\pm$0.51	**2172.71**$\pm$131.14	**2153.02**$\pm$139.46	**12.10**$\pm$0.61	**11.82**$\pm$0.60
KR-IQN(RS)	**1253.93**$\pm$11.36	**1086.62**$\pm$106.02	**16.28**$\pm$0.47	**15.96**$\pm$0.60	2153.88$\pm$228.74	2140.37$\pm$216.99	12.02$\pm$1.01	11.81$\pm$0.99
Environment	Ant ($i=3$)				Finance* ($i=3$)			
Algorithm	EU	EU_{Risk}	HV($\times 10^8$)	HV_{Risk}($\times 10^8$)	EU	EU_{Risk}	HV($\times 10^6$)	HV_{Risk}($\times 10^6$)
EWP	783.20$\pm$161.83	320.84$\pm$93.87	9.70$\pm$4.52	8.80$\pm$3.45	12.67$\pm$12.38	2.18$\pm$8.99	1.06$\pm$1.18	1.07$\pm$1.26
MarginalIQN (N)	1399.39$\pm$308.63	1429.64$\pm$404.34	34.86$\pm$14.21	35.47$\pm$12.69	**51.44**$\pm$6.60	30.16$\pm$3.69	11.59$\pm$0.80	12.93$\pm$0.36
MarginalIQN (RS)	912.61$\pm$77.73	948.00$\pm$101.71	12.44$\pm$2.77	16.95$\pm$4.21	50.41$\pm$7.67	29.94$\pm$4.12	11.68$\pm$0.55	12.98$\pm$0.49
KR-IQN(N)	1513.96$\pm$194.01	1459.33$\pm$166.78	41.81$\pm$12.41	39.14$\pm$7.79	50.55$\pm$6.87	**32.05**$\pm$6.59	11.49$\pm$0.46	13.12$\pm$0.23
KR-IQN(RS)	**1596.25**$\pm$101.15	**1645.08**$\pm$133.61	**45.87**$\pm$7.29	**45.69**$\pm$6.15	46.36$\pm$3.92	27.84$\pm$5.86	11.87$\pm$0.42	12.99$\pm$0.37

SKKU

Thank you for Listening