Geometry-Aware, Adaptive Risk for Agents:
A Coherent-Risk and Wasserstein-Robust Foundation for Learning under Interaction and Evolving Uncertainty

Deep Ganguly

Technical University of Munich

deep.ganguly@tum.de

An agent acting on a user's behalf rarely operates in a fixed environment. Its own decisions reshape the data it will later see, the other agents it interacts with adapt in response, and the uncertainty it must manage changes over time. How well such an agent behaves depends on how it reasons about risk, and we argue that three properties are jointly necessary. Risk should be *coherent*, so that catastrophic tails rather than averages drive behavior; it should be *geometry-aware*, so that model error is measured by how far probability mass actually moves across the state and observation space; and it should be *adaptive*, so that an agent's caution tracks uncertainty as evidence accumulates. We sketch a foundation built on these properties and the questions it raises for learning theory.

From entropic to geometric robustness. A coherent risk measure can always be written as a worst-case expectation over an ambiguity set, and the divergence defining that set fixes both the measure and the geometry it respects. The Entropic Value-at-Risk (EVaR_α), the tightest coherent upper bound on CVaR and VaR, corresponds to a Kullback-Leibler ball of radius $-\ln\alpha$. It is analytically convenient but blind to geometry: KL is infinite as soon as the support shifts and ignores how far mass travels. A Wasserstein ball replaces it with an ambiguity set that charges each perturbation by its transport cost along the ground metric, which is how a misspecified transition kernel actually redistributes probability toward neighboring states. Entropic optimal transport (Sinkhorn regularization) interpolates between KL-based and Wasserstein-based robustness, placing the entropic structure of EVaR and the geometric sensitivity of optimal transport inside one ambiguity model.

Adaptive risk and convergent algorithms. Rather than fixing the risk level and ambiguity radius in advance, we let the agent adjust them online, tightening as evidence accumulates and relaxing when it detects drift, so that risk becomes a learned quantity rather than a tuning constant. Optimizing a risk of this form couples an inner problem over the risk parameter with an outer policy update. We handle it with multi-timescale stochastic approximation that separates the two recursions and establishes asymptotic convergence to locally optimal risk-sensitive policies via the ODE method [1]. Carried out over an augmented state, the analysis continues to hold under partial observability, when the agent acts on belief states rather than the latent state of the environment.

When robustness backfires: the price of paranoia. Interaction changes the picture. When agents learn together, each gradient step an agent takes shifts the distribution of actions its partner will play, turning a cooperating partner into a source of noise exactly where the decision to cooperate is most delicate. We show that a Pareto-dominant cooperative equilibrium can be exponentially unstable under risk-neutral learning, unravelling once this co-learning noise crosses a critical threshold; and, more pointedly, that the instinctive fix of hedging against partner uncertainty with return-level distributional robustness makes things strictly worse, since a risk-averse return objective penalizes the high-variance cooperative action relative to defection and widens the unstable region [2]. The lesson is to be robust to the right quantity: controlling the variance of the co-learning gradient rather than of the return yields a lightweight trust mechanism, Robust Adaptive Trust Region Learning, that sustains cooperation with no communication.

From statistical to formal guarantees. Convergence does not guarantee that a deployed policy will behave, and that gap calls for certificates. We verify learned policies through probabilistic barrier-certificate conditions and hot-started, risk-aware value iteration that certify tail-risk specifications against the worst case in the chosen ambiguity set, KL or Wasserstein. Carrying these certificates to neural policies, by co-designing the objective with the certificate it must satisfy, is the main obstacle to scale.

Open problems. Several questions seem central and within reach: the sample complexity of estimating geometry-aware coherent risk, and when Wasserstein-robust risk-sensitive RL is efficiently solvable; how to define and bound regret when the risk level itself evolves and when feedback is strategic or performative, which the transport view recasts as decision-induced movement of mass within a Wasserstein ball; the optimal-transport geometry of equilibria among interacting risk-sensitive learners; and whether calibrated, geometry-aware risk envelopes can discipline the overconfidence of language-model agents. We would welcome discussion of where learning theory can most sharpen these foundations.

References.

[1] D. K. Ganguly, A. G. Joseph, S. Girotra, and S. Sekhar. Risk-Seeking Reinforcement Learning via Multi-Timescale Entropic Value at Risk Optimization. TMLR, 2025.

[2] D. Ganguly, C. Jonnalagadda, Pratham C., and A. Ananth. The Price of Paranoia: Robust Risk-Sensitive Cooperation in Non-Stationary Multi-Agent Reinforcement Learning. ALA Workshop, AAMAS 2026. arXiv:2604.15695.