A Family Reunion of Three Distant Cousins: Risk, Regularization, and Robustness in Probabilistic Optimal Control

Ajinkya Bhole, Mohammad Mahmoudi Filabadi, Guillaume Crevecoeur and Tom Lefebvre

***Abstract*— The concept of probabilistic policies has deeply permeated optimal control, serving as formal beliefs about the true optimal solution that capture informational uncertainty and guide principled exploration. Consequently, a rich mosaic of distinct modeling paradigms has emerged, including Control as Inference, Risk-Sensitive Control, Active Inference, and Distributionally Robust Control, among many others. While these approaches address deeply related challenges, their differing mathematical foundations and terminology often obscure their underlying connections. The main contribution of this short paper is the introduction of a central control formulation. By toggling specific structural decisions within this formulation, namely activating or deactivating policy regularization and auxiliary transition kernel optimization, one can systematically recover and reason about these seemingly disparate paradigms. We outline a map of these connections and detail its structural properties, offering a consolidated view of the probabilistic optimal control landscape that can be explicitly exploited for algorithm design.**

I. Introduction

Over the past two decades, the growing interplay between control engineering and machine learning has significantly elevated the role of probabilistic policies in control design. Consequently, the field of optimal control has diversified into several distinct modeling paradigms, each employing its own conceptual framework. *Density Matching* approaches frame decision-making as minimizing the divergence between predictive behaviors and target distributions [1], while the closely related framework of *Control as Inference* recasts this problem as Bayesian inference on probabilistic graphical models conditioned on desired goals or optimal outcomes [2]. *Risk-Sensitive Control* evaluates the underlying cost distribution beyond its expected mean [3]. *Active Inference*, rooted in the Free Energy Principle, unifies perception, learning, and control into a single imperative of minimizing variational free energy [4]. Meanwhile, *Distributionally Robust Control* (DRC) uses ambiguity sets to hedge against model mismatch [5], [6].

To formalize the underlying links between these domains, the main contribution of this short paper is the introduction of a central control formulation that serves as a unifying starting point. By explicitly toggling specific structural decisions within this formulation, one can systematically trace how distinct modeling assumptions naturally translate into these seemingly disparate paradigms. Furthermore, this central perspective clarifies how entropy regularization acts across these frameworks as a fundamental mathematical tool to bound information processing and tame computational complexity. Ultimately, establishing these connections bring the three distant cousins: risk, regularization, and robustness to the same table.

II. The Central Problem as a Starting Point

The consolidation is anchored in a central control problem. By explicitly optimizing over both the control policy and an auxiliary system transition kernel, one can formally distinguish between the agent's informational uncertainty regarding action selection and its anticipation of deviations in the system dynamics. The resulting central problem is given as follows:

$$\min_{\underline{\pi}} \operatorname*{opt}^{\lambda^S}_{\underline{\tau}} \Big\{ \mathbb{E}_{p_{(\underline{\pi},\underline{\tau})}}[\underline{c}_T] + \frac{1}{\lambda^P}\mathbb{D}^{\underline{\pi}}_{\underline{\rho}} + \frac{1}{\lambda^S}\mathbb{D}^{\underline{\tau}}_{\underline{\iota}} \Big\} \tag{1}$$

Here, $\underline{\pi}$ is the policy sequence, $\underline{\tau}$ is an auxiliary transition kernel sequence, and $p_{(\underline{\pi},\underline{\tau})}$ is the resulting trajectory distribution. The term $\underline{c}_T$ represents the cumulative cost. The baseline policy $\underline{\rho}$ and nominal dynamics $\underline{\iota}$ provide reference behaviors. The parameters $\lambda^P > 0$ and $\lambda^S \in \mathbb{R}\backslash\{0\}$ weight the Kullback-Leibler ($\mathbb{D}$) divergences. The operator opt toggles between $\max$ (when $\lambda^S < 0$, leading to risk-averse behavior) and $\min$ (when $\lambda^S > 0$, leading to risk-seeking behavior).

This central problem acts as a structural template. By toggling two primary decisions: keeping policy regularization ON/OFF and keeping the auxiliary transition optimization ON/OFF, one can traverse different paradigms (see Fig. 1).

A. Control as Inference and Active Inference

Active Inference (AIF) and Control as Inference (CaI) both fundamentally cast decision-making as a distribution matching problem; however, the mechanism by which value is encoded to attain useful behavior differs significantly across these frameworks [7].

To see their connection to the central problem (1), consider switching OFF the auxiliary transition optimization $\underline{\tau}$. This yields the Soft-Policy Stochastic Optimal Control (SP-SOC) objective (see Fig. 1), which after rearranging reveals equivalence to a Forward KL (Information projection) matching problem $\min_{\underline{\pi}} \mathbb{D}^{p_{(\underline{\pi},\underline{\iota})}}_{p^*}$, where the target distribution is defined as $p^* \propto p_{(\underline{\rho},\underline{\iota})} e^{-\lambda^P \underline{c}_T}$.

AIF also operates within this Forward KL paradigm and, in its more general form, jointly optimizes the policy alongside the agent's internal dynamics and sensor models. It constructs the target distribution by augmenting a generative model with synthetic prior factors encoding goals and epistemic drives [8]. One of the ways the connection to SP-SOC can be recovered is by (i) locking the agent's internal transition and sensor beliefs directly to the true generative models ($\underline{\tau} \triangleq \underline{\iota}$, $\underline{\omega} \triangleq \underline{o}$), which eliminates model complexity and sensory epistemic terms from the KL divergence, and (ii) defining a preference prior jointly over state-action pairs $\hat{p}(x_t, u_t) \propto e^{-\lambda^P c_t(x_t,u_t)}$. Under these assumptions, the AIF objective reduces exactly to SP-SOC.

CaI also operates within this paradigm, but rather than using synthetic priors, it establishes matching by introducing binary optimality variables $\mathcal{O}_t$ to encode value, with $p(\mathcal{O}_t = 1 \mid x_t, u_t) \propto e^{-\lambda^P c_t(x_t,u_t)}$. Consequently, its target distribution corresponds to the baseline trajectory conditioned on universal optimality. Like AIF, deriving SP-SOC from the general CaI inference problem follows from locking the internal transition and sensor beliefs directly to the true system models ($\underline{\tau} \triangleq \underline{\iota}$, $\underline{\omega} \triangleq \underline{o}$).

In the CaI framework, a problem can also be formulated in a fundamentally different way. Rather than matching a predictive model to a target distribution, one can treat the policy sequence $\underline{\pi}$ as unknown parameters of the generative model and seek to maximize the marginal likelihood of observing optimality: $\max_{\underline{\pi}} \log p_{\underline{\pi}}(\underline{\mathcal{O}}_T = \mathbf{1})$, written differently is the *Risk-Sensitive Optimal Control* (RSOC) formulation: $\min_{\underline{\pi}} -\frac{1}{\lambda^S} \log \mathbb{E}_{p_{(\underline{\pi},\underline{\iota})}}[e^{(-\lambda^S \underline{c}_T)}]$, which can be recovered from the central problem (1) by switching policy regularization OFF. A computationally tractable way to solve this problem

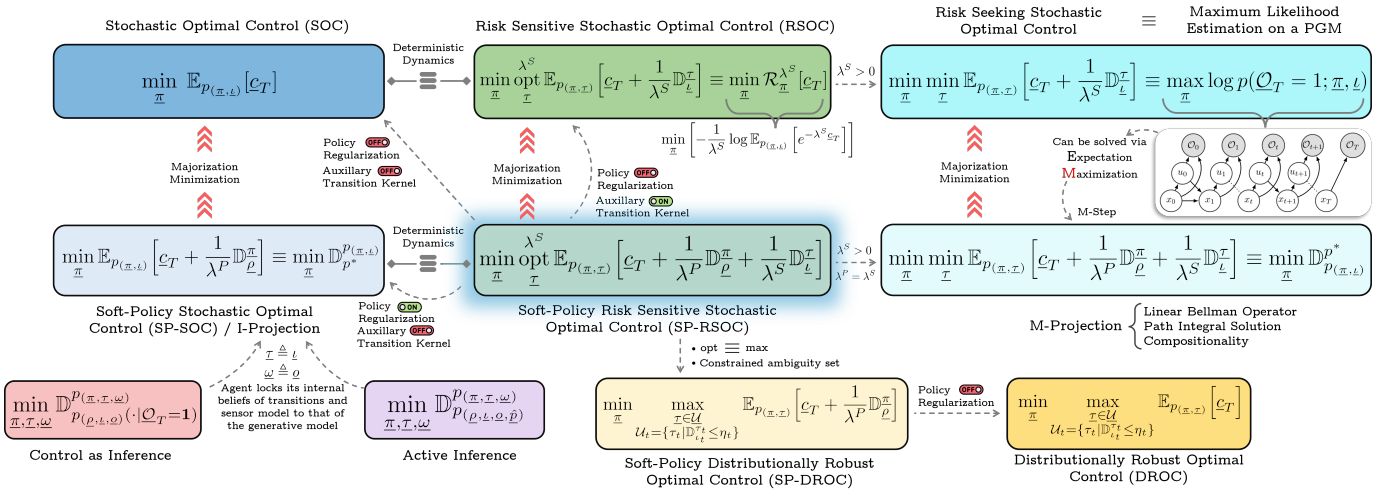

Fig. 1. A visual map of the probabilistic optimal control landscape, illustrating the structural toggles that connect distinct problem formulations.

is using the Expectation-Maximization (EM) algorithm. The E-step infers a surrogate distribution ($p^* \propto p_{(\underline{\rho},\underline{\iota})} e^{-\lambda^S \underline{c}_T}$), and the M-step then updates the policy to match this surrogate, which corresponds to a Reverse KL (Moment projection) formulation: $\min_{\underline{\pi}} \mathbb{D}^{p^*}_{p_{(\underline{\pi},\underline{\iota})}}$ [9]. This formulation can be recovered in the central problem (1) by synchronizing the regularization weights ($\lambda^P = \lambda^S > 0$).

B. Risk-Sensitive Control

By retaining the opt operator over the auxiliary kernel $\underline{\tau}$ in (1), the problem inherently evaluates the underlying cost distribution beyond its mean. The choice of operator toggles the risk attitude, admitting a game-theoretic interpretation where $\text{opt} \equiv \max$ models a demonic nature, evaluating worst-case transitions and yielding risk-averse behavior, while $\text{opt} \equiv \min$ models an angelic nature, yielding an optimistic, risk-seeking policy.

C. Distributionally Robust Control

To map from Risk-Sensitive Control to Distributionally Robust Control, the toggles shift from the penalty space to the constraint space via Lagrangian duality.

$$\min_{\underline{\pi}} \max_{\substack{\underline{\tau}\in\underline{\mathcal{U}} \\ \mathcal{U}_t=\{\tau_t \mid \mathbb{D}^{\tau_t}_{\iota_t} \le \eta_t\}}} \mathbb{E}_{p_{(\underline{\pi},\underline{\tau})}}\left[\underline{c}_T + \frac{1}{\lambda^P}\mathbb{D}^{\underline{\pi}}_{\underline{\rho}}\right]$$

The scalar penalty $\frac{1}{\lambda^S}$ on the transition divergence morphs into a hard ambiguity budget ($\mathbb{D}^{\tau_t}_{\iota_t} \le \eta_t$). While mathematically linked, these formulations exhibit notable operational trade-offs based on how ambiguity is bounded. The regularized formulation utilizes a fixed price of risk, leading to a *dynamic* ambiguity set. This preserves analytical tractability but can possibly leave the agent vulnerable to the *overconfidence trap* in environments where the agent's nominal model $\underline{\iota}$ is a poor representation of reality. Conversely, the constrained DRC formulation utilizes a fixed ambiguity budget. This guarantees strict safety bounds but can possibly leave the agent vulnerable to the *paranoia trap*, bracing for physically impossible worst-case scenarios, even in benign regions.

III. Structural Properties and Features

Formalizing the map (Fig. 1) uncovers properties that provide a consolidated compass for algorithm design.

Majorization-Minimization (MM) Iterations: The soft-policy formulations natively majorize the original crisp optimal control objectives. By recursively feeding the optimal soft policy back in as the new baseline prior, gradually toggling policy regularization from ON to OFF asymptotically, the policy regularization is annealed. This allows an algorithm to tractably optimize a regularized surrogate while systematically converging to the classical, policy-unregularized optimum.

Deterministic Collapse: The map also reveals the geometric boundaries of different paradigms. When the nominal dynamics $\underline{\iota}$ are purely deterministic, toggling the transition optimization ON loses its expressive freedom because the transition kernel is locked to the deterministic Dirac delta. Consequently, the theoretical boundaries dissolve, yielding SOC ≡ RSOC and SP-SOC ≡ SP-RSOC.

Features of M-projection problem: The M-projection problem exhibits uniquely powerful features. The nonlinear Bellman equations collapse into a linear form via desirability functions, admitting a path integral solution and enabling compositionality for modular control design.

IV. Conclusion and Future Frontiers

By consolidating disparate paradigms into a single mathematical topology, researchers are provided with a unified view of the probabilistic control landscape. Rather than operating in isolated silos, practitioners can explicitly trace the connections between these formulations, exploiting properties like MM-iterations or path integral solutions depending on the desired behavioral outcomes or computational constraints.

The mapping presented here represents only the tip of the iceberg. It relies exclusively on Shannon entropy and the Kullback-Leibler divergence. As the field is maturing, a critical future direction will be expanding this map to incorporate other geometries. Investigating how non-additive measures or alternative ambiguity sets alter this central formulation will be essential for mapping the full landscape of robust autonomous systems.

References

[1] M. Kárný, "Towards fully probabilistic control design," *Automatica*, vol. 32, no. 12, pp. 1719–1722, 1996.

[2] H. Attias, "Planning by probabilistic inference," ser. Proceedings of Machine Learning Research. PMLR, 2003. [Online]. Available: https://proceedings.mlr.press/r4/attias03a.html

[3] R. A. Howard and J. E. Matheson, "Risk-sensitive Markov decision processes," *Management science*, vol. 18, no. 7, pp. 356–369, 1972.

[4] K. Friston, F. Rigoli, D. Ognibene, C. Mathys, T. Fitzgerald, and G. Pezzulo, "Active inference and epistemic value," *Cognitive neuroscience*, vol. 6, no. 4, pp. 187–214, 2015.

[5] A. Nilim and L. El Ghaoui, "Robust control of Markov decision processes with uncertain transition matrices," *Operations Research*, vol. 53, no. 5, pp. 780–798, 2005.

[6] G. N. Iyengar, "Robust dynamic programming," *Mathematics of Operations Research*, vol. 30, no. 2, pp. 257–280, 2005.

[7] B. Millidge, A. Tschantz, A. K. Seth, and C. L. Buckley, "On the relationship between active inference and control as inference," 2020. [Online]. Available: https://arxiv.org/abs/2006.12964

[8] B. de Vries, "Active inference for physical AI agents – an engineering perspective," *arXiv preprint arXiv:2603.20927*, 2026. [Online]. Available: https://arxiv.org/abs/2603.20927

[9] T. Lefebvre, "Probabilistic control and majorisation of optimal control," *Syst. Control Lett.*, vol. 190, p. 105837, 2024.