Consolidating Probabilistic Optimal Control

A. Bhole, M. M. Filabadi, G. Crevecoeur, T. Lefebvre
Department of Electromechanical Systems and Metal Engineering, Ghent University, Ghent, Belgium
Core lab MIRO, Flanders Make, Belgium
Email: {ajinkya.bhole, mohammad.mahmoudifilabadi, guillaume.crevecoeur, tom.lefebvre}@ugent.be

Over the past two decades, growing interplay between control engineering and certain areas of machine learning has elevated the role of probabilistic policies in control design. Randomness, once confined to disturbances and uncertainty, is now deliberately employed to promote exploration, robustness, and generalization, with entropy regularization emerging as a principled means to balance exploration and exploitation. This paradigm shift also motivated the search for control formulations in which the optimal or posterior policy admits a expectation-based representation with respect to a prior policy, enabling tractable solutions.

The field of (stochastic) optimal control today, has thus diversified into a rich mosaic of distinct modeling paradigms, each employing its own mathematical tools and conceptual frameworks. *Probabilistic Control Design* frames decision-making as a density-matching problem using KL divergence; *Control as Inference* embeds optimality variables in a probabilistic graphical model; *Risk-Sensitive Control* accounts for the underlying cost distribution beyond the mean; *Distributionally Robust Control* uses ambiguity sets to hedge against model mismatch; *KL-Regularized Control* adds entropy penalties to improve tractability; while *Linearly Solvable Optimal Control* exploits linear Bellman equations under specific structural assumptions. While these approaches often address closely related questions, their differing formulations and solution techniques have led to a fragmented landscape in which connections are obscured and comparisons are nontrivial. We consolidate these diverse perspectives through the common lens of regularization, providing a consistent and self-contained framework that clarifies how modeling assumptions translate into different control formulations.

The unifying framework is anchored in a central control problem that goes beyond standard trajectory-level regularization, by explicitly optimizing over both the control policy and an auxiliary system transition kernel, each penalized by a divergence from a baseline, and weighted independently. This separation makes it possible to distinguish between uncertainty induced by stochastic control actions and that attributed to deviations in system dynamics, thereby providing a principled mechanism to encode and understand robustness and risk sensitivity within a single formulation.

The resulting problem admits a representation of the form

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

where $\underline{\pi} = (\pi_0, \ldots, \pi_{T-1})$ is the policy sequence, $\underline{\tau} = (\tau_0, \ldots, \tau_{T-1})$ is an auxiliary transition kernel sequence, $p_{(\underline{\pi},\underline{\tau})}$ is the resulting trajectory distribution, and c_T is the cumulative cost. The baseline policy $\underline{\rho}$ and baseline dynamics $\underline{\iota}$ provide reference behaviors. The operator $\operatorname{opt}^{\lambda^S}$ denotes minimization when $\lambda^S > 0$ (risk-seeking) and maximization when $\lambda^S < 0$ (risk-averse). The weights $\lambda^P > 0$ and $\lambda^S \neq 0$ independently control the strength of regularization toward the policy and transition baselines, respectively. Serving as an expressive umbrella, (1) provides a systematic way to represent, relate, and compare several classical stochastic control formulations by toggling two structural choices: (i) Whether the control policy is regularized against a baseline reference $\underline{\rho}$ (policy regularization on/off). (ii) Whether the transition kernel sequence is fixed to the true dynamics $\underline{\iota}$ or left free as an optimization variable (transition optimization on/off). Through these toggles we recover:

- **Stochastic Optimal Control (SOC)**: No policy regularization by setting $\underline{\rho} = \underline{\pi}$ and constraining transitions $\underline{\tau} = \underline{\iota}$.
- **Risk-Sensitive Optimal Control (RSOC)**: No policy regularization, while leaving $\underline{\tau}$ free.
- **Soft-Policy SOC (SP-SOC)**: Policy $\underline{\pi}$ regularized w.r.t. baseline $\underline{\rho}$ and fixing $\underline{\tau} = \underline{\iota}$.
- **Soft-Policy RSOC (SP-RSOC)**: Allowing both $\underline{\pi}$ and $\underline{\tau}$ to be regularized and optimized.

Using the Majorization–Minimization (MM) framework it can be shown that the regularized problems SP-SOC and SP-RSOC majorize the original SOC and RSOC objectives, respectively, providing tractable surrogates for iterative algorithm design. When the baseline dynamics are deterministic, the auxiliary transition kernel loses its expressive freedom, and we have: SOC $\equiv$ RSOC, and SP-SOC $\equiv$ SP-RSOC.

For SP-RSOC, the special case of synchronized regularization, where the policy and transition regularization weights coincide ($\lambda^P = \lambda^S = \lambda > 0$), enjoys powerful properties: a linear Bellman equation, a path integral solution, and compositionality that enables modular control design. These properties, previously known only in restricted settings, are thus extended to a broader class.

These insights open up several promising research avenues. Our future work will explore time-varying regularization weights to adapt risk attitudes throughout the horizon, shift from soft to hard constraints, and extend the framework to partially observable settings. Practical algorithmic development could also build on these connections to design efficient solvers for a wide range of optimal control applications.

References

[1] A. Bhole, M. M. Filabadi, G. Crevecoeur, T. Lefebvre, "Unifying Entropy Regularization in Optimal Control: From and Back to Classical Objectives via Iterated Soft Policies and Path Integral Solutions," arXiv:2512.06109, 2025.