Increasing Information for Model Predictive Control with Semi-Markov Decision Processes

Rémy Hosseinkhan Boucher, Stella Douka, Onofrio Semeraro, Lionel Mathelin,
LISN, CNRS, Université Paris-Saclay

How and **when should data be collected along the system trajectory** *for Gaussian Process-Based Model Predictive Control?*

Background

Markov Decision Process (MDP)

$$P\left(dx_0 du_0 dx_1 \ldots\right) = P_{X_0}\left(dx_0\right) \pi\left(du_0 \mid dx_0\right) \mathcal{P}\left(dx_1 \mid dx_0, du_0\right) \ldots$$

$\mathcal{P}$: transition Kernel $\quad u \in \mathcal{U}$: control $\quad x \in \mathcal{X}$: state

Gaussian Process (GP) Model Predictive Control (MPC)

Dynamics model (GP)

$$\hat{\mathcal{P}}_{\mathcal{D}}(\cdot,(x,u)) \sim \mathcal{N}(\mu(x,u), \Sigma((x,u),(x,u)) \mid \mathcal{D})$$

Cost function

$$J^{\pi} = \mathbb{E}^{\pi}\left[\sum_{k=0}^{T} c\left(X_k, U_k\right)\right]$$

Model Predictive Control with Cross-Entropy Method (CEM)

$$\pi^{\mathrm{MPC}}(x) = u_0^*$$

$$s.t. \quad \left(u_0^*, \ldots, u_{T^{\mathrm{MPC}}}^*\right) = \underset{\left(u_0, \ldots, u_{T^{\mathrm{MPC}}}\right)}{\arg\min}\ \mathbb{E}^{\left(u_0, \ldots, u_{T^{\mathrm{MPC}}}\right)}\left[\sum_{k=0}^{T^{\mathrm{MPC}}} c\left(X_k, u_k\right) \mid X_0 = x\right]$$

Objective

Collect minimal $\mathcal{D} = (x_k, u_k)_{k=1}^{n}$ **such that** $\hat{\mathcal{P}}_{\mathcal{D}} \simeq \mathcal{P}$ **from evolving dynamical system**

State-of-the-art strategy: **Expected Information Gain (EIG)** on the optimal trajectory $H_T^* = (X_0^*, U_0^*, \ldots, U_{T-1}^*, X_T^*)$ (under π^*)

$$\mathrm{EIG}_n(x,u) = \mathcal{H}[\hat{H}_T^* \mid \mathcal{D}_n] - \mathbb{E}_{P_{X_{n+1} \mid \mathcal{D}_n,\, X_n = x,\, U_n = u}}[\mathcal{H}[\hat{H}_T^* \mid \mathcal{D}_n, X_n = x, U_n = u, X_{n+1}]]$$

Dataset update: $\mathcal{D}_{n+1} = \mathcal{D}_n \cup (x^*, u^*) \quad (x^*, u^*) = \mathrm{argmax}_{x,u}\, \mathrm{EIG}(x,u)$

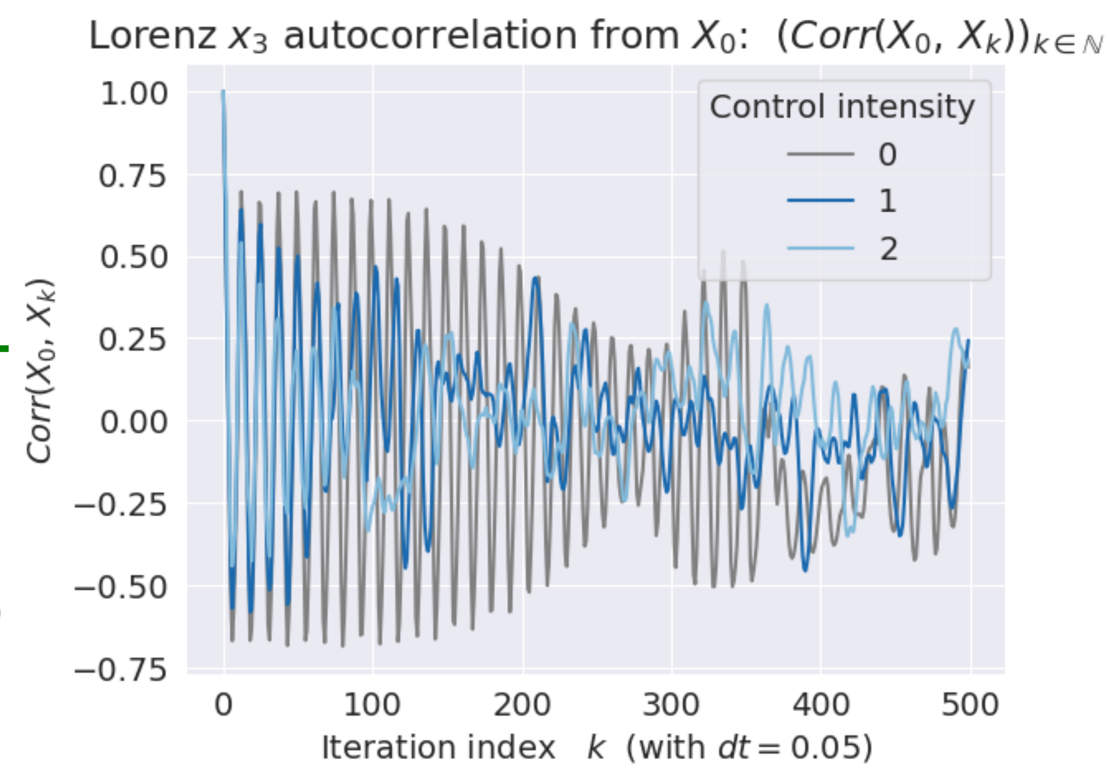
Select point which minimises the uncertainty (entropy) $\mathcal{H}$ on the optimal trajectory

By symmetry of EIG, the entropy of X_{t+1} (more tractable) is considered:

$$\mathrm{EIG}_n(x,u) = \mathcal{H}\left[X_{n+1} \mid \mathcal{D}_n, X_n = x, U_n = u\right] - \underset{P_{\hat{H}_T^* \mid \mathcal{D}_n}}{\mathbb{E}}\left[\mathcal{H}\left[X_{n+1} \mid \mathcal{D}_n, X_n = x, U_n = u, \hat{H}_T^*\right]\right]$$

Question

How to **extend the EIG criterion** to avoid **information redundancy?**



Contribution

Introduction of **temporal abstraction** with Semi-markov Decision Process (options framework)

Temporally extended actions (options) $\quad$ Interdecision time $\longrightarrow$ $t \in \{1, \ldots, t_{\max}\}$

New temporally-extended transition $\mathcal{P}^{\mathrm{SMDP}}\left(dx' \mid (x,(u,t))\right) = P\left(X_{k+t} \mid X_k = x, U_{k:k+t-1} = u\right)$

No system **interaction** for a duration of length t

Look-ahead (non-causal) information criterion

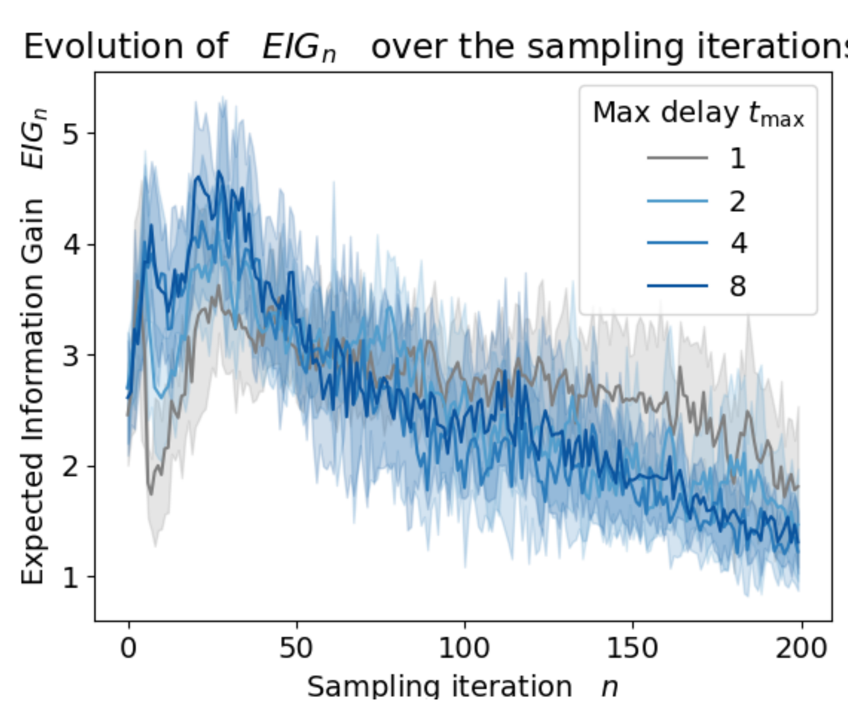
$$\mathrm{EIG}^{\mathrm{SM}}(x,(u,t)) = \mathcal{H}\left[X_{\kappa_n+t+1} \mid \mathcal{D}_n, X_{\kappa_n} = x, U_{\kappa_n:\kappa_n+t} = u, \kappa_n\right] - \mathbb{E}_{P_{\hat{H}_T^* \mid \mathcal{D}_n}}\left[\mathcal{H}\left[X_{\kappa_n+t+1} \mid \mathcal{D}_n, X_{\kappa_n} = x, U_{\kappa_n:\kappa_n+t} = u, \hat{H}_T^*, \kappa_n\right]\right]$$

Dataset update: $\mathcal{D}_{n+1} = \mathcal{D}_n \cup (x^*, u^*) \quad (x^*, u^*) = \mathrm{argmax}_{x,u,t}\mathrm{EIG}^{\mathrm{SM}}(x,(u,t))$

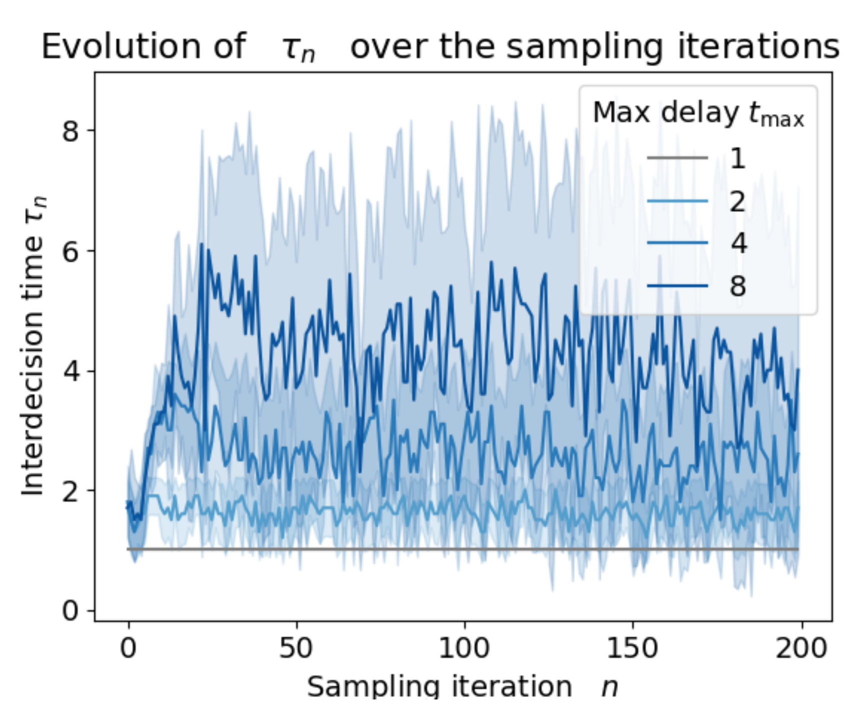
Results

Experiment on the Lorenz system:

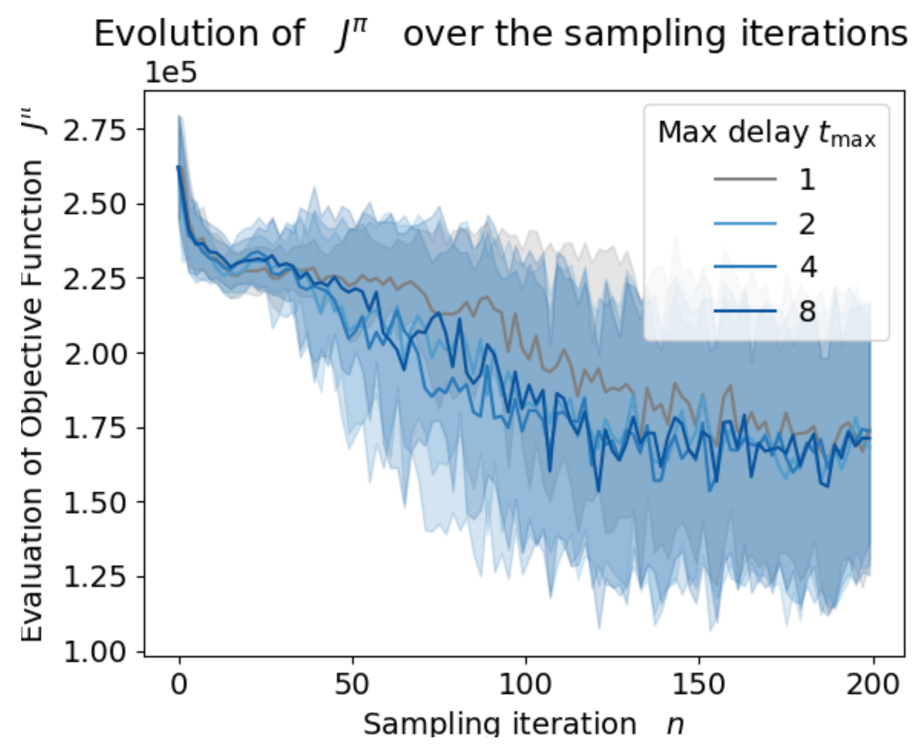
- *max inter-decision time:* $t_{\max} = 8$
- *sampling budget:* $n = 200$



✓ Temporal abstraction increases EIG



✓ Semi-markov interdecision time $t > 1$



✓ Improved MPC

V. Mehta et al. - An Experimental Design Perspective on Model-Based Reinforcement Learning (ICLR, 2022)

R. S. Sutton, D. Precup, S. Singh - Between MDPs and semi-MDPs: A Framework for Temporal Abstraction in Reinforcement Learning (Artifical Intelligence, 1999)

cnrs · université PARIS-SACLAY · LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE · anr agence nationale de la recherche