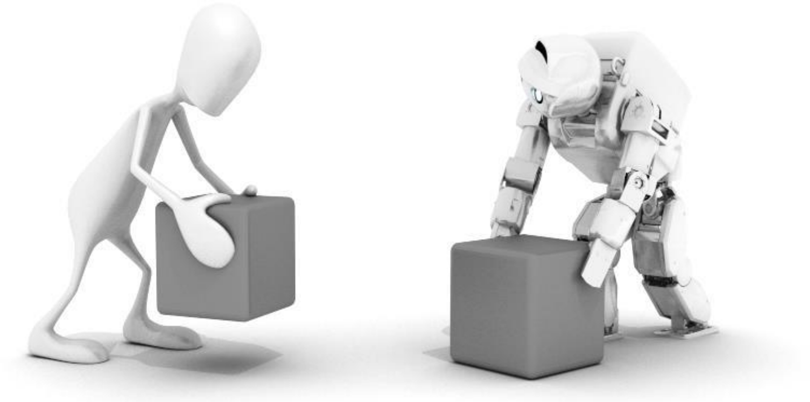
Active Improvement of Control Policies with Bayesian Gaussian Mixture Model

Hakan Girgin, Emmanuel Pignat, Noémie Jaquier and Sylvain Calinon

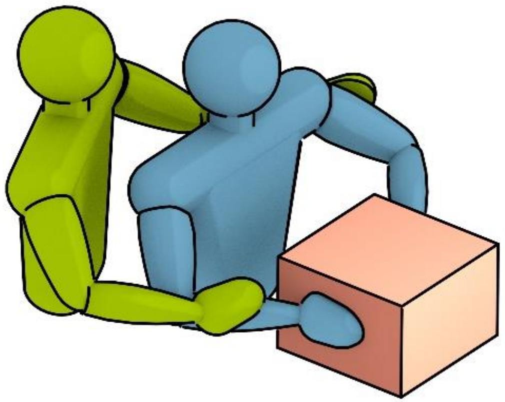
Idiap Research Institute, EPFL, Switzerland

IROS2020

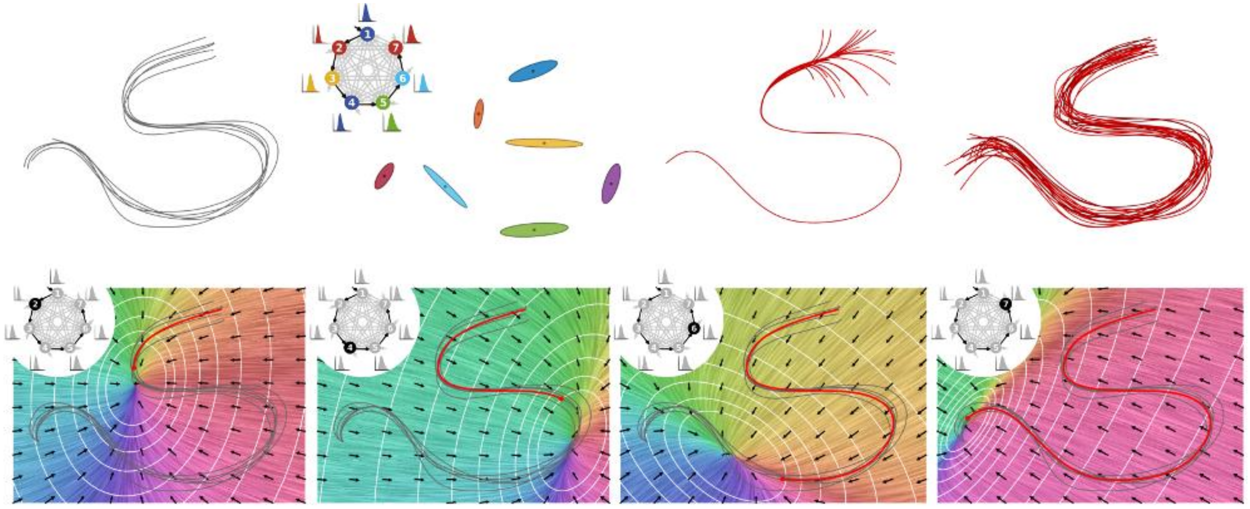
Motivation: Learning from Demonstration (LfD)



User friendly transfer of skills

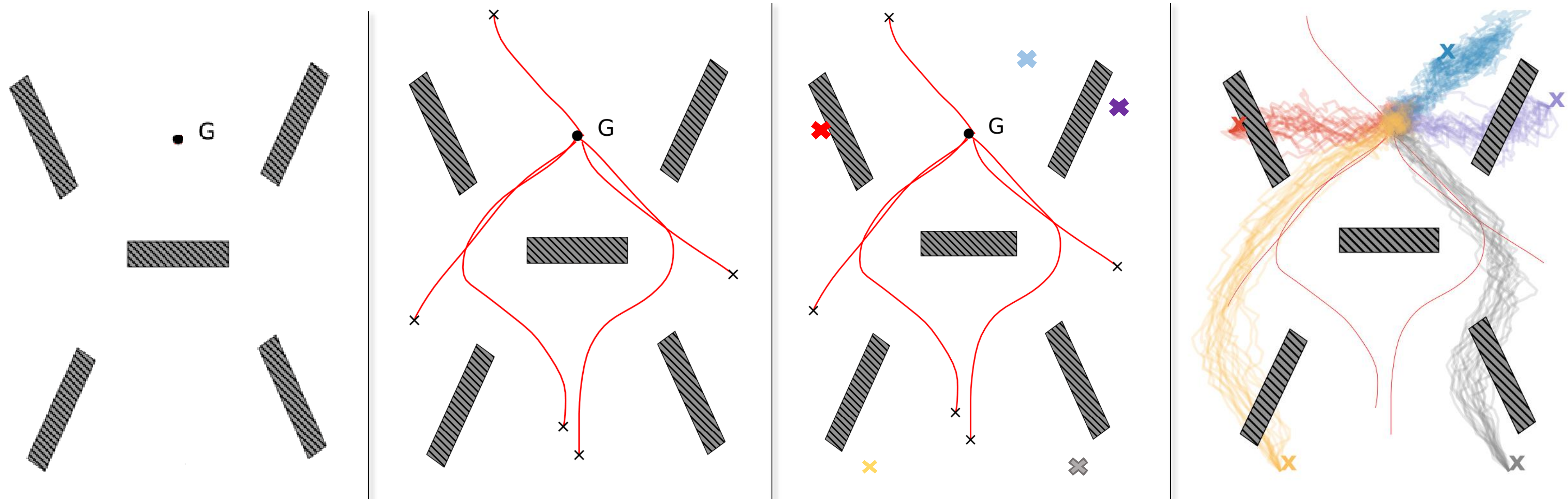


Kinesthetic teaching



Adaptive movement representation

Motivation: Challenges



Obstacle avoidance task | Initial demonstrations | Random generalization test | Sampling from control policy

$p(\boldsymbol{u}_t|\boldsymbol{x}_t)$?

Human Teacher

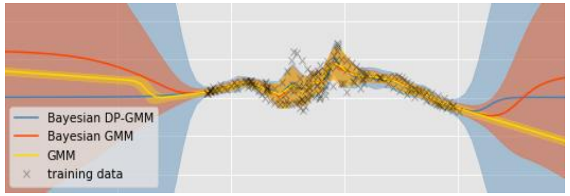
Learning $p(\boldsymbol{u}_t|\boldsymbol{x}_t)$

$\boldsymbol{u}_t \sim p(\boldsymbol{u}_t|\boldsymbol{x}_t) \quad \forall t = 1, ..., T$

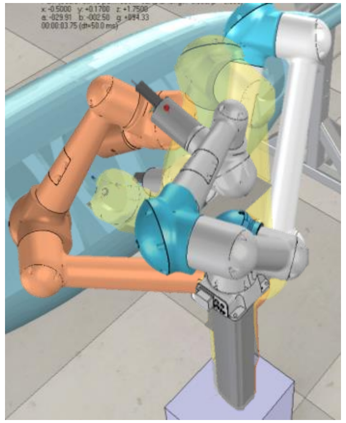
Overview of proposed active learning framework

We propose an **active learning framework** for control policies for
- Good generalization with **few demonstrations**
- Reducing the **cognitive load** on the teacher

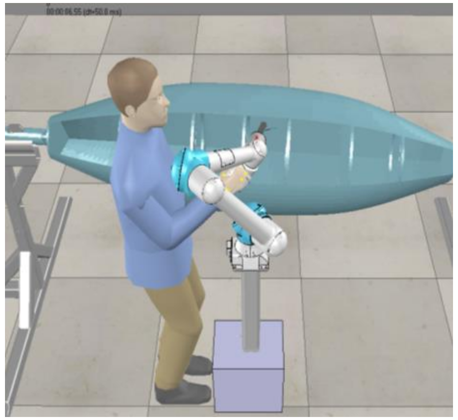
- Learn a Bayesian model which can encode variations in the demonstrations (for compliance) and uncertainties of the model (for exploration)
- Find an uncertainty measure of the learned model and the variable that maximizes it.
- Robot requests a demonstration around the most informative state.



Movement model learning



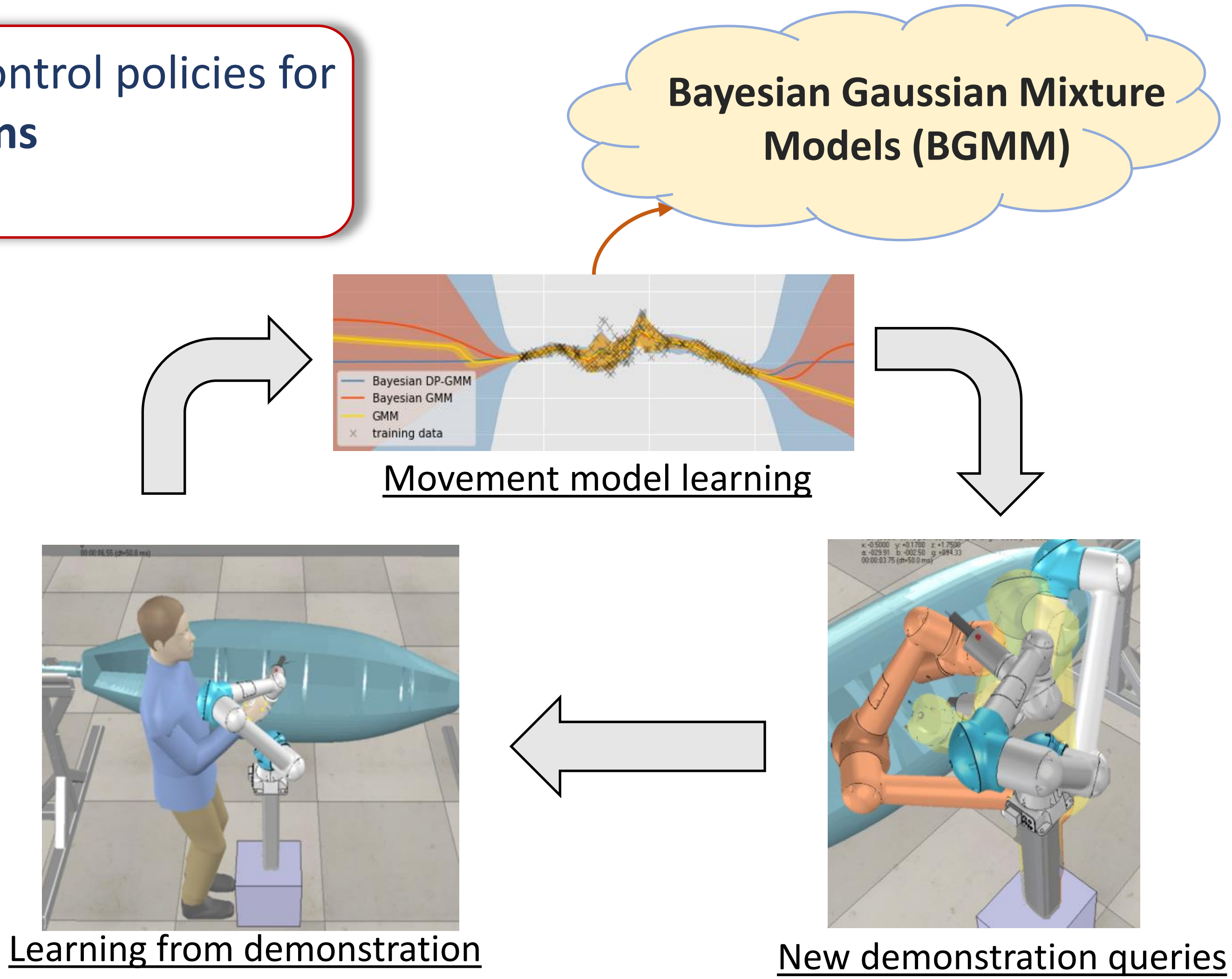
New demonstration queries



Learning from demonstration

Overview of proposed active learning framework

We propose an **active learning framework** for control policies for
- Better generalization with **few demonstrations**
- Reducing the **cognitive load** on the teacher



Contributions

1. **An uncertainty decomposition** in BGMM control policies
2. **Information-weighted** closed-form cost function for **uncertainty maximization**
3. **Active learning** framework with easy monitoring of the uncertainty reduction

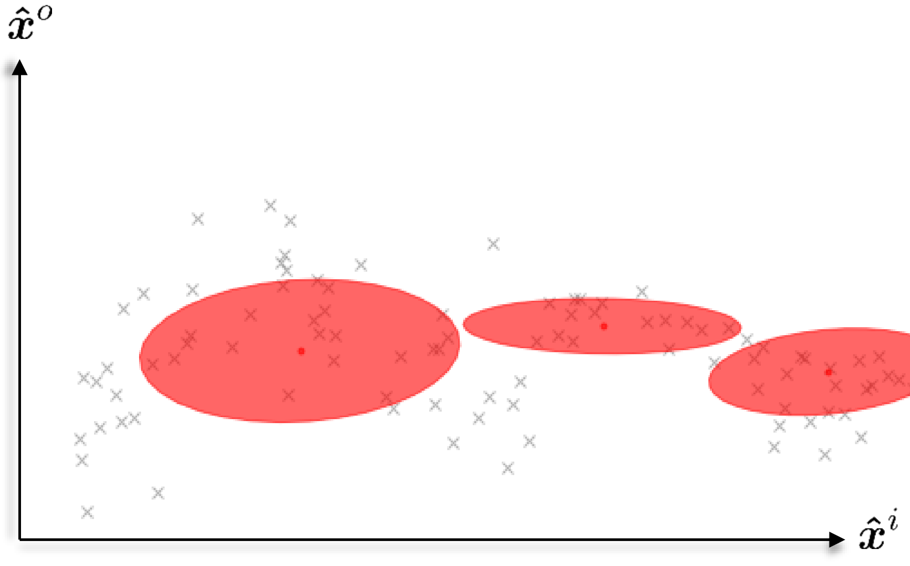
Background: Learning BGMM control policies

Bayesian Gaussian Mixture Models (BGMM)*

Learning the joint distribution $p(\boldsymbol{x}) = p(\boldsymbol{x}^i, \boldsymbol{x}^o)$

State of the robot : x

Control action: u



Model: $$p(\boldsymbol{x}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Lambda}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Lambda}_k^{-1})$$

Posterior : $$p(\hat{\boldsymbol{x}}|\boldsymbol{X}) = \sum_{k=1}^{K} \hat{\pi}_k \mathrm{t}(\hat{\boldsymbol{x}}|\hat{\boldsymbol{m}}_k, \hat{\boldsymbol{L}}_k, \hat{\nu}_k)$$

Posterior Conditional: $$p(\hat{\boldsymbol{x}}^o|\hat{\boldsymbol{x}}^i, \boldsymbol{X}) = \sum_{k=1}^{K} \hat{\pi}_k^{o|i} \mathrm{t}(\hat{\boldsymbol{x}}^i|\hat{\boldsymbol{m}}_k^{o|i}, \hat{\boldsymbol{L}}_k^{o|i}, \hat{\nu}_k^{o|i})$$

Prior on $\boldsymbol{\mu}, \boldsymbol{\Lambda}$ $p(\boldsymbol{\mu}, \boldsymbol{\Lambda})$	$\prod_{k=1}^{K} \mathcal{N}(\boldsymbol{\mu}_k|\boldsymbol{m}_0, (\beta_0 \boldsymbol{\Lambda}_k)^{-1}) \mathcal{W}(\boldsymbol{\Lambda}_k|\boldsymbol{W}_0, \nu_0)$
Prior on $\boldsymbol{\pi}$ $p(\boldsymbol{\pi})$	$\mathrm{Dir}(\boldsymbol{\pi}|\alpha_0)$

Conjugate Priors

*E. Pignat and S. Calinon, “Bayesian Gaussian mixture model for robotic policy imitation,” IEEE Robotics and Automation Letters, 2019.

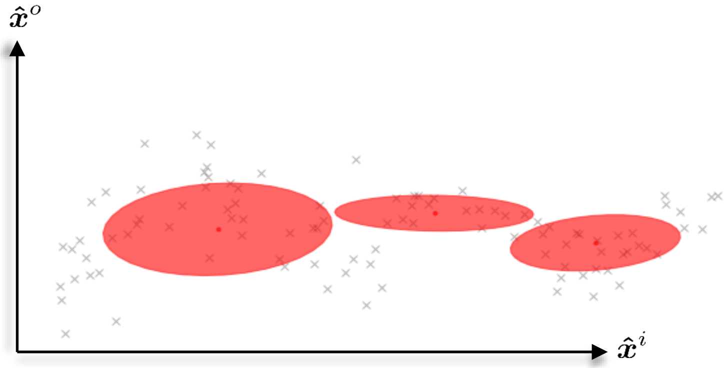
A closer look at the covariance matrix

$$p(\hat{\boldsymbol{x}}^o|\hat{\boldsymbol{x}}^i, \boldsymbol{X}) = \sum_{k=1}^{K} \hat{\pi}_k^{o|i} \mathrm{t}(\hat{\boldsymbol{x}}^i|\hat{\boldsymbol{m}}_k^{o|i}, \boxed{\hat{\boldsymbol{L}}_k^{o|i}}, \hat{\nu}_k^{o|i}),$$

Mixture of multivariate t-distributions

$$\left.\begin{aligned} \boldsymbol{L}_s &= \boldsymbol{L}_k^{oo} - \boldsymbol{L}_k^{oi} {\boldsymbol{L}_k^{ii}}^{-1} {\boldsymbol{L}_k^{oi}}^T \\ \hat{\boldsymbol{L}}_k^{o|i} &= \frac{\hat{\nu}_k + (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)^T {\boldsymbol{L}_k^{ii}}^{-1} (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)}{\hat{\nu}_k^{o|i}} \boldsymbol{L}_s \end{aligned}\right\}$$

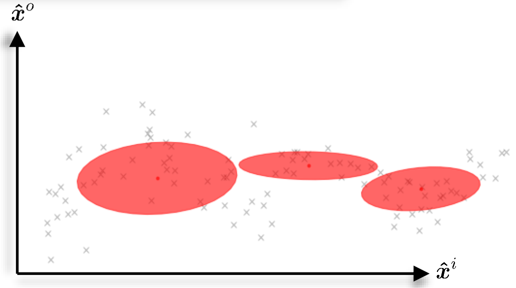
Covariance matrices



Decomposition of the covariance matrix

$$p(\hat{\boldsymbol{x}}^o|\hat{\boldsymbol{x}}^i,\boldsymbol{X}) = \sum_{k=1}^{K} \hat{\pi}_k^{o|i} \mathrm{t}(\hat{\boldsymbol{x}}^i|\hat{\boldsymbol{m}}_k^{o|i}, \hat{\boldsymbol{L}}_k^{o|i}, \hat{\nu}_k^{o|i})$$

Aleatoric $\sim$ Variations
Epistemic $\sim$ Uncertainties

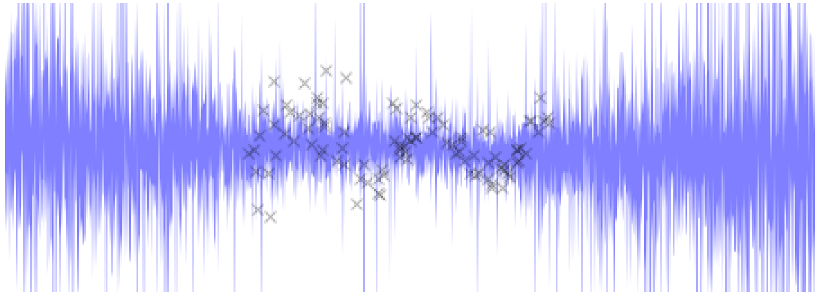


a) Total Covariance Matrix

$$\boldsymbol{L}_s = \boldsymbol{L}_k^{oo} - \boldsymbol{L}_k^{oi} \boldsymbol{L}_k^{ii^{-1}} \boldsymbol{L}_k^{oi^T}$$

- GMM conditioning
- GMR

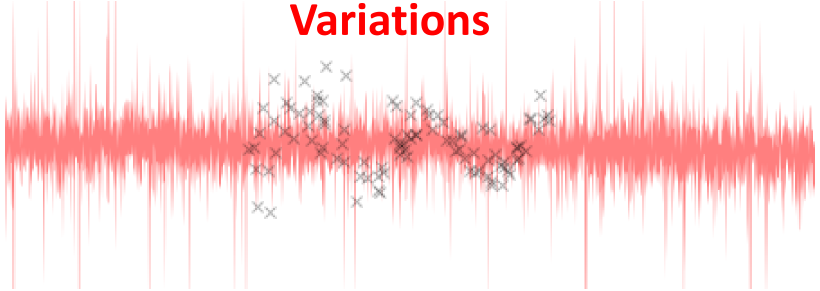
$$\hat{\boldsymbol{L}}_k^{o|i} = \frac{\hat{\nu}_k + (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)^T \boldsymbol{L}_k^{ii^{-1}} (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)}{\hat{\nu}_k^{o|i}} \boldsymbol{L}_s$$



b) Aleatoric Covariance Matrix

$$\hat{\boldsymbol{L}}_k^{\mathrm{al}} = \frac{\hat{\nu}_k}{\hat{\nu}_k^{o|i}} \boldsymbol{L}_s$$

Constant!



c) Epistemic Covariance Matrix

$$\hat{\boldsymbol{L}}_k^{\mathrm{ep}} = \frac{(\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)^T \boldsymbol{L}_k^{ii^{-1}} (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)}{\hat{\nu}_k^{o|i}} \boldsymbol{L}_s$$

Quadratic!

Quadratic Rényi entropy as uncertainty measure

Quadratic Rényi entropy for exponential mixtures:

$$H_2(p(\boldsymbol{u}|\boldsymbol{x})) = -\log \int_K p^2(\boldsymbol{u}|\boldsymbol{x})\mathrm{d}\boldsymbol{u},$$

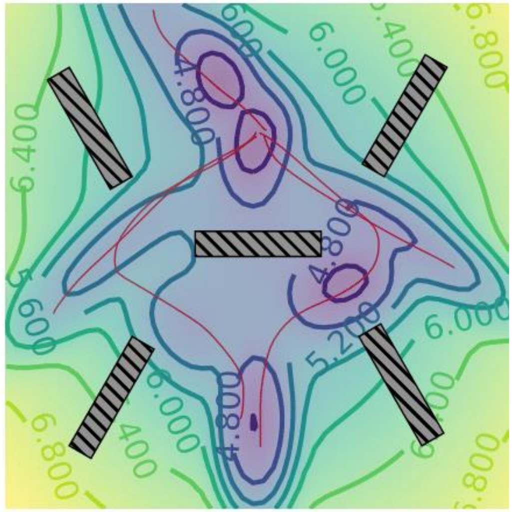
$$H_2(p(\boldsymbol{u}|\boldsymbol{x})) = -\log \sum_{i=1}^{K} \sum_{j=1}^{K} \pi_i(\boldsymbol{x})\pi_j(\boldsymbol{x}) e^{\Delta_{ij}(\boldsymbol{x})}$$

Moment matching of a t-distribution with a Gaussian:

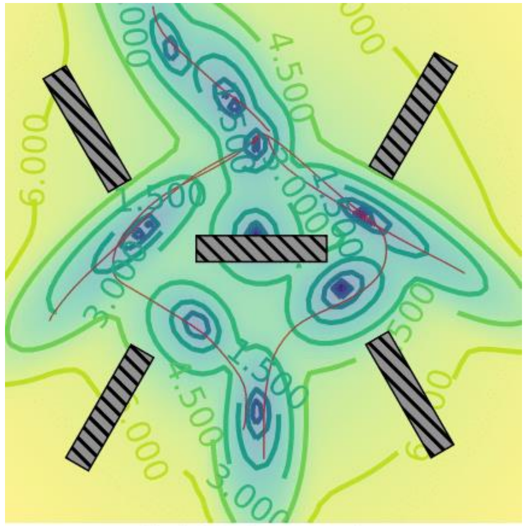
$$t_\nu(\boldsymbol{u}|\boldsymbol{\mu}(\boldsymbol{x}), \boldsymbol{\Sigma} \sim \mathcal{N}(\boldsymbol{u}|\tilde{\boldsymbol{\mu}}(\boldsymbol{x}), \tilde{\boldsymbol{\Sigma}}(\boldsymbol{x})))$$

$$\tilde{\boldsymbol{\mu}}(\boldsymbol{x}) = \boldsymbol{\mu}(\boldsymbol{x}), \qquad \tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}) = \frac{\nu}{\nu - 2}\boldsymbol{\Sigma}(\boldsymbol{x}).$$

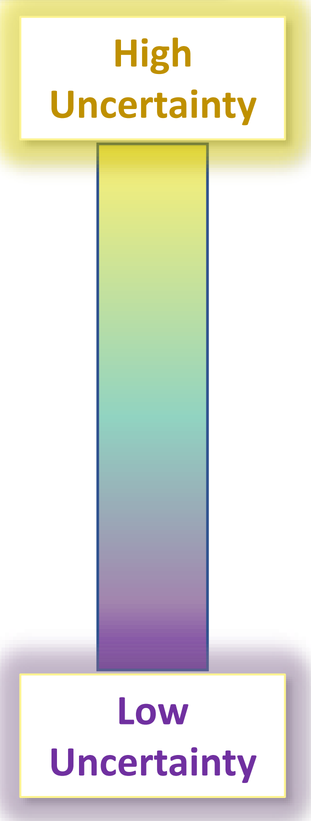
$$\hat{\boldsymbol{L}}_k^{o|i} = \frac{\hat{\nu}_k + (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)^T \boldsymbol{L}_k^{ii^{-1}} (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)}{\hat{\nu}_k^{o|i}} \boldsymbol{L}_s \qquad \hat{\boldsymbol{L}}_k^{\mathrm{ep}} = \frac{(\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)^T \boldsymbol{L}_k^{ii^{-1}} (\hat{\boldsymbol{x}}^i - \hat{\boldsymbol{m}}_k^i)}{\hat{\nu}_k^{o|i}} \boldsymbol{L}_s$$



(a) Total



(c) Epistemic



Uncertainty maximization

Uncertainty maximization for active learning:

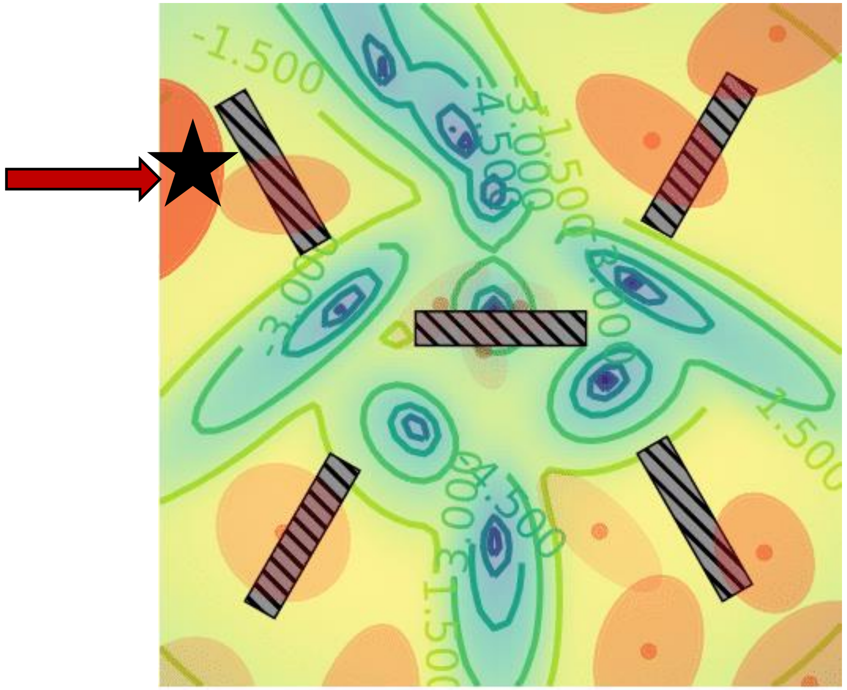
$$\underset{\boldsymbol{x}}{\operatorname{argmin}} -H_2(p(\boldsymbol{u}|\boldsymbol{x}))$$

- Will most certainly diverge if not constrained.
- If constrained, will only find solutions at the borders

Information density approach for active learning:

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} -H_2(p(\boldsymbol{u}|\boldsymbol{x})) - \beta \log p_{\mathrm{sim}}(\boldsymbol{x})$$

- Divergence issue resolved.
- Soft constraint
- **Flat gradients**



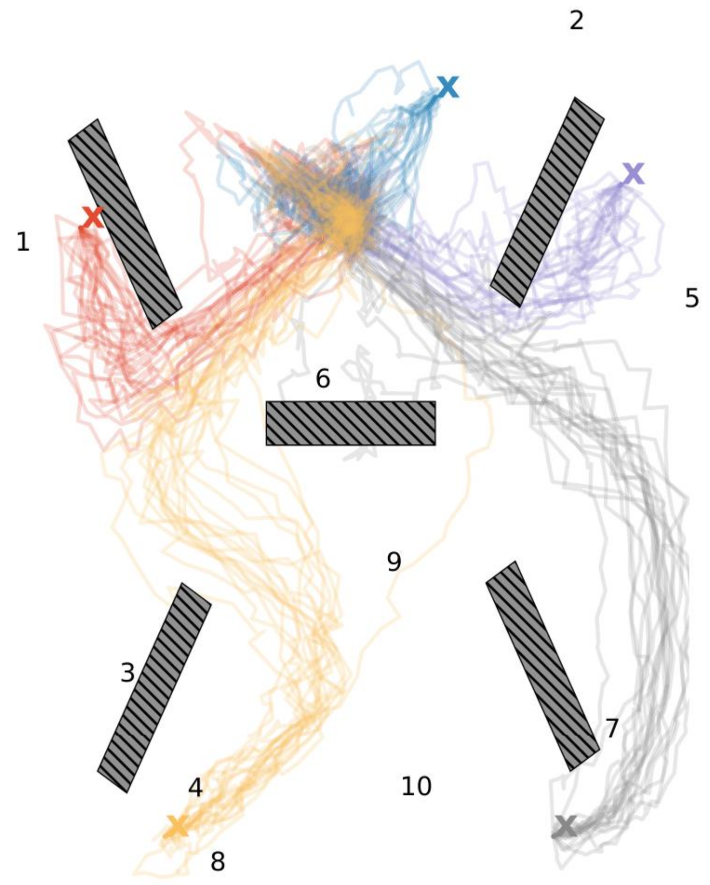
(d) Information

Gaussian Mixture Model (GMM)

Proposed solution: $$\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, KL\Big(q(\boldsymbol{x}) || H_2(p(\boldsymbol{u}|\boldsymbol{x})) + \beta \log p_{\mathrm{sim}}(\boldsymbol{x})\Big)$$

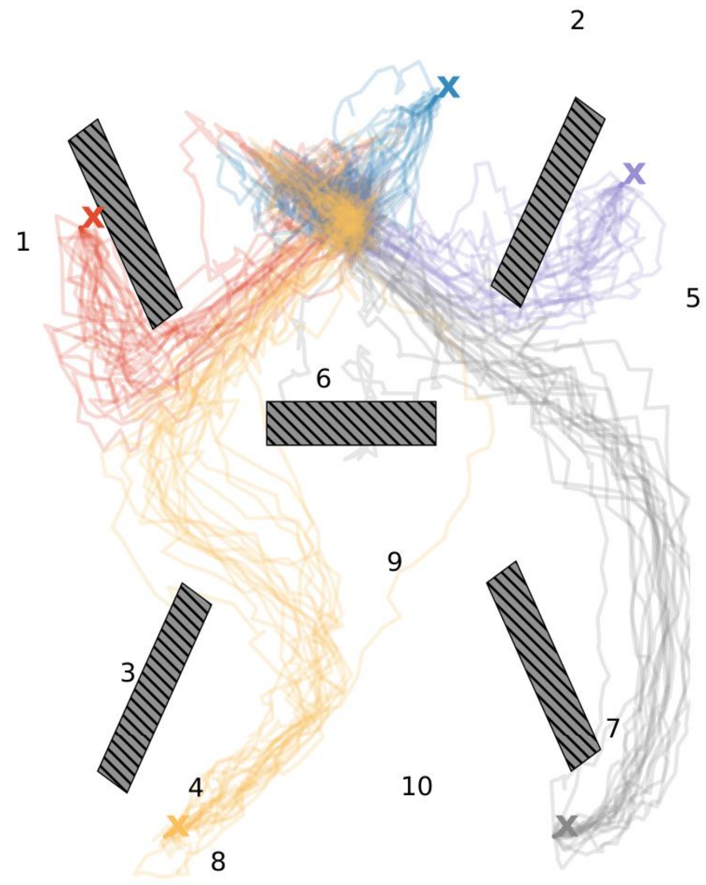
GMM Parameters

Simulated experiment



BGMM policy sampling after active learning

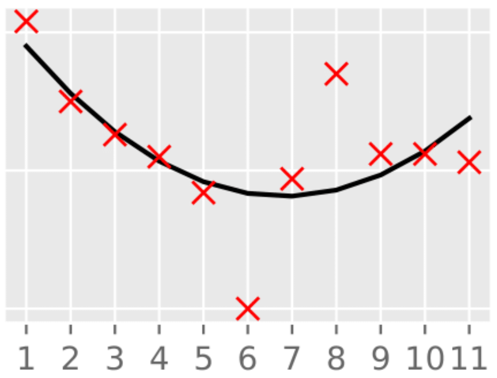
Simulated experiment



BGMM policy sampling after active learning

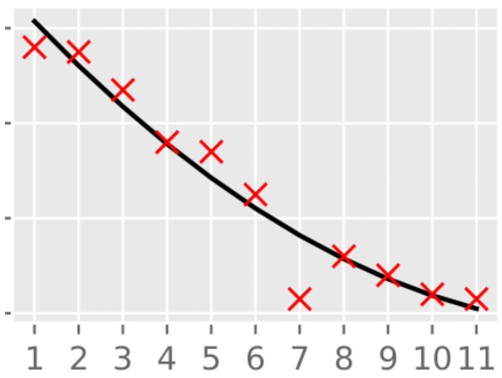
GMM Model

Uncertainty in q(x)



Marginal Model

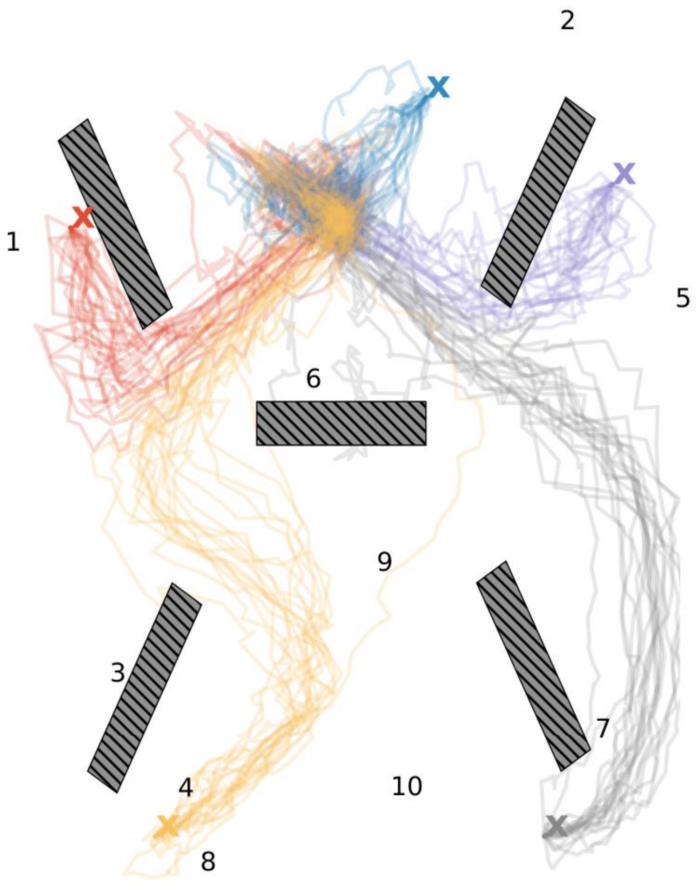
Uncertainty in p(x)



Active query request

$$p(\boldsymbol{u}_t|\boldsymbol{x}_t)$$

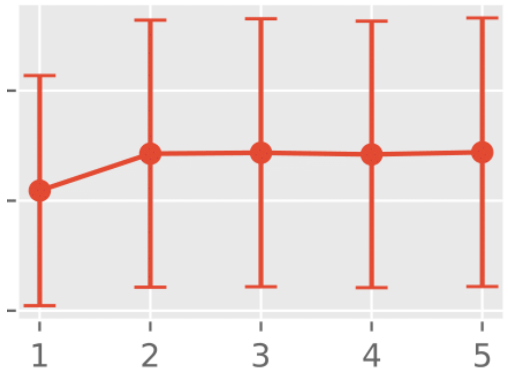
Simulated experiment



BGMM policy sampling after active learning

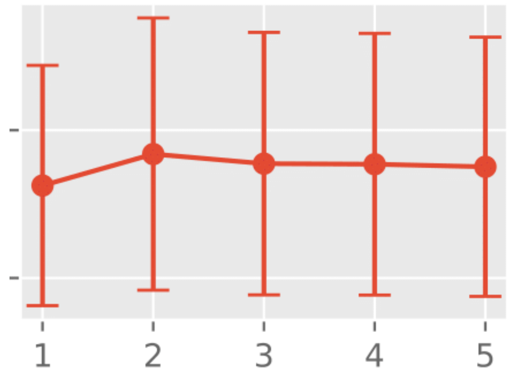
GMM Model

Uncertainty in q(x)



Marginal Model

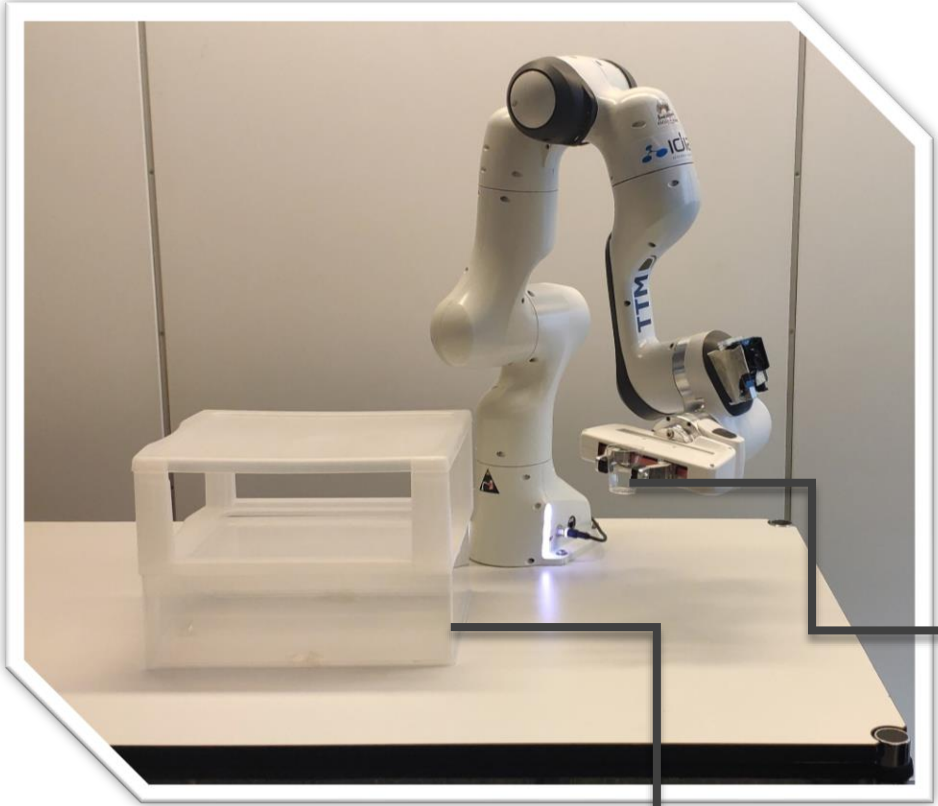
Uncertainty in p(x)

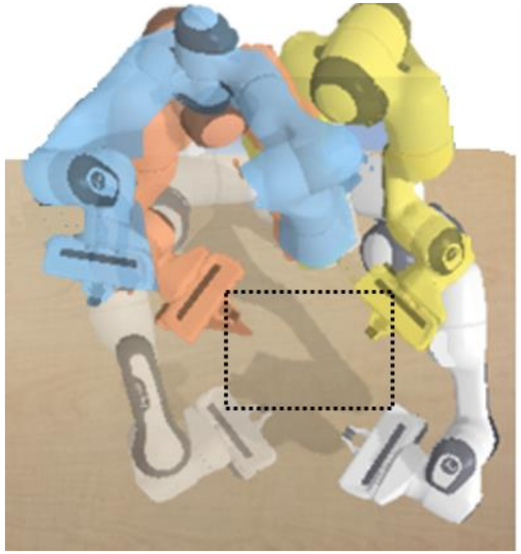


Random query request

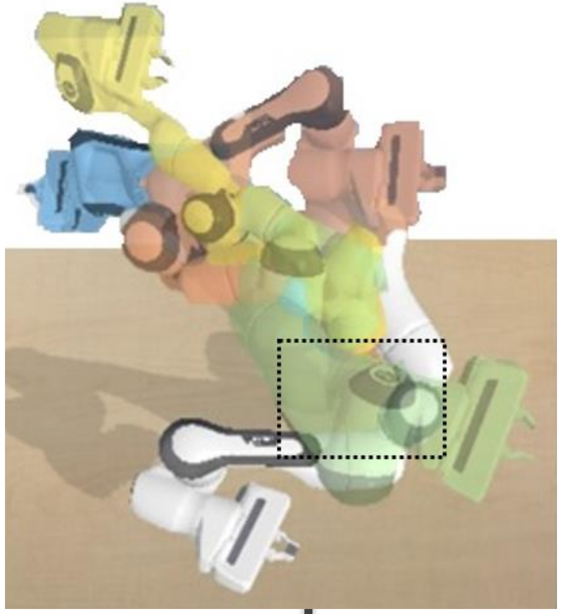
Robot experiment

$$\underset{\boldsymbol{x}}{\operatorname{argmin}}\, KL\Big(q(\boldsymbol{x})||H_2(p(\boldsymbol{u}|\boldsymbol{x}))$$
$$+\beta \log p_{\text{jointlimits}}(\boldsymbol{x})$$
$$+\alpha \log p_{\text{upright}}(\boldsymbol{x})\Big)$$





Initial demonstrations' starting configurations



Requested initial configurations for demonstration

Conclusions

- We presented an active learning framework allowing a robot to ask for informative new demonstrations
- Representation of closed-form epistemic uncertainties in BGMM control policies
- Variational inference approach to capture all the maximal information areas
- Can reduce the cognitive load of the teacher

Future Work

- How to propagate uncertainties in the state-action policies, or on extending it to **trajectory policies**.
- Theoretically determine a threshold to stop the learning process.
- Answer two of the main questions of LfD : i) Where to give demonstrations?
ii) How many demonstrations are required?

Thanks for watching!

Do not hesitate to contact me for more information:

hakan.girgin@idiap.ch