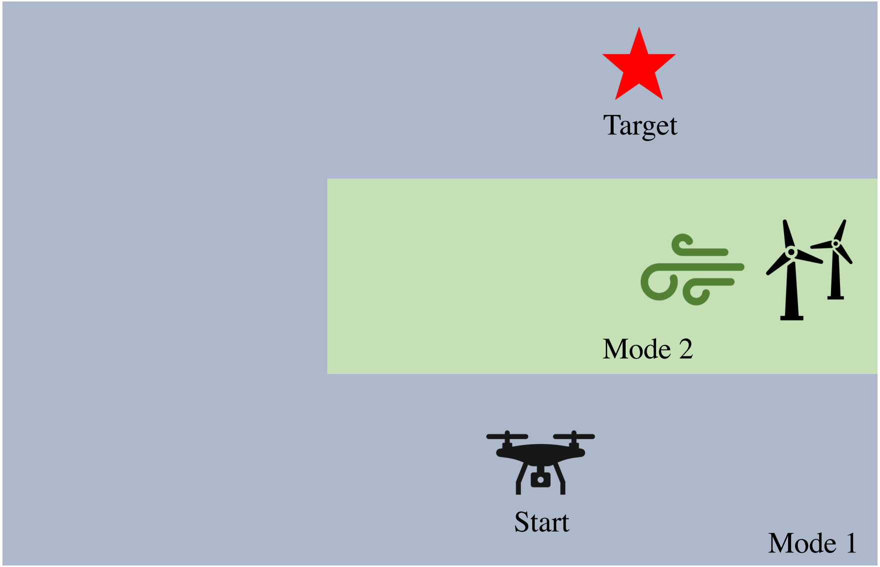
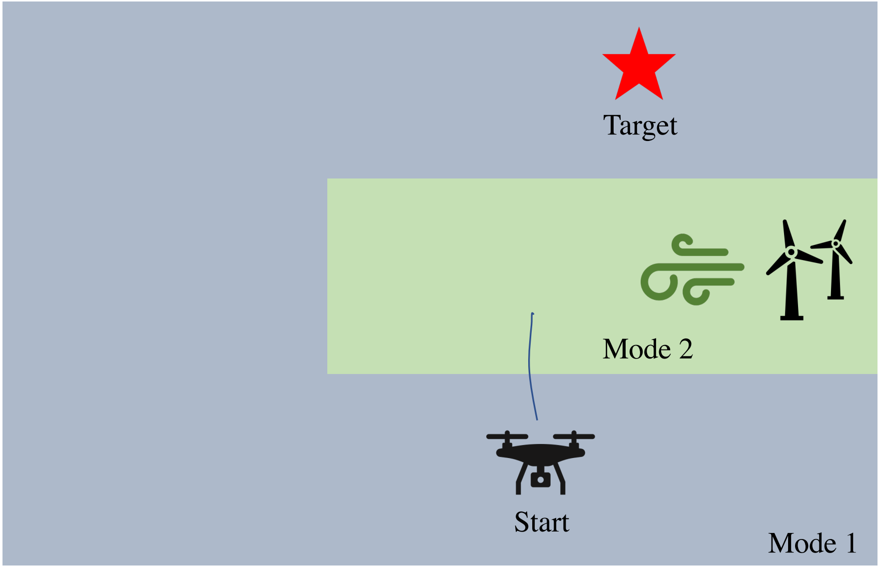
University of BRISTOL

Bayesian Learning for Control in Multimodal Dynamical Systems

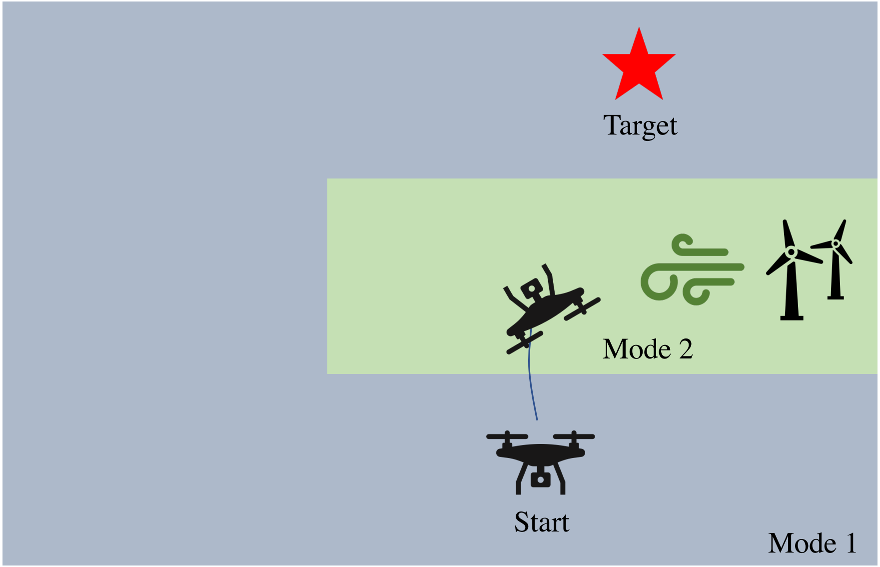
Aidan Scannell | Carl Henrik Ek | Arthur Richards

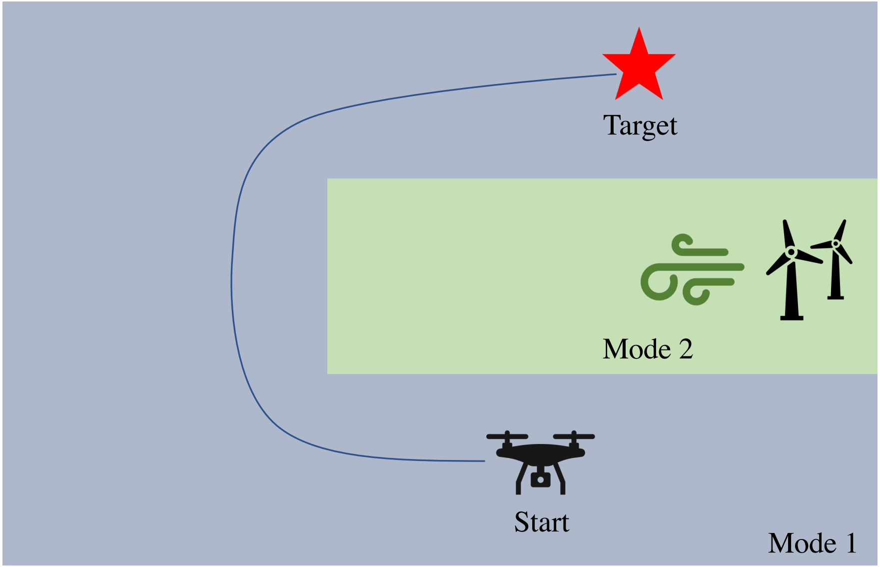
23rd June 2022


Target
Mode 2
Start
Mode 1



Target
Mode 2
Start
Mode 1


Target
Mode 2
Start
Mode 1


Target
Mode 2
Start
Mode 1

Goals

Goal 1 Navigate to the target state x_f

Goals

Goal 1 Navigate to the target state x_f

Goal 2 Remain in the operable, desired dynamics mode k^*

Mode remaining navigation problem

$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}\left[\sum_{t=0}^{T} c(\mathrm{x}_t, \pi(\mathrm{x}_t, t)) \mid \mathrm{x}_0 = \mathrm{x}_0\right] && \text{(1a)} \\
\text{s.t.} \quad & && \text{(1b)} \\
& && \text{(1c)} \\
& && \text{(1d)} \\
& && \text{(1e)}
\end{aligned}
$$

Mode remaining navigation problem

$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}\left[\sum_{t=0}^{T} c(\mathrm{x}_t, \pi(\mathrm{x}_t, t)) \mid \mathrm{x}_0 = \mathrm{x}_0\right] && \text{(1a)} \\
\text{s.t.} \quad & && \text{(1b)} \\
& && \text{(1c)} \\
& \mathrm{x}_0 = \mathrm{x}_0 && \text{(1d)} \\
& && \text{(1e)}
\end{aligned}
$$

Mode remaining navigation problem

$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}\left[\sum_{t=0}^{T} c(\mathrm{x}_t, \pi(\mathrm{x}_t, t)) \mid \mathrm{x}_0 = \mathrm{x}_0\right] && \text{(1a)} \\
\text{s.t.} \quad & && \text{(1b)} \\
& && \text{(1c)} \\
& \mathrm{x}_0 = \mathrm{x}_0 && \text{(1d)} \\
& \mathrm{x}_T = \mathrm{x}_f && \text{(1e)}
\end{aligned}
$$

Mode remaining navigation problem

$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}\left[\sum_{t=0}^{T} c(\mathrm{x}_t, \pi(\mathrm{x}_t, t)) \mid \mathrm{x}_0 = \mathrm{x}_0\right] && \text{(1a)} \\
\text{s.t.} \quad & \mathrm{x}_{t+1} = f_k(\mathrm{x}_t, \pi(\mathrm{x}_t, t)) + \epsilon_k, \quad \text{if } \alpha(\mathrm{x}_t) = k \qquad \forall t \in \{0, \ldots, T-1\} && \text{(1b)} \\
& && \text{(1c)} \\
& \mathrm{x}_0 = \mathrm{x}_0 && \text{(1d)} \\
& \mathrm{x}_T = \mathrm{x}_f && \text{(1e)}
\end{aligned}
$$

Mode remaining navigation problem

$$
\begin{aligned}
\min_{\pi \in \Pi} \quad & \mathbb{E}\left[\sum_{t=0}^{T} c(\mathrm{x}_t, \pi(\mathrm{x}_t, t)) \mid \mathrm{x}_0 = \mathrm{x}_0\right] && && \text{(1a)} \\
\text{s.t.} \quad & \mathrm{x}_{t+1} = f_k(\mathrm{x}_t, \pi(\mathrm{x}_t, t)) + \epsilon_k, \quad \text{if } \alpha(\mathrm{x}_t) = k && \forall t \in \{0, \ldots, T-1\} && \text{(1b)} \\
& \alpha(\mathrm{x}_t) = k^* && \forall t \in \{0, \ldots, T-1\} && \text{(1c)} \\
& \mathrm{x}_0 = \mathrm{x}_0 && && \text{(1d)} \\
& \mathrm{x}_T = \mathrm{x}_f && && \text{(1e)}
\end{aligned}
$$

Mode remaining navigation problem

- Dynamics and mode switching are *unknown a priori*

Mode remaining navigation problem

- Dynamics and mode switching are *unknown a priori*
- So we cannot guarantee mode remaining behaviour...

$$\alpha(\mathrm{x}_t) = k^* \qquad \forall t \in \{0, \ldots, T-1\} \tag{2}$$

Mode remaining navigation problem

- Dynamics and mode switching are *unknown a priori*
- So we cannot guarantee mode remaining behaviour...

$$\alpha(\mathbf{x}_t) = k^* \qquad \forall t \in \{0, \ldots, T-1\} \tag{2}$$

- Solve using model-based reinforcement learning

Mode remaining navigation problem

- Dynamics and mode switching are *unknown a priori*
- So we cannot guarantee mode remaining behaviour...

$$\alpha(\mathbf{x}_t) = k^* \qquad \forall t \in \{0, \dots, T-1\} \tag{2}$$

- Solve using model-based reinforcement learning
  - jointly learn dynamics modes and switching mechanism

Mode remaining navigation problem

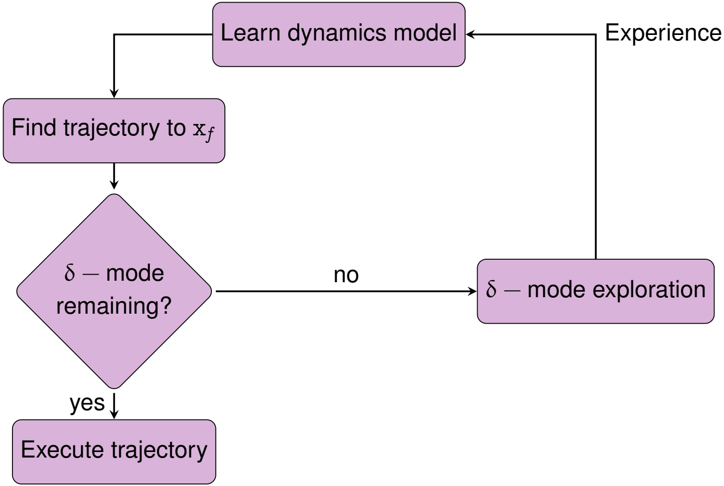
- Dynamics and mode switching are *unknown a priori*
- So we cannot guarantee mode remaining behaviour...

$$\alpha(\mathbf{x}_t) = k^* \qquad \forall t \in \{0, \dots, T-1\} \tag{2}$$

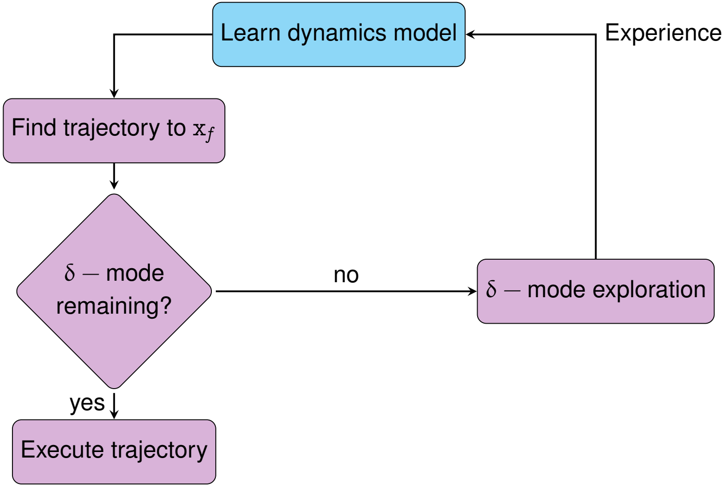
- Solve using model-based reinforcement learning
  - jointly learn dynamics modes and switching mechanism
  - relax mode remaining requirement

$$\Pr(\forall t \in \{0, \dots, T\} : \alpha(\mathbf{x}_t) = k^*, \mathbf{u}_t \in \mathcal{U}) \geqslant 1 - \delta \tag{3}$$

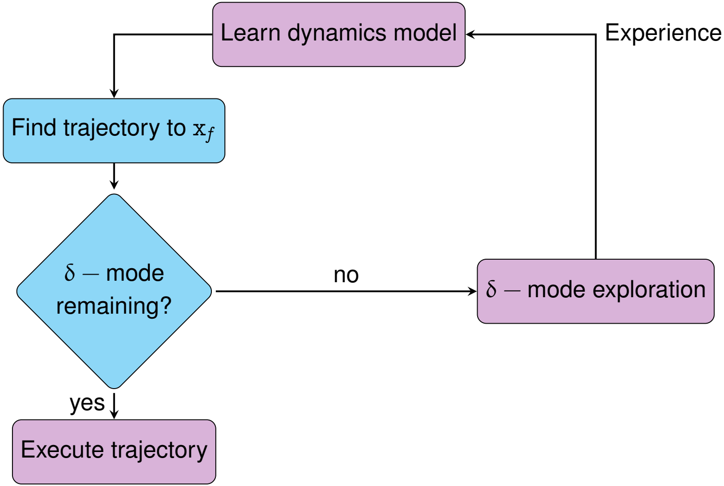
Mode remaining model-based RL?



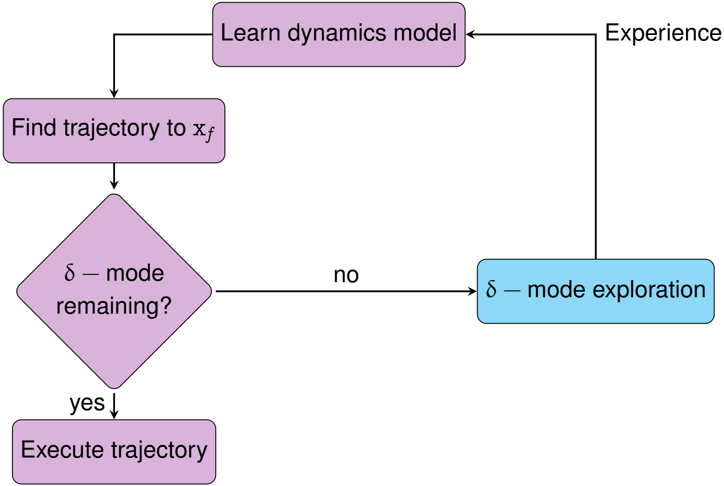
Contributions - Chapter 3



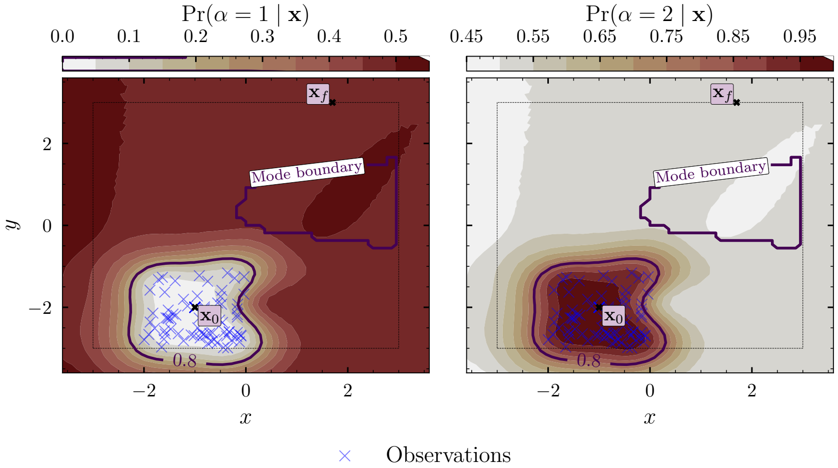
Contributions - Chapter 4



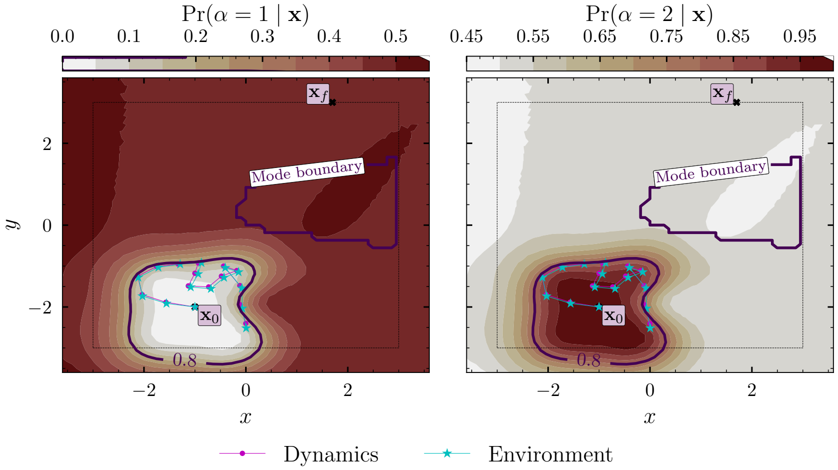
Contributions - Chapter 6



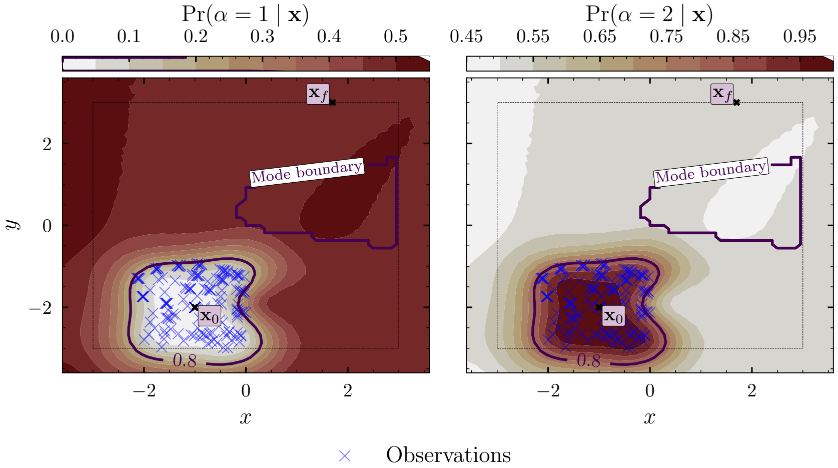
ModeOpt iteration 0



ModeOpt iteration 0

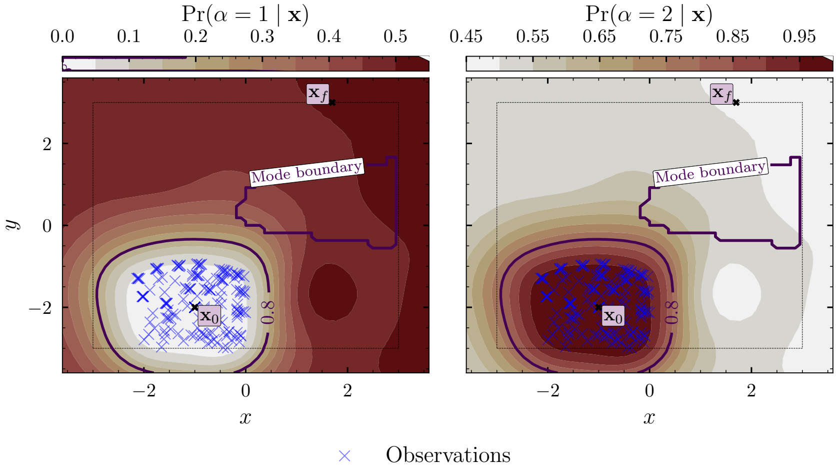


ModeOpt iteration 0

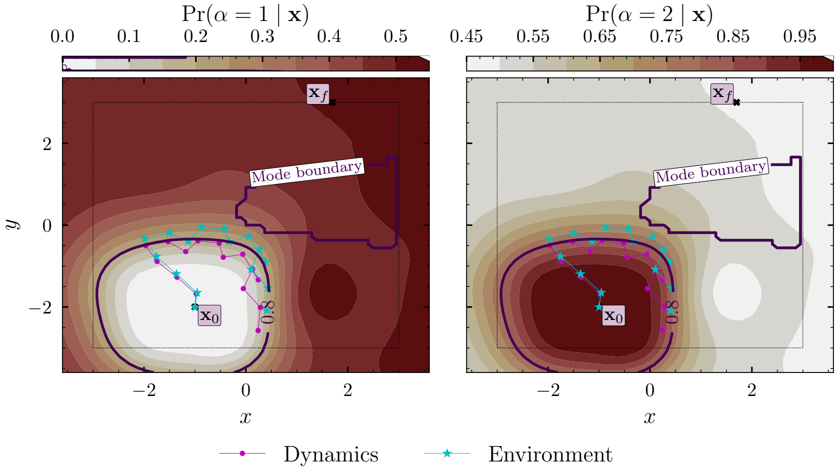


bristol.ac.uk

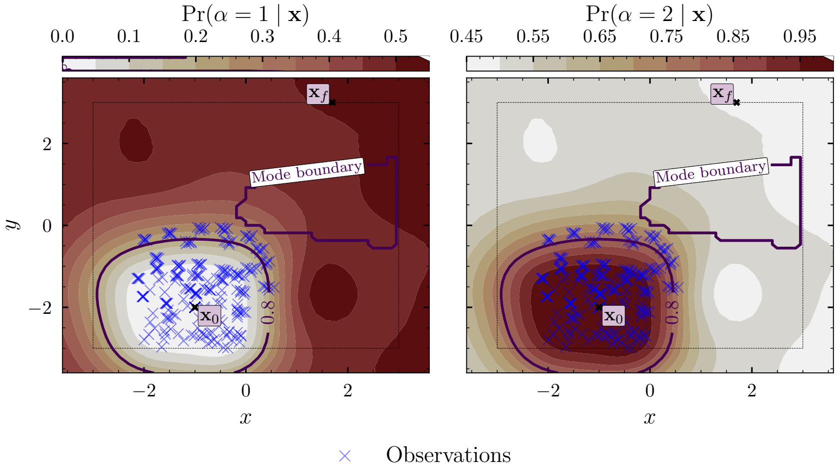
ModeOpt iteration 1



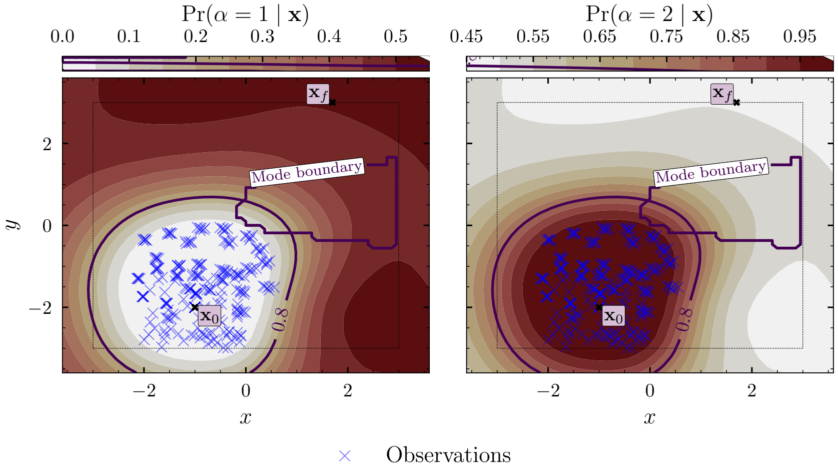
ModeOpt iteration 1



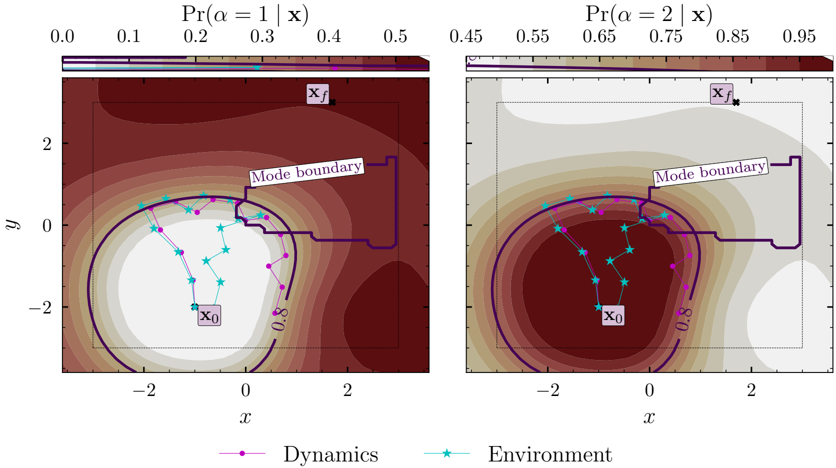
ModeOpt iteration 1



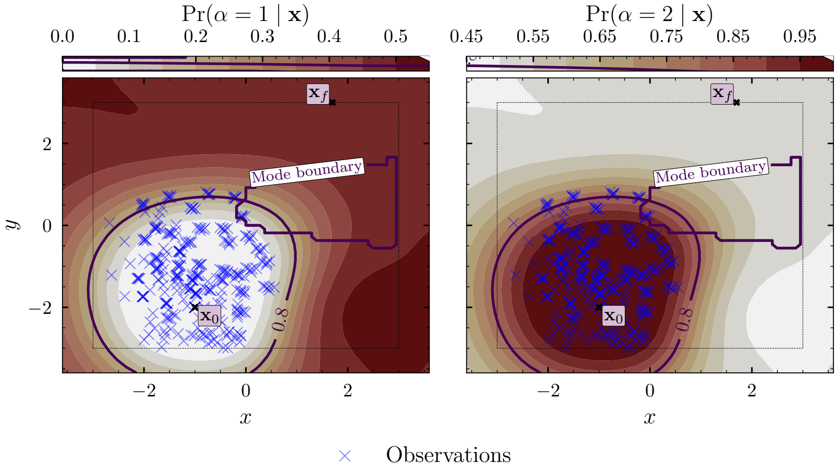
ModeOpt iteration 2



ModeOpt iteration 2

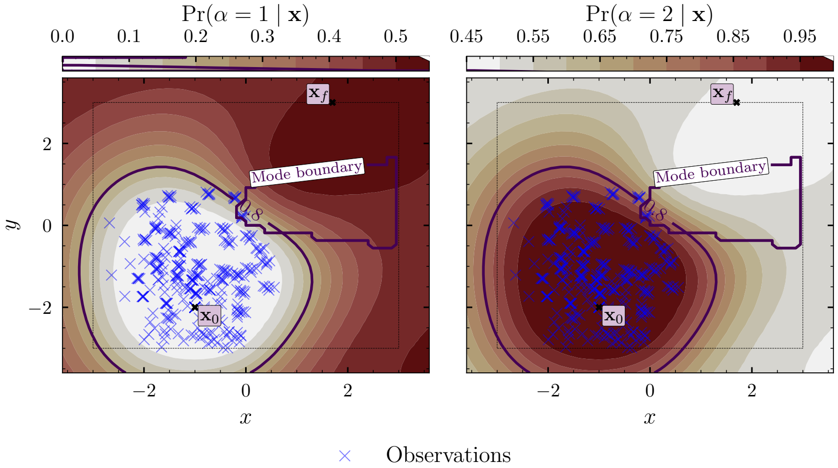


ModeOpt iteration 2

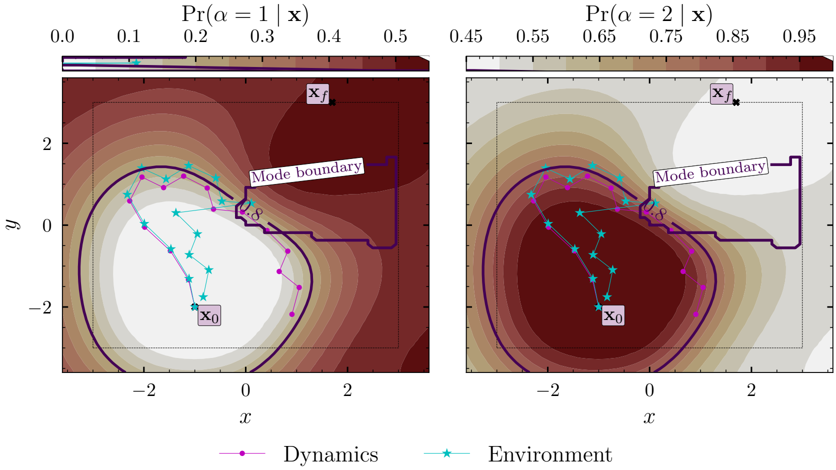


bristol.ac.uk

ModeOpt iteration 3

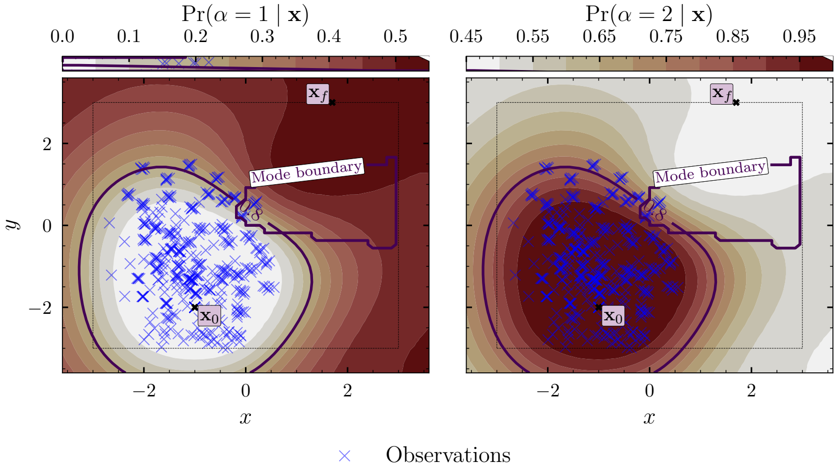


ModeOpt iteration 3



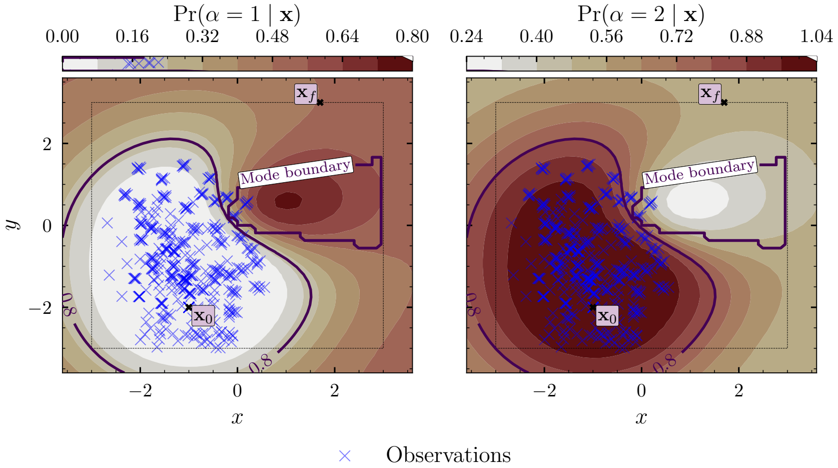
bristol.ac.uk

ModeOpt iteration 3



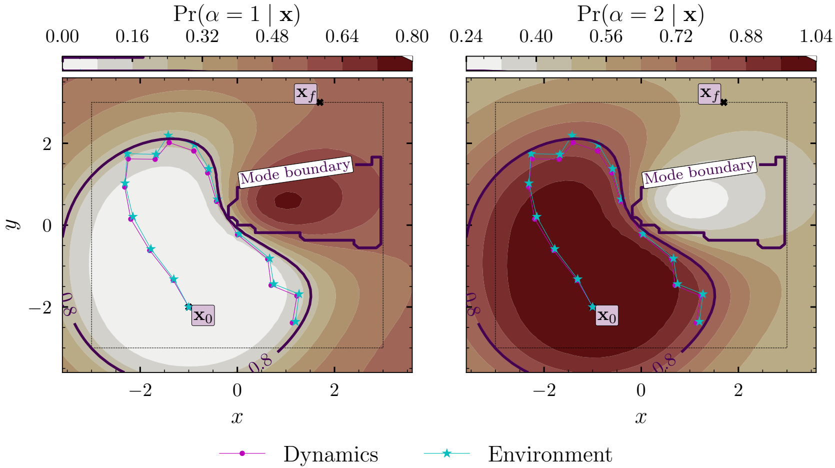
bristol.ac.uk

ModeOpt iteration 4

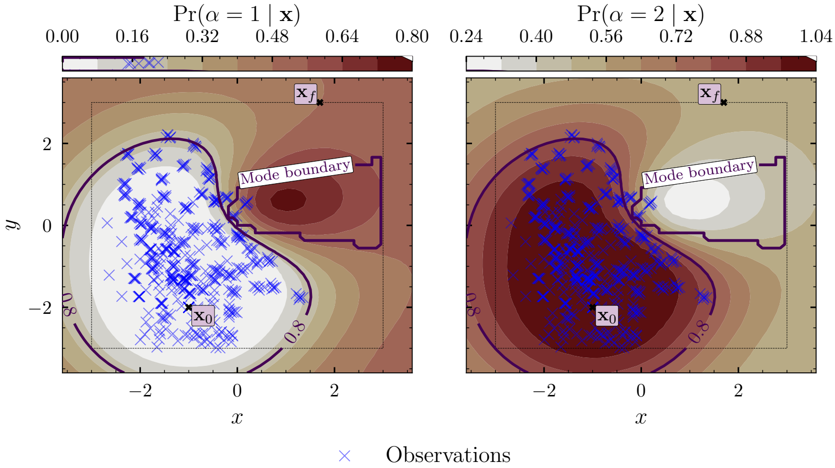


bristol.ac.uk

ModeOpt iteration 4

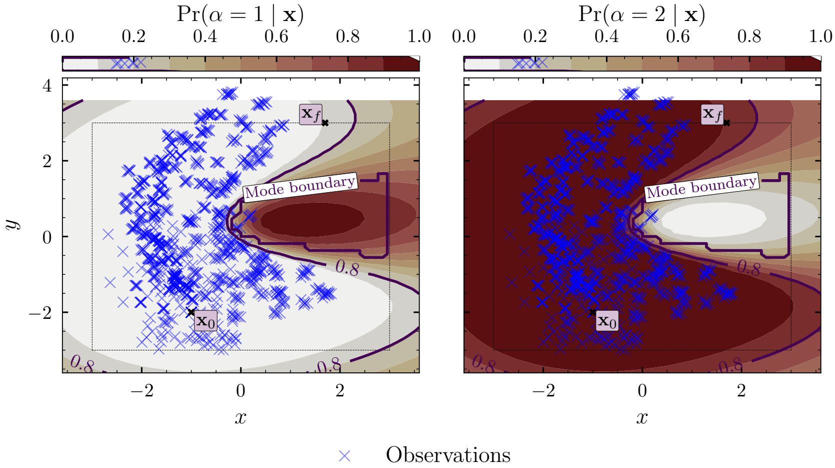


ModeOpt iteration 4



And so on, until

And so on, until



Thanks for listening

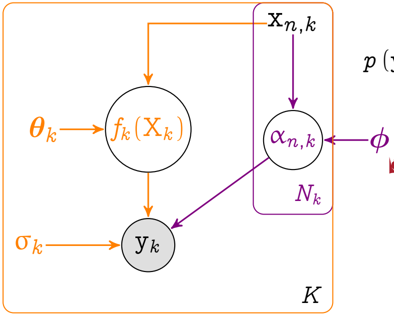
Questions?

Model learning - Mixture models?

MoE marginal likelihood

$$p\left(\mathrm{y} \mid \mathrm{X}, \boldsymbol{\theta}, \boldsymbol{\phi}\right) = \prod_{n=1}^{N} \sum_{k=1}^{K} \underbrace{\Pr\left(\alpha_n = k \mid \mathrm{x}_n, \boldsymbol{\phi}\right)}_{\text{gating network}} \underbrace{p\left(y_n \mid \alpha_n = k, \mathrm{x}_n, \boldsymbol{\theta}_k\right)}_{\text{expert } k}, \tag{4}$$

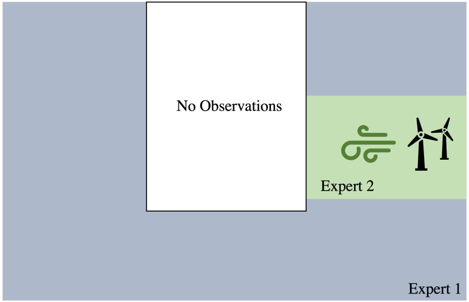
Model learning - Mixtures of Nonparametric Experts

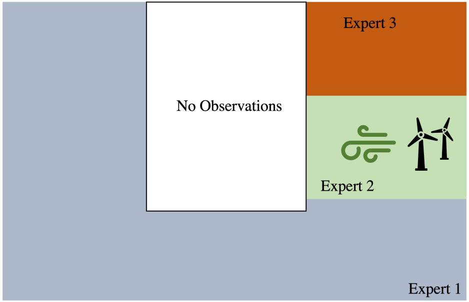


$$p(\mathrm{y} \mid \mathrm{X}, \boldsymbol{\theta}, \phi) = \sum_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha} \mid \mathrm{X}, \phi) \left[\prod_{k=1}^{K} \underbrace{p(\{y_n : \alpha_n = k\} \mid \{\mathrm{x}_n : \alpha_n = k\}, \boldsymbol{\theta}_k)}_{\text{expert } k} \right]$$

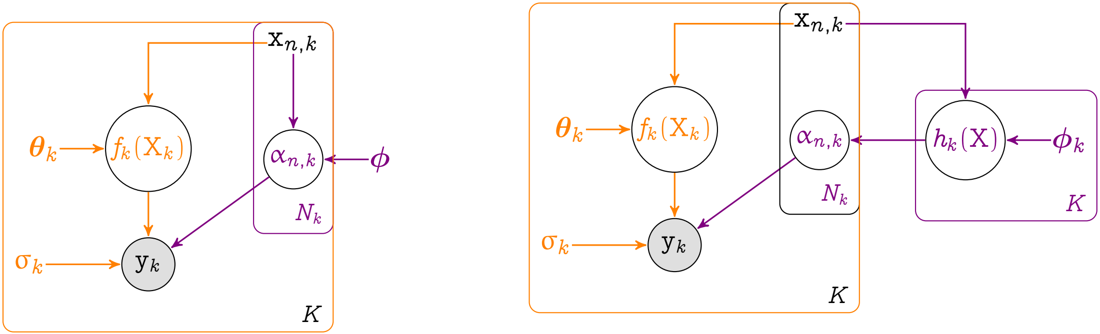
- Sum over exponentially many (K^N) sets of assignments
  - $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_N\}$

Model learning - Identifiability





Model learning - Mixtures of Nonparametric Experts



Model learning - Identifiable Mixtures of Sparse Variational Gaussian Process Experts

$$p(\mathrm{y} \mid \mathrm{X}, \boldsymbol{\theta}, \phi) = \sum_{\boldsymbol{\alpha}} \underbrace{\mathbb{E}_{p(\mathbf{h}(\mathrm{X}))} \left[\prod_{n=1}^{N} P\left(\alpha_n \mid \mathbf{h}(\mathrm{x}_n)\right) \right]}_{\text{GP gating network}} \left[\prod_{k=1}^{K} \underbrace{p\left(\{y_n : \alpha_n = k\} \mid \{\mathrm{x}_n : \alpha_n = k\}, \boldsymbol{\theta}_k\right)}_{\text{expert } k} \right] \tag{5}$$

Model learning - Parameterise the nonparametric model?

- Like a sparse GP parameterises a GP...

Model learning - Parameterise the nonparametric model?

- Like a sparse GP parameterises a GP…
- GP prior where $\mathrm{X}_k = \{\mathrm{x}_n : \alpha_t = k\}$

$$f_k(\mathrm{X}_k) \sim \mathcal{N}\left(\mu_k(\mathrm{X}_k), k_k(\mathrm{X}_k, \mathrm{X}_k)\right) \tag{6}$$

Model learning - Parameterise the nonparametric model?

- Like a sparse GP parameterises a GP...
- GP prior where $\mathrm{X}_k = \{\mathrm{x}_n : \alpha_t = k\}$

$$f_k(\mathrm{X}_k) \sim \mathcal{N}\left(\mu_k(\mathrm{X}_k), k_k(\mathrm{X}_k, \mathrm{X}_k)\right) \tag{6}$$

- Augment with inducing points

$$f_k(\zeta_k) \sim \mathcal{N}\left(\mu_k(\zeta_k), k_k(\zeta_k, \zeta_k)\right) \tag{7}$$

Model learning - Parameterise the nonparametric model?

- Like a sparse GP parameterises a GP...
- GP prior where $\mathrm{X}_k = \{\mathrm{x}_n : \alpha_t = k\}$

$$f_k(\mathrm{X}_k) \sim \mathcal{N}(\mu_k(\mathrm{X}_k), k_k(\mathrm{X}_k, \mathrm{X}_k)) \tag{6}$$

- Augment with inducing points

$$f_k(\zeta_k) \sim \mathcal{N}(\mu_k(\zeta_k), k_k(\zeta_k, \zeta_k)) \tag{7}$$

- Approximate marginal likelihood

$$p(\mathrm{y} \mid \mathrm{X}) \approx \mathbb{E}_{p(\mathrm{h}(\xi))p(\mathrm{f}(\zeta))}\left[\prod_{n=1}^{N}\sum_{k=1}^{K}\Pr(\alpha_n = k \mid \mathrm{h}(\xi))p(y_n \mid f_k(\zeta_k))\right] \tag{8}$$

Model learning - Parameterise the nonparametric model?