The 37th Conference on Neural Information Processing Systems
New Orleans, United States
Dec. 10th – Dec 16th

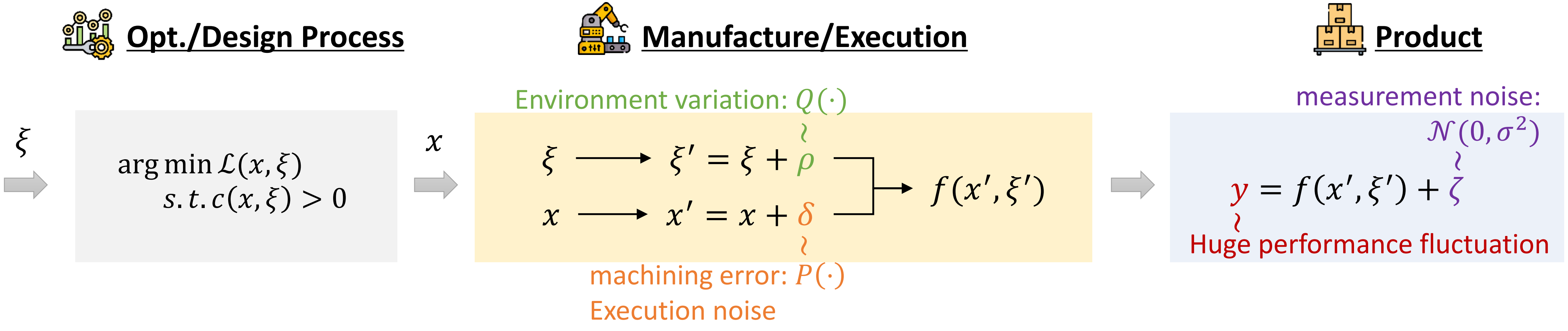
AIRBO

Efficient Robust Bayesian Optimization for Arbitrary Uncertain Inputs

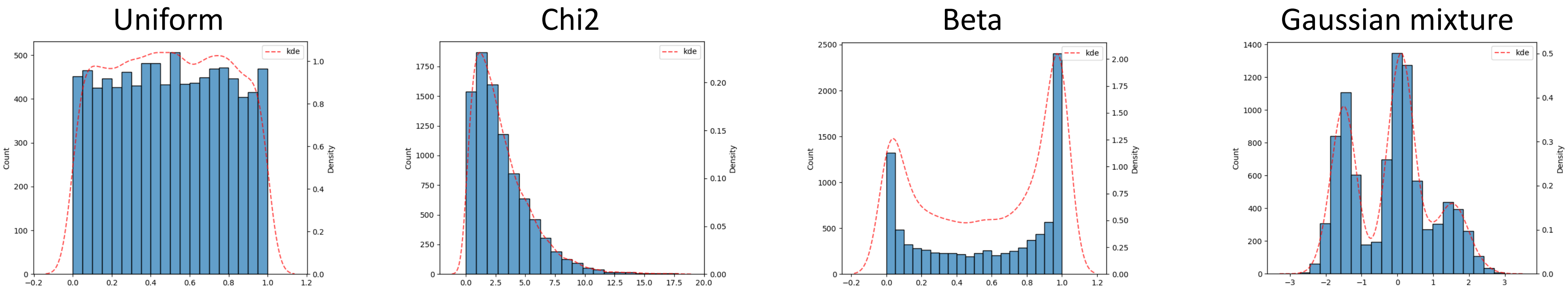
Lin Yang, Junlong Lyu, Wenlong Lyu, Zhitang Chen

Noah's Ark Lab, Huawei

Problem

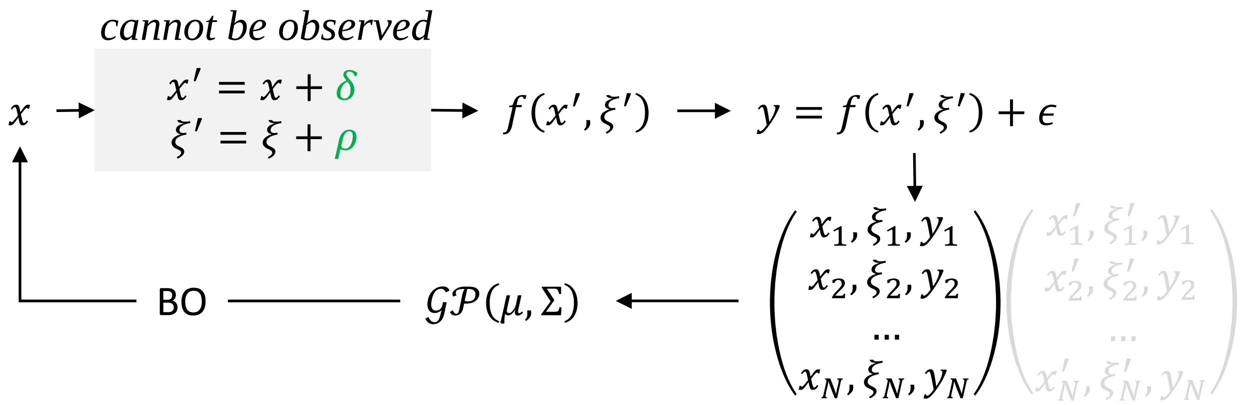


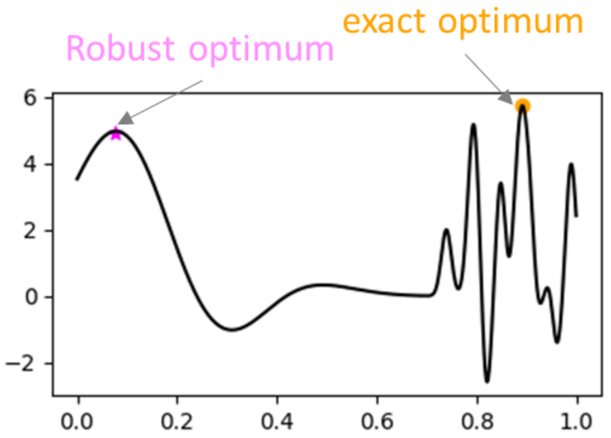
Moreover, depending on the source of randomness, the input uncertainty can be quite complex…



Formulation

➢Robust BO





Objective: Robust optimum

$$\arg\min_{x} \int\int f(x + \delta, \xi + \rho)\, d_\rho\, d_\delta$$

$$s.t.\ C(x') \leq 0$$

➢In this work,

- The input uncertainty can follow arbitrary complex distribution.
- Assume that we can samples from input distribution, which can be done via statistical learning.

Intuition

➢Weight interpretation of $\mathcal{GP}$ [1]

- Starts from Bayesian linear model: $y = x^T w + \zeta, \quad \zeta \sim \mathcal{N}(0, \sigma^2)$

$w \sim \mathcal{N}(0, \Sigma_p)$

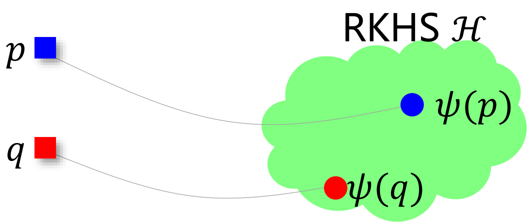
- Posterior:

$$f_* | x_*, X, y \sim \mathcal{N}\left(\phi^T(x_*)\Sigma_p\phi(X)(A + \sigma_n^2 I)^{-1} y,\ \phi^T(x_*)\Sigma_p\phi(x_*) - \phi^T(x_*)\Sigma_p\phi(X)(A + \sigma_n^2 I)^{-1}\phi^T(X)\Sigma_p\phi(x_*)\right)$$

- Apply Kernel to project into feature space

$k(x, x') = \phi^T(x)\Sigma_p\phi(x') = \psi(x) \cdot \psi(x')$

- GP posterior: $f_* | x_*, X, y \sim \mathcal{N}\left(K(X_*, X)(K(X, X) + \sigma_n^2 I)^{-1} y,\ K(X_*, X_*)(K(X, X) + \sigma_n^2 I)^{-1} K(X, X_*)\right)$



The core steps of GP are:

1) project the input x to a high-dim. feature embedding $\psi(x)$

2) compare them in the RKHS defined by the kernel.

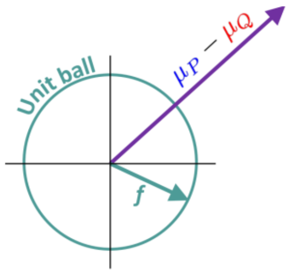
Considering the input uncertainty, how to compare the uncertain inputs?

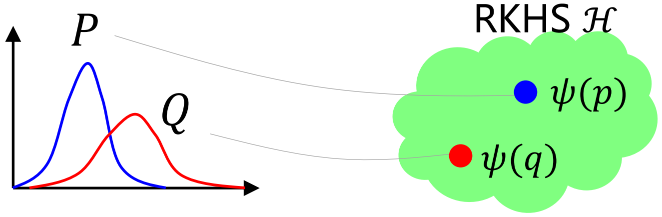
[1] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA, 2006

MMD-based Kernel for Arbitrary Uncertain Inputs

- In general, the Integral Probabilistic Metric (IPM) serves our purpose.
- MMD ⇔ Measuring distance btw prob. distributions in RKHS[*]

$$
\begin{aligned}
\mathrm{MMD}(P,Q) &= \sup_{||g||\le 1} [\mathbb{E}_{X\sim P} g(X) - \mathbb{E}_{Y\sim Q} g(Y)] \\
&= \sup_{||g||\le 1} [\langle g, \mathbb{E}_P \psi(X)\rangle_{\mathcal{G}} - \langle g, \mathbb{E}_Q \psi(Y)\rangle_{\mathcal{G}}] \\
&= \sup_{||g||\le 1} [\langle g, \mu_P\rangle_{\mathcal{G}} - \langle g, \mu_Q\rangle_{\mathcal{G}}] \\
&= \sup_{||g||\le 1} \langle g, \mu_P - \mu_Q\rangle_{\mathcal{G}} \\
&= \langle g^*, \mu_P - \mu_Q\rangle_{\mathcal{G}} \\
&= ||\mu_P - \mu_Q|| \ (\text{the supremum is achieved when } g^* = \frac{\mu_P - \mu_Q}{||\mu_P - \mu_Q||})
\end{aligned}
$$





- MMD-based kernel to propagate input uncertainty to posterior
  - MMD kernel: $\hat{k}(P,Q) = \exp\big(\alpha MMD^2(P,Q,k)\big)$
  - For the MMD estimation, we employ a compositional rational quadratic kernel:

$$k(x,x') = \sum_{\alpha_i \in \mathcal{S}} \left(1 + \frac{(x-x')^2}{2\alpha_i l_i^2}\right)^{-\alpha_i}, \mathcal{S} = \{0.2, 0.5, 1,2,5\}$$

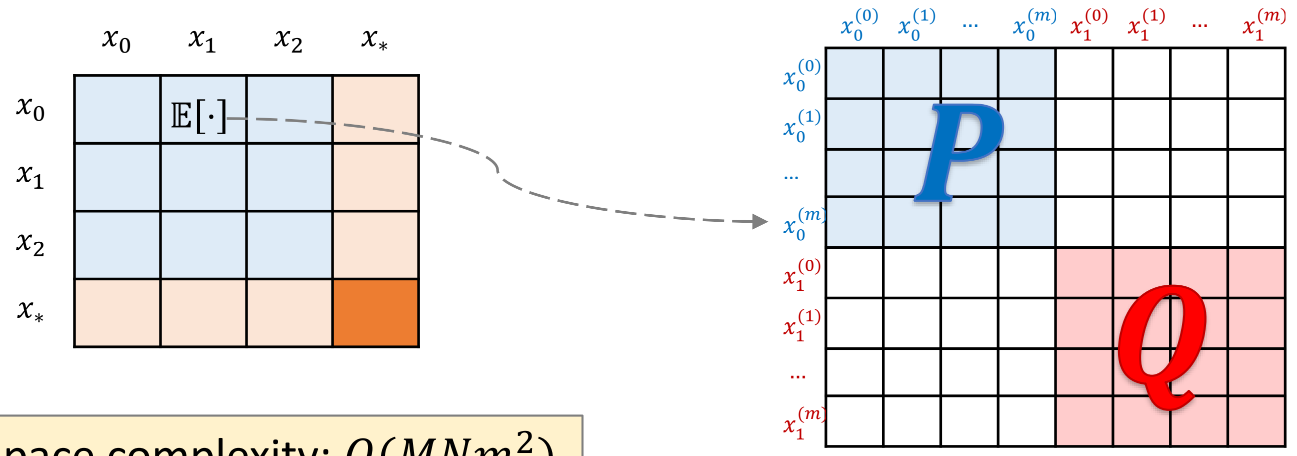
[*] Arthur Gretton, Dougal Sutherland, and Wittawat Jitkrittum. "Interpretable comparison of distributions and models". In NeurIPS *[Tutorial]* (2019)

High Estimation Complexity of MMD

- The empirical Estimation of MMD requires further sampling m samples from the input uncertainty[*]:

$$MMD^2(P,Q) = \mathbb{E}_{u,u' \sim P\otimes P}[k(u,u')] + \mathbb{E}_{v,v' \sim Q\otimes Q}[k(v,v')] - 2\mathbb{E}_{u,v \sim P\otimes Q}[k(u,v)]$$

- This consumes a huge GPU memory and hinders its ability of parallel computation:



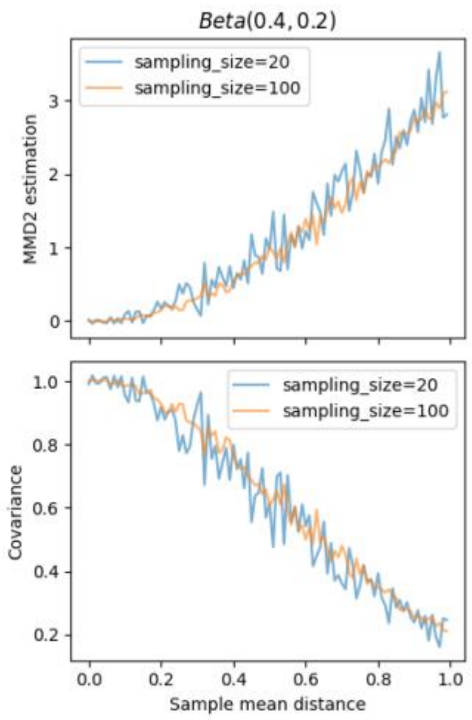
Space complexity: $O(MNm^2)$

M: #training samples
N: #testing samples
m: #sampling size

[*] Note here we only need samples from the input distribution, but not their target function values

Stable MMD Estimation vs. Inference Complexity

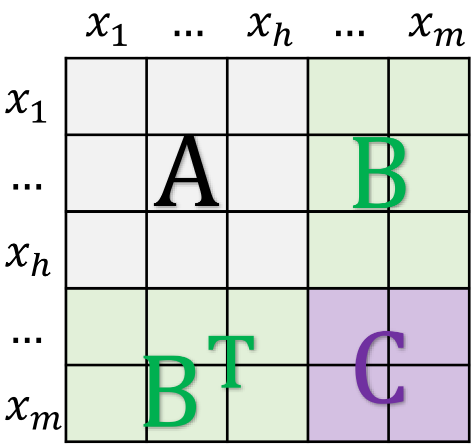
- Insufficient sampling results in a highly-varied posterior.
- A larger sample size can occupy significant GPU memory and reduce the ability of parallel computing.



HUAWEI

Accelerating Posterior Inference via Nyström Approximation

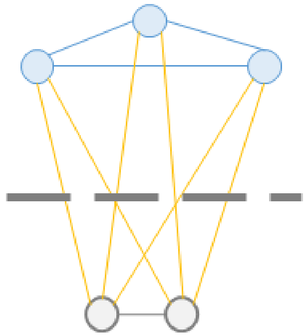
➢Nyström MMD estimator for efficient posterior inference



$C = B^T A^{-1} B$

$$\tilde{\mathrm{MMD}}^2(P,Q) = \mathbb{E}_{X,X'\sim P\otimes P}[k(X,X')] + \mathbb{E}_{Y,Y'\sim Q\otimes Q}[k(Y,Y')] - 2\mathbb{E}_{X,Y\sim P\otimes Q}[k(X,Y)]$$
$$\approx \frac{1}{m^2}\mathbf{1}_m^T U \mathbf{1}_m + \frac{1}{m^2}\mathbf{1}_m^T V \mathbf{1}_m - \frac{2}{m^2}\mathbf{1}_m^T W \mathbf{1}_m$$
$$\approx \frac{1}{m^2}\mathbf{1}_m^T U_{mh} U_h^+ U_{mh}^T \mathbf{1}_n + \frac{1}{m^2}\mathbf{1}_m^T V_{mh} V_h^+ V_{mh}^T \mathbf{1}_m - \frac{2}{m^2}\mathbf{1}_m^T W_{mh} W_h^+ W_{mh}^T \mathbf{1}_m,$$

where $U = K(X,X')$, $V = K(Y,Y')$, $W = K(X,Y)$ are the kernel matrices, $\mathbf{1}_m$ represents a m-by-1 vector of ones, m defines the sampling size and h controls the sub-sampling size.



Empirical estimator
Space complexity: $O(MNm^2)$

⇨

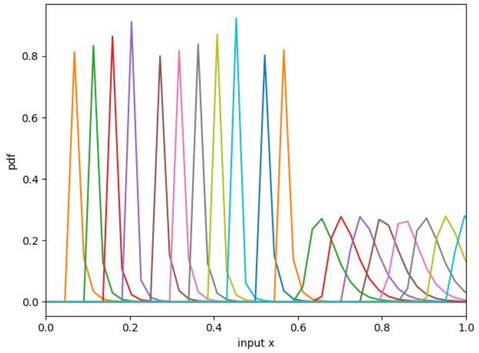
Nystrom estimator
Space complexity: $O(MNmh)$

$h \ll m$
For 1D Gaussian,
$m = 1000, h = 10$

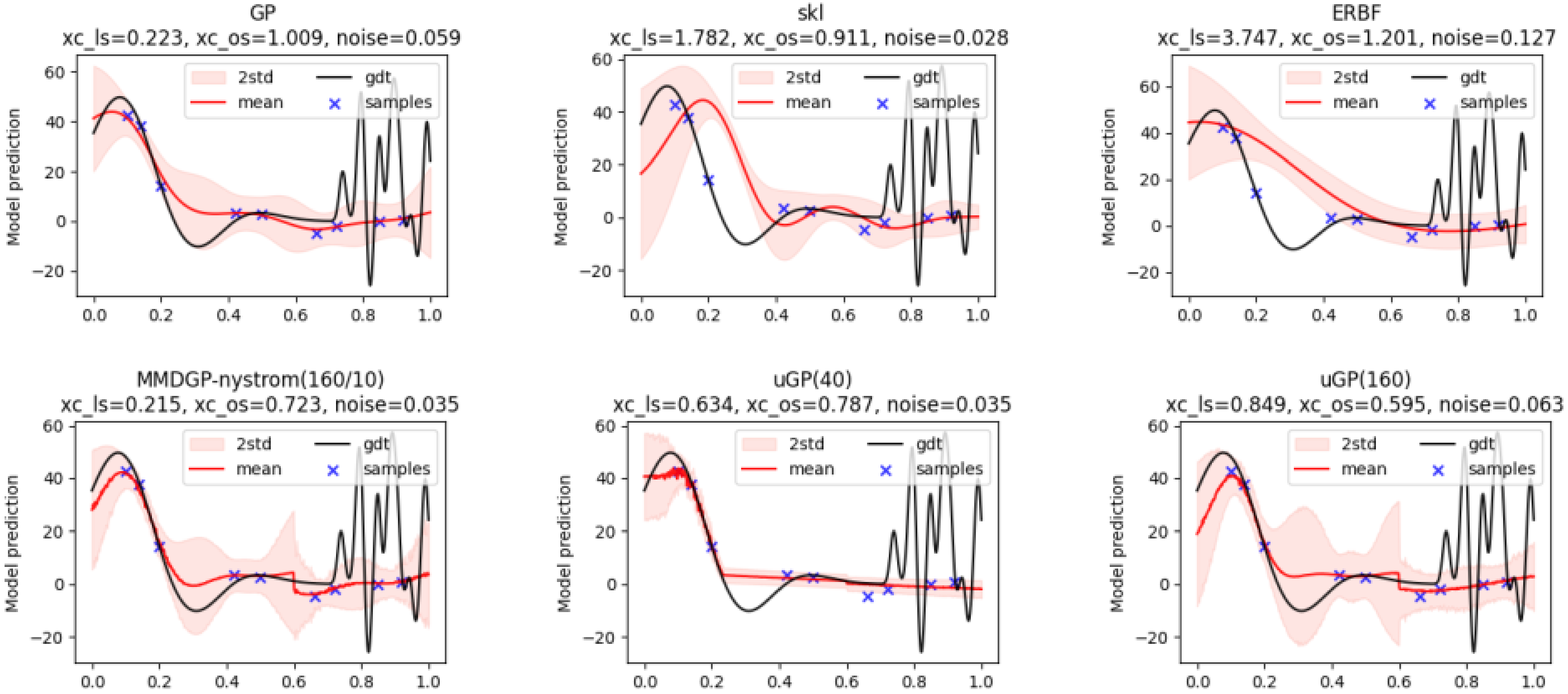
M: #training samples
N: #testing samples
m: #sampling size
h : #sub-sampling size

Evaluation: Modeling Complex Input Uncertainty

➢ Step-changing χ^2 distribution:



- $P_x = \chi^2(k = g(x), \sigma = 0.01), g(x) = \begin{cases} 0.5, & x \in [0.0, 0.6) \\ 7.0, & x \in [0.6, 1.0] \end{cases}$



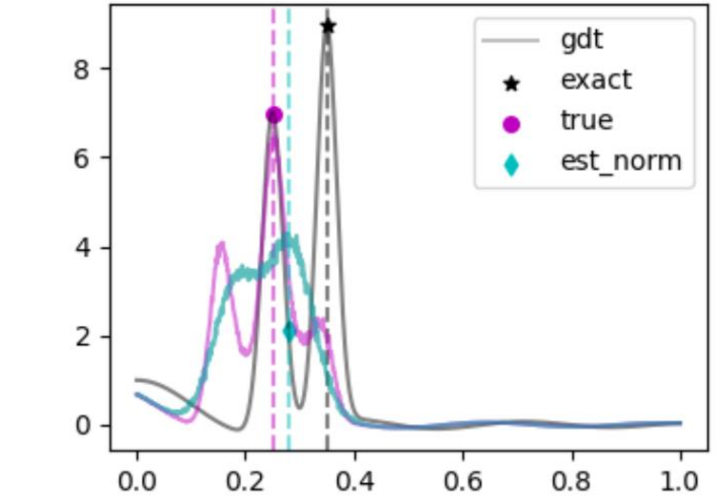
HUAWEI

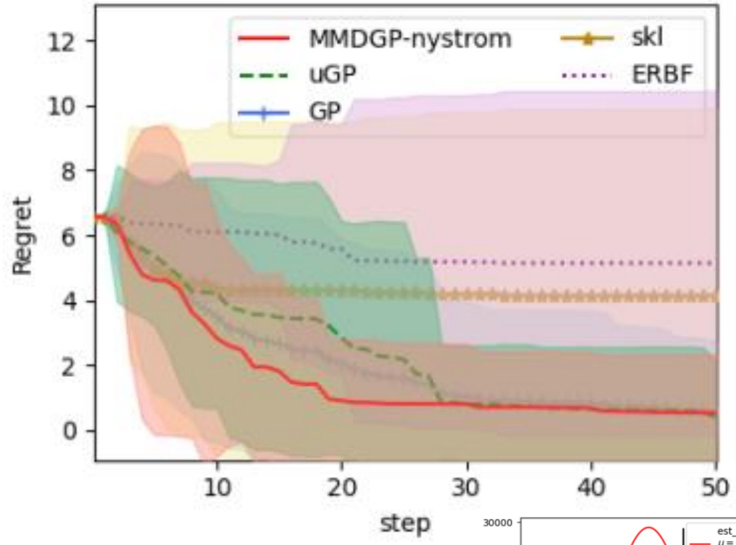
Evaluation: Posterior Inference

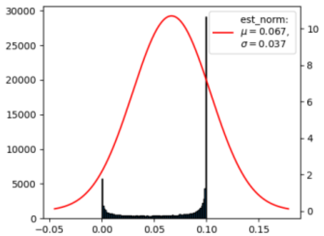
Table 1: Performance of Posterior inference for 512 samples.

Method	Sampling Size	Sub-sampling Size	**Inference Time (seconds)**	**Batch Size (samples)**
Empirical	20	-	1.143 ± 0.083	512
Empirical	100	-	8.117 ± 0.040	128
Empirical	1000	-	840.715 ± 2.182	1
Nystrom	100	10	0.780 ± 0.001	512
Nystrom	1000	100	21.473 ± 0.984	128

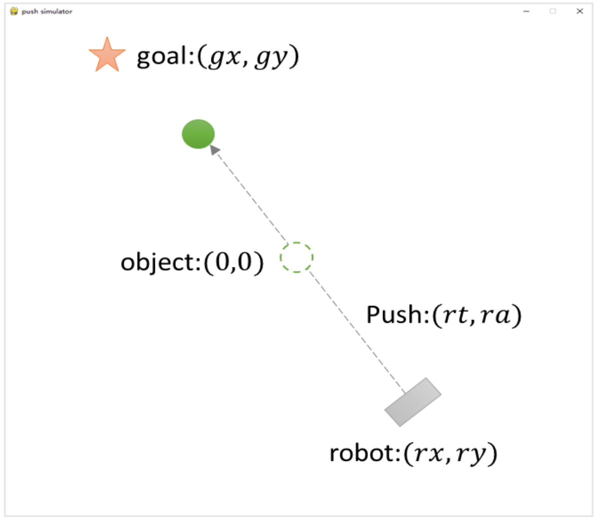
Evaluation: Robust Optimization with Non-Gaussian Inputs

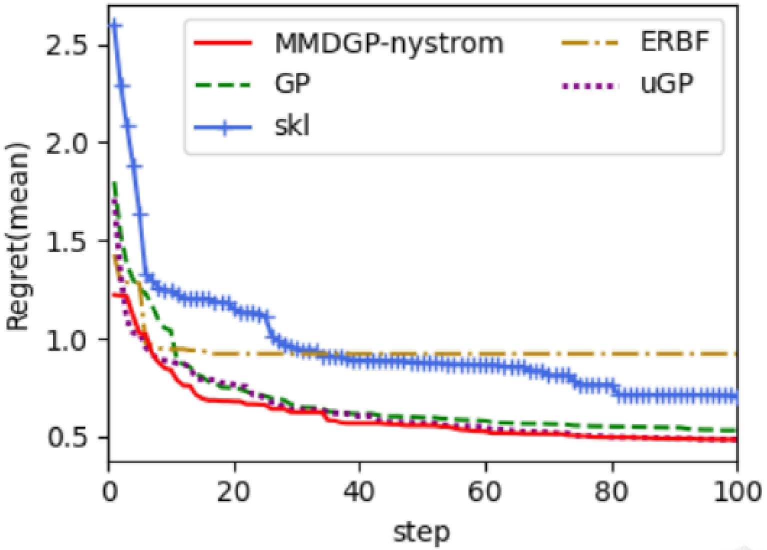




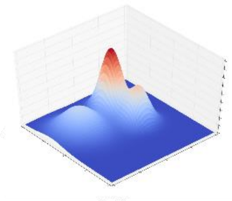


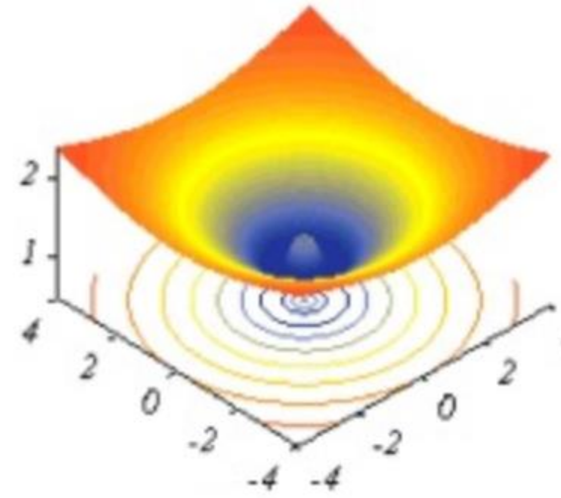
1D double-peak function
w/ ***Beta*** input distribution

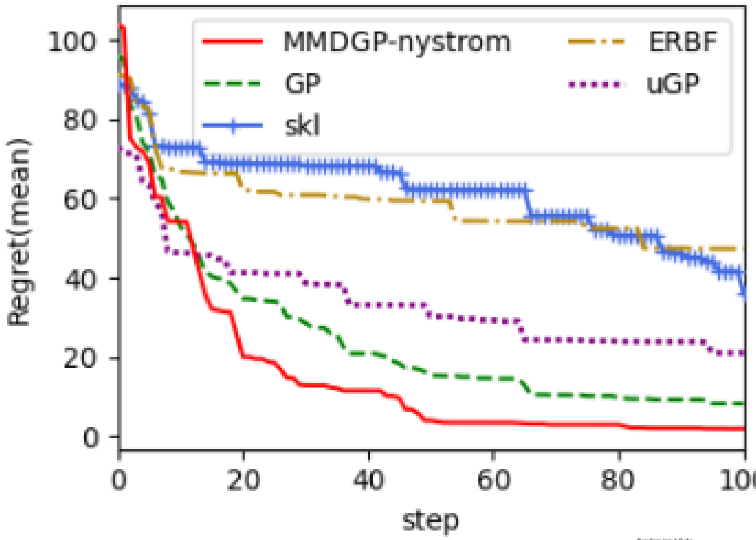




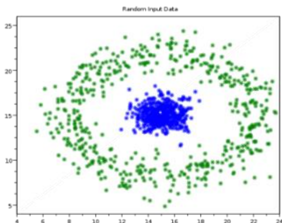
3D Robot pushing
w/ ***Gaussian Mixture*** distribution







10D bumped-bowl
with ***circular*** distribution



Thank you for listening!

Poster session: Great Hall & Hall B1+B2 #1225

Contact: yanglin_jason@qq.com Noah's Ark Lab, Huawei