REGRET BOUNDS FOR META BAYESIAN OPTIMIZATION WITH AN UNKNOWN GAUSSIAN PROCESS PRIOR

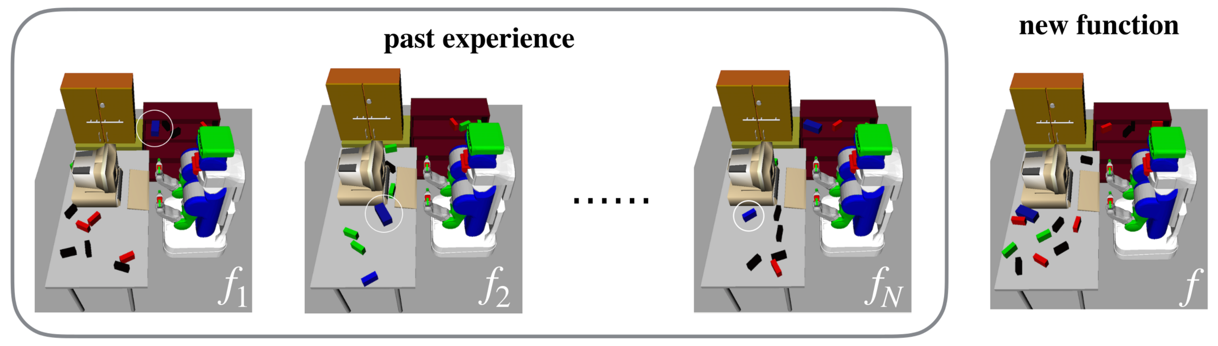
ZI WANG*, BEOMJOON KIM*, LESLIE PACK KAELBLING   ziw, beomjoon, lpk@mit.edu   (*equal contribution)

MAIN CONTRIBUTIONS

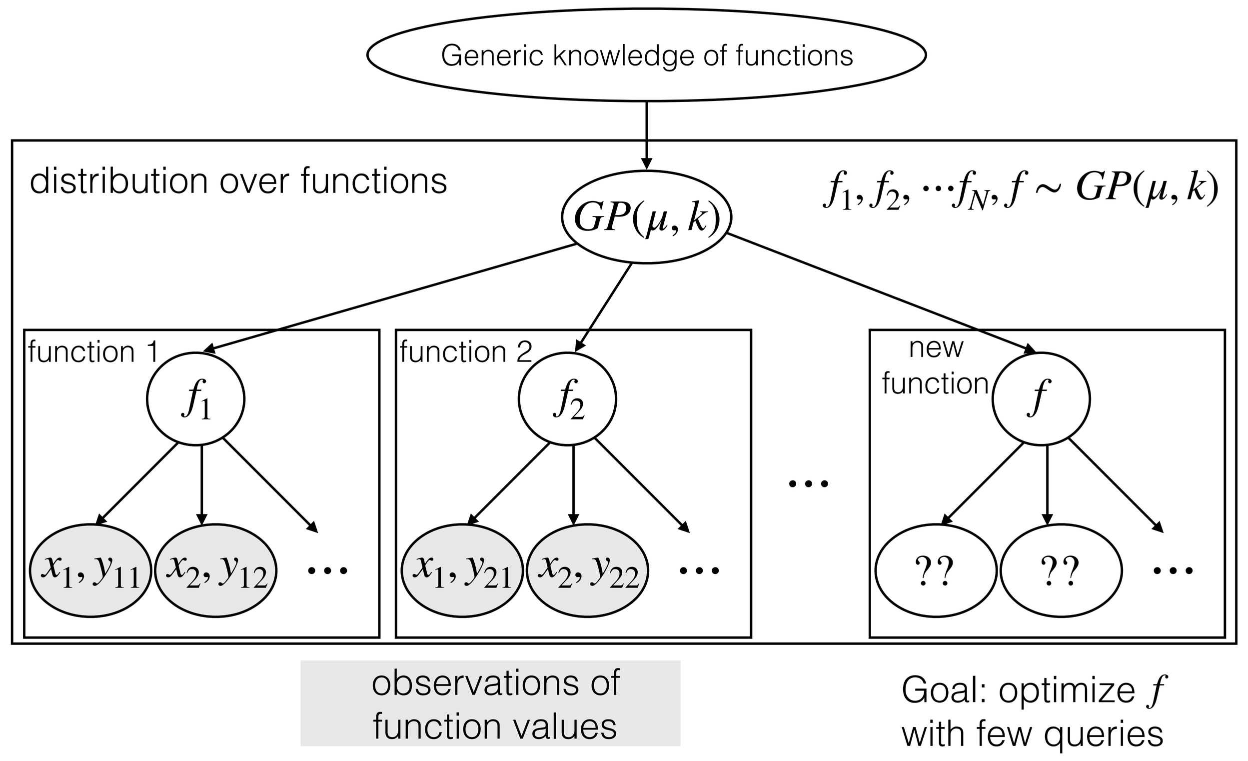
- A stand-alone Bayesian optimization module that takes in only a multi-task training data set as input and then actively selects inputs to efficiently optimize a new function.
- Constructive analyses of the regret of this module.

BAYESIAN OPTIMIZATION

- Maximize an expensive blackbox function $f : \mathfrak{X} \to \mathbb{R}$ with sequential queries $x_1, \cdots, x_T$ and noisy observations of their values $y_1, \cdots, y_T$.
- Assume a Gaussian process prior $f \sim GP(\mu, k)$.
- Use acquisition functions as the decision criterion for where to query.
- Major problem: the prior is unknown. Existing approaches:
  - maximum likelihood;
  - hierarchical Bayes.
- Current theoretical results break down if the prior is not given.
- What if we have experience with similar functions? Ex. Optimizing robot grasps in different environments:



OUR MODEL



Our evaluation criteria:

- best-sample simple regret $r_T = \max_{x \in \mathfrak{X}} f(x) - \max_{t \in [T]} f(x_t)$;
- simple regret $R_T = \max_{x \in \mathfrak{X}} f(x) - f(x_{t^*})$, $t^* = \arg\max_{t \in [T]} y_t$.

GAUSSIAN PROCESSES

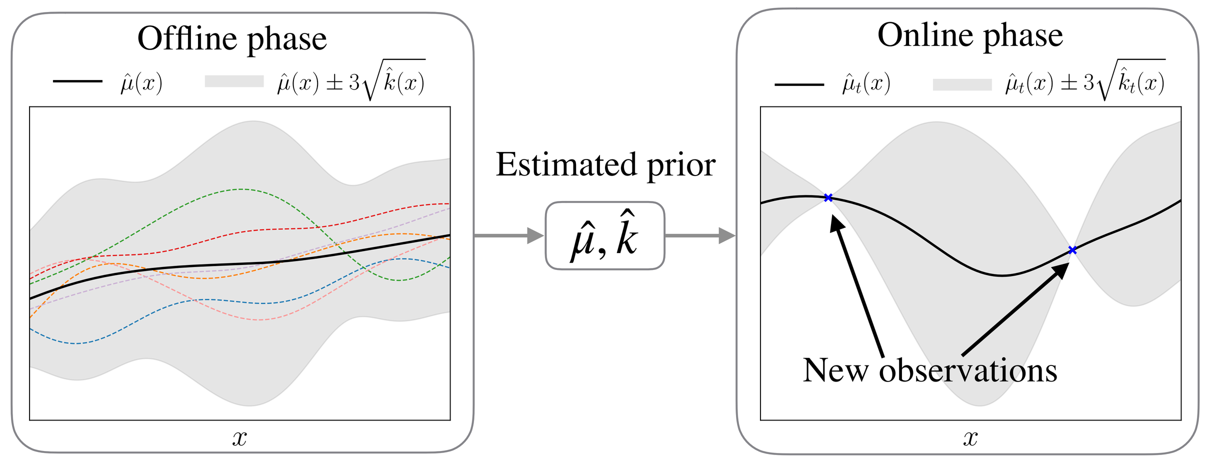
Given observations $D_t = \{(x_\tau, y_\tau)\}_{\tau=1}^t$, $y_\tau \sim \mathcal{N}(f(x_\tau), \sigma^2)$, the posterior $GP(\mu_t, k_t)$ satisfies

$$\mu_t(x) = \mu(x) + k(x, \boldsymbol{x}_t)(k(\boldsymbol{x}_t) + \sigma^2 \boldsymbol{I})^{-1}(\boldsymbol{y}_t - \mu(\boldsymbol{x}_t)),$$

$$k_t(x, x') = k(x, x') - k(x, \boldsymbol{x}_t)(k(\boldsymbol{x}_t) + \sigma^2 \boldsymbol{I})^{-1} k(\boldsymbol{x}_t, x'),$$

where $\boldsymbol{y}_t = [y_\tau]_{\tau=1}^T$ and $\boldsymbol{x}_t = [x_\tau]_{\tau=1}^T$, $\mu(\boldsymbol{x}) = [\mu(x_i)]_{i=1}^n$, $k(\boldsymbol{x}, \boldsymbol{x}') = [k(x_i, x_j')]_{i \in [n], j \in [n']}$, $k(\boldsymbol{x}) = k(\boldsymbol{x}, \boldsymbol{x})$.

MAIN IDEA: USE THE PAST EXPERIENCE TO ESTIMATE THE PRIOR OF f



- Offline phase:
  - collect meta training data: M evaluations from each of the N functions sampled from the same prior, $\bar{D}_N = \{[(\bar{x}_j, \bar{y}_{ij})]_{j=1}^M\}_{i=1}^N$, $\bar{y}_{ij} \sim \mathcal{N}(f_i(\bar{x}_j), \sigma^2)$, $f_i \sim GP(\mu, k)$;
  - estimate the prior mean function $\hat{\mu}$ and kernel $\hat{k}$ from the meta training data.
- Online phase: estimate the posterior mean $\hat{\mu}_t$ and covariance $\hat{k}_t$ and use them for BO on a new function $f \sim GP(\mu, k)$ with a total of T iterations.

ESTIMATE THE PRIOR $GP(\hat{\mu}, \hat{k})$ AND POSTERIOR $GP(\hat{\mu}_t, \hat{k}_t)$: DISCRETE AND CONTINUOUS CASES

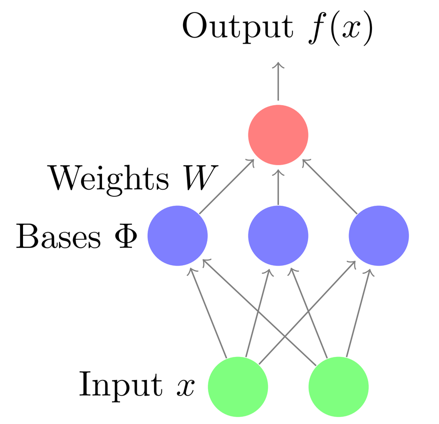
$\mathfrak{X}$ is finite: directly estimate the prior and the posterior [1]

Define the observation matrix as $Y = [Y_i]_{i \in [N]} = [\bar{y}_{ij}]_{i \in [N], j \in [M]}$. Missing entries? Use matrix completion [2].

$$\hat{\mu}(\mathfrak{X}) = \frac{1}{N} Y^{\mathrm{T}} 1_N \sim \mathcal{N}(\mu(\mathfrak{X}), \frac{1}{N}(k(\mathfrak{X}) + \sigma^2 \boldsymbol{I})), \quad \hat{k}(\mathfrak{X}) = \frac{1}{N-1}(Y - 1_N \hat{\mu}(\mathfrak{X})^{\mathrm{T}})^{\mathrm{T}}(Y - 1_N \hat{\mu}(\mathfrak{X})^{\mathrm{T}}) \sim \mathcal{W}(\frac{1}{N-1}(k(\mathfrak{X}) + \sigma^2 \boldsymbol{I}).$$

$$\hat{\mu}_t(x) = \hat{\mu}(x) + \hat{k}(x, \boldsymbol{x}_t)\hat{k}(\boldsymbol{x}_t, \boldsymbol{x}_t)^{-1}(\boldsymbol{y}_t - \hat{\mu}(\boldsymbol{x}_t)), \quad \hat{k}_t(x, x') = \frac{N-1}{N-t-1}\left(\hat{k}(x, x') - \hat{k}(x, \boldsymbol{x}_t)\hat{k}(\boldsymbol{x}_t, \boldsymbol{x}_t)^{-1}\hat{k}(\boldsymbol{x}_t, x')\right).$$

$\mathfrak{X} \subset \mathbb{R}^d$ is compact: using weights to represent the prior



- Assume there exist basis functions $\Phi = [\phi_s]_{s=1}^K : \mathfrak{X} \to \mathbb{R}^K$, mean parameter $\boldsymbol{u} \in \mathbb{R}^K$ and covariance parameter $\Sigma \in \mathbb{R}^{K \times K}$ such that $\mu(x) = \Phi(x)^{\mathrm{T}}\boldsymbol{u}$ and $k(x, x') = \Phi(x)^{\mathrm{T}}\Sigma\Phi(x')$, i.e. $f = \Phi(x)^{\mathrm{T}}W \sim GP(\mu, k), W \sim \mathcal{N}(\boldsymbol{u}, \Sigma)$.
- Assume $M \geq K$, and $\Phi(\bar{\boldsymbol{x}})$ has full row rank. The observation $Y_i = \Phi(\bar{\boldsymbol{x}})^{\mathrm{T}}W_i + \bar{\boldsymbol{\epsilon}}_i \sim \mathcal{N}(\Phi(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{u}, \Phi(\bar{\boldsymbol{x}})^{\mathrm{T}}\Sigma\Phi(\bar{\boldsymbol{x}}) + \sigma^2\boldsymbol{I})$, and we estimate the weight vector as $\hat{W}_i = (\Phi(\bar{\boldsymbol{x}})^{\mathrm{T}})^{+}Y_i \sim \mathcal{N}(\boldsymbol{u}, \Sigma + \sigma^2(\Phi(\bar{\boldsymbol{x}})\Phi(\bar{\boldsymbol{x}})^{\mathrm{T}})^{-1})$. Let $\mathsf{W} = [\hat{W}_i]_{i=1}^N \in \mathbb{R}^{N \times K}$.
- Unbiased GP prior parameter estimator: $\hat{\boldsymbol{u}} = \frac{1}{N}\mathsf{W}^{\mathrm{T}}1_N$ and $\hat{\Sigma} = \frac{1}{N-1}(\mathsf{W} - 1_N\hat{\boldsymbol{u}})^{\mathrm{T}}(\mathsf{W} - 1_N\hat{\boldsymbol{u}})$.
- Unbiased GP posterior estimator: $\hat{\mu}_t(x) = \Phi(x)^{\mathrm{T}}\hat{\boldsymbol{u}}_t$ and $\hat{k}_t(x) = \Phi(x)^{\mathrm{T}}\hat{\Sigma}_t\Phi(x)$ where

$$\hat{\boldsymbol{u}}_t = \hat{\boldsymbol{u}} + \hat{\Sigma}\Phi(\boldsymbol{x}_t)(\Phi(\boldsymbol{x}_t)^{\mathrm{T}}\hat{\Sigma}\Phi(\boldsymbol{x}_t))^{-1}(\boldsymbol{y}_t - \Phi(\boldsymbol{x}_t)^{\mathrm{T}}\boldsymbol{u}), \quad \hat{\Sigma}_t = \frac{N-1}{N-t-1}\left(\hat{\Sigma} - \hat{\Sigma}\Phi(\boldsymbol{x}_t)(\Phi(\boldsymbol{x}_t)^{\mathrm{T}}\hat{\Sigma}\Phi(\boldsymbol{x}_t))^{-1}\Phi(\boldsymbol{x}_t)^{\mathrm{T}}\hat{\Sigma}\right).$$

Lemma 1. *If the size of the training dataset satisfies $N \geq T + 2$, then for any input $x \in \mathfrak{X}$, with probability at least $1 - \delta$,*

$$|\hat{\mu}_t(x) - \mu_t(x)|^2 < a_t(k_t(x) + \bar{\sigma}^2(x)) \text{ and } 1 - 2\sqrt{b_t} < \hat{k}_t(x)/(k_t(x) + \bar{\sigma}^2(x)) < 1 + 2\sqrt{b_t} + 2b_t,$$

where $a_t = \frac{4\left(N-2+t+2\sqrt{t\log(4/\delta)}+2\log(4/\delta)\right)}{\delta N(N-t-2)}$ and $b_t = \frac{1}{N-t-1}\log\frac{4}{\delta}$. For finite $\mathfrak{X}$, $\bar{\sigma}^2(x) = \sigma^2$; for compact $\mathfrak{X}$, $\bar{\sigma}^2(x) = \sigma^2\Phi(x)^T(\Phi(\bar{\boldsymbol{x}})\Phi(\bar{\boldsymbol{x}})^T)^{-1}\Phi(x)$.

NEAR ZERO REGRET BOUNDS FOR BO WITH THE ESTIMATED PRIOR AND POSTERIOR

Acquisition functions: $\alpha_{t-1}^{\mathrm{PI}}(x) = \frac{\hat{\mu}_{t-1}(x) - \hat{f}^*}{\hat{k}_{t-1}(x)^{\frac{1}{2}}}$, $\quad \alpha_{t-1}^{\text{GP-UCB}}(x) = \hat{\mu}_{t-1}(x) + \zeta_t\hat{k}_{t-1}(x)^{\frac{1}{2}}$.

$$\hat{f}^* \geq \max_{x \in \mathfrak{X}} f(x), \quad \zeta_t = \frac{\left(6(N - 3 + t + 2\sqrt{t\log\frac{6}{\delta}} + 2\log\frac{6}{\delta})/(\delta N(N-t-1))\right)^{\frac{1}{2}} + (2\log(\frac{3}{\delta}))^{\frac{1}{2}}}{(1 - 2(\frac{1}{N-t}\log\frac{6}{\delta})^{\frac{1}{2}})^{\frac{1}{2}}}$$

With high probability, the simple regret decreases to a constant proportional to the noise level σ as the number of iterations and training functions increases.

Theorem 2. *Assume there exist constant $c \geq \max_{x \in \mathfrak{X}} k(x)$ and a training dataset is available whose size is $N \geq 4\log\frac{6}{\delta} + T + 2$. With probability at least $1 - \delta$, the best-sample simple regret in T iterations of meta BO with either GP-UCB or PI satisfies*

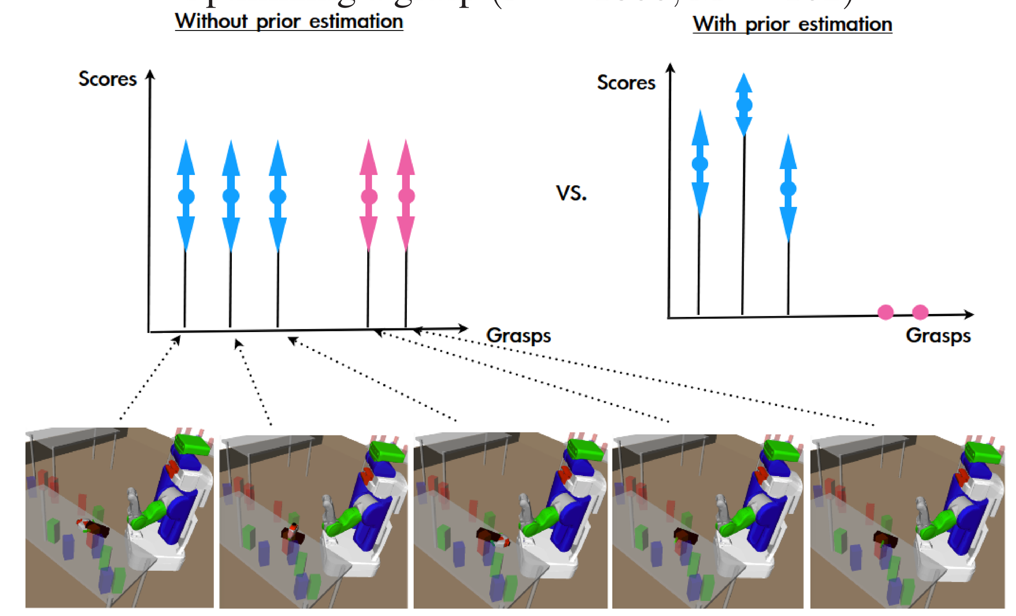
$$r_T^{UCB} < \eta_T^{UCB}(N)\lambda_T, \quad r_T^{PI} < \eta_T^{PI}(N)\lambda_T, \quad \lambda_T^2 = O(\rho_T/T) + \bar{\sigma}(x_\tau)^2,$$

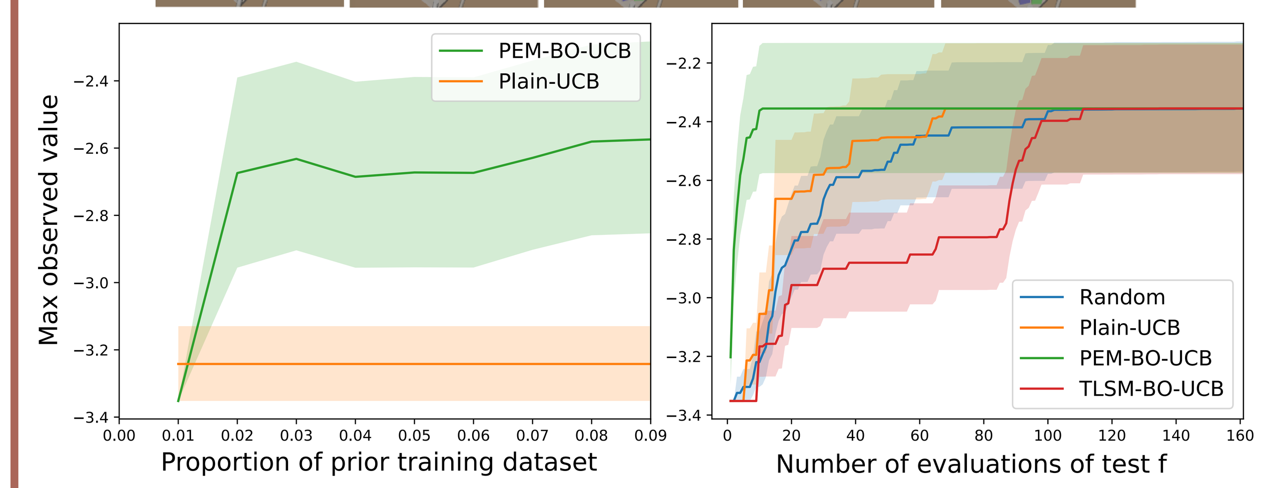
where $\eta_T^{UCB}(N) = (m + C_1)(\frac{\sqrt{1+m}}{\sqrt{1-m}} + 1)$, $\eta_T^{PI}(N) = (m + C_2)(\frac{\sqrt{1+m}}{\sqrt{1-m}} + 1) + C_3$, $m = O(\sqrt{\frac{1}{N-T}})$, $C_1, C_2, C_3 > 0$ are constants, $\tau = \arg\min_{t \in [T]} k_{t-1}(x_t)$ and $\rho_T = \max_{A \in \mathfrak{X}, |A| = T} \frac{1}{2}\log|\boldsymbol{I} + \sigma^{-2}k(A)|$. $\bar{\sigma}$ is defined in the same way as in Lemma 1.

Lemma 3. *With probability at least $1 - \delta$, the simple regret $R_T \leq r_T + 2(2\log\frac{1}{\delta})^{\frac{1}{2}}\sigma$.*

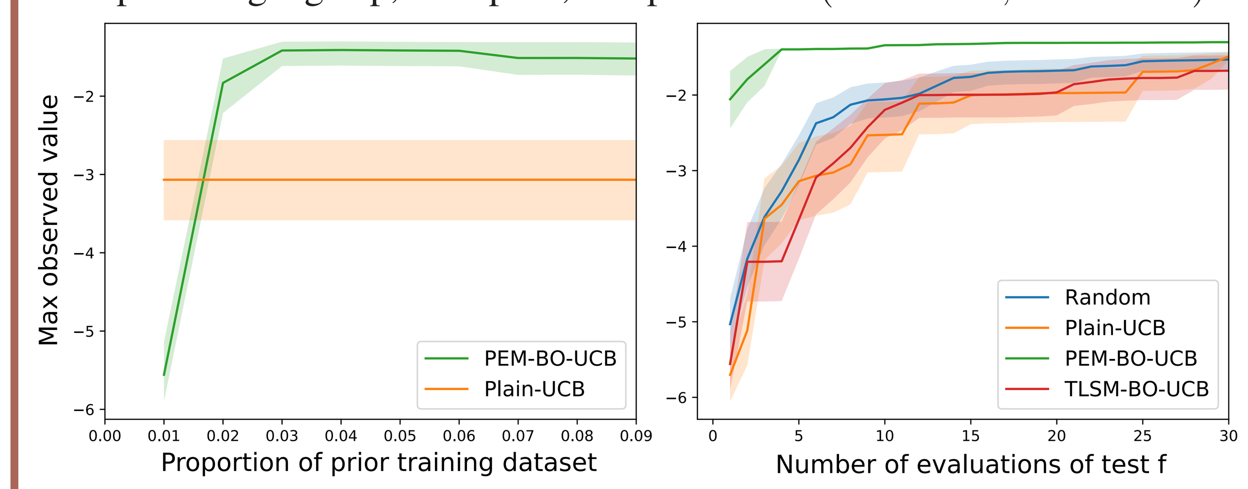
EXPERIMENTS

Optimizing a grasp ($N = 1800$, $M = 162$)

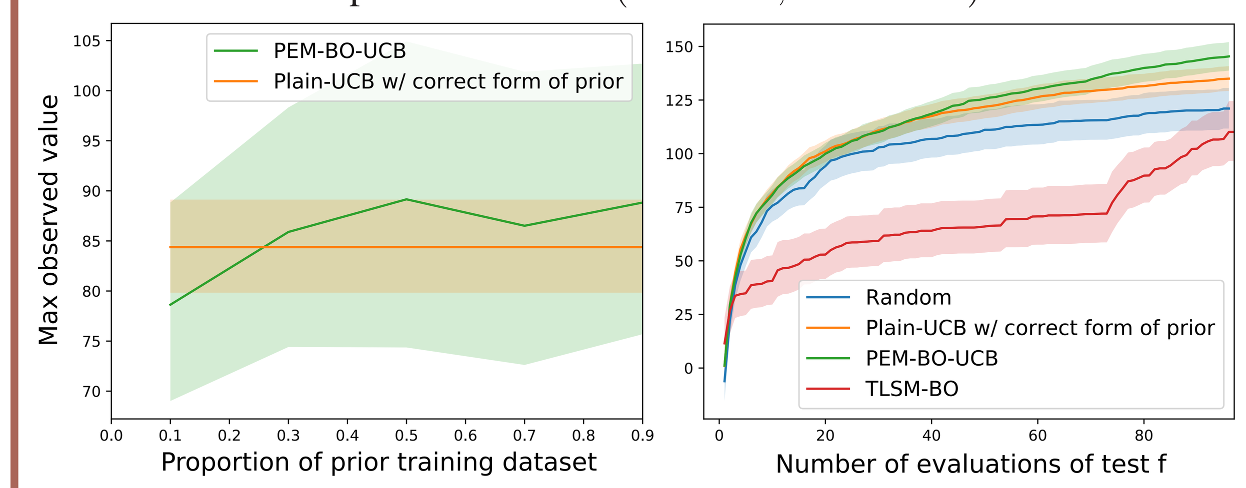




Optimizing a grasp, base pose, and placement ($N = 1500$, $M = 1000$)



Optimizing a continuous synthetic function drawn from a GP with a squared exponential kernel ($N = 100$, $M = 1000$)



- PEM-BO: our point estimate meta BO approach.
- Plain-BO: plain BO using a squared exponential kernel and maximum likelihood for GP hyperparameter estimation.
- TLSM-BO: transfer learning sequential model-based optimization [5].
- Random: random selection.

SELECTED REFERENCES

[1] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley New York, 1958.

[2] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

[3] B. Kim, L. P. Kaelbling, and T. Lozano-Pérez. Learning to guide task and motion planning using score-space representation. In *ICRA*, 2017.

[4] R. Neal. *Bayesian Learning for Neural networks*. Lecture Notes in Statistics 118. Springer, 1996.

[5] D. Yogatama and G. Mann. Efficient transfer learning method for automatic hyperparameter tuning. In *AISTATS*, 2014.