Lifelong Bandit Optimization: No Prior and No Regret

Felix Schur[*1] **Parnian Kassraie**[*1] **Jonas Rothfuss**[1] **Andreas Krause**[1]

[1]ETH Zurich, Switzerland

A PSEUDO-CODES TO ALGORITHMS

Algorithm 1 META-KGL

Require: Data from previous tasks $\mathcal{D}^{\text{exp}}_{1:s}$, threshold parameter $\omega > 0$
  $\hat{\boldsymbol{\beta}} \leftarrow \min_{\boldsymbol{\beta} \in \mathbb{R}^{sd}} \mathcal{L}\left(\boldsymbol{\beta}; \mathcal{D}^{\text{exp}}_{1:s}\right)$ ▷ solves Eq. (1)
  $\hat{J} \leftarrow \{j \leq p \mid \|\hat{\boldsymbol{\beta}}^{(j)}\|_2 \geq \omega\sqrt{s}\}$
  $\hat{k}_s \leftarrow \frac{1}{|\hat{J}|}\sum_{j \in \hat{J}} k_j$

Algorithm 2 LIBO

Require: $n, m \in \mathbb{N}$, $0 < \omega < c_1$, BASEBO
  $\hat{k}_0 \leftarrow \sum_{j=1}^{p} \frac{1}{p} k_j$
  for $s \in \{1, \dots, m\}$ **do**
    $\mathcal{D}^{\text{exp}}_s \leftarrow \emptyset$ ▷ Dataset for kernel prediction
    $\mathcal{D}_s \leftarrow \emptyset$ ▷ Dataset for the base bandit algorithm
    for $i \in \{1, \dots, n\}$ **do**
      if $i \leq \frac{\sqrt{n}}{s^{1/4}}$ **then** ▷ Forced exploration with rate $\sqrt{n}/s^{1/4}$
        Sample $\boldsymbol{x}_{s,i}$ uniformly from $\mathcal{X}$
        Play action $\boldsymbol{x}_{s,i}$ and observe $y_{s,i}$
        $\mathcal{D}^{\text{exp}}_s \leftarrow \mathcal{D}^{\text{exp}}_s \cup \{(\boldsymbol{x}_{s,i}, y_{s,i})\}$ ▷ Add to kernel prediction dataset
      else
        $\boldsymbol{x}_{s,i} \leftarrow \text{BASEBO}(\hat{k}_{s-1})$ ▷ Select action using base bandit algorithm
        Play action $\boldsymbol{x}_{s,i}$ and observe $y_{s,i}$
      end if
      $\mathcal{D}_s \leftarrow \mathcal{D}_s \cup \{(\boldsymbol{x}_{s,i}, y_{s,i})\}$
      Update BASEBO using $\mathcal{D}_s$ ▷ Update base bandit algorithm
    end for
    $\hat{k}_s \leftarrow \text{META-KGL}(\mathcal{D}^{\text{exp}}_{1:s}, \omega)$ ▷ Update $\hat{k}$ using META-KGL and $\mathcal{D}^{\text{exp}}_s$
  end for

[*]Equal contribution.

Accepted for the 39th Conference on Uncertainty in Artificial Intelligence (UAI 2023).

Algorithm 3 F-LiBO

Require: $n, m \in \mathbb{N}$, $0 < \omega < c_1$, BASEBO
  $\hat{k}_0 \leftarrow \sum_{j=1}^{p} \frac{1}{p} k_j$
  for $s \in \{1, \dots, m\}$ **do**
    $\mathcal{D}_s^{\text{exp}} \leftarrow \emptyset$ ▷ Dataset for kernel prediction
    $\mathcal{D}_s \leftarrow \emptyset$ ▷ Dataset for the base bandit algorithm
    for $i \in \{1, \dots, n\}$ **do**
      if $i \leq \sqrt{n}$ **then** ▷ Forced exploration with rate $\sqrt{n}$
        Sample $\boldsymbol{x}_{s,i}$ uniformely from $\mathcal{X}$
        Play action $\boldsymbol{x}_{s,i}$ and observe $y_{s,i}$
        $\mathcal{D}_s^{\text{exp}} \leftarrow \mathcal{D}_s^{\text{exp}} \cup \{(\boldsymbol{x}_{s,i}, y_{s,i})\}$ ▷ Add to kernel prediction dataset
      else
        $\boldsymbol{x}_{s,i} \leftarrow \text{BASEBO}(\hat{k}_s)$ ▷ Select action using base bandit algorithm
        Play action $\boldsymbol{x}_{s,i}$ and observe $y_{s,i}$
      end if
      $\mathcal{D}_s \leftarrow \mathcal{D}_s \cup \{(\boldsymbol{x}_{s,i}, y_{s,i})\}$
      Update BASEBO using $\mathcal{D}_s$ ▷ Update base bandit algorithm
    end for
    $\hat{k}_s \leftarrow \text{F-META-KGL}(\mathcal{D}_{1:s}^{\text{exp}}, \omega)$ ▷ Update $\hat{k}$ using F-META-KGL and $\mathcal{D}_s^{\text{exp}}$
  end for

Algorithm 4 F-META-KGL

Require: $n, m \in \mathbb{N}$, data for each task $\mathcal{D}_s$, $\omega > 0$, $\alpha \in [0, 1]$
  $\text{count}_1, \dots, \text{count}_p \leftarrow 0$
  for $s \in \{1, \dots, m\}$ **do**
    $\hat{\boldsymbol{\beta}}_s \leftarrow \min_{\boldsymbol{\beta} \in \mathbb{R}^d} \mathcal{L}\left(\boldsymbol{\beta}; \mathcal{D}_s^{\text{exp}}\right)$
    for $j \in \{1, \dots, p\}$ **do**
      $\text{count}_j \leftarrow \text{count}_j + \mathbb{1}\{\|\hat{\boldsymbol{\beta}}_s^{(j)}\|_2 \geq \omega\}$
    end for
  end for
  $\hat{J} \leftarrow \{j \leq p \mid \text{count}_j \geq m\alpha\}$
  $\hat{k} \leftarrow \frac{1}{|\hat{J}|} \sum_{j \in \hat{J}} k_j$

B EXTENDED LITERATURE REVIEW

In this section, we present an overview of works that consider learning a potentially low-dimensional reward function by leveraging data of similar bandit tasks.

Linear Contextual Bandits with Shared Representation. The common assumption here is that the reward function for all $s \in [m]$, is linear $f_s(\boldsymbol{x}) = \langle \boldsymbol{\theta}_s, \boldsymbol{x} \rangle$ where $\boldsymbol{\theta}_s = \boldsymbol{B}\boldsymbol{w}_s$. The matrix $\boldsymbol{B} \in \mathbb{R}^{d \times r}$ is a shared representation matrix and $r \ll d$ is an intrinsic dimension. This assumption becomes more intuitive if we re-write the reward as $f_s(\boldsymbol{x}) = \langle \boldsymbol{w}_s, \boldsymbol{B}^T \boldsymbol{x} \rangle$, which implies that there exists a mapping $\boldsymbol{B} : \mathbb{R}^d \to \mathbb{R}^r$ that produces a low-dimensional representation of the actions. Our reward assumption implies that there exists a sparse matrix $\boldsymbol{S} \in \mathbb{R}^{d \times d^\star}$ which satisfies $f_s(\boldsymbol{x}) = \langle \boldsymbol{\theta}_s, \boldsymbol{S}^T \boldsymbol{\phi}(\boldsymbol{x}) \rangle$ and screens the relevant features $\boldsymbol{\phi}_j$ with $j \in J^\star$. The intrinsic dimension r then corresponds to $|J^\star|$.

Recent work on shared representation learning, often consider the contextual setting, where at every step of the bandit problem, actions may only be chosen from a set $\mathcal{A}_{s,t}$. Once the action is chosen, a noisy reward is observed. Regarding the occurrence of the tasks, two scenarios are often studied. The multi-task setting where all the tasks are solved concurrently, and the lifelong setting where the tasks arrive consecutively. Table 1 summarizes these efforts in terms of the obtained regret bounds. Here, $\tilde{\mathcal{O}}$ hides polylogarithmic factors. With the exception of [Hu et al., 2021], these works either 1) require forced exploration to fulfill sufficient exploration assumptions (SE) similar to Assumption 3.2, or 2) design a greedy algorithm assuming that the actions in set $\mathcal{A}_{s,t}$ are sampled from a diverse context distribution (DC) which gives free exploration [c.f. Bastani et al., 2021]. This suggests that for minimax optimality, either the algorithm has to explore, or the presented context should induce exploration for free. To better understand the tightness of the results in Table 1, we recall that the oracle solver which has knowledge of the representation matrix $\boldsymbol{B}$, has a lower-bound of $R^\star(n) = \Omega(\sqrt{rn \log n \log k})$, when $|\mathcal{A}_{s,t}| = k$ [Li et al., 2019]. If $\mathcal{A}_{s,t}$ is an ellipsoid, the lower-bound achievable by the oracle is $\Omega(r\sqrt{n})$ [Li et al., 2021]. Clearly, for $r \ll d$, the algorithms of Hu et al. [2021], Cella and Pontil [2021], and Cella et al. [2022] do not converge to the oracle solver as $m \to \infty$, since $R(m,n)/m \not\to R^\star(n)$.

	$\mathcal{A}_{s,t}$	r	Tasks	Expected Lifelong Regret	Base Policy	Assumptions
Yang et al. [2021]	finite	known	conc.	$\tilde{\mathcal{O}}\left(m\sqrt{rn} + \sqrt{rdnm}\right)$	Greedy	DC
	ellipsoid	known	conc.	$\tilde{\mathcal{O}}\left(mr\sqrt{n} + d^{3/2}r\sqrt{nm}\right)$	ETC	SE
Hu et al. [2021]	compact	known	conc.	$\tilde{\mathcal{O}}\left(m\sqrt{drn} + d\sqrt{rnm}\right)$	OFUL	-
Cella et al. [2022]	finite	unknown	seq.	$\tilde{\mathcal{O}}\left(mr\sqrt{n} + r\sqrt{dnm}\right)$	Greedy	DC
Cella and Pontil [2021]	finite	unknown	seq./conc.	$\tilde{\mathcal{O}}\left(mr\sqrt{n}\right)$	Greedy	DC & SE
Yang et al. [2022]	compact	known	seq.	$\tilde{\mathcal{O}}\left(mr\sqrt{n} + d\sqrt{rnm}\right)$	ETC	SE
LiBO (Ours)	compact	unknown	seq.	$\tilde{\mathcal{O}}\left(mr\sqrt{n} + m^{3/4}\sqrt{n}\right)$	any	SE

Table 1: Overview of recent work on representation learning for contextual linear bandits. Oracle lower-bound is $R^\star(m,n) = \Omega\left(mr\sqrt{n}\right)$ for infinite action set, and $R^\star(m,n) = \Omega\left(m\sqrt{rn\log n \log k}\right)$ for finite set. Polylog terms are not included.

Bayesian Bandits with Shared Prior Distributions. Alternatively, some works consider a Bayesian reward model, but without any assumption on sparsity, or low-dimensional representations. Let $f_s(\boldsymbol{x}) = \langle \boldsymbol{\theta}_s, \boldsymbol{x}\rangle$ where $\boldsymbol{\theta}_s$ are i.i.d. from $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and the parameters $(\boldsymbol{\mu}, \Sigma)$ are shared across all tasks. Peleg et al. [2022] assume that $(\boldsymbol{\mu}, \Sigma)$ are unknown, and estimate it using the exploratory action-reward pairs collected during the first m_0 tasks. The suggested meta-algorithm can be wrapped around any Quasi-Bayesian base policy, such as Thomspon Sampling [Thompson, 1933] or Information Directed Sampling [Russo and Van Roy, 2014], however, the resulting algorithm over-explores as indicated by the $\tilde{\mathcal{O}}(md)$ term in the regret bound (See Table 2).

Taking a hierarchically Bayesian approach, Basu et al. [2021] and Hong et al. [2022] further assume that $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\mu}_0, \Sigma_0)$ where $\boldsymbol{\mu}_0$ is unknown, but both covariance matrices Σ and Σ_0 are known. Prior distribution of $\boldsymbol{\mu}$ is updated after each task, according to the evidence collected during the task. Both papers design a meta-algorithm with Thomspon Sampling as the base solver. While Basu et al. [2021] suffers from over-exploration, Hong et al. [2022] does not require any exploration. Indeed, if Σ the covariance matrix between the actions is known, it helps with inferring rewards of other actions, and reduces the need for uniform exploration.

An overview is given in Table 2, here $R^\star(m,n)$ indicates the Bayesian lifelong regret of the oracle agent who has knowledge of $(\boldsymbol{\mu}, \Sigma)$. Note that Theorem 4.1 gives slightly stronger result, which is a high-probability bound over the regret. Here, we have taken the average to make it comparable with the Bayes regret reported in other works. As m grows, the average single-task regret is upper bounded by $R(m,n)/m$, implying that only Hong et al. [2022] and LiBO can converge to the oracle solver.

	Σ/Σ_0	sparse	Tasks	Bayesian Regret	Policy	Exp
Basu et al. [2021]	known	no	seq.	$\tilde{\mathcal{O}}\left(R^\star(m,n) + \sqrt{dnm} + md\right)$	TS	yes
Peleg et al. [2022]	unknown	no	seq.	$\tilde{\mathcal{O}}\left((1 + d^3/\sqrt{m})R^\star(m,n) + md\right)$	any QB	yes
Hong et al. [2022]	known	no	seq.	$\tilde{\mathcal{O}}\left(R^\star(m,n) + \sqrt{dmn} + d^{3/2}\right)$	TS	no
	known	no	conc.	$\tilde{\mathcal{O}}\left(R^\star(m,n) + \sqrt{dmn} + md^{3/2}\right)$	TS	yes
LiBO (Ours)	-	yes	seq.	$\tilde{\mathcal{O}}\left(R^\star(m,n) + m^{3/4}\sqrt{n} + (mn)^{1/3}\log(md)\right)$	any	yes

Table 2: Overview of recent work on meta-learning Bayesian priors for linear bandits . All works consider compact action set, except for Basu et al. [2021] which requires a finite set of actions selected from $\mathbb{R}^d$. The regret of the oracle solver is denoted by $R^\star(m,n)$.

Overall Landscape of Research. We merge the two lines of work in Table 3, to give an overview of ongoing efforts on meta-learning for linear bandits and the important properties of each method. Column $|\mathcal{A}_t|$ shows if the model holds for compact action sets, or only finite ones. Column "sparse" and "learns r" denote whether the model aims for sparse solution, and if it requires knowledge of the true sparsity r or preferably, it learns it. Column " learns Σ/Σ_0" only applies to Bayesian method, where some assume the covariance matrix of $\boldsymbol{\theta}_s$ is known, and some estimate it from data. Column "Tasks" shows if the method considers simultaneous or sequentially appearing bandit tasks. Column "O-opt" refers to oracle optimality, and has a checkmark only if $R(m,n)/m \to R^\star(n)$. Column "MS Cost" shows the cost of model selection/meta-learning. In particular, shows whether the additional regret, incurred due to not knowing the true representation/features, is logarithmic in dimension d or is it polynomial. Column "Policy" shows the base BO solvers that can be paired with the meta-learning algorithm, "any" indicates that the method can work with any (linear) policy, and "any QB" refers to Quasi-Bayesian methods. Column "Ass." shows the assumptions of the method on diversity of data, SE indicates Sufficient Exploration type assumptions, and DC refers to Diverse Context assumptions. Column "Has no uniform draws" shows if the algorithm requires forces exploration or not.

	$\lvert\mathcal{A}_t\rvert$	Sparse	Learns r	Learns Σ/Σ_0	Tasks	O-opt	MS cost	Policy	Ass.	Has no unif. draws
Yang et al. [2021]	k	✓	✗	-	conc	✓	poly(d)	greedy	DC	✗
	∞	✓	✗	-	conc	✓	poly(d)	ETC	SE	✗
Hu et al. [2021]	∞	✓	✗	-	conc	✗	poly(d)	OFUL	-	✓
Cella and Pontil [2021]	k	✓	✓	-	both	✗	$\log(d)$	greedy	DC SE	✗
Cella et al. [2022]	k	✓	✓	-	seq	✗	poly(d)	greedy	DC	✓
Yang et al. [2022]	∞	✓	✗	-	conc	✓	poly(d)	ETC	SE	✗
Basu et al. [2021]	k	✗	-	✗	seq	✗	poly(d)	TS	SE	✗
Peleg et al. [2022]	∞	✗	-	✓	seq	✗	poly(d)	any QB	SE	✗
Hong et al. [2022]	∞	✗	-	✗	seq	✓	poly(d)	TS	-	✓
	∞	✗	-	✗	conc	✗	poly(d)	TS	SE	✗
LIBO (Ours)	∞	✓	✓	-	seq	✓	$\log(d)$	any	SE	✗
F-LIBO (Ours)	∞	✓	✓	-	seq	✓	$\log(d)$	any	SE	✗

Table 3: Collective pro and cons of related works: LIBO gives an overall pareto-optimal solution. Refer to the corresponding paragraph in Appendix B for information on meaning of each column. Table 1 presents a concise version.

C GENERALITY OF THE KERNEL ASSUMPTION

In Section 2, we claim that, the average of kernels formulation, i.e.

$$k^\star(\cdot,\cdot) = \frac{1}{|J^\star|}\sum_{j\in J^\star} k_j(\cdot,\cdot)$$

is without loss of generality equivalent to assuming a linear combination,

$$k^\star(\cdot,\cdot) = \sum_{j\in J^\star} \alpha_j k_j(\cdot,\cdot).$$

Here, we formally show this claim. Assume there exist $\alpha_1,\dots,\alpha_p \in \mathbb{R}$ and kernels $k_1,\dots,k_p$ such that

$$k^\star(\boldsymbol{x},\boldsymbol{x}') = \sum_{j\in J^\star} \alpha_j k_j(\boldsymbol{x},\boldsymbol{x}'), \quad \forall \boldsymbol{x},\boldsymbol{x}' \in \mathcal{X}.$$

Let $f \in \mathcal{H}_{k^\star}$, then there exists $\beta_1 \in \mathbb{R}^{d_1},\dots,\beta_p \in \mathbb{R}^{d_p}$ such that for all $x \in \mathcal{X}$

$$f(\boldsymbol{x}) = \sum_{j\in J^\star} \sqrt{\alpha_j}\beta_j^\top \phi_j(\boldsymbol{x}).$$

Define $m_j := \max\{k(\boldsymbol{x}, \boldsymbol{x}) \mid \boldsymbol{x} \in \mathcal{X}\}$, $\tilde{\beta}_j := p m_j \beta_j \sqrt{\alpha_j}$ and $\tilde{k}_j := k_j/m_j$ for all $j \in \{1, \ldots, p\}$, then

$$f(\boldsymbol{x}) = \frac{1}{p} \sum_{j \in J^\star} \tilde{\beta}_j^\top \phi_j(\boldsymbol{x})$$

and therefore $f \in \mathcal{H}_{\tilde{k}}$ for

$$\tilde{k}^\star(\boldsymbol{x}, \boldsymbol{x}') = \frac{1}{|J^\star|} \sum_{j \in J^\star} k_j(\boldsymbol{x}, \boldsymbol{x}').$$

This shows that the corresponding Reproducing Kernel Hilbert Spaces are equivalent, i.e. the same functions reside in both, while the norm is scaled. Therefore, we can assume, without loss of generality, that the base kernels are normalized and that the true kernel is an average of base kernels.

D CONSISTENCY OF META-KGL (PROOF OF THEOREM 3.3)

We start by proving the necessary lemmas. During this section we assume a slightly more general setting. More precisely, we assume that we have $n_s \leq n$ samples in task s, which means that the total samples size of the meta-dataset is $N = N_m := |\mathcal{D}^{\text{exp}}| = \sum_{s=1}^{m} n_s$.

Definition D.1 (sub-Gaussian random variables)**.** *Let X be a random variable. We call X a σ* sub-Gaussian random variable *if $\mathbb{E}[X] = 0$ and*

$$\mathbb{P}[|X| \geq t] \leq 2 \exp\left(-\frac{t^2}{2\sigma^2}\right). \tag{D.1}$$

Lemma D.2 (Theorem 6.3.2 of Vershynin [2018])**.** *Let $\epsilon_1, \ldots, \epsilon_n$ be independent, zero mean, unit variance sub-Gaussian random variables. Define $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$. Let $A \in \mathbb{R}^{m \times n}$ and $t \geq 0$. Then*

$$\mathbb{P}\left(\Big| \|A\epsilon\|_2 - \|A\|_F \Big| \geq t\right) \leq \exp\left(-\frac{t^2}{2\|A\|_2^2}\right).$$

Corollary D.3. *Let $\epsilon_1, \ldots, \epsilon_n$ be i.i.d. σ sub-Gaussian random variables and define $\epsilon = (\epsilon_1, \ldots, \epsilon_n)$. Let $A \in \mathbb{R}^{m \times n}$ and $t \geq \sigma\sqrt{\operatorname{tr}(AA^T)}$. Then*

$$\mathbb{P}\left(\|A\epsilon\|_2 \geq t\right) \leq \exp\left(-\frac{\left(t/\sigma - \sqrt{\operatorname{tr}(AA^T)}\right)^2}{2\|AA^T\|_2}\right).$$

Proof. The standard deviation of an σ sub-Gaussian random variable is smaller equal σ. Therefore

$$\begin{aligned}
\mathbb{P}\left(\|A\epsilon\|_2 \geq t\right) &= \mathbb{P}\left(\|A\epsilon\|_2/\sqrt{Var(\epsilon_1)} \geq t/\sqrt{Var(\epsilon_1)}\right) \\
&\leq \mathbb{P}\left(\|A\epsilon\|_2/\sqrt{Var(\epsilon_1)} \geq t/\sigma\right).
\end{aligned}$$

It holds that $\|A\|_F = \sqrt{\operatorname{tr}(AA^T)}$. Define $\tilde{\epsilon} = \epsilon/\sqrt{Var(\epsilon_1)}$. We have

$$\begin{aligned}
\mathbb{P}\left(\|A\epsilon\|_2 \geq t\right) &\leq \mathbb{P}\left(\|A\tilde{\epsilon}\|_2 - \|A\|_F \geq t/\sigma - \sqrt{\operatorname{tr}(AA^T)}\right) \\
&\leq \mathbb{P}\left(|\|A\tilde{\epsilon}\|_2 - \|A\|_F| \geq t/\sigma - \sqrt{\operatorname{tr}(AA^T)}\right).
\end{aligned}$$

Using Lemma D.2 and noting that $\|A\|_2^2 = \|AA^T\|_2$ yields the desired result. □

Lemma D.4. *Consider the model in Eq.* (1) *with σ sub-Gaussian i.i.d. noise. Then, for $\frac{\lambda N}{4\sigma} > \sqrt{\mathrm{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)}$ with probability at least*

$$1 - \sum_{j=1}^{p} \exp\left(- \frac{\left(\frac{\lambda N}{4\sigma} - \sqrt{\mathrm{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)} \right)^2}{2\|\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T\|_2} \right)$$

we have for any solution $\hat{\boldsymbol{\beta}}$ of Eq. (1)

$$\frac{1}{N}\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)\|_2^2 + \frac{\lambda}{2}\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}^{\star(j)}\|_2 \leq$$
$$2\lambda \sum_{j\in J^\star} \min\left(\|\hat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}^{\star(j)}\|_2, \|\boldsymbol{\beta}^{\star(j)}\|_2 \right).$$

Proof. This proof is inspired by the proof of Lemma 3.1 in Lounici et al. [2011].
For all solutions $\hat{\boldsymbol{\beta}}$ of Eq. (1)

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathbb{R}^{dm}}{\arg\min} \frac{1}{N}\|\mathbf{\Phi}\boldsymbol{\beta}^\star + \epsilon - \mathbf{\Phi}\boldsymbol{\beta}\|_2^2 + \lambda\sum_{j=1}^{p}\|\boldsymbol{\beta}^{(j)}\|_2.$$

Therefore for all $\boldsymbol{\beta} \in \mathbb{R}^{dm}$

$$\frac{1}{N}\|\mathbf{\Phi}\boldsymbol{\beta}^\star + \epsilon - \mathbf{\Phi}\hat{\boldsymbol{\beta}}\|_2^2 + \lambda\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}\|_2 \leq \frac{1}{N}\|\mathbf{\Phi}\boldsymbol{\beta}^\star + \epsilon - \mathbf{\Phi}\boldsymbol{\beta}\|_2^2 + \lambda\sum_{j=1}^{p}\|\boldsymbol{\beta}^{(j)}\|_2.$$

This yields

$$\frac{1}{N}\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\|_2^2 \leq \frac{1}{N}\|\mathbf{\Phi}(\boldsymbol{\beta} - \boldsymbol{\beta}^\star)\|_2^2 + \frac{2}{N}\epsilon^T\mathbf{\Phi}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$$
$$+ \lambda\sum_{j=1}^{p}\left(\|\boldsymbol{\beta}^{(j)}\|_2 - \|\hat{\boldsymbol{\beta}}^{(j)}\|_2 \right).$$

And in particular if $\boldsymbol{\beta} = \boldsymbol{\beta}^\star$, then,

$$\frac{1}{N}\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)\|_2^2 \leq \frac{2}{N}\epsilon^T\mathbf{\Phi}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star)$$
$$+ \lambda\sum_{j=1}^{p}\left(\|\boldsymbol{\beta}^{\star(j)}\|_2 - \|\hat{\boldsymbol{\beta}}^{(j)}\|_2 \right).$$

By Corollary D.3 and union bound we have jointly for all $j \leq p$ with probability at least

$$1 - \sum_{j=1}^{p} \exp\left(- \frac{\left(\frac{\lambda N}{4\sigma} - \sqrt{\mathrm{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)} \right)^2}{2\|\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T\|_2} \right)$$

that

$$\|(\epsilon^T\mathbf{\Phi})^{(j)}\|_2 \leq \frac{\lambda N}{4}.$$

Therefore, by Cauchy-Schwarz,

$$\epsilon^T\mathbf{\Phi}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^\star) \leq \sum_{j=1}^{p}\|(\epsilon^T\mathbf{\Phi})^{(j)}\|_2\|\hat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}^{\star(j)}\|_2$$
$$\leq \frac{\lambda N}{4}\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}^{\star(j)}\|_2.$$

This implies that

$$\frac{1}{N}\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{\star})\|_2^2 \leq \frac{\lambda}{2}\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 + \lambda\sum_{j=1}^{p}\left(\|\boldsymbol{\beta}^{\star(j)}\|_2 - \|\hat{\boldsymbol{\beta}}^{(j)}\|_2\right).$$

Therefore

$$\begin{aligned}\frac{1}{N}\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{\star})\|_2^2+\frac{\lambda}{2}\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 &\leq \\ &\lambda\sum_{j=1}^{p}\left(\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 + \|\boldsymbol{\beta}^{\star(j)}\|_2 - \|\hat{\boldsymbol{\beta}}^{(j)}\|_2\right)\end{aligned}$$

and since $\boldsymbol{\beta}^{\star(j)} = 0$ for all $j \notin J^{\star}$

$$\begin{aligned}\frac{1}{N}\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{\star})\|_2^2+\frac{\lambda}{2}\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 &\leq \\ &2\lambda\sum_{j\in J^{\star}}\min\left(\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2, \|\boldsymbol{\beta}^{\star(j)}\|_2\right).\end{aligned}$$

This proves the statement. ☐

Lemma D.5. *Let Assumption 3.1 hold. If $\hat{\boldsymbol{\beta}}$ is a solution of Eq. (1) then we have for $\frac{\lambda N}{4\sigma} > \sqrt{\mathrm{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)}$ with probability at least*

$$1-\sum_{j=1}^{p}\exp\left(-\frac{\left(\frac{\lambda N}{4\sigma}-\sqrt{\mathrm{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)}\right)^2}{2\|\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T\|_2}\right)$$

that

$$\sum_{j\notin J^{\star}}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 \leq 3\sum_{j\in J^{\star}}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2,$$

and

$$\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 \leq \frac{8\lambda m}{\bar{\kappa}^2}, \tag{D.2}$$

where

$$\bar{\kappa} := \frac{\sqrt{m}}{\sqrt{N}}\frac{\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^{\star})\|_2}{\sum_{j\in J^{\star}}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2}. \tag{D.3}$$

Proof. Lemma D.4 implies

$$\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 \leq 4\sum_{j\in J^{\star}}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 \tag{D.4}$$

and therefore

$$\sum_{j\notin J^{\star}}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 \leq 3\sum_{j\in J^{\star}}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2,$$

which yields the first statement of Lemma D.5.
Again, by the first statement of Lemma D.4, we have with probability at least

$$1-\sum_{j=1}^{p}\exp\left(-\frac{\left(\frac{\lambda N}{4\sigma}-\sqrt{\mathrm{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)}\right)^2}{2\|\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T\|_2}\right)$$

that

$$\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star)\|_2 \le \sqrt{2\lambda N}\sqrt{\sum_{j\in J^\star}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2}. \tag{D.5}$$

Therefore

$$\begin{aligned}\sum_{j\in J^\star}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 &\le \frac{\sum_{j\in J^\star}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2}{\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star)\|_2}\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star)\|_2\\ &\le \frac{\sum_{j\in J^\star}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2}{\|\mathbf{\Phi}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}^\star)\|_2}\sqrt{2\lambda N}\sqrt{\sum_{j\in J^\star}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2}.\end{aligned}$$

Solving this yields

$$\sum_{j\in J^\star}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 \le \frac{2\lambda m}{\bar{\kappa}^2},$$

and by Eq. (D.4) we have

$$\begin{aligned}\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 &\le 4\sum_{j\in J^\star}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2\\ &\le \frac{8\lambda m}{\bar{\kappa}^2}.\end{aligned}$$

□

Definition D.6 (compatibility variable)**.** *Let*

$$S := \left\{(J,b)\subset\mathcal{P}(\{1,\dots,p\})\times(\mathbb{R}^d\backslash\{0\})\;\middle|\;|J|\le s^\star, \sum_{j\notin J}\|b^{(j)}\|_2 \le 3\sum_{j\in J}\|b^{(j)}\|_2\right\}.$$

For $n,d\in\mathbb{N}$ and $\mathbf{\Phi}\in\mathbb{R}^{N\times d}$ we define $\kappa(\mathbf{\Phi})$ by

$$\kappa(\mathbf{\Phi}) := \inf_{(J,b)\in S}\frac{\sqrt{m}}{\sqrt{N}}\frac{\|\mathbf{\Phi}b\|_2}{\sum_{j\in J}\|b^{(j)}\|_2}.$$

and call it the compatibility variable of $\mathbf{\Phi}$.

Remark D.7. *It holds that $\kappa\le\bar{\kappa}$.*

Corollary D.8. *Let $0<\omega<c_1$ and let $\hat{\boldsymbol{\beta}}$ is a solution of Eq. (1) with*

$$\lambda \le \frac{\bar{\omega}\kappa^2}{8\sqrt{m}},$$

where $\bar{\omega} := \min\{\omega, c_1-\omega\}$. Then we have for $\frac{\lambda N}{4\sigma} > \sqrt{\operatorname{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)}$ with probability at least

$$1-\sum_{j=1}^{p}\exp\left(-\frac{\left(\frac{\lambda N}{4\sigma}-\sqrt{\operatorname{tr}(\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T)}\right)^2}{2\|\mathbf{\Phi}^{(j)}(\mathbf{\Phi}^{(j)})^T\|_2}\right)$$

that

$$\frac{1}{\sqrt{m}}\sum_{j=1}^{p}\|\hat{\boldsymbol{\beta}}^{(j)}-\boldsymbol{\beta}^{\star(j)}\|_2 \le \bar{\omega}. \tag{D.6}$$

If additionally Assumption 3.1 holds, then we have with the same probability for

$$\hat{J} := \left\{j\in\{1,\dots,p\}\;\middle|\;\|\hat{\boldsymbol{\beta}}^{(j)}\|_2 > \omega\sqrt{m}\right\}$$

that

$$\hat{J} = J^\star.$$

Proof. The first statement follows directly from Lemma D.5.
Assume $j \in J^\star$. Then by Assumption 3.1

$$\|\hat{\boldsymbol{\beta}}^{(j)}\|_2 > \sqrt{m}\Big(c_1 - \|\hat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}^{\star(j)}\|_2\Big) \geq \sqrt{m}(c_1 - \bar{\omega}) \geq \omega\sqrt{m},$$

which implies $J^\star \subset \hat{J}$. Assume $j \notin J^\star$, then

$$\|\hat{\boldsymbol{\beta}}^{(j)}\|_2 \leq \|\hat{\boldsymbol{\beta}}^{(j)} - \boldsymbol{\beta}^{\star(j)}\|_2 + \|\boldsymbol{\beta}^{\star(j)}\|_2 \leq \bar{\omega}\sqrt{m} \leq \omega\sqrt{m},$$

which implies $\hat{J} \subset J^\star$ □

Remark D.9. *Choosing ω optimally yields $\omega = \bar{\omega} = c_1/2$.*

Proof of Theorem 3.3. Note that $\boldsymbol{\Phi}^{(j)} \in \mathbb{R}^{N \times md_j}$ is block-diagonal. Since by assumption $k_j(x, x') \leq 1, \forall j \leq p$ we have

$$\operatorname{tr}\left((\boldsymbol{\Phi}^{(j)})^T \boldsymbol{\Phi}^{(j)}\right) = \operatorname{tr}\left(\boldsymbol{\Phi}^{(j)}(\boldsymbol{\Phi}^{(j)})^T\right) = \sum_{s=1}^{m}\sum_{i=1}^{n_s} k_j\left(x_i^{(s)}, x_i^{(s)}\right) \leq N,$$

and

$$\begin{aligned}
\|(\boldsymbol{\Phi}^{(j)})^T \boldsymbol{\Phi}^{(j)}\|_2 &= \|\boldsymbol{\Phi}^{(j)}(\boldsymbol{\Phi}^{(j)})^T\|_2 \\
&\leq \max_{s \leq m} \operatorname{tr}\left(\boldsymbol{\Phi}_s^{(j)}(\boldsymbol{\Phi}_s^{(j)})^T\right) \\
&\leq \max_{s \leq m} \sum_{i=1}^{n_s} k_j\left(x_i^{(s)}, x_i^{(s)}\right) \\
&\leq n.
\end{aligned}$$

Corollary D.8 yields the result. □

E LIFELONG ANALYSIS (PROOF OF THEOREM 4.1)

We start by proving a generic variant of Theorem 4.1, from which we can obtain the theorem in the main text as a corollary.

Theorem E.1. *Assume that the true reward functions $f_1, \dots, f_m$ satisfy $\|f_i\|_{\mathcal{H}_{k^\star}} \leq B$ for some constant $B > 0$. Assume $\{n_s\}_{s \in \mathbb{N}}$ is a non-increasing sequence with $n_s \leq n, \forall s$. Define $N_m := \sum_{s=1}^m n_s$. Let ν be a distribution on $\mathcal{X}^N$ independent of $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_m$. Let $V \sim \nu$ be the random vector used for forced exploration. Let $\tilde{\boldsymbol{\Phi}}_s \in \mathbb{R}^{N_s \times md}$ be the data matrix obtained by forced exploration. Assume the forced exploration distribution ν and $\{k_j\}_{j \leq p}$ are such that, with probability at least $1 - \delta/4$, there exists $c_\kappa > 0$ such that $\kappa(\tilde{\boldsymbol{\Phi}}_s) \geq c_\kappa, \forall s \leq m$. Assume further that* BaseBO *using the true kernel function for m tasks with n interactions with independent noise achieves with probability at least $1 - \delta/2$ cumulative regret lower than $R^\star(m, n)$ in the worst-case. Then, for $m_0 \in \mathbb{N}$ and $0 < \delta < 1$, if*

$$N_{m_0} \geq \frac{2n_1 \log(4mp/\delta)}{(\sqrt{N_m/m}\frac{c_1 c_\kappa^2}{32\sigma} - 1)^2},$$

with probability at least $1 - \delta$ LiBO *achieves*

$$R(m, n) \leq 2Bm_0 n + 2BN_m + R^\star(m, n).$$

Proof. Denote by $C := \{v \mid \kappa(\tilde{\boldsymbol{\Phi}}_s(v)) \geq c_\kappa, \forall s \leq m\}$ the set of data points such that κ is lower bounded by c_κ. By assumption we have $\mathbb{P}[V \in C] \geq 1 - \delta/4$. Denote by $\hat{J}_s \subset \{1, \dots, p\}$ the sparsity structure predicted by LiBO after the first s tasks. Note that $\tilde{\boldsymbol{\Phi}}_s^{(j)} \in \mathbb{R}^{n_s m \times md_j}$ is block-diagonal. Since by assumption $k_j(x, x') \leq 1, \forall j \leq p$ we have

$$\operatorname{tr}((\tilde{\boldsymbol{\Phi}}_s^{(j)})^T \tilde{\boldsymbol{\Phi}}_s^{(j)}) = \operatorname{tr}(\tilde{\boldsymbol{\Phi}}_s^{(j)}(\tilde{\boldsymbol{\Phi}}_s^{(j)})^T) = \sum_{s=1}^{m}\sum_{i=1}^{n_s} k_j(x_i^{(s)}, x_i^{(s)}) \leq N,$$

and

$$\begin{aligned}
\|(\tilde{\mathbf{\Phi}}_s^{(j)})^T \tilde{\mathbf{\Phi}}_s^{(j)}\|_2 &= \|\tilde{\mathbf{\Phi}}_s^{(j)}(\tilde{\mathbf{\Phi}}_s^{(j)})^T\|_2 \\
&\leq \max_{s\leq m} \operatorname{tr}(\mathbf{\Phi}_s^{(j)}(\mathbf{\Phi}_s^{(j)})^T) \\
&\leq \max_{s\leq m} \sum_{i=1}^{n_s} k_j(x,x) \\
&\leq \max_{s\leq m} n_s \\
&= n_1.
\end{aligned}$$

Since V is independent of $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_m$, we have by Corollary D.8 for all s and $v' \in C$

$$\mathbb{P}\left[\hat{J}_s = J^\star \mid V = v'\right] \geq 1 - p \exp\left(-\frac{N_s}{2n_1}\left(\sqrt{\frac{N_s}{s}}\frac{c_1 c_\kappa^2}{32\sigma} - 1\right)^2\right).$$

By union bound and since $\frac{N_s}{s}$ is non-increasing by assumption we have for $m_0 \leq m$

$$\begin{aligned}
\mathbb{P}\left[\forall m \geq s \geq m_0, \hat{J}_s = J^\star \mid V = v'\right] &\geq 1 - \sum_{s=m_0}^{m} p \exp\left(-\frac{N_s}{2n_1}\left(\sqrt{\frac{N_s}{s}}\frac{c_1 c_\kappa^2}{32\sigma} - 1\right)^2\right) \\
&\geq 1 - mp \exp\left(-\frac{N_{m_0}}{2n_1}\left(\sqrt{\frac{N_m}{m}}\frac{c_1 c_\kappa^2}{32\sigma} - 1\right)^2\right),
\end{aligned}$$

where we defined $N_s := \sum_{i=1}^{s} n_i$. If $m_0 \in \mathbb{N}$ is large enough such that

$$N_{m_0} \geq \frac{2n_1 \log(4mp/\delta)}{(\sqrt{N_m/m}\frac{c_1 c_\kappa^2}{32\sigma} - 1)^2},$$

then for all $v' \in C$

$$\mathbb{P}\left[\forall m \geq s \geq m_0, \hat{J}_s = J^\star \mid V = v'\right] \geq 1 - \delta/4.$$

By assumption we have

$$\mathbb{P}\left[V \in C\right] \geq 1 - \delta/4.$$

Because V is independent of the noise

$$\begin{aligned}
\mathbb{P}\left[\exists m \geq s \geq m_0, \hat{J}_s \neq J^\star\right] &= \int \mathbb{P}\left[\exists m \geq s \geq m_0, \hat{J}_s \neq J^\star \mid V = v\right] p_V(v) dv \\
&= \int_C \mathbb{P}\left[\exists m \geq s \geq m_0, \hat{J}_s \neq J^\star \mid V = v\right] p_V(v) dv + \\
&\quad \int_{C^c} \mathbb{P}\left[\exists m \geq s \geq m_0, \hat{J}_s \neq J^\star \mid V = v\right] p_V(v) dv \\
&\leq \mathbb{P}[V \in C^c] + \mathbb{P}[V \in C]\delta/2 \\
&\leq \delta/4 + \delta/4 = \delta/2.
\end{aligned}$$

For all tasks that happen after task m_0 we have jointly with probability at least $1 - \delta/2$ that $\hat{J}_s = J^\star$.
Denote by $r(k, s)$ the regret that the base bandit algorithm achieves after n interactions with kernel k in task s. By assumption

$\mathbb{P}[\sum_{s=m_0}^{m} r(k^\star, s) \leq R^\star(n, m-m_0)] \geq 1-\delta/2$. Denote by $\hat{k}_s$ the predicted kernel for task s. By union bound

$$\begin{aligned}\mathbb{P}\left[\sum_{s=m_0}^{m} r(\hat{k}_s, s) \leq \mathcal{O}(R^\star(m,n))\right] &\geq \mathbb{P}\Bigg[\forall m_0 \leq s \leq m, \hat{k}_s = k^\star \text{ and } \\ &\qquad \sum_{s=m_0}^{m} r(\hat{k}_s, s) \leq \mathcal{O}(R^\star(m-m_0, n))\Bigg] \\ &\geq 1 - \mathbb{P}\left[\sum_{s=m_0}^{m} r({k^\star}_s, s) > \mathcal{O}(R^\star(m-m_0,n))\right] \\ &\quad - \mathbb{P}\left[\exists m_0 \leq s \leq m, {k^\star}_s \neq k^\star\right] \\ &\geq 1-\delta\end{aligned}$$

Therefore it holds with probability at least $1-\delta$

$$R(m,n) \leq m_0 n L + L N_m + R^\star(m-m_0, n).$$

Here, the first term is an upper bound of the regret in the first m_0 tasks. The other terms are an upper bound on the reward for the other $m-m_0$ tasks. They can be divided into the regret obtained by forced exploration and the regret obtained by the base bandit task. By Lemma E.2 we know that the maximum instantaneous regret L is bounded by $2B$. Therefore

$$\begin{aligned}R(m,n) &\leq m_0 n L + L N_m + R^\star(m-m_0, n) \\ &\leq 2Bm_0 n + 2BN_m + R^\star(m,n).\end{aligned}$$

□

Lemma E.2. *Let k be a kernel with $k(x,x') \leq 1, \forall x, x' \in \mathcal{X}$ and let $f \in \mathcal{H}_k$ with $\|f\|_k \leq B$, then for all $\boldsymbol{x} \in \mathcal{X}$*

$$|f(\boldsymbol{x})| \leq B.$$

Proof. By the reproducing property, we have

$$\begin{aligned}|f(\boldsymbol{x})| &= |\langle f(\cdot), k(x,\cdot)\rangle_k| \\ &\leq \|f\|_{\mathcal{H}_k} k(x,x) \\ &\leq B.\end{aligned}$$

□

A clarification is due, regarding the exact number of exploratory steps taken. In the algorithm design and in the main text, we require that during every task s, purely exploratory actions are taken at every step i where $i \leq n_s$. The number of exploratory steps has to be an integer, while the proposed rate of $n_s = \sqrt{n}/s^{1/4}$ may not be an integer. Therefore, the $i \leq n_s$ condition implies that only the first $\lfloor n_s \rfloor$ steps will be exploratory. In our proofs so far, we have assumed that at least a total of $N_{m_0} = \sum_{s=1}^{m_0} n_s$ exploratory action are chosen, which may be well larger than $\sum_{s=1}^{m_0} \lfloor n_s \rfloor$. To resolve this gap, we accumulate the non-integer remainder $n_s - \lfloor n_s \rfloor$ in a variable r. Whenever r becomes larger than 1, we increase the number forced exploration queries by 1 to $\tilde{n}_s = \lfloor n_s \rfloor + \lfloor r \rfloor$. At every task s, we force exactly $\tilde{n}_s \in \mathbb{N}$ exploratory actions, where $(\tilde{n}_1, \dots, \tilde{n}_s)$ is calculated as described in Algorithm 5. Then to ensure that N_{m_0} exploratory datapoint are available, we calculate the smallest $\tilde{m}_0$ which satisfies:

$$\sum_{s=1}^{\tilde{m}_0} \tilde{n}_s \geq N_{m_0}$$

It is straightforward to show that by construction of Algorithm 5, $m_0 \leq \tilde{m}_0 \leq m_0 + 1$. In other words, by taking exploratory actions according to $\tilde{n}_s$ (which is an integer) we reqiure at most 1 additional task to fulfill the lower bound on the total number of required exploratory actions. In the next two corollaries we give a lower bound on the $\tilde{m}_0$ which satisfies the required dataset size N_{m_0}.

Algorithm 5 Forced Exploration Rate to Integer number of Exploratory Steps

Require: The sequence of $(n_1, \dots, n_m)$
  $r \leftarrow 0$ ▷ r is the sum of fractional residue
  for $s \in \{1, \dots, m\}$ **do**
    $r \leftarrow r + n_s - \lfloor n_s \rfloor$ ▷ Add the fractional part of n_s to the residue sum
    $\tilde{n}_s \leftarrow \lfloor n_s \rfloor + \lfloor r \rfloor$ ▷ If the residue sum is over 1, then add 1 to $\lfloor n_s \rfloor$
    $r \leftarrow r - \lfloor r \rfloor$
  end for
  Output: $(\tilde{n}_1, \dots, \tilde{n}_m)$

Corollary E.3. *Assume the setting of Theorem E.1. Set the rate*

$$n_s = \sqrt{n}$$

for all $s \in \mathbb{N}$, and choose the integer number of forced exploration steps according to Algorithm 5. Then, for all $0 < \delta < 1$, with probability at least $1 - \delta$

$$R(m, n) \leq \mathcal{O}\left(B \log(mp/\delta)\sqrt{n}\right) + 2mB\sqrt{n} + R^\star(m, n).$$

Proof. Taking actions at a $n_s = \sqrt{n}$ rate via Algorithm 5, we can ensure that after $\tilde{m}_0$ many tasks the condition of Theorem E.1 on N_{m_0} is met, where

$$\frac{2\log(4mp/\delta)}{(n^{1/4}\frac{c_1 c_\kappa^2}{32\sigma} - 1)^2} \leq \tilde{m}_0 \leq \frac{2\log(4mp/\delta)}{(n^{1/4}\frac{c_1 c_\kappa^2}{32\sigma} - 1)^2} + 1.$$

Then the proof directly from Theorem E.1 with $n_1 = \sqrt{n}$, $N_m = m\sqrt{n}$. □

Corollary E.4. *Assume the setting of Theorem E.1. Set the rate*

$$n_s = \frac{\sqrt{n}}{s^{1/4}}$$

for all $s \in \mathbb{N}$, and choose the explicit integer number of forced exploration steps according to Algorithm 5. Then, for all $0 < \delta < 1$, with probability at least $1 - \delta$

$$R(m, n) \leq \mathcal{O}\left(Bn^{1/3}\log^{3/4}(mp/\delta)m^{1/3} + B\sqrt{n}m^{3/4}\right) + R^\star(m, n).$$

Proof. We have

$$N_m = \sum_{s=1}^{m} \frac{\sqrt{n}}{s^{1/4}} = \Theta\left(\sqrt{n}m^{3/4}\right)$$

Choose

$$\tilde{m}_0 = \Theta\left(\frac{\log(4mp/\delta)}{(\sqrt{\frac{n}{m^{1/4}}}\frac{c_1 c_\kappa}{32\sigma} - 1)^2}\right)^{4/3}$$

and take exploratory actions according to Algorithm 5 then,

$$N_{m_0} \geq \frac{2n_1 \log(4mp/\delta)}{(\sqrt{N_m/m}\frac{c_1 c_\kappa^2}{32\sigma} - 1)^2}.$$

By Theorem E.1, since $n_1 = \sqrt{n}$, for all $0 < \delta < 1$, with probability at least $1 - \delta$ LIBO achieves

$$\begin{aligned} R(m, n) &\leq 2Bm_0 n + 2BN_m + R^\star(m, n) \\ &\leq \mathcal{O}\left(Bn^{1/3}\log^{3/4}(mp/\delta)m^{1/3} + B\sqrt{n}m^{3/4}\right) + R^\star(m, n). \end{aligned}$$

□

E.1 BACKGROUND ON GP-UCB

To solve task s, GP-UCB first constructs *confidence sets* for $f_s(\boldsymbol{x})$ based on the history $\{(\boldsymbol{x}_{s,t}, y_{s,t})_{t\leq i}\}$ to balance exploration and exploitation at any step i. For any $\boldsymbol{x} \in \mathcal{X}$, the set $\mathcal{C}_{i-1}(\boldsymbol{x})$ defines an interval to which $f(\boldsymbol{x})$ belongs with high probability such that,

$$\mathbb{P}\left(\forall \boldsymbol{x} \in \mathcal{X} : f(\boldsymbol{x}) \in \mathcal{C}_{i-1}(\boldsymbol{x})\right) \geq 1 - \delta.$$

Given a kernel k, GP-UCB builds sets of the form

$$\mathcal{C}_{i-1}(k;\boldsymbol{x}) = [\mu_{i-1}(k;\boldsymbol{x}) - \nu_i \sigma_{i-1}(k;\boldsymbol{x}),\ \mu_{i-1}(k;\boldsymbol{x}) + \nu_i \sigma_{i-1}(k;\boldsymbol{x})]$$

where the exploration coefficient ν_i depends on the desired confidence level $1-\delta$, and is often treated as a hyper-parameter of the algorithm. The functions μ_{i-1} and σ_{i-1} set the center and width of the confidence set as

$$\begin{aligned}\mu_{i-1}(k;\boldsymbol{x}) &= \boldsymbol{k}_{i-1}^T(\boldsymbol{x})(\boldsymbol{K}_{i-1} + \lambda_{\text{ucb}}^2 \boldsymbol{I})^{-1}\boldsymbol{y}_{i-1}\\ \sigma_{i-1}^2(k;\boldsymbol{x}) &= k(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}_{i-1}^T(\boldsymbol{x})(\boldsymbol{K}_{i-1} + \lambda_{\text{ucb}}^2 \boldsymbol{I})^{-1}\boldsymbol{k}_{i-1}(\boldsymbol{x})\end{aligned}$$

where λ_{ucb} is a regularizer, $\boldsymbol{y}_{i-1} = [y_{s,t}]_{t<i}$ is the vector of observed values, $\boldsymbol{k}_{i-1}(\boldsymbol{x}) = [k(\boldsymbol{x}, \boldsymbol{x}_{s,t})]_{t<i}$, and $\boldsymbol{K}_{i-1} = [k(\boldsymbol{x}_{s,t}, \boldsymbol{x}_{s,t'})]_{t,t'<i}$ is the kernel matrix. GP-UCB then chooses an action that maximizes the upper confidence bound, i.e.

$$\boldsymbol{x}_{s,i} = \arg\max_{\boldsymbol{x}\in\mathcal{X}} \mu_{i-1}(\boldsymbol{x}) + \nu_i \sigma_{i-1}(\boldsymbol{x}).$$

The acquisition function balances exploring uncertain actions and exploiting the gained information via parameter ν_i. Chowdhury and Gopalan [2017] show that following this policy, and using $k^\star$ as the kernel function, yields a regret of

$$R^\star(n) = \mathcal{O}\left(Bd^\star\sqrt{n}\log\tfrac{n}{d^\star} + \sqrt{nd^\star\log\tfrac{n}{d^\star}\log\tfrac{1}{\delta}}\right).$$

E.2 LIFELONG REGRET OF GP-UCB PAIRED WITH LIBO (PROOF OF COROLLARY 4.2)

Definition E.5 (maximum information gain)**.** *The maximum information gain after t observations of GP-UCB with kernel k and parameter λ_{ucb} is defined by*

$$\gamma_t(k) = \max_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_t\in\mathcal{X}} \frac{1}{2}\log\det\left(I + \lambda_{\text{ucb}}^{-2}K_t\right),$$

where

$$(K_t)_{ij} = (\boldsymbol{\Phi}^T\boldsymbol{\Phi})_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j).$$

Theorem E.6 (Theorem 3 of Chowdhury and Gopalan [2017])**.** *Let k be a kernel and $f \in \mathcal{H}_k$, where $\mathcal{H}_k$ is the RKHS corresponding to kernel k. Let $\delta \in (0,1)$, $\|f\|_k \leq B$ and assume the errors ϵ_t are conditionally σ-sub-Gaussian. Running GP-UCB with $\lambda_{\text{ucb}} = 1 + 2/n$ for n steps we have with probability at least $1-\delta$ that*

$$R(n) = \mathcal{O}(B\sqrt{n\gamma_n(k)} + \sqrt{n\gamma_n(k)(\gamma_n(k) + \log(1/\delta))})$$

Corollary E.7. *Let k_s be kernels and $f_s \in \mathcal{H}_{k_s}$, where $\mathcal{H}_{k_s}$ is the RKHS corresponding to kernel k_s. Let $\delta \in (0,1)$, $\|f_s\|_{k_s} \leq B$ and assume the errors $\epsilon_{s,t}$ are i.i.d. σ-sub-Gaussian. Assume further that k_s are $\sigma(\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_{s-1})$-measurable. Running GP-UCB with $\lambda_{\text{ucb}} = 1 + 2/n$ for m tasks, each with n steps, we have with probability at least $1-\delta$ that jointly for all $s \in \{1, \dots, m\}$*

$$R_s(n) = \mathcal{O}(B\sqrt{n\gamma_n(k)} + \sqrt{n\gamma_n(k)(\gamma_n(k) + \log(1/\delta))})$$

where $R_s(n)$ denotes the reward in task s after n interactions. In particular

$$R^\star(m,n) \leq \mathcal{O}\left(m\sqrt{n\gamma_n(k)}(B + \sqrt{\gamma_n(k)} + \log(1/\delta))\right).$$

Proof. We will adapt the proof of Theorem 1 in Chowdhury and Gopalan [2017].
Let $\epsilon_1^s, \dots, \epsilon_n^s$ be the noise of task s. Define a function

$$s(t) = \sum_{j=1}^{m} j \mathbb{1}_{\{(j-1)n+1 \le t \le jn\}}$$

and a filtration on $\{1, \dots, mn\}$

$$\mathcal{F}_t = \sigma(\epsilon_{1,1}, \dots, \epsilon_{1,n}, \epsilon_{2,1}, \dots, \epsilon_{2,n}, \dots, \epsilon_{s(t),1}, \dots, \epsilon_{s(t),t-(s(t)-1)n}).$$

Further define for task s a filtration on $\{1, \dots, n\}$

$$\mathcal{F}_t^s = \sigma(\epsilon_{s,1}, \dots, \epsilon_{s,t}).$$

Similar to the proof of Theorem 1 in Chowdhury and Gopalan [2017] define for $t \in \{1, \dots, n\}$, $g : \mathcal{X} \to \mathbb{R}$ and $l_1, \dots, l_n \in \mathbb{N}$

$$M_t^{g,n}(s) = \exp\left((\epsilon_{s,1:t})^T g_{1:t,l} - \frac{\sigma^2}{2} \|g_{1:t,l}\|_2^2 \right)$$

where

$$g_{1:t,l} := [g(\boldsymbol{x}_1) + l_1, \dots, g(\boldsymbol{x}_t) + l_t]^T.$$

Further let $N_1, \dots, N_n$ i.i.d. with distribution $\mathcal{N}(0, \kappa)$ and independent of $\mathcal{F}_n^s$ and let h_s be a random function distributed according to the Gaussian Process measure $GP_{\mathcal{X}}(0, k_s)$ and independent of F_n^s and $N_1, \dots, N_n$. Define

$$M_t(s) = \mathbb{E}[M_t^{h_s,N}(s) \mid \mathcal{F}_n^s].$$

Now by the proof of Theorem 1 of Chowdhury and Gopalan [2017] we have that for all $s \in \{1, \dots, m\}, t \in \{1, \dots, n\}$ and all stooping times τ_s with respect to the filtration $\mathcal{F}_t^s$

$$\mathbb{E}[M_{\tau_s}(s)] \le 1. \tag{E.1}$$

Given stopping times $\tau_1, \dots, \tau_m$ on $\mathcal{F}_t^1, \dots, \mathcal{F}_t^m$ we construct a stopping time τ on $\mathcal{F}_t$

$$\tau(\omega) = \min\{mn \ge t \ge 1 \mid \tau^{s(t)}(\omega) = t - (s(t) - 1)n\}. \tag{E.2}$$

We need to show that τ is a stopping time with respect to the filtration $\mathcal{F}_t$. We have

$$\{\omega \mid \tau(\omega) = t\} = \left(\bigcap_{s < s(t)} \{\omega \mid \tau_s(\omega) > n\} \right) \cap \{\omega \mid \tau^{s(t)}(\omega) = t - (s(t) - 1)n\}.$$

It holds that $\{\omega \mid \tau_s(\omega) > n\} = \{\omega \mid \tau_s(\omega) \le n\}^c \in \mathcal{F}_n^s \subset \mathcal{F}_{sn}$ and $\{\omega \mid \tau^{s(t)}(\omega) = (s(t)-1)n-t\} \in \mathcal{F}_{t-(s(t)-1)n}^{s(t)} \subset \mathcal{F}_t$. This impies that $\{\omega \mid \tau(\omega) = t\} \in \mathcal{F}_t$ and therefore τ is a stopping time with respect to $\mathcal{F}_t$. Define

$$M_t = M_{(s(t)-1)n-t}(s(t)).$$

We have that $M_t = M_{(s(t)-1)n-t}(s(t))$ is measurable with respect to $\mathcal{F}_{(s(t)-1)n-t}^{s(t)} \subset \mathcal{F}_t$, which means M_t is $\mathcal{F}_t$-adapted. Let τ be a stopping time constructed as in Equation E.2. Then by Equation E.1

$$\mathbb{E}[M_\tau] \le 1.$$

Define for $t \in \{1, \dots, n\}$ and $s \in \{1, \dots, m\}$

$$B_t^s(\delta) = \left\{ \omega \mid \|\epsilon_{s,1:t}\|^2_{((K_t^s + \kappa I)^{-1} + I)^{-1}} > 2 \log \left(\sqrt{\det((1+\kappa)I + K_t^s)} / \delta \right) \right\},$$

where K_t^s the design matrix for task s. Further define

$$\tau^s(\omega) = \min\{t \in \{1, \ldots, n\} \mid \omega \in B_t^s(\delta)\}$$

and let τ be the corresponding stopping time on $\mathcal{F}_t$. It holds by the proof of Theorem 1 of Chowdhury and Gopalan [2017] that

$$M_t(s) = \frac{\exp\left(\frac{1}{2}\|\epsilon_{s,1:t}\|^2_{((K_t^s+\kappa I)^{-1}+I)^{-1}}\right)}{\sqrt{\det((1+\kappa)I + K_t^s)}}$$

and therefore

$$M_t = \frac{\exp\left(\frac{1}{2}\|\epsilon_{s(t),1:t-(s(t)-1)n}\|^2_{((K^{s(t)}_{t-(s(t)-1)n}+\kappa I)^{-1}+I)^{-1}}\right)}{\sqrt{\det((1+\kappa)I + K^{s(t)}_{t-(s(t)-1)n})}}.$$

Putting things together yields

$$\begin{aligned}
\mathbb{P}\left[\bigcup_{s\leq m, t\leq n} B_t^s(\delta)\right] &= \mathbb{P}\left[\tau \leq mn\right] \\
&= \mathbb{P}\Bigg[\tau \leq mn, \|\epsilon_{s(\tau),1:\tau-(s(\tau)-1)n}\|^2_{((K_\tau^{s(\tau)}+\kappa I)^{-1}+I)^{-1}} > \\
&\qquad 2\log\left(\sqrt{\det((1+\kappa)I + K^{s(\tau)}_{\tau-(s(\tau)-1)n}}/\delta\right)\Bigg] \\
&= \mathbb{P}\left[\tau \leq mn, M_\tau > 1/\delta\right] \\
&\leq \mathbb{P}\left[M_\tau > 1/\delta\right] \\
&\leq \mathbb{E}[M_\tau]\delta = \delta.
\end{aligned}$$

Now follow the steps of the proof of Theorem 2 of Chowdhury and Gopalan [2017] and the claim follows. □

Lemma E.8. *Let $k : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ be a kernel with $d^{(k)} \in \mathbb{N}$ dimensional feature map and assume $k(x, x') = \boldsymbol{\phi}(\boldsymbol{x})^T\boldsymbol{\phi}(x') \leq 1, \forall x, x' \in \mathcal{X}$. Then the maximum information gain of GP-UCB with kernel k and regularization parameter λ_{ucb} satisfies*

$$\gamma_n(k) \leq \frac{1}{2}d^{(k)}\log(1 + \frac{\lambda_{\text{ucb}}^{-2}n}{d^{(k)}}).$$

Proof. This proofs follows the arguments of Vakili et al. [2021] and Kassraie et al. [2022]. We have that $K_n = \boldsymbol{\Phi}_n\boldsymbol{\Phi}_n^T$ and by the Weinstein-Aronszajn identity

$$\begin{aligned}
\frac{1}{2}\log\det(I_n + \lambda_{\text{ucb}}^{-2}K_n) &= \frac{1}{2}\log\det(I_{d^{(k)}} + \lambda_{\text{ucb}}^{-2}\boldsymbol{\Phi}_n^T\boldsymbol{\Phi}_n) \\
&\leq \frac{1}{2}d^{(k)}\log\left(\text{tr}(I + \lambda_{\text{ucb}}^{-2}\boldsymbol{\Phi}_n^T\boldsymbol{\Phi}_n)/d^{(k)}\right) \\
&\leq \frac{1}{2}d^{(k)}\log\left(1 + \frac{\lambda_{\text{ucb}}^{-2}}{d^{(k)}}\text{tr}(\boldsymbol{\Phi}_n^T\boldsymbol{\Phi}_n)\right).
\end{aligned}$$

Now

$$\begin{aligned}
\text{tr}(\boldsymbol{\Phi}_n^T\boldsymbol{\Phi}_n) &= \sum_{i=1}^n \text{tr}(\boldsymbol{\phi}(\boldsymbol{x}_i)\boldsymbol{\phi}(\boldsymbol{x}_i)^T) \\
&= \sum_{i=1}^n \text{tr}(\boldsymbol{\phi}(\boldsymbol{x}_i)^T\boldsymbol{\phi}(\boldsymbol{x}_i)) \\
&= n
\end{aligned}$$

and therefore

$$\frac{1}{2}\log\det(I_n+\lambda_{\mathrm{ucb}}^{-2}K_n)\le\frac{1}{2}d^{(k)}\log\left(1+\frac{\lambda_{\mathrm{ucb}}^{-2}n}{d^{(k)}}\right).$$

□

Corollary E.9. *Assume we are in the setting of Corollary E.3 with GP-UCB as the base bandit algorithm and $\lambda_{\mathrm{ucb}}=1+2/n$. Then, for all $0<\delta<1$, with probability at least $1-\delta$,*

$$R(m,n)=\mathcal{O}\left(Bmd^\star\sqrt{n}\log\tfrac{n}{d^\star}+m\sqrt{nd^\star\log\tfrac{n}{d^\star}\log\tfrac{1}{\delta}}+B\sqrt{n}(m+\log(mp/\delta))\right).$$

Proof. By Corollary E.7 and Lemma E.8 we have with high probability

$$\begin{aligned}R^\star(m,n)&=\mathcal{O}\left(m\sqrt{n}\left((B+\log(1/\delta))\sqrt{\frac{1}{2}d^\star\log\left(1+\frac{n^3}{d^\star(n+2)^2}\right)}+\frac{1}{2}d^\star\log\left(1+\frac{n^3}{d^\star(n+2)^2}\right)\right)\right)\\&=\mathcal{O}\left(Bmd^\star\sqrt{n}\log\tfrac{n}{d^\star}+m\sqrt{nd^\star\log\tfrac{n}{d^\star}\log\tfrac{1}{\delta}}\right)\end{aligned}$$

where $d^\star:=d^{(k^\star)}=\sum_{j\in J^\star}d_j$. And using Corollary E.3 we have with high probability

$$\begin{aligned}R(m,n)&\le\mathcal{O}\left(B\log(mp/\delta)\sqrt{n}+mB\sqrt{n}\right)+R^\star(m,n)\\&=\mathcal{O}\left(Bmd^\star\sqrt{n}\log\tfrac{n}{d^\star}+m\sqrt{nd^\star\log\tfrac{n}{d^\star}\log\tfrac{1}{\delta}}+Bm\sqrt{n}\right).\end{aligned}$$

□

Corollary E.10. *Assume we are in the setting of Corollary E.4 with GP-UCB as the base bandit algorithm and $\lambda_{\mathrm{ucb}}=1+2/n$. Then, for all $0<\delta<1$, with probability at least $1-\delta$,*

$$R(m,n)=\mathcal{O}\left(Bmd^\star\sqrt{n}\log\tfrac{n}{d^\star}+m\sqrt{nd^\star\log\tfrac{n}{d^\star}\log\tfrac{1}{\delta}}+Bn^{1/3}\log^{3/4}(mp/\delta)m^{1/3}+B\sqrt{n}m^{3/4}\right).$$

Proof. The proof is the same as the proof for Corollary E.9, except that we use Corollary E.4 in place of Corollary E.3. □

Remark E.11. *Compare the results of Theorem 4.1 with the default alternative: not learning k and just setting $\hat{k}=\sum_{j=1}^{p}\frac{1}{p}k_j$. We would then only get a bound of the form*

$$R(m,n)\le\mathcal{O}\left(m\hat{B}\sqrt{n}\log(n)d\right),$$

where $B=\|f\|_{\hat{k}}=\frac{p}{s^\star}B$ and $d=\sum_{j=1}^{p}d_j\ge n$, which is not sublinear in n.

E.3 FORCED EXPLORATION LOWER BOUND (PROOF OF PROPOSITION 4.3)

Assumption E.12. *Assume there exists $c_c,c_{od}>0$*

$$\frac{m}{N}(\mathbf{\Phi}^T\mathbf{\Phi})_{i,i}\ge c_d,\qquad\forall i\in\{1,\dots,md\}$$

and

$$\frac{m}{N}(\mathbf{\Phi}^T\mathbf{\Phi})_{i,j}<c_{od},\qquad\forall i\ne j\in\{1,\dots,md\}.$$

Lemma E.13. *Let Assumption E.12 be satisfied. Then $\kappa\ge\sqrt{c_d/s^\star-5c_{od}}$.*

Proof. Let $(b, J) \in S$. We have by definition of S

$$\left(\frac{\sqrt{m}}{\sqrt{N}}\frac{\|\boldsymbol{\Phi}\boldsymbol{b}\|_2}{\sum_{j\in J}\|b^{(j)}\|_2}\right)^2 = \frac{m}{N}\frac{b^T(\boldsymbol{\Phi}^T\boldsymbol{\Phi})b}{\left(\sum_{j\in J}\|b^{(j)}\|_2\right)^2}$$

$$\geq \frac{m}{N}\frac{\sum_{s=1}^m\sum_{i,j\in J}\boldsymbol{b}_s^{(i)}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})_{i,j}\boldsymbol{b}_s^{(j)}}{\left(\sum_{j\in J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)^2}$$

$$-4\frac{m}{N}\frac{\sum_{s=1}^m\sum_{i\in J, j\in J^c}|\boldsymbol{b}_s^{(i)}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})_{i,j}\boldsymbol{b}_s^{(j)}|}{\left(\sum_{j\in J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)\left(\sum_{j\notin J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)}$$

By Assumption E.12

$$\frac{m}{N}\frac{\sum_{s=1}^m\sum_{i,j\in J}\boldsymbol{b}_s^{(i)}(\boldsymbol{\Phi}^T\boldsymbol{\Phi})_{i,j}\boldsymbol{b}_s^{(j)}}{\left(\sum_{j\in J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)^2} \geq \sum_{s=1}^m\sum_{i\in J}\frac{\boldsymbol{b}_s^{(i)}c_2\boldsymbol{b}_s^{(i)}}{\left(\sum_{j\in J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)^2}$$

$$-\frac{\sum_{s=1}^m\sum_{i\neq j, i,j\in J}|\boldsymbol{b}_s^{(i)}c_{od}\boldsymbol{b}_s^{(j)}|}{\left(\sum_{j\in J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)^2}.$$

Since for $q > 0$ using $\|\cdot\|_1 \leq \sqrt{s}\|\cdot\|_2$

$$\sum_{s=1}^m\sum_{i\in J}\frac{\boldsymbol{b}_s^{(i)}c_2\boldsymbol{b}_s^{(i)}}{\left(\sum_{j\in J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)^2} \geq c_2\sum_{s=1}^m\sum_{i\in J}\frac{(\boldsymbol{b}_s^{(i)})^2}{s^\star\left(\sqrt{\sum_{j\in J}\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)^2} = \frac{c_2}{s^\star}$$

and using Cauchy-Schwarz to prove

$$\sum_{k,l}(x_ky_l)^2 \geq \sum_{k,l}(x_ky_l)(y_kx_l)$$

which implies

$$\sqrt{\sum_{k,l}(x_ky_l)^2} \geq \sum_l |x_ly_l|$$

we get

$$\frac{\sum_{s=1}^m\sum_{i\neq j, i,j\in J}|\boldsymbol{b}_s^{(i)}c_{od}\boldsymbol{b}_s^{(j)}|}{\left(\sum_{j\in J}\sqrt{\sum_{s=1}^m(\boldsymbol{b}_s^{(j)})^2}\right)^2} = c_{od}\frac{\sum_{s=1}^m\sum_{i\neq j, i,j\in J}|\boldsymbol{b}_s^{(i)}||\boldsymbol{b}_s^{(j)}|}{\sum_{i,j\in J}\sqrt{\sum_{k,l}(\boldsymbol{b}_l^{(i)}\boldsymbol{b}_k^{(j)})^2}}$$

$$\leq c_{od}\frac{\sum_{s=1}^m\sum_{i\neq j, i,j\in J}|\boldsymbol{b}_s^{(i)}||\boldsymbol{b}_s^{(j)}|}{\sum_{i,j\in J}\sum_{s=1}^m|\boldsymbol{b}_s^{(i)}||\boldsymbol{b}_s^{(j)}|}$$

$$\leq c_{od}.$$

Also

$$\frac{\sum_{s=1}^m \sum_{i\in J, j\in J^c} |\boldsymbol{b}_s^{(i)} c_{od} \boldsymbol{b}_s^{(j)}|}{\left(\sum_{j\in J} \sqrt{\sum_{s=1}^m (\boldsymbol{b}_s^{(j)})^2}\right)\left(\sum_{j\notin J} \sqrt{\sum_{s=1}^m (\boldsymbol{b}_s^{(j)})^2}\right)}$$
$$= c_{od} \frac{\sum_{s=1}^m \sum_{i\in J, j\in J^c} |\boldsymbol{b}_s^{(i)}||\boldsymbol{b}_s^{(j)}|}{\sum_{i\in J, j\in J^c} \sqrt{\sum_{k,l} (\boldsymbol{b}_l^{(i)} \boldsymbol{b}_k^{(j)})^2}}$$
$$= c_{od} \frac{\sum_{s=1}^m \sum_{i\in J, j\in J^c} |\boldsymbol{b}_s^{(i)}||\boldsymbol{b}_s^{(j)}|}{\sum_{i\in J, j\in J^c} \sum_{s=1}^m |\boldsymbol{b}_s^{(i)} \boldsymbol{b}_s^{(j)}|}$$
$$= c_{od}.$$

Therefore

$$\frac{\sqrt{m}}{\sqrt{N}} \frac{\|\boldsymbol{\Phi b}\|_2}{\sum_{j\in J} \|b^{(j)}\|_2} \geq \sqrt{c_2/s - 5c_{od}}.$$

□

Proposition E.14. *Let μ be the Lebesgue measure and $d = p$. Assume that $\phi_i \in L_\mu^2(\mathcal{X})$, $i \in \{1, \ldots, p\}$ are orthogonal and satisfy $\|\phi_i\|_{L_\mu^2(\mathcal{X})}/Vol(\mathcal{X}) \geq z$, for all $i \in \{1, \ldots, p\}$. Assume also that $k_i(x, x) = \boldsymbol{\phi}_i(\boldsymbol{x})^2 \leq 1$ for all $\boldsymbol{x} \in \mathcal{X}$. Choose $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ i.i.d. uniformly from $\mathcal{X}$ and let*

$$\boldsymbol{\Phi}_s \coloneqq \begin{bmatrix} \boldsymbol{\phi}(\boldsymbol{x}_1) & \ldots & \boldsymbol{\phi}(\boldsymbol{x}_n) \end{bmatrix}^T \in \mathbb{R}^{n\times d} \qquad \forall s \leq m.$$

Then with probability at least $1 - \delta$ Assumption E.12 is satisfied with

$$c_d = \left(z - \sqrt{\frac{1}{2n} \log(4d/\delta)}\right)$$

and

$$c_{od} = \sqrt{\frac{2}{n} \log\left(\frac{4d^2}{\delta}\right)}.$$

Proof. For the second, let X be a random variable uniformly distributed on $\mathcal{X}$ and denote by $v_i \coloneqq \boldsymbol{\phi}_i(\boldsymbol{x}_{1:n})$ the ith column of $\boldsymbol{\Phi}_s$. It holds that

$$\mathbb{E}[\boldsymbol{\phi}_i(\boldsymbol{x})^2] = \frac{1}{\mathrm{Vol}(\mathcal{X})} \int_{\mathcal{X}} \boldsymbol{\phi}_i(\boldsymbol{x})^2 \mathrm{d}\mu(\boldsymbol{x}) \geq z, \quad \forall i \leq d.$$

Therefore

$$\mathbb{E}[\|v_i\|_2^2] = \mathbb{E}\left[\sum_{i=1}^n \boldsymbol{\phi}_i(\boldsymbol{x}_i)^2\right] \geq nz.$$

By union bound and Höffding's inequality

$$\mathbb{P}\left[\exists i \leq d, \left|\|v_i\|_2^2 - \mathbb{E}\left[\|v_i\|_2^2\right]\right| \geq \epsilon\right] \leq 2d \exp(-\frac{2\epsilon^2}{n})$$

or

$$\mathbb{P}\left[\exists i \leq d, \left|\|v_i\|_2^2 - \mathbb{E}\left[\|v_i\|_2^2\right]\right| \geq \sqrt{\frac{n}{2} \log(\frac{4d}{\delta})}\right] \leq \delta.$$

Therefore with probability at least $1 - \delta/2$ for all $i \leq d$

$$\|v_i\|_2^2 \geq \mathbb{E}\left[\|v_i\|_2^2\right] - \sqrt{\frac{n}{2} \log(\frac{4d}{\delta})} \geq nz - \sqrt{\frac{n}{2} \log(\frac{4d}{\delta})}.$$

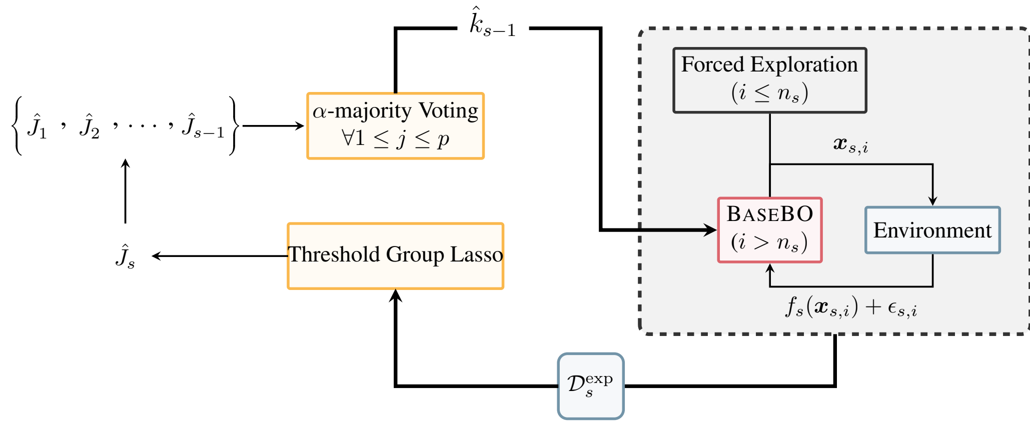

Figure 1: F-LiBO visualized. The yellow boxes corresponds to modules of F-Meta-KGL.

Further, for $i \neq j$

$$\mathbb{E}[\boldsymbol{\phi}_i(\boldsymbol{x})\boldsymbol{\phi}_j(\boldsymbol{x})] = \frac{1}{\text{Vol}(\mathcal{X})}\int_{\mathcal{X}} \boldsymbol{\phi}_i(\boldsymbol{x})\boldsymbol{\phi}_j(\boldsymbol{x})\text{d}\mu(\boldsymbol{x}) = 0,$$

since $\boldsymbol{\phi}_i$ and $\boldsymbol{\phi}_j$ are orthogonal in $L^2_\mu(\mathcal{X})$. By assumption $\boldsymbol{\phi}_i(\boldsymbol{x}) \le 1, \forall i \le d, \forall \boldsymbol{x} \in \mathcal{X}$ and by Höffding's inequality

$$\mathbb{P}[|\langle v_i, v_j\rangle| \ge \epsilon] \le 2\exp(-\frac{\epsilon^2}{2n}).$$

and therefore for $0 \le \delta \le 1$

$$\mathbb{P}\left[\exists i \neq j, |\langle v_i, v_j\rangle| \ge \sqrt{2n\log\left(\frac{4d^2}{\delta}\right)}\right] \le \delta/2.$$

We derived that with probability at least $1-\delta$

$$(\boldsymbol{\Phi}^T\boldsymbol{\Phi})_{ii}/n \ge c_2 = \left(z - \sqrt{\log(4d/\delta)/2n}\right)$$

and for $i \neq j$

$$(\boldsymbol{\Phi}^T\boldsymbol{\Phi})_{ij}/n < \sqrt{\frac{2}{n}\log\left(\frac{4d^2}{\delta}\right)}.$$

□

Corollary E.15. *Assume the setting of Proposition E.14. Then*

$$\kappa \ge \sqrt{z/s^\star - \sqrt{\frac{\log(4d/\delta)/(2s^\star) + 50\log(4d^2/\delta)}{n}}} = \mathcal{O}(1).$$

F FEDERATED ANALYSIS (PROOF OF THEOREM 5.1)

Recall that in the federated setting, each client minimizes the following loss locally.

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{s,\text{prvt}} &\coloneqq \arg\min_{\boldsymbol{\beta}_s \in \mathbb{R}^d} \mathcal{L}\left(\boldsymbol{\beta}_s; \mathcal{D}_s^{\text{exp}}\right) \\ &= \arg\min_{\boldsymbol{\beta}_s \in \mathbb{R}^d} \frac{1}{n_s}\|\boldsymbol{y}_s - \boldsymbol{\Phi}_s\boldsymbol{\beta}_s\|_2^2 + \lambda\sum_{j=1}^p \|\boldsymbol{\beta}_s{}^{(j)}\|_2.\end{aligned} \tag{F.1}$$

In this section, for simplicity we refer to the solution as $\hat{\boldsymbol{\beta}}_s$. We may further omit the subscript s, whenever it can be determined from the context. For our federated analysis, we require a slightly stronger version of the Beta-min assumption.

Assumption F.1 (Beta-min federated). *Assume there exists $c_{1,\mathrm{f}} > 0$ such that for all $s \leq m$ and $j \in J^\star$*

$$\|\boldsymbol{\beta}_s^{\star(j)}\|_2 \geq c_{1,\mathrm{f}}.$$

Remark F.2. *Note that Assumption F.1 implies Assumption 3.1.*

F.1 CONSISTENCY OF THE META-LEARNED KERNEL

In this section we prove the equivalent of Theorem 3.3 in the federated setting.

Theorem F.3 (Consistency of F-META-KGL). *Let $\omega \in (0, c_{1,\mathrm{f}})$ and $\bar{\omega} := \min\{\omega, c_{1,\mathrm{f}} - \omega\}$. Let Assumption F.1 hold. Let $n_s = \underline{n}, \forall s \leq m$ and assume $\underline{n}$ is large enough to satisfy $\bar{\omega} > (\sqrt{\log(p/\bar{\alpha})} + 1)32\sigma/(\sqrt{\underline{n}}c_\kappa^2)$, where $\bar{\alpha} := \max\{\alpha, 1-\alpha\}$. Assume that $\boldsymbol{\Phi}_s \in \mathbb{R}^{n \times d}$ satisfy Assumption 3.2 with c_κ for $s = 1, \ldots, m$. Let $\hat{\boldsymbol{\beta}}$ be a solution of Equation F.1 with regularization parameter $\lambda = \bar{\omega}c_\kappa^2/8$. Then $\hat{J}_\mathrm{f}$ is a consistent estimator in $\underline{n}$ and m, that is*

$$\lim_{\underline{n} \to \infty} \mathbb{P}\left[\hat{J}_\mathrm{f} = J^\star\right] = 1 \qquad \textit{and} \qquad \lim_{m \to \infty} \mathbb{P}\left[\hat{J}_\mathrm{f} = J^\star\right] = 1.$$

We start by proving the necessary lemmas.

Lemma F.4. *Let $\hat{\boldsymbol{\beta}}$ is a solution of equation F.1 and*

$$\lambda \leq \frac{\bar{\omega}\kappa^2}{8},$$

where $\bar{\omega} := \min\{\omega, c_{1,\mathrm{f}} - \omega\}$ for $0 < \omega < c_{1,\mathrm{f}}$. Then we have for $\frac{\lambda n_s}{4\sigma} > \sqrt{\mathrm{tr}(\Phi_s^{(j)}(\Phi_s^{(j)})^T)}$ with probability at least

$$1 - p \max_{j \leq p} \exp\left(-\frac{\left(\frac{\lambda n_s}{4\sigma} - \sqrt{\mathrm{tr}(\Phi_s^{(j)}(\Phi_s^{(j)})^T)}\right)^2}{2\|\Phi_s^{(j)}(\Phi_s^{(j)})^T\|_2}\right)$$

that

$$\sum_{j=1}^{p} \|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}_s^{\star(j)}\|_2 \leq \bar{\omega}. \tag{F.2}$$

If additionally Assumption F.1 holds, then we have with the same probability for

$$\hat{J}_{s,\mathrm{f}} = \left\{ j \in \{1, \ldots, p\} \,\middle|\, \|\hat{\boldsymbol{\beta}}_s^{(j)}\|_2 > \omega \right\}$$

that

$$\hat{J}_{s,\mathrm{f}} = J^\star.$$

Proof. Follows directly from Corollary D.8 with $m = 1$. $\square$

Lemma F.5 (Chernoff-Höffding bound). *Let $X_1, \ldots, X_n$ be i.i.d Bernoulli random variables with $\mathbb{E}[X_i] = p_i$. Define $p := \frac{1}{n}\sum_{i=1}^{n} p_i$, then for $t < np$,*

$$\begin{aligned}
\mathbb{P}\left[\sum_{i=1}^{n} X_i \leq t\right] &\leq \exp\left(-n\left(\frac{t}{n}\log\left(\frac{t}{np}\right) + (1 - t/n)\log\left(\frac{1 - t/n}{1-p}\right)\right)\right) \\
&\leq \exp\left(-n\frac{(p - t/n)^2}{2p(1-p)}\right).
\end{aligned}$$

Lemma F.6. *Let $0 < w < c_{1,\mathrm{f}}$ and let $\hat{\boldsymbol{\beta}}_s$ be the solution of equation F.1 for tasks $s \leq m$ and $\lambda \leq \frac{\bar{\omega}\kappa^2}{8}$. Define for $j \in \{1, \ldots, p\}$*

$$Q_j = \left\{s \in \{1, \ldots, m\} \mid \|\hat{\boldsymbol{\beta}}_s^{(j)}\|_2 > w\right\}$$

and for $\alpha > 0$

$$\hat{J}_{\mathrm{f}} = \left\{j \in \{1, \ldots, p\} \mid |Q_j| > m\alpha\right\}. \tag{F.3}$$

Define for $s \in \{1, \ldots, m\}$

$$v_s := 1 - p \max_{j \leq p} \exp\left(-\frac{\left(\frac{\lambda n_s}{4\sigma} - \sqrt{\mathrm{tr}(\Phi_s^{(j)}(\Phi_s^{(j)})^T)}\right)^2}{2\|\Phi_s^{(j)}(\Phi_s^{(j)})^T\|_2}\right)$$

and

$$v := \frac{1}{m}\sum_{s=1}^{m} v_s.$$

Assume that $\frac{\lambda n_s}{4\sigma} > \sqrt{\mathrm{tr}(\Phi_s^{(j)}(\Phi_s^{(j)})^T)}, \forall s \leq m$ and $v > \bar{\alpha} := \min\{\alpha, 1-\alpha\}$. Then

$$\mathbb{P}\left[J^\star = \hat{J}_{\mathrm{f}}\right] \geq 1 - p\exp\left(-m\frac{(v-\bar{\alpha})^2}{2v(1-v)}\right).$$

Proof. Recall that $\boldsymbol{\epsilon}_s = [\epsilon_{s,i}]_{i=1}^n$. Since $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_m$ are independent,

$$\mathbb{1}_{\left\{\sum_{j=1}^p \|\hat{\boldsymbol{\beta}}_1^{(j)} - \boldsymbol{\beta}_1^{\star(j)}\|_2 \leq \bar{\omega}\right\}}, \ldots, \mathbb{1}_{\left\{\sum_{j=1}^p \|\hat{\boldsymbol{\beta}}_m^{(j)} - \boldsymbol{\beta}_m^{\star(j)}\|_2 \leq \bar{\omega}\right\}}$$

are independent and Bernoulli distributed with coefficient

$$\mathbb{P}\left[\sum_{s=1}^{p} \|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}_s^{\star(j)}\|_2 \leq \bar{\omega}\right] \geq v_s,$$

where we used Lemma F.4 and set $\bar{\omega} := \min\{\omega, c_{1,\mathrm{f}} - \omega\}$. If $j \in J^\star$ and $\|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}_s^{\star(j)}\|_2 < \bar{\omega}$, then by Assumption F.1

$$\|\hat{\boldsymbol{\beta}}_s^{(j)}\|_2 > c_{1,\mathrm{f}} - \|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}^{\star(j)}\|_2 \geq c_{1,\mathrm{f}} - \bar{\omega} \geq \omega,$$

which implies $J^\star \subset \hat{J}$. If $j \notin J^\star$ and $\|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}_s^{\star(j)}\|_2 < \bar{\omega}$, then

$$\|\hat{\boldsymbol{\beta}}_s^{(j)}\|_2 \leq \|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}_s^{\star(j)}\|_2 + \|\boldsymbol{\beta}_s^{\star(j)}\|_2 \leq \bar{\omega} \leq \omega.$$

We have by Lemma F.5 and for $v > \bar{\alpha}$,

$$\begin{aligned}
\mathbb{P}\left[\hat{J}_{\mathrm{f}} = J^\star\right] &\geq \mathbb{P}\left[\forall j \in J^\star, |Q_j| \geq m/x; \forall j \notin J^\star, |Q_j| < m/x\right] \\
&\geq \mathbb{P}\left[\forall j \notin J^\star, \sum_{s=1}^m \mathbb{1}_{\left\{\|\hat{\boldsymbol{\beta}}_s^{(j)}\|_2 \leq w\right\}} \geq m/x; \forall j \in J^\star, \sum_{s=1}^m \mathbb{1}_{\left\{\|\hat{\boldsymbol{\beta}}_s^{(j)}\|_2 > w\right\}} \geq m/x\right] \\
&\geq \mathbb{P}\left[\forall j \in \{1, \ldots, p\}, \sum_{s=1}^m \mathbb{1}_{\left\{\|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}_s^{\star(j)}\|_2 \leq \bar{\omega}\right\}} \geq m\min\{\alpha, 1-\alpha\}\right] \\
&\geq \mathbb{P}\left[\sum_{s=1}^m \mathbb{1}_{\left\{\sum_{j=1}^p \|\hat{\boldsymbol{\beta}}_s^{(j)} - \boldsymbol{\beta}_s^{\star(j)}\|_2 \leq \bar{\omega}\right\}} \geq m\bar{\alpha}\right] \\
&\geq 1 - \exp\left(-m\left(\bar{\alpha}\log\left(\frac{\bar{\alpha}}{v}\right) + (1-\bar{\alpha})\log\left(\frac{1-\bar{\alpha}}{1-v}\right)\right)\right) \\
&\geq 1 - \exp\left(-m\frac{(v-\bar{\alpha})^2}{2v(1-v)}\right).
\end{aligned} \tag{F.4}$$

□

Proof of Theorem F.3. Assume the setting of Lemma F.6 and that there exists $c_\kappa > 0$ such that $\kappa \geq c_\kappa$. Set $\lambda = \frac{\bar{\omega} c_\kappa^2}{8}$, $n_s = \underline{n}, \forall s \leq m$ and assume $\frac{\lambda\sqrt{\underline{n}}}{4\sigma} > 1$ and $v = 1 - p\exp(-(\lambda\sqrt{\underline{n}}/4\sigma - 1)^2/2) > \bar{\alpha}$.

Note that $\Phi_s^{(j)} \in \mathbb{R}^{N \times md_j}$ is block-diagonal. Since by assumption $k_j(x, x') \leq 1, \forall j \leq p$, we have

$$\|(\Phi_s^{(j)})^T \Phi_s^{(j)}\|_2 \leq \operatorname{tr}((\Phi_s^{(j)})^T \Phi_s^{(j)}) = \operatorname{tr}(\Phi_s^{(j)} (\Phi_s^{(j)})^T) = \sum_{i=1}^{\underline{n}} k_j\left(x_i^{(s)}, x_i^{(s)}\right) \leq \underline{n}.$$

Lemma F.6 yields the result. □

F.2 LIFELONG REGRET OF F-LIBO (PROOF OF THEOREM 5.1)

We start by stating Theorem 5.1 more rigorously.

Theorem F.7. *Assume that the true reward functions $f_1, \ldots, f_m$ satisfy $\|f_i\|_{\mathcal{H}_{k^\star}} \leq B$ for some constant $B > 0$. Let $\bar{n}$ be the number of times forced exploration is used in each task. Let ν be a distribution on $\mathcal{X}^{\bar{n}m}$ independent of $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_m$. Let $V \sim \nu$ be the random vector used for forced exploration. Let $\tilde{\Phi}_s \in \mathbb{R}^{\bar{n} \times md}$ be the data matrix obtained by forced exploration in task s. Set $\lambda = \bar{\omega} c_\kappa^2/8$. Assume the forced exploration distribution ν and $\{k_j\}_{j \leq p}$ are such that, with probability at least $1 - \delta/4$, there exists $c_\kappa > 0$ such that $\kappa(\tilde{\Phi}_s) \geq c_\kappa, \forall s \leq m$. Assume further that the base bandit algorithm using the true kernel function achieves on m tasks with independent noise with probability at least $1 - \delta/2$ cumulative regret lower than $R^\star(n, m)$. Define*

$$v := 1 - p\exp\left(-\frac{1}{2}\left(\frac{\bar{\omega} c_\kappa^2 \sqrt{\bar{n}}}{32\sigma} - 1\right)^2\right).$$

and assume for all $s \leq m$

$$v \geq \bar{\alpha}, \qquad\qquad \frac{\bar{\omega} c_\kappa^2 \sqrt{\bar{n}}}{32\sigma} > 1. \tag{F.5}$$

Then with probability at least $1 - \delta$, LIBO *(using* F-META-KGL *to predict the kernel) achieves*

$$R(m, n) \leq \mathcal{O}\left(Bn\log(mp/\delta)/\bar{n} + Bm\bar{n}\right) + R^\star(n, m).$$

Proof. Similar to the proof of Theorem E.1 we have by Equation F.4 for all s and $v' \in C$

$$\mathbb{P}\left[\hat{J}_s = J^\star \mid V = v'\right] \geq 1 - p\exp\left(-s\left(\bar{\alpha}\log\left(\frac{\bar{\alpha}}{v}\right) + (1 - \bar{\alpha})\log\left(\frac{1 - \bar{\alpha}}{1 - v}\right)\right)\right).$$

By union bound we have for $m_0 \leq m$

$$\begin{aligned}
&\mathbb{P}\Big[\forall m \geq s \geq m_0, \hat{J}_s = J^\star \mid V = v'\Big] \\
&\qquad\qquad \geq 1 - \sum_{s=m_0}^{m} p\exp\left(-s\left(\bar{\alpha}\log\left(\frac{\bar{\alpha}}{v}\right) + (1 - \bar{\alpha})\log\left(\frac{1 - \bar{\alpha}}{1 - v}\right)\right)\right) \\
&\qquad\qquad \geq 1 - mp\exp\left(-m_0\left(q - (1 - \bar{\alpha})\log\left(1 - v\right)\right)\right),
\end{aligned}$$

where $q := \bar{\alpha}\log(\bar{\alpha}) + (1 - \bar{\alpha})\log(1 - \bar{\alpha})$. Set

$$m_0 = \left\lceil \frac{\log(4mp/\delta)}{\bar{q} + (1 - \bar{\alpha})(\bar{\omega} c_\kappa^2 \sqrt{\bar{n}}/32\sigma - 1)^2/2} \right\rceil,$$

where $\bar{q} := q - (1 - \bar{\alpha})\log(p)$. Following the same steps as in the proof of Theorem E.1 we get

$$\begin{aligned}
R(m, n) &\leq \mathcal{O}\left(m_0 nL + Lm\bar{n}\right) + R^\star(n, m - m_0) \\
&\leq \mathcal{O}\left(2Bm_0 n + 2Bm\bar{n}\right) + R^\star(n, m) \\
&\leq \mathcal{O}\left(Bn\log(mp/\delta)/\bar{n} + Bm\bar{n}\right) + R^\star(n, m).
\end{aligned}$$

□

Corollary F.8. *Assume the setting of Theorem F.7 and set $\bar{n} = \sqrt{n}$. Then with probability at least $1 - \delta$ we have*

$$R(m, n) \leq \mathcal{O}\left(B\sqrt{n}(\log(mp/\delta) + m)\right) + R^\star(n, m).$$

F.3 PERFORMANCE OF GP-UCB PAIRED WITH F-LIBO

Corollary F.9. *Assume we are in the setting of Corollary F.8 with GP-UCB as the base bandit algorithm and* $\lambda_{\text{ucb}} = 1+2/n$*. Then, for all* $0 < \delta < 1$*, with probability at least* $1-\delta$*,*

$$R(m,n) = \mathcal{O}\left(Bmd^\star\sqrt{n}\log\tfrac{n}{d^\star} + m\sqrt{nd^\star\log\tfrac{n}{d^\star}\log\tfrac{1}{\delta}} + B\sqrt{n}(m+\log(mp/\delta))\right).$$

Proof. The proof is the same as the proof for Corollary E.9, except that we use Corollary F.8 in place of Corollary E.3. □

G EXPERIMENT DETAILS

For the synthetic experiments, we initiate the algorithms with $\omega = c_1/2$. For all experiments, the exploration coefficient of the GP-UCB algorithm is set to $\nu_i = 10$ and $\lambda_{\text{ucb}} = 0.1$. Experiment are all repeated 20 times for difference random seeds, and the plots show the corresponding standard error. The remaining experiment settings are detailed in the following subsections.

G.1 OFFLINE DATA EXPERIMENTS

We generate the reward functions $f_1, ..., f_{30}$ from the synthetic environment. Corresponding to each f_s, we generate a data set $\mathcal{D}_s$ of size $n = 10$ by sampling points $\boldsymbol{x}_{s,1}, \dots, \boldsymbol{x}_{s,n}$ i.i.d from a uniform distribution $\mathcal{U}(\mathcal{X})$ over the domain $\mathcal{X} = [0,1]$ and collecting the corresponding noisy function values $y_{s,i} = f_s(\boldsymbol{x}_{s,i}) + \epsilon$, where the noise is samples from $\mathcal{N}(0, \sigma^2 = 0.01)$. We initiate META-KGL with the lasso regularization parameter of $\lambda = 0.25$ and F-META-KGL with $\lambda = 0.015$. For F-META-KGL, we set the majority vote threshold to $\alpha = 0.25$.

G.2 LIFELONG DATA EXPERIMENTS

For experiments using synthetic data, we set $n = 100$, and for the experiments on GLMNET data, there are $n = 144$ BO steps in each task. To run LIBO on the synthetic environment we set $\lambda = 0.5$ and for F-LIBO we set $\lambda = 0.2$. On the GMLNET environment, we instantiate LIBO with $\omega = 0.25$ and $\lambda = 0.015$, and F-META-KGL with $\alpha = 0.25$, $\omega = 10^{-6}$, $\lambda = 2.6 \times 10^{-6}$.

G.3 FURTHER EXPERIMENTS WITH SYNTHETIC DATA

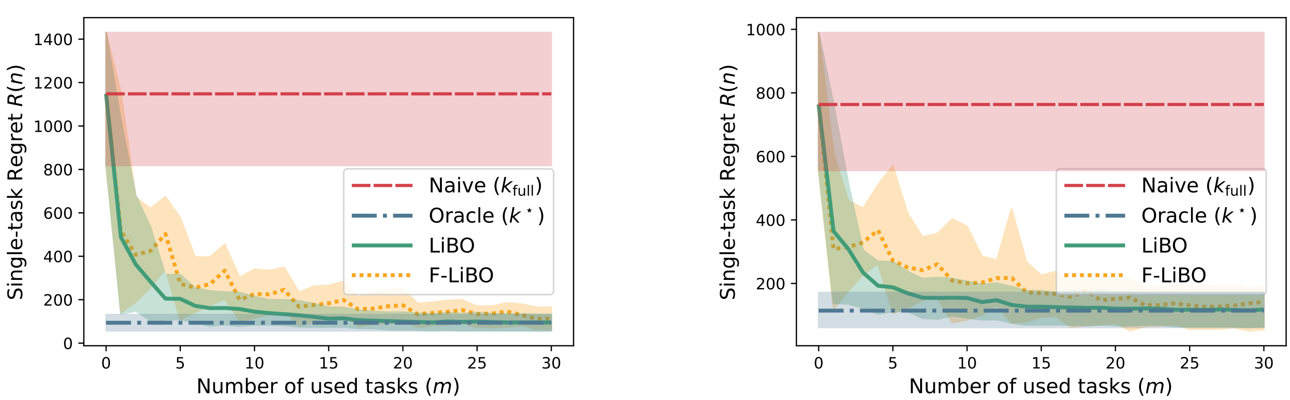

Figure 2: Single task cumulative regret of GP-UCB with meta-learned kernel $\hat{k}$ on an increasing number of meta-training tasks. Left: base kernels constructed with 2-dimensional cosine basis. Right: base kernels constructed with 1-dimensional Legendre polynomials. The BO performance with meta-leaned kernels quickly approaches oracle performance as the number of meta-training task increases.

Offline Data Analogous to the offline data experiments in Section 6.1, we provide additional results for a two-dimensional domain and Legendre polynomials instead of cosine bases in Figure 2. In particular, the left plot corresponds to $\mathcal{X} = [0, 1]^2$ as the domain and the first 50 2-dimensional cosine basis functions, i.e., $\phi_{i,j}(x) = \cos(i\pi x_1)\cos(\pi x_2), \forall x \in \mathcal{X}$, as the feature maps. For the right plot we choose $\mathcal{X} = [-1, 1]$ as the domain and use the first 50 Legendre Polynomials as the feature maps.

Figure 2 shows that both meta-learners converge with increasing number of tasks to the oracle kernel. This holds for different sets of base kernels and kernels with more than 1 input dimension. This empirically validates the theoretical findings of Theorem 3.3 and Theorem F.3. Somewhat peculiar is that we can observe oscillating behavior for the federated algorithm (yellow). This is a result of discrete nature of the voting system. The the total of number of tasks is a multiple of α the value $|\hat{J}_s|$ is large, while for points directly after that $|\hat{J}_s|$ are small. With increasing number of tasks the discretization has a lesser impact on the kernel estimation and the amplitude of the oscillations decreases.

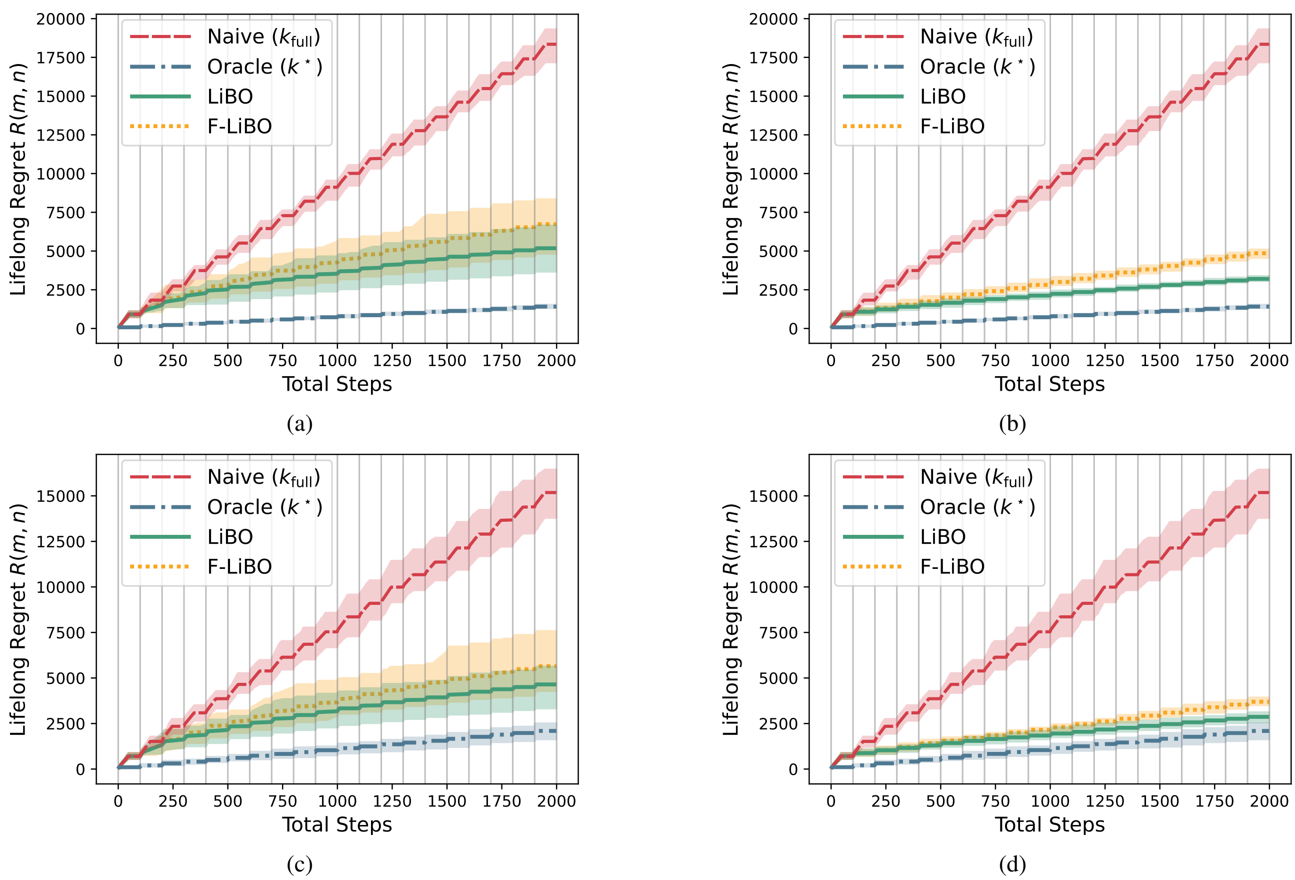

Figure 3: Lifelong regret with cosine basis (**a** & **b**) and Legendre polynomials as feature maps (**c** & **d**). In the plots **(a & c)** on the left, we use only the forced exploration data $\mathcal{D}^{\text{exp}}_{1:s}$ for meta-learning the kernel. For the plots **(b & d)** on the right, $D_{1:s}$, the data from all previous bandit interactions, is used. We observe that convergence is much faster when all interaction data is used in LIBO and F-LIBO.

Lifelong Data We now present modifications of the lifelong BO experiments in Section 6.2. In particular, we consider other base kernels as well as a modification of LIBO where we use all collected data for meta-learning $\hat{k}$ instead of only the forced exploration data $\mathcal{D}^{\text{exp}}_{1:s}$. The results are depicted in Figure 3. Figure **(a)** and **(b)** correspond to 50 cosine basis functions as feature maps for the base kernels. For Figure **(c)** and **(d)**, we use the first 50 Legendre polynomials as feature maps. The plots on the left (i.e. Fig. a, c) are generated with LIBO and F-LIBO, as presented in Algorithm 2 and 3, where only the forced exploration data is used for meta-learning. The plots on the right (i.e. Fig. b, d) correspond to a modified version of LIBO and F-LIBO where we use $D_{1:s}$, i.e., all previous bandit interactions, to meta-learn the kernel.

Generally, we observe that LIBO and F-LIBO substantially outperform the naive method which uses all base kernels. The gray vertical lines in Figure 3 indicate the beginning of a new task. We see that for every new task all algorithms initially experiences high regret, but, over time, as reward estimation improves, the cumulative regret flattens. As the rate of single-task convergence is dependent on the kernel, we see that differences in the performance between the algorithms emerge. When running LIBO, over time, forced exploration decreases and the estimated kernel converges to the true kernel. This means that, over time, the behavior of the agent using the LIBO estimator becomes indistinguishable form the agent using the oracle kernel. This is evident from 3 **(a)** as the slope of the single-task cumulative regret of the meta-agent (green) becomes the same as for the oracle agent (blue). In the federated case (yellow), while the estimated kernel also converges to the true kernel, the more restrictive setting forces us to use a constant exploration rate (see Algorithm 3 and Section 5) which means that the behavior of the federated meta-learner is always slightly sub-optimal. This can be observed by noting that the slope of the single-task cumulative regret of the federated meta-learner (yellow) is higher compared to the oracle agent even after the estimated kernel converges to the true kernel.

When we adjust LIBO and F-LIBO to use all available data to predict the kernel instead of only using $\mathcal{D}_{1:s}^{\mathrm{exp}}$, the lifelong regret decreases. As we would expect, using more data for meta-learning the kernel speeds up the convergence of $\hat{k}$ to k^* which, in turn, makes the BO runs more efficient. In practice, using the data from all interactions, not just the ones obtained by forced exploration, seems to be the best choice. From a theoretical perspective, this comes with additional technical challenges, as we point out in Section 4.2.