No-Regret Approximate Inference via Bayesian Optimisation (Supplement)

Rafael Oliveira[1,2] **Lionel Ott**[3] **Fabio Ramos**[1,2,4]

[1]School of Computer Science, the University of Sydney, Australia
[2]ARC Centre for Data Analytics for Resources and Environments, Australia
[3]Autonomous Systems Lab, ETH Zürich, Switzerland
[4]NVIDIA, Seattle, WA, USA

APPENDIX

This appendix complements the main paper with proofs, additional results and experiment details. In Appendix A, we start with additional theoretical background. Appendix B presents full proofs for the main theoretical results in the paper. Appendix C provides kernel-specific theoretical upper bounds for the KL divergence of posterior candidates obtained by KL-UCB. In Appendix D, we present a sensitivity analysis on dimensionality effects on KL-UCB. Finally, in Appendix E, we conclude with further details on the experiments setup.

A FURTHER BACKGROUND

Definition A.1. *Let Q and P be two probability measures such that Q is absolutely continuous with respect to P. The Kullback-Leibler (KL) divergence between Q and P is defined as:*

$$\mathbb{D}_{\mathrm{KL}}(Q||P) := \int \log\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}Q \,. \tag{1}$$

Case Q and P are both defined on a Euclidean space $\mathbb{R}^d$ and are absolutely continuous with respect to the Lebesgue measure on this space, the equation above can be rewritten as:

$$\mathbb{D}_{\mathrm{KL}}(Q||P) = \mathbb{D}_{\mathrm{KL}}(q||p) := \int_{\mathbb{R}^d} q(\boldsymbol{\theta}) \log\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta} \,, \tag{2}$$

where $q(\boldsymbol{\theta}) := \frac{\mathrm{d}Q(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}}$ and $p(\boldsymbol{\theta}) := \frac{\mathrm{d}P(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}}$ are the probability density functions of Q and P, respectively.

GP predictive equations: The theoretical results are given in terms of a Gaussian process model $\mathcal{GP}(m, k)$ for the log-likelihood function $\ell : \Theta \to \mathbb{R}$ which is learnt with batches of S i.i.d. samples from a probability distribution q_t at each iteration $t \in \{1, \ldots, T\}$. Therefore, at each iteration $t \geq 1$, the GP predictive mean and variance are given by:

$$\mu_t(\boldsymbol{\theta}) := m(\boldsymbol{\theta}) + \mathbf{k}_{N_t}(\boldsymbol{\theta})^{\mathsf{T}}(\mathbf{K}_{N_t} + \eta\mathbf{I})^{-1}(\mathbf{z}_{N_t} - \mathbf{m}_{N_t}) \tag{3}$$

$$k_t(\boldsymbol{\theta}, \boldsymbol{\theta}') := k(\boldsymbol{\theta}, \boldsymbol{\theta}') - \mathbf{k}_{N_t}(\boldsymbol{\theta})^{\mathsf{T}}(\mathbf{K}_{N_t} + \eta\mathbf{I})^{-1}\mathbf{k}_{N_t}(\boldsymbol{\theta}') \tag{4}$$

$$\sigma_t^2(\boldsymbol{\theta}) := k_t(\boldsymbol{\theta}, \boldsymbol{\theta}) \,, \tag{5}$$

where $\mathbf{k}_{N_t}(\boldsymbol{\theta}) := [k(\boldsymbol{\theta}, \boldsymbol{\theta}_1), \ldots, k(\boldsymbol{\theta}, \boldsymbol{\theta}_{N_t})]^{\mathsf{T}}$, $[\mathbf{K}_{N_t}]_{i,j} := k(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)$, $i, j \in \{1, \ldots, N_t\}$, and $N_t := tS$.

B PROOFS

We here present full proofs for the theoretical results in the paper. We start with a few auxiliary results and then follow with the proofs for the main results.

B.1 AUXILIARY RESULTS

Lemma B.1. *In our settings, the GP posterior variance is always bounded, i.e.:*

$$\forall t \geq 0, \quad \|\sigma_t\|_\infty \leq 1. \tag{6}$$

Proof. As the kernel k is bounded, with $k(\boldsymbol{\theta}, \boldsymbol{\theta}) \leq 1, \forall \boldsymbol{\theta} \in \Theta$, the definition in Equation 5 leads to:

$$\forall t \geq 1, \quad \sigma_t^2(\boldsymbol{\theta}) \leq k(\boldsymbol{\theta}, \boldsymbol{\theta}) \leq 1, \quad \forall \boldsymbol{\theta} \in \Theta \,. \tag{7}$$

□

Lemma B.2. *Assuming a bounded mean function $m : \Theta \to \mathbb{R}$, the GP posterior mean is bounded, i.e.:*

$$\forall t \geq 0, \quad \|\mu_t\|_\infty < \infty \,. \tag{8}$$

Proof. As defined in Equation 3, for $t \geq 1$, we have:

$$\begin{aligned}\mu_t(\boldsymbol{\theta}) &:= m(\boldsymbol{\theta}) + \mathbf{k}_{N_t}(\boldsymbol{\theta})^\mathsf{T}(\mathbf{K}_{N_t} + \eta\mathbf{I})^{-1}(\mathbf{z}_{N_t} - \mathbf{m}_{N_t}) \\ &\leq m(\boldsymbol{\theta}) + \eta^{-1}\mathbf{k}_{N_t}(\boldsymbol{\theta})^\mathsf{T}(\mathbf{z}_{N_t} - \mathbf{m}_{N_t}) \\ &\leq \|m\|_\infty + \eta^{-1}\|\mathbf{k}_{N_t}(\boldsymbol{\theta})\|_1\|\mathbf{z}_{N_t} - \mathbf{m}_{N_t}\|_\infty \\ &\leq \|m\|_\infty + \eta^{-1}t(\|\mathbf{z}_{N_t}\|_\infty + \|m\|_\infty) \\ &< \infty\,, \quad \forall \boldsymbol{\theta} \in \Theta\,,\end{aligned} \tag{9}$$

where we first used the fact that $(\mathbf{K}_{N_t} + \eta\mathbf{I})^{-1} = \eta^{-1}(\eta^{-1}\mathbf{K}_{N_t} + \mathbf{I})^{-1} \preccurlyeq \eta^{-1}\mathbf{I}$ due to positive-definiteness, then applied Hölder's inequality, and finally the observation that $\|\mathbf{k}_{N_t}(\boldsymbol{\theta})\|_1 \leq t$, due to the kernel bound. □

Lemma B.3 (Srinivas et al. (2010, Lemma 5.3))**.** *The information gain for a sequence of evaluations $\{\boldsymbol{\theta}_i, z_i\}_{i=1}^N$, where $z_i = f(\boldsymbol{\theta}_i) + \epsilon_i$, $\epsilon_i \sim \mathcal{N}(0, \eta)$, can be expressed in terms of the predictive variances. Namely, if $f \sim \mathcal{GP}(m, k)$, then:*

$$I(\mathbf{z}_N, \mathbf{f}_N|\Theta_N) = \frac{1}{2}\sum_{i=1}^N \log(1 + \eta^{-1}\sigma_{i-1}^2(\boldsymbol{\theta}_i))\,. \tag{10}$$

Lemma B.4 (Chowdhury and Gopalan (2017, Lemma 4))**.** *Following the setting of Lemma B.3, the sum of predictive standard deviations at a sequence of evaluation points is bounded in terms of the maximum information gain:*

$$\sum_{i=1}^N \sigma_{i-1}(\boldsymbol{\theta}_i) \leq \sqrt{4(N+2)\gamma_N}\,. \tag{11}$$

Lemma B.5. *Let $\mathcal{A} \subset \Theta$ be a set of points where a function $f \sim \mathcal{GP}(m, k)$ was evaluated, so that the GP posterior covariance function and the corresponding variance are given by:*

$$k_\mathcal{A}(\boldsymbol{\theta}, \boldsymbol{\theta}') := k(\boldsymbol{\theta}, \boldsymbol{\theta}') - k(\boldsymbol{\theta}, \mathcal{A})^\mathsf{T}(\mathbf{K}(\mathcal{A}) + \eta\mathbf{I})^{-1}k(\mathcal{A}, \boldsymbol{\theta}') \tag{12}$$

$$\sigma_\mathcal{A}^2(\boldsymbol{\theta}) := k_\mathcal{A}(\boldsymbol{\theta}, \boldsymbol{\theta})\,, \quad \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta, \tag{13}$$

where $k(\boldsymbol{\theta}, \mathcal{A}) := [k(\boldsymbol{\theta}, \mathbf{a})]_{\mathbf{a}\in\mathcal{A}}$ and $\mathbf{K}(\mathcal{A}) := [k(\mathbf{a}, \mathbf{a}')]_{\mathbf{a},\mathbf{a}'\in\mathcal{A}}$. Then, for any given set $\mathcal{B} \supset \mathcal{A}$ of evaluations of f, we have:

$$\sigma_\mathcal{A}^2(\boldsymbol{\theta}) \leq \sigma_\mathcal{B}^2(\boldsymbol{\theta}), \quad \forall \boldsymbol{\theta} \in \Theta\,. \tag{14}$$

Proof. The result follows by observing that the GP posterior given observations at $\mathcal{A}$ is a prior for the GP with the new observations at the complement $\mathcal{C} := \mathcal{B}\setminus\mathcal{A}$. Then we obtain, for all $\boldsymbol{\theta} \in \Theta$:

$$\begin{aligned}\sigma_\mathcal{B}^2(\boldsymbol{\theta}) &:= k(\boldsymbol{\theta}, \boldsymbol{\theta}) - k(\boldsymbol{\theta}, \mathcal{B})(\mathbf{K}(\mathcal{B}) + \eta\mathbf{I})^{-1}k(\mathcal{B}, \boldsymbol{\theta}) \\ &= \sigma_\mathcal{A}^2(\boldsymbol{\theta}) - k_\mathcal{A}(\boldsymbol{\theta}, \mathcal{C})(\mathbf{K}_\mathcal{A}(\mathcal{C}) + \eta\mathbf{I})^{-1}k_\mathcal{A}(\mathcal{C}, \boldsymbol{\theta}) \\ &\leq \sigma_\mathcal{A}^2(\boldsymbol{\theta})\,,\end{aligned} \tag{15}$$

since $k_\mathcal{A}(\boldsymbol{\theta}, \mathcal{C})(\mathbf{K}_\mathcal{A}(\mathcal{C}) + \eta\mathbf{I})^{-1}k_\mathcal{A}(\mathcal{C}, \boldsymbol{\theta})$ is non-negative. □

B.2 PROOF OF LEMMA 1

We prove Lemma 1 by applying Jensen's inequality to Theorem 1. Specifically, consider that the following holds with probability greater than $1 - \delta$ simultaneously over all Θ:

$$\forall t \geq 0, \quad |\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})| \leq \beta_t(\delta)\sigma_t(\boldsymbol{\theta})\,. \tag{16}$$

With the same probability, we then have that:

$$\begin{aligned}|\mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})]| &\leq \mathbb{E}_{\boldsymbol{\theta}\sim q_t}[|\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})|] \\ &\leq \beta_t(\delta)\mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\sigma_t(\boldsymbol{\theta})]\,, \\ &\forall \boldsymbol{\theta} \in \Theta\,, \forall t \geq 0\,,\end{aligned} \tag{17}$$

since the absolute value $|\cdot|$ is convex. Lastly, we note that the GP in Algorithm 1 is taking batches of samples per iteration. Therefore, we have to replace the original $\mathbf{K}_t$, which holds t observations, by $\mathbf{K}_{\mathcal{D}_t}$, which holds all the tS observations collected up to iteration t.[1] This concludes the proof.

B.3 PROOF OF THEOREM 2

We prove Theorem 2 in two parts, one for each component of the result. In the first part we establish a bound on the instant regret. We then propagate this bound to the cumulative regret. For these derivations, we will make use of Lemma 1 and some of the auxiliary results in Section B.1.

Proof of Theorem 2. We start by proving that, with $h(q|\mathcal{D}_{t-1})$ given by KL-UCB and $\beta_t := \beta_t(\delta)$ according to Theorem 1, we obtain a bound on the instant regret:

$$r_t \leq 2\beta_{t-1}(\delta)\mathbb{E}_{q_t}[\sigma_{t-1}]\,. \tag{18}$$

By Lemma 1, uniformly over all $t \geq 1$ with probability at least $1 - \delta$, we have that:

$$\begin{aligned}\max_{q\in\mathcal{Q}} \mathbb{E}_q[\ell] - \mathbb{D}_{\mathrm{KL}}(q||p) &\leq \max_{q\in\mathcal{Q}} \mathbb{E}_q[u_{t-1}] - \mathbb{D}_{\mathrm{KL}}(q||p) \\ &= \mathbb{E}_{q_t}[u_{t-1}] - \mathbb{D}_{\mathrm{KL}}(q_t||p)\,.\end{aligned} \tag{19}$$

Applying this bound to the definition of instant regret yields

[1] This replacement does not violate the conditions under which Theorem 1 holds, as we may define a filtration $\mathfrak{F}_{t-1} := \sigma(\mathcal{D}_{t-1}, \{\boldsymbol{\theta}_{t,i}\}_{i=1}^S)$, i.e., the σ-algebra generated by the random variables in the dataset $\mathcal{D}_{t-1}$ and the selected batch, $t \geq 1$. In this case, the noise $\{\epsilon_{t,i}\}_{i=1}^S$ in the observations at iteration $t \geq 1$ is σ_ϵ-sub-Gaussian when conditioned on $\mathfrak{F}_{t-1}$.

the first part of Theorem 2:

$$\begin{aligned} r_t &:= \mathbb{D}_{\mathrm{KL}}(q_t||p_{\mathbf{x}}) - \mathbb{D}_{\mathrm{KL}}(q^*||p_{\mathbf{x}}) \\ &= \max_{q\in\mathcal{Q}} \mathbb{E}_q[\ell] - \mathbb{D}_{\mathrm{KL}}(q||p) - \mathbb{E}_{q_t}[\ell] + \mathbb{D}_{\mathrm{KL}}(q_t||p) \\ &\leq \mathbb{E}_{q_t}[u_{t-1}] - \mathbb{E}_{q_t}[\ell] \\ &= \mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\mu_{t-1}(\boldsymbol{\theta}) + \beta_{t-1}\sigma_{t-1}(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta})] \\ &\leq 2\beta_{t-1}\mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\sigma_{t-1}(\boldsymbol{\theta})]\,, \end{aligned} \tag{20}$$

which holds with probability at least $1-\delta$.

For the second part, we apply the bound to the cumulative regret and initally obtain:

$$\begin{aligned} R_T := \sum_{t=1}^{T} r_t &\leq 2\sum_{t=1}^{T} \beta_{t-1}\mathbb{E}_{\tilde{\boldsymbol{\theta}}_t\sim q_t}[\sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)] \\ &\leq 2\beta_T \sum_{t=1}^{T} \mathbb{E}_{\tilde{\boldsymbol{\theta}}_t\sim q_t}[\sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)] \\ &\leq 2\beta_T \mathbb{E}_{\tilde{\boldsymbol{\theta}}_1\sim q_1,\ldots,\tilde{\boldsymbol{\theta}}_T\sim q_T}\left[\sum_{t=1}^{T}\sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)\right], \end{aligned} \tag{21}$$

since $\beta_t \geq \beta_{t-1}, \forall t \geq 1$, and expectations are linear operations. Considering the predictive variances above, recall that, at each iteration $t \geq 1$, the algorithm selects a batch of i.i.d. points $\mathcal{B}_t := \{\boldsymbol{\theta}_{t,i}\}_{i=1}^{S}$, sampled from q_t, where to evaluate the log-likelihood function ℓ. The predictive variance σ_{t-1}^2 is conditioned on all previous observations, which are grouped by batches. We can then decompose, for any $t \geq 1$:

$$\begin{aligned} \sigma_t^2(\boldsymbol{\theta}) &= \sigma_{t-1}^2(\boldsymbol{\theta}) \\ &\quad - k_{t-1}(\boldsymbol{\theta},\mathcal{B}_t)(\mathbf{K}_{t-1}(\mathcal{B}_t) + \eta\mathbf{I})^{-1}k_{t-1}(\mathcal{B}_t,\boldsymbol{\theta})\,, \end{aligned} \tag{22}$$

where we use the notation introduced in Lemma B.5, and:

$$\begin{aligned} k_t(\boldsymbol{\theta},\boldsymbol{\theta}') &= k_{t-1}(\boldsymbol{\theta},\boldsymbol{\theta}') \\ &\quad - k_{t-1}(\boldsymbol{\theta},\mathcal{B}_t)(\mathbf{K}_{t-1}(\mathcal{B}_t) + \eta\mathbf{I})^{-1}k_{t-1}(\mathcal{B}_t,\boldsymbol{\theta}') \end{aligned} \tag{23}$$

$$k_0(\boldsymbol{\theta},\boldsymbol{\theta}') := k(\boldsymbol{\theta},\boldsymbol{\theta}')\,. \tag{24}$$

Therefore, the predictive variance of the batched algorithm is not the same as the predictive variance of a sequential algorithm, and we cannot direcly apply Lemma B.4 to bound the last term in Equation 21.

Lemma B.5 tells us that the predictive variance given a set of observations is less than the predictive variance given a subset of observations. Selecting only the first point from within each batch and applying Lemma B.5, we get, for $t \geq 1$:

$$\begin{aligned} \sigma_t^2(\boldsymbol{\theta}) &\leq s_t^2(\boldsymbol{\theta}) \\ &:= k(\boldsymbol{\theta},\boldsymbol{\theta}) - k(\boldsymbol{\theta},\Theta_t)(\mathbf{K}(\Theta_t) + \eta\mathbf{I})^{-1}k(\Theta_t,\boldsymbol{\theta})\,, \end{aligned} \tag{25}$$

where $\Theta_t := \{\boldsymbol{\theta}_{i,1}\}_{i=1}^{t}$, with $\boldsymbol{\theta}_{i,1} \in \mathcal{B}_i$, $i \in \{1,\ldots,t\}$. Note that the right-hand side of the equation above is simply the non-batched GP predictive variance. Furthermore, sample points within a batch are i.i.d., so that $\boldsymbol{\theta}_{t,1} \sim q_t$ and $\tilde{\boldsymbol{\theta}}_t \sim q_t$ are identically distributed. We can now apply Lemma B.4, yielding:

$$\begin{aligned} &\mathbb{E}_{\tilde{\boldsymbol{\theta}}_1\sim q_1,\ldots,\tilde{\boldsymbol{\theta}}_T\sim q_T}\left[\sum_{t=1}^{T}\sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)\right] \\ &\leq \mathbb{E}_{\tilde{\boldsymbol{\theta}}_1\sim q_1,\ldots,\tilde{\boldsymbol{\theta}}_T\sim q_T}\left[\sum_{t=1}^{T}s_{t-1}(\tilde{\boldsymbol{\theta}}_t)\right] \\ &\leq 2\sqrt{(T+2)\gamma_T}\,. \end{aligned} \tag{26}$$

Combining this result with Equation 21, we obtain:

$$R_T \leq 4\beta_T\sqrt{(T+2)\gamma_T} \in \mathcal{O}(\beta_T\sqrt{T\gamma_T})\,. \tag{27}$$

Lastly, from the definition of $\beta_t(\delta)$, we have:

$$\beta_T(\delta) := b + \sigma_\epsilon\sqrt{2\eta^{-1}\log(|\mathbf{I}+\eta^{-1}\mathbf{K}_{\mathcal{D}_T}|^{1/2}/\delta)}\,, \tag{28}$$

where:

$$\log(|\mathbf{I}+\eta^{-1}\mathbf{K}_{\mathcal{D}_T}|^{1/2}) = I(\mathbf{z}_{N_T},\mathbf{g}_{N_T}) \leq \gamma_{N_T} = \gamma_{ST}\,, \tag{29}$$

for $g \sim \mathcal{GP}(m,k)$. Therefore, the KL-UCB cumulative regret is such that:

$$R_T \in \mathcal{O}(\sqrt{T}(b\sqrt{\gamma_T} + \sqrt{\gamma_T\gamma_{ST}}))\,, \tag{30}$$

which concludes the proof. □

B.4 PROOF OF LEMMA 2

Lemma 2 states that, for a bounded kernel k, maximising h at any $t \geq 1$ is equivalent to:

$$\operatorname*{argmax}_{q\in\mathcal{Q}} h(q|\mathcal{D}_{t-1}) = \operatorname*{argmin}_{q\in\mathcal{Q}} \mathbb{D}_{\mathrm{KL}}(q||\hat{p}_{t-1})\,, \tag{31}$$

where h is defined as:

$$h(q|\mathcal{D}_{t-1}) := \mathbb{E}_{\boldsymbol{\theta}\sim q}[u_{t-1}(\boldsymbol{\theta})] - \mathbb{D}_{\mathrm{KL}}(q||p)\,, \tag{32}$$

with $u_t(\boldsymbol{\theta}) := \mu_t(\boldsymbol{\theta}) + \beta_t\sigma_t(\boldsymbol{\theta})$, $0 \leq \beta_t < \infty$, and μ_t and σ_t^2 define GP posterior mean and variance at iteration t. The proof follows by the same argument which turns the general KL divergence minimisation problem into an ELBO maximisation. The only part to verify is whether $\hat{p}_{t-1}$ defines a valid probability density function.

Proof of Lemma 2. In general, let $f : \Theta \to \mathbb{R}$ be a bounded function on $\Theta \subset \mathbb{R}^d$, i.e. $\|f\|_\infty < \infty$, and p a probability density function on Θ. Then we have that:

$$\mathbb{E}_p[f] := \int_\Theta f(\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \leq \|f\|_\infty \int_\Theta p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = \|f\|_\infty\,. \tag{33}$$

Combining Lemma B.1 and Lemma B.2, we have that u_t is always bounded, since:

$$\begin{aligned}\forall t \geq 1, \quad u_t(\boldsymbol{\theta}) &:= \mu_t(\boldsymbol{\theta}) + \beta_t \sigma_t(\boldsymbol{\theta}) \\ &\leq \|\mu_t\|_\infty + \beta_t \|\sigma_t\|_\infty < \infty, \quad \forall \boldsymbol{\theta} \in \Theta\,.\end{aligned} \tag{34}$$

Then the normalisation constant ζ_t for $\hat{p}_t(\boldsymbol{\theta}) = \frac{1}{\zeta_t} p(\boldsymbol{\theta}) \exp u_t(\boldsymbol{\theta})$ is bounded, for:

$$\begin{aligned}\forall t \geq 0, \quad \zeta_t &:= \mathbb{E}_p[\exp u_t] \\ &= \int_\Theta p(\boldsymbol{\theta}) \exp u_t(\boldsymbol{\theta}) \,\mathrm{d}\boldsymbol{\theta} \leq \|\exp u_t\|_\infty \\ &< \infty\,.\end{aligned} \tag{35}$$

Now, for the lemma's main result, we obtain:

$$\begin{aligned}&\operatorname*{argmax}_{q \in \mathcal{Q}} \mathbb{E}_q[u_{t-1}] - \mathbb{D}_{\mathrm{KL}}(q||p) \\ &= \operatorname*{argmax}_{q \in \mathcal{Q}} \mathbb{E}_q[u_{t-1} + \log p - \log q] \\ &= \operatorname*{argmin}_{q \in \mathcal{Q}} \log \mathbb{E}_p[\exp u_{t-1}] - \mathbb{E}_q[u_{t-1} + \log p - \log q] \\ &= \operatorname*{argmin}_{q \in \mathcal{Q}} \int_\Theta q(\boldsymbol{\theta}) \log \left(\frac{q(\boldsymbol{\theta}) \mathbb{E}_p[\exp u_{t-1}]}{p(\boldsymbol{\theta}) \exp u_{t-1}(\boldsymbol{\theta})} \right) \mathrm{d}\boldsymbol{\theta} \\ &= \operatorname*{argmin}_{q \in \mathcal{Q}} \mathbb{D}_{\mathrm{KL}}(q||\hat{p}_{t-1})\,,\end{aligned} \tag{36}$$

which concludes the proof. □

B.5 PROOF OF COROLLARY 1

The corollary is simply a restatement of Theorem 2 in terms of $q_t = \hat{p}_{t-1}$. In this case, the KL divergence with respect to the optimal solution is $\mathbb{D}_{\mathrm{KL}}(q^*||p_{\mathbf{x}}) = 0$, since we are considering a class of non-parametric distributions which can recover arbitrary distributions via MCMC. By Theorem 2, we then know that:

$$\min_{t \leq T} r_t \leq \frac{1}{T} R_T \in \mathcal{O}(\gamma_{ST} T^{-1/2})\,. \tag{37}$$

If $\gamma_{ST} \in \mathcal{O}(T^\alpha)$, for some $\alpha < 1/2$, so that $\alpha - 1/2 < 1$, then:

$$\frac{1}{T} R_T \in \mathcal{O}(T^{\alpha - 1/2}) \implies \lim_{T \to \infty} \min_{t \leq T} r_t \leq \lim_{T \to \infty} \frac{1}{T} R_T = 0\,. \tag{38}$$

Replacing $r_t = \mathbb{D}_{\mathrm{KL}}(\hat{p}_{t-1}||p_{\mathbf{x}})$ above concludes the proof.

Alternative proof. Another way of proving the same bound on the KL divergence is the following. The KL divergence from $\hat{p}_t$ to $p_{\mathbf{x}}$ is bounded via Lemma 1. Namely, with probability greater than $1 - \delta$, we have that:

$$\begin{aligned}\mathbb{D}_{\mathrm{KL}}(\hat{p}_t||p_{\mathbf{x}}) &= \mathbb{E}_{\boldsymbol{\theta} \sim \hat{p}_t}[\log \hat{p}_t(\boldsymbol{\theta}) - \log p_{\mathbf{x}}(\boldsymbol{\theta})] \\ &= \mathbb{E}_{\boldsymbol{\theta} \sim \hat{p}_t}[\log p(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta}) + \beta_t \sigma_t(\boldsymbol{\theta}) \\ &\quad - \log \zeta_t - \log p_{\mathbf{x}}(\boldsymbol{\theta})] \\ &= \mathbb{E}_{\boldsymbol{\theta} \sim \hat{p}_t}[\mu_t(\boldsymbol{\theta}) + \beta_t \sigma_t(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta})] + \log p(\mathbf{x}) \\ &\quad - \log \zeta_t \\ &\leq 2\beta_t \mathbb{E}_{\boldsymbol{\theta} \sim \hat{p}_t}[\sigma_t(\boldsymbol{\theta})] + \log p(\mathbf{x}) - \log \zeta_t\,.\end{aligned} \tag{39}$$

For the second part of the last term in the right-hand side above, we also have that:

$$p(\mathbf{x}) = \int_\Theta p(\boldsymbol{\theta}) \exp \ell(\boldsymbol{\theta}) \,\mathrm{d}\boldsymbol{\theta} \leq \int_\Theta p(\boldsymbol{\theta}) \exp u_t(\boldsymbol{\theta}) \,\mathrm{d}\boldsymbol{\theta} = \zeta_t\,, \tag{40}$$

which holds with the same probability as Equation 39. Therefore, we conclude the proof with:

$$\begin{aligned}&\log p(\mathbf{x}) - \log \zeta_t \leq 0 \\ &\implies \mathbb{D}_{\mathrm{KL}}(\hat{p}_t||p_{\mathbf{x}}) \leq 2\beta_t \mathbb{E}_{\boldsymbol{\theta} \sim \hat{p}_t}[\sigma_t(\boldsymbol{\theta})]\,.\end{aligned} \tag{41}$$

□

C KL DIVERGENCE BOUNDS FOR KL-UCB WITH SPECIFIC KERNELS

Corollary 1 connects the regret bound in Theorem 2 with the KL divergence of posterior approximations by KL-UCB when sampling directly from the posterior surrogate induced by UCB. In this case, we can bound the KL divergence of the posterior approximations by KL-UCB with respect to the posterior as:

$$\min_{t \leq T} \mathbb{D}_{\mathrm{KL}}(\hat{p}_t||p_{\mathbf{x}}) \leq \frac{R_T}{T}\,, \quad T \geq 1\,. \tag{42}$$

According to Theorem 2, we have $R_T \in \mathcal{O}(\gamma_{ST}\sqrt{T})$. Therefore, to bound the KL divergence, we need kernel-specific upper bounds for the maximum information gain γ_T (Srinivas et al., 2010; Vakili et al., 2021). In particular, we consider the case of two popular stationary kernel classes. The first one is the squared-exponential kernel, used in our experiments. As previously mentioned, this kernel yields a bound $\gamma_T \in \mathcal{O}\left(\log^{d+1}(T)\right)$ (Srinivas et al., 2010).

The second type of kernel is the Matérn class with smoothness parameter $\nu > 1/2$:

$$k(\boldsymbol{\theta}, \boldsymbol{\theta}') := \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{\sqrt{2\nu}\rho_{\boldsymbol{\theta},\boldsymbol{\theta}'}}{l} \right)^\nu B_\nu \left(\frac{\sqrt{2\nu}\rho_{\boldsymbol{\theta},\boldsymbol{\theta}'}}{l} \right)\,, \tag{43}$$

where $\rho_{\boldsymbol{\theta},\boldsymbol{\theta}'} := \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2$, $l > 0$ is a length-scale parameter controlling the smoothness of the functions in the RKHS, Γ

Kernel class	$\min_{t \leq T} \mathbb{D}_{\mathrm{KL}}(\hat{p}_t \| p_{\mathbf{x}})$
Squared exponential	$\mathcal{O}\left(T^{-1/2} \log^{d+1}(ST)\right)$
Matérn $\nu > 1/2$	$\mathcal{O}\left(T^{\frac{1}{2\nu+d}\left(\frac{d}{2}-\nu\right)} \log^{\frac{2\nu}{2\nu+d}}(ST)\right)$

Table 1: KL divergence bounds for KL-UCB

is the gamma function, and B_ν is the modified Bessel function of the second kind, for $\boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta$. This kernel leads to a maximum information gain $\gamma_T \in \mathcal{O}\left(T^{\frac{d}{2\nu+d}} \log^{\frac{2\nu}{2\nu+d}}(T)\right)$ according to recent results (Vakili et al., 2021).

Table 1 presents upper bounds for the KL divergence of posterior approximations by KL-UCB. As the table shows, both kernels lead to an asymptotically vanishing KL divergence in the approximations with respect to the true posterior in general. An exception to asymptotic convergence, however, is that convergence does not necessarily hold for $d \geq 2\nu$ in the case of the Matérn kernel, noticing the exponent in its rate. For instance, a Matérn kernel with $\nu = 3/2$ would not guarantee convergence in a problem whose parameter space has dimension $d > 3$. The guarantees for the squared-exponential kernel do not suffer from this drawback, though the rates are possibly worse due to the exponential dependence on d via the logarithmic term.

In terms of approximation bounds with respect to the number of likelihood evaluations $N = ST$, we have a KL divergence bound of $\tilde{\mathcal{O}}\left(N^{-1/2}\right)$ and $\tilde{\mathcal{O}}\left(N^{\frac{1}{2\nu+d}\left(\frac{d}{2}-\nu\right)}\right)$ for the squared-exponential and the Matérn kernels, respectively. Here the $\tilde{\mathcal{O}}$-notation suppresses logarithmic factors. Compared to the exponential convergence rates in Kanagawa and Hennig (2019), notice that their results are for a noise-free setting, while we consider settings with (sub-Gaussian) noise. In the noise-free setting, one is usually able to obtain tighter concentration bounds for the GP approximation (see de Freitas et al., 2012).

D DIMENSIONALITY EFFECT

In this section, we present a short analysis on the effect of dimensionality on the regret of the KL-UCB algorithm. Figure 1 presents the KL-UCB regret for the problem in Section 7.1 when we increase the dimensionality of the parameter space $\Theta = \mathbb{R}^d$. As the plot shows, the regret has an exponential dependence on the dimensionality of the parameter space. Therefore, the practitioner might need to run the method for longer to obtain reasonable posterior approximations or apply dimensionality reduction methods.

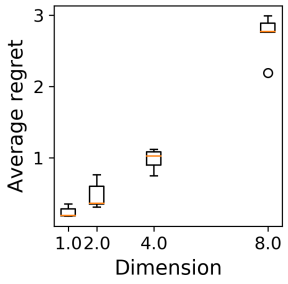

Figure 1: Dimensionality effect on the mean regret of KL-UCB on the RKHS log-likelihood problem after 20 iterations as a function of the dimensionality of the parameter space. The results of 5 independent runs were combined to produce the box plot.

E DETAILS OF THE EXPERIMENTAL SETUP

In this section, we present further details on the experimental setup for the empirical results in the paper. In particular, we describe the settings for KL-UCB. For VBMC, we used an implementation provided by its author.[2]

KL-UCB setup: For the GP model in KL-UCB, we used GPyTorch (Gardner et al., 2018) with adaptations to perform fast rank-1 Cholesky updates on the GP covariance matrix (see Rasmussen and Williams, 2006, Algorithm 2.1). Given the theoretical nature of this work, we did not perform online hyper-parameters learning, contrary to what is usual in other GP-based approximate inference methods (Gutmann and Corander, 2016; Acerbi, 2018). The kernel lengthscale was set as 0.5, a value which provided fitting GP estimates for the generated problem scenarios. The GP was configured as zero mean $m := 0$ for the RKHS-based problem in Section 7.1, while we used the log-prior probability as the mean function $m := \log p$ for the problems in Section 7.2. The latter allows the GP to provide low likelihood estimates for parameters of low prior probability, avoiding excessive exploration of the parameter space. In terms of noise settings, observations in the RKHS problem (Section 7.1) were added with Gaussian noise $\sigma_\epsilon := 0.01\|\ell\|_k$, while the problems with comparisons against VBMC were configured with basically no noise ($\sigma_\epsilon := 10^{-6}$), since this algorithm was not originally designed to handle noise (see Acerbi, 2020, for a recent noise-adapted version). The GP noise parameter was correspondingly set as $\eta := 10^{-2}$ for the RKHS problem, and $\eta := 10^{-4}$ for the noise-free problems. Having a small, non-zero η avoids numerical issues with matrix inversions. Lastly, for the setting of β_t, the RKHS norm for the log-likelihood function in Section 7.1 is available in closed form, but for the non-RKHS functions we set $b := 3$ which yields

[2]VBMC experiments code: `https://github.com/lacerbi/infbench`

the common 3 standard deviations UCB parameter (also in VBMC (Acerbi, 2018)) in a noiseless setting.

MCMC setup: We configured the EMCEE sampler (Foreman-Mackey et al., 2013) with 25 walkers and a burn-in of 400 samples. We selected the $S := 5$ evaluation points in each BO iteration's batch out of 500 samples drawn by EMCEE. Sub-sampling from a larger batch of samples reduces the correlation between samples from the chain.

Estimation of KL divergence: To verify theoretical bounds, we measured the KL divergence between KL-UCB's posterior approximations and the true posterior distribution, which is unknown to KL-UCB. As the posterior distribution approximations from KL-UCB are sampled-based MCMC estimates, we had to estimate the KL divergence based on samples. We also took into account that MCMC samples are usually correlated. To decorrelate the samples, we built a kernel density estimator out of the MCMC samples using a Gaussian kernel with the rule by Scott (1992) for kernel bandwidth selection[3] and then sampled from this continuous density. For a large enough number of MCMC samples (2000 for the likelihood-free inference problem, 10000 for the circular likelihood, and 1000 for the other problems), the i.i.d. samples from the density estimator approximately follow the stationary distribution of the MCMC chain, i.e. its target posterior. We then applied a k-nearest-neighbours method (Szabó, 2014)[4] to obtain the KL divergence estimates.

References

Luigi Acerbi. Variational Bayesian Monte Carlo. In *Advances in Neural Information Processing Systems*, volume 31, Montréal, Canada, 2018.

Luigi Acerbi. Variational Bayesian Monte Carlo with Noisy Likelihoods. In *Advances in Neural Information Processing Systems*, volume 33, Online, 2020.

Sayak Ray Chowdhury and Aditya Gopalan. On Kernelized Multi-armed Bandits. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, Sydney, Australia, 2017.

Nando de Freitas, Alex Smola, and Masrour Zoghi. Exponential regret bounds for Gaussian process bandits with deterministic observations. In *Proceedings of the 29th International Conference on Machine Learning (ICML-12)*, pages 1743–1750, 2012.

D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman. emcee: The MCMC hammer. *PASP*, 125:306–312, 2013. doi: 10.1086/670067.

Jacob R Gardner, Geoff Pleiss, David Bindel, Kilian Q Weinberger, and Andrew Gordon Wilson. Gpytorch: Blackbox matrix-matrix gaussian process inference with gpu acceleration. In *Advances in Neural Information Processing Systems*, volume 31, Montréal, Canada, 2018.

Michael U. Gutmann and Jukka Corander. Bayesian Optimization for Likelihood-Free Inference of Simulator-Based Statistical Models. *Journal of Machine Learning Research*, 17:1–47, 2016.

Motonobu Kanagawa and Philipp Hennig. Convergence guarantees for adaptive Bayesian quadrature methods. In *Advances in Neural Information Processing Systems*, volume 32, Vancouver, Canada, 2019.

Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, Cambridge, MA, 2006.

David W Scott. *Multivariate density estimation: theory, practice, and visualization*. Wiley, 1992.

Niranjan Srinivas, Andreas Krause, Sham M. Kakade, and Matthias Seeger. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design. In *Proceedings of the 27th International Conference on Machine Learning (ICML 2010)*, pages 1015–1022, 2010.

Zoltán Szabó. Information theoretical estimators toolbox. *Journal of Machine Learning Research*, 15:283–287, 2014.

Sattar Vakili, Kia Khezeli, and Victor Picheny. On information gain and regret bounds in gaussian process bandits. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 130 of *Proceedings of Machine Learning Research*, pages 82–90. PMLR, 13–15 Apr 2021.

[3] Implementation available with the KDEpy package: `https://github.com/tommyod/KDEpy`

[4] We used the Python version of the ITE toolbox: `https://bitbucket.org/szzoli/ite-in-python/`