No-Regret Approximate Inference via Bayesian Optimisation

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

Abstract

We consider Bayesian inference problems where the likelihood function is either expensive to evaluate or only available via noisy estimates. This setting encompasses application scenarios involving, for example, large datasets or models whose likelihood evaluations require expensive simulations. We formulate this problem within a Bayesian optimisation framework over a space of probability distributions and derive an upper confidence bound (UCB) algorithm to propose non-parametric distribution candidates. The algorithm is designed to minimise regret, which is defined as the Kullback-Leibler divergence with respect to the true posterior in this case. Equipped with a Gaussian process surrogate model, we show that the resulting UCB algorithm achieves asymptotically no regret. The method can be easily implemented as a batch Bayesian optimisation algorithm whose point evaluations are selected via Markov chain Monte Carlo. Experimental results demonstrate the method's performance on inference problems.

1 INTRODUCTION

Bayesian inference problems have been traditionally solved via exact inference methods such as Markov chain Monte Carlo (MCMC) (Andrieu et al., 2003) when the likelihood function is reasonably cheap to evaluate. In these settings, often with the aid of other information, including gradients (Neal, 2011), MCMC methods explore posterior surfaces by performing typically thousands of evaluations until attaining a suitable sample-based posterior approximation. As a lightweight alternative, variational inference methods (Bishop, 2007) tradeoff approximation accuracy by sample-efficiency with methods which can reduce the number of likelihood evaluations by orders of magnitude (Ranganath et al., 2014). However, most state-of-the-art variational inference algorithms rely on gradient information from the probabilistic model within automatic inference frameworks (Salvatier et al., 2016; Bingham et al., 2019).

In some application scenarios in science and engineering, such as geophysical processes (Wellmann et al., 2010), ecological systems (Beaumont, 2010) and robotics (Ramos et al., 2019), likelihood evaluations are often expensive or intractable due to the use of complex simulations of physical phenomena. In the simulation of high-energy collisions in particle physics, for example, a single event may involve sampling of millions of random variables, making likelihood evaluations completely intractable (Brehmer et al., 2018).

When likelihood functions are not available, one resorts to likelihood-free inference methods, such as approximate Bayesian computation (ABC) (Robert, 2016), conditional density estimation (Papamakarios and Murray, 2016), synthetic likelihoods (Ong et al., 2018), etc. Simulations might still be expensive and limited by resource constraints, imposing challenges to traditional approaches, like ABC. Recent methods address the efficiency problem by sequentially learning approximations to the likelihood (Gutmann and Corander, 2016; Papamakarios et al., 2019), or the posterior directly (Greenberg et al., 2019), from simulation data. In particular, Gutmann and Corander (2016) derive an active learning approach using Bayesian optimisation (BO) (Shahriari et al., 2016) to propose simulation parameters.

In this paper, we address the problem of learning posterior approximations via Bayesian optimisation. We consider settings with black-box likelihood functions which might be noisy and expensive to evaluate. Contrary to the traditional point optimisation for BO, which only seeks extreme values, we address the problem of learning the distribution as a whole via non-parametric, sample-based approximations. For this purpose, we derive an upper confidence bound (UCB) algorithm to sample approximate posteriors learnt by a Gaussian process (GP) (Rasmussen and Williams, 2006) model on the log-likelihood. Compared to

Accepted for the 37th Conference on Uncertainty in Artificial Intelligence (UAI 2021).

previous approaches using GP-based methods for approximate inference (Gutmann and Corander, 2016; Acerbi, 2018; Järvenpää et al., 2020), our method proposes a sequence of non-parametric sample-based posterior approximations which allow us to bound the Kullback-Leibler (KL) divergence between the estimate and the true unknown posterior. Theoretical results demonstrate that the algorithm is able to converge with sub-linear regret bounds towards the posterior distribution. We also provide experimental results, assessing the method's performance in practice on approximate inference problems.

2 RELATED WORK

Bayesian optimisation approaches for black-box approximate inference have recently been proposed in the literature. In the context of ABC, Gutmann and Corander (2016) proposed learning a likelihood function from discrepancy evaluations of simulations using a UCB criterion for simulation parameters selection, with other acquisition rules later developed (Järvenpää et al., 2017).

Adaptive Bayesian quadrature (ABQ) methods (Kanagawa and Hennig, 2019) are also closely related to BO when considering a search space consisting of probability distributions. Acerbi (2018) presented variational Bayesian Monte Carlo (VBMC), an ABQ framework which sequentially learns variational approximations to a posterior distribution on top of a GP surrogate of the log-joint probability density between parameters and observations. More recently, Järvenpää et al. (2020) proposed a GP-based method to sequentially select informative batches of parameters for parallel likelihood evaluations. Their method selects batch elements based on an optimal posterior approximation derived from the current GP log-likelihood estimate. Although both ABQ methods do not explicitly involve an optimisation objective, the intermediate posterior approximations at each iteration can be seen as a sequence of candidate solutions optimising an implicit distributional objective.

Non-BO methods based on sequential approximations to a posterior are also plentiful in the likelihood-free inference literature (Papamakarios and Murray, 2016; Papamakarios et al., 2019; Greenberg et al., 2019; Durkan et al., 2020; Cranmer et al., 2020). These methods rely on neural density, or density-ratio, estimators to learn surrogate posterior densities from simulated data. Although our framework can also be applied to likelihood-free inference, we do not assume direct access to simulated data, but only to evaluations of a synthetic likelihood or discrepancy function.

Regarding theoretical results on posterior approximation errors, Kanagawa and Hennig (2019) analyse the convergence rate of adaptive Bayesian quadrature methods, which applies, for example, to Acerbi (2018, 2020). However, in this work, we are concerned with finite-time upper bounds on the posterior approximation error, defined by the KL divergence. We derive an algorithm which not only comes equipped with such finite-time guarantees, but which is also shown to achieve asymptotic optimality.

3 PROBLEM FORMULATION

Our goal is to estimate a posterior distribution over parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^d$ given observed data $\mathbf{x} \in \mathcal{X}$:

$$p_{\mathbf{x}}(\boldsymbol{\theta}) := p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{x})} \,, \tag{1}$$

where we consider a likelihood $p(\mathbf{x}|\boldsymbol{\theta})$ only available via noisy and/or expensive evaluations, the evidence $p(\mathbf{x})$ is assumed intractable, but we have a prior $p(\boldsymbol{\theta})$.

We formulate the posterior estimation in Equation 1 as an optimisation problem over the space $\mathcal{P}$ of probability distributions on Θ:

$$q^* \in \operatorname*{argmin}_{q \in \mathcal{Q}} \mathbb{D}_{\mathrm{KL}}(q||p_{\mathbf{x}}) \,, \tag{2}$$

where $\mathcal{Q} \subset \mathcal{P}$ is a set of probability distributions over Θ, and $\mathbb{D}_{\mathrm{KL}}(q||p_{\mathbf{x}}) \geq 0$ denotes the Kullback-Leibler (KL) divergence between q and $p_{\mathbf{x}}$, which is such that $\mathbb{D}_{\mathrm{KL}}(q||p_{\mathbf{x}}) = 0 \iff q = p_{\mathbf{x}}$. The problem above can be shown to be equivalent to:

$$q^* \in \operatorname*{argmax}_{q \in \mathcal{Q}} \mathbb{E}_{\boldsymbol{\theta} \sim q}[\ell(\boldsymbol{\theta})] - \mathbb{D}_{\mathrm{KL}}(q||p) \,, \tag{3}$$

where $\ell(\boldsymbol{\theta}) := \log p(\mathbf{x}|\boldsymbol{\theta})$ denotes the log-likelihood. The objective function in Equation 3 is known as the evidence lower bound (ELBO), as it lower bounds $\log p(\mathbf{x})$.

Finding the global optimum in Equation 3 is usually intractable for an arbitrary log-likelihood function ℓ and distributional search space $\mathcal{Q}$. In addition, we also consider cases where the estimates of ℓ are possibly noisy. Therefore, we make a few regularity assumptions on the log-likelihood function and its estimates.

Regularity assumptions: We assume $\ell : \Theta \to \mathbb{R}$ lies in a reproducing kernel Hilbert space (RKHS) (Schölkopf and Smola, 2002). Given a positive-definite kernel $k : \Theta \times \Theta \to \mathbb{R}$, a RKHS $\mathcal{H}_k$ is a Hilbert space of functions with inner product $\langle \cdot, \cdot \rangle_k$ and norm $\|\cdot\|_k = \sqrt{\langle \cdot, \cdot \rangle_k}$, such that $f(\boldsymbol{\theta}) = \langle f, k(\cdot, \boldsymbol{\theta}) \rangle_k$, for any $f \in \mathcal{H}_k$ and any $\boldsymbol{\theta} \in \Theta$. We assume k is continuous and bounded with $k(\boldsymbol{\theta}, \boldsymbol{\theta}) \leq 1, \forall \boldsymbol{\theta} \in \Theta$, and that $\|\ell\|_k \leq b$, where $b > 0$ is known.[1]

Likelihood estimates: We consider a general sequential setting where we may not have direct access to the log-likelihood ℓ, but only to noisy estimates $\hat{\ell}$. In particular, we

[1]These assumptions are met by most of the popular kernels and are common in the online learning literature.

assume that the approximation error:

$$\epsilon_t := \hat{\ell}(\boldsymbol{\theta}_t) - \ell(\boldsymbol{\theta}_t) \tag{4}$$

is zero-mean σ_ϵ-sub-Gaussian for a given $\sigma_\epsilon > 0$ (Boucheron et al., 2013; Chowdhury and Gopalan, 2017). The sub-Gaussian assumption implies unbiased log-likelihood estimates with an error which is bounded or whose distribution tails decay at least as fast as a Gaussian. Examples include unbiased synthetic log-likelihood models (Ong et al., 2018) and mini-batching estimates when considering models with bounded log-likelihood functions.

Regret: Our performance analysis is based on the instant regret of an algorithm choosing a sequence of distributions $\{q_t\}_{t\geq 1}$. The instant regret at iteration $t \geq 1$ is defined as:

$$r_t := \mathbb{D}_{\mathrm{KL}}(q_t||p_{\mathbf{x}}) - \mathbb{D}_{\mathrm{KL}}(q^*||p_{\mathbf{x}}) \,. \tag{5}$$

If our class of distributions $\mathcal{Q}$ is flexible enough, so that $p_{\mathbf{x}} \in \mathcal{Q}$, then $\mathbb{D}_{\mathrm{KL}}(q^*||p_{\mathbf{x}}) = 0$, setting $q^* = p_{\mathbf{x}}$, and:

$$r_t = \mathbb{D}_{\mathrm{KL}}(q_t||p_{\mathbf{x}}) \,. \tag{6}$$

An upper bound on the cumulative regret $R_T := \sum_{t=1}^{T} r_t$ tells us how far the candidates q_t picked by the algorithm are to the global optimum q^*. In particular, the regret for the best candidate up to iteration T is bounded by:

$$\min_{t\leq T} r_t \leq \frac{1}{T} R_T \,. \tag{7}$$

Therefore, if R_T grows sub-linearly with T, the term above vanishes as $T \to \infty$ and the choices of the algorithm get arbitrarily close to the optimal q^* in terms of KL divergence with respect to the posterior $p_{\mathbf{x}}$.

4 BACKGROUND

This section briefly reviews background on Gaussian processes, Bayesian optimisation via upper confidence bounds, and MCMC algorithms, which we apply for our results.

4.1 GAUSSIAN PROCESS MODELS

A Gaussian process with mean function $m : \Theta \to \mathbb{R}$ and a positive-definite covariance function $k : \Theta \times \Theta \to \mathbb{R}$ represents a prior over functions $f : \Theta \to \mathbb{R}$ so that, for any finite collection of query points $\{\boldsymbol{\theta}_i\}_{i=1}^N \subset \Theta$, the vector $\mathbf{f}_N = [f(\boldsymbol{\theta}_1), \ldots, f(\boldsymbol{\theta}_N)]^\mathsf{T}$ is Gaussian, $\mathbf{f}_N \sim \mathcal{N}(\mathbf{m}_N, \mathbf{K}_N)$, where $[\mathbf{K}_N]_{ij} = k(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j)$. Given a set of observations $\mathcal{D}_N = \{(\boldsymbol{\theta}_i, z_i)\}_{i=1}^N$, where $z_i = f(\boldsymbol{\theta}_i) + \epsilon_i$, and $\epsilon_i \sim \mathcal{N}(0, \eta)$, the posterior over f is:

$$f(\boldsymbol{\theta}_*)|\mathcal{D}_N \sim \mathcal{N}(\mu_N(\boldsymbol{\theta}_*), \sigma_N^2(\boldsymbol{\theta}_*)) \,, \tag{8}$$

where:

$$\mu_N(\boldsymbol{\theta}_*) = m(\boldsymbol{\theta}_*) + \mathbf{k}_N(\boldsymbol{\theta}_*)^\mathsf{T}(\mathbf{K}_N + \eta\mathbf{I})^{-1}(\mathbf{z}_N - \mathbf{m}_N) \tag{9}$$

$$k_N(\boldsymbol{\theta}, \boldsymbol{\theta}') = k(\boldsymbol{\theta}, \boldsymbol{\theta}') - \mathbf{k}_N(\boldsymbol{\theta})^\mathsf{T}(\mathbf{K}_N + \eta\mathbf{I})^{-1}\mathbf{k}_N(\boldsymbol{\theta}') \tag{10}$$

$$\sigma_N^2(\boldsymbol{\theta}_*) = k_N(\boldsymbol{\theta}_*, \boldsymbol{\theta}_*) \,, \tag{11}$$

where $\mathbf{k}_N(\boldsymbol{\theta}_*) := [k(\boldsymbol{\theta}_1, \boldsymbol{\theta}_*), \ldots, k(\boldsymbol{\theta}_N, \boldsymbol{\theta})]^\mathsf{T}$.

To model log-likelihoods, we need to consider that in log-scale low-likelihood values tend to $-\infty$. Following previous works (Acerbi, 2018; Järvenpää et al., 2020), we use a negative quadratic mean function and, as covariance function, the squared-exponential kernel:

$$k(\boldsymbol{\theta}, \boldsymbol{\theta}') := \sigma_f^2 \exp\left(-\frac{1}{2l^2}\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_2^2\right) \,, \quad \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta \tag{12}$$

where $l > 0$ and $\sigma_f > 0$ are hyper-parameters (see Rasmussen and Williams, 2006, Ch. 4), and $\|\cdot\|_2$ denotes the Euclidean 2-norm of a vector. However, our algorithm and results are not restricted to this particular choice of kernel.

4.2 THE UPPER CONFIDENCE BOUND ALGORITHM

The upper confidence bound (UCB) algorithm has been a popular method in the multi-armed bandits and Bayesian optimisation literature due to its theoretical properties and simplicity (Abbasi-Yadkori et al., 2010; Srinivas et al., 2010; Shahriari et al., 2016; Chowdhury and Gopalan, 2017). The algorithm selects queries to solve a sequential optimisation problem by maximising an upper confidence bound over the objective function. In the case of Gaussian processes, the algorithm selects queries by:

$$\boldsymbol{\theta}_t \in \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} \mu_{t-1}(\boldsymbol{\theta}) + \beta_{t-1}\sigma_{t-1}(\boldsymbol{\theta}) \,, \tag{13}$$

where β_t is a parameter which can be adjusted based on the information gain of the GP model, and μ_t and σ_t^2 define the GP posterior mean and variance, respectively, at time $t \geq 0$. UCB algorithms enjoy bounded regret under reasonable assumptions in a point optimisation setting (Abbasi-Yadkori et al., 2010; Durand et al., 2018). In this paper, however, we consider a distribution optimisation setting.

4.3 MARKOV CHAIN MONTE CARLO

MCMC constitutes a broad class of sampling-based Bayesian inference algorithms, most of which enjoy asymptotic convergence guarantees (Andrieu et al., 2003). These methods formulate sampling as a Markov process whose stationary distribution is given by the desired target distribution, i.e. the posterior in Bayesian inference. The main building

block of the majority of current MCMC approaches is the Metropolis-Hastings (MH) algorithm (Hastings, 1970). In the Bayesian inference formulation, given a proposal distribution $\pi(\boldsymbol{\theta}'|\boldsymbol{\theta})$, the MH algorithm accepts or rejects a sample $\boldsymbol{\theta}'$ given the previous $\boldsymbol{\theta}$ with the following acceptance probability:

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}) := \min\left\{1, \frac{p(\mathbf{x}|\boldsymbol{\theta}')p(\boldsymbol{\theta}')\pi(\boldsymbol{\theta}|\boldsymbol{\theta}')}{p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta})\pi(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right\}. \quad (14)$$

Among the most popular choices of MCMC algorithms, we have Hamiltonian Monte Carlo (Neal, 2011) and its related self-adjusting variant, the no-U-turn sampler (NUTS) (Hoffman and Gelman, 2014). In addition, we also have ensemble samplers, which often deal better with multi-modal distributions (Foreman-Mackey et al., 2013).

5 DISTRIBUTIONAL BAYESIAN OPTIMISATION

We propose a BO algorithm for approximate Bayesian inference problems (Equation 2) with black-box and/or expensive-to-evaluate likelihood functions. The algorithm iterates between estimating posterior approximations and collecting likelihood evaluations, which are then passed on to update an internal GP model. In its general form, the algorithm may operate with any class of variational posteriors $\mathcal{Q}$. We, however, focus on the class of non-parametric sample-based approximations via MCMC, which offer theoretical and practical benefits, detailed in the next sections.

The algorithm is outlined in Algorithm 1. At each iteration, the algorithm selects a variational posterior by maximising a distributional acquisition function consisting of an upper-confidence bound on the ELBO (cf. Equation 3). From the variational posterior, we sample a batch of S parameters to evaluate the log-likelihood function ℓ via its estimator $\hat{\ell}$. The data from the evaluations is fed back to update the algorithm's internal GP model, and the cycle repeats for a given number of T iterations. In the following, we present the proposed distributional acquisition function and a method to directly sample from its maximiser.

5.1 KULLBACK-LEIBLER UPPER CONFIDENCE BOUND

As the log-likelihood $\ell(\boldsymbol{\theta})$ is the only unknown term in Equation 3, a GP model on $\ell : \Theta \to \mathbb{R}$ allows us to derive an upper confidence bound on the ELBO. From this observation, we define the following acquisition function:

$$h(q|\mathcal{D}_t) := \mathbb{E}_{\boldsymbol{\theta}\sim q}[u_t(\boldsymbol{\theta})] - \mathbb{D}_{\mathrm{KL}}(q||p)\,, \quad t \geq 0\,, \quad (15)$$

where $u_t(\boldsymbol{\theta}) := \mu_t(\boldsymbol{\theta}) + \beta_t\sigma_t(\boldsymbol{\theta})$ corresponds to a pointwise UCB (Srinivas et al., 2010; Durand et al., 2018), and we

Algorithm 1: KL-UCB

$\mathcal{D}_0 := \varnothing$
for $t \in \{1, \ldots, T\}$ **do**
  $q_t \leftarrow \operatorname{argmax}_{q\in\mathcal{Q}} h(q|\mathcal{D}_{t-1})$
  Sample $\{\boldsymbol{\theta}_{t,i}\}_{i=1}^{S} \sim q_t$
  Evaluate $z_{t,i} = \hat{\ell}(\boldsymbol{\theta}_{t,i})$, $i \in \{1, \ldots, S\}$
  Update GP with $\mathcal{D}_t := \mathcal{D}_{t-1} \cup \{\boldsymbol{\theta}_{t,i}, z_{t,i}\}_{i=1}^{S}$

derive settings for $\beta_t > 0$ in Section 6. With this acquisition function, we select candidate distributions as:

$$q_t \in \operatorname*{argmax}_{q\in\mathcal{Q}} h(q|\mathcal{D}_{t-1}), \quad t \geq 1\,. \quad (16)$$

Comparing h in Equation 15 with the main ELBO formulation (Equation 3), we notice that Equation 15 defines an ELBO with respect to a surrogate posterior obtained by taking the pointwise UCB u_t as a log-likelihood function:

$$\hat{p}_{t-1}(\boldsymbol{\theta}) := \frac{p(\boldsymbol{\theta})\exp u_{t-1}(\boldsymbol{\theta})}{\int_\Theta p(\boldsymbol{\theta}')\exp u_{t-1}(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{\theta}'}\,, \quad t \geq 1\,. \quad (17)$$

The equation above defines a valid probability density function as long as u_t is bounded, which is usually the case for practical GP models. We formalise this result in Section 6 (Lemma 2). Due to its connection with the KL divergence, we refer to the acquisition function in Equation 15 as *Kullback-Leibler upper confidence* bound (KL-UCB).

5.2 OPTIMISATION OVER PROBABILITY DISTRIBUTIONS VIA MCMC

We now address the choice of the class $\mathcal{Q}$ of variational posteriors and how to sample from the optimal candidates $q_t \in \mathcal{Q}$ at each BO iteration (cf. Equation 16). As discussed in the previous section, KL-UCB as defined in Equation 15 corresponds to an ELBO with respect to $\hat{p}_t$ in Equation 17, while also being an upper confidence bound on the main ELBO in Equation 3. Therefore, for a large enough class of distributions $\mathcal{Q}$, the maximiser q_t of Equation 16 is $\hat{p}_{t-1}$ itself, and we may sample directly from it, bypassing the optimisation step in Equation 16.

To sample from the candidates $q_t := \hat{p}_{t-1}$ at each BO iteration, we need to consider that the denominator in Equation 17 involves a possibly intractable integral. We address this issue by applying MCMC to sample, which only requires the unnormalised density $p(\boldsymbol{\theta})\exp u_{t-1}(\boldsymbol{\theta})$.

Due to its asymptotic convergence guarantees (Andrieu et al., 2003), running MCMC with $\hat{p}_{t-1}$ as our target distribution gives us samples which converge in distribution to $\hat{p}_{t-1}$, the optimiser of Equation 16. We may see MCMC in this case as a form of optimisation over a space of non-parametric probability distributions. Further discussions on

the topic of MCMC as optimisation are available in the literature of gradient-based methods (Wibisono, 2018).

Note that we do not need to draw an infinite amount of samples from the MCMC chain. We only need to run it so that a batch of $S < \infty$ samples coming out of the chain approximately follow $\hat{p}_t$, i.e. after a long enough *burn-in* period, or by initialising the chain at a high-probability region, such as a mode of $\hat{p}_t$ (van Ravenzwaaij et al., 2018). In our experiments, we use the EMCEE sampler (Foreman-Mackey et al., 2013) which can deal with the multi-modality of the UCB surface at reasonably fast mixing times.

6 THEORETICAL RESULTS

In this section, we present theoretical results on the regret of Algorithm 1 with respect to the objective in Equation 2. UCB's theoretical guarantees are based on uniform concentration bounds. In particular, we use the following result for point-wise approximations.

Theorem 1 (Durand et al. (2018))**.** *Let $k(\boldsymbol{\theta}, \boldsymbol{\theta}) \leq 1$ for all $\boldsymbol{\theta} \in \Theta$, $\|\ell\|_k \leq b$ and ϵ_t be conditionally σ_ϵ-sub-Gaussian. Then, for any $\delta \in (0, 1]$, with probability at least $1 - \delta$, uniformly over all $t \geq 0$ and $\boldsymbol{\theta} \in \Theta$,*

$$|\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})| \leq \beta_t(\delta)\sigma_t(\boldsymbol{\theta}) \,,$$

where $\beta_t(\delta) = b + \sigma_\epsilon \sqrt{2\eta^{-1} \log(|\mathbf{I} + \eta^{-1}\mathbf{K}_t|^{1/2}/\delta)}$.

Notice that the confidence bound in Theorem 1 is simultaneous over all Θ and is currently the tightest known bound in the UCB literature. In our case, we extend the result above to a concentration bound over the space of probability distributions $\mathcal{P}$ with support on Θ.

Lemma 1. *Under our regularity assumptions (cf. Section 3), with probability at least $1 - \delta$, $\delta \in (0, 1)$, we have that the following holds for all $t \geq 0$:*

$$|\mathbb{E}_{\boldsymbol{\theta}\sim q}[\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})]| \leq \beta_t(\delta)\mathbb{E}_{\boldsymbol{\theta}\sim q}[\sigma_t(\boldsymbol{\theta})] \,, \quad \forall q \in \mathcal{P} \,, \tag{18}$$

where $\beta_t(\delta) := b + \sigma_\epsilon \sqrt{2\eta^{-1} \log(|\mathbf{I} + \eta^{-1}\mathbf{K}_{\mathcal{D}_t}|^{1/2}/\delta)}$, $\mathbf{K}_{\mathcal{D}} := [k(\boldsymbol{\theta}, \boldsymbol{\theta}')]_{\boldsymbol{\theta},\boldsymbol{\theta}' \in \mathcal{D}}$ represents the GP covariance matrix for a set of observations $\mathcal{D}$, and $|\mathbf{A}|$ denotes the determinant of a square matrix $\mathbf{A}$.

Proof sketch. This result follows by an application of Jensen's inequality to the confidence bound in Durand et al. (2018, Theorem 1) for GP-UCB. □

The result in Lemma 1 allows us to place a high-probability upper bound on the instant regret of Algorithm 1 when using the acquisition function in Equation 15 and the UCB parameter set to $\beta_t := \beta_t(\delta)$. In addition to the regularity assumptions in Section 3, we need to place conditions on the sampling process $\{\boldsymbol{\theta}_i\}_{i=1}^S$ for the bounds to hold. We defer full proofs to the supplementary material.

The main result depends on the information gain of the GP, which is bounded by:

$$\gamma_N := \sup_{\Theta_N \subset \Theta : |\Theta_N| = N} I(\mathbf{z}_N, \mathbf{g}_N | \Theta_N) \,, \tag{19}$$

where $I(\mathbf{z}_N, \mathbf{g}_N|\Theta_N)$ denotes the mutual information between $\mathbf{z}_N = \mathbf{g}_N + \boldsymbol{\epsilon}'_N$ and $\mathbf{g}_N \sim \mathcal{N}(\mathbf{0}, \mathbf{K}_N)$, with $[\mathbf{K}_N]_{ij} = k(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j), \boldsymbol{\theta}_i, \boldsymbol{\theta}_j \in \Theta_N$ and $\boldsymbol{\epsilon}'_N \sim \mathcal{N}(\mathbf{0}, \eta\mathbf{I})$. For a compact parameter space Θ, or equivalently when dealing with a compactly supported prior $p(\boldsymbol{\theta})$, the supremum in Equation 19 corresponds to a maximum, which is referred to as the *maximum information gain*. For popular choices of kernels, information gain bounds are available in the literature, such as the squared exponential kernel, which is $\mathcal{O}((\log T)^{d+1})$ (Srinivas et al., 2010). We can now provide the following bound on the cumulative regret of Algorithm 1.

Theorem 2. *Under the same assumptions in Lemma 1, running Algorithm 1 with $h(q|\mathcal{D}_{t-1})$ given by Equation 15 and $\beta_t := \beta_t(\delta)$, we obtain a bound on the instant regret:*

$$r_t \leq 2\beta_{t-1}(\delta)\mathbb{E}_{q_t}[\sigma_{t-1}] \,, \quad t \geq 1 \,, \tag{20}$$

and on the cumulative regret at $T \geq 1$:

$$R_T \leq 4\beta_T(\delta)\sqrt{(T+2)\gamma_T} \in \mathcal{O}(\sqrt{T}(b\sqrt{\gamma_T} + \sqrt{\gamma_T \gamma_{ST}})) \tag{21}$$

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

Proof sketch. The first result follows by applying the upper confidence bound on ℓ to the regret defined in Equation 5 for an optimal choice according to Equation 16. The second result follows a similar derivation to previous bounds in the literature (cf. Chowdhury and Gopalan, 2017, Theorem 2). □

For common choices of kernel functions, such as the squared exponential and the Matérn kernels, the maximum information gain has sub-linear growth (Srinivas et al., 2010). Consequently, Theorem 2 shows that Algorithm 1 achieves vanishing regret under these settings, which allows the algorithm to find arbitrarily close approximations to q^* within a finite amount of time. In particular, let us consider the setting where we run exact inference on Equation 16.

Lemma 2. *For any bounded kernel k and bounded GP mean function m, maximising h in Equation 15 at any $t \geq 1$ is equivalent to:*

$$\operatorname*{argmax}_{q \in \mathcal{Q}} h(q|\mathcal{D}_{t-1}) = \operatorname*{argmin}_{q \in \mathcal{Q}} \mathbb{D}_{\mathrm{KL}}(q||\hat{p}_{t-1}) \,. \tag{22}$$

Lemma 2 tells us that the maximiser of $h(q|\mathcal{D}_{t-1})$ is unique and given by $\hat{p}_{t-1}$ when $\hat{p}_{t-1} \in \mathcal{Q}$. In fact, we can sample from $\hat{p}_{t-1}$ via MCMC, which leads us to the final result.

Corollary 1. *Under the regularity assumptions in Section 3, with probability greater than* $1-\delta$*,* $\delta \in (0,1)$*, we have:*

$$\forall t \geq 0, \quad \mathbb{D}_{\mathrm{KL}}(\hat{p}_t || p_{\mathbf{x}}) \leq 2\beta_t(\delta)\mathbb{E}_{\hat{p}_t}[\sigma_t]\,. \tag{23}$$

In particular, for $\gamma_T \in \mathcal{O}(T^\alpha)$ *with* $\alpha < 1/2$*, we have:*

$$\lim_{T\to\infty} \min_{t\leq T} \mathbb{D}_{\mathrm{KL}}(\hat{p}_t || p_{\mathbf{x}}) = 0\,. \tag{24}$$

This result provides a finite-sample bound on the KL divergence of posterior approximations (see Appendix C for kernel-specific bounds) and confirms that the GP-based posterior approximation asymptotically approaches the true posterior $p_{\mathbf{x}}$ for the case of popular kernels, such as the squared exponential and the Matérn family, which lead to sub-linear regret bounds (Srinivas et al., 2010). The result also tells us that a sequence of MCMC samples over a UCB-based likelihood leads us to optimal posterior approximations and a no-regret approximate inference algorithm.

6.1 CONNECTION WITH OTHER APPROACHES

Contrast with point-based approach: In contrast with applying UCB to learn a log-likelihood model via a point-maximisation problem (Gutmann and Corander, 2016), Theorem 2 establishes a regret bound on distribution choices. The distributional regret bound is $\mathcal{O}(\sqrt{T\gamma_T} + \sqrt{\gamma_T\gamma_{ST}})$, essentially smaller than the regret bound of a GP-UCB algorithm given the same number of likelihood evaluations $N = ST$, which is $\mathcal{O}(\sqrt{ST\gamma_{ST}} + \gamma_{ST})$ (Chowdhury and Gopalan, 2017) by a factor dependent on the sample batch size S, since $\gamma_{ST} \geq \gamma_T$ by monotonicity. In addition, samples from q_t consider the likelihood surface as a whole via the expectation within the ELBO, while a pointwise UCB algorithm would be ultimately concerned with finding the parameter of maximum likelihood.

Connection with VBMC: VBMC (Acerbi, 2018) selects variational posteriors by maximising a lower confidence bound on the ELBO based on GP estimates. For a GP model on the log-likelihood[2], VBMC optimises:

$$h_{\mathrm{VBMC}}(q|\mathcal{D}_t) := \mathbb{E}_q[\mu_t] - \mathbb{D}_{\mathrm{KL}}(q||p) - \beta\sigma_t(q)\,, \tag{25}$$

where $\beta \geq 0$ is fixed (e.g. $\beta = 3$), and $\sigma^2_{t-1}(q) := \int_\Theta \int_\Theta k_{t-1}(\boldsymbol{\theta},\boldsymbol{\theta}')q(\boldsymbol{\theta})q(\boldsymbol{\theta}')\,\mathrm{d}\boldsymbol{\theta}\,\mathrm{d}\boldsymbol{\theta}'$ corresponds to the variance of the expected log-likelihood. From the variational posterior q_t^{VBMC} that maximises $h_{\mathrm{VBMC}}(q|\mathcal{D}_{t-1})$, VBMC selects a sequence of points for evaluation:

$$\boldsymbol{\theta}_{t+i}^{\mathrm{VBMC}} \in \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} \sigma^2_{t+i-1}(\boldsymbol{\theta}) q_t^{\mathrm{VBMC}}(\boldsymbol{\theta}) \exp \mu_{t+i-1}(\boldsymbol{\theta})\,, \tag{26}$$

where each point is chosen after updating the GP with the evaluation at the previous point for a total of S_{VBMC} points (set to 5 by default). VBMC uses a class of non-parametric variational posteriors consisting of mixtures of Gaussians with an adaptible number of components and performs a warm-up run to optimise GP hyper-parameters.

Equation 25 corresponds to a *lower* confidence bound on the ELBO (Acerbi, 2018) in contrast to KL-UCB (cf. Equation 15), which sets an *upper* confidence bound. Therefore, if the lower confidence bound in VBMC is replaced by KL-UCB, the algorithm should have a similar distributional regret bound to the one in Theorem 2, since the stochastic process producing the batch does not violate the assumptions in Durand et al. (2018) for Theorem 1 (see also the assumptions in Chowdhury and Gopalan (2017)).

6.2 LIMITATIONS

Our distributional Bayesian optimisation algorithm KL-UCB comes with strong theoretical guarantees allowing one not only to learn approximations of a posterior, but also to quantify the error in the approximation. However, the algorithm also comes with its own limitations. Although not explicit in the regret bounds, the dimensionality of the parameter space Θ affects the bounds via the dependence of the maximum information gain γ_T on it. For instance, the squared-exponential kernel (Equation 12) yields $\gamma_T \in \mathcal{O}(\log^{d+1} T)$ (Srinivas et al., 2010), which is exponential in the dimension of the parameter space. In addition, the inversion of the kernel matrix required for GP predictions scales as $\mathcal{O}(N^3)$ in time complexity. Therefore, our method is more suitable for applications with low-dimensional parameter spaces and where we are limited to a small number of likelihood evaluations. Sparse approximations to GPs (Gijsberts and Metta, 2013; Hensman et al., 2013) and advances in BO methods for high-dimensional search spaces (Wang et al., 2018; Mutný and Krause, 2018) may alleviate these issues, but adapting these frameworks to this problem setup is objective of future research.

7 EXPERIMENTS

In this section, we present empirical assessments of the proposed KL-UCB algorithm to complement the theoretical results in the previous section. We start with experiments assessing the theoretical regret bounds in situations where the modelling assumptions hold. We then follow with experiments comparing KL-UCB against VBMC (Acerbi, 2018), as an adaptive Bayesian quadrature baseline, in two test cases where the log-likelihood function is not in the RKHS defined by the GP kernel. Lastly, we present an experiment with a likelihood-free inference problem using a single-point UCB strategy as a baseline (Gutmann and Corander, 2016).

[2]The GP model in VBMC is learnt over the log-joint probability $p(\mathbf{x},\boldsymbol{\theta})$, which is still compatible with our setup, with the exception that the prior is known.

For all experiments, KL-UCB was configured with a standard normal prior and a squared exponential kernel. To perform inference over the UCB-based posterior approximation $\hat{p}_{t-1}$, we run the ensemble-based MCMC sampler EMCEE (Foreman-Mackey et al., 2013) set with 25 walkers to draw each iteration's batch of $S := 5$ after a burn-in period. We set $\delta := 0.2$ and adjust the kernel hyper-parameters accordingly. Appendix E provides additional experiment details.

Results are reported in terms of the number of evaluations of the log-likelihood. As performance metrics, we use both the KL divergence and the Gaussian symmetrised KL divergence (gsKL) used by Acerbi (2018). The latter is defined as the symmetric KL divergence between the Gaussian distributions formed by the first two moments of the approximate and the true posterior distribution. In cases where the posterior distribution is not available in closed form, we also run MCMC and kernel density estimation to recover a reference posterior density.

Ablation experiments in the supplementary material analyse the effect of dimensionality in the current framework (see Appendix D). Code and further ablation experiments are also to be made available online.[3]

7.1 THEORY ASSESSMENT

We first assess how the theoretical results reflect in terms of practical performance in settings where the regularity assumptions are known to hold. For this experiment, we generate log-likelihood functions in the RKHS $\mathcal{H}_k$:

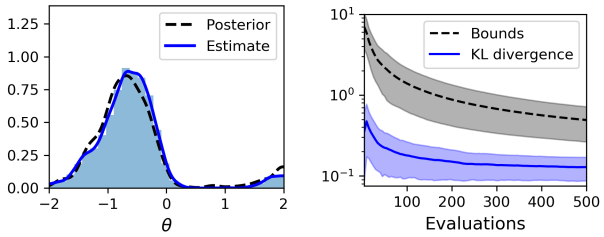
$$\ell := \sum_{i=1}^{M} w_i k(\cdot, \boldsymbol{\theta}_i) \in \mathcal{H}_k \,, \tag{27}$$

where we sample $w_i \sim \mathcal{N}(0,1)$, $\boldsymbol{\theta}_i \sim p(\boldsymbol{\theta})$, for $i \in \{1, \dots, M\}$, with a fixed $M := 40$.

Figure 1 presents the posterior approximation results for a 1-D RKHS log-likelihood. As seen in Figure 1a, the final posterior approximation, which uses the GP posterior mean as a log-likelihood, mostly matches the true posterior. The exception is a slight under-estimation of the second, smaller mode. This, however, is a well-known effect of using the reverse KL, which induces a mode-seeking behaviour (Wang and Titterington, 2005; Bishop, 2007).

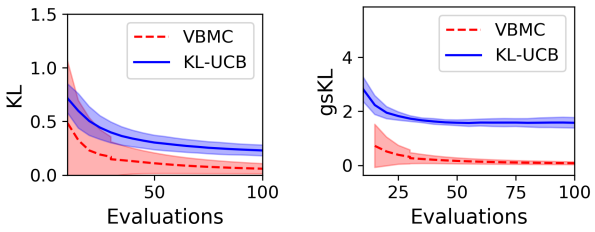
In Figure 1b, the regret plot shows that the bound in Theorem 2 is large enough to accommodate for the measured KL divergence between the proposed distributions and the true posterior. The regret bound is calculated based on the formulation of β_t in Theorem 1 using the norm of ℓ in $\mathcal{H}_k$. Details about the KL divergence estimation are available in the supplementary material (see Appendix E).

[3]Code repository: `https://github.com/rafaol/no-regret-approximate-inference-via-bo`



(a) Approximation (b) Regret

Figure 1: Results for theory assessment experiment. Plot (a) shows the final posterior approximation by KL-UCB (solid line) alongside the true synthetic posterior (dashed line) for one of the trials. Plot (b) shows the averaged cumulative regret R_T/T as a function of the number of likelihood evaluations. Results were averaged over 10 trials, and the shaded area corresponds to ± 1 standard deviation.



(a) KL divergence (b) gsKL

Figure 2: Performance on 2D lumpy likelihood. Results were averaged over 5 trials. Shaded areas correspond to ± 1 standard deviation.

7.2 COMPARISONS

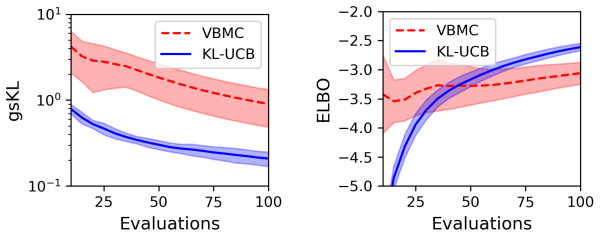
In this section, we test KL-UCB on settings where the main assumption $\ell \in \mathcal{H}_k$ does not hold. As a related approximate inference baseline, we use VBMC (Acerbi, 2018).

Lumpy likelihood: This likelihood function is formulated as a mixture of equally weighted Gaussians:

$$p(\mathbf{x}|\boldsymbol{\theta}) = \exp \ell(\boldsymbol{\theta}) = \frac{1}{M} \sum_{i=1}^{M} \mathcal{N}(\boldsymbol{\theta}; \boldsymbol{\theta}_i, \boldsymbol{\Sigma}_i) \,, \tag{28}$$

where we sample $\boldsymbol{\theta}_i \sim \mathcal{U}[0,1]^d$ and $\boldsymbol{\Sigma}_i = \mathrm{diag}(\sigma_{i,1}^2, \dots, \sigma_{i,d}^2)$ with $\sigma_{i,j} \sim \mathcal{U}[0.2, 0.6]$, $j \in \{1, \dots, d\}$ for a fixed number of components $M := 12$, following the setup in Acerbi (2018). Placing a Gaussian prior $p(\boldsymbol{\theta}) := \mathcal{N}(\boldsymbol{\theta}; \mathbf{0}, \sigma_p^2 \mathbf{I})$ with $\sigma_p := 0.5$ on $\boldsymbol{\theta}$, the resulting posterior is available in closed-form as also a mixture of Gaussians with properly adjusted means and covariances.

Figure 2 presents the results for the experiment on the lumpy likelihood function. As we can see, KL-UCB's performance



(a) gsKL (b) ELBO

Figure 3: Performance on experiment with circular likelihood function. Results were averaged over 5 trials. Shaded areas correspond to ± 1 standard deviation.

is surpassed by that of VBMC. As this problem setting involves a mixture of Gaussians as a likelihood, VBMC has a natural advantage, since this algorithm uses variational posteriors based on mixtures of Gaussians as well.

Circular likelihood: We define a circular 2-dimensional log-likelihood function by:

$$\ell(\boldsymbol{\theta}) := -(\|\boldsymbol{\theta} - \boldsymbol{\theta}_c\|_2 - \rho_c)^2 / s_c \,, \tag{29}$$

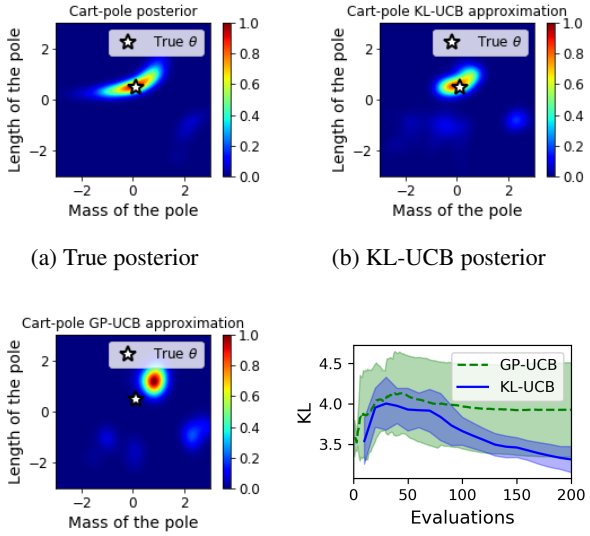
where $\boldsymbol{\theta}_c := \mathbf{0}$ denotes the circle centre, $\rho_c := 1.5$ is the circle radius, and $s_c := 0.25$ is a length-scale.

In Figure 3, we see the performance results for the circular likelihood. This time KL-UCB noticeably outperforms VBMC. One of the reasons for this outcome is that to be able to cover the narrow circle defined by the likelihood function, VBMC needs a large number of mixture components (reaching 50 in our tests). KL-UCB instead samples on top of the non-parametric surface defined by the GP directly.

7.3 LIKELIHOOD-FREE INFERENCE

We applied KL-UCB to a likelihood-free inference problem. In this experiment, we infer the parameters of a classic control environment in OpenAI Gym[4]: the cart-pole. We fix a given setting for its physics parameters $\boldsymbol{\theta}^{\text{real}}$ and generate a dataset of 10 trajectories by executing randomly sampled actions. Summary statistics $\boldsymbol{\xi}$ were the same as Ramos et al. (2019), which consist of state-action inner products and the first two moments of the trajectory state-differences distribution. The discrepancy was set to $\Delta_{\boldsymbol{\theta}} := \|\boldsymbol{\xi}_{\boldsymbol{\theta}} - \boldsymbol{\xi}^{\text{real}}\|_2^2 / \sigma^2$, and we evaluate $\hat{\ell}(\boldsymbol{\theta}) := -\Delta_{\boldsymbol{\theta}}$. We also placed a Gaussian prior $p(\boldsymbol{\theta}) := \mathcal{N}(\boldsymbol{\theta}; \mathbf{0}, \mathbf{I})$ on the parameters.

Our GP model was configured with the Matérn kernel set with smoothness parameter $\nu := 3/2$ (Rasmussen and Williams, 2006). Other hyper-parameters include the kernel length-scales, signal variance and noise variance, and were



(a) True posterior (b) KL-UCB posterior

(c) GP-UCB posterior (d) KL regret

Figure 4: Cart-pole likelihood-free inference problem. Plot (a) shows the reference true posterior obtained by ABC. Plot (b) shows a posterior estimated via the batch-based KL-UCB, and plot (c) shows an estimate by the point-to-point GP-UCB. Plot (d) shows the KL-divergence regret of each method averaged over 5 runs, with shaded areas corresponding to ± 1 standard deviation.

adjusted based on maximum a posteriori estimates from previous runs. The GP mean function was set as the log-prior probability density $m(\boldsymbol{\theta}) := \log p(\boldsymbol{\theta})$.

We compared KL-UCB against GP-UCB in this experiment. Both methods were configured with a fixed $\beta_t := 3$, $\forall t \geq 0$. The main difference between the methods is that GP-UCB sequentially selects single points for evaluations, which maximise the pointwise UCB, while KL-UCB selects an i.i.d. batch of points according to the UCB-based posterior approximation. The GP-UCB approach in this case corresponds to the approach by Gutmann and Corander (2016), providing a baseline as a point-based likelihood-learning method.

Figure 4 presents the final posterior approximations obtained by KL-UCB and the KL-divergence-based regret for each method. As a reference for the "true" posterior, we ran rejection ABC[5] with an adaptable threshold parameter to obtain 1000 samples, and then fit a kernel density estimator to measure the KL divergences. As the plots show, KL-UCB was able to obtain approximations much closer to the true posterior when compared to GP-UCB. It is worth noting that the true posterior's mass is highly concentrated on a narrow region around the true parameters, making the inference problem relatively hard for a point-by-point method.

[4]OpenAI Gym: `https://gym.openai.com`

[5]ABC code: `https://github.com/elfi-dev/elfi`

In terms of regret, we measured the KL divergence between the corresponding UCB-based posterior approximation and the true posterior for both methods. These results are also in Figure 4. Even though GP-UCB and KL-UCB had the same settings for their GP models, the results show that, within the given budget of 200 evaluations of the discrepancy function, GP-UCB was not able to obtain reasonable posterior approximations. Its performance is also more unstable.

8 CONCLUSION

This paper presented an approach for approximate Bayesian inference via Bayesian optimisation in settings where we need to evaluate black-box likelihood functions. The algorithm is composed of an acquisition function formulated as an upper confidence bound over the ELBO of a variational objective. In contrast to traditional variational inference approaches, the algorithm samples the non-parametric posterior approximation defined by a GP model over the log-likelihood function to propose evaluation points at each iteration. The use of a flexible non-parametric class of distributions allows us to propose a sequence of optimal distributional candidates, which theoretical results have shown to lead to a vanishing KL divergence. A set of experiments also demonstrated the method in practice. As future work, the method offers a few avenues for improvement, such as possible batch design approaches (Järvenpää et al., 2020) on top of the MCMC samples, scalability to high-dimensional spaces and applications to general optimisation problems.

Bibliography

Yasin Abbasi-Yadkori, David Pal, and Csaba Szepesvari. Improved Algorithms for Linear Stochastic Bandits. In *Advances in Neural Information Processing Systems*, volume 23, 2010.

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.

Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50(1-2):5–43, 2003.

Mark A Beaumont. Approximate bayesian computation in evolution and ecology. *Annual review of ecology, evolution, and systematics*, 41:379–406, 2010.

Eli Bingham, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman. Pyro: Deep Universal Probabilistic Programming. *Journal of Machine Learning Research*, 20, 2019.

Christopher M. Bishop. Approximate Inference. In *Pattern Recognition and Machine Learning*, chapter 10, pages 461–522. Springer, 2007.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

Johann Brehmer, Kyle Cranmer, Gilles Louppe, and Juan Pavez. A guide to constraining effective field theories with machine learning. *Phys. Rev. D*, 98, Sep 2018.

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.

Kyle Cranmer, Johann Brehmer, and Gilles Louppe. The frontier of simulation-based inference. *Proceedings of the National Academy of Sciences*, 117(48):30055–30062, 2020.

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.

Audrey Durand, Odalric-Ambrym Maillard, and Joelle Pineau. Streaming kernel regression with provably adaptive mean, variance, and regularization. *Journal of Machine Learning Research*, 19(1):650–683, 2018.

Conor Durkan, Iain Murray, and George Papamakarios. On Contrastive Learning for Likelihood-free Inference. In *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, Vienna, Austria, 2020. PMLR: v. 119.

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.

Arjan Gijsberts and Giorgio Metta. Real-time model learning using Incremental Sparse Spectrum Gaussian Process Regression. *Neural Networks*, 41:59–69, 2013.

David Greenberg, Marcel Nonnenmacher, and Jakob Macke. Automatic posterior transformation for likelihood-free inference. In Kamalika Chaudhuri and Ruslan

Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 2404–2414, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

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.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1): 97–109, 04 1970.

James Hensman, Nicolo Fusi, and Neil D. Lawrence. Gaussian Processes for Big Data. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pages 282–290, Bellevue, WA, USA, 2013.

Matthew D Hoffman and Andrew Gelman. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15 (1):1593–1623, 2014.

Marko Järvenpää, Michael U. Gutmann, Arijus Pleska, Aki Vehtari, and Pekka Marttinen. Efficient acquisition rules for model-based approximate Bayesian computation. In *Advances in Approximate Bayesian Inference, NIPS 2017 Workshop*, 2017.

Marko Järvenpää, Michael U. Gutmann, Aki Vehtari, and Pekka Marttinen. Parallel Gaussian Process Surrogate Bayesian Inference with Noisy Likelihood Evaluations. *Bayesian Analysis*, pages 1–32, 2020.

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

Mojmír Mutný and Andreas Krause. Efficient high dimensional Bayesian optimization with additivity and quadrature Fourier features. In *Advances in Neural Information Processing Systems*, volume 31, Montreal, Canada, 2018.

Radford M. Neal. MCMC using Hamiltonian dynamics. In Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, editors, *Handbook of Markov Chain Monte Carlo*, pages 113–162. Chapman & Hall, 2011.

Victor M.H. Ong, David J. Nott, Minh Ngoc Tran, Scott A. Sisson, and Christopher C. Drovandi. Variational Bayes with synthetic likelihood. *Statistics and Computing*, 28 (4):971–988, 2018.

George Papamakarios and Iain Murray. Fast ϵ-free Inference of Simulation Models with Bayesian Conditional Density Estimation. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

George Papamakarios, David Sterratt, and Iain Murray. Sequential neural likelihood: Fast likelihood-free inference with autoregressive flows. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89 of *Proceedings of Machine Learning Research*, pages 837–848. PMLR, 2019.

Fabio Ramos, Rafael Carvalhaes Possas, and Dieter Fox. BayesSim : adaptive domain randomization via probabilistic inference for robotics simulators. In *Robotics: Science and Systems (RSS)*, Freiburg im Breisgau, Germany, 2019.

Rajesh Ranganath, Sean Gerrish, and David M. Blei. Black Box Variational Inference. In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS)*, Reykjavik, Iceland, 2014.

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

Christian P. Robert. Approximate Bayesian Computation: A Survey on Recent Results. In *Monte Carlo and Quasi-Monte Carlo Methods*, volume 163. Springer, Cham, 2016.

John Salvatier, Thomas V Wiecki, and Christopher Fonnesbeck. Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2:e55, 2016.

Bernhard Schölkopf and Alexander J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge, MA, 2002.

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

Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.

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.

Don van Ravenzwaaij, Pete Cassey, and Scott D. Brown. A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic Bulletin and Review*, 25:143–154, 2018.

Bo Wang and D. M. Titterington. Inadequacy of interval estimates corresponding to variational Bayesian approximations. In Robert G. Cowell and Zoubin Ghahramani, editors, *Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics (AISTATS), Jan 6-8, 2005, Savannah Hotel, Barbados*, pages 373–380. Society for Artificial Intelligence and Statistics, 2005.

Zi Wang, Clement Gehring, Pushmeet Kohli, and Stefanie Jegelka. Batched Large-scale Bayesian Optimization in High-dimensional Spaces. In *Proceedings of the 21st International Conference on Artificial Intelligence and Statistics (AISTATS)*, Lanzarote, Spain, 2018. JMLR: W&CP.

J Florian Wellmann, Franklin G Horowitz, Eva Schill, and Klaus Regenauer-Lieb. Towards incorporating uncertainty of structural data in 3d geological inversion. *Tectonophysics*, 490(3-4):141–151, 2010.

Andre Wibisono. Sampling as optimization in the space of measures: The Langevin dynamics as a composite optimization problem. In *31st Annual Conference on Learning Theory (COLT)*, volume 75. PMLR, 2018.

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 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:*

$$\mathbb{D}_{\mathrm{KL}}(Q||P) := \int \log\left(\frac{\mathrm{d}Q}{\mathrm{d}P}\right) \mathrm{d}Q \,. \tag{30}$$

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:

$$\mathbb{D}_{\mathrm{KL}}(Q||P) = \mathbb{D}_{\mathrm{KL}}(q||p) := \int_{\mathbb{R}^d} q(\boldsymbol{\theta}) \log\left(\frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta})}\right) \mathrm{d}\boldsymbol{\theta} \,, \tag{31}$$

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:

$$\mu_t(\boldsymbol{\theta}) := m(\boldsymbol{\theta}) + \mathbf{k}_{N_t}(\boldsymbol{\theta})^{\mathsf{T}}(\mathbf{K}_{N_t} + \eta\mathbf{I})^{-1}(\mathbf{z}_{N_t} - \mathbf{m}_{N_t}) \tag{32}$$

$$k_t(\boldsymbol{\theta}, \boldsymbol{\theta}') := k(\boldsymbol{\theta}, \boldsymbol{\theta}') - \mathbf{k}_{N_t}(\boldsymbol{\theta})^{\mathsf{T}}(\mathbf{K}_{N_t} + \eta\mathbf{I})^{-1}\mathbf{k}_{N_t}(\boldsymbol{\theta}') \tag{33}$$

$$\sigma_t^2(\boldsymbol{\theta}) := k_t(\boldsymbol{\theta}, \boldsymbol{\theta}) \,, \tag{34}$$

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

$$\forall t \geq 0, \quad \|\sigma_t\|_\infty \leq 1. \tag{35}$$

Proof. As the kernel k is bounded, with $k(\boldsymbol{\theta}, \boldsymbol{\theta}) \leq 1, \forall \boldsymbol{\theta} \in \Theta$, the definition in Equation 34 leads to:

$$\forall t \geq 1, \quad \sigma_t^2(\boldsymbol{\theta}) \leq k(\boldsymbol{\theta}, \boldsymbol{\theta}) \leq 1, \quad \forall \boldsymbol{\theta} \in \Theta \,. \tag{36}$$

□

Lemma 4. *Assuming a bounded mean function $m : \Theta \to \mathbb{R}$, the GP posterior mean is bounded, i.e.:*

$$\forall t \geq 0, \quad \|\mu_t\|_\infty < \infty \,. \tag{37}$$

Proof. As defined in Equation 32, for $t \geq 1$, we have:

$$\begin{aligned}
\mu_t(\boldsymbol{\theta}) &:= m(\boldsymbol{\theta}) + \mathbf{k}_{N_t}(\boldsymbol{\theta})^{\mathsf{T}}(\mathbf{K}_{N_t} + \eta\mathbf{I})^{-1}(\mathbf{z}_{N_t} - \mathbf{m}_{N_t}) \\
&\leq m(\boldsymbol{\theta}) + \eta^{-1}\mathbf{k}_{N_t}(\boldsymbol{\theta})^{\mathsf{T}}(\mathbf{z}_{N_t} - \mathbf{m}_{N_t}) \\
&\leq \|m\|_\infty + \eta^{-1}\|\mathbf{k}_{N_t}(\boldsymbol{\theta})\|_1\|\mathbf{z}_{N_t} - \mathbf{m}_{N_t}\|_\infty \\
&\leq \|m\|_\infty + \eta^{-1}t(\|\mathbf{z}_{N_t}\|_\infty + \|m\|_\infty) \\
&< \infty \,, \quad \forall \boldsymbol{\theta} \in \Theta \,,
\end{aligned} \tag{38}$$

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 5 (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:*

$$I(\mathbf{z}_N, \mathbf{f}_N|\Theta_N) = \frac{1}{2}\sum_{i=1}^{N} \log(1 + \eta^{-1}\sigma_{i-1}^2(\boldsymbol{\theta}_i)) \,. \tag{39}$$

Lemma 6 (Chowdhury and Gopalan (2017, Lemma 4))**.** *Following the setting of Lemma 5, the sum of predictive standard deviations at a sequence of evaluation points is bounded in terms of the maximum information gain:*

$$\sum_{i=1}^{N} \sigma_{i-1}(\boldsymbol{\theta}_i) \leq \sqrt{4(N+2)\gamma_N} \,. \tag{40}$$

Lemma 7. *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:*

$$k_{\mathcal{A}}(\boldsymbol{\theta}, \boldsymbol{\theta}') := k(\boldsymbol{\theta}, \boldsymbol{\theta}') - k(\boldsymbol{\theta}, \mathcal{A})^{\mathsf{T}}(\mathbf{K}(\mathcal{A}) + \eta\mathbf{I})^{-1}k(\mathcal{A}, \boldsymbol{\theta}') \tag{41}$$

$$\sigma_{\mathcal{A}}^2(\boldsymbol{\theta}) := k_{\mathcal{A}}(\boldsymbol{\theta}, \boldsymbol{\theta}) \,, \quad \boldsymbol{\theta}, \boldsymbol{\theta}' \in \Theta, \tag{42}$$

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:*

$$\sigma^2_{\mathcal{A}}(\boldsymbol{\theta}) \leq \sigma^2_{\mathcal{B}}(\boldsymbol{\theta}), \quad \forall \boldsymbol{\theta} \in \Theta\,. \tag{43}$$

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 $\boldsymbol{\theta} \in \Theta$:

$$\begin{aligned}\sigma^2_{\mathcal{B}}(\boldsymbol{\theta}) &:= k(\boldsymbol{\theta},\boldsymbol{\theta}) - k(\boldsymbol{\theta},\mathcal{B})(\mathbf{K}(\mathcal{B}) + \eta\mathbf{I})^{-1}k(\mathcal{B},\boldsymbol{\theta}) \\ &= \sigma^2_{\mathcal{A}}(\boldsymbol{\theta}) - k_{\mathcal{A}}(\boldsymbol{\theta},\mathcal{C})(\mathbf{K}_{\mathcal{A}}(\mathcal{C}) + \eta\mathbf{I})^{-1}k_{\mathcal{A}}(\mathcal{C},\boldsymbol{\theta}) \\ &\leq \sigma^2_{\mathcal{A}}(\boldsymbol{\theta})\,,\end{aligned} \tag{44}$$

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:

$$\forall t \geq 0, \quad |\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})| \leq \beta_t(\delta)\sigma_t(\boldsymbol{\theta})\,. \tag{45}$$

With the same probability, we then have that:

$$\begin{aligned}|\mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})]| &\leq \mathbb{E}_{\boldsymbol{\theta}\sim q_t}[|\ell(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta})|] \\ &\leq \beta_t(\delta)\mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\sigma_t(\boldsymbol{\theta})]\,, \\ &\forall \boldsymbol{\theta}\in\Theta\,, \forall t \geq 0\,,\end{aligned} \tag{46}$$

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.[6] 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.

[6]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 $\mathfrak{F}_{t-1}$.

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:

$$r_t \leq 2\beta_{t-1}(\delta)\mathbb{E}_{q_t}[\sigma_{t-1}]\,. \tag{47}$$

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

$$\begin{aligned}\max_{q\in\mathcal{Q}} \mathbb{E}_q[\ell] - \mathbb{D}_{\mathrm{KL}}(q||p) &\leq \max_{q\in\mathcal{Q}} \mathbb{E}_q[u_{t-1}] - \mathbb{D}_{\mathrm{KL}}(q||p) \\ &= \mathbb{E}_{q_t}[u_{t-1}] - \mathbb{D}_{\mathrm{KL}}(q_t||p)\,.\end{aligned} \tag{48}$$

Applying this bound to the definition of instant regret yields the first part of Theorem 2:

$$\begin{aligned}r_t &:= \mathbb{D}_{\mathrm{KL}}(q_t||p_{\mathbf{x}}) - \mathbb{D}_{\mathrm{KL}}(q^*||p_{\mathbf{x}}) \\ &= \max_{q\in\mathcal{Q}} \mathbb{E}_q[\ell] - \mathbb{D}_{\mathrm{KL}}(q||p) - \mathbb{E}_{q_t}[\ell] + \mathbb{D}_{\mathrm{KL}}(q_t||p) \\ &\leq \mathbb{E}_{q_t}[u_{t-1}] - \mathbb{E}_{q_t}[\ell] \\ &= \mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\mu_{t-1}(\boldsymbol{\theta}) + \beta_{t-1}\sigma_{t-1}(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta})] \\ &\leq 2\beta_{t-1}\mathbb{E}_{\boldsymbol{\theta}\sim q_t}[\sigma_{t-1}(\boldsymbol{\theta})]\,,\end{aligned} \tag{49}$$

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

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

$$\begin{aligned}R_T := \sum_{t=1}^T r_t &\leq 2\sum_{t=1}^T \beta_{t-1}\mathbb{E}_{\tilde{\boldsymbol{\theta}}_t\sim q_t}[\sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)] \\ &\leq 2\beta_T \sum_{t=1}^T \mathbb{E}_{\tilde{\boldsymbol{\theta}}_t\sim q_t}[\sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)] \\ &\leq 2\beta_T \mathbb{E}_{\tilde{\boldsymbol{\theta}}_1\sim q_1,\ldots,\tilde{\boldsymbol{\theta}}_T\sim q_T}\left[\sum_{t=1}^T \sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)\right]\,,\end{aligned} \tag{50}$$

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^2_{t-1} is conditioned on all previous observations, which are grouped by batches. We can then decompose, for any $t \geq 1$:

$$\begin{aligned}\sigma^2_t(\boldsymbol{\theta}) &= \sigma^2_{t-1}(\boldsymbol{\theta}) \\ &\quad - k_{t-1}(\boldsymbol{\theta},\mathcal{B}_t)(\mathbf{K}_{t-1}(\mathcal{B}_t) + \eta\mathbf{I})^{-1}k_{t-1}(\mathcal{B}_t,\boldsymbol{\theta})\,,\end{aligned} \tag{51}$$

where we use the notation introduced in Lemma 7, and:

$$\begin{aligned}k_t(\boldsymbol{\theta},\boldsymbol{\theta}') &= k_{t-1}(\boldsymbol{\theta},\boldsymbol{\theta}') \\ &\quad - k_{t-1}(\boldsymbol{\theta},\mathcal{B}_t)(\mathbf{K}_{t-1}(\mathcal{B}_t) + \eta\mathbf{I})^{-1}k_{t-1}(\mathcal{B}_t,\boldsymbol{\theta}')\end{aligned} \tag{52}$$

$$k_0(\boldsymbol{\theta},\boldsymbol{\theta}') := k(\boldsymbol{\theta},\boldsymbol{\theta}')\,. \tag{53}$$

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 6 to bound the last term in Equation 50.

Lemma 7 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 7, we get, for $t \geq 1$:

$$\begin{aligned}\sigma_t^2(\boldsymbol{\theta}) &\leq s_t^2(\boldsymbol{\theta}) \\ &:= k(\boldsymbol{\theta},\boldsymbol{\theta}) - k(\boldsymbol{\theta},\Theta_t)(\mathbf{K}(\Theta_t) + \eta\mathbf{I})^{-1}k(\Theta_t,\boldsymbol{\theta})\,,\end{aligned} \tag{54}$$

where $\Theta_t := \{\boldsymbol{\theta}_{i,1}\}_{i=1}^t$, with $\boldsymbol{\theta}_{i,1} \in \mathcal{B}_i$, $i \in \{1,\dots,t\}$. 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 6, yielding:

$$\begin{aligned}&\mathbb{E}_{\tilde{\boldsymbol{\theta}}_1\sim q_1,\dots,\tilde{\boldsymbol{\theta}}_T\sim q_T}\left[\sum_{t=1}^T \sigma_{t-1}(\tilde{\boldsymbol{\theta}}_t)\right] \\ &\leq \mathbb{E}_{\tilde{\boldsymbol{\theta}}_1\sim q_1,\dots,\tilde{\boldsymbol{\theta}}_T\sim q_T}\left[\sum_{t=1}^T s_{t-1}(\tilde{\boldsymbol{\theta}}_t)\right] \\ &\leq 2\sqrt{(T+2)\gamma_T}\,.\end{aligned} \tag{55}$$

Combining this result with Equation 50, we obtain:

$$R_T \leq 4\beta_T\sqrt{(T+2)\gamma_T} \in \mathcal{O}(\beta_T\sqrt{T\gamma_T})\,. \tag{56}$$

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

$$\beta_T(\delta) := b + \sigma_\epsilon\sqrt{2\eta^{-1}\log(|\mathbf{I}+\eta^{-1}\mathbf{K}_{\mathcal{D}_T}|^{1/2}/\delta)}\,, \tag{57}$$

where:

$$\log(|\mathbf{I}+\eta^{-1}\mathbf{K}_{\mathcal{D}_T}|^{1/2}) = I(\mathbf{z}_{N_T},\mathbf{g}_{N_T}) \leq \gamma_{N_T} = \gamma_{ST}\,, \tag{58}$$

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

$$R_T \in \mathcal{O}(\sqrt{T}(b\sqrt{\gamma_T} + \sqrt{\gamma_T\gamma_{ST}}))\,, \tag{59}$$

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:

$$\operatorname*{argmax}_{q\in\mathcal{Q}} h(q|\mathcal{D}_{t-1}) = \operatorname*{argmin}_{q\in\mathcal{Q}} \mathbb{D}_{\mathrm{KL}}(q||\hat{p}_{t-1})\,, \tag{60}$$

where h is defined as:

$$h(q|\mathcal{D}_{t-1}) := \mathbb{E}_{\boldsymbol{\theta}\sim q}[u_{t-1}(\boldsymbol{\theta})] - \mathbb{D}_{\mathrm{KL}}(q||p)\,, \tag{61}$$

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:

$$\mathbb{E}_p[f] := \int_\Theta f(\boldsymbol{\theta})p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \leq \|f\|_\infty \int_\Theta p(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} = \|f\|_\infty\,. \tag{62}$$

Combining Lemma 3 and Lemma 4, we have that u_t is always bounded, since:

$$\begin{aligned}\forall t \geq 1, \quad u_t(\boldsymbol{\theta}) &:= \mu_t(\boldsymbol{\theta}) + \beta_t\sigma_t(\boldsymbol{\theta}) \\ &\leq \|\mu_t\|_\infty + \beta_t\|\sigma_t\|_\infty < \infty, \quad \forall\boldsymbol{\theta}\in\Theta\,.\end{aligned} \tag{63}$$

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:

$$\begin{aligned}\forall t \geq 0, \quad \zeta_t &:= \mathbb{E}_p[\exp u_t] \\ &= \int_\Theta p(\boldsymbol{\theta})\exp u_t(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta} \leq \|\exp u_t\|_\infty \\ &< \infty\,.\end{aligned} \tag{64}$$

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

$$\begin{aligned}&\operatorname*{argmax}_{q\in\mathcal{Q}} \mathbb{E}_q[u_{t-1}] - \mathbb{D}_{\mathrm{KL}}(q||p) \\ &= \operatorname*{argmax}_{q\in\mathcal{Q}} \mathbb{E}_q[u_{t-1} + \log p - \log q] \\ &= \operatorname*{argmin}_{q\in\mathcal{Q}} \log\mathbb{E}_p[\exp u_{t-1}] - \mathbb{E}_q[u_{t-1} + \log p - \log q] \\ &= \operatorname*{argmin}_{q\in\mathcal{Q}} \int_\Theta q(\boldsymbol{\theta})\log\left(\frac{q(\boldsymbol{\theta})\mathbb{E}_p[\exp u_{t-1}]}{p(\boldsymbol{\theta})\exp u_{t-1}(\boldsymbol{\theta})}\right)\mathrm{d}\boldsymbol{\theta} \\ &= \operatorname*{argmin}_{q\in\mathcal{Q}} \mathbb{D}_{\mathrm{KL}}(q||\hat{p}_{t-1})\,,\end{aligned} \tag{65}$$

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:

$$\min_{t\leq T} r_t \leq \frac{1}{T}R_T \in \mathcal{O}(\gamma_{ST}T^{-1/2})\,. \tag{66}$$

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

$$\frac{1}{T} R_T \in \mathcal{O}(T^{\alpha - 1/2}) \implies \lim_{T\to\infty} \min_{t \leq T} r_t \leq \lim_{T\to\infty} \frac{1}{T} R_T = 0 \,. \tag{67}$$

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:

$$\begin{aligned}
\mathbb{D}_{\mathrm{KL}}(\hat{p}_t || p_{\mathbf{x}}) &= \mathbb{E}_{\boldsymbol{\theta}\sim\hat{p}_t}[\log \hat{p}_t(\boldsymbol{\theta}) - \log p_{\mathbf{x}}(\boldsymbol{\theta})] \\
&= \mathbb{E}_{\boldsymbol{\theta}\sim\hat{p}_t}[\log p(\boldsymbol{\theta}) - \mu_t(\boldsymbol{\theta}) + \beta_t \sigma_t(\boldsymbol{\theta}) \\
&\quad - \log \zeta_t - \log p_{\mathbf{x}}(\boldsymbol{\theta})] \\
&= \mathbb{E}_{\boldsymbol{\theta}\sim\hat{p}_t}[\mu_t(\boldsymbol{\theta}) + \beta_t \sigma_t(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta})] + \log p(\mathbf{x}) \\
&\quad - \log \zeta_t \\
&\leq 2\beta_t \mathbb{E}_{\boldsymbol{\theta}\sim\hat{p}_t}[\sigma_t(\boldsymbol{\theta})] + \log p(\mathbf{x}) - \log \zeta_t \,.
\end{aligned} \tag{68}$$

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

$$p(\mathbf{x}) = \int_{\Theta} p(\boldsymbol{\theta}) \exp \ell(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \leq \int_{\Theta} p(\boldsymbol{\theta}) \exp u_t(\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} = \zeta_t \,, \tag{69}$$

which holds with the same probability as Equation 68. Therefore, we conclude the proof with:

$$\begin{aligned}
&\log p(\mathbf{x}) - \log \zeta_t \leq 0 \\
&\implies \mathbb{D}_{\mathrm{KL}}(\hat{p}_t || p_{\mathbf{x}}) \leq 2\beta_t \mathbb{E}_{\boldsymbol{\theta}\sim\hat{p}_t}[\sigma_t(\boldsymbol{\theta})] \,.
\end{aligned} \tag{70}$$

□

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:

$$\min_{t \leq T} \mathbb{D}_{\mathrm{KL}}(\hat{p}_t || p_{\mathbf{x}}) \leq \frac{R_T}{T} \,, \quad T \geq 1 \,. \tag{71}$$

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).

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

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

$$k(\boldsymbol{\theta}, \boldsymbol{\theta}') := \frac{1}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{\sqrt{2\nu}\rho_{\boldsymbol{\theta},\boldsymbol{\theta}'}}{l} \right)^{\nu} B_{\nu} \left(\frac{\sqrt{2\nu}\rho_{\boldsymbol{\theta},\boldsymbol{\theta}'}}{l} \right) , \tag{72}$$

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 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 5 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.

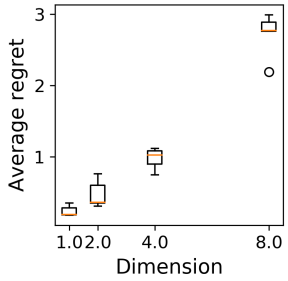


Figure 5: Dimensionality effect on the mean regret of KL-UCB on the RKHS log-likelihood problem after 20 iterations as a function of the dimensionality of the parameter space. The results of 5 independent runs were combined to produce the box plot.

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.[7]

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 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[8] 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)[9] to obtain the KL divergence estimates.

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

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

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